Eurographics 2007
September 3-7, 2007
Prague, Czech Republic

Tutorial Notes

Volume 2

Content: T3, T4, T7, T8, T11, T12

Tuesday, 4th September 2007

Karol Myszkowski and Vlastimil Havran
Tutorial Chairs, Eurographics 2007

Published by
The Eurographics Association
ISSN 1017-4656

©2007	The Eurographics Association
ISSN	1017-4656
Produced by:	Computer Graphics Group Department of Computer Science and Engineering Faculty of Electrotechnical Engineering Czech Technical University in Prague Karlovo nám. 13, 12135 Prague 2, Czech Republic
Print Preparation:	Vlastimil Havran
Cover Design:	Diana Delévová
Printed by:	APOLYS, Kolbenova 609/40, Prague 9

Preface

Over the years, tutorials have built a strong reputation of being an important part of the Eurographics conference. In these tutorials, conceptual and implementation aspects of recent techniques are analyzed in depth by leading experts in the field.

This year the tutorial program is composed of 12 half-day tutorials, which span a wide range of topics including: shape modeling, physically-based modeling and simulation, haptic interfaces, virtual crowds, 3D video, reflectance acquisition, computational photography, rendering, high-performance parallel computing, and visual data mining. While the selected topics are very broad, several tutorials are provided in the context of virtual human modeling. Another group of tutorials present new trends in imaging and rendering. Thus, we informally split the tutorial program into three tracks: Virtual Humanoid, Imaging and Rendering, and Miscellaneous.

We are grateful to everyone who has submitted a tutorial proposal to EG'07. We also would like to thank all reviewers for their valuable comments and expertise, which greatly helped in the tutorial selection process. We want to express our gratitude to the members of the technical team Stefanie Behnke and René Berndt for their continuous support with the on-line submission system. Finally, we are indebted to Jiří Žára, EG'07 co-chair, for his continuous support and advice.

Karol Myszkowski and Vlastimil Havran

August 2007

List of Reviewers

Aubrecht, Petr
Baciu, George
Barbagli, Federico
Becks, Andreas
Belyaev, Alexander
Benthin, Carsten
Berka, Roman
Čadík, Martin
Cheung, German
Chrysanthou, Yiorgos
Decoret, Xavier
Gagalowicz, Andre
Goesele, Michael
Gumhold, Stefan
HO, Edmond S. L.
Klatzky, Roberta
Klíma, Martin
Kordík, Pavel
Lensch, Hendrik
Magnor, Marcus
Matusik, Wojciech
Mikovec, Zdenek
Orkisz, Maciej
Otaduy, Miguel A.
Reiterer, Harald
Scholz, Volker
Slater, Mel
Smolic, Aljoscha
Tecchia, Franco
Teschner, Matthias
Theisel, Holger
Ward, Greg
Žára, Jiří
Ziegler, Gernot

Table of Contents - Volume 2

Table of Contents - Volume 1

EG:593

Towards the Virtual Physiological Human

Eurographics 2007 Tutorial T3

Organizers

Nadia Magnenat-Thalmann (MIRALab, University of Geneva)

Speakers

Nadia Magnenat-Thalmann (MIRALab/University of Geneva)
Benjamin Gilles (MIRALab/University of Geneva)
Hervé Delingette (INRIA, Asclepios)
Andrea Giachetti (CRS4, Visual computing group)
Marco Agus (CRS4, Visual computing group)

EG:594

Towards the virtual physiological human

N. Magnenat-Thalmann [1], B. Gilles[1], H. Delingette[2], A. Giachetti[3] and M. Agus[3]

[1] MIRALab - University of Geneva, Switzerland
[2] INRIA - Asclepios, France
[3] CRS4 - Visual computing group, Italy

Abstract

The objective of this tutorial is to train students and researchers in the various domains involving the modelling and simulation of the human body for medical purposes. Human body representations have been used for centuries to help in understanding and documenting the shape and function of its compounding parts. Nowadays, medical acquisition devices especially medical scanners are able to produce a large amount of information, such as high-resolution volumes, temporal sequences or functional images, more-and-more difficult to analyse and visualise. In other words, we measure much more than we understand. In this context, higher-level information such as anatomical and functional models is increasingly required to support diagnosis and treatment. Three levels of complexity can be distinguished (geometry, function and control) where modelling and simulation methods take place. In this tutorial, we will present the current research issues towards the patient-specific reconstruction of virtual models and their functional simulation. We will focus on the computer graphics aspects involved in medical modelling/ simulation: deformable model-based segmentation, mesh optimisation, data fusion, interactive mechanical simulation, cost-efficient visualisation, etc. Examples will be given in the musculoskeletal, cardiac and neurological domains. All speakers are partners of the EU project "3D Anatomical Human" led by MIRALab - University of Geneva.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Curve, surface, solid, and object representations. I.3.5 [Computer Graphics]: Physically based modelling. I.3.3 [Computer Graphics]: Viewing algorithms. I.3.8 [Computer Graphics.]: ApplicationsComputer-aided diagnosis and treatment.

1. Anatomical and kinematical reconstruction from medical data

1.1. Introduction

Nowadays, medical imaging devices are able to produce a large amount of information, such as high-resolution volumes and temporal sequences, more-and-more difficult to analyse and visualise. In this context, higher-level information such as 3D or 4D models is increasingly required to support medical diagnosis. Such models can have predictive capabilities (for computer-aided treatment, prevention, rehabilitation and ergonomics). For simulation, the use of medical images is important, as they allow the production of patient-specific models (anatomical models), expressing subtle inter-subject variability (possible cause of pathology), and the validation of mechanical prediction through temporal studies (kinematical models).

1.2. Acquisition modalities

When modelling and simulating real objects, data acquisition is essential. It aims at extracting model parameters and validating simulation output. Data can be divided into geometric data (e.g. medical images), kinematical data (e.g. optical motion capture data), mechanical parameters (e.g. strain-stress curves) and physiological data (e.g. muscle activation). Due to measurement difficulties (for instance, the access to bone kinematics, internal muscle activation patterns) and musculoskeletal system complexity, it is common to make assumptions about this data and use prior knowledge. Relying on these, computer models need to be validated correctly. Obviously, geometric models (shape) deal with few assumptions, whereas functional models require a lot of prior information for their simulation. In this section, we focuss on musculoskeletal macroscopic acquisition,

which is more relevant for the computer graphics community.

1.2.1. Static acquisition

1.2.1.1. Cadaveric studies: Until XX century, cadaver analysis has been the only one way for assessing organ shapes and interrelationships (e.g. Gray's anatomy [Gra00]). The visible human [NLM99], a mixture of medical images and histological cuts of the same individual, is up to now the most complete dataset for anatomical description. It has supported many projects in the fields of medical image processing and biomechanics. The measurement of submillimeter anatomical features such as in synovial joints still requires specimens. However, medical imaging modalities, more and more accurate and standardised, tend to supplant these kind of studies.

1.2.1.2. Radiography and Computed Tomography (CT): Historically, the radiography has been the first modality for in-vivo anatomical imaging of internal structure (Roentgen, end of XIX century). It is based on the photo measurement of the intensity of X-rays traversing the body. Sensors are generally phosphor plates, that can be subsequently digitalised (Computed Radiography - CR), but new digital sensors are emerging (Digital Radiography - DR), though they are still expensive. The resulting 2D projections highlight anatomical structures, absorbing X-rays differently. Combining several projections (Computed Tomography - CT), volumes (typically 512x512x128 voxels) can be reconstructed with a resolution of about 0.5x0.5x1 mm in 30s. CT is a tomographic modality, meaning that output images, reconstructed from several projections, are cross-sections. The intensity in images is measured towards the Hounsfield scale, leading to a simple segmentation of tissues. However, different tissues can share the same value, and artefacts disturb this identification. It is known that CT is more dedicated to chest and bone investigations (about 70% of all imaging examinations), as bone-to-soft tissue contrast is high. Thresholding, few regional constraints and simple post-processing techniques (e.g. morphological operations) are generally sufficient to segment bones, or structures highlighted with suited contrast agents (e.g. angiography). Direct volume rendering is also popular in the clinical environment [HPP*03]. In orthopaedics, where most of pathologies rely on bones (e.g. fracture, cartilage ossification), CT is the privileged modality, although it is invasive (high dose X-ray absorption can damage cells and cause cancer).

1.2.1.3. Ultra-sound (US): Ultra-sounds propagate in homogeneous tissues but are partly reflected at tissue interfaces. The principle of US is to measure this reflection in time, the emission/ reception time interval being proportional to the depth of the interface (assuming few wave velocity variations). US scanners generally operate in brightness mode (B-mode), that is grey-scale in images represent an amplitude map of the reflected sound. 2D cross-sections and 3D volumes can be interpolated by casting rays in different directions (tomography). Either mechanical and free-hand probes are available [FDC01], allowing for some of them position tracking. General interpolation techniques may be used to resample unparallel slices. US is low-cost, fast and non-invasive. Unfortunately, image quality is low: speckle is present due to wave interference, geometric distortions occur due to wave velocity changes and ultra-sound can be totally absorbed by some tissues (e.g. air, bones) and at large depth, leading to occlusions. US is typically used for heart and foetus dynamic imaging, but is also adequate for superficial musculoskeletal imaging. Particularly, muscle fibers direction can be visualised in-vivo.

Figure 1: *CT imaging of the hip*

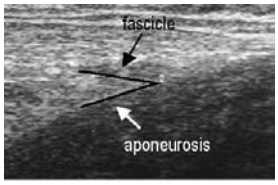


Figure 2: *Ultrasound assessment of the pennation angle of the gastrocnemius muscle in a relaxed state, from University of California, Davis*

1.2.1.4. Magnetic Resonance Imaging (MRI): Atomic nuclei have a magnetic moment that can be oriented by an external magnetic field. This moment rotates about the axis of the field, with a frequency depending on the type of nucleus and on the strength of the magnetic field. In this condition, atoms can briefly jump from one equilibrium state (spin $+1/2$) to another one (spin $-1/2$), under radio impulses emitted at this frequency (resonance). When returning to the initial state, atoms generate a damped radio frequency signal (relaxation). MRI aims at measuring this signal from the resonance of hydrogen atoms, contained in all tissues (water). Indeed, tissues react differently, depending on the concentration of water (proton density), but also on the way that protons revert back to their resting states after the initial RF pulse (relaxation time). Relaxation time is measured

in two directions: T1 (longitudinal) and T2 (transversal). In addition, fluid flow may decrease the signal. These four tissue parameters: proton density, T1 relaxation time, T2 relaxation time and fluid flow; determine the signal intensity on the MR images. By weighting these parameters, MRI is capable of producing various types of contrast, conferring to this modality a great versatility. The self-diffusion of water is restricted by membranes, and consequently relies on tissue geometry (e.g. fiber orientation). By applying different gradient directions, fluid flow/ fiber direction can be measured (six directions: diffusion tensor MRI; three directions: diffusion weighted MRI). Local, voxel-based directions can be combined with a fiber-tracking algorithm (tractography) to reconstruct complete paths [MZ02]. It is admitted that MRI does not produce any harmful ionizing radiation [SC04]. Typical resolution is about $1mm^3$. Strong magnets ($>$1 Tesla), which are horizontal tubes of about 60cm diameter, lead to a higher signal-to-noise ratio (SNR), and consequently to a higher resolution and shorter scan times than feeble magnets. In research, the spatial resolution can attain $1\mu m^3$, using powerful magnets (e.g. 4 Tesla, 7 Tesla).

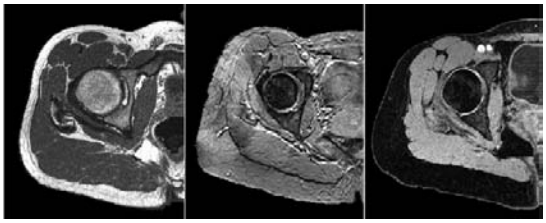


Figure 3: *Different MRI sequences for imaging the hip: a) Spin-Echo T1 b) Gradient-Echo T2* c) Gradient-Echo T1*

1.2.1.5. Body scanning: For computer graphics and animation purposes, accurate skin models of the complete body are generally digitalised using commercial optical scanners. The reflection of white light or eye-safe laser is measured by mobile cameras, providing a point cloud of the scene that need to be subsequently triangulated (resolution: $\simeq 1mm$, acquisition time: $\simeq 30s$). Another approach uses stereo photogrammetry from at least two camera views. This technique is faster ($\simeq 1.5ms$) and more accurate ($0.5mm$) than laser scanning when using efficient mesh reconstruction techniques, however this has not been applied to large scale (e.g. full body) yet.

1.2.2. Kinematics and dynamics acquisition

1.2.2.1. Ultra-sound (US): US is a fast imaging modality that can provide useful kinematic information, with a frequency up to 100 frames/s (e.g. foetal echography). 1D temporal analysis (x axis: time; y axis: depth), called M-mode, is useful for visualising highly sampled (about 1000 pulses/sec) superficial muscle contraction patterns or heart valves movements. Direct tissue/ liquid velocity visualisation is also possible with US, exploiting the Doppler effect. Colour Doppler aims at visualising flow direction and velocity, whereas color power Doppler (CPD) only deals with velocity amplitude.

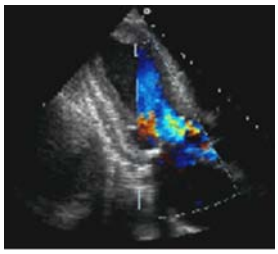


Figure 4: *B-mode ultrasound slice with superimposed colour Doppler showing blood flow inside the heart (from AnaesthesiaUK)*

1.2.2.2. Fluoroscopy and Computed Tomography (CT): Fluoroscopy produces real-time images of internal structures in a similar fashion to radiography. X-ray fluoroscopy is faster than CT, as it performs only one projection from a continuous input of X-rays, while CT requires multiple projections to reconstruct cross-sections. CT volumes are converted into 2D projections using digitally radiographs reconstruction (DRR) techniques. Fluoroscopy has been widely used for intra-operative bone tracking and biomechanical studies [YSAT01]. Implanted pins are used for accurate bone localisation, but non-invasive bone motion tracking can be obtained through 2D/ 3D registration between fluoroscopic images and CT data. New technical advances in CT scanner including large 2D detectors, fast rotating devices and fast systems for transmission and reconstruction, have recently permitted 4D CT acquisition (e.g. 256-slices, field of view along $z = 12cm$, resolution $= 1mm \times 1mm \times 0.5mm \times 1sec$ [MET*04]).

1.2.2.3. Magnetic Resonance Imaging (MRI): Several MRI techniques has been made available for dynamic tissue analysis, through the decrease of MRI acquisition time. Particularly, joint kinematic studies with MRI have been reported. The passive sequential (or stepped) acquisition technique (i.e. loaded stationary joint during acquisition) is well-established but has limited use. However, the problem of acquiring volumetric image data in real-time with MRI during active motion remains to be solved due to inherent trade-off in the MR imaging technique between Signal-to-Noise Ratio (SNR), spatial resolution and temporal resolution. Specific imaging sequence (trueFISP: $6frames/s$ [QLH*02]) and parallel imaging techniques exploiting spatiotemporal correlations with reference images (e.g. SENSE, Philips Medical Systems, $38frames/s$ [TBP03]) can greatly speed up the acquisition. Motion periodicity can be exploited by combining information across several cycles (*cine-MRI*), which

is particularly suited for cardiac imaging. Because tissues keep their magnetic orientation for a time (∼T1 time), it is possible to create temporary magnetic fiducial markers, or tags, and hence trace tissue deformations (*tagged MRI*) [ZPR*88]. Velocity imaging is possible using phase contrast sequences (*pcMRI*) [PHSE91]: two datasets are acquired, one being flow sensitive and the other flow compensated. The raw data (phase information) from the two data sets are then subtracted. pcMRI can be coupled with cine-MRI as in [DP94].

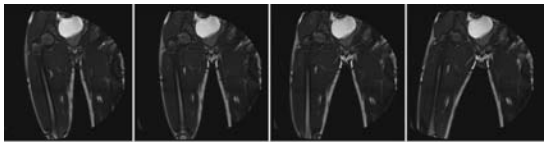


Figure 5: *Real-time dynamic MRI of the thigh [GPMTV04]*

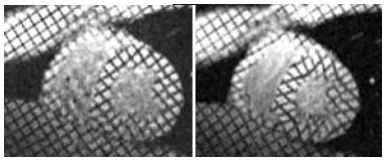


Figure 6: *Cine tagged-MRI of the heart*

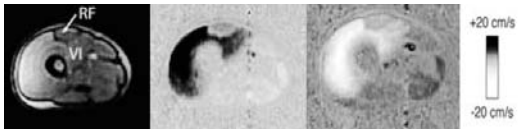


Figure 7: *Cine phase-Contrast MRI of the thigh during knee flexion/extension [ABGD02]*

1.2.2.4. Motion Capture (MoCap): While medical imaging provides precise but local kinematic measurements, motion capture (MoCap) techniques aim at acquiring global skin motion (larger scales and ranges of motion). Avoiding the burden inherent to medical scanners, MoCap provides in-situ kinematical data (e.g. gait), and is compatible to other modalities for force (e.g. force plates) and physiological (e.g. EMG) measurement. Motion capture of the complete body is achieved through the recording of landmarks attached to the skin. In a post-processing phase, landmark positions are reconstructed in 3D, and rigid motions of the different body segments can be established, approximation skeletal movements and joint centers [LO99]. Three main motion capture systems can be distinguished [Men00]: the ones based on electromagnetic trackers (*flock-of-bird*) ($\sim 140Hz$, $\sim 2mm$); the ones based on mobile electromechanical devices such as inclinometers, goniometers, accelerometers and gyroscopes ($\sim 90Hz$, $\sim 0.5deg/2mm$); and the most used ones based on reflective optical markers ($< 1000Hz$, ∼0.5mm). With optical systems, a post-processing (interactive) phase is necessary to manage occlusions. As the relative skin/ bone motion introduces errors in joint center and skeletal kinematics computation, implantable pins and error minimisation techniques have been used for biomechanical studies.

1.2.2.5. Video-based methods: Instead of MoCap, raw video images or laser scans can be used to assess body kinematics. Whereas MoCap provides accurate and high-rate motion of the skin at sample points, it can be useful to assess the skin with more complete spatial information. However, video-based methods have a lower accuracy and frame-rate, and require efficient post-processing methods (computer vision techniques). In [DHW*04], a dynamic body scanner system ($15Hz$) is used for spine dynamic analysis where internal anatomical structure are estimated for the shape of the skin. Commercial products using stereo photogrammetry can achieve around $60Hz$ for imaging body parts. In [ACP02], skin animation is performed from scattered scan data, trough interpolation techniques.

1.2.2.6. Force plates: For dynamic simulations of the musculoskeletal system, it is necessary to evaluate external forces acting on the body, such as ground reaction forces. MoCap, EMG and force plates are typically associated. Force plates can record ground reaction forces (6 DOF) [DP02]. Alternatively, one can use pressure measurement soles [WC03], leading to slightly different results (shoe/ground vs. foot/shoe interaction). Both systems can record at approximately $500Hz$, with a resolution of about $1cm$.

1.2.3. Physiology acquisition

1.2.3.1. Electromyography (EMG): The measurement of muscle stimulation patterns is essential to understand musculoskeletal dynamics and control, and consequently the causes of abnormal movements. Electromyography (EMG) consists in assessing electrical signals ($0-6mV$, $10-500Hz$) being generated by a muscle at the same time that it contracts [Cut93]. Surface EMG (sEMG), where electrodes are placed at the skin above muscles, is a non-invasive approach (contrary to needles) for muscle stimulation recording. It has been used in most studies regarding the forward dynamic simulation of the musculoskeletal system, the validation of inverse dynamic simulation, and the diagnosis of neuro-musculoskeletal disabilities [DL95]. Due to differences in electrode properties, configuration and location, which is problematic for the use and exchange of sEMG data, standard protocols have been recently established [HFDKR00]. Indeed, a good electrode/skin contact need to be insured to reduce noise, and a precise electrode placement to minimise crosstalk from neighbouring muscles. To tackle this, researchers have investigated

more sophisticated electrodes such as bipolar electrodes and more recently electrode arrays. EMG data need to be post-processed to extract relevant information and to allow intra and inter subject comparison [DeL97]. Because there is no absolute signal value, EMG are generally normalised according to the maximum voluntary contraction or MVC (highest peak during maximum effort). In addition, EMG analysis in the Fourier domain is particularly useful to separate the contribution from the different muscles, and analyse physical fatigue. Indeed, fatigue is characterised by an average frequency shift towards lower frequencies.

1.2.3.2. Mechanomyography (MMG): MMG is a non-invasive technique for measuring muscle contraction speed of muscle groups, and is the mechanical equivalent to EMG (electrical activity). Its principle is to measure skin displacement under neuromuscular stimulation. Measurements are commonly made through a variety of transducers (condenser microphones, piezoelectric contact sensors or accelerometers) leading to various output names: muscle sound, soundmyogram, acousticmyogram, phonomyogram, accelerometermyogram or vibromyogram.

1.2.4. Mechanical testing

1.2.4.1. Mechanical devices: Biological tissues mechanical parameters can be established from various modalities measuring the uniaxial/biaxial tissue strain response in relation to mechanical or ultrasound impulses. Indeed, the relationship between the strain (tension, compression, bending or torsion) and the stress (pressure or force) provides the elastic property of the material, while viscoelasticity can be analysed from stress and strain rates (stress-relaxation and creep testing) over multiple loading/unloading cycles. The biaxial stress-strain relation of a soft tissue can not be obtained by superimposing two individual uniaxial test results due to the large deformation, but data regression methods are usually performed to fit the constitutive model to data. Commonly, tissue specimens need preconditioning (several loading/unloading cycles) before obtaining reproducible results. Then, parameters are established from the loading and unloading curves. The use of dead specimens is problematic, as they may behave differently than living tissues: they need to be correctly hydrated (e.g. saline baths) and kept at a suitable temperature. In addition, tissue initial milieu is difficult to reproduce (e.g. pressurised synovial joints). In-vivo mechanical measurement of internal tissues poses extreme experimental difficulties, but is becoming possible through the development of miniaturised devices. Buckle transducers, telemetered pressure sensors and strain gauges have been investigated for in-vivo force, pressure and deformation measurement. Non-invasive approaches are also available for isotropic materials such as aspiration tools [NMK*03] or torsional resonator devices [VM04]. Highly anisotropic materials (e.g. ligaments) are, however, difficult to measure in situ. Implantable force transducers and video-based measurement techniques have been used, but the reference state (zero stress) remains problematic to determine [WG01].

1.2.4.2. Magnetic Resonance Elastography (MRE): MRE is a new technique to measure material internal stiffness in vivo and non-invasively [MRL*96]. It is based on the measurement of tissue response under shear waves, propagated from a transducer attached to the skin surface. The stiffer is the tissue, the more rapidly waves are propagated (longer wavelength). Wave displacement velocity is measured with Phase-Contrast MRI (pcMRI), whereas in ultrasound electrography, the measurement is based on Doppler techniques. MRE has been applied to measure muscle actuation, exploiting the fact that the more contracted is a muscle, the stiffer it is. In [PBHS05] muscle anisotropy (3 parameters) has been measured with MRE, based on a transversely isotropic linear material. MRE is a very promising approach allowing tension distribution within a muscle to be imaged. In addition, contrary to EMG, MRE can provide information about passive properties of muscle.

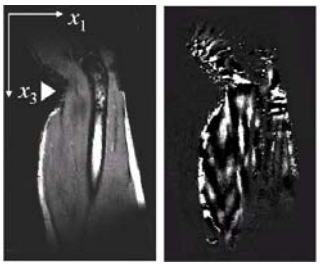


Figure 8: *MRE measurement based on a harmonic shear excitation in x1-direction, from [PBHS05]*

1.3. Image segmentation and registration

1.3.1. Introduction

Anatomical and kinematical modelling deal with image segmentation, that is the identification, in 2D, 3D or 4D images, of the region or contour of interest. Segmentation has been an intensive research field for many years, and is generally associated to registration (the search of spatial correspondences across datasets). Indeed, registration can indirectly perform segmentation as soon as a generic/ reference dataset has been segmented (the generic model is iteratively deformed/ registered to images). On the contrary, direct segmentation can be achieved with few assumptions about the problem (organ shape, image intensities, deformation, etc.). Direct approaches such as thresholding, region growing or edge detection are generally ad-hoc techniques that fail to handle noisy images with a large amount of textural information such as medical images, while indirect approaches use prior information (reference segmented datasets, generic models, anatomical and statistical information, texture database) to perform the segmentation.

Given ordinary clinical images (low-resolution images with anisotropic voxels, noise and partial volumes), the output of segmentation (binary volume), even manually obtained, is a more-or-less correct representation of organ shapes. A post-processing step is necessary to reconstruct a model, from the binary map, that can be simulated and efficiently visualised. The Marching Cubes algorithm, originally proposed by Lorensen et al. [LC87], is considered to be a standard approach to the problem of extracting isosurfaces from a volumetric dataset. However it produces aliased meshes, with a poor geometrical quality (irregular triangles) and an excessive number of triangles. Constrained reconstruction techniques (such as [Del99]) where surface smoothness and mesh topology can be controlled are more efficient. In this context, deformable models are nowadays a widely spread approach for segmentation, since they allow geometrically constrained image segmentation (no post-processing).

Figure 9: *marching cubes reconstruction of the knee (left) and constrained reconstruction using deformable models (right)*

1.3.2. Direct segmentation

Direct segmentation of anatomical structures in medical images is particularly challenging because medical images are noisy (e.g. speckle in ultrasound) and generally complex from a textural point of view. Moreover, the ratio between image resolution and the size of structures to segment can be high. Low resolution images produce partial volume effects, making frontiers between organs unclear. Direct segmentation makes very little assumption about organ shape, organ size or image intensity, so is very generic. For complex problems, direct segmentation is noise-sensitive, not robust and quite inaccurate. However it can help in speeding up manual segmentation and enhancing/ extracting image features for further algorithms (e.g. registration).

Generally speaking, direct segmentation is an iterative process involving a "detection" step where regions are identified in images and a *classification* step where regions are combined to create new regions. A pre-processing step may initially filter images to normalise them, reduce artefacts and noise and enhance specific features. After segmentation, the resulting atlas can be converted into surfacic or volumetric model through reconstruction techniques.

1.3.2.1. Pre-processing: Well-known filtering algorithm may be applied to images in a pre-processing step. Noise can be attenuated through low pass filtering. Anisotropic filtering [PM90] is particularly suited, as it does not alter edge sharpness. When segmenting several datasets from different patients or sessions, intensity normalisation may be applied to get more robust results (because most algorithms rely on image intensity). This is frequent to fit image contrast and brightness to a reference image. Normalisation is generally performed by an (piece-wise) affine transform of the intensities. Between cross-sections, intensity bias can be present in medical volumes and attenuated through intensity normalisation across slices.

1.3.2.2. Region detection: Regions in images can be defined by their internal voxels or by their boundaries. Similarly, segmentation methods aim at discriminating regions, looking at image intensity levels and their derivatives. Another way is to assess the probability of voxels to belong to a region, considering local intensity values.

- **Edge-based approaches:** Organ interfaces are characterised by a more or less steep intensity level variation in images and are quantified through gradient amplitude and orientation. Various convolution filters have been proposed for gradient measurement. Among them, 1st order derivative filters such as the Sobel or Prewitt filter, and 2nd order derivative filters such as the Laplacian filter are particularly used.
- **Region-based approaches:** When regions are more or less homogeneous, voxels and their neighbours look similar, except at boundaries. Voxel intensities, standard deviations or gradients are computed and compared to the ones of neighbouring voxels (classification step) in an ascending or descending fashion (e.g. split and merge algorithm). When regions are not homogeneous, but present repetitive patterns, texture analysis may be performed. It is based on local parameters (e.g. frequency or statistics), extracted within a specific window. Alternatively, cross-correlation with prior texture primitives may be performed. An example of the use of statistical parameters is the cooccurence method.
- **Statistical approaches:** Bayesian approaches try to maximise the knowledge of the scene (classification) given the image, through its probability density function $p(S|I)$. According to the Bayes rule: $p(S|I) = p(I|S)p(S)/p(I)$, this is an optimisation process involving the likelihood (probability of the image given the scene) and the prior (probability of the scene). It leads to $\tilde{S} = argmax(log(p(S)) + log(p(I|S)))$ (Maximum A Posteriori or MAP solution). For expressing the likelihood, additive measurement noise (usually Gaussian) is considered as well as a signal model (such as the Gibbs distribution, considering that the signal is a Markov field). It is also common to incorporate a model of spatial interaction to account for signal continuity in the image. It is a very general framework that can be used also in deformable segmentation and registration [SD92], as we will see later.

1.3.2.3. Classification: Detected regions, or voxels are classified into classes (e.g. tissue type) using homogeneity criterions. When using one parameter, thresholding is performed at a specific level. The criterion is generally image intensity level or probability density. The threshold may be automatically computed through histogram analysis (e.g. Gaussian modelling of grey level distribution). Alternatively, classification can be performed through multiple parameters (multidimensional classification), such as the region mean level, the standard deviation, the contour mean curvature or the object principal axis. Unsupervised classification tries to delineate automatically the different classes in the parameter space using proximity criterions, whereas supervised classification makes use of prior knowledge about parameter values of the different classes. Two strategies are possible regarding the detection/ classification loop. The ascending strategy (split algorithm) starts from under-segmented regions and make them grow iteratively by detecting similar neighbouring regions. In the descending strategy (merge algorithm) the image is split into many regions (over-segmentation) which are subsequently merged through homogeneity/ proximity evaluation and decision making (e.g. quadtree partition). The two may be mixed (split and merge algorithm) [HP78].

1.3.3. Image Registration

1.3.3.1. Introduction Image registration, that is the computation of spatial correspondences between images, is one of the main fields in image processing. Indeed, registration is required in various domains such as image fusion (multimodal registration), segmentation (template registration), motion tracking (temporal registration) and navigation (viewpoint registration). Image registration aims at finding the displacement field that maximises the similarity between images, while preserving a certain regularity of the displacement. Hence, it can be seen as an energy optimisation problem, where the external energy represents the attraction potential field from images, and the internal energy, the regularity of the displacement. Alternatively, image registration can be viewed as a force (energy derivative) equilibrium process where external and internal forces self-balance. The diversity of problems, approaches and comparisons has made any attempt to exhaustively classify registration methods difficult. The various existing reviews [Bro92] [MV98b] [FHM00] have focused on different more-or-less independent criterions such as:

- **The nature of features:** Features, also called primitives or landmarks, are the information, either geometric or iconic, extracted from images to be registered.
- **The similarity measure:** this is the criterion upon which the registration is performed.
- **The problem regularity:** the type of transformation that we are looking for (e.g. rigid, elastic, etc.)
- **The resolution method:** how the optimal transformation, maximising the similarity between features, is found.

1.3.3.2. Registration features: We commonly distinguish geometric features (shape) and icons (image), but it is possible to combine them.

Iconic features: Icons rely on the (pre-processed) photometric information of a voxel (e.g. optical flow [HS81]) or a region (e.g. template, intensity profile [MD05]). With iconic registration, there is no pre-segmentation or geometric feature extraction. However, images may be pre-processed to remove noise and bias, and normalise intensities. Iconic registration has shown to be very efficient in multimodal registration (higher relevance of intensity distribution correlation over extracted features matching).

Geometric features and models: There are two main approaches when using geometric features for image registration. On one hand, they can be segmented separately in the two datasets and then, registered geometrically (using geometrical distance as the similarity measure) [AFP00]. On the other hand, they can be extracted from the source dataset, and subsequently used to segment the target one. With this approach, geometric features (the model) are rigidly or elastically matched, in an iterative process. With deformable models, segmentation, registration and reconstruction are performed at the same time, leading to a possible combination between iconic (image-based) and geometric registration. Moreover, they allow spatially coherent regularisation mechanisms.

Geometric features of lowest dimension are **points**. We differentiate artificial (or extrinsic) and anatomical (or intrinsic) fiducial points [FHM00]. Artificial markers are fixed invasively (e.g. bone pins) or non-invasively (e.g. skin markers) to organs for inter-patient image registration, intra-patient image registration or motion tracking in image sequences. Their discrete locations can be extracted from marker regions using image processing methods or directly registered through iconic registration (block matching). Intrinsic landmarks rely on salient anatomical features. They can be extracted from images, but robust point extraction (automatic or manual) is often a difficult task [RFS99]. Point registration can lead to rigid or elastic registration: from known spatial correspondences at discrete locations, it is possible to estimate the global rigid movement or the surrounding elastic displacements using error minimisation and adequate interpolation methods such as radial basis functions (e.g. thin plate splines). Interpolation methods are usually not realistic at locations distant from landmarks. So that accurate point-based registration is generally limited to rigid registration (tracking applications). For rigid registration, a **coordinate system** can be used instead of a set of landmarks. This can be obtained through gravity center, principal axis and moment calculation from grey-scale or binary images, but also from organ shape (intrinsic anatomical axes). To overcome unrealistic point-based elastic registration, landmarks orientation and anisotropic error computation have been incorporated in the classical Thin-Plate-Spline (TPS)

interpolation method [RFS99]. This provides better results, especially at organ interfaces where the registration transformation is not continuous.

Higher dimensional geometric features, such as curves, surfaces and volumes, have various mathematical representations. The type of representation is a key-point with deformable models, as it influences geometric descriptors, the evolution process and the visualisation methods associated to them. McInerney et al. [MT96], Montagnat et al. [MD05], Singh et al. [SGT98] present detailed reviews and classifications of deformable models. *Curves*, such as crest lines, are perceptive features that may be used for registration; however, *surfaces* remain the most used geometric primitives, since boundary is commonly the most noticeable object feature. In images, homogeneous objects, are, in fact, only characterised by their shape. We broadly present the various representations used in image registration, and to a larger extent in tissue modelling and simulation.

There are three distinct families of deformable models: **continuous models** are definite through the mapping of parameter ranges in space; **discrete models** are defined at discrete locations in space and connectivity relationships; and **implicit models** are indirectly defined by function isovalues. Continuous models offer the ability to compute differential quantities (e.g. normals, curvatures) easily and are well controllable through their limited number of parameters (intrinsic geometric regularisation) but their shape is limited by the parameterisation. On the contrary, discrete models are more flexible but difficult to constrain geometrically. Implicit models can benefit from their volumetric nature, and merging capabilities, but their abstract nature makes spatial interactions with them, difficult.

The **continuous** representation has been the first proposed one, for deformable models. In 1988, Kass et al. [KWT88] propose the well-known *snake* representation (also called active contour, or explicit contour). A snake is an explicit mapping of a real parameter to 2D coordinates. This has been generalised to 3D case by Terzopoulos et al. [TF88b] and adapted by McInerney et al. to handle topology change [MT99] (T-snake). Parametric models are continuous models where the mapping is made through particular mathematical functions, defined by few parameters (degrees of freedom of the model). Continuous models are popular in medical image analysis since their regularities are simply evaluated; examining mapping function derivatives with regards to material coordinates. Moreover, parametric models are intrinsically regularised thanks to their limited number of parameters (DOFs). B-splines, (bicubic) Hermite surfaces, and superquadrics are the most used parametric model/ functions, since they can describe a wide range of shapes, while offering physically significant regularisation/ deformation. Shape limitation can be overcome by adding local shape perturbations [TM91], or by applying FFD subsequently to registration [BCA96]. Continuous models are particularly suited for statistical analysis, because their mapping functions can be efficiently decomposed into a linear combination of variation modes. Different basis (e.g. Fourier, harmonic, principal components) have been exploited in image segmentation [SD92] [CT01] for constraining model evolution into relevant deformations, that can be learnt from examples. Explicit continuous models need to be discretised for evolving (Section 1.3.4), while, with parametric models, control points (part of the parameters) may be directly used. The main problem with this kind of models remains the inverse problem (defining shape functions and parameters from a set of locations), which is required for reconstructing/ initialising models and interacting with them. This is carried out through energy functional minimisation, as a geometric registration process.

Figure 10: *topology adaptive continuous model (T-snake) based segmentation from [MT99]*

A **discrete** mesh is represented by vertices and connectivity relationships. Connected points form faces and cells. In discrete contours (respectively triangle meshes, tetrahedral meshes and rectangular grids), each cell contain exactly 2 (resp. 3, 4 and 8) vertices. They are characterised by a constant cell connectivity. Dually, simplex meshes [Del99] are characterised by a constant point connectivity (in p-simplex meshes, each vertex is connected to exactly $p+1$ vertices). Discrete surfaces are very popular in the field of computer graphics and image processing, as they are flexible (correlation with anatomical structures) and memory cheap. Among the numerous registration methods, [PMTK01] use triangle meshes, while [FWG*99] use tetrahedral meshes and [MD05] use 2-simplex meshes. Abstract volumetric deformation grid have shown to be an efficient support for image deformation (FFD) and hence registration. In [LRMK99] Lötjönen et al. present a coarse-to-fine approach with multiresolution grids. Szeliski et al. [SL96] propose an octree grid, to ensure better fitting with anatomical structure. However, FFD-based registration methods, often fail to handle displacement discontinuities at boundaries.

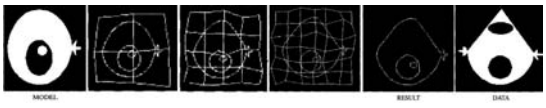


Figure 11: *Multi-resolution FFD-based registration from [LRMK99]*

Implicit Models are defined by the zero of a potential field F taking spatial coordinates as parameters. Com-

plex shapes can be obtained by merging simple implicit primitives, defined by polynomial function, such as superquadrics or hyperquadrics. Ellipsoids are part of superquadrics. Superquadrics also have a parametric formulation and have been widely used in image segmentation such as in [BCA96]. Level-set [OS88] is the most used implicit model representation in medical image analysis. It has become popular since they allow a great flexibility in terms of shape and topology, through higher dimensional parameters. However, as emphasised by Montagnat et al. [MD05], level-sets are computationally heavy and generally difficult to constrain spatially.

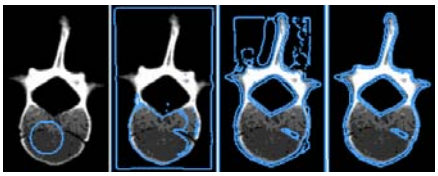


Figure 12: *Level-set based segmentation from [MD05]*

1.3.3.3. Similarity measure: The similarity (or external energy) between images (or between geometric features) is a measure of the matching quality. It is an important parameter upon which the searched transformation is computed. Global external forces and moments can be defined through the examination of the global energy variation according to transformation parameters (**standard approach**). In opposite, local external forces, subsequently converted into the closest global transformation, correspond to the **pair and smooth approach** (e.g. Iterative Closest Point (ICP) algorithm [BM92]).

In **geometric registration**, the similarity is derived from the distance between features: for instance, the Euclidian distance or the p-order Minkowski distance between two points, the Hausdorff distance or Mahanalobis distance between polygons, etc. A review of geometric registration techniques in medical image processing can be found in [AFP00].

A common methodology in segmentation is to derive the external energy from **image gradient** values [KWT88] around models. In this case, models converge to locally maximal gradients, but may be attracted by undesired noisy edges. This can be partly corrected by regularising the gradient vector field. Nevertheless, a more robust approach is to use prior information about image intensity and gradient through **intensity (or iconic) registration**. In iconic registration, the similarity relies on the correlation between the intensities of aligned voxels or regions (in the source reference dataset and the target dataset). Depending on the assumptions about this correlation [RMA00], many similarity measures have been proposed [HHD*00] [MV98b]. Intensity conservation along displacement, which can be valid in motion analysis, is assumed with the sum of absolute differences (SAD), the sum of square differences (SSD), or the cross-correlation (CC) measures. Between these measures, CC is the least noise-sensitive but the heaviest in terms of computation. In mono-modal registration, the hypothesis of affine relationship between intensities from the two registered datasets improves robustness with regards to global intensity change across acquisition protocols and sessions. In this case, the cross-correlation normalised by the standard deviation (NCC) measure [HHD*00] [RMA00] eliminates the affine terms. In multimodal registration, assuming that intensities values can be mapped between datasets through a certain function (functional relationship), measures have been proposed, such as the correlation ratio (CR) or the variance of intensity ratio (VIR) measure [WCM92]. Finally, for multimodal registration, information theory has lead to popular and flexible measures, based on the joint entropy. Minimising the joint entropy (or joint histogram spreading) is equivalent to optimise intensity value prediction from one image to another, without a-priori knowledge of the functional relationship. The (normalised) mutual information (NMI, MI) results from the normalisation of the joint entropy [CMD*95] [Vio95]. MI is robust with regards to change in the size of the overlapping region. In [RMA00], Roche et al. unify the MI and CR measures within a maximum likelihood (ML) framework.

Prior to similarity computation, intensity filtering can be performed to decrease noise (e.g. anisotropic smoothing) or enhance particular features by convolution with discrete masks. Particularly, some authors have used image gradient magnitudes [MD05] or vectors [GPMTV04] instead of basic intensities.

1.3.3.4. Regularisation: Due to noise, local solutions and the aperture problem (apparent motion), the registration process need to be constrained through degrees of freedom minimisation (**parameterisation**) and **internal regularisation**. Making hypothesis about the regularity of the problem and the form of the solution (initial distance between the two datasets) is a key-point in a registration process, for at least two reasons: 1) it determines the solution search space (dimension, order of magnitude) and consequently rules most of the computational charge (parameterisation of the resolution); 2) it decreases the uneven effect of external forces and energy through intrinsic constraints. Assumptions are generally application-dependent: for instance, rigid transformations are involved in intra-patient multimodal fusion, bone tracking or viewpoint registration problems.

Parameterisation: Parameterisation of the resolution aims at limiting solution search to relevant degrees of freedom. Indeed most of the difference between datasets can be recovered from a limited number of parameters. The researched transformations can be global or local, whether which subset of images is concerned. Coarse-to-fine approaches increase the number of DOF and the localness of

the transformation during the registration process, avoiding fall into local minima of the energy. There are basically two ways or parameterising a registration process: the standard approach maximises the similarity directly in the space of admitted transformations (global parameterisation), while the Pair and Smooth (P&S) approach locally maximises the similarity (pairing) according to a local parameterisation, and subsequently smooth the result (through standard smoothing of the displacement or explicit resolution).

In the standard approach, global parameterisation can be performed using **homogeneous transforms**. Indeed, standard 3D geometric transformations can be formulated with 4×4 homogeneous matrices derived from successive basic transforms (e.g. rotation around one axis, shear in one direction, etc.).

Geometrically continuous elastic deformations can be derived (interpolated) from the **structured deformations** of a regular lattice (e.g. FFD [SP86]). The interpolant function is a triple tensor product of 1D weighting functions. The use of cubic splines is particularly popular since they allow various continuity constraints: Hermite, Bezier and Catmull-Rom enforce C1 continuity, while B-Splines are C2 continuous. Hence, B-splines are widely used in image registration [RSH*99]. The use of a regular lattice can be limiting for handling displacement discontinuities at organ boundaries. Indeed, a careful matching between model geometry and anatomy provides more realistic deformations. As extensions of B-splines, octree B-splines [SL96] and triangular B-splines [WHQ05] allow the use of more flexible lattice in terms of mesh topology.

Besides, **unstructured deformations** from irregular control points can be performed through radial basis function - RBF. The interpolated displacement is defined as the weighted combination of non-linear functions (or radial basis) of the distance from control points plus an affine or polynomial term. Weights are pre-computed according to control point displacements (boundary conditions). The TPS radial basis [RSS*96] is popular for interpolating displacement since it minimises the bending energy; moreover it can handle different local continuity constraints [RFS99] and is well adapted when the number of control points (landmarks) is small (however, it turns to heavy computation when it grows). The use of RBF is difficult in a variational framework, because, weights need to be recomputed at each control point displacement. In general, with RBF, scattered spatial correspondences are manually selected, or semi-automatically segmented; and not iteratively found in an optimisation framework. However, in [RAD03] Rohde et al. use a compactly supported radial basis allowing weights to be optimised locally and independently for each group of 8 points.

In P&S registration approaches, feature local pairing is constrained through **local parameterisation**. Because image registration highly relies on object boundaries registration, it is particularly suitable to constrain deformation in the normal direction of these boundaries and within a certain range (similarity is maximised in this search space) [Del99] [MD05] [LM01]. In [HS81] [Thi95], pairing is done in the direction of image gradient, with a certain magnitude (optical flow). With the DFD [HS81], the displacement is computed as $\mathbf{u} = (T - S)\nabla\mathbf{T}$. While, with the demon algorithm [Thi95] we obtain $\mathbf{u} = (T - SoM)\nabla\mathbf{T}/(\|\nabla\mathbf{T}\|^2 + \alpha^2(T - SoM)^2)$ (SoM is the current transformed source image within the iterative process), leading to a normalisation of the displacement (its amplitude remains below the constant $1/(2\alpha)$). Ghanei et al. [GSZW98] use both surface normal and image gradient directions.

The last class of parameterisation techniques is the one based on **statistics**. That is new shapes are obtained by a linear combination of N sample shapes. Consequently, deformations are limited in a space of N dimensions. Relevant samples are generally obtained by PCA. The example-based approach is quite popular in image segmentation and registration (brain) since it takes into account shape and appearance (image intensity) variability within the population [CT01] (Active Shape Model - ASM and Active Appearance Model - AAM) [WS00] [DSMK01].

During the registration process, a widespread approach is to progressively increase the number of DOF and the localness of the transformation in order to improve robustness and computational speed; this is the **coarse-to-fine approach**. In [MD05] and [RSH*99], geometric transforms, from rigid to affine, are successively applied. In [LM01], deformation spheres become smaller and smaller. Multi-resolution lattices [SL96] [RAD03] and multi-resolution surfaces [PMTK01] [GMMT06] are also used.

Internal forces/ energy: Intrinsic regularisation aims at ensuring a certain continuity of shapes and displacements by penalising other configurations during optimisation. Because the image registration problem is ill-conditioned (no stability) due to noisy external terms, extra regularisation terms (stabilisers) are needed within the resolution process. Either forces or regularisation energy terms can be applied depending on the chosen evolution process (see Section 1.3.4). Basically, they are of three types: smoothing, physically-based and example-based.

Smoothing energy or forces enforce shape continuity through curvature, elastic energy and/or bending energy minimisation. It relies on the assumption that biological tissues are relatively smooth. Most of regularisation terms, used in the literature, are based on 2nd order Tikhonov differential stabilisers [TPBF87], since they are a generic expression of shape deformation through differential terms up to the second order. 1^{st} order terms are related to the elasticity of the model (minimisation of the curvature, Laplacian smoothing), while 2^{nd} order terms deal with bending (averaging of the curvature). Forces derived from the minimisation of the energy can be obtained from the energy using

the Euler-Lagrange theorem. When using explicit or parametric models, the differential terms can be easily computed due to their continuous nature. This is more complicated for discrete models where geometric approximations need to be made. Standard Laplacian smoothing forces (attraction of vertices towards the barycenter of their neighbors) minimise the elastic energy but produce model shrinking. This well known problem can be tackled by compensating them with surfacic balloon forces [Coh91] or removing their normal contributions [MD05]. Their tangential contributions prevent from surface stretching. Bending forces rely on the averaging, in the normal direction, of the discrete curvature. There are several ways for estimating the curvature: Park et al. [PMTK01] average the normal contribution of Laplacian forces; Montagnat et al. [MD05] smooth the simplex angle; Ghanei et al. [GSZW98] smooth the Gaussian curvature. Similarly to shape, deformation can be regularised [TPBF87] [LM01], by enforcing continuous variation of the deformation with Tikhonov stabilisers. Deformable models, fitting to organ boundaries, implicitly handle deformation discontinuities at boundaries. However, this is not the case with FFD-based registration, where boundaries are embedded into the deformable volumetric lattice, so that Tikhonov stabilisers would not suit (isotropic regularisation). Different anisotropic stabilisers, reviewed in [DKA95], have been developed for handling discontinuities. They basically provide anisotropic smoothing perpendicularly to image gradient. Radial regularisation of surfaces is useful for constraining models possessing a medial representation. Indeed, this provides to surfaces a volumetric aspect and higher level control parameters through simple mechanisms such as radial springs [PFJ*03] [GMMT06].

Smoothing forces, minimising the bending and/or the elastic energy can be considered as weak prior information [MD05] because biological tissue generally satisfy this condition. On the contrary, **example-based regularisation** provides strong prior information as it may be specific to a particular problem, organ, individual, instant, etc. Montagnat et al. [MD05] add shape and trajectory constraints through forces attracting shapes towards a reference configuration. In contrast, example-based parameterisation combines sample deformations to compute model deformation.

The minimisation of the elastic strain energy ($E_{reg} = \int W dv$) has been used in several image registration studies [Chr94] [WS00] [BN96] [VGW06], since it provides **physically-based regularisation** (one-to-one mapping, no negative volume). Equivalently, one can use body forces (divergence of the stress), in a force equilibrium evolution process 1.3.4, where an equilibrium with external forces (image-based) is iteratively found. In the linear elasticity framework, body forces are expressed as (linear elasticity operator):

$$div(\sigma) = \mu\Delta\mathbf{u} + (\lambda+\mu)\nabla(div(\mathbf{u})) = -\mathbf{f_{ext}}$$

where λ and μ are the Lamé coefficients and $\mathbf{u}$ the displacement. Since infinitesimal displacement is not a valid assumption, Christensen et al. [Chr94] introduced the fluid registration, where the internal stress is reinitialised at each iteration. To speed up the rather slow algorithm from Christensen et al. (several hours for 2D registration) Bro-Nielsen et al. [BN96] implement a fast filtering method from the impulse response of the linear elasticity operator. As a natural extension, authors have investigated hyperelastic registration to allow large displacement in the elastic framework (no stress re-initialisation). Nowadays, sophisticated constitutive models are used for registration such as in [VGW06], where Veress et al. incorporate an anisotropic hyperviscoelastic constitutive model for cardiac registration. In addition to the large computational weight of such methods, there are some restrictions in using mechanical parameters, making physically-based regularisation not necessarily more robust than geometric regularisation: first, external forces from images do not have any physical meaning, so that they are difficult to blend with body forces. There are also topological problems due to the linearization of the elasticity operator. In addition, inter-patient registration based on mechanics does not have much sense (there is no matter deformation). Finally, the mechanical model is difficult to fit to the data because of assumptions (small displacement, linear isotropic behaviour, etc.) and unknown mechanical parameters. The problem can be over constrained such as the over penalisation of large rotations when using linear elasticity.

In multiple object segmentation, it is valuable to add extra constraints, preventing from interpenetrations. Few authors have actually incorporated neighbouring constraints and collision handling within an image registration process. In [MT99] McInerney et al. check self-penetration of their T-snake, using spatial hashing, to allow topology change. Non self-intersecting forces are applied in [PMTK01] on discrete models. Recent advances in collision handling from the computer graphics community [TKH*05] and topological constraints (organ attachment) have being applied in [GMMT06] on simplex surfaces.

1.3.4. Model evolution (numerical resolution):

The solution of the registration problem is found by minimising the energy of the model, composed of an internal/ regularisation term and an external/ image-based potential energy: $\tilde{\mathbf{P}} = argmin_{\mathbf{P}} E(\mathbf{P})$ ($\mathbf{P}$: parameters). At minimum, competing forces equilibrate and we have $\mathbf{F} = -\nabla\mathbf{E} = \mathbf{0}$. In most cases, because of the non-convex aspect of the energy (noise, multiple local minima, etc.), there is no analytical solution. After discretisation, the energy is minimised iteratively using traditional multivariate optimisation strategies. Alternatively, when conferring a mechanical aspect to energy/ forces, the iterative process can be performed through a dynamic evolution. Another class of evolution method is the Eulerian evolution, where the space is transformed rather than the model itself. The solution (model shape) is hence

implicitly defined by the space state. Eulerian evolution has been investigated for the front propagation of implicit models (level-sets [OS88]) and will not be discussed here.

1.3.4.1. Explicit resolution: In some simple cases, where the system has a small number of DOFs, and the external energy a simple form, it is possible to analytically find the global solution. In [AHB87], the best rigid, similarity or affine transform that minimise the distance (error in the least square sense) between paired point-sets is found, by applying the quaternion method. This is particularly useful in body motion capture, where the rigid motion of the different body segments need to be estimated from noisy marker motion. When not available, pairing can be performed by minimising locally the distance (e.g. closest point). The pairing is then updated at each iteration, where the optimal transform is found, until the convergence (P&S approach). This is the principle of the popular Iterative Closest Point algorithm [BM92].

1.3.4.2. Energy minimisation: The minimisation, aimed at seeking the equilibrium, is driven by forces (energy spatial derivatives). Applying the Euler-Lagrange equation, the energy minimisation problem is turned into a force equilibrium equation (stationary equation) that can be use to derive the evolution of the model. Space discretisation is performed by the finite difference method [TPBF87], the finite element/volume method [MT95] or any parameterisation scheme (e.g. B-Splines) that lumps system degrees of freedom to control nodes. After discretisation, we obtain $-\nabla \mathbf{E}(\mathbf{P}) = \mathbf{F_{int}}(\mathbf{P}) + \mathbf{F_{ext}}(\mathbf{P}) = \mathbf{KP} + \mathbf{F_{ext}}(\mathbf{P})$ ($\mathbf{K}$: the stiffness matrix) which is zero at equilibrium (local energy minimum). Instead of minimising the energy (maximum likelihood scheme), some authors maximise the probability of model shape given the image (maximum a posteriori scheme), through a Bayesian framework [SD92] [WS00]. This can be viewed as an extra layer to which optimisation is performed. The Bayesian approach is interesting for incorporating statistical shape information (example-based approach) and image noise.

Local optimisation: to find the minimum of the energy, *exhaustive research* methods, or quasi-exhaustive methods (e.g. multigrid) may suit when the number of degrees of freedom is small (e.g. translation). In most cases, the complete inspection of the search space is too costly, and the search must be oriented. A popular oriented search method is the *Downhill Simplex method* [NM65]. The associated figure (simplex) is successively transformed through reflections, expansions and contractions, until a tolerance (simplex volume) is reached. Another way is to descend along to the energy gradient [Thi95] (the model evolves in the steepest energy direction with a certain speed). However, the convergence might require a large number of iterations (oscillations often occur around the solution). To speed up the convergence, one may adapt the time-step and the direction of the evolution. In the *conjugate gradient* algorithm, the direction of the evolution is forced to be orthogonal to the previous ones. The *Powell's method* provides conjugate directions without computing the derivatives. In the *Newton algorithm*, the time-step is computed, assuming that the energy is quadratic (2nd order Taylor development of the energy). The *Levenberg-Marquardt method* smoothly combines the *Newton* and the *gradient descent* algorithm (the Newton method is more suited near a local minimum, whereas the gradient descent works better far away from the solution). Newton method involves the expensive evaluation of the Hessian matrix of the energy. *Quasi-Newton methods* have been proposed to update an Hessian matrix estimate at each iteration, considering energy and energy gradient change. Considering the first order development of the energy that we want to be null in the next time step and assuming a displacement along the gradient we get the *Newton-Raphson method* where only first order energy derivatives (forces) need to be computed. In [MHHR06], Müller et al. use this method within a mechanical system to derive its dynamic evolution.

Global methods: Above methods are local, meaning they work well when the energy has few local minima (convex energy). If not, *stochastic optimisation* methods such as the simulated annealing or evolutionary algorithms can be used to find a global solution, despite their significant computational load. The simulated annealing (or stochastic relaxation) randomly changes parameter values at each iteration, and allows energy increase depending on the temperature (the higher the temperature, the higher rise is allowed). The temperature decreases until the instant freezing, where allowed perturbations only make a reduction of the energy. *Evolutionary algorithms* (e.g. genetic algorithms) are based on the Darwinian principle where a population of agents (parameters) evolves according to a fitness function (energy) and gene crossing principles. In [MH06], a genetic algorithm is used along with deformable models for medical image segmentation. *Dynamic programming* can be used to find a global solution. It is based on the decomposition of the energy into independent terms. Solving these sub-problems, the global solution can be recovered by finding the shortest path in the energy graph. Amini et al. [AWJ90] have applied dynamic programming to 2D snakes. But, this has not been extended in 3D, the critical point being the ability to decompose the energy.

1.3.4.3. Dynamic evolution: In a dynamic system, forces drive the model, that is the position and velocity of control points. This is particularly suited for discrete models that are commonly considered as lumped mass particles, moving according to forces. Dynamic systems are judged against their stability and accuracy, resulting from the numerical integration of the differential equation. The Newtonian law of motion leads to a first-order differential equation system relating the force vector $\mathbf{F}$ to the particle velocity vector $\mathbf{V}$ and position vector $\mathbf{P}$:

$$\frac{\partial \mathbf{P}}{\partial t} = \mathbf{V} \qquad \mathbf{M}\frac{\partial \mathbf{V}}{\partial t} = \mathbf{F}(\mathbf{P},\mathbf{V})$$

The force vector $\mathbf{F}(\mathbf{P},\mathbf{V})$ depends on particle position (internal and external forces/ energy derivative) and velocity (damping representing the viscosity of the milieu/ energy dissipation). When discretising the system with finite differences, we obtain:

$$\begin{aligned} \mathbf{u} = \mathbf{P_{t+dt}} - \mathbf{P_t} &= \mathbf{V}dt \\ \mathbf{V_{t+dt}} - \mathbf{V_t} &= \mathbf{M^{-1}F}(\mathbf{P},\mathbf{V})dt \end{aligned}$$

The Lagrangian evolution, that assumes massless particles and unitary isotropic viscosity [TM91] ($\mathbf{F}(\mathbf{P},\mathbf{V}) = -\nabla\mathbf{E}(\mathbf{P_t}) - \mathbf{IV} = \mathbf{0}$) leads to $\mathbf{P_{t+dt}} - \mathbf{P_t} = -\nabla\mathbf{E}(\mathbf{P_t})dt$, which is equivalent to the gradient descent algorithm.

In the **explicit (or forward) methods**, velocities and forces are taken at time t ($\mathbf{V} = \mathbf{V_t}$ and $\mathbf{F}(\mathbf{P},\mathbf{V}) = \mathbf{F}(\mathbf{P_t},\mathbf{V_t})$), so that the new state vector (P_{t+dt}, V_{t+dt}) is explicitly defined in the system 1. However it goes forward blindly (no notice of force derivatives), so that the stability is ensured for very small time steps dt. A higher order forward method is the well known Runge-Kutta method [PTVF92]. Its principle it to use force multiple evaluation to better extrapolate the new state vector through higher order Taylor expansion. A simple way for improving stability, with no extra computational cost, is to take the implicit formulation of the first equation of 1: $\mathbf{P_{t+dt}} - \mathbf{P_t} = \mathbf{V_{t+dt}}dt$, where $\mathbf{V_{t+dt}}$ can be obtained with the forward step described above.

In the **implicit methods**, the idea is to estimate state vector temporal derivatives at $t+dt$ using its spatial derivatives [BW98] [VMT05]. Implicit methods are called backward since a forward step from new state vector at $t+dt$ brings it back to the initial state vector at t. Hence, results are always consistent, allowing larger time-steps. Since [BW98], implicit methods have become popular in computer graphics. However they are not necessarily more accurate than explicit methods, due to numerical damping. A comparison of explicit/ implicit methods in terms of stability and accuracy is done in [VMT05] [HE01]. The implicit resolution is equivalent to the resolution of a large sparse linear equation system $\mathbf{V_{t+dt}} - \mathbf{V_t} = \mathbf{H^{-1}Y}$. The more stiff (large number of eigenvalues) is the system, the more complicated is the resolution (more solver iterations). In [DSB99], only the linear part is resolved, through pre-computed inverse matrix. In [EEH00], explicit resolution is applied to the non-stiff nonlinear part, while implicit resolution is applied to the stiff linear part (IMEX method), which improves accuracy (less numerical damping due to implicit resolution) and computational speed. To increase accuracy, authors have used history states to better approximate the new state. Particularly, Backward Differentiation Formulas (BDF) have been recently applied to deformable models [HE01]. They are computationally inexpensive, since they only require one system resolution at each step, contrary to Runge-Kutta methods. BDF are less robust in the cases of nonlinear systems, since they are sensible to sudden state change (e.g. collision response) due to their dependence with previous states. In practice, BDF are limited to their second order (first order is equivalent to the implicit Euler).

The resolution of the system always involves the inversion of a large sparse matrix. Terzopoulos et al. [TPBF87] resolve the linear system with a Choleski decomposition and a relaxation method (Gauss-Seidel). Recent applications use the Conjugate Gradient (CG) algorithm, which is particularly adapted to sparse systems [PTVF92]. The CG algorithm works on symmetric positive-definite systems which is not the case with implicit schemes, due to mass-modified particles. Post-multiplication [VMT00] or pre-multiplication [BW98] by the inverse mass matrix can be performed to recover global symmetry. In [HE01], the CG algorithm is integrated in the Newton minimisation method, allowing error control and adaptive time stepping. The Conjugate Gradient can be accelerated using matrix preconditioning [She94]. System stiffness is reduced through matrix multiplication: $\mathbf{V_{t+dt}} - \mathbf{V_t} = (\mathbf{U^{-1}H})^{-1}\mathbf{U^{-1}Y}$ where $\mathbf{U}$ is the symmetric positive-definite preconditioning matrix. Indeed, if $\mathbf{U^{-1}H}$ contains less non-zero elements than $\mathbf{H}$, its inversion is cheaper. Popular preconditioners are the Jacobi preconditioner ($\mathbf{U_J}$ = diagonal of $\mathbf{H}$) and SSOR (Symmetric Successive Overrelaxation Method) conditioner ($\mathbf{U_{SSOR}} = (\mathbf{U_J} + \mathbf{L})\mathbf{U_J^{-1}}(\mathbf{U_J} + \mathbf{L})^T$ where $\mathbf{L}$ is the lower triangular part of $\mathbf{H}$). Following the idea of Terzopoulos et al. [TW88], forces can be applied on rigidly registered models (non-rotated references), leading to more sparse system, especially when using linear Cauchy strain [MG04]. This can be viewed as a particular preconditioning method.

1.3.5. Conclusion

Most proposed registration techniques remain generic and fairly independent from the targeted applications. The lack of specificity makes methods difficult to parameterise, difficult to evaluate and generally holds back their implementation in the clinical environment. However, it is reducing due to more-and-more established physical-based simulation methods, available data for example-based approaches and cross-validation studies. Unifying tissue simulation and segmentation could lead to a better specificity for daily clinical use. Moreover, physically-based constraints, taking into account not only pixel intensity but high-level parameters such as the shape, the function and the mechanics of tissues, improve the accuracy and the robustness of registration methods by reducing problem degrees of freedom. Transformations can be limited to admissible deformations through examples [DSMK01] or/ and prediction methods [VGW06]. This is particularly relevant for intra-patient registration, while, for inter-patient registration or longitudinal studies, it

is required to study how high-level parameters change across individuals or through time. For inter-patient registration, we expect that external forces would play a more significant role than for intra-patient registration (image-driven versus physics-driven). In this context, deformable surfaces offer a moderate complexity and a good flexibility to tune intrinsic, geometrical and data constraints. Besides, physically-based volumetric models are more suited for patient-specific tissue simulation and intra-patient registration.

1.4. Example: musculoskeletal analysis from MRI

1.4.1. Introduction

For the diagnosis, the surgical planning and the post-operative assessment of musculoskeletal disorders, the automatic segmentation of the patient musculoskeletal system is important for orthopaedists, biomechanicians and kinesiologists that would like to simulate, visualize and navigate through articulations with a minimum amount of manual tasks. As stressed by [TSB*05], [NTH00] and [BD05], usual simplified models (stick-figures, muscle action lines) are not able to take into account large attachment areas, as well as global constraints such as volume preservation and non-penetration; although they are important biomechanical parameters. Therefore, the relationship between musculoskeletal dynamics and organ shapes need to be better studied through image segmentation. Magnetic Resonance Imaging (MRI) is a flexible modality for imaging both soft and bony tissues non-invasively. However, due to the large amount of textural information, noise, low-resolution, organ imbrication, and large spatial variability; automatic, fast and robust musculoskeletal segmentation is a difficult task. As a consequence, existing methods for musculoskeletal modelling are interactive [NTH00] [AT01] [TSB*05] [BD05] and therefore time-consuming. A common approach to constrain a segmentation process is to use prior information: shape constraints rely on assumptions about surface regularity (smoothness, curvature) and variability across the population, while topological constraints exploit prior knowledge about organ interrelationships. Contrary to traditional pixel-based segmentation (e.g. level-sets) and registration (e.g. FFD) methods, these constraints can be efficiently applied on deformable models [TWK87a]. Physical-based simulation approaches (volumetric), including finite element or finite volume are appropriate for enforcing mechanical constraints, but their computational cost would exclude time-efficient and user-controlled segmentation. In this context, we propose to use scalable discrete deformable surfaces (simplex meshes), that can benefit from efficient geometric methods, popular in computer graphics applications.

Simplex meshes were first described by H. Delingette [Del99] for constrained 3D shape reconstruction and segmentation, and extended in 4D by J. Montagnat [MD05], with application to the heart (single model and resolution). A k-simplex mesh is defined by a set of vertices and a connectivity function (each vertex is connected to exactly k+1 neighbors). In this paper, we use 2-simplex meshes (dual to triangle meshes). Mesh geometric quality (uniformity of vertices repartition) and topological quality (uniformity of edge number among faces) are improved using simple topological operators. The most interesting property is its simple geometric description: three parameters (two barycentric coordinates plus the curvature) uniquely define vertex positions from their three neighbors. Based, on these parameters, smoothing and shape constraint forces are computed to regularize the segmentation. External forces are obtained through 1D registration of intensity profiles (at vertex positions and in normal direction) with generic profiles from a reference segmentation. The external force field is regularized using a local smoothing and global regularization based on the closest affine transformation.

Discrete models are commonly considered as punctual masses evolving under the Newtonian law of motion. The Newton equation leads to a first-order differential equation system relating forces to particle state (velocity and position). After forces evaluation and time discretisation, particle state can be explicitly resolved (forward Euler, Runge-Kutta) with tight time step restrictions for ensuring stability. We prefer the more stable implicit (or backward) scheme [BW98] [VMT00], that however requires the resolution of a large sparse linear system. To simplify force derivatives evaluation, we consider that forces have a independent anisotropic action on each particle. Non-penetration constraints are based on collision handling techniques [TKZ*04]. In this paper, we apply a hierarchical collision detection scheme based on 18-discrete oriented polytope(DOP) quadtrees [TKZ*04]. We perform collision correction and response on particle positions, speeds and accelerations, such as in [VMT00].

From these background studies, we propose improvements in terms of computational speed and robustness by extending the simplex mesh framework with a multi-resolution scheme, topological constraints and medial surfaces. We show that the medial axis is suited for muscle shape analysis.

1.4.2. Methods

1.4.2.1. Multi-resolution Scheme The use of levels of details (LODs) aims at reducing system complexity and sensitiveness to local solutions. As shape constraints spatial influence depends on the resolution level, the idea is to quickly propagate forces from lower resolutions to a current simulation LOD in order to get multiscale regularization. In addition, collision detection (often considered as the bottleneck for simulation) is more efficiently performed on coarse LODs, collision response being passed to fine LODs. Indeed, exact contact computation is most of the time not relevant as fat separates organs. The tessellation of dual triangle meshes leads to a systematic and computationally efficient

LOD generation scheme for simplex meshes (linear combination of vertex positions). Shape features are preserved as low level vertices are contained in higher levels. The number of vertices is quadrupled when increasing the resolution. During the simulation, forces from lower resolutions are linearly combined like vertex positions and added to current resolution forces. Even if this is not exact (vertices having moved relatively to their neighbors, since resolution increase), the estimation is still relevant, assuming that shape constraints have enforced mesh local regularity.

1.4.2.2. Topological constraints The human musculoskeletal anatomy exhibits various organs interrelationships: muscles are attached to bones, they can merge into common tendon units, fascia binds muscles and enforces frictionless contact between them. We have developed a spline-based method for generating attachment areas, so that the placement and adjustment of areas have a reduced number of degrees of freedom. Spline control points are projected onto bone surfaces, while soft-tissues vertices are attached to the spline through curvilinear coordinates. These vertices are constrained using mass modification [BW98]: $\mathbf{M^{-1}} = \mathbf{0}$. For individualization, we wrap splines from a generic model using spline control point barycentric coordinates on bone surface. As shown in [KdH04], this approach is valid for most attachments as they rely on bone geometrical features. For merging deformable models, we sum forces and masses of the attached vertices. This is interesting for modelling common tendons (e.g. quadriceps femoris tendon), but also specific parts with high curvature, where smoothing forces are not appropriate (e.g. attachment between the adductor magnus and its inferior tendon). Fascia is modelled through collision handling: once generic collisions and proximities have been detected, they are subsequently used as a reference (springs) to smoothly enforce the relative position between models. In other words, we deform generic contacts.

1.4.2.3. Radial constraints Muscles generally have a smooth and tubular shape. Hence, they can be efficiently represented by an underlying piece-wise action line [TSB*05] [AT01], where isotonic contractions are modelled through action line shrinking/stretching, and isometric contractions through radial constraints applied to a wrapped surface. This has been applied in biomechanics for calculating joint moment arms [BD05]. However, muscles with large attachment areas and/or several origins/insertions require many action lines. We propose a continuous representation using medial surfaces, leading to a continuous radial regularization of the overlying surface and enhanced mechanisms for detecting collisions and characterizing shapes. Anchored medial surfaces (M-reps) have been successfully used in [TWK87a] [PFJ*03] for constraining deformable model-based segmentation. The medial axis transform (computation of maximal ball centers and radii inside an object [Blu67]) is an invertible transform that allows an efficient reduction of parameters dimension. Several methods have been presented for approximating the medial axis (MA), based on Voronoi diagrams, on distance maps, or on thinning.

We propose an iterative method based on forces applied to constrained deformable surfaces. Hook's spring interaction forces $\mathbf{f_j^*}$ are applied to medial surface vertices for MA approximation (j indexes medial vertices). Reciprocally, forces $\mathbf{f_i}$ are applied to overlying surface vertices for MA-based shape regularization (i indexes model vertices). Each model vertex $\mathbf{P_i}$ is associated to a medial surface point $\mathbf{P_{i\perp}}$ with a certain radius r_i. Radii are linearly interpolated between medial surface vertices $\mathbf{Q_j}$ of radius R_j to allow continuous force definition (leading to a smooth reconstructed surface from the MA). Given the weights w_{ij}, we have $\mathbf{P_{i\perp}} = \sum_j w_{ij}.\mathbf{Q_j}$ and $r_i = \sum_j w_{ij}.R_j$. From a reference state (Figure 13a), weights are obtained by projecting orthogonally $\mathbf{P_i}$ onto the MA, and radii by the weighted mean $R_j = \sum_i w_{ij}.P_iP_{i\perp}/\sum_i w_{ij}$. Weights w_{ij} and radii R_j are subsequently memorized to compute radial forces for deformed states (Figure 13b). Model forces $\mathbf{f_i}$ are simply defined by $\mathbf{f_i} = k.(1 - r_i/P_iP_{i\perp})\mathbf{P_iP_{i\perp}}$ where k is the spring stiffness (or force weight). By applying the momentum conservation law [VMT00], we calculate the force contribution $\mathbf{f_{ij}}$ for one spring i at $\mathbf{Q_j}$ such as $\mathbf{f_{ij}} = -w_{ij}.\mathbf{f_i}/\sum_j w_{ij}^2$. We average all spring forces associated to a MA vertex: $\mathbf{f_j^*} = mean_{\,i}(\mathbf{f_{ij}})$. To obtain the reference medial surface for a given model, we first construct and attach a simplex plane, based on muscle attachments(Figure 15). After cropping, we simulate the medial surface with radial and smoothing forces, while projecting model vertices and updating radii at each timestep. The medial surface quickly converges to the medial axis. We measure the error between the reconstructed surface from the medial axis and the desired initial model shape by calculating the mean of $|r_i - P_iP_{i\perp}|$. The average error for all tested models is 0.6mm. This value remains stable during the segmentation when applying radial forces to both models and axis.

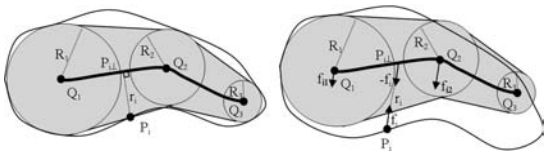


Figure 13: *Surface model and its associated medial axis in: A) A reference state; B) A deformed state; In grey, the reconstructed surface from the medial axis*

In addition, we use medial axis for collision detection: 18-DOP quadtrees hierarchies, inflated according to radii, are generated/updated for medial axis cells. Muscles vertices at a certain resolution are subsequently tested towards DOPs and projected to medial axis for the inclusion test (comparison with interpolated radii r_i). Independently of the model resolution, medial axis based-collision detection is around 3

times faster than surface-based collision detection. On top of MA based-collision detection, multi-resolution collision handling is used to improve computation time (by a factor of 7 per LOD). Given a current resolution, we use its first coarser LOD for collision handling in order to minimize errors.

1.4.3. Results

1.4.3.1. Automatic segmentation of the musculoskeletal system MRI protocol definition has been done in close collaboration with physicians from radiology and orthopaedics. The goal was to obtain images carrying sufficient information with clinically achievable protocols (fast). The final protocol was T1-weighted spin echo with TR/TE=578/18ms, Matrix/ FOV=512x512/40cm and slice thickness=2mm to 10mm between the foot the the iliac crests (five series in total, $\sim$150 slices, $\sim$30 minutes).

Based on interactive segmentation, we have reconstructed a generic model of the hip and the thigh, composed of a skin, 20 muscles, 4 bones and the corresponding attachment splines. After topological optimization, the different LODs (3 for soft-tissues, 4 for bones) and medial surfaces were generated. The final surface model (highest resolution) is composed of 71328 vertices for muscles and 85100 vertices for bones. Alternatively, muscles can be represented by their medial surfaces (1 resolution, 3821 vertices) with an average error of 0.6mm. The compression factor (parameter dimension reduction) is equal to (3x71328)/(4x3821)=14 (3 dimensions for model points, 4 for axis points, including radii).

Before automatic individualization, the generic model is coarsely initialized using ten manually placed landmarks corresponding to anatomical landmarks, and using thin-plate-spline (TPS) interpolation. After this step, bones are deformed (Euler implicit integration) from the coarse level to the fine level using multi-resolution internal forces (shape and smoothing constraints), non-penetration constraints (multi-resolution collision handling) and intensity profile-based external forces. Muscles are subsequently individualized (Figure 14) as follows: A) attachment splines are initialized on bone surfaces from their generic barycentric coordinates; B) a skeleton-driven deformation algorithm [KMTM*98] (skinning) is applied to generic muscles and medial surfaces according to joint angles; C) soft-tissues/medial surfaces are deformed using internal forces (radial/shape and smoothing constraints), proximity constraints (deforming contacts), and skin surface matching (gradient-based external forces); D) soft-tissues are deformed from the coarse level to the fine level using internal forces, proximity/non-penetration constraints and intensity profile-based external forces. The best intensity profile size and resolution have been experimentally defined from interactively segmented models by minimizing excursions. Being more variable, the external part of the profile is less relevant (thus shorter). The intensity profile search depth decreases from 20mm to 5mm during segmentation (Figure 14). We found that normalized cross-correlation is the most robust metric for intensity profile similarity measure. The affine regularization contribution decreases from 100% to 20%. For fine resolutions, internal forces from coarse levels are propagated (global/local constraints) as presented in **2.1**. Muscle shape constraints are derived from their medial surfaces with generic radii and weights. However, radii are updated for fine levels to give more freedom to surfaces. Radial forces are applied simultaneously to surfaces and to medial axis with an equal contribution. For coarse levels, generic proximity constraints are applied (deforming contacts). These constraints are released for the finest level, where multi-resolution medial axis-based collision detection is applied to allow surface sliding.

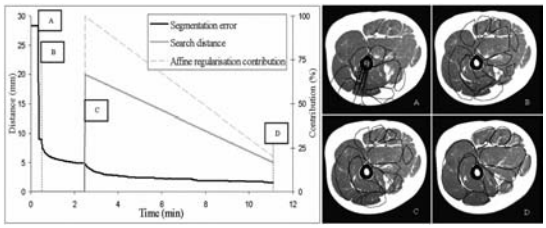


Figure 14: *Automatic muscle segmentation process and result on a sample slice: A) Generic model with wrapped attachments; B) Skeleton-driven model initialization; C) Segmentation using internal forces and skin matching (min. resolution); D) Final segmentation using external forces (max. resolution)*

During the segmentation, it is possible to interactively place constraint points on the images, to get a faster matching and a more accurate segmentation. Using our collision detection scheme, deformable models are forced to include or exclude these constraint points. In addition to the generic model, this method has been applied by a medical student to segment accurately 4 datasets from 4 different healthy subjects (2 females and 2 males). We compared automatically segmented models to these reference models (Figure 14). The average distance (std. dev.) was 1.25mm (1mm) for bones and 1.7mm (1.8mm) for muscles. The overall computation time for the automatic method is around 15min on a standard PC, for which 3/4 of the time is spent on external forces computation. Computation time for each timestep, including visualization, is around 0.5sec.

1.4.3.2. Shape analysis Musculoskeletal shape characterization is important for anthropometric comparison between individuals, and deformation analysis (temporal and longitudinal studies). For this purpose, our method registers anatomical features through shape and topological constraints. High-level descriptors such as the medial axis convey more information than local descriptors (curvature).

Muscle thickness can be simply analyzed through medial axis radii comparison, as shown in Figure 15. Using geodesic distances to attachments, we compute normalized coordinates X and Y along medial surfaces, from which a thickness profile can be extracted (maximum radius in Y direction). For some muscles showing thickness steep changes, tendons lengths (which is an important biomechanical parameter) can be automatically extracted (Figure 15).

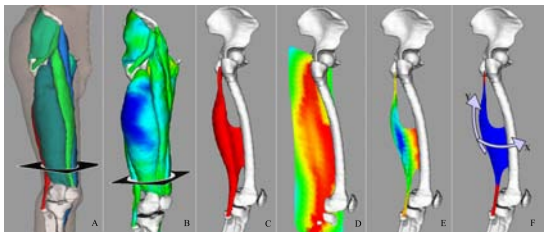


Figure 15: *A) Reconstructed generic model; B) Muscle thickness comparison with an individualized model (blue:≥5mm; red:≤5mm); C) Biceps femoris generic model; D) Initialized medial axis (blue: max radius, red: min radius); E) Medial axis after cropping and fitting; F) Tendon selection (in red) after thickness profile analysis along Y*

1.4.4. Discussion and future work

Prior low-level (e.g. curvature) and high-level (e.g. medial axis) shape information and topological relationships (e.g. proximities, attachments) are relevant for musculoskeletal modelling, and complexity can be efficiently decreased using a multi-resolution approach for force and contact computation. By constraining the problem, our goal is to get a fast and accurate segmentation from a minimum amount of information: we want to extend our previous work on bone motion extraction from real-time dynamic MRI [GPMTV04] (6 low resolution slices) by extracting soft tissue deformation. We believe that we can get a higher accuracy through a deeper study of intensity profile forces (combination of several metrics, weighting according to profile relevance, etc.) and by adding statistical constraints applied to high-level descriptors (e.g. medial axis radii) according to joint angles. The next step will be to relate these descriptors with dynamic (e.g. moment arms) and physiological parameters (e.g. muscle activation from EMG). This will provide useful information for validating functional biomechanical models.

2. Physically-based simulation of biological tissues

2.1. Constraints for Soft Tissue Modeling

There is no such thing as a good soft tissue model. The value of model lies in its adequacy with its applicative context. Thus, we first overview the constraints that arise from the use of biological tissue models inside a medical simulator.

2.1.1. Biomechanical Model

The obvious constraint for soft tissue modeling is that it correctly represents the deformability of real tissue. More precisely, we need to control the accuracy of deformation and therefore compare the computed soft tissue model with the actually deformed tissue. The level of accuracy typically depends on the application. For surgery training for instance, realistic visual and haptic display is more important than the accuracy of deformation. However, if the difference of behavior is too great under large deformations, it could result in learning inappropriate procedures.

In all cases, it is vital to have quantitative knowledge of the biomechanical behavior of soft tissue. We distinguish between knowledge of soft tissue deformable properties and knowledge of interaction with surrounding tissues.

The study of soft tissue deformability belongs to the field of biomechanics. There exists a large bibliography on the study of soft tissue deformation [Fun93]. Those studies range from the determination of qualitative behavior of tissues to the recovery of quantitative parameters governing their deformation. The tissues that have been mainly studied include skin [Lar86], bones [Cow01], vessels [Fun93], muscles [McM87], brain [CM96] and heart [HS88].

Many mathematical models of soft tissue deformations have been proposed [Fun93], and an extensive description would fall outside the scope of this paper. However, we briefly present the simplified models that are commonly encountered for surgery simulation purposes.

The simplest model of static reversible elastic deformation corresponds to the *linear elastic* model. According to this model, a cylindrical tissue sample of height H and diameter L under a load τ (see figure 16) follows Hooke's law :

$$\varepsilon_1 = \frac{\Delta H}{H} = \frac{1}{E}\tau \quad \varepsilon_2 = \frac{\Delta L}{L} = \frac{\sigma}{E}\tau \tag{1}$$

where :

- ε_1 and ε_2 represent the longitudinal and transverse elongation of the tissue under the load τ.
- E and σ are the Young modulus and the Poisson coefficient that characterize the nature of the material.

Therefore a linear elastic material has a linear *stress/strain* relationship and the trajectories of physical points under a load of varying intensity are straight lines. Because of this linear relationship, linear elastic materials have been used for fast surgery simulation [CDA99a, PRZ92, KGG96].

For most materials, the linear elastic model is only valid for small displacements. For large displacements, more complex non-linear models have been been introduced such as the Mooney-Rivlin model [BN96, SBM*94, TBHF03], the Veronda materail [HFS*01] or the St. Venant Kirchoff [TF88c, BNC96, Kai96, PDA03] where the stress/strain

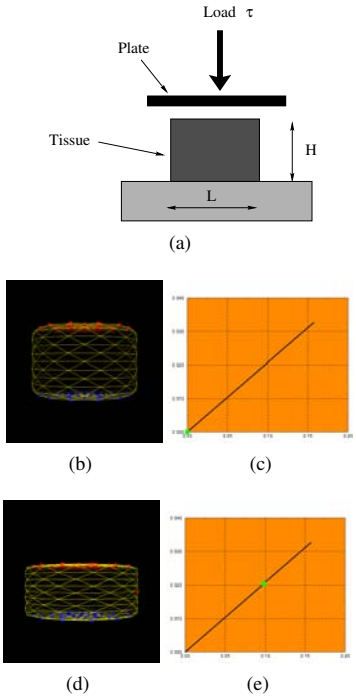


Figure 16: *Behavior of linear elastic materials.*

or stress/displacement relationships are no longer linear. Additional physical constraints may be considered, such as isotropy, transverse isotropy (e.g. for fibrous tissues) and incompressibility.

Furthermore, for many tissues, plastic deformations, where the strain does not reverse to zero after unloading (see figure 17 (a)), occur when the material reaches its elastic limit. Similarly, the material may have a non-reversible elastic behavior (see figure 17 (b)).

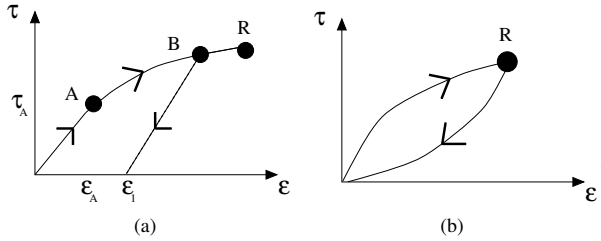


Figure 17: *Stress/strain relationships for plastic material (a) and hysteresis elastic material (b).*

The stress/strain relationship characterizes the static behaviors of a material. However, in general, the stress τ is also related to the speed of deformation $\dot{\varepsilon}$ (the *strain rate*) and therefore tissue material can be considered *viscous*. This implies that the deformation of a viscous material depends on the history of the applied forces and not only on their instantaneous values. The simplest viscous material is the *Newtonian* viscous fluid, where the stress is proportional to the strain rate (Equation of Newton) :

$$\tau = \eta\dot{\varepsilon} \tag{2}$$

Most soft tissues are *viscoelastic*, combining elastic and viscous behaviors. For instance, the Maxwell (serial dashpot and spring) and the Kelvin-Voigt (parallel dashpot and spring) viscoelastic models associate linear elasticity with constant viscosity. Those materials are characterized by their *creep* and *relaxation* functions occuring when the material is subjected to a constant load or deformation (see figure 18). Finally, many anatomical tissues can be considered as biphasic, where the tissue is a mixture of a solid porous matrix and an incompressible fluid [MKLA80].

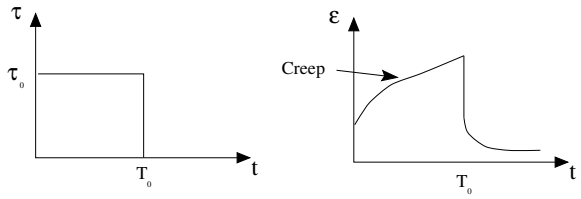


Figure 18: *Deformation response of a linear viscoelastic material under loading and unloading*

2.1.2. Interaction with rigid or soft bodies

The interaction between a soft tissue and surrounding bodies (surgical instruments, bones, soft tissue,..) can be decomposed into two different tasks : collision detection and computation of interaction forces (or displacements). Many algorithms have been proposed for detecting the collision between a moving point object and a static mesh. They are sufficiently efficient to sustain a 30Hz refresh rate with meshes of reasonable size (between 10 000 to 40 000 polygons).

However, in surgical simulation, those algorithms are not applicable because soft tissue models cannot be considered as static since several surgical instrument interact at the same time. Furthermore, for many surgical instruments, the interaction with tissue not only occurs at their extremity, but also along their length. Therefore, it is necessary to consider the surgical instruments as rigid moving meshes and soft tissue as deformable moving meshes. When detecting the collision between two moving objects, pre-computation is not possible and the computational time rises sharply. greater and it constitutes a bottleneck for

When considering the collisions between two deformable

meshes or the self-collisions of a deformable mesh, the computational complexity becomes prohibitive with current algorithms for real-time processing. However, recent techniques [TKH*05] offer possible research directions to overcome those major limitations. Distance fields (or implicit models), well suited for rigid objects, allow to quickly check if primitives lie inside/outside an object. Spatial subdivision methods store primitives/ grid correspondances within a hash table, and collisions are detected by comparing object entries. Image-space techniques use graphics hardware to project (render) images of the scene, and check collisions in 1D within the layered depth image. Bounding volume hierarchies (probably the most popular technique) use simple enclosing volumes for fast overlapping tests. Finally, stochastic methods are inexact methods that ensure control over computational time and collision detection quality.

Once a collision has been detected, the second task consists of computing the resulting interaction forces (collision response) to reach a non-penetration state. We usually distinguish *constraint dynamics* methods that incorporate reaction forces into the system [WGW90], *penalty methods* that enforce non-penetration through stiff springs [HFS*01], *exact methods* [Bar92] that analytically compute penalty terms by resolving non-penetration equations involving the relative acceleration between bodies, and finally *impulse-based methods* [MC94] that apply velocity impulses to guaranty non-penetration. The main limitation is that very little is known about the real interaction between biological tissues. Unlike material characterization, the experiments on tissue interaction must be performed *in vivo*.

2.1.3. Real Time Deformation

The real-time deformation of soft tissue is an important constraint for medical virtual reality systems. It has been clearly established that the immersion of the operator, therefore its ability to learn from a computer simulated system, is directly linked to the bandwidth of the simulator. An acceptable bandwidth for visual display is in the range of 20-60Hz while the acceptable bandwidth for haptic display is on the range if 300-1000Hz (300Hz is the free hand gesture frequency).

Two numbers are particularly important for accurate perception by the user : *latency* and *computation time* . Latency measures the time between sensor acquisition (for instance the position of the surgical instrument) and action (visual or haptic display). The computation time is that needed to update the geometric model. On multi-processor computers, those two numbers are not necessarily correlated.

Latency is critical for the user immersion. The hardware configuration of the system can greatly influence latency since communication between elements may be responsible for additional delays. In figure 19, we show the architecture of the simulation system used at INRIA [CDA96]. It is composed of one haptic display, a PC and a graphics workstation. There are several contributing causes to latency: communication between the haptic display and the PC, communication between the PC and the graphics workstation, the time taken by the graphics display, the computation time for collision detection, force feedback and deformation. Since much of the communication between elements are asynchronous, the total latency is not the sum of those delays but it is important to reduce them to their minimum values. The latency depends greatly on hardware, specifically on computation and graphics performance.

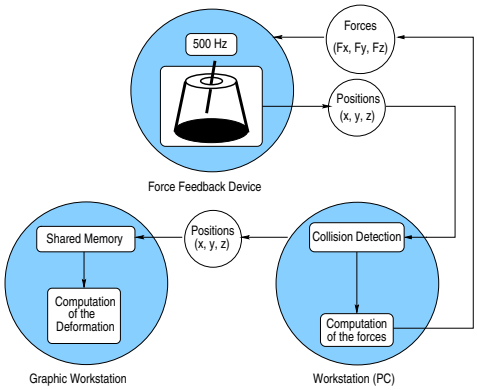


Figure 19: *Example of simulation systems.*

The computation time depends on the choice of the geometric and physical model of soft tissue (see section 2.2). If we write X_t as the position of the tissue model at iteration t, we will write C_t the computation time needed to compute the new position X_{t+1}. The computation time must be bounded ($C_t < C^\star$) in order to guarantee minimal bandwidth.

We distinguish between a *static* equilibrium equation and a *dynamic* law of motion depending whether X_{t+1} depends on the previous position X_t. When using static equations of the form $F(X) = 0$, the computed shape corresponds to the state of equilibrium, so takes no account of inertia or viscoelasticity.

With dynamic laws of motion, the current position of the tissue has an influence on its future position. Mostly, this can be modeled using a Newtonian law of motion having the corresponding differential equation :

$$m\frac{\partial^2 X}{\partial t^2} = -\gamma\frac{\partial X}{\partial t} + F(X) \qquad (3)$$

Depending on the complexity of the deformation model, the computation of X_{t+1} may be performed interactively (see Section 1.3.4). The number of iterations must be bounded since the computation time C_t must be less than $C^\star$. A major difficulty consists of ensuring *synchronicity*, *i.e.* that the *numerical time* used for computation matches the *user time*

used for interaction. For instance to correctly estimate the speed of a vertex x_t, one should divide the position difference $x_t - x_{t-1}$ by the real time spent between two iterations. Adaptive approaches have been proposed to ensure bounded computational load for computing deformations [DDCB01].

Static equations have two advantages. First, they are faster to compute because no time integration is needed. Second, they are well suited for parallel algorithms or for asynchronous computation. For instance, in Cotin *et al.* [CDA96], we use a static formulation motion that enable us to decouple the force with the deformation computation. This is highly advantageous because, as noted before, the force computation must run at least at 300Hz but requires few operations (product of a matrix with a vector) whereas the deformation computation must run at 30Hz with a large amount of computing.

However, static laws are not able to model realistic deformations such as inertia or viscoelasticity. Dynamic laws of motion on the other hand, can model more accurate deformations but are more difficult to handle and computationally more expensive.

2.1.4. Tissue Cutting and Suturing

The ability to cut and suture tissue is of primary importance for designing a surgery simulation system. The impact of those operations in terms of tissue modeling is considerable. In fact, they imply that the geometric representation of tissue must change its topology over time [CDA99a]. The cost of such a topological change depends largely on the chosen representation (see section 2.2).

In addition, the behavioral model of the tissue must be adapted at parts where cutting or suturing occurs. Little is known about the stress/strain relationship occuring during and after cutting. The basic assumption is that the physical properties of tissue are only modified locally. However, in practise, cutting can greatly modify the boundary conditions between tissue and the surrounding organs which entails considerable change in terms of deformability.

Finally, when cutting volumetric or surface models, it is very likely that the new geometric and physical representation of tissue leads to self-intersections. The detection of self-intersections is computationally extremely expensive, therefore repulsive force between neighboring vertices are sometimes added to prevent self-intersections.

2.1.5. Force-Feedback Computation

Haptic display serves at least two purposes in a surgical simulator : kinesthetic and cognitive. First, it provides the sensation of movement to the user and therefore it greatly enhances it surgical performance. Second, it is used to distinguish between tissues by testing their mechanical properties.

However, the addition of a haptic display in a simulation system increases by a large factor its complexity and the required computational power [MRF*96] : an increase by a factor 10 of the required bandwidth, synchronisation between visual and haptic displays, force computation, ... Few papers have assessed the importance of haptic feedback in surgery [Mar96] [CDA99a] [DDCB01]. In general, it is accepted that the combination of visual and haptic displays are optimal for surgery training or pre-planning.

In video-surgery, the surgical instruments slide inside a trocard and are constrained to go through a fixed point. This entails substantial friction, specifically in laparoscopy where airtightness must be enforced. The friction of the instruments inside trocards perturbes the sensing of forces by the end-user. Despite those perturbations, it appears that it is still necessary to provide force-feedback for realistic user immersion.

The force computation depends on the chosen deformation model. Deformation models based on biomechanics naturally lead to physically meaningful forces and are very likely to provide an intuitive feeling to the end-user. On the other hand, with more *ad hoc* deformation models, for instance spring models, the force computed at each iteration may not correspond to intuitive sensation, especially if there is no continuous representation of the tissue.

Another problem with the use of haptic displays is the small number of commercially available systems. They often correspond to modified versions of joysticks dedicated to different applications (teleoperation, video games, general user interface,...). Therefore, they rarely meet the specific constraints of open and video surgery, in terms of workspace or encoder resolution. Current commercial equipment read the position of the end-effector and are force-controlled.

2.1.6. Visualization

Visual feedback is the most powerful perception channel. The quality of the visual rendering greatly influences user immersion and therefore the effictiveness of the simulator. In the past few years, substantial technological advances have enabled sharp decreases in the price of efficient graphics boards. With the existence of standard programming environment (such as OpenGL, Direct3D,...) and standard platform configurations, this trend should grow over the next five years to deliver high performance at reasonable price. It will have the effect of increasing the number of polygons that can be drawn at each frame. By combining improved graphics performance with anti-aliased texture mapping with large texture memory, a major improvement of the realism can be foreseen.

However, there are many problems in visualization for surgery simulation that cannot be solved simply by improving the raw graphics processing. For instance, after cutting or suturing the tissue model, modification of the texture coordinates must occur to reflect the change of topology. The development of volumetric texture images could be used instead of two-dimensional texture images. When simulating

video-surgery, the optical aberration of large angled endoscopes, as well as the intense light of the optical fiber, must be modeled for more realistic display. Finally, techniques for depicting bleeding, and the rendering of semi-translucid and filandrous structures, must be improved.

2.2. Deformable Tissue Models

In this section, we review the main existing models of soft tissue.

2.2.1. Surface or Volumetric Tissue Models

The geometric representation of deformable tissue may consist of surfaces or volumes. The choice between surface and volume based models is governed by two factors: computer efficiency and physical accuracy. In terms of computation, surface models are advantageous because they have less vertices than volumetric models for representing the same shapes.

However, most biomechanical models such as those presented in section 2.1.1, naturally call for volumetric representations rather than surface models. Surface models tend to give physically invalid deformations especially in regions that are thin. In figure 20, we show the deformation of a volumetric plate model. Similar experiments with a surface plate model would have entailed self-intersections. Furthermore, the behavior of volumetric models may take into account physical inhomogeneities, for instance due to the presence of lesions.

Finally, volumetric models are better-suited to the simulation of cutting or suturing operations. This is because those operations changes the geometrical and the physical nature of the model.

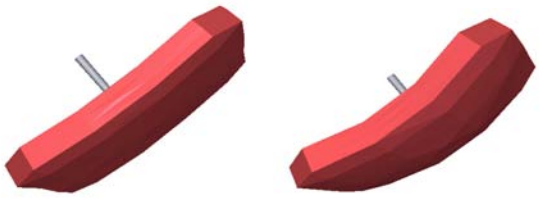


Figure 20: *Volumetric deformable model behavior.*

However, surface models may be relevant for modeling cavernous tissues such as vessels or the gallblader. In this case, the physical model of deformation can incorporate a representation of a liquid or of a gaseous pressure combined with surface tensions.

Here, we briefly review the main existing approaches, focusing of solid object expressed from the Lagrangian point of view. For more details, please refer to the recent exhaustive review on physically based deformable models by Nealen et al. [NMK*05].

2.2.2. Springs and Particles

The discretisation of the spatial continuous domain leads to a set of nodes that can be considered as particles with a certain mass, evolving in space under forces. As opposed to continuum-based methods, particle systems solve motion equation independently for each particle. This is a critical difference, allowing particle systems to be much more efficient in terms of computational speed. Indeed, particle systems, as they rely on a local description, lead to a much more sparse system that FEM, thus to a simpler resolution.

A particular implementation of particle systems is the well known mass-spring model, in which internal forces are straightforwardly computed from independent springs between particles. Spring models consist of a set of points linked by springs and dampers. In the simplest formulation, the equation of motion of a point i is :

$$\mu \frac{d^2\mathbf{r}_i}{dt^2} = -\gamma \frac{d\mathbf{r}_i}{dt} + \sum_{j \in N(i)} K_{i,j} \frac{(l^0_{i,j} - \|\mathbf{r}_i\mathbf{r}_j\|)\mathbf{r}_i\mathbf{r}_j}{\|\mathbf{r}_i\mathbf{r}_j\|} \quad (4)$$

where μ is the mass, γ the damping factor and $K_{i,j}$ the stiffness of the spring connecting point i and points j in the neighborhood $N(i)$ of point i.

Spring models have been used extensively for simulating the elasticity of soft tissue. Waters [Wat92] has defined springs on regular lattices for modeling facial tissue. He derived the two stiffness parameters of biphasic springs from stress/strain curves described in a biomechanical study [KGEB75]. Similarly, Delingette *et al.* [DSCP94] represented fat tissue elasticity as a network of springs on a 3-simplex mesh. Keeve *et al.* [KGG96] proposed a similar approach for modeling fat tissue in a craniofacial surgery simulation system but with the addition of a volume-preservation force intended to model the incompressibility of human tissue. Koch *et al.* [KG96] combined a finite element model for representing the skin surface, with a spring model to represent the fat tissue. The stiffnesses of the springs are derived from the intensity of voxels in a CT-scan image. The underlying assumption is that stiffness is proportional to tissue density and therefore to the Hounsfield units.

The main advantage of spring models is their ease of implementation since they do not require continuous parameterization. They have been used for static as well as for dynamic computation. Another advantage is their ability to model cutting or suturing simply by removing or adding connections between vertices. Mass-spring simulation is fast and can be easily parallelised and GPU accelerated [Mr05] (e.g. 100 000 particles at 100 frames/s).

However, for the soft tissue simulation, they suffer from the following problems :

- **Topological design**. The topology of springs and masses

is of great importance. Since a spring constrains the length between two vertices, the number of springs per vertex conditions the global behavior of the system. If the system is under-constrained, several rest positions are possible, and the system can fall into unwanted local minima. If the system is over-constrained, it tends to decrease the range of deformation. For tetrahedral meshes, the number of springs per vertex should be as close as possible to 6. Because of this difficulty in designing topologically a network of springs and masses, authors have usually organized the springs on regular lattices [Wat92] or on prisms with triangular base [KG96,KGG96]. Those choices however, impose the restriction of the geometric representation that it be organized in sets of parallel layers.

- **Validity of deformations**. The deformation induced by springs cannot easily be compared with those given by biomechanical studies because springs do not rely on continuum mechanics. However, for small deformations, a spring model behaves similar to a linear elastic finite element model, as verified by Keeve *et al.* [KGG96]. By linearizing equation 4, the stiffness parameter can be identified with the stiffness of a linear elastic model. However, for large deformations, spring models do not behave like a linear elastic material and comparison with nonlinear elastic finite element models is difficult. In order to identify spring parameters, given a behavior model, several algorithms have been proposed. In [JL95], Louchet uses a genetic algorithm for identifying the spring parameters to represent a cloth animation model. Deussen *et al.* [DKT95] bases the search for optimal parameters on simulated annealing.
- **Dynamic Behavior**. For dynamic spring models consisting of n nodes, for a given time step Δt and a given mass $\mu = m_{total}/n$, there is a critical stiffness K_c above which the numerical system is divergent. The relationship between K_c and the time step Δt is :

$$K_c \approx \frac{\mu}{\pi^2(\Delta t)^2} \approx \frac{m_{total}}{n\pi^2(\Delta t)^2} \tag{5}$$

  This relation, which is also valid for explicit linear elastic models, implies that in order to increase the stiffness of the model, it is necessary to decrease the time step. Since the computation time C_t is independent of the time step Δt, to maintain the refresh rate as high as possible, it is necessary either to decrease the stiffness K or to decrease the number of nodes n. Stiffness controls the propagation of constraints along the tissue model. High stiffness models tend to exhibit global behavior which is desirable for most soft tissues. In practise, we have found that greater time steps could be used with explicit linear elastic finite element models than for spring models. This implies that the range of possible dynamic behaviors of spring models is more limited than those of finite element models.
- **Visualization** Spring models are composed of mass-points and edges represented by springs. To visualize the tissue surface, it is necessary to define polygons from the set of edges and vertices. However, when cutting or suturing occurs it is necessary to update the set of visible faces. This suggests that the network of springs and masses should be built upon a manifold (surface or volume). A complete data structure of the manifold must be used in a way similar to finite element models.

Luciani [LJF*91] developed an animation system (CORDIS/ANIMA) based on particle systems for modeling complex physical phenomenon. The system is coupled with a force-feedback device for real-time interaction. Provot *et al.* [Pro95] defined super-elongated springs to increase the stiffness of its cloth model but with no guarantee of convergence. Cover *et al.* [CEO93] used springs to represent the surface of the gall bladder. They combined *home* forces with internal forces to enforce shape constraints. Similarly, Kuhn *et al.* [KKKN96], in the KISMET project, developed a surface model of the gall bladder based on springs, dampers and plastic elements where mass-points are connected to parent nodes. In that system, a NURBS representation is attached to points, giving a realistic rendering of the tissue. Stone *et al.* [RRF96] added a slip and a split threshold to a spring network, in order to model plasticity and fracture. Meseure *et al.* [MC97] proposed a surface representation of tissue similar to the hybrid model proposed by Terzopoulos [TF88a]. Springs of zero length are attached to a virtual rigid component. During deformation, both the positions of mass-points and the translation/rotation parameters of the rigid component are updated. This framework has the advantage of decoupling stiff behavior modeled by the rigid component and local deformation modeled by springs and dampers. Similarly, to reduce distortions due to linear strain formulation, Müller et al. [MG04] use non-rotated configuration to allow large deformations through simple computation. They also present a meshless implementation [MHTG05a] that makes use of point explicit registration (Section 1.3.4) to derive geometric internal forces and a stable integration scheme.

Gibson *et al.* [GSM*97] proposed a "ChainMail" model that does not derive from the equation of dynamics 3. Instead, the deformation model governs the displacement of all nodes, given a displacement on its boundary. Tissue is represented as a set of deformable voxels linked to their 6 nearest neighbors. When a node is pulled or pushed, neighboring links absorb the movement by moving slightly. If a link between two nodes is stretched or compressed to its limit, displacements are transferred to neighboring links. Within this framework, stiff behavior can be modeled for large displacements whereas compliant motion will be observed for small displacements. The "ChainMail" model is well-suited for real-time deformation and cutting operations, However, it is not clear how realistic the deformation is nor how feasible the real-time rendering of the tissue model is.

Other methods consider forces acting on elements such as tetrahedrons [CDA99a], or triangles [BW98] [VMT00], taking into account more that particle pair interactions, in

the evaluation of forces and force derivatives. They have shown promising results, in reproducing the various deformation modes of real materials, thanks to the formulation of the strain (small deformation strain or linearized version of the Green-Lagrange strain) at the element level.

2.2.3. Continuum-based methods

The finite-element method (FEM), the finite-difference method (FDM) and the boundary-element method (BEM) are well-established mechanical engineering techniques that have been widely used since the 70's for simulating tissues deformation. Their main idea is to discretise the continuous equation describing energy equilibrium (virtual work principle) inside an object [Zie77]. The spatial domain is subdivided into elements such as triangles, bricks or more frequently tetrahedrons [CDA99b] [OH99] [HFS*01] [TSB*05], and boundary conditions are defined (Dirichlet or Neumann boundary conditions). While the change of variables across a regular spatial grid is straightforwardly evaluated by discrete derivatives with FDM [TPBF87], FEM makes use of an interpolation shape function with limited support (e.g. B-splines, piecewise polynomial, linear basis, Bernstein-Bézier etc.), which is integrated in the energy equation. Finite elements with C^0 continuity, where the shape node consists of a vertex position are similar to finite difference methods. In the general case, FEM exploits higher order derivatives than FDM and requires fewer nodes, not necessarily regularly sampled (better approximation of organ boundaries). This leads to continuous representations with varying levels of continuity. A finite element model is fully defined by the choice of its elements, its shape function and its global parameterization between parameter space Ω and $\mathbb{R}^3$ ($\Omega \subset \mathbb{R}^2$ for surfaces and $\Omega \subset \mathbb{R}^3$ for volumes). For surfaces that are neither topologically planar or cylindrical, the parameterization can be problematic.

A finite element model is represented by the node vector $\mathbf{X}$. For static computation, the stress-strain relationships leads to $f(\mathbf{X}) = 0$, whereas for dynamic computation, the following Newtonian formulation is often used :

$$m\ddot{\mathbf{X}} + \gamma\dot{\mathbf{X}} + f(\mathbf{X}) = \mathbf{0} \tag{6}$$

Those equations can equivalently be derived by minimization of the bending energy through the principle of virtual work. The stiffness matrix expresses the strain-stress relationship and is consequently based on the underlying constitutive mechanical model. The integration of this differential equation can be performed using an semi-implicit or explicit scheme (see Section 1.3.4). In general, implicit schemes are unconditionally stable whereas explicit schemes are conditionally stable. This implies that smaller time steps must be used with explicit schemes. However, explicit schemes are simpler to compute and may not require the matrix inversion.

Membrane and thin plate energies have been largely used in computer vision [TWK87b], computer graphics [TPBF87] and for modeling elastic tubular surface tissue [Meg96]. These two belong to a family of regularizing energies : the controlled-continuity generalized spline kernels [Ter86]. Those quadratic energies have been used extensively because of their numerical properties (they lead to linear elastic forces). However they do not correspond to physical elastic energies. The membrane energy is a linearized version of the surface tension energy on soap films [Hil44]. The thin plate energy is a linearized version of the isotropic thin shell flexion energy [Ber96]. In particular, those energies are not invariant with respect to a change of parameterization. However, for small deformations, they can be considered as valid approximations. In [TF88a], Terzopoulos proposes a more general bending energy consisting of the sum of the square of the metric tensor and the curvature variation. He defines a hybrid formulation of deformation that includes a rigid component and a deformable component. With this new parameterization of deformation, invariant to rigid transformation, thin plate bending energy is used to model large deformations. In [CYTT92], a dynamic model similar to [TF88a] is proposed but the damping factor is replaced by the time derivative of the strain tensor. In [TF88c] a Maxwell and Voigt viscoelastic model has been implemented using a semi-implicit scheme with finite differences on a regular grid. The addition of plastic units enabled non reversible behavior to be modeled. However, those results required substantial computational power for relatively small grids (around 30×30).

Biological tissues are, in general, non-linear, anisotropic, hyperviscoelastic and quasi-incompressible materials. For simplicity and because of time-step restriction, linear elastic models have been widely used to model the deformation of soft tissues. In such cases, the stress/strain relationship is represented by a linear equation : $\mathbf{F} = \mathbf{KX}$. The rigidity matrix $\mathbf{K}$ depends on the rest shape geometry, the Young modulus E and the Lamé parameter λ. In most cases, only C^0 elements are used, leading to simple shape functions. However, for large deformations (Green strain), strain non-linearities are paid by high computational cost (the stiffness matrix depends on the displacement).

Pieper [PRZ92] and Keeve [KGG96] simulated fat tissue elasticity for plastic surgery with prismatic finite elements. Chen [CZ92] built a sophisticated muscle model based on biomechanical data where a linear elastic muscle is submitted to non-linear tendon forces. Despite the approximation on the muscle elastic behavior of the muscle, good correspondence with biomechanical has been observed. Gourret [GMTT89] computes the skin deformation of human fingers during the grasping of a soft object. The simulation models the interaction between two deformable bodies.

More complex elastic behaviors have been proposed by Terzopoulos. In [TF88c], he defines the square norm of

the metric tensor as the potential energy, which can be shown [BN96] to be equivalent to a St. Venant Kirchoff deformation model. Bro-Nielsen [BN95] used this non-linear elastic model for simulating craniofacial surgery with a finite difference scheme defined on cubic lattices. In [Kai96], Kaiss *et al.* proposes a complete model of the eye/trepan contact for predicting eye deformations under surgery. The sclera and cornea are modeled as St. Venant-Kirchoff materials and transversely isotropic materials. The penetration of the trepan requires the re-meshing of the cornea represented by hexahedrical elements with 27 nodes. The computation time was of the order of an hour on a powerful workstation.

To simplify FEM formulation, the boundary element method (BEM) has been used [JP99]. Its priciple is to transform the differential operator defined in the domain to integral operators defined on the boundary (divergence theorem) by assuming that the material is linear and homogeneous. Using this idea for individual tetrahedrons, the finite volume method (FVM) [TBHF03] [TSB*05] makes use of the divergence theorem to simplify force and strain computation by handling tetrahedron triangles instead of the boundary of the Voronoi region associated to each node. Assumption in FVM is the constant strain tetrahedra and linear variables change (linear basis).

Continuum-based models have been used widely to compute soft tissue deformations under mechanical constraints. In the past years, real-time models have been developed as an alternative to spring models. Sagar [SBM*94] developed a virtual environment for eye surgery simulation where the cornea deformation is modeled as a non-linear elastic material (Mooney-Rivlin material). The finite element solver computed the cornea deformation every second while the graphics module was able to provide a 10Hz refresh rate. In [CDA96], Cotin *et al.* describes a hepatic surgery simulator where the liver is represented as a linear elastic volumetric model with static constraints. By precomputing the response of surface vertices to position constraints, the liver model can be deformed in real-time. Furthermore, the force-feedback computation and the liver deformation computation can be decoupled to achieve optimal haptic display (500Hz) and visual display (30Hz). Similarly, Bro-Nielsen [BNC96, BN96] decreased the computation time of a linear elastic model with a semi-implicit scheme, by condensing and explicitly inverting the reduced stiffness matrix in a preprocessing stage. Video frame-rates of 15-20 frames/second were obtained with this method. Koch et al. [KRG*02] have used a high order finite element model for maxillofacial surgery planning from medical images. By taking advantage of the linear nature of the static or dynamic equation, the methods of Cotin [CDA99b] and Bro-Nielsen [BNC96] decreases the computation time of finite element models by at least factor of 100.

In figure 21 we show an example of real-time hepatic simulator developed at INRIA [CDA96, Cot97]. The user manipulates a force-feedback simulation platform and feels the contact between the virtual instrument and a liver model.

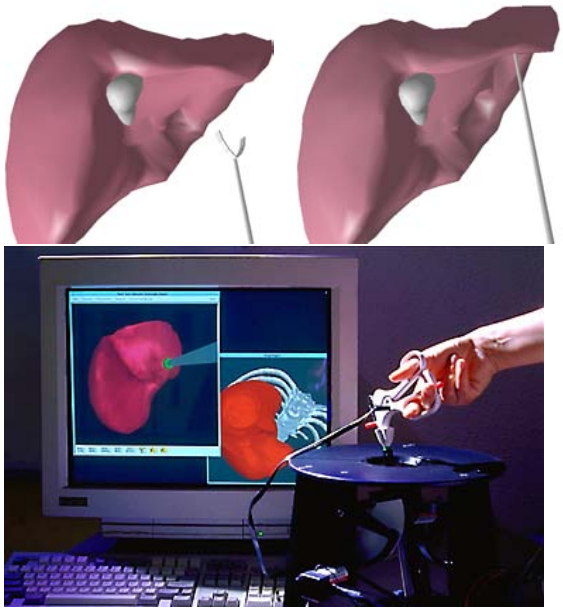


Figure 21: *Hepatic surgery simulation platform at INRIA*

As a conclusion, finite element models are well suited to compute accurate and complex deformation of soft tissue. However, it is extremely difficult to get real time performance on a moderately powerful workstation using finite element models. But for linear elastic models, only valid for small displacements, it is possible to achieve real-time deformations. Unlike spring models (see section 2.2.2), there is no restriction on the stiffness value of the model with respect to the time step Δt when using semi-implicit or static schemes.

The cutting or suturing operation requires the finite element model to be re-meshed. When using structured elements such as rectangular, prismatic or hexaedral elements, the cutting is often constrained to occur along a given direction [Kai96, SR95]. With unstructured elements such as triangular or tetrahedral elements, more general cut planes may be designed at the cost of greater complexity. Finally, visualization of finite elements is well suited for graphics hardware since it consists a rendering visible elements. After cutting or suturing a volumetric model, it is necessary to update the list of visible facets. Figure 22 shows the real-time cutting of a tetrahedral finite element mesh of a kidney [Cot97]. The kidney has a linear elastic behaviour and an explicit computation scheme prevents the re-computation of the global stiffness matrix. To handle cuts and mesh topology changes, Bro-Nielsen and Cotin combine their method with a particle

system ("mass-tensor" network), where internal forces are formulated with respect to the linear elasticity framework (as opposed to mass-spring networks) [CDA99b]. This work has been extended by Pincibono et al. [PDA03] for handling anisotropy and non-linear elasticity.

Another approach by Debunne et al. [DDCB01] is to use multi-resolution (space/time adaptive sampling) with a weighted finite difference integration technique to approximate the Laplacian and the gradient of the divergence operators of the linear elasticity operator. They use a linearized version of the Green-Lagrange strain (invariant to rigid motion) allowing large deformations. Similarly, a local approximation of the resolution is performed in [OH99], through mass lumping, to reduce computational charge. This is known as *explicit FEM*. All these techniques are encouraging since they try to merge the simplicity of particle systems and the accuracy of finite element methods. However, no method currently allows realistic simulation (anisotropic non-linear viscoelastic materials under large deformations) of large multi-organs complexes (neighbourhood constraints). This is enhanced by the difficulty of validating and comparing such methods.

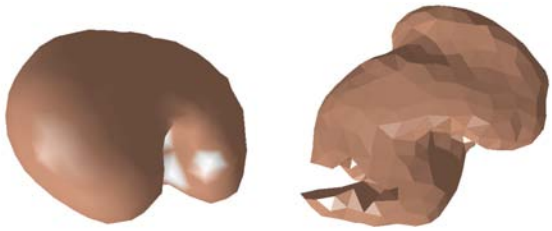


Figure 22: *Cutting of a kidney finite element model.*

2.2.4. Other Deformable Models

Other models of deformable bodies have been proposed in computer graphics. Even if such methods have seldom been applied to medical simulation, they are relevant to the development of real time simulation of soft tissue.

For instance, implicit surfaces defined by potential fields attached to skeletons is a representation well-suited for collision detection computation and the modeling of very soft objects. In [CGD97], Gascuel shows that material stiffness is related to the gradient of potential field and that such models can support cutting or suturing. Other global transformations such as extended free-form deformations [Coq90] or modal analysis [SP91] are potentially of interest, for instance for modeling the effect on abdominal tissues of breathing.

3. Medical visualization and applications

3.1. Introduction

The large amount of data provided by the modern clinical modalities can give a lot of diagnostic information to radiologists and clinicians. Data usually encode maps of physical parameter values in 2D, 3D or 4D and it is not obvious in general, how to show them to the doctors allowing a clear understanding of organs morphology or functionality and eventually obtain quantitative diagnostic values.

To obtain visual representations really helping diagnosis and useful for a variety of new applications like image guided surgery, augmented reality in operating room, surgical training etc. is a big challenge for the Computer Graphics community. In the same way, usability and the easiness of use of new interfaces implementing these applications is another big challenge for experts in Human Computer Interaction.

The simplest data analysis method, still widely applied by radiologists, consists of just showing 2D images (printed or on screen), acquired or resliced, using ad hoc color mapping and windowing to enhance the structures of interest.The intrinsic 3D or 4D nature of most of the data, however, suggests the necessity of finding effective ways of displaying the whole dataset in order to give an immediate interpretation of the organ morphology and functionality.

Standard Computer Graphics techniques, like volume rendering, surface extraction/rendering, can be applied to reach this goal, obtaining photo-realistic representation of organs in 3D. Several visualization enhancements can then be applied to give the physicians useful diagnostic information in addition to the simple morphology. It is possible, for example, to show selectively different image features and to add to the scene results of segmentation procedures, or additional functional parameters related to the morphological information (i.e. pathological region localization, motion information, functional information) and represented using classical visualization methods (color codes, transparency, vectors, streamlines, etc.). The importance of medical data rendering is not, however, limited to a simple diagnostic use. 3D/4D visualizations can be applied in surgical planning, to increase the understanding of organs morphology and or to perform accurate measurements. 3D virtual objects can be also superimposed to real images to provide the physicians with relevant information during surgical procedures (Augmented Reality).

Another relevant application of advanced visualization is related to human body simulation, and to the realization of surgical simulators for medical training. Especially in this last case, the visualization task is particularly critical due to the fact that the model rendered is interactively changed by the user and therefore it is not possible, for example, to use pre processing methods often applied to speed up the rendering process.

Furthermore, diagnostic images are often used to build anatomy teaching tools for students and doctors, requiring user friendly visualization interfaces. For this kind of applications, and also for simple simulations, web based 3D data visualization tools have also been proposed and tested [Joh07]. Standard web3D technologies, like VRML and X3D have been used for the visualization of medical models and also to realize simple clinical applications. The Medical Working Group of the X3D consortium is developing an open interoperable standard for human anatomy representation.

All these specific application of 3D data visualization (pre-operative planning, augmented reality in operating rooms, surgical simulation, teaching, etc.) require not only the application of advanced visualization techniques, but also a careful design of the interaction in order to give the end user the possibility of easily navigating the scene, selecting the views and performing and qualitative and quantitative evaluation.

The section reviews the main techniques and trends in medical visualization and a few example applications. It is organized as follows: subsection 3.2 presents the state of the art of the 3D applications currently available in clinical departments. Subsection 3.3 describes the basic rendering techniques (volume rendering, surface rendering, etc.) used ti display medical data. Time critical visualization, rendering acceleration and large model visualization are discussed as well. Subsection 3.4 shows how a new generation of 3D displays can help the analysis of medical data. Subsection 3.5 presents applications of real time visualization in surgical simulators. Subsection 3.6 discusses generic issues related to performing quantitative measurements on volume data and an example application for measuring reconstructed vascular structures for interventional planning. Subsection 3.7 discusses recent advances in virtual endoscopy techniques and related clinical applications. Subsection 3.8 deals with visualization issues related to medical applications of augmented reality.

3.2. Visualization of diagnostic imaging in clinical centers

CT and MRI scanners are since their introduction computerized digital machines, but for several years (and even now) radiologists used to print the images to sheets of film in order to hang them to the light-boxes. The recent introduction of PACS archives and visualization workstations, however, gives the possibility of displaying 3D information and performing more complex data analysis.

Several advanced tools are already available in clinical environments, even if their use in real applications is still limited.

Medical workstations provided by company like GE, Siemens, Barco/Voxar, Vital Images, Viatronix allow the visualization of 2D slices (with windowing and color mapping options and measurement tools) and several pre-processing and volume reconstruction methods. Different volume rendering methods (see section 3.3.1) are usually implemented, supporting 3D windowing, i.e. the interactive mapping of voxel values to transparency values. 2D analysis can be applied not only on original slices (acquired once, usually in axial direction, but also, with the Multi Planar Reconstruction (MPR), along arbitrary cutting planes (standard coronal, sagittal views, oblique MPR. Also Curved Planar Reconstructions are sometimes available, mapping curved surfaces to viewing planes, in order, for example, to show arteries, or the human spine or teeth. Of course this option requires particular procedures to define the surfaces, usually based on human interaction. MPR can be extremely useful to avoid measurement errors due to tortuosity of vessels and organs. Ideally, it would, of course preferable to perform quantitative evaluation during the 3D data analysis; however the interaction with 3D world, especially if not performed with immersive Virtual Reality environments, presents relevant problems. Problems related on measurements in 3D are discussed briefly in subsection 3.6. Clinical examples of 3D quantitative evaluation in vascular analysis and virtual endoscopy are also described.

Other useful visualization methods available in clinical workstations are those related to multi-modal analysis. Different image registration methods can be applied (see [MV98a]for a review), and relevant problems may arise, especially if markers are not available and nonrigid transformation are required. From the visualization point of view the problem is then the graphical representation of the multiple sources of information (data fusion).

Finally, packages available to clinicians offer specialized packages for a selected number of clinical studies (e.g. Vascular analysis, Virtual Endoscopy, Heart). They implement several advanced visualization methods and segmentation/reconstruction tools (i.e. vessel tracking, surface reconstruction, etc). The potential applications of the 3D visualization and analysis are, however, quite limited by the necessity of specific education and training for the effective use of the 3D segmentation/visualization/analysis systems. Diagnostic applications also require a careful analysis of the medical work-flow and of the user interaction in order to have an effective use of the systems. It is difficult, in fact, to realize easy to use segmentation procedures and it is difficult to implement optimal 3D interaction paradigms, especially if standard monitors and not Virtual Reality immersive displays (see subsection 3.4) are used.

3.3. Medical data rendering techniques

3.3.1. Volume rendering

Volume rendering in general covers the creation of images from volumetric data. These three-dimensional data sets can

be said to represent one or more functions over a three-dimensional space. A volumetric data set can, for example, be created from a stack of images from an image acquisition system, such as a medical Computed Tomography (CT) scanner. This technique is computationally demanding and has only recently begun to be applied in medical visualization tasks. Generally speaking, algorithms for rendering of volume data can be divided into two main categories: Image order and object order approaches. In the past, both of these techniques have proven to be useful for specific applications, though image order approaches seemed to be more generally applicable because the computational complexity scales with image rather than object size, which often makes it a better choice for applications where the exact size and structure of the volume is unknown.

Image order approaches As the name suggests, image order approaches iterate through the parts of the image that should finally be generated and try to find all possible contributions to this part. With these parts being screen pixels most of the time, the algorithm tries to find all objects that change the appearance of a certain pixel. The most widely known image order approach is ray-casting [Lev88], where a ray is cast through each pixel and sampled at regular intervals. The contributing samples of the volume are composited to a final pixel color, which in the end is a (more or less exact) approximation of the volume rendering integral along that ray. Primary advantage of this algorithm is the fact that it is much more dependent on screen resolution than on object size, which also makes it easily scalable by reducing the resolution for quick in-between renders. Also, the number of objects in the volume does not influence rendering speed as much as with object order approaches (this situation changes a little bit with the implementation of techniques like early ray termination and empty space skipping). Obvious disadvantage of image order approaches is that, if no further measures are taken, very sparsely populated volumes will be rendered a lot slower than with object order approaches, because a lot of pixels will be checked (and thus a lot of rays started) that never even hit an object. On the other hand, techniques like early ray termination make sure that in the case of very dense volumes, only those objects are rendered that really contribute to the final image.

Object order approaches In contrast to image order approaches, object order approaches iterate through all parts of the object - in most cases voxels - and determine their contribution to the final image. The most popular object order approach is splatting [LH91, Wes90], where a footprint of the current object is generated and splatted onto the image plane. This works particularly well if there are only very few non-empty voxels inside a large volume. In all other cases, the probability is quite high that a substantial part of the objects will not even be visible in the final image, thus wasting a lot of computational effort. Apart from speed concerns, object order approaches can have a substantial advantage in terms of memory consumption. This is mainly due to the fact that virtually all image based approaches rely on some kind of regular grid to store the data, thus reserving the same amount of memory for empty and non-empty voxels. In the case of object order approaches, usually only non-empty voxels are stored, which can reduce memory demands significantly if the volume is not heavily populated. However, a drawback of this approach is that the appearance of the footprint limits the quality of the final image - zooming in on a splatted image will reveal the structure of the footprint quickly, making the quality of the precomputed kernel essential for good results [Wes90]. Though various extensions of this algorithm have been published [XX06,CRZP04], the quality of rendering is still inferior to image-based approaches in most cases.

Ray-casting fundamentals The most common software ray-casting algorithm, was proposed by Marc Levoy in his initial publication on ray-casting [Lev88]. It divides the process of image generation into six distinct steps:

- preparation of volume densities along a regular grid, resulting in voxel values for each discrete position;
- classification of voxels, mapping each voxel density to a respective opacity value;
- resampling of sample opacities at the discrete sampling positions along the ray;
- shading, mapping each voxel density to a color value;
- resampling of voxel colors at the discrete sampling positions along the ray;
- compositing step, by computing a final pixel color from the vector of shaded samples and respective opacities.

The first step is to prepare the volume densities for further processing, arranging the acquired values at certain positions inside the volume along a regular grid. This step might include correction for nonorthogonal sampling grids, patient motion while scanning or even contrast enhancements, interpolation of additional samples or pre-filtering of noisy data. The output of this step is an array of prepared values which is again used as input for the shading and classification steps. In the case of shading, phong shading is used regularly because it represents a good trade-off between speed and quality. The classification performs the essential step of assigning each voxel a respective opacity value. This opacity value can be a function of various parameters, like voxel density, normal vector direction or gradient magnitude. Standard ray-casting modes would include setting the opacity above a certain threshold to 1. This results in rendering of the first intersection with a value above the threshold along the ray, commonly referred to as iso-surface ray-casting or first-hit-ray-casting. Another common classification strategy is the simple definition of opacities for all density values via a transfer function, resulting in a visualization of translucent tissue that is used primarily for direct volume rendering. Including the normal vector or gradient magnitude into the classification function is primarily used for nonphotorealistic renderings and thus mostly found in specialized ap-

plications where these strategies provide a better insight into certain structures. With shading and classification strategies defined, the actual algorithm is performed by casting rays into the volume and resampling the voxel densities at evenly spaced locations along that ray. The color and opacity values are usually trilinearly interpolated from the eight voxels closest to each sample location. This provides a good trade-off between simple nearest-neighbor interpolation (always take the value from the closest voxel) and more complex filter kernels like tricubic interpolation, that yield better results at higher computational demands.

Finally, these color and opacity values have to be composited to the final pixel color. In order to exploit strategies like early ray termination front-to-back compositing is usually used, starting the ray at the viewing plane and casting rays through every single pixel until a certain alpha threshold near 1 is reached. Front-to-back compositing calculates the summed pixel color by adding further samples according to the following formula:

$$C_{out} = C_{in} + (1 - O_{in}) C_v O_v \quad (7)$$

$$O_{out} = O_{in} + (1 - O_{in}) O_v \quad (8)$$

where C_{in} and O_{in} are the input color and opacity values before adding the current sample, $Cout$ and O_{out} are the output color and opacity values after adding the current sample, and C_v and O_v are the color and opacity values of the sample point (i.e. the result of the trilinear interpolation of classified and shaded samples). Depending on the compositing strategy, different rendering modes can be achieved, as shown in figure 23. For medical visualization, the most common compositing strategies are Maximum Intensity Projection and X-Ray simulation.

Maximum Intensity Projection Maximum Intensity Projection (MIP) is a compositing strategy of direct volume rendering, where, instead of composing optical properties, the maximum value encountered along a ray is used to determine the color of the corresponding pixel [ME05] (see fig. 23).

An important application area of such a rendering mode, are medical datasets obtained by MRI (magnetic resonance imaging) or CT (Computer Tomography) scanners. Such data sets usually exhibit a significant amount of noise that can make it hard to extract meaningful iso-surfaces, or define transfer functions that aid the interpretation. MIP is considered very useful for visualizing angiography data sets, since the data values of vascular structures are higher than the values of the surrounding tissues. The biggest drawback is that intensity channel completely unmasks depth information; so, in normal 2D displays users are not able to disambiguate features and recognize which objects are at the front and which are at the back. Figure 24 shows a rotational angiography

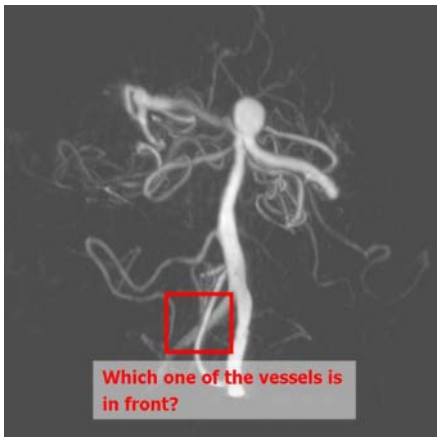


Figure 24: ***Depth oblivious MIP angiography rendering.*** *MIP volume rendering of a rotational angiography scan of a head with aneurysm. The positions and the crossings of vascular structures are not detectable, or wrongly interpreted.*

scan of a head with aneurysm. By using MIP volume rendering technique, only the contrasted vessels are visible. In a 2D view, the positions and the crossings of vascular structures are not detectable, or can be wrongly interpreted. This is because the technique does not provide any depth information in a single image and does not support occlusion. In subsection 3.4 we show how new generation spatial 3D displays can help solving these kind of ambiguities.

X-ray In classical X-ray volume rendering, a viewing ray is cast through the center of each pixel and the line integral of the intensity is evaluated along the given ray [ME04].

Figure 25: ***Depth oblivious X-ray CT rendering.*** *X-ray volume rendering of a CT Scan of abdomen and pelvis. In 2D views, it is not possible to distinguish between front and rear.*

In this case internal parts of the volume are visible, but

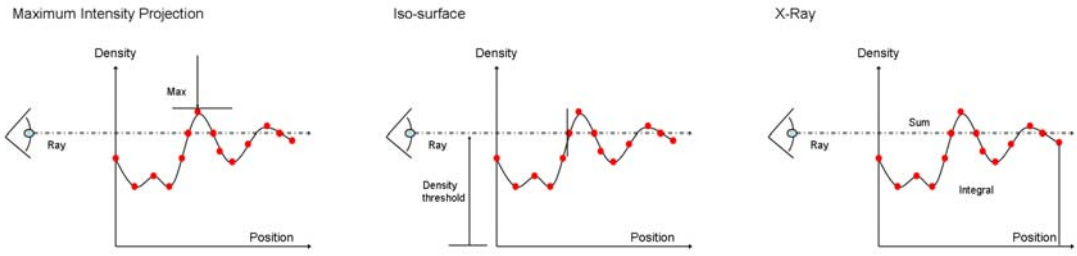


Figure 23: ***Various compositing strategies****. In ray casting different schemes can be chosen, in order to compose color and opacities to final pixel color. Maximum Intensity Projection consists of taking the maximum density along the ray, while iso–surface or first–hit strategy consists of taking the first density above a given threshold along the ray; finally x–ray is obtained by integrating voxel contribution along the ray.*

depth information is not maintained. Hence, view disambiguation tasks are very hard to accomplish using 2D. Figure 25 shows a CT scan of abdomen and pelvis, containing also a stent in the abdominal aorta. In this case no contrast agent was used to enhance blood vessels. A X-ray volume visualization is able to highlight bone and vascular structures, but in 2D views it is not possible to distinguish between front and rear.

Transfer functions Transfer functions are fundamental to direct volume rendering because their role is essentially to make the data visible: by assigning optical properties like color and opacity to the voxel data, the volume can be rendered with traditional computer graphics methods. Good transfer functions reveal the important structures in the data without obscuring them with unimportant regions. Commonly, in medical practice transfer functions are limited to one-dimensional (1D) domains, meaning that the 1D space of scalar data value is used to assign color and opacity. Often, there are features of interest in volume data that are difficult to extract and visualize with 1D transfer functions. Many medical datasets created from CT or MRI scans contain a complex combination of boundaries between multiple materials. This situation is problematic for 1D transfer functions because of the potential for overlap between the data value intervals spanned by the different boundaries. Nowadays, the usage of multi-dimensional transfer functions is considered, but the application to volume rendering is really complicated. Even when the transfer function is only 1D, finding an appropriate transfer function is generally accomplished by trial and error. This is one of the main challenges in making direct volume rendering an effective visualization tool. Levoy first introduced two styles of transfer functions, both two-dimensional, and both using gradient magnitude for the second dimension [Lev88]. One transfer function was intended for the display of interfaces between materials, the other for the display of isovalue contours in more smoothly varying data. Other work involving multi-dimensional transfer functions uses various types of second derivatives in order to distinguish features in the volume according to their shape and curvature characteristics [KKH02].

Accelerating techniques Real-time methods to directly convey the information contents of large volumetric scalar fields are still considered a challenge to the computer graphics community. Based on the observation that the capability of a single general-purpose CPU is not sufficient to achieve interactivity or even real-time for large data sets in general, considerable effort has been spent on the development of acceleration techniques for CPU-based volume rendering and on the design and exploitation of dedicated graphics hardware. Among others, one research direction has lead to volume rendering techniques that exploit hardware assisted texture mapping. Fundamentally, these systems re-sample volume data, represented as a stack of 2D textures or as a 3D texture, onto a sampling surface or so-called proxy geometry [EKE01, RSEB*00]. The most common surface is a plane that can be aligned with the data, aligned orthogonal to the viewing direction, or aligned in other configurations (such as spherical shells). The ability to leverage the embedded trilinear interpolation hardware is at the core of this acceleration technique. Furthermore, ray-casting approaches can be accelerated by employing GPU implementations supporting perspective projection without decreasing performance [KW03]. The entire volume is stored in a single 3D texture, and resampling is performed by fetching samples from this texture with trilinear texture filtering. For each pixel, a fragment program steps from sample location to sample location in a loop and performs compositing until the volume is exited or full opacity is reached. For further information related to GPU based volume rendering techniques, refer to [EHK*04].

Non-photorealistic techniques Historically, most volume rendering techniques are based on an approximation of a realistic physical model. It was noticed, however, that tradi-

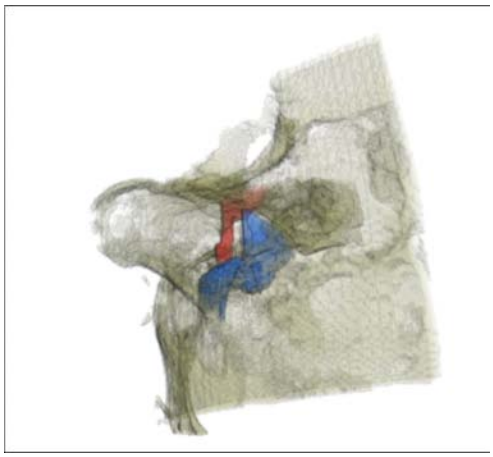


Figure 26: ***Illustrative rendering.*** *Illustrative rendering of temporal bone CT scan dataset. Data courtesy of Prof. Sellari Franceschini, University of Pisa.*

tional depictions of the same types of data as found in medical textbooks, for example, deliberately use non-realistic techniques in order to focus the viewer attention to important aspects. Using abstraction, visual overload is prevented leading to a more effective visualization. Recent approaches have considered this fact, leading to an increased interest in illustrative volume visualization [BG07]. Approaches for illustrative volume visualization frequently employ nonphotorealistic rendering techniques to mimic the style of traditional illustrations [RE01]. They take advantage of the illustrators century-long experience in depicting complex structures in an easily comprehensible way [SES05, LMT*03]. Figure 26 shows an example of illustrative rendering obtained by employing transparent direct volume rendering and exaggerated color map in order to highlight vascularity inside a temporal bone CT scan dataset.

3.3.2. Surface rendering

Surface rendering consists of visualizing data-sets by generating polygonal surfaces representing the anatomical part of interest. This option simplifies the rendering complexity, but requires pre-processing and causes loss of relevant data. Triangulated surface rendering with textures and lighting effects is optimally implemented by the graphics hardware, so the major problems of surface rendering applications are related to the extraction of the polygons representing the model (segmentation) and the simplification of the representation that is necessary when the number of triangles becomes huge (large scale visualization). In medical applications huge surface meshes are not usual; several techniques are however available for mesh simplifications and to build multi-resolution data structures allowing efficient visualization of complex surfaces [CGG*04]. For the creation of surfaces from medical data variants of "marching cubes" algorithm [LC87] are commonly employed. They are able to extract iso-surfaces representing ideally organs boundaries. To recover more accurate and topologically controlled surface representation of organs from diagnostic images, several advanced 3D segmentation techniques are applied (deformable surfaces, front propagation, model based methods), usually specialized for each particular image type and medical task (i.e. vascular segmentation, liver segmentation etc.). Current state of the art in image segmentation techniques is described in section 1 of this tutorial.

3.4. Non-standard visualization on future displays

The representation of 3D scenes on standard 2D monitors limits the possible applications of advanced visualization. This is the reason why 3D displays have been proposed for clinical applications.

Stereo viewers using standard technology (polarized glasses or head mounted displays) are used to improve depth perception in particular medical data visualization tasks and in clinical applications like Virtual Endoscopy. The overhead for the visualization software only consists in the fact that two images should be rendered at interactive frame rates. Some displays use head/eye tracking to automatically adjust the two displayed images to follow the viewer eyes as he moves the head.

Auto-stereoscopic displays do not require googles or head mounted devices. They are usually based on the use of lenticular lenses or parallax barriers that allow the user who keeps his head in a particular viewing positions, to perceive a different image with each eye giving a stereo image. 3D displays of such kind are already available also in commercial medical systems: the SpatialView 3D Medical Workstation offers, in fact an auto-stereoscopic monitor with head tracking coupled with a vision based freehand intuitive gesture based 3D navigation and analysis (SpaceTool). New generations of volume displays are becoming available and are expected to be widely used in the near future. One way to obtain 3D images is to display 2D images on rapidly rotating surfaces. The Perspecta Display by Actuality systems [Fav05], creates 3-D imagery by projecting a sequence of 2x198 images onto a screen rotating at or above 900 rpm. The company offers to research institutions a medical application of the display that is able to visualize DICOM datasets (limited in depth and size) with the 3D method. The LightSpaces Technology DepthcCube uses a similar principle but with a different geometry: different LCD panels stacked in a cube and switching from transparency to images.

A powerful kind of volumetric display can be however realized similarly to auto-stereoscopic screens, but synthesizing with multiple projections 3-D light fields nearly equivalent to volumetric images. Continuous parallax monitors

have been obtained in this way by Holografika (HoloVizio), which realized high resolution real 3D spatial displays allowing users to see a 3D scene moving freely (horizontally) within the viewing angle (up to 70 degrees), and supporting real time operation.

3D spatial display in details The largest display Holografika display [ABF*06] uses an array of 64 XGA projectors and a holographic screen with a diagonal of 72 inches. Smaller screens(32, 26 inches) are also produced.

In these systems, module images generated by the projectors, are determined by geometry and not associated with specific view directions. Each module emits light beams toward a subset of the points of the holographic screen and each point of the holographic screen is hit by more light beams arriving from different modules. The light beams propagate to address fixed spatial positions that are independent from the viewer's position. Many modules contribute to each view of the 3D image, thus no sharp boundary occurs between views, and the display offers continuous and smooth change at different image areas, resulting in a truly 3D experience with continuous horizontal parallax. The holographic screen transforms the incident light beams into an asymmetrical pyramidal form. The cut of this light distribution is a long rectangle, where the vertical size of the rectangle is the vertical field of view, while the horizontal size corresponds to the neighboring emitting directions. This configuration corresponds to the horizontal-only-parallax capability of the display. With the same principles it would to possible to provide vertical parallax as well, at the cost, however, of another order of magnitude increase in data size, rendering times, and system complexity, for little gain in the visual performance in standard settings. The screen is a holographically recorded, randomized surface relief structure that enables high transmission efficiency, controlled angular distribution profile and eliminates moiré and chromatic effects.

The angular light distribution profile introduced by the holographic screen, with a wide plateau and steep Gaussian slopes precisely overlapped in a narrow region, results in a highly selective, low scatter hat-shaped diffuse characteristics. The result is a homogeneous light distribution and continuous 3D view with no visible crosstalk within the field of depth determined by the angular resolution (see figure 27).

Multiple video streams are generated by an array of 16 PCs, connected to the display through four DVI connections each. Each PC runs a server that controls a graphics frame buffer. The server is responsible for generating images associated to a fixed subset of the display rendering modules. In the 72 inches prototype, each PC generates 4 XGA images using two double head NVIDIA boards controlling a 4096x768 frame buffer. In order to support legacy graphics programs and to simplify the development of new holographic applications, an OpenGL compatible front-end has been developed. The front-end runs on a client PC and

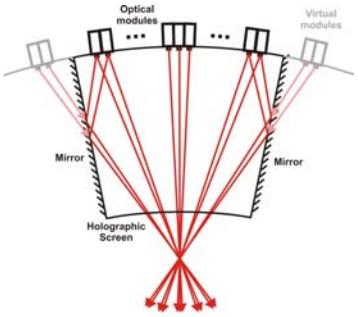


Figure 27: *In the Holovizio display a large number of light beams create a spatial point.*

looks to applications like an ordinary OpenGL library which transparently broadcasts the graphics command stream to the dedicated cluster driving the holographic display. The client PC is connected to the cluster through dual Gbit ethernet links. The back-end servers listen to the network and decode the stream of multicast graphics commands coming from the client. Once decoded, commands are interpreted and sent to the local renderer. The interpretation of the graphics commands involves modifying the way they are generated according to parameters available from the local configuration service, in order to transform the original central view into the view associated with each of the associated optical modules. For each of the optical module views, the graphics commands of the current frame are re-executed, with the following modifications: the original perspective matrix is replaced with a matrix that matches the module's specific position and viewing frustum; a geometrical calibration is performed, to correct nonlinearities in the display/optical geometry; a light calibration is performed to correct the intensity and contrast differences response of the optical modules; an angular resolution correction is performed for depth dependent anti-aliasing.

Volume rendering of diagnostic data on the holographic display New trends in medical visualization research focus on advancing medical visualization by combining 3D rendering techniques with novel spatial 3D displays able to provide all the depth cues exploited by the human visual system [ABGP07].

For example, we can consider as test cases two depth oblivious techniques commonly employed in medical visualization: Maximum Intensity Projection and X-Ray volume rendering. These techniques were originally developed for visualizing images on 2D displays, and have the drawback that 2D views are really ambiguous, since no depth cues are present in the intensity channel, the occlusion is not considered, and shading is not present. The addressed question is whether a 3D display is able to recover lost 3D info, and it was shown that it is possible. Figure 28 shows a MIP ren-

Figure 28: ***MIP angiography rendering.*** *Direct capture from the spatial 3D display taken from different positions in order to disambiguate vascular structures.*

Figure 29: ***X-ray CT rendering.*** *Direct capture from the spatial 3D display taken from different positions in order to immediately understand spatial information.*

dering of an angiography CT scan dataset visualized on the Holovizio spatial display, while figure 29 shows a sequence of pictures taken from direct observation of an X-ray rendering of the CT abdomen and pelvis dataset in the Holovizio spatial 3D display. The set of pictures were taken from different positions in the display workspace, and try to convey part of the visual information provided by the display. The pictures here are merely illustrative, by the way they show that 3D spatial displays are very effective in resolving depth relationships of overlapping, if combined with direct order-indipendent volume rendering approaches (like X-ray and MIP).

An important issue related to the use of immersive 3D displays is that it is necessary to develop new interaction techniques in order to make their use simple and effective. While in simulation tasks the interaction design is simple, because the interaction is designed mimicking the real procedure, generic paradigms for 3D model analysis (that can be based, for example on special devices like spacemouses, gloves or on vision based tracking) did not emerge clearly.

3.5. Application: time-critical rendering for surgical simulation

A real time interactive surgical simulator presents different challenges respect to simple interactive rendering with changing viewpoints. The scene, in fact, changes with time and, even if the frequency constraints are not strict as for the haptic response, it could be difficult to reproduce a reasonable reproduction of the operation scene in real time. As anticipated in the description of basic techniques, specialized approaches can be applied for accelerated volume rendering and surface rendering of dynamically changing data. We describe here two examples of virtual reality training systems employing these techniques: the first is a training system for temporal bone surgery developed at CRS4 in the context of the European project IERAPSI (IST-1999-12175)), while the second is a simulator for the Phaco-emulsification and capsulorhexis tasks in the context of cataract extraction training.

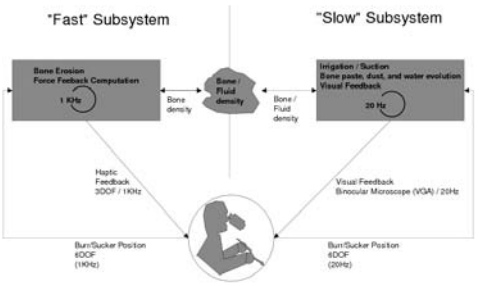


Figure 30: ***Temporal bone system architecture****. The system is divided in a "fast" subsystem, responsible for the high-frequency tasks (surgical instrument tracking, force feedback computation, and bone erosion), and a "slow" one, which is essentially dedicated to the production of data for visual feedback.*

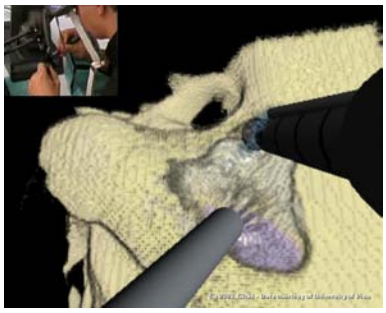


Figure 31: *Snapshot taken from IERAPSI surgical system, where a ENT trainee is performing a mastoidectomy. The inset shows the hardware setup, composed by two PHANToM haptic interfaces and a NVision binocular display*

3.5.1. IERAPSI temporal bone simulator

The IERAPSI prototype virtual reality training system has been developed to simulate completely the temporal bone dissection procedure. Force feedback on haptic devices (Sensable Phantom) is realized using a physically motivated burr-bone interaction model, loosely based on Hertz contact theory, that includes haptic forces evaluation, the bone erosion process and the resulting debris and a multi-resolution representation of the bone characteristic function to adaptively trade simulation quality with speed [AGG*02]. Visual rendering (showing the bone erosion and a particle system based simulation of blood, water and mud) is obtained through a dynamic direct volume rendering technique based on OpenGL register combiners that is able to render shaded representations of dynamically changing rectilinear scalar volume in parallel to simulation threads. The system is based on patient-specific volumetric object models derived from 3D CT and MR imaging data. The performance constraints dictated by the human perceptual system are met by exploiting parallelism via a decoupled simulation approach on a multi-processor PC platform. Specifically, the system is divided in a "fast" subsystem, responsible for the high-frequency tasks (surgical instrument tracking, force feedback computation, and bone erosion), and a "slow" one, which is essentially dedicated to the production of data for visual feedback (see figure 30). The system has been evaluated by experts (Senior ENT Surgeons) and demonstrated realistic haptic and visual rendering, and the ability of simulating complete mastoidectomy procedures [AGG*04]. Figure 31 shows a snapshot taken from IERAPSI surgical system, where a ENT trainee is performing a mastoidectomy. The inset shows the hardware setup, composed by two PHANToM haptic interfaces and a NVision binocular display.

3.5.2. The Cataract surgical simulator

Cataract is a clouding of the eye's natural lens, normally due to natural aging changes, and involving at least half of the population over 65 years. Cataract extraction is the only solution for restoring a clear vision, and nowadays is probably the most frequently practiced surgical procedure. To train surgeons to perform this task, a specific virtual reality simulation system for cataract surgery, involving the capsulorhexis and phaco-emulsification tasks has been developed. The simulator runs on a multiprocessing PC platform and provides realistic physically-based visual simulations of tools interactions. The current setup employs SensAble PHANToM for simulating the interaction devices, and a binocular display for presenting images to the user [AGP*06, AGG*06]. Similarly to IERAPSI temporal bone simulator, the system is divided in a fast subsystem, responsible for the surgical instrument tracking (100 Hz), and a slower one, essentially dedicated to the production of data for visual feedback. The slow subsystem is responsible for the global simulation of the eye, and for the interaction of the devices with the cornea, the anterior camera membrane and the crystalline lens. The algorithms used to control the simulations are local in character, leading naturally to a further break-up of the slow subsystem in components, each dedicated to the generation of a specific visual effect, and thus possibly to a parallel implementation on a multiprocessor architecture. The system runs on two multiprocessor machines connected with a 100 Mbit Ethernet link. The first machine is dedicated to the high-frequency devices tracking tasks (100 Hz), while the second machine concurrently runs the low frequency task (20-25 Hz): eye simulation, anterior camera simulation, and crystalline lens simulation. Since the low-frequency tasks do not influence high-frequency ones, the two machines are synchronized using one-way message passing. Differently from the bone drilling case, since force feedback returned by surgical instruments is here nearly imperceptible and the main way the surgeon have to under-

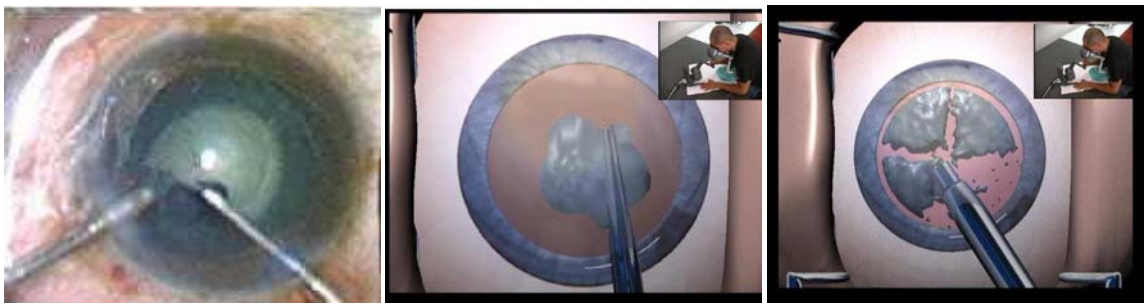


Figure 32: *Comparison between real (left picture) and virtual (center picture) capsulorhexis, and a snapshot from a virtual emulsification (right picture)*

stand the effort really exerted on the patient eye is the visual feedback provided by the tool position related to the environment and the eye globe displacement. This fact enhance the need for a realistic real time visualization. Figure 32 shows a comparison between a picture taken from a real capsulorhexis and a snapshot taken from the simulator (left and center pictures), as well as a snapshot taken from the virtual emulsification task (right picture). The inset shows the hardware setup for the virtual reality training simulator system. In this case, visualized data have no relationships with images acquired on real patients, but accurate reproduction of realistic textures have been superimposed to the dynamic polygonal scene rendered by exploiting hardware acceleration. However, the time critical part of the system is the physically based simulation of the interaction between anatomy and devices during the virtual intervention. The anterior camera for the capsulorhexis was modeled as triangular faceted mesh moving under forces generated by a spring-mass model joining different particles with several effects (following the spring-mass technique described in section 2.2). The emulsification task was simulated by modeling the eye lens as a collection of simplexes built from a tetrahedral mesh. Particles are eroded during the procedure and a shape matching approach [MHTG05b] is used to animate the residual part. The model parameters are tuned by expert surgeons, and overall realism is considered sufficient for training purposes. The system is currently used in specialization courses.

3.6. Model measurement for diagnosis and pre-operative planning

Diagnostic images are not only used for qualitative analysis, but can also be quantitatively measured in order to define the relevance of a pathology or to plan a surgical procedure. Measure can be static (i.e. tumor size, vessel calcification volumes, etc.) or dynamic (i.e. left ventricle ejection fraction) and can be performed on 2D slices or on 3D volumes. The use of 3D measurements is extremely important to avoid errors that are due to slicing, for example on vessel length and sections (fig. 33). The use of dynamic and 3D quantita-

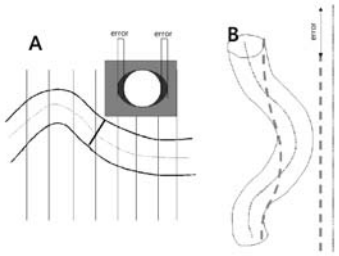


Figure 33: *Large errors are often caused by the use of 2D measurements on wrong planes to evaluate vascular lengths, diameters and tortuosity*

tive measurements is indeed important to improve the quality of diagnosis operative planning, but, in order to introduce such new techniques in the work-flow, two important things should be taken into account:

- A clinical validation of the reconstruction/measurement procedure is required
- The interaction with 3D scenes is not simple and specific interaction modes should be designed to obtain an effective use
- It is often necessary to change the approach of radiologist and surgeons to the particular procedure and make them appreciate the benefits of the new method

Commercial 3D systems are widely used only for very few precise tasks, e.g. intra-cranial or surgery. One of the problems for using 3D quantitative analysis is the difficulty in interacting with 3D scenes.

User interfaces based on the classic WIMP (Windows, Icons, Menus, Pointing devices) paradigm are not well suited for navigating/exploring a 3D scene and a post-WIMP 3D interaction general paradigm has not yet clearly emerged.

3D measurement modes supported by advanced medical workstation modules (i.e Voxar Cardia Metrix, Vessel

Metrix or Colon Metrix) usually limit the interactivity and perform pre defined analysis and visualizations given some input hints in 2D images.

It is not simple to develop, as would be required for the creation of a fully immersive virtual archive environment, user friendly tools avoiding complex user interaction modes, keeping at the same time the advantages of showing 3D paths and spatial relationships between brain areas. Furthermore, it is not easy for all users to move in the 3D world with situational awareness without using immersive virtual environments.

Discussion and general tests on 3D measurement and interaction methods have been presented in recent works. In [PTSP02] 3D measurement tools like distance lines, rulers, angular measurements, interactive volume approximations are presented. In [RSBB06] the interaction style when 3D measurements are performed is discussed and the importance of stereoscopic view and haptic feedback is discussed. Authors present a 3D measurement toolkit developed for the Virtual Liver Surgery Planning and also measured accuracy and time necessary to perform particular tasks (angles, volumes and distances estimation) and conclude that "3D measurements which build an integrated part of a surgical planning environment can be better carried out in a VR environment rather than on a 2D desktop-based application".

3.6.1. Pre-operative measurements with Web3D technology: an example

In the AQUATICS European project (IST 1999-20226), a tool for the web based measurement of Abdominal Aortic Aneurysms morphology has been developed, tested and validated with the help of interventional radiologists. Measurements were performed on a simple 2D web interface with ad hoc solutions making easy to perform 3D length volume and angles measurements without the use of 3D displays. The measurements tool developed have been, however, also modified to be used with the HoloVizio 3D collaborative displays as a test of the new technology.

Endovascular repair of abdominal aortic aneurysms requires a preliminary accurate assessment of patient's specific anatomy. The measurements of the geometry of the aorta is therefore extremely important. The true aneurysm diameter and its growth ratio are fundamental parameters to evaluate the risk of a rupture and to compare it with the risk of a surgical intervention. Fig. 34 shows distances and angles that have to be measured with sufficient precision for the planning of the surgical interventions and the selection/design of prostheses.

Reconstruction of the aortic model The model used for aortic measurement was a complex one defined by a watertight surface mesh representing the vascular lumen, other

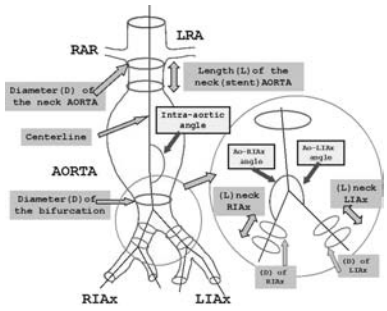


Figure 34: *A large number of measurements are necessary, including the length and diameter of the proximal aortic "neck", the length and diameter of the distal "cuffs", and the length of the two graft limbs. Measurements for an endovascular tube graft are only slightly less complex.*

Figure 35: *Example of skeleton extraction. White line: Shortest path at high resolution. Black line: Final snake-based centerline extracted.*

surface meshes representing calcification and external vascular walls including thrombus and a curve skeleton with a tree structure. The vascular lumen was extracted using a deformable surface algorithm (fast simplex mesh balloon). The user of the reconstruction tool just initialize the surface as a small sphere inside the lumen and the forces applied make it grow to detect the borders. A reparametrization of the surface with topological changes can be obtained using marching cubes in the case the surface becomes auto-intersecting. The curve skeleton extraction is realized with a voxel coding method [ZT99] improved by using a snake-based regularization.

Let us describe shortly the procedure, that is similar to those used to recover centerline paths in Virtual Endoscopy (section 3.7. It is based on the computation of two distance maps (or potential fields), one computing the distance from borders (BSC), and the other from a seed point (SSC). Taking as starting points local maxima of the SSC with high values, shortest paths reaching the seed or the previously extracted branches are found and then centered using the BSC map. In our implementation we also applied a regularization step driven by the BSC map, making the lines centered

and smooth. With this method we obtained results compliant with the requirements: continuous curves connected in a tree structure and locally centered in the volume. Fig. 35 shows an example of the procedure, that was sufficiently fast and reliable for our needs. The external surface of the aorta was recovered by slicing perpendicularly to the vessel direction the data set and using constrained snakes initialized outside the lumen in order to reach the external boundaries. Shape constraints have been introduced due to the presence of tissues with HU value close to the thrombus outside the vessel. To recover calcifications we applied therefore the well known "marching cubes" algorithm [LC87] with a threshold chosen to represent the calcification boundaries. Even in this case, however, it is necessary to customize the procedure and to consider some peculiarities of the problem. Details on different methods to recover vascular trees ad centerlines can be found in [GZ06]

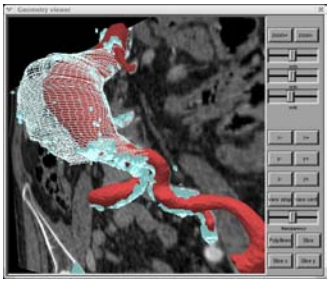


Figure 36: *Hybrid visualization of different vascular components(lumen, thrombus, calcium) superimposed to the CT data.*

Measurement interface Patient specific models built from CT data could then be remotely and collaboratively measured with a web based tool realized with open standards (VRML, External Authoring Interface, now replaceable with X3D and Scene Authoring Interface). Arterial models have been coded in VRML, including Javascript code to support guided navigation, measurement procedures, image visualization, etc. The VRML scene was included in dynamic

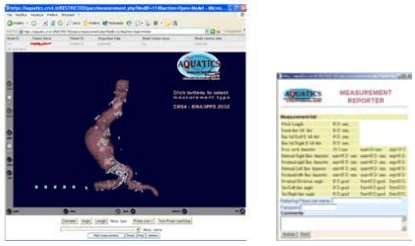


Figure 37: *Example of model visualization and measurements on the web interface.*

Figure 38: *Aneurysm phantom realized at University of Innsbruck and used to validate the aortic measurement system.*

web pages, including the measurement interface, the Electronic Patient record and the measurement reporting interface. The Aquatics Measurement application shows the segmented vessel and give the possibility of interactively hide or show its part, navigate the scene or choose pre defined views and perform 3D measurements of lengths and centerline paths, vessel diameters and volumes. The interface allows the interventional radiologist to perform easily all the quantitative measurements required to design the endovascular prosthesis to be inserted in the aorta 37.

The tool is a good examples of integration of 3D contents in standard web applications. Despite the large diffusion of Web applications, the use of Web 3D standards and of, in general of 3D applications on the web is still limited due to several reasons, including compatibility problems between different browser implementations and platforms. However, the example show the possibility of creating 3D visualization and analysis systems that can be used collaboratively by different remote users to obtain better diagnosis and procedure planning.

Validation The web based aortic measurement tool has been validated in several ways. Different technicians performed reconstructions using the segmentation tool and three clinical specialists used the web interface to measure the parameters for endovascular procedure planning. More than 40 patient specific aortic models have been reconstructed and models of a synthetic phantom 38 have been recovered from CT scans for validation. AQUATICS measurements resulted compatible with phantom's true data and patient data measurements done manually by radiologists using standard methods. The t-test showed a very good correlation between the measurements obtained on phantom with the Aquatics system and the true measurements of the phantom ($p < 0.0001$), demonstrating the reliability of the system. The correlation between different observers and reconstructing operator was also tested with the Spearman rank test and again a statistical significant correlation was proved ($p < 0.0001$), proving the reproducibility of the measurements.

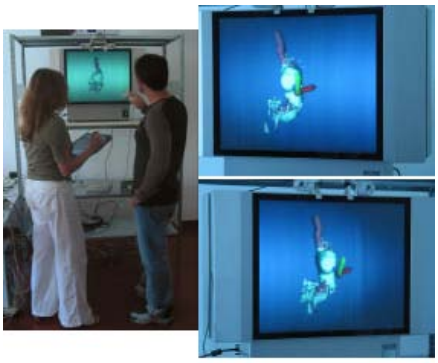


Figure 39: *Left: the collaborative aneurysm analysis session: an operator selects and check measurement types, while the other interacts with the 3D model. Right: Aneurysm rendered on the holographic display, imaged at the same time from two different angles.*

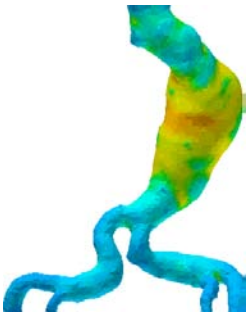


Figure 40: *Wall stress mapped on the aneurysm surface.*

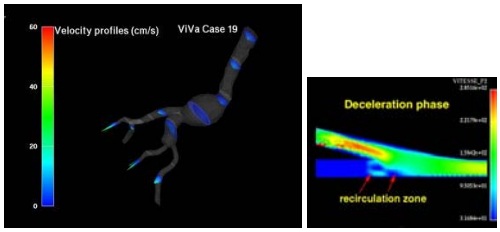


Figure 41: *Blood flow simulation results, displayed on 3D models or on 2D projections with standard methods like profiles, streamlines or color codes, can be used to show anomalous regions correlated with high risk of rupture or plaques formation*

3.6.2. Collaborative measurement/evaluation on the 3D spatial display

The analysis of abdominal aortic aneurysms has been chosen also as a demonstration of the possible collaborative analysis performed with the holographic display. We realized a test application where the measurement interface is implemented as a dynamic web page, and the user can access it through a PDA [BFG*06]. When a patient-specific reconstruction is selected with the touchscreen, the rendering application (messages are passed through wireless connection and HTML protocol) loads the model and display it on the holographic screen. Since objects rendered on the holographic display appear floating in fixed positions, it is possible to naturally interact with them with a 3D user interface that supports direct manipulation in the display space. This is achieved by using tracked 3D cursors manipulated by users. Multiple cursor control interfaces have been developed, using both commercial 3D trackers (Logitech 3D mouse) and custom-made wireless solutions (camera based tracking of pointers, using a wireless USB interface for buttons). With this interface the same measurements for the endovascular repair planning of the AQUATICS application are supported, and a collaborative analysis is made possible.

3.6.3. Diagnostic use of 3D models: Wall stress and blood flow simulations

The enhancement of morphological models is not necessarily limited to selective structural visualization or geometrical measurement support. Even without acquiring functional data through the use of MRI or PET/CT scanners, it is possible to recover functional information with diagnostic relevance through the use of physical simulation. In the vascular field, for example, it is possible to simulate, given geometrical models and material behaviors, the deformation of the vascular boundaries when the blood pressure is applied. Fig. 40 shows the result of a finite element structural simulation showing the wall stress when a blood pressure cycle was applied on a geometry reconstructed for the AQUATICS project.

Figure 41 show results of some test blood flow simulations realized with a Navier Stokes finite element solver developed at CRS4 for the ViVa project [GSG*98]. The solver was parallelized with a domain decomposition technique.

Recent advances in simulation methods should give the possibility of modeling also the fluid-structure interaction in order to reconstruct a more reliable flow field and accurate wall stress estimation. For example in [FLBW06] a turbulent fluid-structure interaction simulation is then run on a commercial solver where the blood is treated as a non-Newtonian fluid. The transient velocity profile at the inlet boundary is taken from Doppler-ultrasound measurements. Other values, like density, Young‘s modulus, and Poisson coefficient, are based on material-specific data that have been adapted during test simulations.

Local simulation can be also coupled wit 1D full body circulation models recently developed and simulating the propagation of pressure and velocity waveforms in 1-D arterial models [KQFL03]. Accurate flow and wall shear stress

evaluation have a relevant diagnostic value, because they have been proved to be an indicator of vessel rupture and plaque formation. Furthermore, pre-operative simulation can be used to chose an optimal geometry for an endovascular prosthesis. In [FLBW06] the simulation results have been used to investigate the pressure and flow patterns in patient specific models of AAAs before and after stent-graft implantation.

3.7. Interaction techniques for medical applications: Virtual Endoscopy

Virtual endoscopy is one of the few well established application where virtual navigation in the 3D acquired data with clinical evaluation and measurement is effectively applied in the clinical work-flow. The basic idea is to simulate the clinical procedure of inserting a catheter inside the tubular structure lumen rendering the scene viewed by the real camera.

Current imaging modalities (i.e. Helical CT scanners, MRI) provide data at a sufficient resolution to make the reconstructed scene sufficiently accurate for several purposes. It should, however, considered that the accuracy of the rendered scene depends also on the quality of the segmentation, often user dependent, and the risk of imaging artifacts should be taken always into account. Different goals can be obtained through the use of Virtual endoscopy, from education, diagnosis, interventional planning and also enhancement of real endoscopy by registering volume data and virtual views with the real endoscopic view. A detailed review on Virtual Endoscopy techniques applied in research and clinical practice is presented in [Bar05].

Applications of Virtual Endoscopy have been mainly developed for the investigation of the colon, blood vessels (virtual angioscopy), of the trachea (virtual bronchoscopy). Similar applications have also been tested for the planing of endonasal interventions.

The advantages of virtual procedures are the non-invasivity, the lower cost, the complete control of lighting and orientation, the absence of access limitations. The real procedure, however, presents the advantage of offering a better resolution, texture information, and the possibility of interacting with tissues and directly perform the actual intervention. Furthermore, it is possible that for virtual procedures requiring a subsequent real intervention planned on their basis, changes happen between the image acquisition and the intervention.

Especially for Virtual Colonoscopy, that already applied in clinical environments, advanced processing techniques offer then several add-ons like the possibility of performing automatic polyp detection and highlighting in colonoscopy, or the possibility of visualizing unfolded surfaces [VWKG01].

While for other applications only research prototypes are available, several commercial systems are available for CT based colonoscopy (Siemens Syngo, Philips EasyVision Endo 3D, Ge Avantage Windows, Viatronix V3D, Vital Images Vitrea, Voxar 3D ColonMetrix, Tiani Jvision).

The examination is extremely useful due to the high incidence of colorectal cancer and the necessity of early detection in order to have successful treatment and the fact that the optical procedure is invasive and expansive. Virtual colonoscopy consists of pumping air into the colon and acquiring multiple high resolution CT scans in different positions to remove the effects of fecal residuals.

Once the elaborated volume are rendered, it is possible to interact with the virtual colon simulating the real procedure using automatic, guided or free navigation along the tubular structure centerline path, usually extracted semi-automatically with algorithms similar to that described in section 3.6.

Different rendering methods have been applied for generic virtual endoscopy. Surface rendering would be the simplest solution for efficiency, but the algorithm has disadvantages such as the low quality of visualized image and the loss of the volume data. All the non-segmented data are not represented, and this means that if the procedure is used for a complex surgical planning activity, all the interested structures must be reconstructed (with difficult, user dependent and time consuming techniques) before the navigation. Clinical validation showed also the superiority of Volume Rendering techniques over Surface Rendering for diagnostic purposes [HLS*03].

Fortunately, algorithmic solutions, like the use of accelerated empty space transversal [WKB99] and technological evolution of current generation of GPUs with advanced pixel shaders allow the realization of GPU-based ray-casting delivering high frame rates and interactivity for either perspective or orthogonal projections. These cards allow the realization of real time perspective volume rendering and hybrid volume/surface representations [SHN*06].

Older systems used also 2D/3D texture mapping for fast rendering, but could not exploit shading effects. Figure 42 show a snapshot of the research virtual endoscopy system developed at CRS4 in 1998, where the visualization was realized in this way [GPZT98]. The system implemented also stereoscopic rendering and measurement tools to quantify vessel stenoses. A particularity of the system was the physically based virtual camera control: we interpreted input from the six-degree-of-freedom device used to control the camera (spaceball) as forces and torques applied to a virtual camera model. A viscous friction force field proportional to volume opacity is used to avoid penetration in opaque areas. This way, the endoscopic camera was always confined in the interior of a cavity. To detect walls during ray-casting, we used an accumulated-opacity algorithm tracing the ray from front to back and stopping when the accumulated opacity was larger than a user settable threshold.

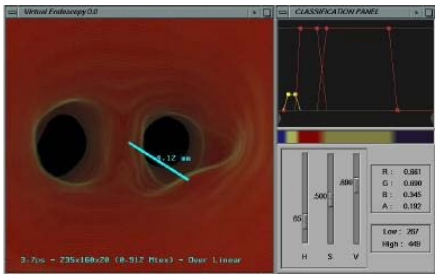


Figure 42: *The Virtual endoscopy prototype realized at CRS4.*

This method is effective and can overcome the difficulty of freely navigating without constraints. Another approach for effective navigation allowing real time interaction and full colon view are use centerline constraint for camera path an free camera orientation. Nonphotorealistic rendering methods can improve the efficacy of Virtual Colonoscopy. For example, to remove false detection of fecal residuals as polyps a recent technique consists of using a translucent rendering mapping colors related to the density over the colon walls, reducing the interpretation time required for the virtual analysis (one of the drawbacks of the 3D analysis) [KLKB05]. Recent studies showed that with Virtual colonoscopy it is possible to measure polyps better than with optical colonoscopy [PCL*07].

3.8. Image guided surgery and augmented reality in operating rooms

Another relevant application of advanced visualization techniques consists in using the 3D data to help the surgeon in performing his tasks. Some surgical procedures, in orthopaedics, or neurosurgery are indeed already planned and executed on the basis of image analysis and reconstruction and commercial system to plan and realize these interventions are available. For example, in neurosurgical tumor resection, surgeons can decide how to access the region through image analysis, and can then use devices guiding the instrument on the basis of the images thanks to the rigid registration of the patients skull with respect to the devices and the image reconstructed volume. Devices can be tracked with magnetic or optical systems. The rigid registration is usually obtained with fiducial markers. Deformations of tissues with respect to the acquired morphology limits the use of such techniques (Brain-shift problem).

In operating room surgeons can also be helped by the use of intra operative imaging (Ultrasound, MRI or angioscopy, C-Arm) These systems are extremely useful, even if the visualization tasks related to them are not particularly challenging for computer scientists.

Different is the case of the clinical use of Augmented Reality. Augmented Reality (AR) consists in overlying the real scene and the virtual image based one, with an interactive real time visualization requiring a perfect registration of the two scene and the tracking of the operator egocentric view. Different kinds of displays are applied for this effect: the most popular are probably head mounted displays (HMD), integrating virtual objects into the real environment in the original field of view of the user. With this technique hand-eye coordination is easy, in particular for keyhole surgery. Drawbacks can be found in the weight and display resolution. Another possibility for in-situ visualization is to use a display panel mounted on a swivel arm and placed between the surgeon and the patient, as for example done in the Medarpa project [SS04]. Accurate tracking of patient user and display position (performed with magnetic or IR systems) is in this case necessary to have an accurate result. If operating microscopes or binoculars are used in the surgical procedure, virtual images can be inserted into the optical paths of the devices and no additional display is necessary. A head mounted see-trough displays like the Varioscope AR [BFH*00] is a generic tool that adds two VGA displays to optical binoculars, and tracking the position of the head can realize the projection transformation (with Tsai's calibration method) of the virtual scene over the real one in real time. In other procedures it is possible to fix cameras and mirrors to the gantry of an imaging device obtaining augmentation without a tracking system. Intra-operative devices which can be augmented this way are endoscopes or laparoscopes used in the majority of minimally invasive surgeries.

Video see-through is an alternative to optical see through that overlay the virtual scene to a video acquisition, registering in this case camera position and orientation with the patient and the data [FNFB04].

The main technological challenge of AR is the accurate real time calibration of the scenes that is necessary to avoid latency or jitter of the virtual objects. Different focus for real world and computer graphics scenes can also cause eye strain and irritation. Due to these problems, augmented reality is still a research field and there are few clinical applications of widespread use. Several initiatives are, however, carried on by research groups to solve problems and to introduce AR in the clinical practice. Among these it is of particular interest the activity of the Marie Curie action ARISER (http://www.ariser.info/), involving eight relevant European partners with strong experience in the field.

3.9. Discussion

We presented a brief description of 3D medical data visualization methods and an overview of example applications. From this overview it appears clear that the application are heterogeneous and the visualiztion methods are quite different according to the different end users. Surgical simulators presents time critical aspects, related to the necessity of re-

alizing real time volume rendering and/or real time tracking and calibration. Other applications usually presents weaker time constraints and their problems rely mainly in their usability, or in the use of new generations of volumetric displays. An interesting remark on the diffusion of volume data in clinical centers is that it created a crisis in the radiology work-flow that requires a complete reorganization of the activity of radiologists and physicians. 3D processing and visualization are necessary to exploit and correctly interpret volume data. The role of technologists in this process can be relevant also in the clinical practice. 3D laboratories where radiologists and experts in image processing and visualization work together are already active (and billing for 3D reconstructions). Radiologists probably prefer to have the full control on the diagnostic process, but the complexity of the 3D reconstruction procedures (reformations, segmentation, visualization) requires new competences and often the use of expensive hardware. Advances in visualization algorithms and new generations of graphic hardware will help to develop easy to use workstations allowing simple 3D analysis and measurements of volume data, moving to the 3D labs only complex procedures like surgical planning [Rid07].

Finally, for the realization of Virtual Humans and physical simulators, that is the main subject of the tutorial, it is expected that the complexity of the reconstructed models will grow relevantly in the future and the large size of the data sets will require specific solution for visualization. Multi resolution schemes have been recently proposed to speed up the rendering of huge data sets on commodity hardware, also for generic triangular meshes [CGG*04]. For large arbitrary surface models, a multiresolution scheme for out-of-core coding and interactive inspection is presented in [GM05]. These kind of techniques, already applied with success in industrial applications, could be used to implement methods for interactive visualization of large and complex human models (static or with pre computed animations).

4. References

References

[ABF*06] AGOCS T., BALOGH T., FORGÁCS T., BETTIO F., GOBBETTI E., ZANETTI G.: A large scale interactive holographic display. In *Proc. IEEE VR 2006 Workshop on Emerging Display Technologies* (Conference Held in Alexandria, VA, USA, March 26 2006, 2006). CD ROM Proceedings.

[ABGD02] ASAKAWA D., BLEMKER S., GOLD G., DELP S.: In vivo motion of the rectus femoris muscle after tendon transfer surgery. *J. of Biomechanics 35*, 8 (2002), 1029–1037.

[ABGP07] AGUS M., BETTIO F., GOBBETTI E., PINTORE G.: Medical visualization with new generation spatial 3d displays. In *Eurographics Italian Chapter Conference* (Conference held in Trento, Italy, February 2007), Eurographics Association. To appear.

[ACP02] ALLEN B., CURLESS B., POPOVIC Z.: Articulated body deformation from range scan data. *ACM Transactions on Graphics 21*, 3 (2002), 612–619.

[AFP00] AUDETTE M., FERRIE F., PETERS T.: An algorithmic overview of surface registration techniques for medical imaging. *Medical Image Analysis 4*, 3 (2000), 201–217.

[AGG*02] AGUS M., GIACHETTI A., GOBBETTI E., ZANETTI G., ZORCOLO A.: Real-time haptic and visual simulation of bone dissection. In *IEEE Virtual Reality Conference* (Feb. 2002), pp. 209–216.

[AGG*04] AGUS M., GIACHETTI A., GOBBETTI E., ZANETTI G., ZORCOLO A.: Hardware-accelerated dynamic volume rendering for real–time surgical simulation. In *Workshop in Virtual Reality Interactions and Physical Simulations (VRIPHYS 2004)* (September 2004). Conference held in Colima, Mexico, September 20-21, 2004.

[AGG*06] AGUS M., GIACHETTI A., GOBBETTI E., ZANETTI G., ZORCOLO A.: Real time simulation of phaco-emulsification for cataract surgery training. In *Workshop in Virtual Reality Interactions and Physical Simulations (VRIPHYS 2006)* (November 2006), Eurographics Association. Conference held in Madrid, Spain, November 6-7.

[AGP*06] AGUS M., GOBBETTI E., PINTORE G., ZANETTI G., ZORCOLO A.: Real-time cataract surgery simulation for training. In *Eurographics Italian Chapter Conference* (Conference held in Catania, Italy, 2006), Eurographics Association.

[AHB87] ARUN K., HUANG T., BLOSTEIN S.: Least-squares fitting of two 3-d point sets. *IEEE Transactions on pattern analysis and machine intelligence 5* (1987), 698–700.

[AT01] AUBEL A., THALMANN D.: Interactive modeling of the human musculature. *Proc. of Computer Animation* (2001).

[AWJ90] AMINI A., WEYMOUTH T., JAIN R.: Using dynamic programming for solving variational problems in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence 12*, 9 (1990), 855–867.

[Bar92] BARAFF D.: Dynamic simulation of non-penetrating rigid bodies. *PhD thesis, Department of Computer Science, Cornell University* (1992).

[Bar05] BARTZ D.: Virtual endoscopy in research and clinical practice. *Computer Graphics Forum 24*, 1 (March 2005), 111–126.

[BCA96] BARDINET E., COHEN L., AYACHE N.: Tracking and motion analysis of the left ventricle with deformable superquadrics. *Medical Image Analysis 1*, 2 (1996), 129–149.

[BD05] BLEMKER S., DELP S.: Three-dimensional representation of complex muscle architectures and geome-

tries. *Annals of Biomedical Engineering 33*, 5 (2005), 661–673.

[Ber96] BERNADOU M.: *Finite element methods for thin shell problems*. Masson, 1996. ISBN 0-471-95647-3/2-225-84940-4.

[BFG*06] BETTIO F., FREXIA F., GIACHETTI A., GOBBETTI E., PINTORE G., ZANETTI G., BALOGH T., FORGACS T., AGOCS T., BOUVIER E.: A holographic collaborative medical visualization system. In *Medicine Meets Virtual Reality 2006* (January 2006), Westwood J. D., (Ed.), IOS, Amsterdam, The Netherlands.

[BFH*00] BIRKFELLNER W., FIGL M., HUBER K., WATZINGER F., WANSCHITZ F., HANEL R., WAGNER A., RAFOLT D., EWERS R., BERGMANN H.: The varioscope ar - a head-monted operating microscope for augmented reality. In *MICCAI '00: Proceedings of the Third International Conference on Medical Image Computing and Computer-Assisted Intervention* (London, UK, 2000), Springer-Verlag, pp. 869–877.

[BG07] BRUCKNER S., GRÖLLER M. E.: Style transfer functions for illustrative volume rendering. *Computer Graphics Forum (accepted for publication) 26*, 3 (Sept. 2007), to be presented at Eurographics 2007.

[Blu67] BLUM H.: A transformation for extracting new descriptors of shape. *Models for the Perception of Speech and Visual Form* (1967).

[BM92] BESL P., MCKAY N.: A method for registration of 3-d shapes. *IEEE Trans. PAMI 14*, 2 (1992), 239–256.

[BN95] BRO-NIELSEN M.: Modelling elasticity in solids using active cubes - application to simulated operations. In *Computer Vision, Virtual Reality and Robotics in Medecine* (Apr. 1995), vol. 905 of *Lecture Notes in Computer Science*, Springer, pp. 535–541.

[BN96] BRO-NIELSEN M.: Medical image registration and surgery simulation. *Ph.D. Thesis, Deptartment of Mathematical Modelling, Technical University of Denmark* (1996).

[BNC96] BRO-NIELSEN M., COTIN S.: Real-time volumetric deformable models for surgery simulation using finite elements and condensation. *Computer Graphics Forum (Proc. of Eurographics Š96)* (1996), C57–C66.

[Bro92] BROWN L.: A survey of image registration techniques. *ACM Computing Surveys 24*, 4 (Dec. 1992), 325–376.

[BW98] BARAFF D., WITKIN A.: Large steps in cloth simulation. *Proc. of SIGGRAPHŠ98, Computer Graphics 32* (1998), 106–117.

[CDA96] COTIN S., DELINGETTE H., AYACHE N.: Real Time Volumetric Deformable Models for Surgery Simulation. In *Vizualisation in Biomedical Computing* (1996), Hohne K., Kikinis R., (Eds.), vol. 1131 of *Lecture Notes in Computer Science*, Springer, pp. 535–540.

[CDA99a] COTIN S., DELINGETTE H., AYACHE N.: A hybrid elastic model allowing real-time cutting, deformations and force-feddback for surgery training and simulation. *Proc. of Computer Animation Š99* (1999).

[CDA99b] COTIN S., DELINGETTE H., AYACHE N.: Real-time elastic deformations of soft tissues for surgery simulation. *IEEE Transactions on Visualization and Computer Graphics 5*, 1 (1999), 62–73.

[CEO93] COVER S. A., EZQUERRA N. F., O'BRIEN J. F.: Interactively Deformable Models for Surgery Simulation. *IEEE Computer Graphics and Applications* (1993), 68–75.

[CGD97] CANI-GASCUEL M.-P., DESBRUN M.: Animation of deformable models using implicit surfaces. *IEEE Transactions on Visualization and Computer Graphics 3*, 1 (Mar. 1997).

[CGG*04] CIGNONI P., GANOVELLI F., GOBBETTI E., MARTON F., PONCHIO F., SCOPIGNO R.: Adaptive TetraPuzzles – efficient out-of-core construction and visualization of gigantic polygonal models. *ACM Transactions on Graphics 23*, 3 (August 2004), 796–803. Proc. SIGGRAPH 2004.

[Chr94] CHRISTENSEN G.: Deformable shape models for anatomy. *Washington University Ph.D. Thesis* (1994).

[CM96] CHINZEI K., MILLER K.: Compression of swine brain tissue ; experiment in vitro. *Journal of Mechanical Engineering Laboratory* (July 1996), 106–115.

[CMD*95] COLLIGNON A., MAES F., DELAERE D., VANDERMEULEN D., SEUTENS P., G.MARCHAL: Automated multimodality image registration using information theory. *in Information Processing in Medical Imaging* (1995), 263–274.

[Coh91] COHEN L.: On active contour models and balloons. *Computer Vision, Graphics, and Image Processing: Image Understanding 53*, 2 (1991), 211–218.

[Coq90] COQUILLART S.: Extended free-form deformation: A sculpturing tool for 3D geometric modeling. In *Computer Graphics (SIGGRAPH '90 Proceedings)* (Aug. 1990), Baskett F., (Ed.), vol. 24, pp. 187–196.

[Cot97] COTIN S.: *Modèles déformables temps-réel pour la simulation de chirurgie avec retour d'effort*. PhD thesis, Université de Nice Sophia-Antipolis, Nice, France, Nov. 1997. in French.

[Cow01] COWIN S.: Bone mechanics handbook, second ed. *Cambridge University Press* (2001).

[CRZP04] CHEN W., REN L., ZWICKER M., PFISTER H.: Hardware-accelerated adaptive ewa volume splatting. In *VIS '04: Proceedings of the conference on Visualization '04* (Washington, DC, USA, 2004), IEEE Computer Society, pp. 67–74.

[CT01] COOTES T., TAYLOR C.: Statistical models of ap-

pearance for medical image analysis and computer vision, in medical imaging. *Proc. of SPIE 4322* (2001), 238–248.

[Cut93] CUTTS A.: Muscle physiology and electromyography. *Mechanics of Human Joints: Physiology, Pathophysiology, and Treatment, ed. by V. Wright, E.L. Radin* (1993).

[CYTT92] CARIGNAN M., YANG Y., THALMANN N. M., THALMANN D.: Dressing animated synthetic actors with complex deformable clothes. In *Computer Graphics (SIGGRAPH '92 Proceedings)* (July 1992), Catmull E. E., (Ed.), vol. 26, pp. 99–104.

[CZ92] CHEN D., ZELTZER D.: Pump it up: Computer animation of a biomechanically based model of muscle using the finite element method. *Computer Graphics (Proc. of SIGGRAPH Š92)* (1992), 89–98.

[DDCB01] DEBUNNE G., DESBRUN M., CANI M.-P., BARR A.: Dynamic real-time deformations using space and time adaptive sampling. *Computer Graphics (Proc. of SIGGRAPH Š01)* (2001).

[DeL97] DELUCA C.: the use of surface electromyography in biomechanics. *J. App. Biomech. 13* (1997), 135–163.

[Del99] DELINGETTE H.: General object reconstruction based on simplex meshes. *International Journal of Computer Vision 32*, 2 (1999), 111–146.

[DHW*04] DENIS K., HUYSMANS T., WILDE T. D., FORAUSBERGER C., RAPP W., HAEX B., SLOTEN J. V., AUDEKERCKE R. V., DER PERRE G. V., HEITMANN K., DIERS H.: A 4d-optical measuring system for the dynamic acquisition of anatomical structures. *Proc. of MICCAI'04 2* (2004), 259–266.

[DKA95] DERICHE R., KORNPROBST P., AUBERT G.: Optical flow estimation while preserving its discontinuities: A variational approach. *Proc. of the 2nd Asian Conference on Computer Vision* (1995), 71–80.

[DKT95] DEUSSEN O., KOBBELT L., TUCKE P.: Using simulated annealing to obtain a good approximation of deformable bodies. In *Proc. Eurographics Workshop on Animation and Simulation* (Maastricht (NL), Sept. 1995), springer.

[DL95] DELP S., LOAN J.: A graphics-based software system to develop and analyze models of musculoskeletal structures. *Comput. Biol. Med. 25*, 1 (1995), 21–34.

[DP94] DRACE J., PELC N.: Measurement of skeletal muscle motion in vivo with phase-contrast mr imaging. *J. of Magnetic Resonance Imaging 4* (1994), 157–163.

[DP02] DEROY K., PEERAER L.: Visualising the pressure distribution on the plantar surface of the foot during stance, walking and running using a plantar pressure measurement system. *the encyclopedia of imaging science and technology, chapter 8* (2002).

[DSB99] DESBRUN M., SHRÖDER P., BARR A.: Interactive animation of structured deformable objects. *Proc. of Graphics Interface* (1999), 1–8.

[DSCP94] DELINGETTE H., SUBSOL G., COTIN S., PIGNON J.: A Craniofacial Surgery Simulation Testbed. In *Visualization in Biomedical Computing (VBC'94)* (octobre 1994).

[DSMK01] DAVATZIKOS C., SHEN D., MOHAMED A., KYRIACOU S.: A framework for predictive modeling of anatomical deformations. *IEEE Trans. Med. Imaging 20*, 8 (2001), 836–843.

[EEH00] EBERHARDT B., ETZMUSS O., HAUTH M.: Implicit-explicit schemes for fast animation with particles systems. *Eurographics workshop on Computer Animation and Simulation* (2000).

[EHK*04] ENGEL K., HADWIGER M., KNISS J. M., LEFOHN A. E., SALAMA C. R., WEISKOPF D.: Real-time volume graphics. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Course Notes* (New York, NY, USA, 2004), ACM Press, p. 29.

[EKE01] ENGEL K., KRAUS M., ERTL T.: High-quality pre-integrated volume rendering using hardwareaccelerated pixel shading, 2001.

[Fav05] FAVALORA G. E.: Volumetric 3d displays and application infrastructure. *Computer 38*, 8 (2005), 37–44.

[FDC01] FENSTER A., DOWNEY D., CARDINAL H.: Three-dimensional ultrasound imaging. *Phys. Med. Biol. 46* (2001), 67–99.

[FHM00] FITZPATRICK J., HILL D., MAURER C.: Image registration, chapter 8. *Handbook of Medical Imaging - Volume 2, Medical Image Processing and Analysis* (2000), 447–513.

[FLBW06] FRAUENFELDER T., LOTFEY M., BOEHM T., WILDERMUTH S.: Computational fluid dynamics: Hemodynamic changes in abdominal aortic aneurysm after stent-graft implantation. *CardioVascular and Interventional Radiology 29* (2006), 613–623.

[FNFB04] FISCHER J., NEFF M., FREUDENSTEIN D., BARTZ D.: Medical Augmented Reality based on Commercial Image Guided Surgery. In *Eurographics Symposium on Virtual Environments (EGVE 2004)* (2004), Coquillart S., Göbel M., (Eds.).

[Fun93] FUNG Y. C.: *Biomechanics - Mechanical Properties of Living Tissues*, second ed. Springer-Verlag, 1993.

[FWG*99] FERRANT M., WARFIELD S., GUTTMANN C., JOLESZ F., KIKINIS R.: 3d image matching using a finite element based elastic deformation model. *Medical Image Computing and Computer Assisted Intervention (MICCAI) 1679* (1999), 202–209.

[GM05] GOBBETTI E., MARTON F.: Far Voxels – a multiresolution framework for interactive rendering of huge

complex 3d models on commodity graphics platforms. *ACM Transactions on Graphics 24*, 3 (August 2005), 878–885. Proc. SIGGRAPH 2005.

[GMMT06] GILLES B., MOCCOZET L., MAGNENAT-THALMANN N.: Anatomical modelling of the musculoskeletal system from mri. *Proc. of MICCAI'06 4190* (2006), 289–296.

[GMTT89] GOURRET J., MAGNENAT-THALMANN N., THALMANN D.: Simulation of object and human skin deformations in a grasping task. *Proc. of SIGGRAPH'89, Computer Graphics 23* (1989), 21–30.

[GPMTV04] GILLES B., PERRIN R., MAGNENAT-THALMANN N., VALLÉE J.-P.: Bones motion analysis from dynamic mri: acquisition and tracking. *Proc. of MICCAI'04 2* (2004), 942–949.

[GPZT98] GOBBETTI E., PILI P., ZORCOLO A., TUVERI M.: Interactive virtual angioscopy. In *Proceedings IEEE Visualization* (Conference held in Research Triangle Park, NC, USA, October 1998), IEEE Computer Society Press, pp. 435–438.

[Gra00] GRAY H.: Anatomy of the human body. *20th edition. Philadelphia: Lea and Febiger 2* (2000).

[GSG*98] G.ABDULAEV, S.CADEDDU, G.DELUSSU, M.DONIZELLI, GIACHETTI L. A., E.GOBBETTI, LEONE A., MANZI C., P.PILI, A.SCHEININE, TUVERI M., A.VARONE, A.VENEZIANI, AND.ZORCOLO G.: Viva: The virtual vascular project. *IEEE trans. on Information Technology in medicine 14*, 1 (1998), 34–48.

[GSM*97] GIBSON S., SAMOSKY J., MOR A., FYOCK C., GRIMSON E., KANADE T., KIKINIS R., LAUER H., MCKENZIE N.: Simulating arthroscopic knee surgery using volumetric object representations, real-time volume rendering and haptic feedback . In *Proceedings of the First Joint Conference CVRMed-MRCAS'97* (Mar. 1997), Troccaz J., Grimson E., Mosges R., (Eds.), vol. 1205 of *Lecture Notes in Computer Science*, pp. 369–378.

[GSZW98] GHANEI A., SOLTANIAN-ZADEH H., WINDHAM J.: A 3d deformable surface model for segmentation of objects from volumetric data in medical images. *Computers in Biology and Medicine 28* (1998), 239–253.

[GZ06] GIACHETTI A., ZANETTI G.: Vascular modeling from volumetric diagnostic data: A review. *Current Medical Imaging Reviews 2*, 4 (November 2006), 415–423.

[HE01] HAUTH M., ETZMUSS O.: A high performance solver for the animation of deformable objects using advanced numerical methods. *Proc. of Eurographics* (2001).

[HFDKR00] HERMENS H., FRERIKS B., DISSELHORST-KLUG C., RAU G.: Development of recommendations for semg sensors and sensor placement procedures. *J Electromyography and Kinesiology 10* (2000), 361–374.

[HFS*01] HIROTA G., FISHER S., STATE A., LEE C., FUCHS H.: An implicit finite element method for elastic solids in contact. *Proc. of Computer Animation* (2001).

[HHD*00] HOLDEN M., HILL D., DENTON E., JAROSZ J., COX T., ROHLFING T., GOODEY J., HAWKES D.: Voxel similarity measures for 3-d serial mr brain image registration. *IEEE Transactions on Medical Imaging 19*, 2 (2000), 94–102.

[Hil44] HILBERT D.: *Geometry and the Imagination*. Dover, 1944.

[HLS*03] HEUSCHMID M., LUZ O., SCHAEFER J. F., STUECKER D., VONTHEIN R., LUBOLDT W., CLAUSSEN C. D., SEEMANN M. D.: Comparison of volume-rendered and surface-rendered mr colonography. *Technology in Cancer Research and Treatment 2*, 1 (2003), 13–17.

[HP78] HOROWITZ S., PAVLIDIS T.: A graph-theoretic approach to picture processing. *CGIP 7*, 2 (1978), 282–291.

[HPP*03] HÖHNE K., PFLESSER B., POMMERT A., PRIESMEYER K., RIEMER M., SCHIEMANN T., SCHUBERT R., TIEDE U., FREDERKING H., GEHRMANN S., NOSTER S., SCHUMACHER U.: Voxel-man 3d navigator: Inner organs. regional, systemic and radiological anatomy. *Springer-Verlag Electronic Media, Heidelberg, ISBN 3-540-40069-9* (2003).

[HS81] HORN B., SCHUNCK B.: Determining optical flow. *Artificial Intelligence 17* (1981), 185–203.

[HS88] HUNTER P., SMAIL B.: The analysis of cardiac function : A continuum approach. *Journal of Prog. Biophysic. molec. Biology 52* (1988), 101–164.

[JL95] JEAN LOUCHET XAVIER PROVOT D. C.: Evolutionary identification of cloth animation model. In *Workshop on Computer Animation and Simulation (Eurographics'95)* (1995), pp. 44–54.

[Joh07] JOHN N. W.: The impact of web3d technologies on medical education and training. *Comput. Educ. 49*, 1 (2007), 19–31.

[JP99] JAMES D., PAI D.: Artdefo: Accurate real time deformable objects. *Proc. of SIGGRAPH Š99, Computer Graphics* (1999).

[Kai96] KAISS, M. AND LE TALLEC, P.: La Modélisation numérique du contact œil-trépan. *Revue Européenne des Eléments Finis 5*, 3 (Feb. 1996), 375–408.

[KdH04] KAPTEIN B., DER HELM F. V.: Estimating muscle attachment contours by transforming geometrical bone models. *Journal of Biomechanics 37* (2004), 263–273.

[KG96] KOCH R., GROSS M.: Simulating facial surgery using finite element models. *Proc. Siggraph '96, Computer Graphics 30* (1996), 421–428.

[KGEB75] KENEDI R., GIBSON T., EVANS J., BARBENEL J.: Tissue mechanics. *Physics in Medecine and Biology 20*, 5 (1975), 163–169.

[KGG96] KEEVE E., GIROD S., GIROD B.: Craniofacial surgery simulation. *Proc. of VBC'96* (1996), 541–546.

[KKH02] KNISS J., KINDLMANN G., HANSEN C.: Multidimensional transfer functions for interactive volume rendering. *IEEE Transactions on Visualization and Computer Graphics 8*, 3 (2002), 270–285.

[KKKN96] KUHN C., KÜHNAPFEL U., KRUMM H.-G., NEISIUS B.: A 'virtual reality' based training system for minimally invasive surgery. In *Proc. Computer Assisted Radiology (CAR '96)* (Paris, June 1996), pp. 764–769.

[KLKB05] KAUFMAN A. E., LAKARE S., KREEGER K., BITTER I.: Virtual colonoscopy. *Commun. ACM 48*, 2 (2005), 37–41.

[KMTM*98] KALRA P., MAGNENAT-THALMANN N., MOCCOZET L., SANNIER G., AUBEL A., THALMANN D.: Real-time animation of realistic virtual humans. *Computer Graphics and Applications 18*, 5 (1998), 42–56.

[KQFL03] KUMAR B. V. R., QUATERONI A., FORMAGGIA L., LAMPONI D.: On parallel computation of blood flow in human arterial network based on 1-d modelling. *Computing 71*, 4 (2003), 321–351.

[KRG*02] KOCH M., ROTH S., GROSS M., ZIMMERMANN A., SAILER H.: A framework for facial surgery simulation. *Proc. of ACM SCCG* (2002).

[KW03] KRUGER J., WESTERMANN R.: Acceleration techniques for gpu-based volume rendering. In *VIS '03: Proceedings of the 14th IEEE Visualization 2003 (VIS'03)* (Washington, DC, USA, 2003), IEEE Computer Society, p. 38.

[KWT88] KASS M., WITKIN A., TERZOPOULOS D.: Snakes: Active contour models. *International Journal of Computer Vision 1* (1988), 321–331.

[Lar86] LARRABEE W.: A finite element model of skin deformation: Biomechanics of skin and tissue : a review. *Laryngoscope 96* (1986), 399–405.

[LC87] LORENSEN W., CLINE H.: Marching cubes: A high resolution 3d surface construction algorithm. *Proc. of SIGGRAPH'87, Computer Graphics 21*, 39 (1987), 163–169.

[Lev88] LEVOY M.: Display of surfaces from volume data. *IEEE Comput. Graph. Appl. 8*, 3 (1988), 29–37.

[LH91] LAUR D., HANRAHAN P.: Hierarchical splatting: a progressive refinement algorithm for volume rendering. In *SIGGRAPH '91: Proceedings of the 18th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1991), ACM Press, pp. 285–288.

[LJF*91] LUCIANI A., JIMENEZ S., FLORENS J., CADOZ C., RAOULT O.: Computational physics: a modeler simulator for animated physical objects. In *Eurographics Workshop on Animation and Simulation* (Vienna, 1991), pp. 425–437.

[LM01] LÖTJÖNEN J., MÄKELÄ T.: Segmentation of mr images using deformable models: Application to cardiac images. *International Journal of Bioelectromagnetism 3*, 2 (2001), 37–45.

[LMT*03] LU A., MORRIS C. J., TAYLOR J., EBERT D. S., HANSEN C., RHEINGANS P., HARTNER M.: Illustrative interactive stipple rendering. *IEEE Transactions on Visualization and Computer Graphics 9*, 2 (2003), 127–138.

[LO99] LU T., O'CONNOR J.: Bone position estimation from skin marker co-ordinates using global optimisation with joint constraints. *J. of Biomechanics 32* (1999), 129–134.

[LRMK99] LÖTJÖNEN J., REISSMAN P., MAGNIN I., KATILA T.: Model extraction from magnetic resonance volume data using the deformable pyramid. *Medical Image Analysis 3*, 4 (1999), 387–406.

[Mar96] MARCUS B.: Hands on : Haptic feedback in surgical simulation. In *Proc. of Medecine Meets Virtual Reality IV (MMVR IV)* (San Diego, CA, Jan. 1996), pp. 134–139.

[MC94] MIRTICH B., CANNY J.: Impulse-based dynamic simulation. *The Algorithmic Foundations of Robotics* (1994).

[MC97] MESEURE P., CHAILLOU C.: Deformable Body Simulation with Adaptative Subdivision and Cuttings. In *Proceedings of the WSCG'97* (Pizen (Czech Republic), Feb. 1997), pp. 361–370.

[McM87] MCMAHON T.: Muscle mechanics. *Handbook of Bioengineering 8*, 2 (1987), 187–203.

[MD05] MONTAGNAT J., DELINGETTE H.: 4d deformable models with temporal constraints: application to 4d cardiac image segmentation. *Medical Image Analysis 9*, 1 (2005), 87–100.

[ME04] MORA B., EBERT D. S.: Instant volumetric understanding with order-independent volume rendering. In *Eurographics 2004* (2004), vol. 23.

[ME05] MORA B., EBERT D. S.: Low-complexity maximum intensity projection. *ACM Transactions on Graphics, 24*, 4 (October 2005), 1392–1416.

[Meg96] MEGLAN D.: Making surgical simulation real. In *Computer Graphics* (Nov. 1996), Addison Wesley, pp. 37–39.

[Men00] MENACHE A.: Understanding motion capture for computer animation and video games. *Academic Press, San Diego* (2000).

[MET*04] MORI S., ENDO M., TSUNOO T., KANDATSU S., TANADA S., ARADATE H., SAITO Y., MIYAZAKI H., SATOH K., MATSUSHITA S., KUSAKABE M.: Physical performance evaluation of a 256-slice ct-scanner for four-dimensional imaging. *Medical Physics 31*, 6 (2004), 1348–1356.

[MG04] MÜLLER M., GROSS M.: Interactive virtual materials. *Proc. of Graphics Interface'04* (2004), 239–246.

[MH06] MCINTOSH C., HAMARNEH G.: Genetic algorithm driven statistically deformed models for medical image segmentation. *Proc. of MedGEC and GECCO* (2006).

[MHHR06] MÜLLER M., HEIDELBERGER B., HENNIX M., RATCLIFF J.: Position based dynamics. *Proc. of Virtual Reality Interactions and Physical Simulations (VRIPhys)* (2006), 6–7.

[MHTG05a] MÜLLER M., HEIDELBERGER B., TESCHNER M., GROSS M.: Meshless deformations based on shape matching. *Proc. of SIGGRAPH'05* (2005), 471–478.

[MHTG05b] MULLER M., HEIDELBERGER B., TESCHNER M., GROSS M.: Meshless deformations based on shape matching. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers* (New York, NY, USA, 2005), ACM Press, pp. 471–478.

[MKLA80] MOW V., KUEI S., LAI W., ARMSTRONG C.: Biphasic creep and stress relaxation of articular cartilage: Theory and experiments. *ASME J Biomech Eng 102* (1980), 73–84.

[Mr05] MOSEGAARD J., RENSEN T. S.: Gpu accelerated surgical simulators for complex morphology. *Proc. of IEEE Virtual Reality '05* (2005), 147–153.

[MRF*96] MARK W., RANDOLPH S., FINCH M., VAN VERTH J., TAYLOR II R. M.: Adding force feedback to graphics systems: Issues and solutions. In *SIGGRAPH 96 Conference Proceedings* (Aug. 1996), Rushmeier H., (Ed.), Annual Conference Series, ACM SIGGRAPH, Addison Wesley, pp. 447–452.

[MRL*96] MUTHUPILLAI R., ROSSMAN P., LOMAS D., GREENLEAF J., RIEDERER S., EHMAN R.: Magnetic resonance imaging of transverse acoustic strain waves. *Magn. Reson. Med 36* (1996), 266–274.

[MT95] MCINERNEY T., TERZOPOULOS D.: A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4d image analysis. *Comp. Med. Imag. Graph. 19*, 1 (1995), 69–83.

[MT96] MCINERNEY T., TERZOPOULOS D.: Deformable models in medical image analysis: a survey. *Medical Image Analysis 1*, 2 (1996), 91–108.

[MT99] MCINERNEY T., TERZOPOULOS D.: Topology adaptive deformable surfaces for medical image volume segmentation. *IEEE Transactions on Medical Imaging 18*, 10 (1999), 840–850.

[MV98a] MAINTZ J., VIERGEVER M.: A survey of medical image registration. *Medical Image Analysis 2*, 1 (April 1998), 1–36.

[MV98b] MAINTZ J., VIERGEVER M.: A survey of medical registration. *Medical image analysis 2*, 1 (1998), 1–36.

[MZ02] MORI S., ZIJL P. V.: Fiber tracking: principles and strategies, a technical review. *NMR Biomed 15* (2002), 468–480.

[NLM99] NLM T. V. H. P.: National library of medicine, http://www.nlm.nih.gov/research/visible/visible_human.html.

[NM65] NELDER J., MEAD R.: A simplex method for function minimization. *Computer Journal 7* (1965), 308–313.

[NMK*03] NAVA A., MAZZA E., KLEINERMANN F., AVIS N. J., MC-CLURE J.: Determination of the mechanical properties of soft human tissues through aspiration experiments. *Proc. of MICCAI'03 1* (2003), 222–229.

[NMK*05] NEALEN A., MÜLLER M., KEISER R., BOXERMAN E., CARLSON M.: Physically based deformable models in computer graphics. *In Eurographics State of-the-Art Report (EG-STAR) 3* (2005), 71–94.

[NTH00] NG-THOW-HING V.: Anatomically-based models for physical and geometric reconstruction of humans and other animals. *Ph.D. Thesis, Department of Computer Science, University of Toronto* (2000).

[OH99] OŠBRIEN J., HODGINS J.: Graphical models and animation of brittle fracture. *Proc. of SiggraphŠ99, Computer Graphics* (1999), 137–146.

[OS88] OSHER S., SETHIAN J.: Fronts propagating with curvature dependent speed: Algorithms based on hamilton-jacobi formulation. *Journal of Computational Physics 79* (1988), 12–49.

[PBHS05] PAPAZOGLOU S., BRAUN J., HAMHABER U., SACK I.: Two-dimensional waveform analysis in mr elastography of skeletal muscles. *Phys. Med. Biol. 50* (2005), 1313–1325.

[PCL*07] PARK S. H., CHOI E. K., LEE S. S., BYEON J.-S., JO J.-Y., KIM Y. H., LEE K. H., HA H. K., HAN J. K.: Polyp measurement reliability, accuracy, and discrepancy: Optical colonoscopy versus ct colonography with pig colonic specimens. *Radiology online* (2007).

[PDA03] PICINBONO G., DELINGETTE H., AYACHE N.: Non-linear anisotropic elasticity for real-time surgery simulation. *Graph. Models 65*, 5 (2003), 305–321.

[PFJ*03] PIZER S., FLETCHER P., JOSHI S., THALL A., CHEN J., FRIDMAN Y., FRITSCH D., GASH A., GLOTZER J., JIROUTEK M., LU C., MÜLLER K., TRACTON G., YUSHKEVICH P., CHANEY E.: Deformable m-reps for 3d medical image segmentation. *International Journal of Computer Vision 55*, 2 (2003), 85–106.

[PHSE91] PELC N., HERFKENS R., SHIMAKAWA A., ENZMANN D.: Phase contrast cine magnetic resonance imaging. *Magnetic Resonance Quarterly 7* (1991), 229–254.

[PM90] PERONA P., MALIK J.: Scale-space and edge detection using anisotropic diffusion. *IEEE trans. on Pattern Analysis and Machine Intelligence 12*, 7 (1990), 629–639.

[PMTK01] PARK J., MCINERNEY T., TERZOPOULOS D., KIM M.: A non-self-intersecting adaptive deformable surface for complex boundary extraction from volumetric images. *Computers and Graphics 25*, 3 (2001), 421–440.

[Pro95] PROVOT X.: Deformation constraints in a mass-spring model to describe rigid cloth behavior. In *Graphics Interface '95* (May 1995), Davis W. A., Prusinkiewicz P., (Eds.), Canadian Information Processing Society, Canadian Human-Computer Communications Society, pp. 147–154. ISBN 0-9695338-4-5.

[PRZ92] PIEPER S., ROSEN J., ZELTZER D.: Interactive graphics for plastic surgery: A task-level analysis and implementation. *Computer Graphics (1992 Symposium on Interactive 3D Graphics) 25*, 2 (Mar. 1992), 127–134.

[PTSP02] PREIM B., TIETJEN C., SPINDLER W., PEITGEN H. O.: Integration of measurement tools in medical 3d visualizations. In *VIS '02: Proceedings of the conference on Visualization '02* (Washington, DC, USA, 2002), IEEE Computer Society, pp. 21–28.

[PTVF92] PRESS W., TEUKOLSKY S., VETTERLING W., FLANNERY B.: Numerical recipies in c (2nd ed.). *Cambridge University Press* (1992).

[QLH*02] QUICK H., LADD M., HOEVEL M., BOSK S., DEBATIN J., LAUB G., SCHROEDER T.: Real-time mri of joint movement with truefisp. *Journal of Magnetic Resonance Imaging 15*, 6 (2002), 710–715.

[RAD03] ROHDE G., ALDROUBI A., DAWANT B.: The adaptive bases algorithm for intensity-based nonrigid image registration. *IEEE Trans. Med. Imag. 22* (2003), 1470–1479.

[RE01] RHEINGANS P., EBERT D.: Volume illustration: Nonphotorealistic rendering of volume models. *IEEE Transactions on Visualization and Computer Graphics 7*, 3 (2001), 253–264.

[RFS99] ROHR K., FORNEFETT M., STIEHL H.: Approximating thin-plate splines for elastic registration: Integration of landmark errors and orientation attributes. *Lecture Notes in Computer Science 1613* (1999), 252–265.

[Rid07] RIDLEY E. L.: Radiologists may be better off taking hands-on approach to 3d workflow. AuntMinnie.com, 2007.

[RMA00] ROCHE A., MALANDAIN G., AYACHE N.: Unifying maximum likelihood approaches in medical image registration. *International Journal of Imaging Systems and Technology: Special Issue on 3D Imaging 11*, 1 (2000), 71–80.

[RRF96] R.J.STONE, R. F M.: Virtual environment training systems for laparoscopic surgery; activities at the uk's wolfson centre for minimally invasive therapy. *The journal of Medicine and Virtual Reality 1*, 2 (1996), 42–51.

[RSBB06] REITINGER B., SCHMALSTIEG D., BORNIK A., BEICHEL R.: Spatial analysis tools for virtual reality-based surgical planning. In *3DUI '06: Proceedings of the 3D User Interfaces (3DUI'06)* (Washington, DC, USA, 2006), IEEE Computer Society, pp. 37–44.

[RSEB*00] REZK-SALAMA C., ENGEL K., BAUER M., GREINER G., ERTL T.: Interactive volume on standard pc graphics hardware using multi-textures and multi-stage rasterization. In *HWWS '00: Proceedings of the ACM SIGGRAPH/EUROGRAPHICS workshop on Graphics hardware* (New York, NY, USA, 2000), ACM Press, pp. 109–118.

[RSH*99] RUECKERT D., SONODA L., HAYES C., HILL D., LEACH M., HAWKES D.: Nonrigid registration using free-form deformations: Application to breast mr images. *IEEE Trans. Med. Imag. 18* (1999), 712–721.

[RSS*96] ROHR K., STIEHL H., SPRENGEL R., BEIL W., BUZUG T. M., WEESE J., KUHN M.: Point-based elastic registration of medical image data using approximating thin-plate splines. *Visualization in Biomedical Computing* (1996), 297–306.

[SBM*94] SAGAR M. A., BULLIVANT D., MALLINSON G., HUNTER P., HUNTER I.: A Virtual Environment and Model of the Eye for Surgical Simulation. In *Computer Graphics (SIGGRAPH'94)* (1994), Annual Conference Series, pp. 205–212.

[SC04] SHELLOCK F., CRUES J.: Mr procedures: Biologic effects, safety, and patient care. *Radiology 232* (2004), 635–652.

[SD92] STAIB L., DUNCAN J.: Deformable fourier models for surface finding in 3d images. *Proc. of Visualization in Biomedical Computing (VBC)* (1992), 90–104.

[SES05] SVAKHINE N., EBERT D. S., STREDNEY D.: Illustration motifs for effective medical volume illustration. *IEEE Comput. Graph. Appl. 25*, 3 (2005), 31–39.

[SGT98] SINGH A., GOLDGOF D., TERZOPOULOS D.: Deformable models in medical image analysis. *IEEE Computer Society* (1998).

[She94] SHEWCHUK J.: An introduction to the conjugate gradient method without the agonizing pain. *Technical report, Carnegie Mellon University* (1994).

[SHN*06] SCHARSACH H., HADWIGER M., NEUBAUER A., WOLFSBERGER S., BUHLER K.: Perspective isosurface and direct volume rendering for virtual endoscopy applications. In *EUROVIS - Eurographics /IEEE VGTC Symposium on Visualization* (2006), pp. 315–322.

[SL96] SZELISKI R., LAVALLÉE S.: Matching 3-d anatomical surfaces with non-rigid deformations using octree-splines. *J. Comp. Vision 18* (1996), 171–186.

[SP86] SEDERBERG T., PARRY S.: Free form deformations of solid geometric models. *Proc. of SIGGRAPH'86, Computer Graphics 20*, 4 (1986), 150–161.

[SP91] SCLAROFF S., PENTLAND A.: Generalized implicit functions for computer graphics. In *Computer Graphics (SIGGRAPH '91 Proceedings)* (July 1991), Sederberg T. W., (Ed.), vol. 25, pp. 247–250.

[SR95] SONG G. J., REDDY N. P.: Tissue Cutting In Virtual Environment. In *Medecine Meets Virtual Reality IV* (1995), IOS Press, pp. 359–364.

[SS04] SCHWALD B., SEIBERT H.: Registration tasks for a hybrid tracking system for medical augmented reality, 2004.

[TBHF03] TERAN J., BLEMKER S., HING V. N. T., FEDKIW R.: Finite volume methods for the simulation of skeletal muscle. *Proc. of Eurographics/SIGGRAPH Symposium on Computer Animation* (2003).

[TBP03] TSAO J., BOESIGER P., PRUESSMANN K.: k-t blast and k-t sense: dynamic mri with high frame rate exploiting spatiotemporal correlations. *Magnetic Resonance in Medicine 50*, 5 (2003), 1031–1042.

[Ter86] TERZOPOULOS D.: Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1986), 413–424.

[TF88a] TERZOPOULOS D., FLEISCHE K.: Deformable models. *The Visual Computer 4* (1988), 306–311.

[TF88b] TERZOPOULOS D., FLEISCHER K.: Deformable models. *The Visual Computer 4*, 6 (1988), 306–331.

[TF88c] TERZOPOULOS D., FLEISCHER K.: Modeling inelastic deformation: Viscoelasticity, plasticity, fracture. *Computer Graphics (SIGGRAPH'88) 22*, 4 (1988), 269–278.

[Thi95] THIRION J.: Fast non-rigid matching of 3d medical images. *INRIA Technical report 2547* (1995).

[TKH*05] TESCHNER M., KIMMERLE S., HEIDELBERGER B., ZACHMANN G., RAGHUPATHI L., FUHRMANN A., CANI M.-P., FAURE F., MAGNETAT-THALMANN N., STRASSER W., VOLINO P.: Collision detection for deformable objects. *Computer Graphics Forum 24*, 1 (2005), 61–81.

[TKZ*04] TESCHNER M., KIMMERLE S., ZACHMANN G., HEIDELBERGER B., RAGHUPATHI L., FUHRMANN A., CANI M.-P., FAURE F., MAGNENAT-THALMANN N., STRASSER W., VOLINO P.: Collision detection for deformable objects. *Proc. of Eurographics State-of-the-Art Report* (2004), 119–135.

[TM91] TERZOPOULOS D., METAXAS D.: Dynamic 3d models with local and global deformations: Deformable superquadrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence archive 13*, 7 (1991), 703–714.

[TPBF87] TERZOPOULOS D., PLATT J., BARR A., FLEISCHER K.: Elastically deformable models. *Computer Graphics 21*, 4 (1987), 205–214.

[TSB*05] TERAN J., SIFAKIS E., BLEMKER S., HING V. N. T., LAU C., FEDKIW R.: Creating and simulating skeletal muscle from the visible human data set. *IEEE TVCG 11* (2005), 317–328.

[TW88] TERZOPOULOS D., WITKIN A.: Physically based models with rigid and deformable components. *IEEE Computer Graphics and Applications 8*, 6 (1988), 41–51.

[TWK87a] TERZOPOULOS D., WITKIN A., KASS M.: Symmetry-seeking models and 3d object reconstruction. *International Journal of Computer Vision 1* (1987), 211–221.

[TWK87b] TERZOPOULOS D., WITKIN A., KASS M.: Symmetry-seeking models for 3d object reconstruction. *International Journal of Computer Vision 1(3)* (1987), 211–221.

[VGW06] VERESS A., GULLBERG G., WEISS J.: Measurement of strain in the left ventricle during diastole with cine-mri and deformable image registration. *ASME Journal of Biomechanical Engineering* (2006).

[Vio95] VIOLA P.: Alignment by maximization of mutual information. *Ph.D. thesis, Massachusetts Inst. Technol.* (1995).

[VM04] VALTORTA D., MAZZA E.: Dynamic measurements of soft tissue viscoelastic properties with a torsional resonator device. *Proc. of MICCAI'04 2* (2004), 284–292.

[VMT00] VOLINO P., MAGNENAT-THALMANN N.: Implementing fast cloth simulation with collision response. *Proc. of the International Conference on Computer Graphics* (2000), 257–268.

[VMT05] VOLINO P., MAGNENAT-THALMANN N.: Implicit midpoint integration and adaptive damping for efficient cloth simulation: Collision detection and deformable objects. *Computer Animation and Virtual Worlds 16* (2005), 163–175.

[VWKG01] VILANOVA BARTROLÍ A., WEGENKITTL R., KÖNIG A., GRÖLLER E.: Virtual colon unfolding, 2001.

[Wat92] WATERS K.: A physical model of facial tissue and muscle articulation derived from computer tomography data. In *Visualization in Biomedical Computing (VBC'92)* (Chappel Hill, NC, 1992), vol. 574.

[WC03] WILSSENS J., COCK A. D.: Barefoot, insole and ground pressure measurement comparison during 5m/s running. *Proc. of the 6th Symposium on footwear biomechanics* (2003).

[WCM92] WOODS R., CHERRY S., MAZZIOTTA J.: Rapid automated algorithm for aligning and reslicing pet images. *J. Comp. Assisted Tomogr. 16*, 4 (1992), 620–633.

[Wes90] WESTOVER L.: Footprint evaluation for volume rendering. In *SIGGRAPH '90: Proceedings of the 17th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1990), ACM Press, pp. 367–376.

[WG01] WEISS J., GARDINER J.: Computational modeling of ligament mechanics. *Critical Reviews in Biomedical Engineering 29*, 4 (2001), 1–70.

[WGW90] WITKIN A., GLEICHER M., WELCH W.: Interactive dynamics. *Computer Graphics 24*, 2 (1990), 11–22.

[WHQ05] WANG K., HE Y., QIN H.: Incorporating rigid structures in non-rigid registration using triangular b-splines. *VLSM* (2005), 235–246.

[WKB99] WAN M., KAUFMAN A., BRYSON S.: High performance presence-accelerated ray casting. In *VIS '99: Proceedings of the conference on Visualization '99* (Los Alamitos, CA, USA, 1999), IEEE Computer Society Press, pp. 379–386.

[WS00] WANG Y., STAIB L.: Physical model-based non-rigid registration incorporating statistical shape information. *Medical Image Analysis 4* (2000), 7–20.

[XX06] XIAO H., XIAO D.-G.: An accelerating splatting algorithm based on multi-texture mapping for volume rendering. In *GRAPHITE '06: Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia* (New York, NY, USA, 2006), ACM Press, pp. 205–208.

[YSAT01] YOU B., SIY P., ANDERST W., TASHMAN S.: In vivo measurement of 3-d skeletal kinematics from sequences of biplane radiographs: Application to knee kinematics. *IEEE Transactions on medical imaging 20*, 6 (2001), 514–524.

[Zie77] ZIENKIEWICZ O.: The finite element method. *McGraw-Hill, London, 3rd edition* (1977).

[ZPR*88] ZERHOUNI E., PARISH D., ROGERS W., YANG A., SHAPIRO E.: Human heart: Tagging with mr imaging, a method for noninvasive assessment of myocardial motion. *Radiology 169* (1988), 59–63.

[ZT99] ZHOU Y., TOGA A. W.: Efficient skeletonization of volumetric objects. *IEEE Transactions on Visualization and Computer Graphics 5*, 3 (/1999), 196–209.

EG:643

Inverse Kinematics and Kinetics for Virtual Humanoids

Eurographics 2007 Tutorial T4

Organizers & Speakers

Ronan Boulic (EPFL)
Richard Kulpa (M2S)

EG:644

Inverse kinematics and kinetics for virtual humanoids

R. Boulic[1] and R. Kulpa[2,3]

[1]VRLab - EPFL - Switzerland
[2]Bunraku - IRISA/INRIA Rennes - France
[3]M2S - University of Rennes 2 - France

Abstract

The present tutorial deals with the use of inverse kinematics and kinetics for postural adaptations of virtual humanoids to different kind of constraints. It first proposes an overview of the problematic of this thematic and then of the existing techniques. Then it technically describes two key approaches: the prioritized inverse kinematics for accurate and realistic adaptation and a CCD-like algorithm based on groups for fast and realistic adaptation of hundreds of characters in real-time.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Animation

1. Keywords

Inverse kinematics, inverse kinetics, postural adaptation, numeric IK, prioritized IK, validity domain of the linearization, joint limits, singularity, pseudo-inverse, kernel, jacobian, balance control

2. Information about the tutorial

2.1. Contribution

Intuitive and efficient postural control is important to master, so this tutorial offers first an overview and a classification of the inverse kinematics and kinetics methods and of the respective advantages/drawbacks of various approaches (analytic, iterative, Jacobian transpose, pseudo-inverse, CCD...). Then it intuitively details the technical aspects of what make two key approaches very useful and how to implement them:

- the prioritized Inverse Kinematics can combine multiple conflicting constraints (useful for interaction in complex environments)
- the CCD combines both IK and balance control very efficiently (useful for large number of virtual humans)

2.2. Benefit for the audience

The tutorial is illustrated with abundant conceptual illustrations and animation examples to explain all elements, starting from the basics (effector, constraints, jacobians, convergence) to more difficult features such as how the priority levels are enforced, how to handle singularity and stability, why can the CCD be used for realistic human adaptation and can be combined with balance control. So this tutorial concerns beginners as well as advanced users that want to have extra technical information about how it works and how to implement it.

2.3. Necessary background

Basic knowledge of articulated structures animation (composition of transformations, notion of joint and joint limits), basic notion of Linear Algebra (rank, range space, null space, SVD). The course is tailored for a non-specialist audience interested in understanding and implementing numeric IK without enduring its pitfalls. Some technical parts of the presentation are also proposed for advanced users.

2.4. Course material

The course notes consists in:

- the present summary with a short list of key annotated references,
- the powerpoint handouts organized in four parts,
- two reference articles [BB04, KMA05],
- and animation examples presented during the course.

2.5. Presenters

Ronan Boulic is a senior researcher, lecturer and PhD advisor at the Swiss Federal Institute of Technology, Lausanne (EPFL). He is working in the Virtual Reality Lab and his research interests include 3D interactions, motion capture, virual prototyping, and realistic motion synthesis for virtual humans and robot. He received the PhD degree in Computer Science in 1986 from the University of Rennes, France, at the INRIA-IRISA research institute. He received the Habilitation degree from the University of Grenoble, France, in march 1995. Ronan Boulic is co-author of 102 research papers among which 27 appeared in international peer-reviewed journals. He is senior member of IEEE, senior member of ACM and member of Eurographics.

Ronan Boulic has been working in the area of Inverse Kinematics since 1985, first in Robotics at INRIA-Rennes, France then in Animation at EPFL, Switzerland. He was the advisor of five PhD students who have contributed to extend the IK technology in various directions:

- Balance control (Inverse Kinetics) with Ramon Mas (1996)
- Efficient Prioritized IK with Paolo Baerlocher (2001)
- Modeling fatigue with Inmaculada Rodriguez Santiago (2003)
- Motion Editing with Benoit Le Callennec (2006)
- Movement recovery from Mocap data and collision avoidance with Manuel Peinado (on-going)

Richard Kulpa is an Assistant Professor at the M2S laboratory (previously LPBEM) of the University of Rennes 2, France. He is also an external collaborator of the Bunraku team of the IRISA-INRIA Rennes research institute. His research interests include biomechanics, realistic humanoids animation and virtual reality.

He worked since 1996 as a research engineer in the SIAMES team (now Bunraku) where he worked on behavioral animations of characters in informed environments in collaboration with French multimedia companies. Since 2000, in parallel to these previous works, he has integrated the LPBEM laboratory where he worked on biomechanics and the understanding of the human movement. He incorporated biomechanical knowledge in IK methods in order to produce fast and realistic animations. At last, he received his PhD in 2005 for a work that combines biomechanics and iterative inverse kinematics and kinetics solver in order to have realistic animation of hundreds of characters in real-time. He received a price at Eurographics 2005 concerning this work. By the way, he co-created the MKM animation engine that incorporates all these works. This software is currently used by an international company.

3. Content of the tutorial

3.1. Plan

First module:

1. Part A - Overview
   a. Motivation
   b. Problem
      i. Definition of inverse kinematics
      ii. Definition of the constraints
   c. Overview of the IK methods
      i. A taxonomy of the IK methods
      ii. Analytical method
      iii. Iterative methods
2. Part B - Iterative methods (pseudo-inverse based)
   a. Concept of prioritized control
   b. From one to P priority levels
      i. General form
      ii. Incremental constraint achievement
      iii. Example 1: the simplest redundant case
      iv. Example 2: two prioritized constraints
      v. Generalization to multiple priority levels
      vi. Convergence: weighted vs prioritized
      vii. Fostering synergies
      viii. Handling singularities with DLS
      ix. Application: collision avoidance
      x. Conclusion

Second module:

1. Part C - Iterative methods (CCD-based)
   a. CCD
   b. Hybrid CCD and analytical
      i. Representation of data
      ii. Kinematic solver
      iii. Kinetic solver
      iv. Kinematic and kinetic solver
      v. results
      vi. Conclusion
2. Part D - Conclusion

3.2. First module

The first module is organized in two sections, an overview and a detailed presentation of the prioritized IK technology.

The overview is intended to set the stage, recall the terminology, offer a quick survey of the IK landscape and highlight the relative advantages of the two complementary Inverse Kinematics technologies detailed afterward.

Indeed, mastering the full body postural control of complex articulated structures such as human-like bodies is intrinsically difficult for two reasons:

- First the postural space is large, and usually much larger than the "action" space (e.g. reach the door knob with the right hand); as a consequence an infinite number of valid postures exists but retaining the "best" one is uneasy to express.
- Second, being humans ourselves we are very sensitive to the posture properties, e.g. whether it is balanced or not; hence uncorrect, or uncomfortable, or unusual postures cannot be accepted as they appear as unbelievable. Similarly, a partial postural control may produce uncorrect postures for the evaluation of complex systems (e.g. in Virtual Prototyping) and lead to design errors.

This tutorial targets the problematic of the intuitive production of full body posture from the specification of high-level goal-oriented constraints such as "remain balanced", "reach this location with this body part" (cf. figure 1), "look at", "orient"... For that purpose the definitions of key concepts are briefly recalled, i.e. kinematic chain, effector, constraints, direct kinematic model, inverse kinematic model. A brief survey then proposes a classification of the existing methods according to their convergence structure [Chi96, Wel93, ZB94, BB04, KMA05]. The major IK approaches are compared in terms of advantages and drawbacks; our major criteria for this comparison is complementary such as versatility and efficiency. Regarding data-driven IK approaches such as [GMHP04], we briefly compare them with other approaches but the focus of this tutorial is data-independent IK technologies.

Figure 1: *Fast posture adjustment to the ground with an hybrid CCD IK approach [KMA05].*

The second part of the first module provides an in-depth presentation of the Prioritized Inverse Kinematics. This technology has been retained for the detailed presentation as it allows the enforcement of strict priority levels to sort the various constraints according to their importance. The interest of this strategy is conceptually illustrated before entering the description of simple case studies highlighting the properties of the pseudo-inverse and the associated Null space projection operator. The key theoretical formula is explained graphically on the two-priority levels case study. It is then generalized to an arbitrary number of priority levels with a special attention to the algorithm implementation and its computing cost. The description includes the joint limits management through a clamping loop (cf. figure 2).

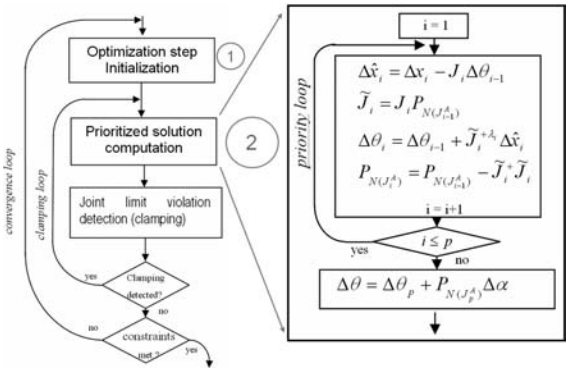


Figure 2: *Accumulating the contributions of each priority levels in the IK priority loop.*

The following discussion and examples illustrate the convergence behavior resulting from prioritized IK. For example, it produces intermediate believable postures as the most important constraints are enforced first, and generally very quickly. This is in contrast with the weighted minimization approach where no constraint is absolutely guaranteed when they conflict with one other. This can lead to local minima or unplausible intermediate/final postures. Another advantage of the prioritized architecture is the possibility to recruit the full body for achieving all constraints resulting in synergistic solutions (cf. figure 3) that could not be found otherwise [BB04].

We discuss then on how to exploit the SVD to handle the critical issue of singularities (geometric and algorithmic). Although the Damped Least Square inverse [BK05] is the key for ensuring the stability, it should not be exploited to enforce priority levels (as often reported in the literature). The correct formulation ensuring both stability and priority levels is highlighted.

After a few examples of reachable space identification, the section ends with the management of collision avoidance (cf. figure 4) through the dynamic creation of partial repulsion effectors [BB04]. Automatic local collision anticipation and avoidance concludes the module [PBCM05].

3.3. Second module

The part C of the second module provides an in-depth presentation of the CCD-like algorithms and the way to produce realistic adaptations of many characters in real-time.

The CCD (Cyclic Coordinate Descent) is an iterative method that modifies only one DOF at a time. This way, the jacobian is simplified as a vector and the inversion is cheap.

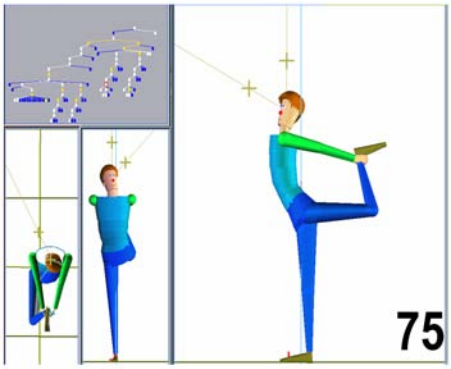


Figure 3: *Full body postural synergy achieving a dance posture where numerous joint limits are enforced with the inequality constraint management (clamping loop).*

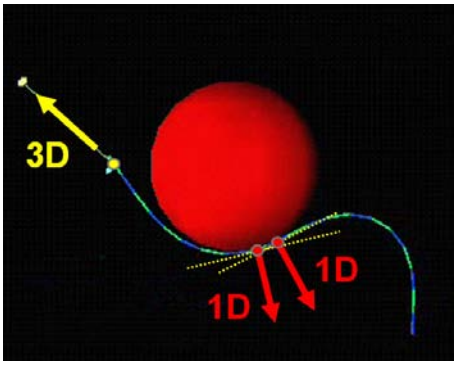


Figure 4: *Dynamic creation of 1D repulsion effectors to ensure obstacle avoidance.*

A geometrical version of this method is also available rotating one joint at a time. This technique is consequently very fast and can easily handle joint limits but has some important drawbacks that makes it incompatible with human postural adaptation. Indeed, the repartition of the deformation is not homogeneous leading to unrealistic postures: first joints are more modified than the following ones.

We propose to take advantage of this drawback by using analytical solutions on the chain ends and by dividing the skeleton into groups to order the adaptations between these groups and within them (ordering the joints). To this end, two representations of data are detailed:

- Constraints: in order to take user high-level constraints into account while having a generic and efficient inverse kinematics and kinetics solver, two levels of constraints are considered: the external one represents the user constraints while the internal one is a unique representation handled by the algorithm.
- Skeleton: in order to create analytical solutions that works with weighted or prioritized constraints, a suitable representation of the skeleton for adaptation is needed. A normalized representation is detailed with some examples of possible easy adaptations.

Based on these representations, analytical solutions are used for the limbs decreasing the number of segments in the kinematic chain and consequently increasing the performance of the method. It also allows to find instantaneously their best configuration according to the constraints. This means that the adaptation applied to the remainder of the body is minimized. Combined with a correct adaptation order for the iterative algorithm, realistic postures can be found very rapidly.

This correct order is obtain by dividing the body into groups. The goal is to take advantage of the difference between human body parts (more or less movable and heavy). Then, a common hierarchy of these groups is proposed to order the adaptation between the groups for the kinematic adaptation as well as for the kinetic adaptation. The correct adaptation order for inverse kinematics goes from the groups with the largest amplitude to the lowest one. On the opposite, inverse kinetics adaptation modifies the groups from the heaviest to the lightest ones. A detailed algorithm is proposed in the tutorial.

Finally, during kinematics, after adapting all the groups, root adaptation is made to manage the constraints placed on the pelvis but also to reach far targets. Indeed, the global error between the constraints and their targets can be reduced translating the root. Using a percentage parameter, it allows to have different adaptations: high percentage for decided characters that want to reach their goals and low percentage for tired characters for example. Additional laws can also be added to have different adaptations such as the joint range to simulate the suppleness as suggested in figure 5.

Figure 5: *Additional laws allow to have different adaptations.*

To have realistic adaptation, the balance of the character must be controlled: in 3D in aerial phases and in 2D when the character is on the ground. In this last case, the goal is to ensure that the projections of the current and the desired COM are at the same position. To do so, we propose an inverse kinetics method based on exactly the same representation and algorithm than described before. As the position of the center of mass (COM) is obtained as a weighted sum of the local COM of the segments, its computation can easily be subdivided into groups. According to the previously

defined hierarchy of groups, the inverse kinetics solver proposed can iterate on these groups in order to verify the desired COM. The error on the COM position is consequently transformed as an error on the local COM of the currently modified group. Placing the group correctly means that the global COM is at the desired position. As for kinematic adaptation, analytical solutions are used for the limbs in order to place them instantaneously at the correct posture in order to verify the local COM. For the other groups, rotations are used instead. Finally, the root adaptation consists in its translation to correct a percentage of the error of the global COM position. This percentage is really important since it allows to have different kind of adaptations mixing those of the root or/and the groups (cf. figure 6).

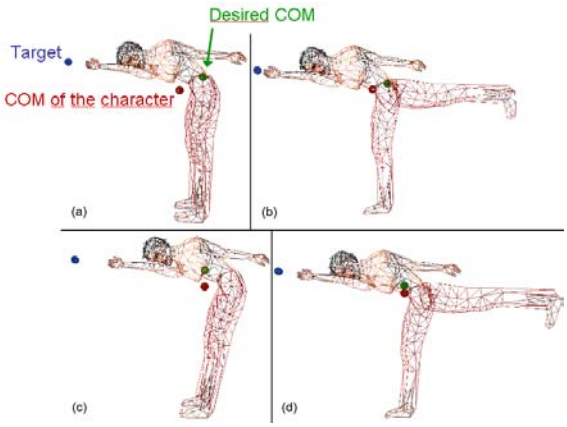


Figure 6: *Different kinetic adaptations: a) reference posture, b) using groups, c) using root, d) using both.*

Finally, inverse kinematics and kinetics are inserted inside a unique iterative loop. Of course, in order to be compatible with the kinematic adaptation, only the groups that are not already constrained can be used for inverse kinetics. And only a subset can be selected by the user in order to have different kind of adaptations. The complete algorithm always converges because the modified groups are distinct and the kinematic and kinetic adaptations are antagonistic ones. Demonstrations are proposed in the tutorial for postural and motion adaptation.

Finally, the part D of this tutorial quickly summarizes the different IK methods and their advantages/drawbacks. It also indicates how these IK methods can also be used for motion adaptations.

4. Structured bibliography

Part A: Overview, Basis and Taxonomy

- CS IK course from Steve Rotenberg (USD) [Rot05]
- Jacobian transpose and CCD approaches [Wel93]
- Weighted IK methods for Virtual human postural control [BMW87, PZB90, PB91, BPW93, ZB94]
- IK control of the center of mass [BMT96, BMST97]
- Analytic IK [Chi96, TB96, TGB00]
- Computer Puppetry [SLSG01]
- Motion-Model based IK [GMHP04]

Part B: Prioritized Inverse Kinematics

- First form of Prioritized IK by Liegeois [Lie77]
- Two priority levels by Hanafusa et al. [HYN81]
- Arbitrary number of priority levels [SS91]
- Prioritized IK with reduced computing cost [BB04]
- Damped Pseudo-Inverse for singularity management [Mac90, BK05]

Part C: CCD and fast Hybrid approach

- Origin of the CCD [Lue84, WC91]
- CCD applied on human [BMB87]
- CCD using geometrical resolution [Lan98]
- Fast IK and balance control [KMA05, KM05]

Part D: Conclusion

- IK used to adapt motions [BW95, Gle97, Gle01, lCB04]
- Prioritized IK demonstrations
  http://vrlab.epfl.ch/ boulic/demos.html
- Fast CCD-like IK demonstrations
  http://www.irisa.fr/bunraku/Richard.Kulpa

5. Acknowledgments

Research on Prioritized Inverse Kinematics led by Ronan Boulic have been realized with the support of the Swiss National Science Foundation, the European Union Network of Excellence ENACTIVE, the University of Alcala and EPFL.

Research on Hybrid-CCD Inverse Kinematics and Kinetics led by Richard Kulpa have been realized with the support of the French National Science Foundation, the "Conseil Regional and General" of Brittany and Rennes city.

References

[BB04] BAERLOCHER P., BOULIC R.: An inverse kinematics architecture enforcing an arbitrary number of strict priority levels. *The Visual Computer 20*, 6 (2004), 402–417.

[BK05] BUSS S., KIM J.: Inverse kinematics with selectively damped least squares. *Journal of Graphics Tools 10*, 3 (2005), 37–49.

[BMB87] BADLER N. I., MANOOCHEHRI K. H., BARAFF D.: Multi-dimensional input techniques and articulated figure positioning by multiple constraints. In *SI3D'86: Proceedings of the 1986 workshop on Interactive 3D graphics* (New York, NY, USA, 1987), ACM Press, pp. 151–169.

[BMST97] BOULIC R., MAS-SANSO R., THALMANN D.: Complex character positioning based on a compatible flow model of multiple supports. *IEEE Transactions on Visualization and Computer Graphics 3*, 3 (jul–sep 1997), 245–261.

[BMT96] BOULIC R., MAS R., THALMANN D.: A robust approach for the control of the center of mass with inverse kinetics. *Computers and Graphics 20*, 5 (Sept. 1996), 693–701.

[BMW87] BADLER N. I., MANOOCHEHRI K. H., WALTERS G.: Articulated figure positioning by multiple constraints. *IEEE Computer Graphics and Applications 7*, 6 (June 1987), 28–38.

[BPW93] BADLER N. I., PHILLIPS C. B., WEBBER B. L.: *Simulating humans: computer graphics animation and control*. Oxford University Press, Inc., New York, NY, USA, 1993.

[BW95] BRUDERLIN A., WILLIAMS L.: Motion signal processing. In *Proceedings of SIGGRAPH'95* (Aug. 1995), Computer Graphics Proceedings, Annual Conference Series, pp. 97–104.

[Chi96] CHIN K. W.: *Closed-form and generalized inverse kinematic solutions for animating the human articulated structure*. Master's thesis, Curtin University of Technology, 1996.

[Gle97] GLEICHER M.: Motion editing with spacetime constraints. In *Symposium on Interactive 3D Graphics'97* (Apr. 1997), pp. 139–148.

[Gle01] GLEICHER M.: Comparing constraint-based motion editing methods. *Graphical models 63*, 2 (2001), 107–134.

[GMHP04] GROCHOW K., MARTIN S. L., HERTZMANN A., POPOVIĆ Z.: Style-based inverse kinematics. *ACM Trans. Graph. 23*, 3 (2004), 522–531.

[HYN81] HANAFUSA H., YOSHIKAWA T., NAKAMURA Y.: Analysis and control of articulated robot with redundancy. In *IFAC, 8th Triennal World Congress* (1981), vol. 4, pp. 1927–1932.

[KM05] KULPA R., MULTON F.: Fast inverse kinematics and kinetics solver for human-like figures. In *Proceedings of IEEE Humanoids* (Tsukuba, Japan, december 2005), pp. 38–43.

[KMA05] KULPA R., MULTON F., ARNALDI B.: Morphology-independent representation of motions for interactive human-like animation. *Computer Graphics Forum, Eurographics 2005 special issue 24*, 3 (2005), 343–352.

[Lan98] LANDER J.: Oh my god, i inverted kine! *Game Developer Magazine* (Sept. 1998).

[lCB04] LE CALLENNEC B., BOULIC R.: Interactive motion deformation with prioritized constraints. In *SCA '04: Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation* (New York, NY, USA, 2004), ACM Press, pp. 163–171.

[Lie77] LIEGEOIS A.: Automatic supervisory control of the configuration and behevior of multibody mechanisms. *IEEE Transactions on Systems, Man and Cybernetics SMC-7*, 12 (Dec. 1977).

[Lue84] LUENBERGER D. G.: *Linear and nonlinear programming*. Addison-Wesley, 1984.

[Mac90] MACIEJEWSKI A. A.: Dealing with the ill-conditioned equations of motion for articulated figures. *IEEE Computer Graphics and Applications 10*, 3 (May 1990), 63–71.

[PB91] PHILLIPS C. B., BADLER N. I.: Interactive behaviors for bipedal articulated figures. *Proceedings of SIGGRAPH'91 on Computer Graphics 25*, 4 (July 1991), 359–362.

[PBCM05] PEINADO M., BOULIC R., CALLENNEC B. L., MEZIAT D.: Progressive cartesian inequality constraints for the inverse kinematics control of articulated chains. *Short Presentation Proc. of Eurographics'05 4656* (2005).

[PZB90] PHILLIPS C. B., ZHAO J., BADLER N. I.: Interactive real-time articulated figure manipulation using multiple kinematic constraints. *Symposium on Interactive 3D Graphics'90 24*, 2 (Mar. 1990), 245–250.

[Rot05] ROTENBERG S.: Inverse kinematics, part 1 and 2. *http://graphics.ucsd.edu/courses/cse169_w06* (2005).

[SLSG01] SHIN H. J., LEE J., SHIN S. Y., GLEICHER M.: Computer puppetry: An importance-based approach. *ACM Transactions on Graphics 20*, 2 (2001), 67–94.

[SS91] SICILIANO B., SLOTINE J.-J.: A general framework for managing multiple tasks in highly redundant robotic systems. In *ICAR'91* (1991), vol. 2, p. 1211Ű1215.

[TB96] TOLANI D., BADLER N. I.: Real-time inverse kinematics of the human arm. *Presence 5*, 4 (1996), 393–401.

[TGB00] TOLANI D., GOSWAMI A., BADLER N. I.: Real-time inverse kinematics techniques for anthropomorphic limbs. *Graphical Models 62*, 5 (2000), 353–388.

[WC91] WANG L.-C. T., CHEN C. C.: A combined optimization method for solving the inverse kinematics problem of mechanical manipulators. *IEEE Transactions on Robotics and Applications 7*, 4 (Aug. 1991), 489–499.

[Wel93] WELMAN C.: *Inverse Kinematics and Geometric Constraints For Articulated Figure Manipulation*. PhD thesis, Simon Fraser University, 1993.

[ZB94] ZHAO J., BADLER N. I.: Inverse kinematics positioning using nonlinear programming for highly articulated figures. *ACM Transactions on Graphics 13*, 4 (Oct. 1994), 313–336.

Inverse kinematics and kinetics for virtual humanoids

Ronan Boulic
VRLab – EPFL
http://vrlab.epfl.ch

Richard Kulpa
Bunraku – INRIA / IRISA
www.irisa.fr/bunraku

M2S – Univ. Rennes 2
www.uhb.fr/labos/m2s

Plan

- Part A - Overview
  - Motivation
  - Problem
  - Overview of the IK methods
- Part B - Iterative methods (pseudo-inverse based)
  - Concept of prioritized control
  - From one to P priority levels
- Part C - Iterative methods (CCD-based)
  - CCD
  - Hybrid CCD and analytical
- Part D - Conclusion

Part A - Overview

- Motivation
- Problem
  - Definition of inverse kinematics
  - Definition of the constraints
- Overview of the IK methods
  - A taxonomy of the IK methods
  - Analytical method
  - Iterative methods

Motivation

- Many applications use virtual humanoids
- Need realistic postures and motions
  - Postures
    - Ergonomics
    - Maintainability
  - Motions
    - Virtual reality
    - Video games
    - Multimedia productions

Motivation

- Can capture postures and motions
- Problem: dependent on original conditions of capture (e.g. morphology, position of the target during grasping…)

⇒ Need to adapt postures and motions

5

Motivation

- Inverse kinematics adapts postures
- Motions can be considered as a succession of postures

⇒ So inverse kinematics can also be used for motion adaptation
  - Directly for small adaptations
  - Using filtering for large adaptations

6

Motivation

- This tutorial deals with inverse kinematics methods to adapt postures
  - Advantages
  - Drawbacks
- Moreover, it deals with inverse kinetics
  - Takes masses into account
  - Preserves balance

7

Part A - Overview

- Motivation
- Problem
  - Definition of inverse kinematics
  - Definition of the constraints
- Overview of the IK methods
  - A taxonomy of the IK methods
  - Analytical method
  - Iterative methods

8

Definition of Inverse Kinematics

- Problem: given the goal position and/or orientation of some body parts, noted $\mathbf{x_g}$, find the "best" joint state θ of the full body
- Direct Kinematic model

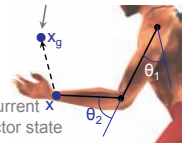
$x = F(\theta, \text{body skeleton})$

Effector state

Joint parameters

Kinematic chain

(here a position constraint goal)



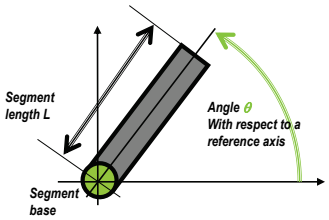
9

Illustration on a 2D chain

- Kinematic chain F composed of
  - Segments
  - Joints (considered as ideal)
- Each joint angle is relative to the parent segment
- For the first segment the joint angle is relative to a reference axis

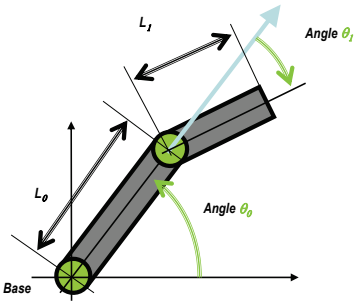
10

First segment



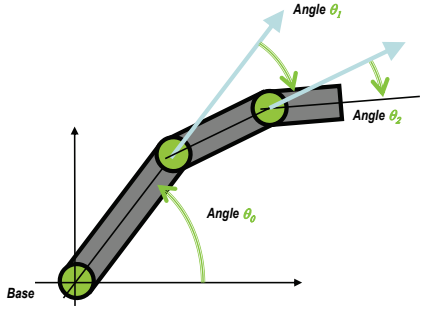
11

Second segment

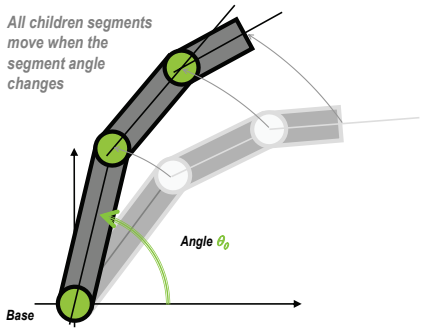


12

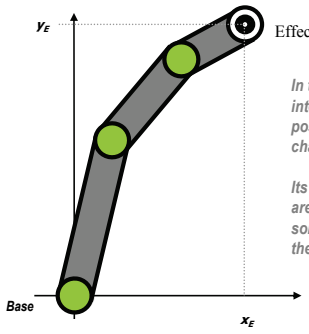
Third segment



13



14



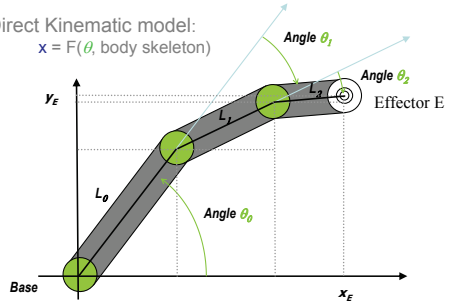
In this example we are interested to control the position of the tip of the chain, called the Effector

Its coordinates X (x_E,y_E) are important to specify some task to achieve with the chain

15

Direct Kinematic model:

x = F(θ, body skeleton)

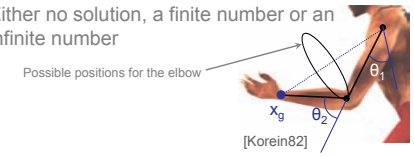


$$x_E = L_0 \cos\theta_0 + L_1 \cos(\theta_0 + \theta_1) + L_2 \cos(\theta_0 + \theta_1 + \theta_2)$$
$$y_E = L_0 \sin\theta_0 + L_1 \sin(\theta_0 + \theta_1) + L_2 \sin(\theta_0 + \theta_1 + \theta_2)$$

16

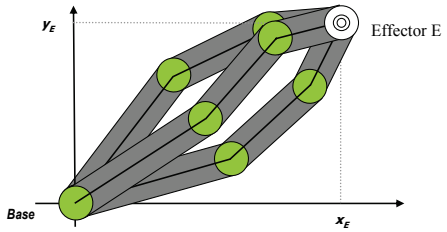
Definition of Inverse Kinematics

- What is θ for a goal value x_g ?
- Inverse Kinematic model
  $\theta = F^{-1}(x_g,$ body skeleton$)$
- Generally redundant: dim(x_g) < dim(θ)
  - Less constraints than degrees of freedom
  - Either no solution, a finite number or an infinite number



[Korein82]

Inverse Kinematic model:
$\theta = F^{-1}(x_g,$ body skeleton$)$?



Either find directly the most appropriate solution
Or choose it amongst all the correct solutions

Part A - Overview

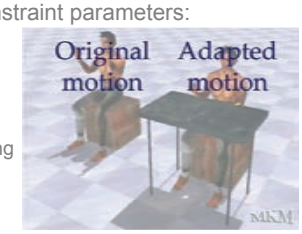
- Motivation
- Problem
  - Definition of inverse kinematics
  - ⇨ Definition of the constraints
- Overview of the IK methods
  - A taxonomy of the IK methods
  - Analytical method
  - Iterative methods

Constraints

- Inverse kinematics is controlled by the x_g parameter
  $\theta = F^{-1}(x_g,$ body skeleton$)$
- x_g represents:
  - a type of high-level control
    - Position
    - Orientation, Gaze
    - Balance
  - the goal defined by the user or automatically depending on the context
    - Interaction tasks (floor, reach, grasp…)
    - Collision avoidance
  - Remark: "***task***" and "***constraint***" are used as synonyms

Constraint definition

- Example of constraint parameters:
  - Type
    - Equality
    - Inequality
  - Position
  - Range/recruiting
  - Importance

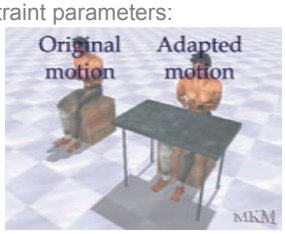


A point on the skin of the forearm placed on the edge of the table

21

Constraint definition

- Example of constraint parameters:
  - Type
    - Equality
    - Inequality
  - Position
  - Range/recruiting
  - Importance



Wrists above the table

22

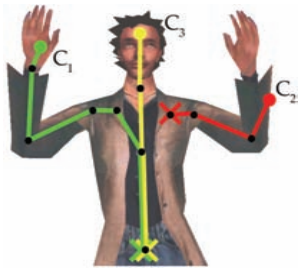
Constraint definition

- Example of constraint parameters:
  - Type
    - Equality
    - Inequality
  - Position
  - Range/recruiting
  - Importance

23

Constraint definition

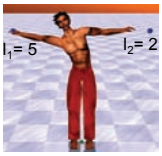
- Example of constraint parameters:
  - Type
    - Equality
    - Inequality
  - Position
  - Range/recruiting
  - Importance



24

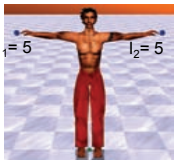
Constraint definition

- Example of constraint param:
  - Type
    - Equality
    - Inequality
  - Position
  - Range/recruiting
  - Importance



Priority

or



Weight

25

Conflicting constraints

- Frequent when interacting with the environment
- Two ways to handle conflicting constraints
  - Weights
    - Find a compromise using a weigthed sum of the constraints [Zhao93, Gleicher98, Gleicher02]
  - Priorities
    - Verify the most important constraints first.
    - Try to verify the others only if possible [Liégeois77, Hanafusa81, Yamane03, Slotine91, Baerlocher98]

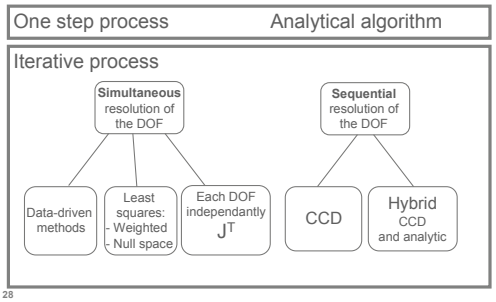
26

Part A - Overview

- Motivation
- Problem
  - Definition of inverse kinematics
  - Definition of the constraints
- Overview of IK methods
  - A taxonomy of IK methods
  - Analytical method
  - Iterative methods

27

Overview of IK methods



28

Part A - Overview

- Motivation
- Problem
  - Definition of inverse kinematics
  - Definition of the constraints
- Overview of the IK methods
  - A taxonomy of the IK methods
  - ⇨ Analytical method
  - Iterative methods

29

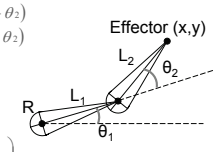
Analytical methods

- Invert the kinematic equations
- Example for a limb
  - Direct kinematics

$$x = f_1(\theta_1, \theta_2) = l_1 \cos\theta_1 + l_2 \cos(\theta_1 + \theta_2)$$
$$y = f_2(\theta_1, \theta_2) = l_1 \sin\theta_1 + l_2 \sin(\theta_1 + \theta_2)$$

  - Inverse kinematics

$$\theta_2 = \arccos\left(\frac{x^2 + y^2 - L_1^2 - L_2^2}{2L_1L_2}\right)$$
$$\theta_1 = \arctan\left(\frac{y}{x}\right) - \arctan\left(\frac{L_2 \sin\theta_2}{L_1 + L_2 \cos\theta_2}\right)$$



30

Analytical methods (2)

- Major references:
  - Robotics: [Craig03]
  - Arm/leg postural control: [Korein82, Paul88, Watt92, Chin96, Tolani00]
  - Other authors used it in combination with other methods [Lee99, Shin01]

31

Summary: Analytical IK

- Advantages
  - Global solution
  - Reliable, exploited in Robotics
  - Strong (no singularity)
  - Fast
  - « Simple »
- Drawbacks
  - Only for a few DOFs
  - Not suited for redundant systems more complex than the arm/leg

32

Part A - Overview

- Motivation
- Problem
  - Definition of inverse kinematics
  - Definition of the constraints
- Overview of the IK methods
  - A taxonomy of the IK methods
  - Analytical method
  - Iterative methods
    - Simultaneous resolution of the DOF
    - Sequential resolution of the DOF

Iterative methods

Simultaneous resolution of the DOF

- General approach: considering the problem as a cost function to be minimized
  - Non-linear Optimization method
    [Badler86, Badler87, Zhao94]

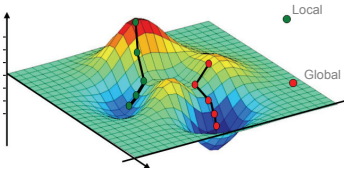
$$\text{Cost}(\theta) = K_{main}\,[\,f(\theta) - X_d\,]^2 + K_{secondary}\,g(\theta)$$

Main task:
error on the position

Secondary task

Summary: non-linear optimization

- No singularity
- Problem to choose the weights
- Difficult to enforce strict priority levels
- Can fall into a local minimum



Iterative methods

Simultaneous resolution of the DOF

- Other focus: Linearization
  - Exploiting the Jacobian
    - Transpose
    - Pseudo-inverse : part B
  - Dealing with small displacements

Principle of the linearization

1) First linearizes the function:

x = F(θ, body skeleton)

by storing its first derivatives

in the jacobian matrix *J*.

$$J = \begin{pmatrix} \frac{\partial Fx_1}{\partial \theta_1} & \frac{\partial Fx_1}{\partial \theta_2} & \cdots & \frac{\partial Fx_1}{\partial \theta_N} \\ \frac{\partial Fx_2}{\partial \theta_1} & \frac{\partial Fx_2}{\partial \theta_2} & \cdots & \frac{\partial Fx_2}{\partial \theta_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial Fx_M}{\partial \theta_1} & \frac{\partial Fx_M}{\partial \theta_2} & \cdots & \frac{\partial Fx_M}{\partial \theta_N} \end{pmatrix}$$

Principle of the linearization

2) Then, we use the inverse of the Jacobian as a local approximation of

θ = F^{-1}(x, body skeleton)

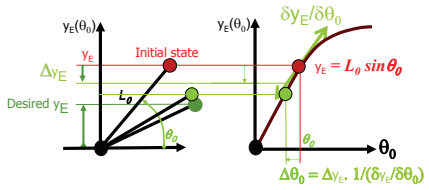
with :

$$\Delta\theta = J^{-1}(\Delta x)$$

Linearization: the 1 D case

(one segment only)

- Linearizing the equation(s) $y_E = f(\theta)$
  - The resolution converges towards *one* possible solution through successive state variations
  - Local solution: great importance of the initial state

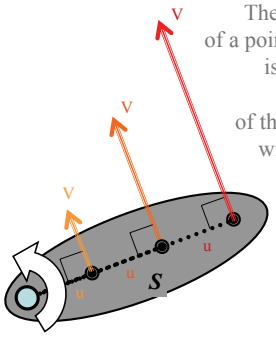


How to build the jacobian?

- Consider the influence of each joint independently of the other joints

  The joints ignore each other ⇔ it's a first order approximation

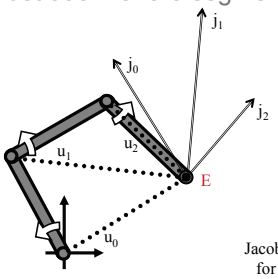
- When one joint changes, all the children segments are viewed as ONE UNIQUE RIGID SOLID
- It is easy to evaluate the instantaneous velocity of a point of a rigid solid:
  - With the cross product of the rotation velocity vector by the position vector locating the point with respect to the joint (see next slide).
  - Use a unit rotation velocity for building the Jacobian



The velocity vector V of a point from the solid ***S*** is obtained with the cross product of the rotation velocity with the *lever arm* u

41

Illustration for a 3 segment arm, for point E



j_0	j_1	j_2
Xj_0	Xj_1	Xj_2
Yj_0	Yj_1	Yj_2

Jacobian J of the effector E for the chain current state

See [Rotenberg06] for other jacobian computation

42

Jacobian transpose

- The Jacobian matrix *J*, of dimension m x n, gathers all the partial derivatives of the m constraints x_i with respect to the n joint variables θ_j

$$J = [\delta x_i / \delta \theta_j]$$

- The Jacobian transpose method [Welman93]:
  - exploits the absolute influence of each joint variable θ_j for contributing to the constraint error vector Δx
  - the influence of a joint j is given by the scalar product of Δx with the column j of J

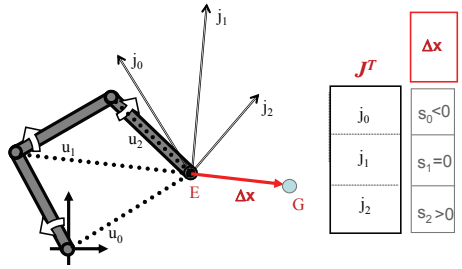
43

Jacobian transpose (2)

- The method is fast to compute
  $\Delta\theta = \beta J^T \Delta x$, with β small
- But not that fast to converge
  - The relative contribution of joint variables is not taken into account; joint dimensions are "blind" to each other…
  - β must be small due to the non-linearity of the direct kinematics model.
  - No possibility to define strict priorities among constrained dimensions
  - Full extension singularity

44

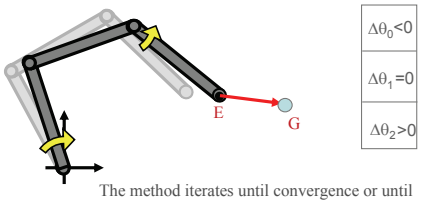
Ex: the effector E is attracted toward the goal G (1)



Translation Jacobian *J* of the effector E for the chain current state

45

Ex: the effector E is attracted toward the goal G (2)



The method iterates until convergence or until the error norm does not decrease any more

46

Part A - Overview

- Motivation
- Problem
  - Definition of inverse kinematics
  - Definition of the constraints
- Overview of the IK methods
  - A taxonomy of the IK methods
  - Analytical method
  - Iterative methods
    - Simultaneous resolution of the DOF
    - Sequential resolution of the DOF

⇨

47

Sequential resolution of the DOF

- Exactly the same concept of small displacement and linearization
- But avoid the Jacobian, dealing only one DOF at a time
  - Computation of one iteration is very fast
  - Number of iterations can be important
- Described in Part C

48

Comparison of analytic and iterative IK

- Analytic is best suited for simple case like isolated arm, leg, etc.
- Iterative is more general but requires multiple steps to converge towards the solution
  - Due to the non-linearity of the problem
  - If big steps are used, it becomes unstable

49

Inverse kinematics and kinetics for virtual humanoids

Ronan Boulic
VRLab – EPFL
http://vrlab.epfl.ch

Richard Kulpa
Bunraku – INRIA / IRISA
www.irisa.fr/bunraku

M2S – Univ. Rennes 2
www.uhb.fr/labos/m2s

Plan

- Part A - Overview
  - Motivation
  - Problem
  - Overview of the IK methods
- ⇨ Part B - Iterative methods (pseudo-inverse based)
  - Concept of prioritized control
  - From one to P priority levels
- Part C - Iterative methods (CCD-based)
  - CCD
  - Hybrid CCD and analytical
- Part D - Conclusion

2

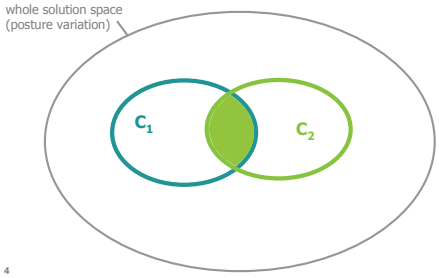
Part B - Iterative methods (pseudo-inv.)

- ⇨ Concept of prioritized control
- From one to P priority levels
  - General form
  - Incremental constraint achievement
  - Example1: the simplest redundant case
  - Example 2: two prioritized constraints
  - Generalization to multiple priority levels
  - Fostering synergies
  - Handling singularities with DLS
  - Application: collision avoidance
  - Conclusion

3

Conceptual analogy

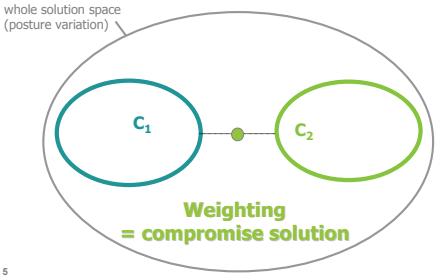
Visualizing the constraints solution sub-spaces: $\mathbf{C}_i$



4

Conceptual analogy

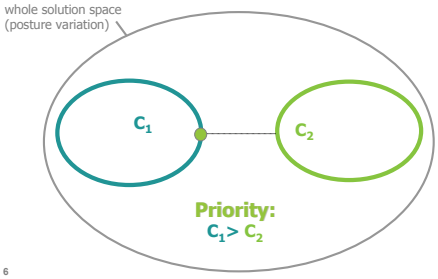
Visualizing the constraints solution sub-spaces: C_i



5

Conceptual analogy

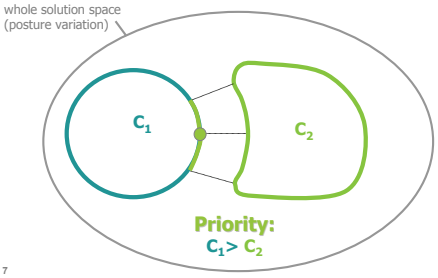
Visualizing the constraints solution sub-spaces: C_i



6

Conceptual analogy

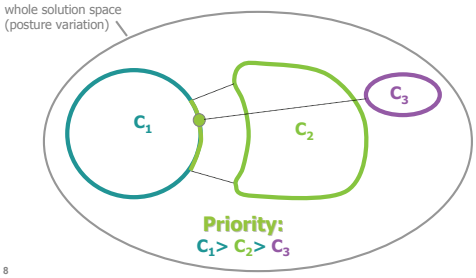
Visualizing the constraints solution sub-spaces: C_i



7

Conceptual analogy

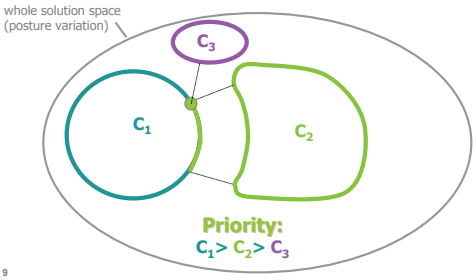
Visualizing the constraints solution sub-spaces: C_i



8

Conceptual analogy

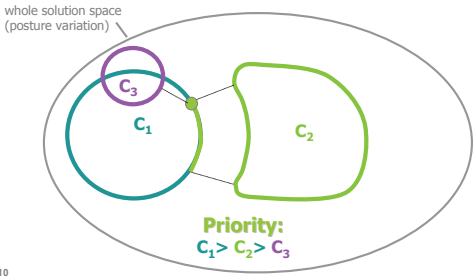
Visualizing the constraints solution sub-spaces: C_i



9

Conceptual analogy

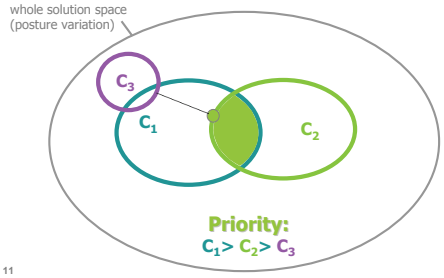
Visualizing the constraints solution sub-spaces: C_i



10

Conceptual analogy

Visualizing the constraints solution sub-spaces: C_i



11

Part B - Iterative methods (pseudo-inv.)

- Concept of prioritized control
- From one to P priority levels
  - General form
  - Incremental constraint achievement
  - Example1: the simplest redundant case
  - Example 2: two prioritized constraints
  - Generalization to multiple priority levels
  - Fostering synergies
  - Handling singularities with DLS
  - Application: collision avoidance
  - Conclusion

12

General form

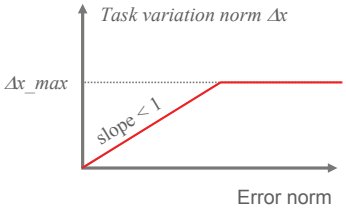
- The pseudo-inverse J^+ , of dimension n x m, provides the minimum norm solution $\Delta\theta$ for a desired task variation Δx [Klein83]

$$\Delta\theta = J^+\Delta x$$

- The validity of the solution is also limited to the neighborhood of the current state, hence the norm of Δx is restricted to an increment toward the desired goal state.

13

Incremental task achievement



- The slope ensures stability (e.g. 0.7)

14

General Form of the Solution

- The solution is given by: $\Delta\theta = J^+\Delta x + P_{N(J)}\Delta\alpha$
- where P_N is the n x n projection operator on the Null space of J :

$$P_{N(J)} = I - J^+J$$

- Any vector belonging to $N(J)$ is mapped by J on the null vector
- $\Delta\alpha$: of dimension n, *can be any* vector belonging to the joint variation space. It is often proportional to the gradient of a cost function expressed in the joint space g(θ).
  - remark: this term is often noted z in the Robotics litterature, e.g. [Liegeois77]

15

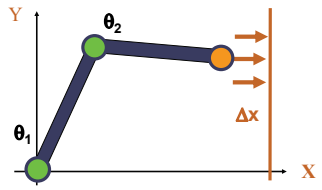
Part B - Iterative methods (pseudo-inv.)

- Concept of prioritized control
- From one to P priority levels
  - General form
  - Incremental constraint achievement
  - ⇨ Example1: the simplest redundant case
  - Example 2: two prioritized constraints
  - Generalization to multiple priority levels
  - Fostering synergies
  - Handling singularities with DLS
  - Application: collision avoidance
  - Conclusion

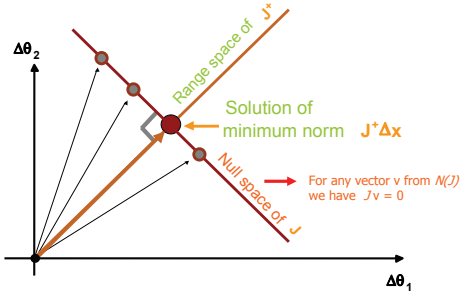
16

The simplest redundant case

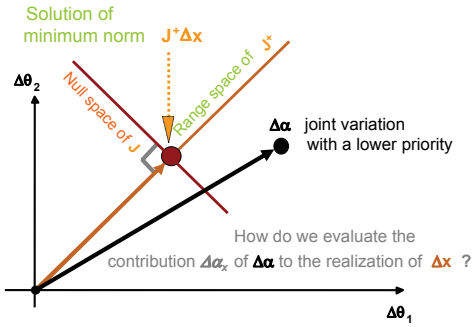
- A two DOF chain has to constrain the **X** dimension of the orange end effector



17



18

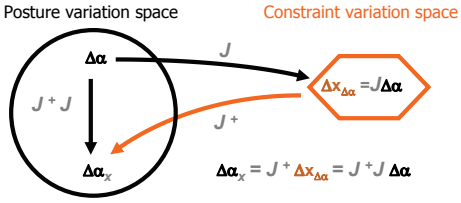


19

– How do we evaluate the contribution $\Delta\alpha_x$ of $\Delta\alpha$ to the realization of Δx ?

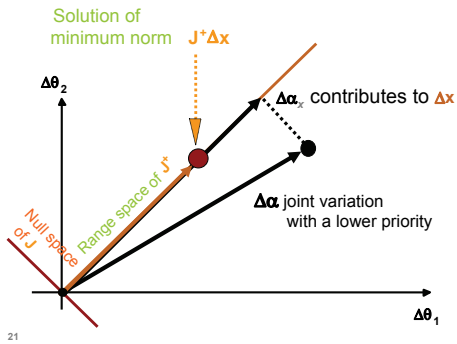
First the influence of $\Delta\alpha$ in the constraint space $\Delta x_{\Delta\alpha}$ is given by multiplying $\Delta\alpha$ with the Jacobian J :

$\Delta x_{\Delta\alpha} = J\,\Delta\alpha$

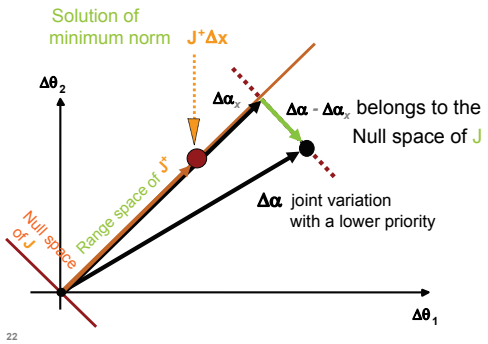


$\Delta\alpha_x = J^+ \Delta x_{\Delta\alpha} = J^+ J\,\Delta\alpha$

20



21



22

$\Delta\alpha - \Delta\alpha_x$ belongs to the Null Space of J

it is mapped to the null vector by J:

$$J(\Delta\alpha - \Delta\alpha_x) = J(\Delta\alpha - J^+ J\,\Delta\alpha)$$

$$= J\Delta\alpha - J\,J^+ J\,\Delta\alpha$$

$$= \Delta x_{\Delta\alpha} - J\,J^+ \Delta x_{\Delta\alpha}$$

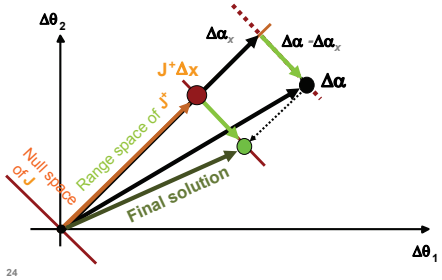
$$= \Delta x_{\Delta\alpha} - J\,\Delta\alpha_x$$

$$= \Delta x_{\Delta\alpha} - \Delta x_{\Delta\alpha}$$

$$= 0$$

23

The *orthogonal projection on the Null space N(J)* ($\Delta\alpha - \Delta\alpha_x$) can be safely added to the minimum norm solution



24

– Any vector $\Delta\alpha$ projected on the Null space N(J) does not perturb the constraint achievement:
  - it is mapped to the null vector by J

$$J\,(\Delta\alpha - \Delta\alpha_x) = 0$$

– The projection operator on the Null Space is noted $P_{N(J)}$:

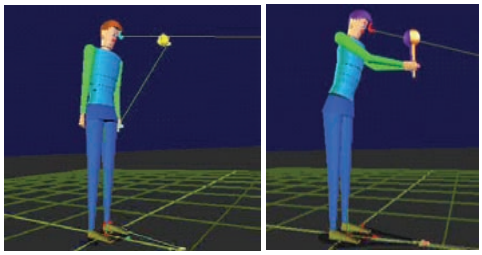
$$(\Delta\alpha - \Delta\alpha_x) = \Delta\alpha_{N(J)} = P_{N(J)}\,\Delta\alpha$$

$$\Delta\alpha - \Delta\alpha_x = \Delta\alpha - J^{+}J\,\Delta\alpha = (I - J^{+}J)\,\Delta\alpha$$

$$P_{N(J)} = (I - J^{+}J)$$

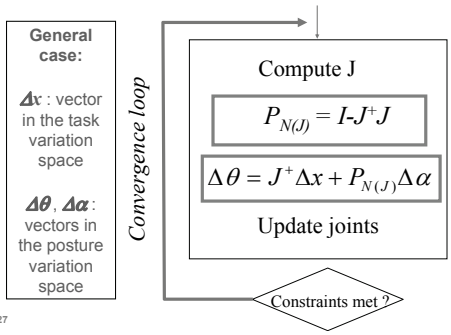
25

Example in interactive context [Baerlocher04]



26

Basic Prioritized IK architecture



27

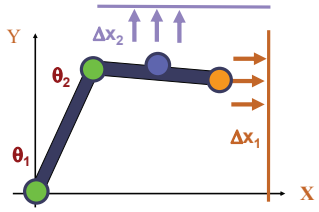
Part B - Iterative methods (pseudo-inv.)

- Concept of prioritized control
- From one to P priority levels
  - General form
  - Incremental constraint achievement
  - Example1: the simplest redundant case
  - ⇨ Example 2: two prioritized constraints
  - Generalization to multiple priority levels
  - Fostering synergies
  - Handling singularities with DLS
  - Application: collision avoidance
  - Conclusion

28

Towards Hierarchical Constraints: two 1D constraints (1)

- A two dofs chain has to constrain effector1 on a line and effector2 on another line.



29

Towards Hierarchical Constraints: two 1D constraints (2)

- Context: the constraints are conflicting; they cannot be achieved simultaneously
- Objective: give the highest priority to the achievement of Δx_1 and achieve as much as possible Δx_2 without perturbating Δx_1
- Specific issue: take into account the fact that achieving the high priority task Δx_1 may interfere with the low priority task Δx_2

30

Towards Hierarchical Constraints: two 1D constraints (3)

- Let $\Delta\theta_i$ denote the solution vector in the posture variation space for priority level i
- The high priority solution for the task Δx_1 is :

$$\Delta\theta_1 = J_1^+ \Delta x_1$$

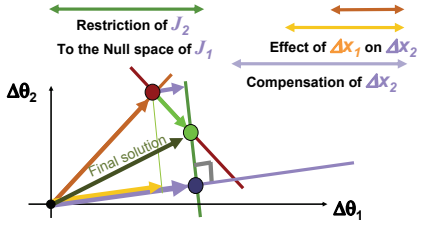
- The low priority term $\Delta\alpha$ has to be projected on, or *to belong to*, the Null space of J_1

31

Two conflicting Constraints [Hanafusa81]

Construction of the contribution of second priority level :

$$\Delta\alpha = [J_2(I - J_1{}^+ J_1)]^+ (\Delta x_2 - J_2(J_1{}^+ \Delta x_1))$$



32

Towards Hierarchical Kinematic Constraints

- Let us rewrite the total solution vector $\Delta\theta$ for two priority levels :

$\Delta\theta$ = $\Delta\theta_1$ + contribution of priority level 2

$$\Delta\theta = \Delta\theta_1 + [J_2(I - J_1^+ J_1)]^+ (\Delta x_2 - J_2(J_1^+ \Delta x_1))$$

$$\Delta\theta = \Delta\theta_1 + [J_2 P_{N(J_1)}]^+ (\Delta x_2 - J_2 \Delta\theta_1)$$

33

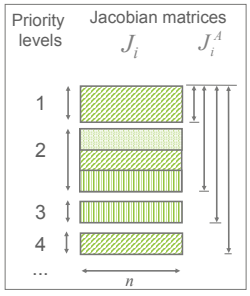
Part B - Iterative methods (pseudo-inv.)

- Concept of prioritized control
- From one to P priority levels
  - General form
  - Incremental constraint achievement
  - Example1: the simplest redundant case
  - Example 2: two prioritized constraints
  - ⇨ Generalization to multiple priority levels
  - Fostering synergies
  - Handling singularities with DLS
  - Application: collision avoidance
  - Conclusion

34

Extension to *P* priority levels [Slotine91]

- Each priority level i is characterized by one jacobian J_i :
  - It can aggregate multiple constraint Jacobians by piling them into a single matrix, resulting in a variable constrained dimension.
  - All have the same number of columns n , dimension of the joint space.



35

Extension to *P* priority levels (2)

- The general solution vector $\Delta\theta$ accumulates :
  - the contribution vectors $\Delta\theta_i$ of each priority level i,
    - for i = 1 to P
  - AND, at the lowest priority level, the projection of $\Delta\alpha$ on the Null space of all constraints $N(J_p^A)$

$$\Delta\theta = \sum_{i=1}^{p} \Delta\theta_i + P_{N(J_p^A)} \Delta\alpha$$

36

Extension to *P* priority levels (3)

- Initializations: $\Delta\theta_0 = 0$ $P_{N(J_0^A)} = I_n$
- Each priority level exploit the Augmented jacobian noted J_i^A that piles all Jacobians from levels 1 to level ***i***
- The contribution of the level ***i*** is given by:

$$\Delta\theta_i = (J_i P_{N(J_{i-1}^A)})^{+\lambda_i}(\Delta x_i - J_i \Delta\theta_{i-1})$$

37

Extension to *P* priority levels (4)

- Computation of the projection operators P_{N_i} [Slotine91] $P_{N(J_i^A)} = I_n - J_i^{A+} J_i^A$
- Incremental computation of the projection operators P_{N_i} [Baerlocher98]

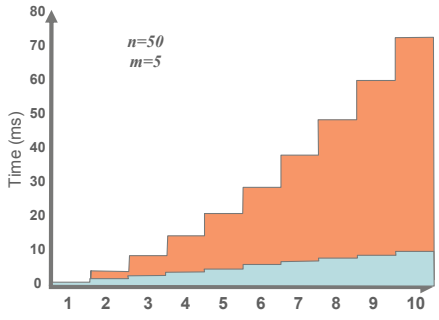
$$P_{N(J_i^A)} = P_{N(J_{i-1}^A)} - (J_i P_{N(J_{i-1}^A)})^{+}(J_i P_{N(J_{i-1}^A)})$$

Cost in *O(p)* compared to previous approach in $O(p^2)$

speedup factor=(p+1)/2

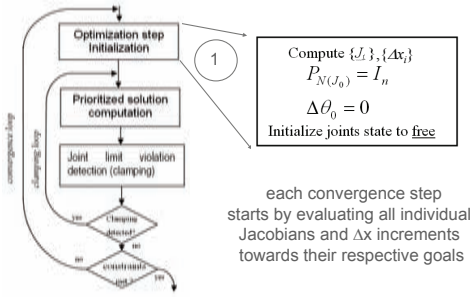
38

Computing cost vs Number of priority levels (p)



39

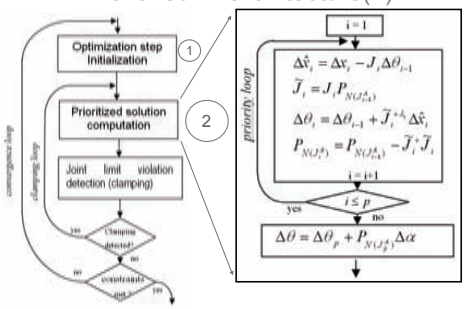
Prioritized IK architecture(1)



each convergence step starts by evaluating all individual Jacobians and Δx increments towards their respective goals

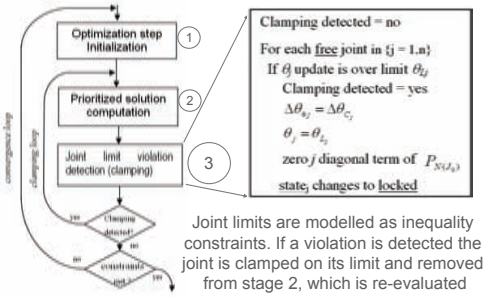
40

Prioritized IK architecture(2)



41

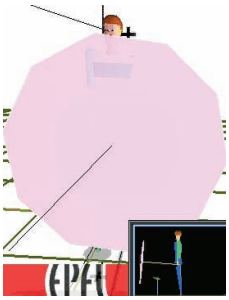
Prioritized IK architecture(3)



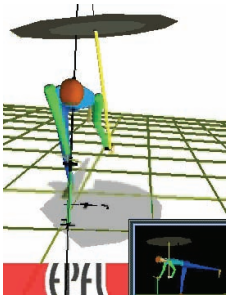
Joint limits are modelled as inequality constraints. If a violation is detected the joint is clamped on its limit and removed from stage 2, which is re-evaluated

42

From the start



with $\Delta\alpha$ term

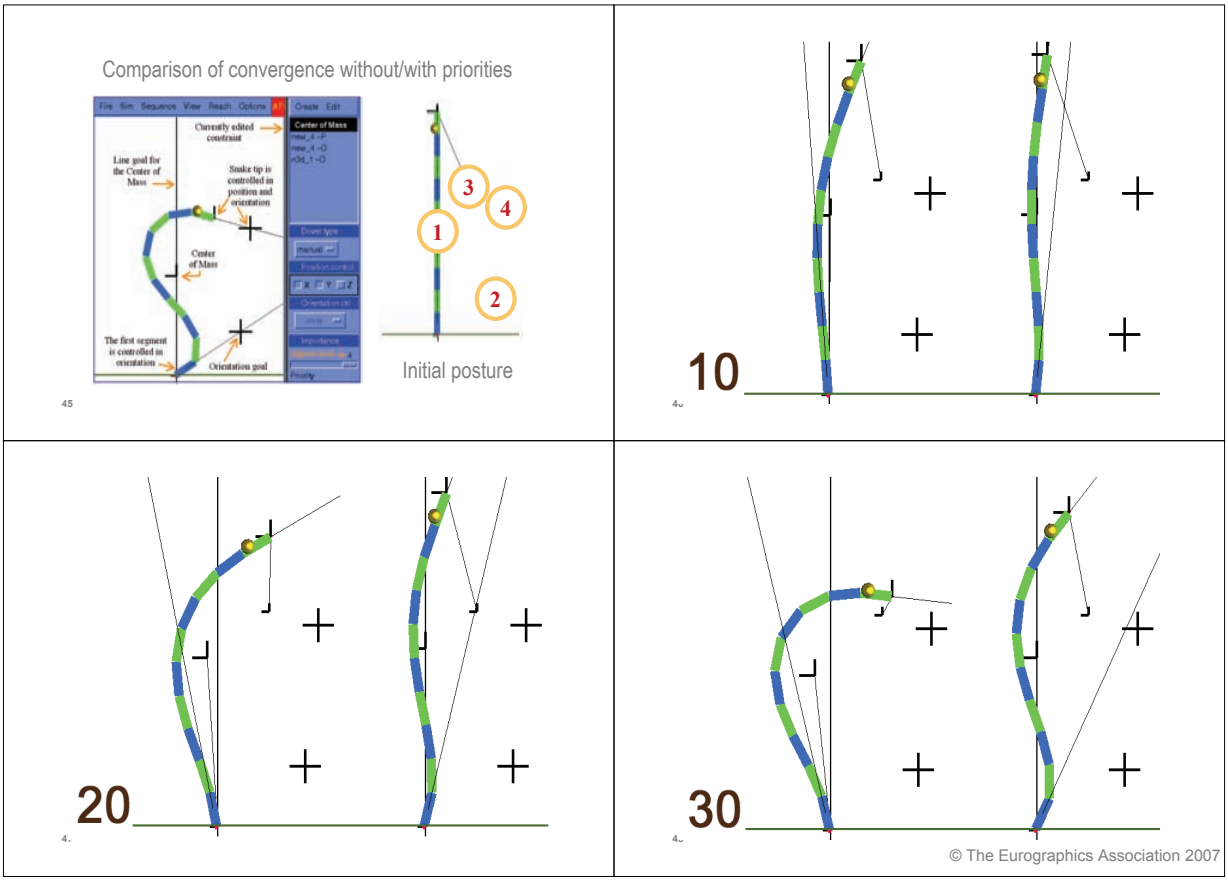


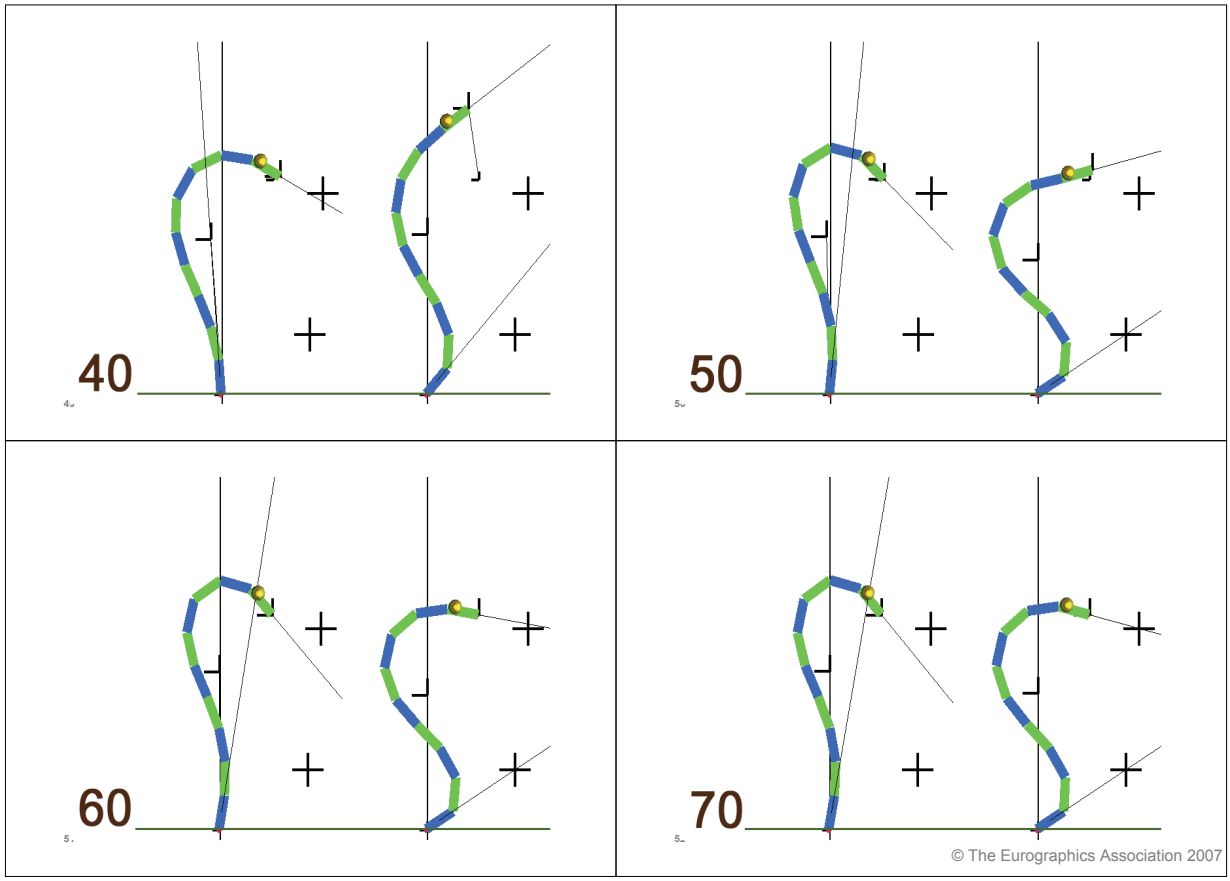
43

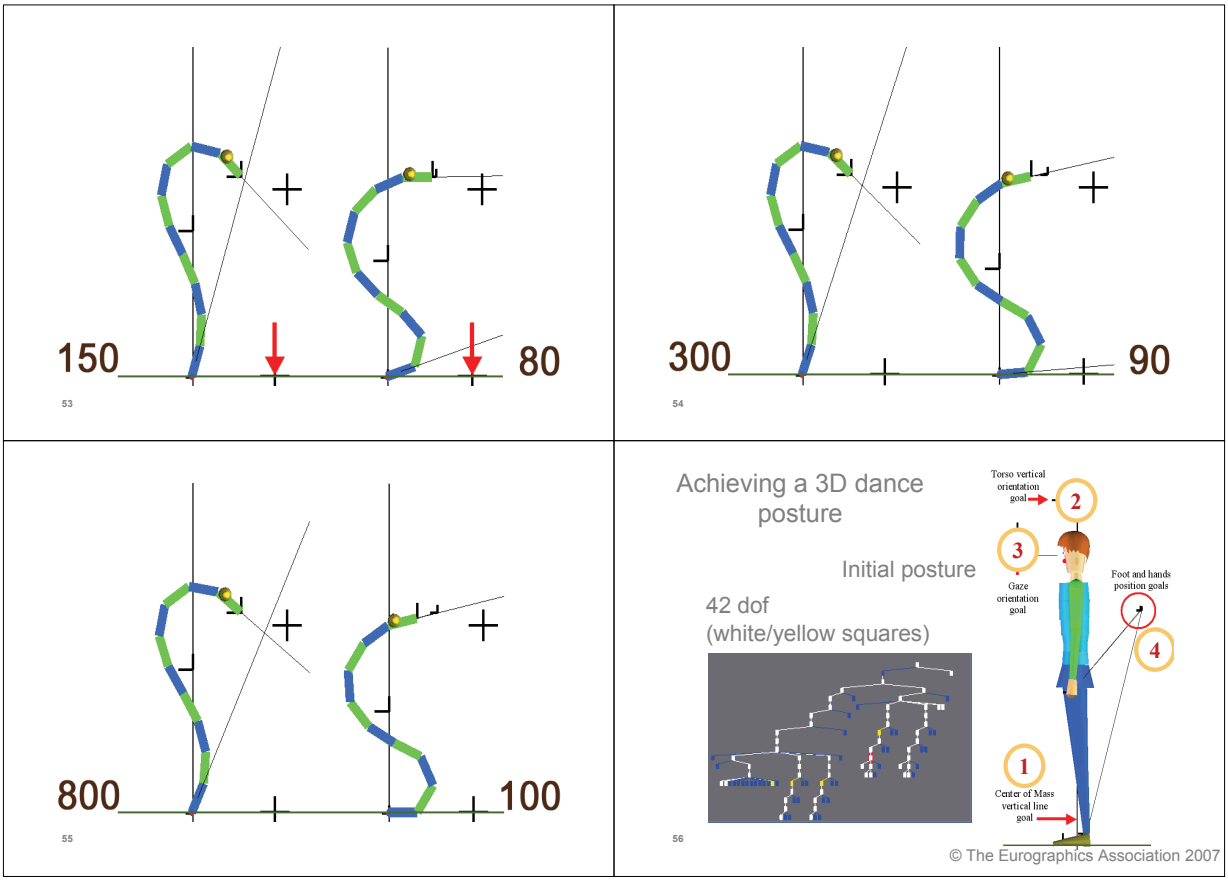
Convergence properties of the prioritized IK

- By construction, highest priority constraints are met first
- This property guides the local optimization through more realistic intermediate postures (compared to a weighting approach).
- Less likely to be stuck in local minima

44


Comparison of convergence without/with priorities
Currently edited constraint
Line goal for the Center of Mass
Snake tip is controlled in position and orientation
Center of Mass
The first segment is controlled in orientation
Orientation goal
1
2
3
4
Initial posture
10
20
30
© The Eurographics Association 2007

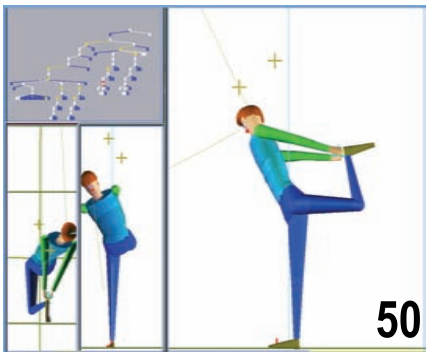




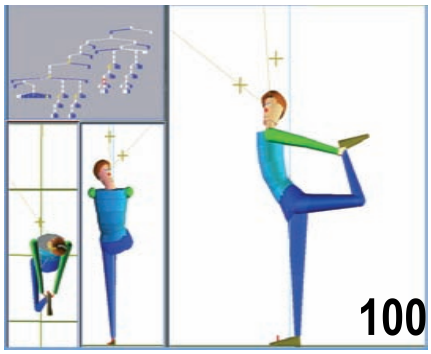
Weighting strategy

- 50 and 100 iterations
- Low realism of intermediate postures
- The final posture does not achieve all constraints

57



58

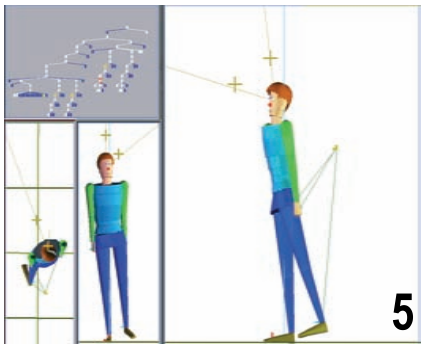


59

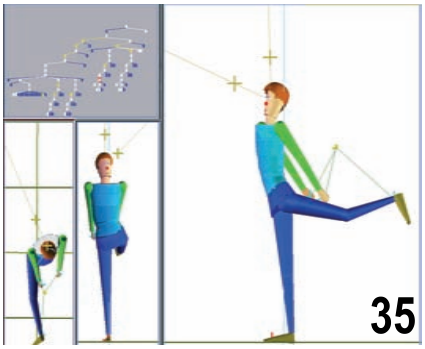
Priority strategy

- 5, 35 and 75 iterations
- Intermediate postures ensure balance
- The final posture achieves all constraints

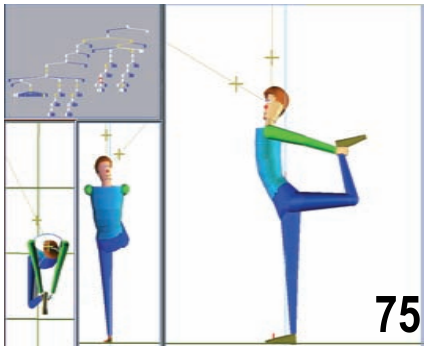
60



61



62



63

Part B - Iterative methods (pseudo-inv.)

- Concept of prioritized control
- From one to P priority levels
  - General form
  - Incremental constraint achievement
  - Example1: the simplest redundant case
  - Example 2: two prioritized constraints
  - Generalization to multiple priority levels
  - ⇨ Fostering synergies
  - Handling singularities with DLS
  - Application: collision avoidance
  - Conclusion

64

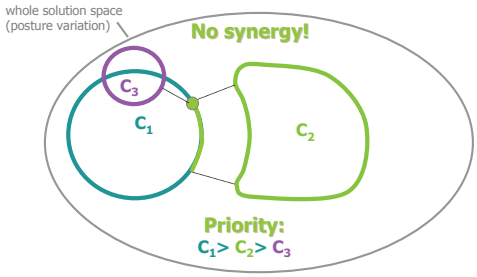
Adopting a **synergistic** approach

- Involve all the joints to satisfy all the constraints
  - larger solution space
  - can find optimal solution because all the joints can contribute to the satisfaction of the constraints
  - computing cost

65

Conceptual analogy

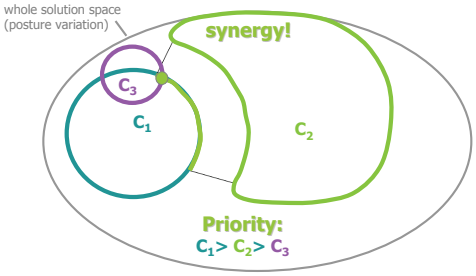
visualizing the constraints solution sub-spaces: C_i



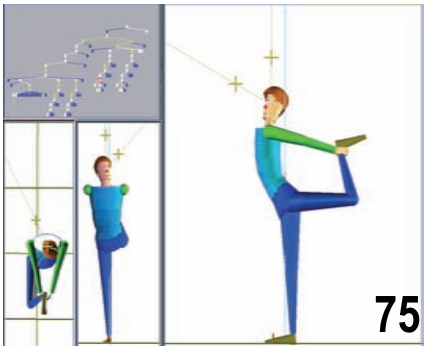
66

Conceptual analogy

visualizing the constraints solution sub-spaces: C_i

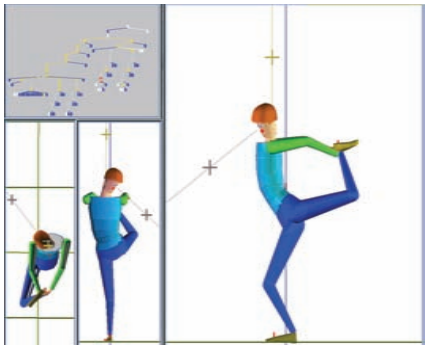


67



75

68



69

Part B - Iterative methods (pseudo-inv.)

- Concept of prioritized control
- From one to P priority levels
  - General form
  - Incremental constraint achievement
  - Example1: the simplest redundant case
  - Example 2: two prioritized constraints
  - Generalization to multiple priority levels
  - Fostering synergies
  - ⇨ Handling singularities with DLS
  - Application: collision avoidance
  - Conclusion

70

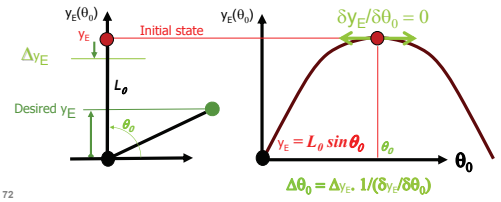
Handling geometric and algorithmic singularities

- Context: mismatch between the requested task Δx and the, temporary lower, rank of any Jacobian
- The solution is still valid in pure singular context :
  - When the rank of a jacobian J decreases, the dimension of its Null Space increases
  - The n-dimensional $\Delta\alpha$ vector can be used to avoid, or to get out of the singularity
  - Attention : Around the singularity, the norm of the solution increases: instability

71

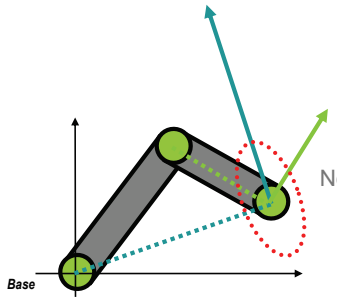
Geometric singularity: the 1 D case

- When the derivative is null
  - The inverse is undefined: ok for the pseudo-inverse
- When the derivative is very small
  - The inverse is very big: instability



72

Geometric singularity: the 2 D case

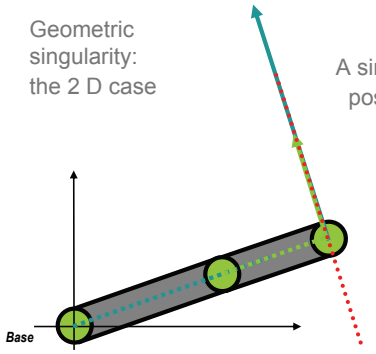


Non-singular posture

73

Geometric singularity: the 2 D case

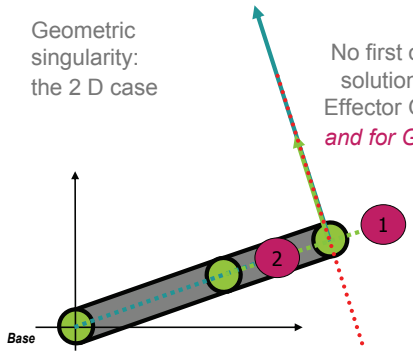
A singular posture



74

Geometric singularity: the 2 D case

No first order solution for Effector Goal1 *and for Goal2!*

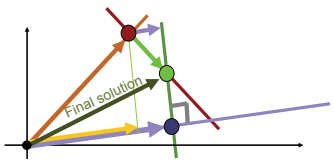


75

Algorithmic singularity (1)

Two conflicting Constraints [Hanafusa81]:

NOT SINGULAR

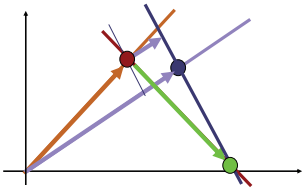


76

Algorithmic singularity (2)

When the low priority task wants to do the same as the high priority task

nearly SINGULAR



77

Managing singularities

- What is causing instability ?
  - Singular Value Decomposition of J

$$J = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

  - The Pseudo-Inverse J^+ is given by:

$$J^+ = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^T$$

78

Managing singularities (2)

- Regularization Technique [Maciejewsky90]

- The damped Pseudo-Inverse ensures that the norm of the solution is under a predefined value.

$$J^{+\lambda} = \sum_{i=1}^{r} \frac{\sigma_i}{\sigma_i^2 + \lambda^2} v_i u_i^T$$

  - The λ term damps the singular behavior of J in the neighborhood of the singularity
  - *The projectors on the Null space have however to be built with the J^+ pseudo-inverse*

79

Extension to *P* priority levels [Slotine91]

damping

$$\Delta\theta_i = (J_i P_{N(J_{i-1}^A)})^{+\lambda_i} (\Delta x_i - J_i \Delta\theta_{i-1})$$

80

NO damping

$$P_{N(J_i^A)} = P_{N(J_{i-1}^A)} - (J_i P_{N(J_{i-1}^A)})^+ (J_i P_{N(J_{i-1}^A)})$$

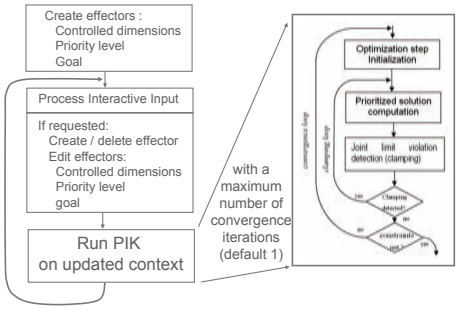
81

Part B - Iterative methods (pseudo-inv.)

- Concept of prioritized control
- From one to P priority levels
  - General form
  - Incremental constraint achievement
  - Example1: the simplest redundant case
  - Example 2: two prioritized constraints
  - Generalization to multiple priority levels
  - Fostering synergies
  - Handling singularities with DLS
  - ⇨ Application: collision avoidance
  - Conclusion

82

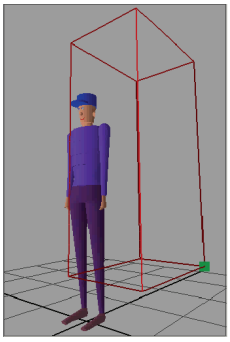
Interactive application architecture



83

Visualization of off-line use for identifying the lateral reachable space under balance constraints

[Peinado & al. 04]

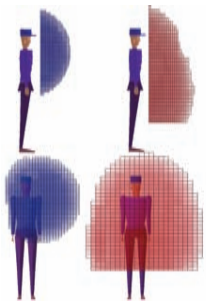


84

Example: standing reachable space of the left hand

blue: with only the arm

red: whole body

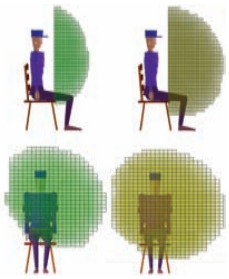


85

Example: seated reachable space of the left hand

green: naive approach

brown: with the hip joint

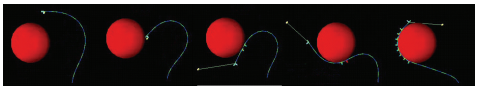


86

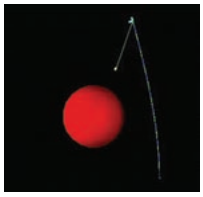
Collision avoidance

- Management of inequality constraint
- Dynamic creation of repulsion effectors when there is an effective collision
- A Repulsion effector :
  - Is a one-dimensional position control to push out the colliding point
  - has a higher priority compared to the chain tip attraction toward a user-guided goal

87



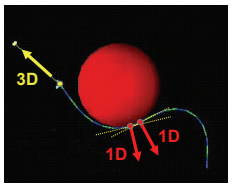
- Single effector
  - Chain tip



- Dynamically created 1D repulsion effectors with higher priority
  - may prevent the chain tip effector to reach its goal

88

- Chain tip effector
  - 3D position constraint toward an interactively set goal location

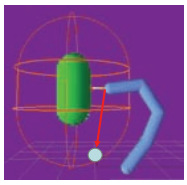


- 2 one dimensional repulsion effectors
  - Pushes away along the local normal
  - Free movement in the tangent plane

89

Smooth Collision avoidance [Peinado et al 05]

- Obstacles are surrounded by a damping zone
- Special points on the articulated structure are "observed"



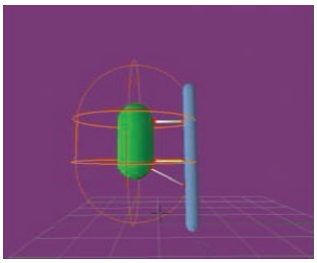
- An observer is promoted as a constrained one-dimensional dynamic damping effector whenever it is inside the damping zone and moves towards the obstacle

90

Smooth Collision avoidance [Peinado et al 05]

The chain tip is interactively guided towards successive goal position

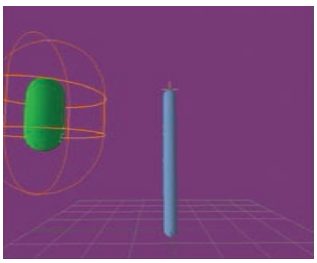
Links highlight the dynamic creation of 1D damping effectors to prevent collisions



91

Smooth Collision avoidance [Peinado et al 06]

Taking into account the relative speed along the normal to the obstacle allows to anticipate the arrival of a mobile obstacle



92

Part B - Iterative methods (pseudo-inv.)

- Concept of prioritized control
- From one to P priority levels
  - General form
  - Incremental constraint achievement
  - Example1: the simplest redundant case
  - Example 2: two prioritized constraints
  - Generalization to multiple priority levels
  - Fostering synergies
  - Handling singularities with DLS
  - Application: collision avoidance
  - ⇨ Conclusion

93

Conclusions & future work

- The prioritized IK provides a powerful way to combine various conflicting requirements such as collision avoidance, balance, gaze, reach:
  - When correctly chosen, the hierarchy of priorities guides the convergence through believable intermediate states, even if only a small subset of constraints are met.
- On-going work: integrating motion models to guide the convergence.

94

Inverse kinematics and kinetics for virtual humanoids

Ronan Boulic
VRLab – EPFL
http://vrlab.epfl.ch

Richard Kulpa
Bunraku – INRIA / IRISA
www.irisa.fr/bunraku

M2S – Univ. Rennes 2
www.uhb.fr/labos/m2s

Plan

- Part A - Overview
  - Motivation
  - Problem
  - Overview of the IK methods
- Part B - Iterative methods (pseudo-inverse based)
  - Concept of prioritized control
  - From one to P priority levels
- ⇨ Part C - Iterative methods (CCD-based)
  - CCD
  - Hybrid CCD and analytical
- Part D - Conclusion

Part C - Iterative methods (CCD-based)

- ⇨ CCD
- Hybrid CCD and analytical method
  - Representation of data
  - Overview of the method
  - Kinematic solver
  - Kinetic solver
  - Kinematic and kinetic solver
  - Results
  - Conclusion

CCD

- CCD (Cyclic Coordinate Descent)
- Goal of this kind of approach
  - Adaptation of the animation of hundreds of characters in real-time
- Compared to Jacobian
  - Solve only one DOF at a time

CCD

- Only one column remains in the Jacobian

$$\begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_M \end{pmatrix} = \begin{pmatrix} \frac{\partial fx_1}{\partial \theta_1} & \frac{\partial fx_1}{\partial \theta_2} & \cdots & \frac{\partial fx_1}{\partial \theta_N} \\ \frac{\partial fx_2}{\partial \theta_1} & \frac{\partial fx_2}{\partial \theta_2} & \cdots & \frac{\partial fx_2}{\partial \theta_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial fx_M}{\partial \theta_1} & \frac{\partial fx_M}{\partial \theta_2} & \cdots & \frac{\partial fx_M}{\partial \theta_N} \end{pmatrix} \begin{pmatrix} \cancel{\Delta\theta_1} \\ \Delta\theta_2 \\ \vdots \\ \cancel{\Delta\theta_N} \end{pmatrix}$$

- Inversion is fast: low computation cost

5

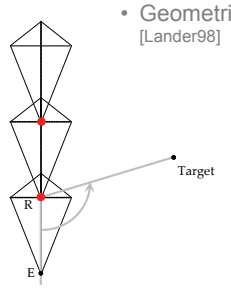
CCD

- General algorithm
  For all the DOF θ_i
    Compute variation $\Delta\theta_i$ of the DOF θ_i
    Add this variation: $\theta_i = \theta_i + \Delta\theta_i$
  End
- Issued from robotics works [Luenberger84, Wang91]
- Animation: recursive algorithm [Badler87]

6

CCD

- Geometrical resolution by CCD [Lander98]



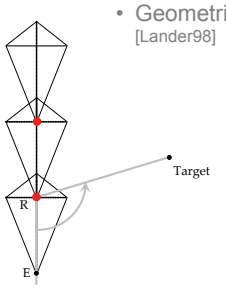
Example with 2 DOFs

E: effector
R: currently modified DOF

7

CCD

- Geometrical resolution by CCD [Lander98]



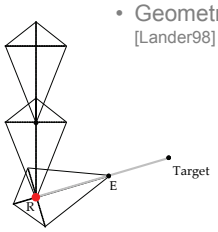
Example with 2 DOFs

E: effector
R: currently modified DOF

8

CCD

- Geometrical resolution by CCD
  [Lander98]



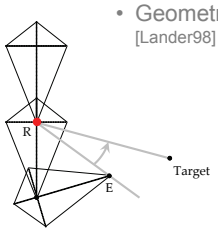
Example with 2 DOFs

E: effector
R: currently modified DOF

9

CCD

- Geometrical resolution by CCD
  [Lander98]



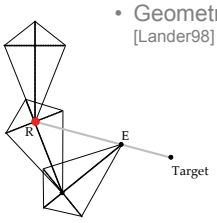
Example with 2 DOFs

E: effector
R: currently modified DOF

10

CCD

- Geometrical resolution by CCD
  [Lander98]



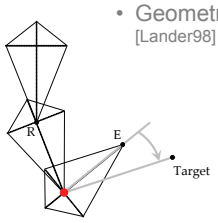
Example with 2 DOFs

E: effector
R: currently modified DOF

11

CCD

- Geometrical resolution by CCD
  [Lander98]



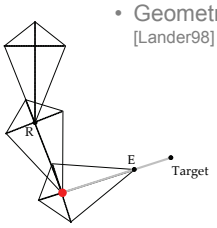
Example with 2 DOFs

E: effector
R: currently modified DOF

12

CCD

- Geometrical resolution by CCD
  [Lander98]



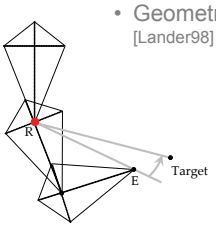
Example with 2 DOFs

E: effector
R: currently modified DOF

13

CCD

- Geometrical resolution by CCD
  [Lander98]



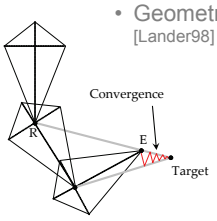
Example with 2 DOFs

E: effector
R: currently modified DOF

14

CCD

- Geometrical resolution by CCD
  [Lander98]



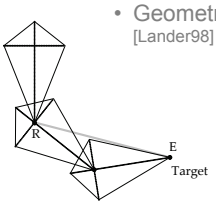
Example with 2 DOFs

E: effector
R: currently modified DOF

15

CCD

- Geometrical resolution by CCD
  [Lander98]



Example with 2 DOFs

E: effector
R: currently modified DOF

16

CCD

- Advantages
  - Fast computation of one iteration
  - Easy to implement
  - Handling of joint limits

17

CCD

- Drawbacks
  - Slow convergence
  - Bad distribution of the adaptation
  - Unnatural posture

Original posture | Desired posture | Resulting posture

18

CCD

- Drawbacks
  - Slow convergence
  - Bad distribution of the adaptation
  - Unnatural posture

⇒ First joints are more modified than the following ones!

19

CCD

- First solution: use damping
  - Threshold on the variation of the joint parameters
  - Minimizes the adaptation of each joint
  - But increases the number of iterations
- More homogeneous adaptation
  ⇒ Bigger computation cost

20

Not suitable for humanoids

- Our goal
  - Find natural postures
  - Computation time compatible with interactive animation of hundreds of characters
- CCD is not suitable for postural adaptation of humanoids
- Here, we present an hybrid CCD and analytical solver based on body groups

21

Part C - Iterative methods (CCD-based)

- CCD
- ⇨ Hybrid CCD and analytical method
  - Representation of data
  - Overview of the method
  - Kinematic solver
  - Kinetic solver
  - Kinematic and kinetic solver
  - Results
  - Conclusion

22

Hybrid CCD – Analytical method

- Idea 1:
  - Improve its main advantage: fast computation

⇒ Use analytical solutions
⇒ Decrease the number of segments

23

Hybrid CCD – Analytical method

- Idea 2:
  - Take advantage of its main drawback: bad distribution of the adaptation

⇒ Most modified segments: first adapted ones
⇒ Use body groups
⇒ Order the adaptation of these groups
⇒ Lead to natural postures

24

Hybrid CCD – Analytical method

- Idea 2:
  - Take advantage of its main drawback: bad distribution of the adaptation

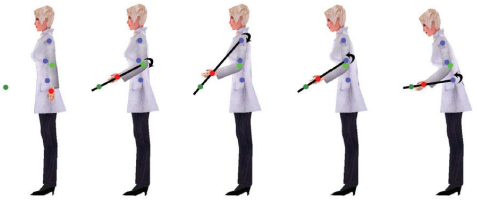
⇒ Use analytical solutions for the kinematic chain ends

⇒ Instantaneously have its best configuration

25

Hybrid CCD – Analytical method

- Without analytical solutions for the arm



26

Hybrid CCD – Analytical method

- With analytical solutions for the arm

27

Hybrid CCD – Analytical method

- With analytical solutions for the arm

⇒ Minimizes the adaptation of the remainder of the body

28

Hybrid CCD – Analytical method

- Here, we present an hybrid algorithm
  - CCD-like algorithm
  - Body groups to order the adaptations
  - Analytical solutions

29

Part C - Iterative methods (CCD-based)

- CCD
- Hybrid CCD and analytical method
  - ⇨ Representation of data
  - Overview of the method
  - Kinematic solver
  - Kinetic solver
  - Kinematic and kinetic solver
  - Results
  - Conclusion

30

Representation of data

- Representation of the constraints
  - How to handle many kind of constraints such as position, orientation, space
- Representation of the skeleton
  - Find a suitable representation of the kinematic chain for adaptations

31

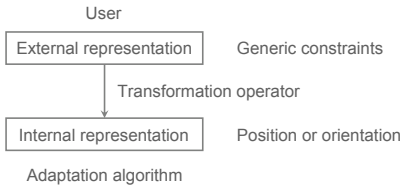
Representation of constraints

- For performance purpose, algorithm deals only with geometric constraints (position / orientation)
- Motion: succession of postures
  => also deals with kinematic constraints (such as speed for example)

32

Representation of constraints

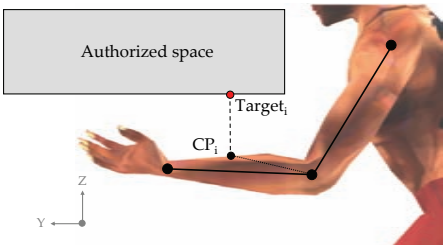
- In order to take all type of constraints into account:
  - Two levels of representation



33

Internal representation

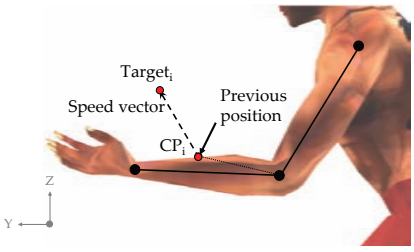
- Example of transformation operator for an authorized space constraint



34

Internal representation

- Example of transformation operator for a speed constraint



35

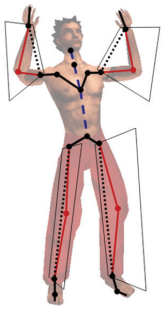
Representation of data

- Representation of the constraints
  - How to handle many kind of constraints such as position, orientation, space
- Representation of the skeleton
  - Find a suitable representation of the kinematic chain for adaptations

36

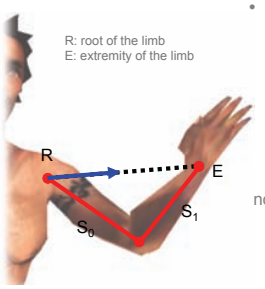
Normalized skeleton

- Really suitable for adaptations [Kulpa05a]
- Skeleton representation independent from morphology
- Skeleton subdivided in:
  - Limbs
  - Spine
  - Other segments



37

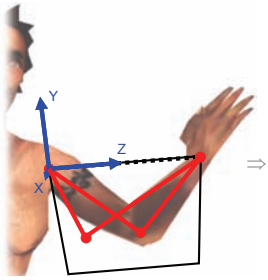
Normalized skeleton



- Limbs are independent from the initial length
  - Use the vector from the shoulder to the wrist
  - Normalize it by the arm length

$$\text{normalizedKC} = \frac{E - R}{\sum_j \text{length}(S_j)}$$

Normalized skeleton



- Limbs are independent from the ratio between segment lengths
  - Store the half-plane that contains the elbow

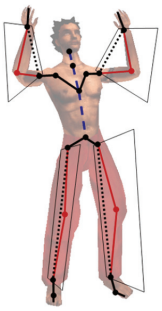
⇒ Limbs are defined by:

- Frame set: define the vector and the half-plane
- Scalar: percentage of the total limb length

Normalized skeleton

- The spine is represented by a spline
  - ⇒ Can be subdivided into any number of vertebraes
- Other segments are normalized by the original length

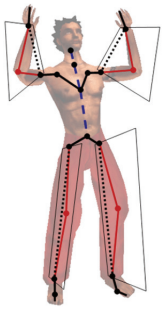
$$\text{normalizedKC} = \frac{E - R}{\sum_j \text{length}(S_j)}$$



40

Normalized skeleton

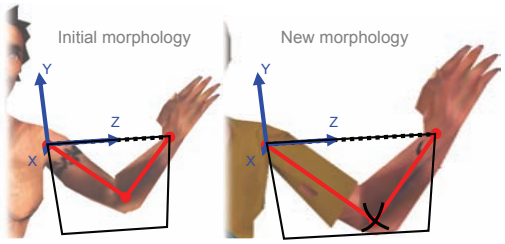
- Root position is normalized by the leg length (should be extended if walking on the hands for example)
- Root orientation is subdivided into :
  - Global orientation
    Ex: direction of the walk
  - Local orientation
    Ex: twist of the pelvis around the global direction



41

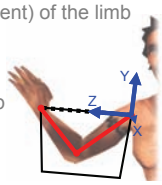
Retargeting

- Invert the normalization process



Representation of the skeleton

- All the analytical solutions are based on this representation
- Example 1: extend a limb
  - Change the scalar (length percent) of the limb
- Example 2: reach a target
  - Change the Z vector of the limb
  - Change the scalar of the limb

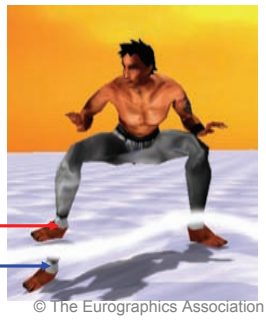


43

Representation of the skeleton

- Example 3: placing the feet on the ground
  [Kulpa05a]

Original posture

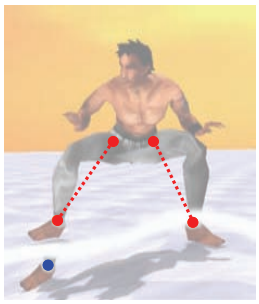


44

Representation of the skeleton

- Example 3: placing the feet on the ground [Kulpa05a]

Normalized limbs

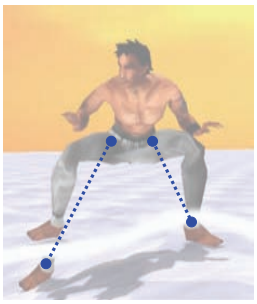


45

Representation of the skeleton

- Example 3: placing the feet on the ground [Kulpa05a]

Changing:
- Vector of the limb
- Scale of the limb



46

Representation of the skeleton

- Example 3: placing the feet on the ground [Kulpa05a]

Final posture

47

Part C - Iterative methods (CCD-based)

- CCD
- Hybrid CCD and analytical method
  - Representation of data
  - Overview of the method
  - Kinematic solver
  - Kinetic solver
  - Kinematic and kinetic solver
  - Results
  - Conclusion

48

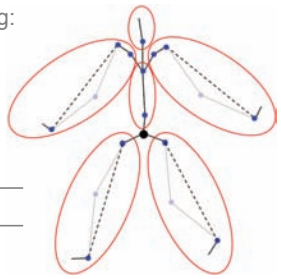
Adaptation algorithm

- Idea: take advantage of the different body parts
  - More or less movable
  - More or less heavy
- Subdivide the skeleton into groups
- Order the adaptation in order to take advantage of the CCD unhomogeneous distribution
- Unique representation compatible with :
  - Inverse kinematics
  - Inverse kinetics

49

Hierarchy of groups

- Split the skeleton into groups
- 2 levels of ordering:
  - Between groups
  - Within groups
- Hierarchy:



1

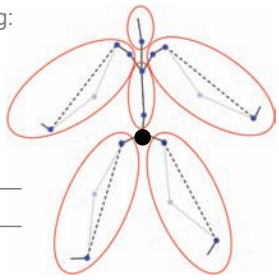
2

3

50

Hierarchy of groups

- Split the skeleton into groups
- 2 levels of ordering:
  - Between groups
  - Within groups
- Hierarchy:



1 Root

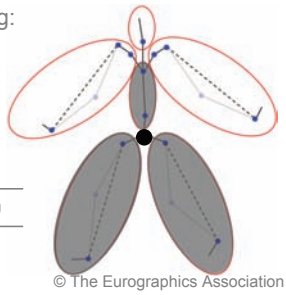
2

3

51

Hierarchy of groups

- Split the skeleton into groups
- 2 levels of ordering:
  - Between groups
  - Within groups
- Hierarchy:



1 Root

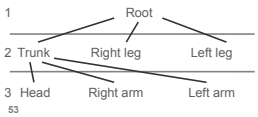
2 Trunk Right leg Left leg

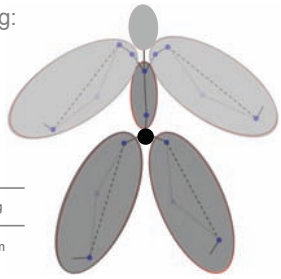
3

52

Hierarchy of groups

- Split the skeleton into groups
- 2 levels of ordering:
  - Between groups
  - Within groups
- Hierarchy:





53

Adaptation order of the groups

- Minimize the adaptation error
- Inverse kinematics
  - Groups with the largest range first
  - Level 3 to 1
- Inverse kinetics
  - Heaviest groups first
  - Level 1 to 3
- Order used between and within the groups

54

Part C - Iterative methods (CCD-based)

- CCD
- Hybrid CCD and analytical method
  - Representation of data
  - Overview of the method
  - ⇨ Kinematic solver
  - Kinetic solver
  - Kinematic and kinetic solver
  - Results
  - Conclusion

55

Kinematic adaptation order

- Bad kinematic order:
  heaviest segments first

56

Kinematic adaptation order

- Bad kinematic order: heaviest segments first

57

Kinematic adaptation order

- Bad kinematic order: heaviest segments first

58

Kinematic adaptation order

- Good kinematic order: lightest segments first

59

Kinematic adaptation order

- Good kinematic order: lightest segments first

60

Kinematic adaptation order

- Good kinematic order: lightest segments first

61

Kinematic adaptation order

- Good kinematic order: lightest segments first

62

Kinematic adaptation order

- Good kinematic order: lightest segments first

63

Inverse kinematics algorithm

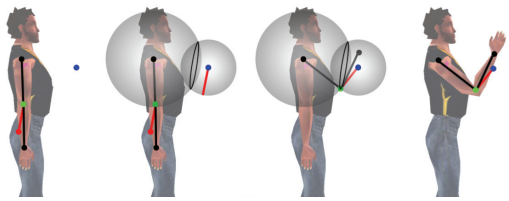
- Inverse kinematics algorithm
- Do
  - groupAdaptation(Head)
  - groupAdaptation(Right arm)
  - groupAdaptation(Left arm)
  - groupAdaptation(Right leg)
  - groupAdaptation(Left leg)
  - groupAdaptation(Trunk)
  - rootAdaptation()
- While (¬ended)

Level 3: groupAdaptation(Head), groupAdaptation(Right arm), groupAdaptation(Left arm)
Level 2: groupAdaptation(Right leg), groupAdaptation(Left leg), groupAdaptation(Trunk)
Level 1: rootAdaptation()

64

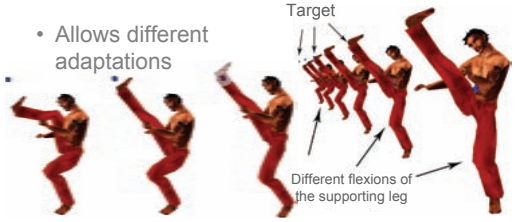
Group adaptation

- Analytical solutions for limbs (for weighted or prioritized constraints)
- Rotations for the remainder of the body



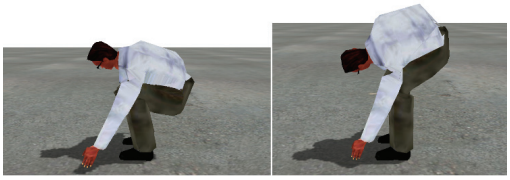
Root adaptation

- For the constraints applied on the pelvis
- To reach far constraints (additional parameter)
- Allows different adaptations



Root adaptation

- Allows the use of biomechanical laws
  - Use of individual parameters (such as joint range…)



67

Part C - Iterative methods (CCD-based)

- CCD
- Hybrid CCD and analytical method
  - Representation of data
  - Overview of the method
  - Kinematic solver
  - ⇨ Kinetic solver
  - Kinematic and kinetic solver
  - Results
  - Conclusion

68

Kinetic adaptation

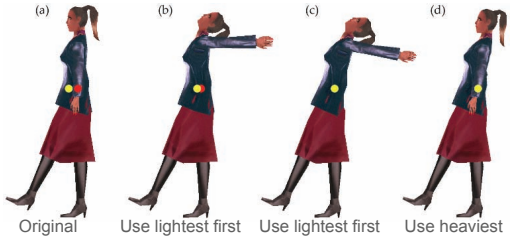
- Inverse kinetics only considers the groups not constrained (not used by the inverse kinematics solver)
- Among these groups, the user can select only a subset

⇒ Can tune the inverse kinetics adaptation

69

Kinetic adaptation order

- Good kinetic order: heaviest first

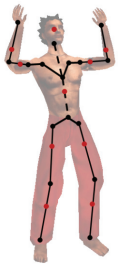


Original posture | Use lightest first (head then arms) | Use lightest first (finally trunk) | Use heaviest first (trunk)

70

Center of mass hierarchy

- Different levels of COM [Kulpa05b]



Weigthed sum so can be subdivided according to the groups

$$COM = \frac{\sum m_S COM_S}{\sum m_S}$$

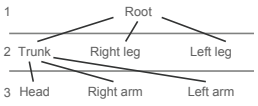
71

Center of mass hierarchy

- Different levels of COM [Kulpa05b]

Weigthed sum so can be subdivided according to the groups

$$COM = \frac{\sum m_S COM_S}{\sum m_S}$$



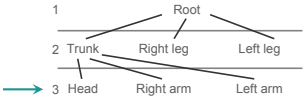
72

Center of mass hierarchy

- Different levels of COM [Kulpa05b]

Weigthed sum so can be subdivided according to the groups

$$COM = \frac{\sum m_S COM_S}{\sum m_S}$$



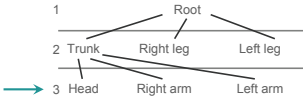
73

Center of mass hierarchy

- Different levels of COM [Kulpa05b]

Weigthed sum so can be subdivided according to the groups

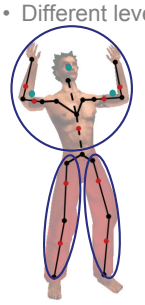
$$COM = \frac{\sum m_S COM_S}{\sum m_S}$$



74

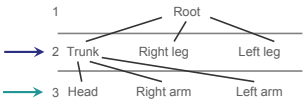
Center of mass hierarchy

- Different levels of COM [Kulpa05b]



Weigthed sum so can be subdivided according to the groups

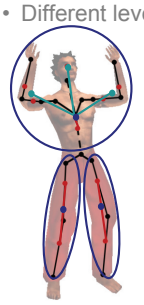
$$COM = \frac{\sum m_S COM_S}{\sum m_S}$$



75

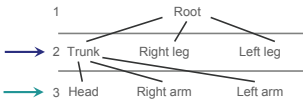
Center of mass hierarchy

- Different levels of COM [Kulpa05b]



Weigthed sum so can be subdivided according to the groups

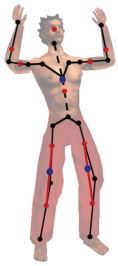
$$COM = \frac{\sum m_S COM_S}{\sum m_S}$$



76

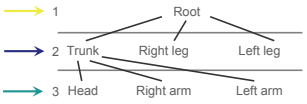
Center of mass hierarchy

- Different levels of COM [Kulpa05b]



Weigthed sum so can be subdivided according to the groups

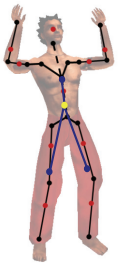
$$COM = \frac{\sum m_S COM_S}{\sum m_S}$$



77

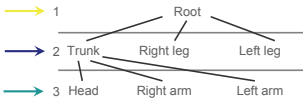
Center of mass hierarchy

- Different levels of COM [Kulpa05b]



Weigthed sum so can be subdivided according to the groups

$$COM = \frac{\sum m_S COM_S}{\sum m_S}$$



78

Adaptation of the group G

1. Error between COM and COM'

79

Adaptation of the group G

1. Error between COM and COM'
2. Computation of desired COM'$_G$

$$COM' = \frac{\sum_g (m_g COM_g)}{\sum_g m_g}$$

80

Adaptation of the group G

1. Error between COM and COM'
2. Computation of desired COM'$_G$

$$COM' = \frac{\sum_g (m_g COM_g)}{\sum_g m_g}$$

$$COM' = \frac{\sum_{g \neq G} (m_g COM_g)}{\sum_g m_g} + \frac{m_G COM'_G}{\sum_g m_g}$$

81

Adaptation of the group G

1. Error between COM and COM'
2. Computation of desired COM'$_G$

$$COM' = \frac{\sum_g (m_g COM_g)}{\sum_g m_g}$$

$$COM' = \frac{\sum_{g \neq G} (m_g COM_g)}{\sum_g m_g} + \frac{m_G COM'_G}{\sum_g m_g}$$

$$COM'_G = \frac{\sum_g m_g * COM' - \sum_{g \neq G} (m_g COM_g)}{m_G}$$

82

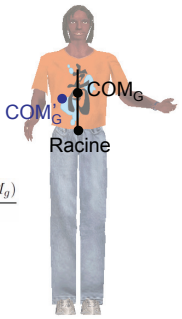
Adaptation of the group G

1. Error between COM and COM'
2. Computation of desired COM'$_G$

$$COM' = \frac{\sum_g (m_g COM_g)}{\sum_g m_g}$$

$$COM' = \frac{\sum_{g \neq G} (m_g COM_g)}{\sum_g m_g} + \frac{m_G COM'_G}{\sum_g m_g}$$

$$COM'_G = \frac{\sum_g m_g * COM' - \sum_{g \neq G} (m_g COM_g)}{m_G}$$



83

Adaptation of the group G

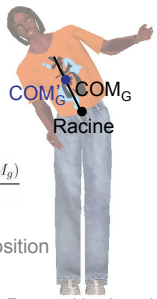
1. Error between COM and COM'
2. Computation of desired COM'$_G$

$$COM' = \frac{\sum_g (m_g COM_g)}{\sum_g m_g}$$

$$COM' = \frac{\sum_{g \neq G} (m_g COM_g)}{\sum_g m_g} + \frac{m_G COM'_G}{\sum_g m_g}$$

$$COM'_G = \frac{\sum_g m_g * COM' - \sum_{g \neq G} (m_g COM_g)}{m_G}$$

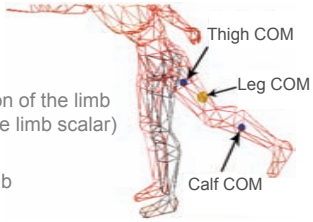
3. Adaptation of the group G
   to place the COM at the right position



84

Adaptation of the groups

- Analytical solution for limbs



- Two steps:
  - Find the flexion of the limb (defined by the limb scalar)
  - Rotate the limb

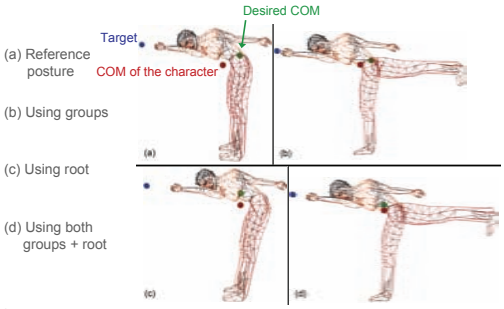
85

Root adaptation

- Translation of:
  - 3D vector (COM' – COM) for aerial phases: can handle real COM trajectory such as parabola
  - 2D vector (COM' – COM) for support phases: can preserve balance handling the projection of the COM on the ground
- With:
  - COM: current COM of the posture
  - COM': desired COM

86

Parameterization of the adaptations



87

Part C - Iterative methods (CCD-based)

- CCD
- Hybrid CCD and analytical method
  - Representation of data
  - Overview of the method
  - Kinematic solver
  - Kinetic solver
  - ⇨ Kinematic and kinetic solver
  - Results
  - Conclusion

88

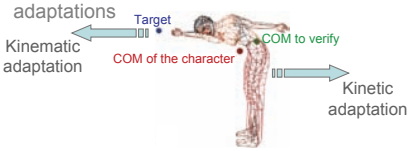
IKK algorithm

- If the user activates the kinetic adaptation, we consider that the balance must be preserved
- Do
  KinematicAdaptation()
  If (error COM > threshold)
  kineticAdaptation()
  End of if
  While (¬ended)

89

IKK algorithm

- Always converges because:
  - Kinematic and kinetic adaptations don't modify the same group (kinetic adaptation only applied on available groups: with no constraints on them)
  - Inverse kinematics and kinetics are antagonist adaptations



90

Part C - Iterative methods (CCD-based)

- CCD
- Hybrid CCD and analytical method
  - Representation of data
  - Overview of the method
  - Kinematic solver
  - Kinetic solver
  - Kinematic and kinetic solver
  - Results
  - Conclusion

91

Results

- Comparison with a real posture

92

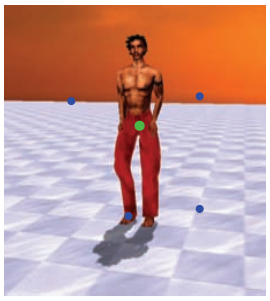
Results

- Comparison with a real posture

93

Results

- Comparison with a real posture



94

Results

- Comparison with a real posture

95

Results

- Comparison with a real posture

96

Results

- Comparison with a real posture

Importance of the reference posture!

97

Results

- The reference posture is also used to find the knees and the elbows
- Works even if the limbs are straightened in the reference posture
- Indeed, limb is represented by a frame and a scalar
  - The frame gives the reference orientation of the limb

98

Results

- Real-time adaptation of many characters

99

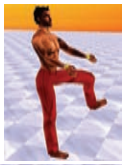
Results

- Additional masses

100

Performances

- Performances at 25 Hz (laptop - P4 2.8GHz)
  - Simple case (3 groups):
    - 454 characters (kinematics)
    - 212 characters (kinematics and kinetics)
  - Complex case (4 groups):
    - 90 characters (kinematics)
    - 35 characters (kinematics and kinetics)



101

Part C - Iterative methods (CCD-based)

- CCD
- Hybrid CCD and analytical method
  - Representation of data
  - Overview of the method
  - Kinematic solver
  - Kinetic solver
  - Kinematic and kinetic solver
  - Results
  - Conclusion

102

Conclusion

- Hybrid CCD on groups – Analytical
- Advantages
  - Very fast
  - Give natural postures
  - Easy to implement
  - Different adaptations for each characters
- Drawback
  - Cannot deal with biomechanical laws such as energy minimization

103

Inverse kinematics and kinetics for virtual humanoids

Ronan Boulic
VRLab – EPFL
http://vrlab.epfl.ch

Richard Kulpa
Bunraku – INRIA / IRISA
www.irisa.fr/bunraku

M2S – Univ. Rennes 2
www.uhb.fr/labos/m2s

Plan

- Part A - Overview
  - Motivation
  - Problem
  - Overview of the IK methods
- Part B - Iterative methods (pseudo-inverse based)
  - Concept of prioritized control
  - From one to P priority levels
- Part C - Iterative methods (CCD-based)
  - CCD
  - Hybrid CCD and analytical
- ⇨ Part D - Conclusion

Conclusion

- Many IK methods
  - Analytical
    - Fast
    - Easy to implement
    - For a maximum of 7 DOF
  - Optimization
    - No singularity
    - Can fall to a local minimum (not the best solution)
    - Cannot handle prioritized constraints

Conclusion

- Many IK methods
  - Iterative method (simultaneous)
    - Jacobian transpose
      - Fast
      - Unrealistic results
    - Jacobian pseudo-inverse
      - Precise
      - Can handle biomechanical laws
      - Can preserve balance
      - Manage prioritized and weighted constraints
      - For a few characters at real-time

Conclusion

- Many IK methods
  - Iterative method (sequential)
    - CCD
      - Fast
      - Unrealistic adaptation
    - Hybrid CCD and analytical method on a hierarchy of groups
      - Fast
      - Realistic results
      - Can preserve balance
      - Manage prioritized and weighted constraints
      - Easy to implement
      - Cannot handle all the biomechanical laws

5

Conclusion

- Choice of IK method depends on several parameters:
  - Speed (interactive applications vs. simulations)
  - Natural result (humans vs. robots)
  - Complexity of the implementation

6

Conclusion

- Comparison of the two key approaches able to manage prioritized constraints

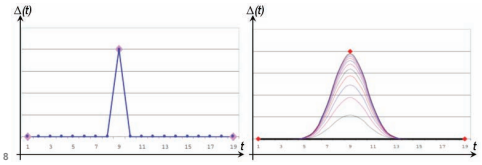
Method	Performance	Realism	COM handling	Biomechanical laws
J Pseudo-inverse	+	+++	+++	+++
CCD + analytical	+++	+++	+++	+

Analytical not presented since it is limited to 7DOF

7

Conclusion

- Previous IK methods can also be used for motion adaptations
  - Directly for small adaptations
  - Using displacement maps for large adaptations
    [Bruderlin95, Gleicher97, Gleicher01, Le Callennec04]



8

The Visual Computer

An inverse kinematics architecture enforcing an arbitrary number of strict priority levels

Paolo Baerlocher,
Ronan Boulic

[1] NEKO Entertainment, 12-16 rue de Vincennes, 93100 Montreuil, France
E-mail: pbaerlocher@e-neko.com
[2] VRlab, EPFL, 1015 Lausanne, Switzerland
E-mail: ronan.boulic@epfl.ch

Published online: 22 June 2004

An efficient inverse kinematics solver is a key element in applications targeting the on-line or off-line postural control of complex articulated figures. In the present paper we progressively describe the strategic components of a very general and robust inverse kinematics architecture. We then present an efficient recursive algorithm enforcing an arbitrary number of strict priorities to arbitrate the fulfillment of conflicting constraints. Due to its local nature, the moderate cost of the solution allows this architecture to run within an interactive environment. The algorithm is illustrated on the postural control of complex articulated figures.

Key words: Articulated figures – Synergistic posture control – Inverse kinematics – Conflicting tasks – Priority strategy

Abbreviations: IK, inverse kinematics; SVD, singular value decomposition

1 Introduction

The production of believable animations is an expensive process. For some classes of animation with high self-interaction and/or interaction with the environment, automated techniques are required to propose solutions to the intricate problem of finding the most believable posture for a set of conflicting constraints. A large number of approaches, both off-line and on-line, have been described. In the present paper we focus on inverse kinematics (IK) postural control as one essential component of the motion manipulation pipeline. While it does not solve the whole animation design problem, we advocate reconsidering its potential in light of its efficiency. The present article progressively describes a robust architecture handling complex cases – including in the singular context.

Our approach deals with general tree-structured articulated systems relying on a minimal set of coordinates, i.e., including only mechanical degrees of freedom between adjacent bodies. The corresponding space is called the joint space (as used in robotics). The major criterion behind this choice is performance. For generality reasons, the IK technique we exploit requires matrix inversions, thus it is important to minimize the dimension of the controlled system. An additional aspect to master for believability is to produce postures within each joint's anatomic range of motion. Thus we integrate a clamping algorithm within our architecture.

The second essential requirement for our IK architecture is the causal nature of its solution – depending only on the present and past states – allowing its use in an on-the-fly fashion. Within that framework, it is possible to add and remove any number of constraints at any time. However, constraints are seldom orthogonal, leading to conflicts regarding to their fulfillment. One major novelty of the present paper is to describe an efficient resolution scheme enforcing an arbitrary number of priority levels among constraints. Figure 1 illustrates a postural control application based on this architecture (details in Sect. 4).

A key argument of this paper is the following: The performance of an IK algorithm must be examined together with the quality of the convergence. At first sight, the computing cost of our architecture might appear to be too high, but this is the cost of guaranteeing essential properties for fast convergence towards constraints enforcement. We show in the results section that our IK solver can interactively handle the postural control of a human-like structure

The Visual Computer (2004) 20:402–417
Digital Object Identifier (DOI) 10.1007/s00371-004-0244-4

Fig. 1. Squatting posture combining four levels of conflicting constraints

with forty degrees of freedom and a 15-dimensional set of constraints.
The next section briefly covers the state of the art in inverse kinematics. The architecture of our multiple priority levels IK is presented in Sect. 3 with a special focus on performance analysis. Then, Sect. 4 illustrates the potential of this architecture within a test bed application dedicated to interactive postural control. Section 5 concludes the paper.

2 Previous work

2.1 Trends in the animation production process

The goal of any animation technique is to generate believable motion. Technically, the simplest solution is to smoothly interpolate key postures over time to generate the full set of frames required for an animation, the animator being free to adjust the key postures if the animation is not satisfying. This is known as the keyframing technique [27]. Although very tedious, this technique is still popular among animators because they can master the creative process up to the finest details of the joint trajectories. However, the time and skill required to produce convincing animations become too critical when trying to animate a large variety of characters. In addition, it is an off-line process. The production cost problem can be partly solved by capturing and retargeting motions to different articulated figures [4, 10, 12, 14, 22, 25]. The requirement of on-line techniques is necessary for the control of autonomous characters; it asks for causal methods relying only on the present and past states of the system. Few approaches presently display this property [7].
Among the numerous sub-problems lying within this short overview of the animation production challenge, we have chosen to focus on the inverse kinematics component. This component is responsible for proposing a believable postural state for a given set of constraints. Whether this proposed state is adopted depends on other aspects of the control strategy such as the dynamics of the system, but this is beyond the scope of the present paper.

2.2 Background in inverse kinematics

Ideally, an analytic formulation should express the relation between the constraints and the joint variables. General closed forms do not exist, but when considered independently, the limbs of 3-D characters can exploit a robust and fast analytical solution [13, 14, 26].
In the most general case, IK can be formulated as a constrained optimization problem and thus solved with nonlinear programming methods. For real-time or interactive applications, local optimization methods are preferred to the much more expensive global optimization methods even if there is a risk of getting stuck in a local optimum of the objective function. In computer graphics, Zhao and Badler [30] use an optimization method for the manipulation of an articulated figure. A nonlinear function (describing the degree of satisfaction of all constraints) is minimized under a set of linear equality and inequality constraints describing joint limits. An application of this method for the manipulation of articulated figures is integrated within the Jack system [21].
Among the qualities of any method, it is highly desirable to simultaneously manage multiple constraints. Very often, some constraints cannot be satisfied at the same time, whereas they can be satisfied independently. This conflicting situation is solved with either of the two following strategies (Fig. 2):

- Weights can be assigned to each task to define their relative importance and a compromise is found that satisfies none of the tasks exactly.

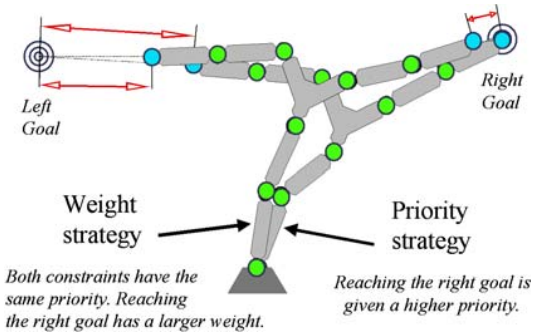


Fig. 2. Two strategies for solving conflicting constraints

- Tasks are optimized at different priority levels. In this mode, every task is satisfied as much as possible but without affecting the satisfaction of more important tasks.

For the positioning and animation of articulated figures, the weighting strategy is most frequently employed. Some typical examples are given by Badler et al. [1, 21, 30]. In the field of robotics however, researchers have developed *task-priority strategies* to precisely arbitrate conflicts by establishing a clear priority order among the tasks [11, 15, 16, 20, 24]. In this family of approaches, the architecture of the optimization solver is simpler but still allows one to build the projection operator on the Jacobian null space if the inverse has the required properties. The partition of the joint variation space into $N(J)$ and its complementary offers the potential for two levels of priority; this was first exploited by Liégeois [15].
One of the first applications of this strategy in computer animation was by Girard and Maciejewski [9]. The feet of multi-legged figures are constrained to remain in contact with the floor, while a kinematic locomotion model controls the coordination of the legs. Later, Maciejewski [18] highlighted the ill-conditioning of the Jacobian matrix near singular configurations and its frequent occurrence in computer animation (whenever there are unreachable goals). The proposed damped least squares inverse exploits the singular value decomposition (SVD) [23]. Instead of inverting small singular values, the method proposes to build a smooth function converging to zero when the singular value tends towards zero. Such a property has allowed the use of a two-priority-levels IK architecture for off-line motion retargeting [19]. The idea is to let the user specify a temporary modification of the trajectory of some end effectors (in fact any part of the articulated structure). These are the high priority constraints while the original joint space trajectory is realized as much as possible at the lowest priority level. Apart from being off-line, this approach has difficulty handling the problem of continuity in the case of multiple effectors sharing a common subset of joints. Another work, restricted to posture optimization, proposed a multiple-priority-levels architecture for combining end effector control and center of mass position control [5] (i.e., to ensure the static balance of the figure). However, the number of strictly independent priority levels is limited to two [2]. A similar approach was used by Yamane and Nakamura for interactive editing of postures [29].
In the following section, we progressively introduce our efficient IK architecture with multiple priority levels that handles singularities and joint limits.

3 The multiple priority levels IK

The present section first explains the most elementary IK architecture, then moves on to two priority levels and finally generalizes to p priority levels. In what follows we use the terms *constraint* and *task* interchangeably. The term *effector* designates a frame (attached to one of the bodies or to the center of mass of the articulated structure) to which we apply a constraint (either of position, orientation or both). In the next subsections, the key elements of the IK architecture will be illustrated on simple case studies in parallel with the outlining of the general equations.

3.1 The two priority levels architecture

Figure 3 shows the simplest redundant case with a two degrees of freedom chain $\{\theta_1, \theta_2\}$ and a one-dimensional constraint along the x dimension for the chain tip, i.e., the *effector*. The equation

$$x = f(\theta) \tag{1}$$

giving the tip x coordinate as a function of the current posture is nonlinear. It is easy to establish such a function in the general case but, as pointed out in Sect. 2.2, its inversion is possible only in specific, non-redundant contexts.

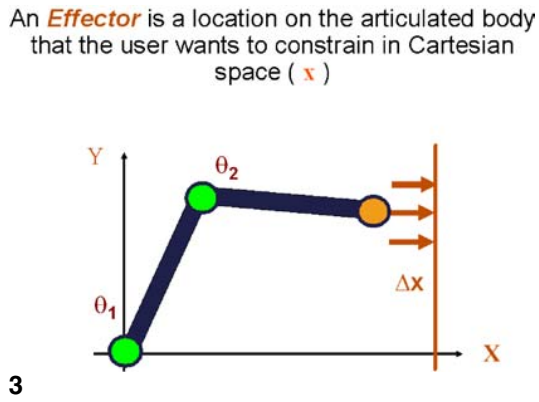


Fig. 3. Simplest redundant case
Fig. 4. The range space of J^+ and the null space of J are orthogonal

In the present example, the system is redundant, thus the IK control scheme is based on a linearization of Eq. 1. The resulting Jacobian matrix $\boldsymbol{J}$ is inverted to map the desired constraint variation Δx to a corresponding posture variation $\Delta\theta$. We use the pseudo-inverse, noted $\boldsymbol{J}^+$, as it ensures desirable properties to build powerful projection operators. Among them, the norm of the solution mapped by $\boldsymbol{J}^+$ is minimal, i.e., it is the smallest posture variation realizing the desired constraint variation (Fig. 4).

The validity of this operation is limited to small variations. Consequently, the initial gap between the tip position and the desired x value is broken down into small increments. The IK algorithm proceeds iteratively by looping through the linearization process for each update of the chain state. The construction of the Jacobian is an easy step for classical position or orientation constraints [27], so the major cost in this algorithm is the Jacobian inversion (in $O(m.n^2)$ where m is the dimension of the constraint space and n is the dimension of the joint space). However, on the whole, it compares favorably with the IK algorithm based on the Jacobian transpose [28] because the quality of the convergence is much higher.

By definition, the null space of the Jacobian, noted $N(J)$, is mapped by J onto the null vector in the constraint variation space. Rephrased from a practical point of view, this means that a variation vector belonging to $N(J)$ has no effect on the constraint.

As already mentioned, Liégeois [15] proposed a two-priority-levels architecture by constructing a projection operator onto $N(J)$ (Fig. 5). This allows the minimization of an additional scalar criterion $h(\theta)$, from which a posture variation $\Delta\alpha = -k\nabla h(\theta)$ is derived (k is a positive gain factor). The general form is:

$$\Delta\theta = J^+\Delta x + P_{N(J)}\Delta\alpha \tag{2}$$

$$P_{N(J)} = I_n - J^+J \tag{3}$$

with the notation defined as shown below.

$\Delta\theta$	n-dimensional posture variation vector
Δx	m-dimensional high priority constraints
J	$m \times n$ Jacobian matrix
J^+	$n \times m$ pseudo-inverse of J
$P_{N(J)}$	$n \times n$ projection operator on $N(J)$
I_n	$n \times n$ identity matrix
$\Delta\alpha$	n-dimensional posture variation vector

3.2 Handling singularities

A problem with Eq. 2 is its instability around a singularity (e.g., in Fig. 3 when the x target value is about to become unreachable for the chain tip). More precisely, the norm of the first term grows to infinity in the immediate vicinity of the singularity. The singular value decomposition [23] clearly highlights this problem.

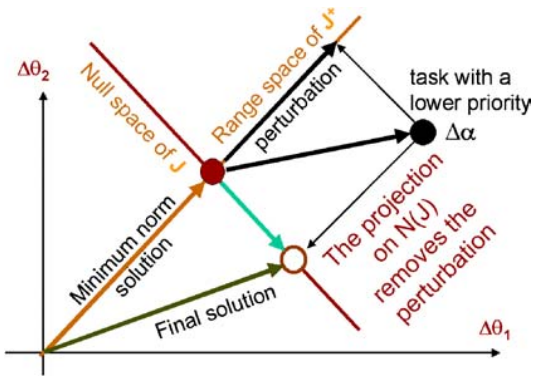


Fig. 5. The perturbation induced by the lower priority task $\Delta\alpha$ is removed, resulting in the homogeneous solution belonging to $N(J)$

The SVD of a $m \times n$ Jacobian matrix J of rank r is:

$$J = \sum_{i=1}^{r} \sigma_i u_i v_i^T \tag{4}$$

where σ_i are the singular values (strictly positive), $\{u_i\}$ and $\{v_i\}$ are the basis respectively spanning the range space of J and the complementary space of $N(J)$. The expression of the pseudo inverse J^+ shows the strong influence of any small singular values, thus explaining the instability of the solution *around* the singularity:

$$J^+ = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^T. \tag{5}$$

The solution [18, 20] is to introduce a damping factor λ transforming the ill-behaved inverse term in Eq. 5 into a damped term converging smoothly to zero when the singular value becomes small:

$$J^{+\lambda} = \sum_{i=1}^{r} \frac{\sigma_i}{\sigma_i^2 + \lambda^2} v_i u_i^T. \tag{6}$$

When λ is zero, Eq. 6 reduces to Eq. 5. By construction, the damped least-squares inverse $J^{+\lambda}$ trades the quality of the constraints satisfaction for an upper bound to the solution. A high value of λ guarantees a stable but slow convergence. The choice of a suitable value for λ is thus one of the subtle difficulties to yield a good compromise between robustness and efficiency. The low-cost under-optimal formula described by Maciejewski and Klein [17] gives good results.

A question remains regarding the computation of the projection operator $P_{N(J)}$. It is a frequent mistake to use $J^{+\lambda}$ for its construction as the damped least-squares inverse lacks various vital properties [2]. Instead, it must be built according to Eq. 3, i.e., based on the pseudo-inverse J^+. A low-cost evaluation formula is obtained by combining Eqs. 4 and 5:

$$P_{N(J)} = I_n - \sum_{i=1}^{r} v_i v_i^T. \tag{7}$$

A recurrent criticism found in the literature about IK is its high computational cost. Indeed, one SVD is required per iteration but efficiency and smoothness of the convergence is guaranteed. In addition, two points are worth noting to improve performance:

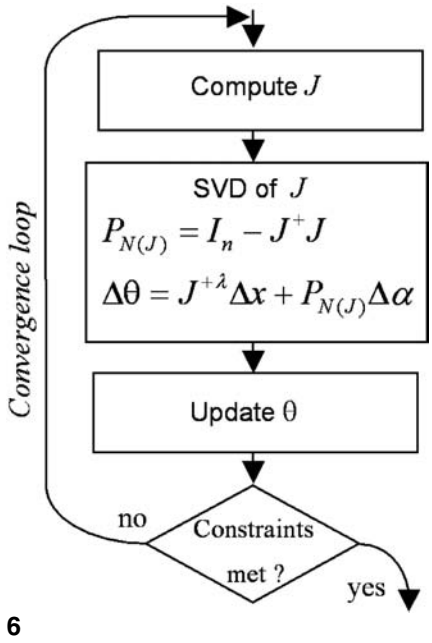
- A trimmed-down version of the SVD, the *thin* SVD [23], provides sufficient elements to compute Eqs. 6 and 7 at a lower cost.
- In the control of highly articulated figures, the dimension of the constraint space is usually much smaller than the dimension of the joint space ($m \ll n$). In this context, the thin SVD of J^T provides the necessary basis for our needs while being faster to compute than the thin SVD of J. A significant speed-up factor of $(n/m)^{3/4}$ can be obtained in this way [3].

Figure 6 provides an overview of the IK convergence loop integrating the previously mentioned concepts. Its structure is now refined to take into account joint limits.

3.3 Handling joint limits

In Fig. 6, the resulting joint variation is simply added to the current joint state. In real life, a joint value is generally bounded to a validity domain of the form [min, max]. A trivial approach to enforce the corresponding inequality constraints is to *clamp* the joint on the limit value whenever the IK update moves it beyond its validity domain.

Unfortunately, this technique converges to a non-optimal final state [6] in singular contexts. Figure 7 illustrates the phenomenon on a simple chain when joint θ_1 reaches its limit (against the cube). This time, the task is two-dimensional for clarity; the tip has



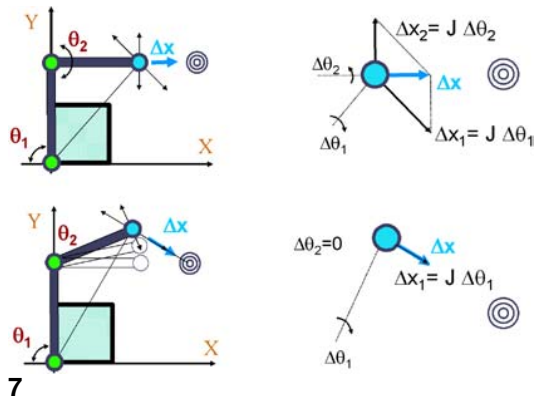


Fig. 6. Structure of the simplest IK algorithm

Fig. 7. Pathologic convergence with simple joint clamping in a singular context (*top row*: initial state, *bottom row*: final state)

to reach a point on the plane. The drawing indicates the directions of the instantaneous effector variations induced by joint variations. IK provides the smallest joint variation ensuring that a desired Δx is the vector sum of such effector variations. The top right corner shows the solution found for the desired constraint variation Δx. However, as clamping leads to ignoring the $\Delta\theta_1$ variation, the forearm moves up over the next few iterations (bottom left corner) until the $\Delta\theta_2$ contribution is reduced to zero (bottom right corner).

The correct handling of joint limits requires an internal loop that checks and removes the joints that reach their limits (Fig. 8). A small adjustment is also applied to the projector computation to make the new algorithm structure more readable: first a projector $P_{N(J_0)}$ is initialized with the identity matrix and all the joints are set to the *free* state, then we enter the clamping loop. After the IK solution evaluation, only the joints in the *free* state are examined. If a limit violation is detected for joint i, its value θ_i is clamped *on* the limit value θ_{Li}. There might be a small joint variation between the previous and the clamped joint value: we call it the clamping variation $\Delta\theta_{Ci}$. The constraint Δx has to be compensated for this clamping variation by subtracting the vector $\underline{\boldsymbol{J}}_i \cdot \Delta\theta_{Ci}$ from its current value (where $\underline{\boldsymbol{J}}_i$ is the Jacobian column vector of joint i). Then joint i is removed from the problem by zeroing the Jacobian column $\underline{\boldsymbol{J}}_i$ and the i diagonal term of $P_{N(J_0)}$. Finally its state is changed to *locked* and the clamping loop continues until no *free* joint violates its limits (Fig. 8). The algorithm ensures the constraint minimization for an arbitrary number of simultaneously clamped joints.

3.4 Handling two conflicting constraints

An alternate two-priority-levels architecture has been proposed to simultaneously solve two Cartesian constraints [11, 16]. The optimal low-priority solution is derived from the following two key factors:

- The solution of the high-priority constraint has some influence on the fulfillment of the low-priority constraint
- The low-priority constraint can be best optimized when we know how the high priority constraint restricts its potential

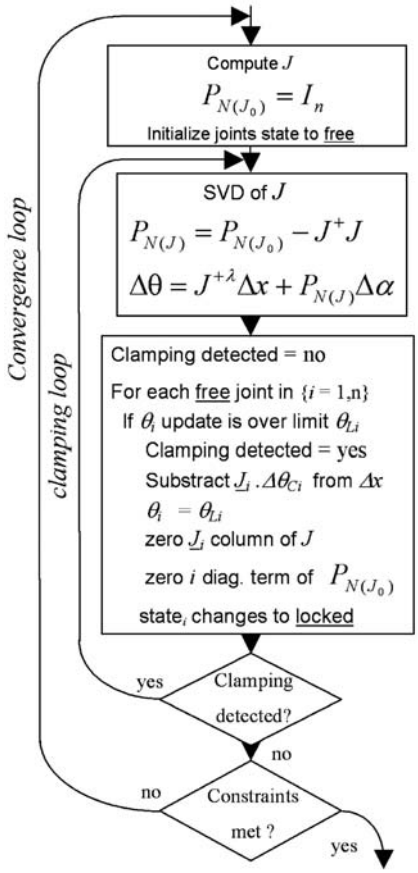


Fig. 8. IK loop including the management of inequality constraints (joint limits)

Fig. 9. The simplest redundant case with two conflicting tasks

We can illustrate this important case in a simple 2-D drawing (Fig. 9). An additional effector, at mid-segment on the forearm, is associated to a second one-dimensional constraint with lowest priority. It has to reach a desired Y coordinate (the corresponding linearized task is noted Δx_2 for generality). Both tasks cannot be satisfied simultaneously. Note too that a posture variation realizing Δx_1 may have some influence on Δx_2.

The linearization leads to the construction of two Jacobians J_1 and J_2, respectively. Two mechanisms are used to build the low-priority solution [11, 20] (Fig. 10):

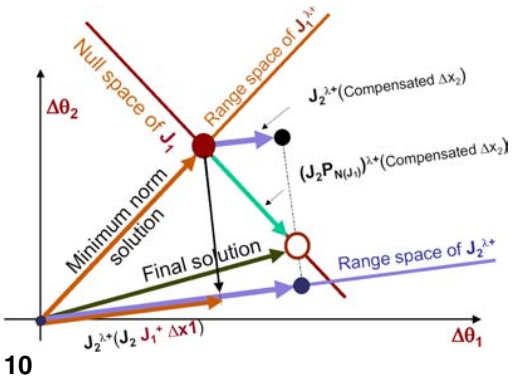
- Compensate the influence of the high priority task by removing its contribution from the secondary task: $J_2(J_1^+ \Delta x_1)$.
- Restrict the domain of J_2 to $N(J_1)$, with: $J_2 P_{N(J_1)}$. By construction, its pseudo-inverse maps the desired variation Δx_2 onto $N(J_1)$, thus it is unnecessary to again apply the null space projector $P_{N(J_1)}$.

The general expression takes into account the singularities, among which we find an algorithmic singularity that appears when the low-priority task can no longer be achieved when projected onto $N(J_1)$:

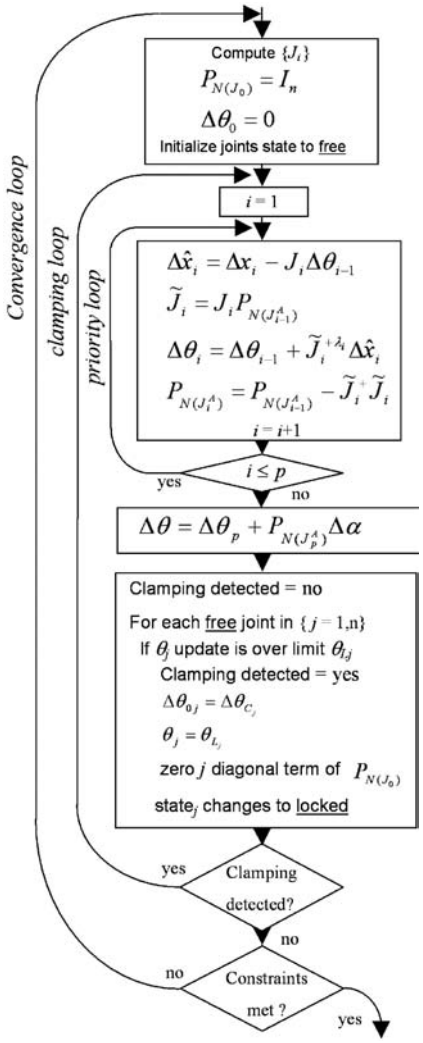
$$\Delta\theta = J_1^{+\lambda_1}\Delta x_1 + \left[J_2 P_{N(J_1)}\right]^{+\lambda_2}(\Delta x_2 - J_2(J_1^{+\lambda_1}\Delta x_1)). \quad (8)$$

3.5 Generalizing to p priority levels

The human structure being highly redundant, there exists a large solution space for the realization of more than two tasks. In the following, the index i denotes the priority, 1 being the priority of highest rank and p being the total number of priority levels. The first general framework was described by Siciliano and Slotine [24]. Our approach is based on this architecture. Nevertheless, we describe it with our more efficient evaluation of the projection operators. We also made some small adjustments to the clamping loop: the constraint compensation is now integrated in the first iteration of the priority loop. The priority management exploits a new Jacobian matrix, called the *augmented* Jacobian and noted J_i^A. It piles up all the individual task Jacobians J_i from level one to level i. The structure appears in Fig. 11. Note the initialization of a partial solution vector $\Delta\theta_0$ to zero and the priority index i to one. Then we enter the priority loop that adds the contribution of each priority level, from highest to lowest, into the partial solution vector $\Delta\theta_i$. After the loop, we add the criterion optimization term.



10



11

Fig. 10. Compensation and restriction of J_2 allow the exact realization of both constraints

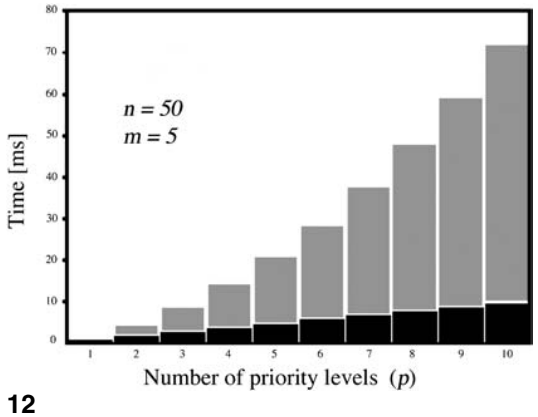
Fig. 11. General structure of the multiple priority levels IK solver

The contribution of each priority level i mimics Eq. 8:

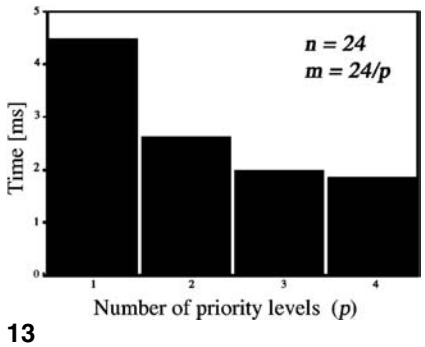
- $\Delta\hat{x}_i$ is the compensated task: the influence of the higher priority levels, from 1 to $i-1$, is removed
- $\tilde{J}_i$ is the Jacobian J_i restricted to $N(J^A_{i-1})$.

The projection operator $P_{N(J^A_i)}$ is also updated for the next priority level $i+1$ or for the criterion optimization, using our recursive formula (see the Appendix for a demonstration). When expressed as a function of the number of priority levels p, our recursive formulation has a linear computational cost compared to the quadratic cost of the formulation of Siciliano and Slotine [24]. Its use results in a speed-up factor of roughly $(p+1)/2$, as shown by Baerlocher [3].

In its multiple-priority-levels version, the clamping algorithm nicely integrates with the priority loop owing to the $\Delta\theta_0$ vector. This vector serves as the compensation of the highest priority constraints Δx_1 and is integrated within the partial solution vector $\Delta\theta_1$. Thus its effect is propagated to the lower priority levels by the priority loop.



12



13

Fig. 12. Comparison of computing costs: $O(p^2)$ in *gray* vs. $O(p)$ in *black*
Fig. 13. Effect of introducing multiple priority levels for a problem of constant dimension

3.6 Performance evaluation

Figure 12 compares measurements of the priority loop cost for Siciliano's versions and ours (in black). In this comparison, each additional priority level adds a five dimensional constraint space (for a fifty-dimensional joint space). Measurements were carried out on an SGI Octane with an R10000 processor (195 MHz).
Other tests have also compared the computing costs when a given number of constraints are gathered on one or more priority levels. Figure 13 shows a cost reduction when the constraints are distributed on several priority levels. This is due to the fact that the cost of the *transpose* Jacobian SVD (see Sect. 3.2) is a function of m^2. Thus splitting m into p priority levels leads to a cost of $O(p.n.(m/p)^2) \sim O(n.m^2/p)$. This advantage is quickly compensated for by the additional $p-1$ projector updates.

4 Interactive optimization framework

The efficiency and robustness of the multiple priority levels IK algorithm have been thoroughly tested with an interactive optimization framework where any aspect of the optimization context can change while the convergence runs in the background. The user is able to add and remove end effectors, as well as change their relative priorities, at any time (multiple tasks can also share the same priority level). Similarly the lowest level cost function is activated or deactivated whenever necessary.
The interactive test application producing the present results runs on an SGI ONYX on one MIPS R10000 processor (195 MHz). In the following 2-D and 3-D examples, we compare the convergence of both weighted IK and prioritized IK. The first two illustrations demonstrate the quantitative and qualitative superiority of the prioritized strategy.

4.1 A simple articulated chain

The posture of a simple ten degree of freedom snake-like chain is optimized to minimize a continuity cost function under four Cartesian constraints (Fig. 14). The continuity cost function is simply the sum of the squared joint angles, which is null when the chain is a straight line.
Both strategies are compared with the same set of four constraints (center of mass position, base orientation, tip position and orientation), the same initial posture and the same goals (Fig. 14b). In the weighted strategy, all constraints have the same priority. In the priority strategy, the decreasing priority order is: center of mass position, base orientation, tip position, tip orientation (note that the tip position and orientation constraints can have distinct priority levels).

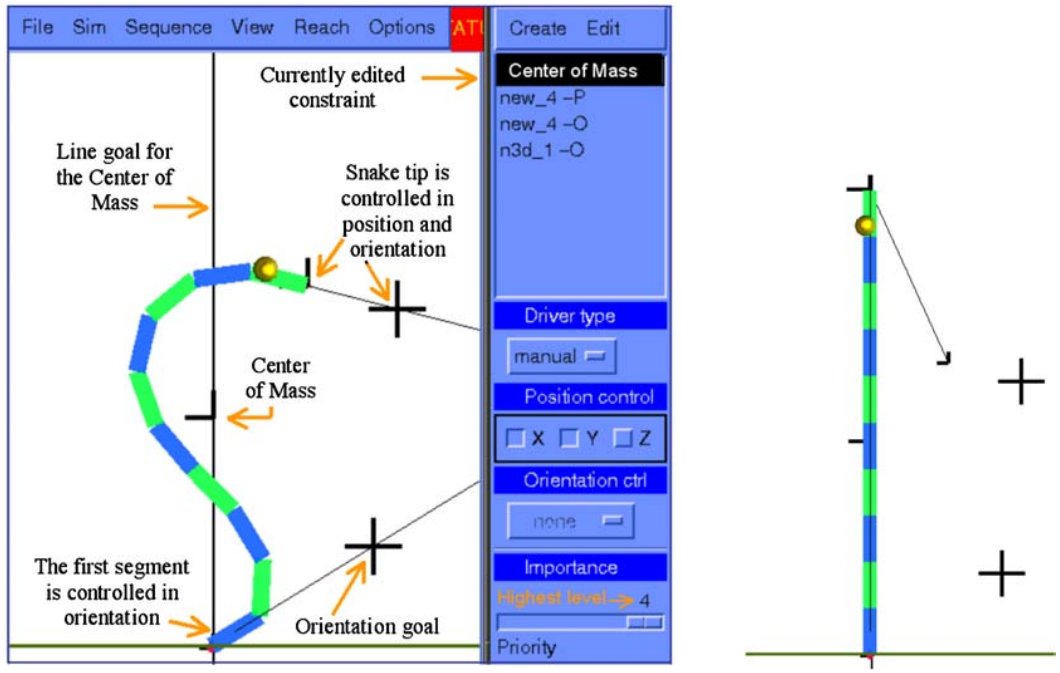


Fig. 14. a The snake-like chain. The center of mass has to reach the vertical line passing through the first segment base. The other constraints are set on the first segment and the snake tip. **b** The initial posture and the goals

Figure 15 first shows the convergence of both strategies every ten steps up to 70 iterations.

The weighted approach is shown in the left column; its convergence leads to a compromise where only half of the constraints are met after 70 iterations. It is close to a local minima due to the two conflicting orientation constraints. As a consequence, the convergence is very slow.

On the other hand, the prioritized IK preserves the fulfillment of the center of mass constraint during the entire convergence. The tip position and the base orientation goals are reached after 50 iterations. The lowest priority tip orientation constraint is met after 70 iterations.

In a second stage, the snake base orientation goal is changed leading to a singular context because some constraints cannot be met. However, the hierarchy of the constraints is still preserved. Thus the center of mass goal and the base orientation goals are still fulfilled. The last three images stress the effectiveness of the error minimization on the tip position and orientation (final posture after 30 iterations). The weighted approach is even slower in this second context; more than 700 additional steps are not sufficient to converge to a posture with a similarly small error norm.

At that stage it is interesting to note that enforcing a hierarchy of constraint priorities may help to avoid some classes of local minima. This can be explained by the fact that each priority layer introduces additional limitations in the solution space. For this reason the highest priority constraint has an important role in the determination of the convergence direction. Our examples stress the usefulness of the center of mass constraint. More tests should be performed to derive general guidelines for systematically obtaining believable human postures.

4.2 A human articulated structure

In the following examples, we use a simplified human model with forty-two degrees of freedom (no hand mobility). Each degree of freedom appears as a white square in the top left corner of Figs. 17–19.

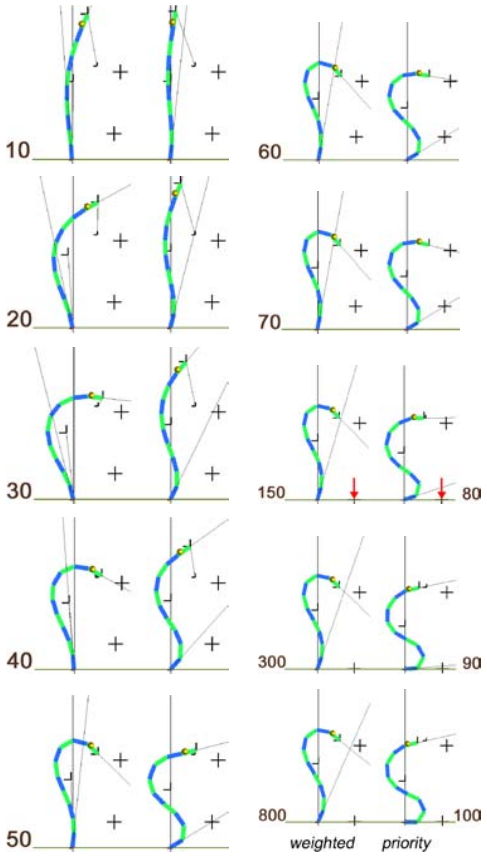
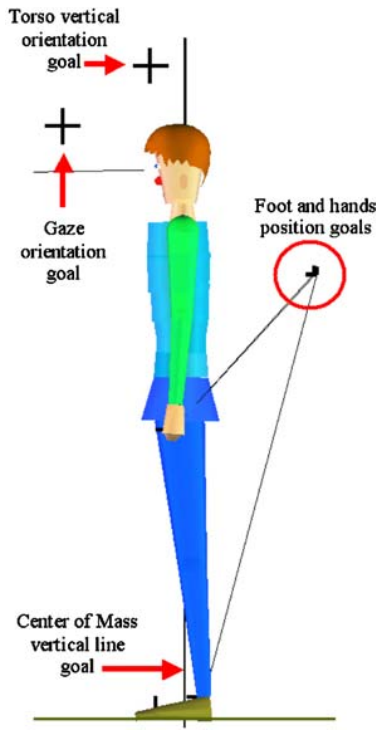


Fig. 15. Convergence comparison

It may temporarily show up in yellow when reaching its range limit.
The mass distribution is most visible in Fig. 16 over the segment volumes given a constant homogeneous density. The articulated figure is rooted at the right foot, which remains fixed in a world coordinate system (the choice of the root is interactive).
It is important to note that all joints may participate in the satisfaction of all the constraints as opposed to approaches relying on skeleton partitioning with body segments dedicated to specialized tasks (e.g., arm for reaching, neck for the gaze, torso for balance, etc.). By construction, our architecture fosters the convergence towards synergistic postures, e.g., the gaze constraint is distributed over the whole body. If needed, we can select and associate a sub-set of joints for a given task, but this is not the case in the following examples.

4.2.1 Convergence comparison

First, we compare the convergence of both approaches on the same set of constraints with the same set of goals, as illustrated in Fig. 16.
The character has to achieve a balanced dance posture requiring a high suppleness. As a consequence, some joints like the left hip, the spine or the neck may reach their range limit during the convergence (indicated with the yellow colored squares in the top left corner). For that reason, the clamping loop is the key for obtaining a correct and stable convergence. As it is exploited for both approaches, the only difference in the convergence behavior comes from the presence or absence of priority levels associated with the constraints.
Figure 17 shows the weighted IK approach where all constraints are handled at the same priority level. The intermediate image, after 50 iterations, still betrays an unbalanced posture. It is also difficult to predict which constraint will be met first. Finally, the torso vertical orientation constraint is only partially met after 100 iterations.
Figure 18 confirms the enforcement of the hierarchy of prioritized constraints during convergence together with its smaller computational cost. The center of mass constraint is satisfied right at the beginning as it has the highest priority. The torso and gaze constraints have distinct decreasing priority levels, themselves being higher than the hands and left foot constraints (all sharing the same low priority level). The middle and last images, after 35 and 75 iterations, respectively, reflect the priority hierarchy. The final posture better corresponds to a classic dance posture.
From a qualitative point of view, it is vital to be able to enforce visually important constraints both at first *and* in a synergistic fashion. A sequential enforcement of prioritized constraints would be a waste of computing resources and would lead to the most primitive robotic-like behavior. On the contrary, we want to offer an optimization scheme where constraints can be activated on the fly with time varying goals and/or priority levels. Our proposition goes in this direction as illustrated in Fig. 19 where the gaze goal is changed to a lower position. As a consequence, the whole posture is adjusted synergistically so that the gaze constraint is met without perturbing the higher priority constraints (center of mass and torso). This explains why the leg is flexed. The



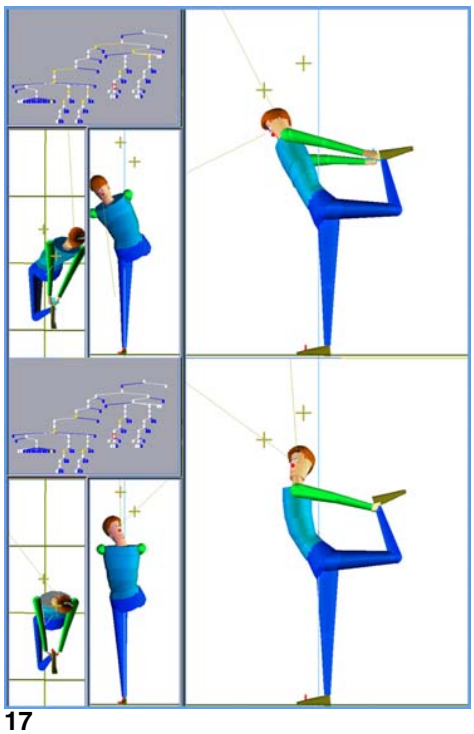


Fig. 16. Six constraint goals for editing a dance posture. The center of mass must remain on a vertical line passing through the right foot, the gaze must focus on a cross, a vertical line attached to the torso must pass through another cross, both hands and the left foot must reach goal positions
Fig. 17. Convergence of the weighted approach after 50 and 100 iterations. Note the intermediate unbalanced posture and the final compromise solution for the torso orientation

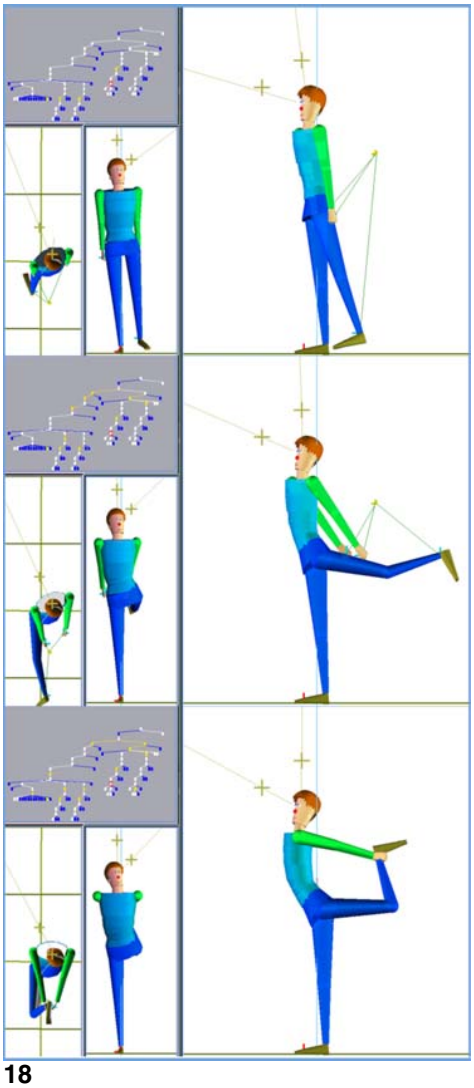
neck and spine degrees of freedom, having reached their range limit, are removed by the clamping loop. Therefore, only right leg mobility remains an option to allow the gaze to pass through the lowered cross.

4.2.2 Complex constrained postures

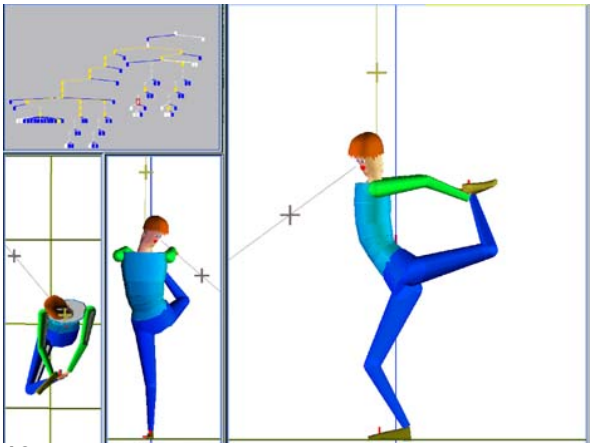
The second 3-D case study deals with reachability issues. Table 1 lists all the end effectors and the associated constraints activated permanently or temporarily during that session. In addition, the lowest priority term $\Delta\alpha$ (expressed in the joint space) was activated from time to time. For the present case study we have exploited the following cost functions to globally adjust the posture: leg-flexion, leg-extension or leg-symmetry.
The initial posture is standing up. The posture in Fig. 1 is obtained from a first convergence stage compliant with the specification of Table 1. Then, starting from that posture, we explore the right hand reach potential under the same set of constraints. The resulting trace appears in Fig. 20. This example stresses the potential of our architecture for ergonomic evaluations.
The last illustration (Fig. 21) illustrates the possibility to apply constraints to accessories such as an umbrella, work tools, swords, etc. In this specific case the umbrella handle orientation is constrained to remain vertical.

5 Discussion and conclusion

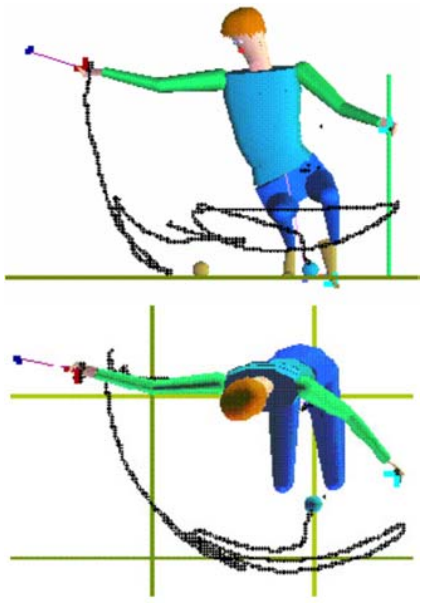
There are numerous methods to solve the inverse kinematics problem. Each has its own advantages



18



19



20

Fig. 18. Convergence of the prioritized IK successively after 5, 35 and 75 iterations
Fig. 19. Lowering the gaze constraint goal leads to flexing the right leg
Fig. 20. Interactive optimization with four priority levels and cost function optimization

and drawbacks. We have retained the criteria of generality, robustness with respect to singularities, and efficiency.

Compared to simpler analytical methods, the price to pay for this generality is a higher computational cost and complexity of the resolution methods (due to their iterative nature), and a lower reliability since the convergence towards a globally optimal solution is not always guaranteed. However, our coherent handling of joint limits and singularities ensures a stable convergence. Our contribution to complexity reduction allows our IK solver to run at interactive rates for the manipulation of complex articulated structures.

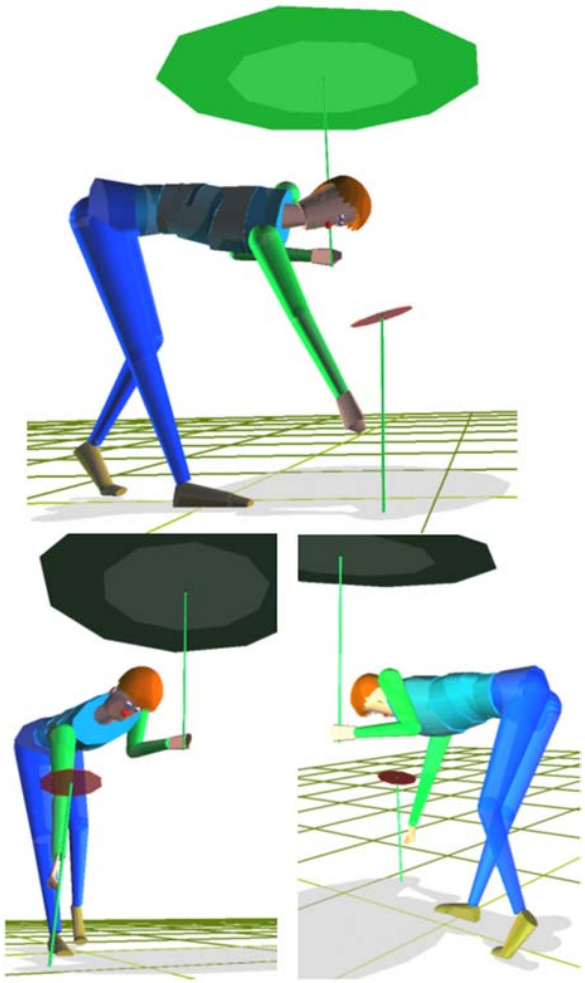


Fig. 21. A combined set of constraints involving balance, reach, and gaze while holding the umbrella vertically (with four priority levels)

Table 1. Constraints with their rank (rank is analogous to the i index in Fig. 11; rank 1 represents the highest priority)

End effector	Purpose of the constraint	Rank
Left toe	Maintain contact with the floor (position)	1
Eyes	Oriented towards the right hand (lookat-orientation)	1
Left hand center	Holding onto the vertical bar (position and orientation)	1
Center of mass	Projecting on a point or a line between the feet	2
Right hand center	Reaching the sphere (Fig. 1) or interactively following a target (Fig. 20) (position)	3
Pelvis, elbow, knee	Temporarily locally readjusting the posture (position)	4
Optimization term $\Delta\alpha$	Temporarily globally readjusting the posture	5

Regarding the specific case of human-like character positioning, realistic results are obtained if a sufficiently detailed model is provided (e.g., with joint limits). It is therefore vital to have good data and models for the joint limits. This modeling task requires additional effort but is rewarding since such data can guide the convergence towards more believable postures.

A specific difficulty has to be mentioned for human limbs: it sometimes happens that, while interacting, the legs or arms converge to a fully extended posture (a locally singular configuration). Our experience is that such a fully extended posture has a tendency to perdure unless we explicitly act on it through the lowest level optimization term $\Delta\alpha$ (hence the purpose of our leg-flexion, leg-extension cost functions). This behavior is due to the first order nature of our architecture. We have to introduce other mechanisms to overcome locally locked postures. Once this is done, we plan to explore the potential of the architecture for motion retargeting.

Acknowledgements. The authors would like to thank Lorna Herda for proofreading. This work was made possible through support from the Swiss National Foundation of Research with grants N°2000-053809 and N°20-61999.00.

Appendix

Proposition: Given an $m \times n$ matrix J_i^A partitioned as $\left[\frac{J_{i-1}^A}{J_i}\right]$, the following identity holds:

$$I_n - J_i^{A+} J_i^A = \left(I_n - J_{i-1}^{A+} J_{i-1}^A\right) - \tilde{J}_i^+ \tilde{J}_i$$

where $\tilde{J}_i = J_i(I_n - J_{i-1}^{A+} J_{i-1}^A)$.

Demonstration: A result due to Cline [8] yields the pseudo-inverse of a partitioned matrix. It is applied to J_i^A:

$$J_i^{A+} = \left(J_{i-1}^{A+} - T_i J_i J_{i-1}^{A+}, T_i\right)$$

where $T_i = \tilde{J}_i^+ + X(I - \tilde{J}_i \tilde{J}_i^+)$ and X is a complex term not useful here.
The following identity can then be established:

$$\begin{aligned} I_n - J_i^{A+} J_i^A &= I_n - \left(J_{i-1}^{A+} - T_i J_i J_{i-1}^{A+}, T_i\right) \left[\frac{J_{i-1}^A}{J_i}\right] \\ &= I_n - J_{i-1}^{A+} J_{i-1}^A - T_i J_i (I_n - J_{i-1}^{A+} J_{i-1}^A) \\ &= I_n - J_{i-1}^{A+} J_{i-1}^A - T_i \tilde{J}_i \\ &= I_n - J_{i-1}^{A+} J_{i-1}^A \\ &\quad - \left(\tilde{J}_i^+ + X\left(I - \tilde{J}_i \tilde{J}_i^+\right)\right) \tilde{J}_i \\ &= \left(I_n - J_{i-1}^{A+} J_{i-1}^A\right) - \tilde{J}_i^+ \tilde{J}_i. \end{aligned}$$

In the last step, the pseudo-inverse property $\tilde{J}_i \tilde{J}_i^+ \tilde{J}_i = \tilde{J}_i$ is used.

References

1. Badler N, Manoochehri KH, Walters G (1987) Articulated figure positioning by multiple constraints. IEEE Comput Graph Appl 7(6):28–38
2. Baerlocher P, Boulic R (1998) Task-priority formulations for the kinematic control of highly redundant articulated structures. Proc of IEEE IROS 98, Victoria, BC, pp 323–329
3. Baerlocher P (2001) Inverse kinematics techniques for the interactive posture control of articulated figures. Dissertation N°2383, Swiss Federal Institute of Technology, Lausanne, EPFL, Switzerland
4. Bindiganavale R, Badler N (1998) Motion abstraction and mapping with spatial constraints. Proc of Captech98. Lecture notes in computer science (Lecture notes in artificial intelligence), vol 1537. Springer, Berlin Heidelberg New York, pp 70–82
5. Boulic R, Mas R, Thalmann D (1996) A robust approach for the center of mass position control with inverse kinetics. J Comput Graph 20(5):693–701
6. Boulic R, Mas R, Thalmann D (1997) Interactive identification of the center of mass reachable space for an articulated manipulator. Proc of International Conference of Advanced Robotics ICAR'97, Monterey, CA, pp 589–594
7. Choi KJ, Ko H-S (2000) Online motion retargeting. J Vis Comput Anim 11(5):223–235
8. Cline RE (1964) Representation for the generalized inverse of a partitioned matrix. J Soc Ind Appl Math XII:588–600
9. Girard M, Maciejewski AA (1985) Computational modeling for the computer animation of legged figures. Proc of SIGGRAPH'85, Computer Graphics, vol 19, pp 263–270
10. Gleicher M (2001) Comparing constraint-based motion editing methods. Graph Models 63:107–134, Academic Press
11. Hanafusa H, Yoshikawa T, Nakamura Y (1981) Analysis and control of articulated robot with redundancy. IFAC, 8th Triennal World Congress, vol 4, pp 1927–1932
12. Hodgins JK, Pollard NS (1997) Adapting simulated behaviors for new characters. Proc of SIGGRAPH'97, Los Angeles, pp 153–162
13. Korein JU (1985) A geometric investigation of reach. MIT Press, Cambridge
14. Lee J, Shin SY (1999) A hierarchical approach to interactive motion editing for human-like figures. Proc of SIGGRAPH'99, Los Angeles
15. Liégeois A (1977) Automatic supervisory control of the configuration and behavior of multibody mechanisms. IEEE Trans Syst Man Cybern 7(12):868–871
16. Maciejewski AA, Klein CA (1985) Obstacle avoidance for kinematically redundant manipulators in dynamically varying environments. Int J Robot Res 4(3):109–117
17. Maciejewski AA, Klein CA (1988) Numerical filtering for the operation of robotic manipulators through kinematically singular configurations. J Robot Syst 5(6):527–552
18. Maciejewski AA (1990) Dealing with the ill-conditioned equations of motion for articulated figures. IEEE CGA 10(3):63–71
19. Monzani J-S, Baerlocher P, Boulic R, Thalmann D (2000) Using an intermediate skeleton and inverse kinematics for

motion retargeting. Proc Eurographics 2000, Interlaken, Switzerland
20. Nakamura Y, Hanafusa H (1986) Inverse kinematic solutions with singularity robustness for robot manipulator control. J Dyn Syst Meas Control 108:163–171
21. Phillips CB, Badler N (1991) Interactive behaviors for bipedal articulated figures. Comput Graph 25(4):359–362
22. Popovic Z, Witkin A (1999) Physically based motion transformation. Proc of SIGGRAPH'99, Los Angeles
23. Press W, Teukolsky S, Vetterling WT, Flannery BP (1992) Numerical recipes in C, 2nd edn. Cambridge University Press, Cambridge, pp 59–70
24. Siciliano B, Slotine J-J (1991) A general framework for managing multiple tasks in highly redundant robotic systems. Proc of ICAR'91, vol 2, pp 1211–1215
25. Tak S, Song O-Y, Ko H-S (2000) Motion balance filtering. Proc of Eurographics'2000, Computer Graphics Forum, vol 19(3). Blackwell, Oxford
26. Tolani D, Goswami A, Badler N (2000) Real-time inverse kinematics techniques for anthropomorphic arms. Graph Model 62:353–388
27. Watt A, Watt M (1992) Advanced animation and rendering techniques. Addison-Wesley
28. Welman C (1993) Inverse kinematics and geometric constraints for articulated figure manipulation. Dissertation, Simon Fraser University
29. Yamane K, Nakamura Y (2003) Natural Motion Animation through Constraining and Deconstraining at Will. IEEE TVCG 9(3):352–360, July–September 2003, ISSN: 1077-2626
30. Zhao J, Badler N (1994) Inverse kinematics positioning using nonlinear programming for highly articulated figures. ACM Trans Graph 13(4):313–336

PAOLO BAERLOCHER received his PhD degree in Computer Science in 2001 at the Swiss Federal Institute of Technology, Lausanne (EPFL). Currently he is game engine developer at Neko Entertainment, a french video-games company. His main research interests include animation of articulated bodies, inverse kinematics, dynamics, and real-time rendering.

RONAN BOULIC is a Senior Researcher, Lecturer and PhD Director at the Swiss Federal Institute of Technology, Lausanne (EPFL). He is working in the Virtual Reality Lab and his research interests include realistic motion synthesis for virtual humans and robot. He received the PhD degree in Computer Science in 1986 from the University of Rennes, France, at the INRIA-IRISA research institute. He received the Habilitation degree from the University of Grenoble, France, in march 1995. Ronan Boulic is co-author of 75 research papers among which 20 appeared in international peer-reviewed journals. He is member of IEEE, ACM and Eurographics.

EUROGRAPHICS 2005 / M. Alexa and J. Marks
(Guest Editors)

Volume 24 (*2005*), *Number 3*

Morphology-independent representation of motions for interactive human-like animation

R. Kulpa[1,2], F. Multon[2,1] and B. Arnaldi[1]

[1] SIAMES Project - IRISA, Rennes, France
[2] LPBEM - University of Rennes 2, France

Abstract

This paper addresses the problem of human motion encoding for real-time animation in interactive environments. Classically, a motion is stored as a sequence of body postures encoded as a set of joint rotations (quaternions, Euler-like angles or rotation matrices). As a consequence, Cartesian constraints must be solved using inverse kinematics and/or optimization. Those processes involve computation costs that do not allow real-time animation of several characters in interactive environments. To solve such a problem with a minimum computation time, we designed a motion representation independent from the morphology and containing the constraints intrinsically linked to the motion such as feet contacts with the ground. With such a description, a unique motion can be shared by several characters with different morphologies and in different environments. We also adapted a Cyclic Coordinate Descent algorithm that takes advantages of this representation in order to rapidly deal with complex tunable spacetime constraints. For example, this method enables to interactively control at least eight characters with different morphologies that interact each other during a fight training. Hence, each character has to deal with geometric constraints that can change at every time, depending on the opponents' morphology and gestures.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism - Animation

Figure 1: *Eight characters dealing with several constraints in real-time.*

1. Introduction

In interactive environments, designing a set of behaviors for an avatar is difficult primarily due to the real-time constraint, especially if we wish to use a relatively poor motion capture database. In such interactive environments, selecting behaviors in a database requires to pre-calculate all the possible transitions between them by using motion graphs [KGP02] [LL04]. An alternative consists in merging several elementary motions thanks to frame-space interpolation [GR96]. Nevertheless those techniques generally require large motion capture databases. An alternative is to use motion adaptation that enables to create new behaviors with a small motion capture database. The main point is then to ensure that adaptations preserve the style and realism of the original motion. The motion must be adapted to the skeleton of the new character. The environment should also be considered to adjust the motion to new constraints.

To solve this problem, two main approaches could be identified: kinematic or dynamic based techniques. While the former allows low computation time, the latter enables to take into account balance, force magnitude control, etc. Dynamic methods are not yet suitable for interactive environments where a user can interfere with several synthetic char-

acters concurrently. Moreover, techniques involving spacetime constraints and optimization cannot be considered because they require a complete knowledge of the constraints whereas a user can interact at any time. Previous works are generally based on the modification of angular trajectories whereas constraints are generally given in the Cartesian frame. The resulting process consequently involves using inverse kinematics.

We propose a specific description of motion in order to accelerate and optimize the process of motion adaptation. To this end, instead of using angular trajectories we propose to deal with data that are not linked to the character's anthropometric properties. This method enables us to animate in real-time several different characters who have to deal with constraints that can change at every frame (see figure 1).

2. Related works

Adapting motions to new constraints has been investigated by adding displacement maps to the original trajectories [Gle98] [LCR*02] [LS99]. Although those techniques are very efficient to edit motions, a complete knowledge on spacetime constraints is required for all the sequence. As a consequence, modifying those constraints in real-time is not possible. Let us consider a kick that must drive the foot to the opponent's head while the other character is also moving. If no assumption of the opponent's head trajectory is possible, displacement maps calculated at the beginning of the sequence may not verify the constraints at the end of the sequence given that the opponent's head is moving unpredictably.

An alternative is to use statistical analysis [BH00] [LWS02] [SHP04] in order to decrease the number of degrees of freedom that are controlled and consequently to simplify the control problem. Learning a search-space is also possible to perform so-called style-based inverse kinematics [GMHP04]. Whatever the technique, a large database of motions is required to obtain accurate controllers whereas we aim at using a small motion capture database.

As constraints are generally expressed in the Cartesian frame whereas joints are controlled using rotations, another solution consists in using inverse kinematics at each frame. This technique allows interactive motion deformation with prioritized constraints [CB04] including the control of the center of mass trajectory [BB04]. In order to decrease computation time, the method can be applied on only subsets of the entire kinematic chain [BCHB03]. A dynamic filter can also be used to ensure that the resulting motion verifies dynamic constraints [TySK02]. This approach is an alternative to the search of a controller that drives a dynamic system by computing the forces and torques that minimize a given objective function [PW99] [FP03] [SP05].

Although those methods provide a powerful tool to adapt motions to new constraints rapidly, the computation time does not allow the control of several characters in real-time. Improvements have been proposed to apply inverse kinematics with less computation time, such as using specific formulations for anthropometric limbs [TGB00]. However, controlling sub-parts of the skeleton independently is not always adequate to solve constraints that require the displacement of all the body. [SLSG01] extent this approach by using an iterative process in order to deal with the complete body. Their goal was to map the movements of a performer to an animated character only considering constraints on the end-effectors. Another improvement consists in using an intermediate skeleton [MBBT00] that has less degrees of freedom and then in recalculating the remaining degrees of freedom analytically. This improvement was also used to control a simplified mechanical system [PW99] while inverse kinematics was used to recover the remaining degrees of freedom. Those works demonstrate the interest of using a specific representation of motion (through simplified skeletons) and we propose to go further those steps. Hence, we propose to define a data structure especially dedicated to motion adaptation: it is supposed to decrease the computation time required to solve complex Cartesian constraints that could change at each frame.

To accelerate the constraints solver based on this data structure, we propose to use the Cyclic Coordinate Descent technique (denoted CCD in the remainder of the paper) [WC91] [Lan98]. Although this method offers rapid solving of inverse kinematics problems, it is not possible to add secondary tasks and consequently to use specific laws. As a consequence, some artifacts may occur in the resulting sequence, such as an inhomogeneous repartition of the deformation along the kinematic chain [Wel93]. Several heuristics were proposed to make the system converge to an acceptable solution. For example, a threshold can be used at each time step in order to limit the quantity of rotation taken by the first joints. This technique, called damping, makes the solution be more homogeneous but it engenders more iterations to reach it. Joint limits can also be considered [Lan98]. However, as CCD does not consider movement continuity, it must be improved to provide acceptable solutions for motion adaptation. Moreover, it does not take comfort information such as using preferably the arm and not the trunk during grasping.

In this paper, we propose to use the advantages of several of the above techniques by introducing a specific data structure to encode the motion. We also propose an enhancement of the CCD that enables to have a more realistic adaptation and to respect the center of mass (called COM in the remainder of the paper) as well as to control every points of the character's body.

3. Overview

As stated above, classical approaches dealing with joint angles require optimization and inverse kinematics to solve

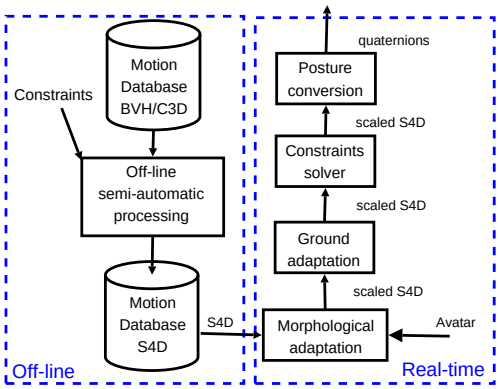


Figure 2: *Overview of the motion adaptation process.*

Cartesian spacetime constraints, such as moving on an uneven terrain, grasping objects at various positions in space... Even for flat grounds, scaling data to new skeletons requires inverse kinematics or optimization to ensure feet contacts with the environment. To take those requirements into account without using complex inverse kinematics methods, we propose to subdivide the skeleton into subparts. Human body is considered as a hierarchy of subparts (such as an arm, the trunk...) expressed relatively to their parent, coupling Cartesian and angular data. For example, instead of storing all the joint angles, we select only the angles that are not directly linked to anthropometric data, such as the angle of the plane containing the shoulder, the elbow and the wrist. The Cartesian data are normalized so that only adimensional data are stored in order to simplify the scaling process.

However, motion is not limited to a sequence of postures but also intrinsically contains constraints, such as contacts between parts of the skeleton and the environment. Foot contact can be semi-automatically identified by selecting the times where the foot height and velocity are below a given threshold [MKMA04]. Other constraints depend on the nature of the motion and must be edited by a user. Those tasks are performed off-line for each captured motion in a so-called *Off-line semi-automatic process* (see figure 2). The inputs are captured files in C3D or BVH format and the outputs are files stored in our format called S4D.

Given a file stored in this format and an avatar (described with its geometry and anthropometric data), a three-steps process is performed in our real-time animation engine (called MKM) in order to adapt the motion to a new character and to the environment. First, the adimensional data stored in the file are scaled to take the new character's anthropometric data into account (called *Morphological adaptation* module). The output is called scaled S4D given that it is still based on our motion representation. Second, the feet that must be in contact with the ground are displaced to a convenient position (called *Ground adaptation* module). As a consequence, the root is recalculated to be compatible with the new feet positions. The data are still in the scaled S4D format. Third, the other constraints are solved in the *Constraints solver* module. Those constraints are the ones edited by the user in the *Off-line semi-automatic process* and others that can be added interactively in the real-time environment. At this step, the scaled S4D format is still used but other points directly linked to the geometric constraints are calculated when needed. All the constraints require geometrical data provided by the real-time environment (such as actual position of a target that have to be reached or actual size of an object that must be held). Finally, joint angles are calculated and sent to visualization through the *Posture conversion* module. Only at this step, our specific format is abandoned to provide visualization with a classical posture representation.

Section 4 describes the data structure used to encode motion. Then section 5 explains the constraints associated to this representation. Section 6 shows that this adimensional representation allows simple and rapid morphological adaptation. Section 7 describes how the motion is adapted to the ground. The way all the joints are calculated according to our data structure is presented in section 8. Next, section 9 deals with the adapted CCD algorithm that makes it possible to control several characters in real-time while respecting the COM. Some results are presented in section 10.

4. Normalized skeleton

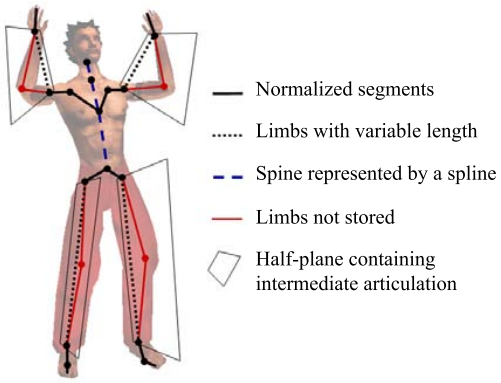


Figure 3: *Normalized representation of the skeleton.*

The motion is stored using a normalized representation of the skeleton and a set of associated constraints. In this section, we describe the representation of a posture. Classically, the human body is considered as a set of kinematic chains starting from the root to extremities. In our representation, the human body is also subdivided into kinematic subchains that

describe parts of the skeleton (see figure 3). Those kinematic chains are divided into three main parts:

- *the normalized segments* are composed of only one body segment (such as the hands, the feet, the clavicle and the scapula),
- *the limbs with variable length* that encode upper and lower limbs; in this representation, the intermediate joints (elbows and knees) are not encoded because their position is directly linked to the character's anthropometric properties,
- and *the spine* represented with a spline that could be subdivided into as segments as wishes in the real-time animation module.

Whatever the kind of kinematic chain KC composed of segments S_j, only normalized values are stored. Let E and R be respectively the extremity and the root of the kinematic chain KC. In order to obtain a normalized representation of KC we can use the following equation:

$$normalizedKC = \frac{E - R}{\sum_j length(S_j)} \quad (1)$$

The *limbs* are stored using only two data: a reference frame attached to the root of the corresponding kinematic chain (shoulder or hip) and a scalar. The axes of the frame are depicted in figure 4: they are defined in order to represent the half-plane in which the intermediate articulation (elbow or knee) takes place. The scalar represents the normalized length of this limb, that is to say a percentage of the maximum limb length.

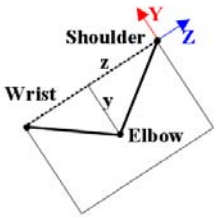


Figure 4: *Specification of the half-plane represented by the frame (X,Y,Z) which contains the elbow. The local coordinates (y, z) of the elbow are computed analytically.*

In this last representation, each point of the skeleton can be retrieved relatively to the root position. Contrary to classical representation, the instantaneous orientation of the root is divided into two main components. The global orientation deals with the global direction of the motion. The local orientation is the additional rotation applied to the global orientation in order to obtain the actual root orientation. During a walk, it represents the pelvis oscillations around the global direction. In order to encode movements without taking anthropometric properties into account, the position of the root is normalized by the leg length.

5. Constraints

Figure 5: *Specification of the point that is constrained.*

As introduced in section 3, constraints are also stored with the motion. The feet contacts with the ground are necessary to avoid sliding effects or unwanted penetration of the feet into the ground. Additional constraints can also be added by the user. For example, considering a clapping motion, at least one constraint must be added to the motion: the contact between the two hands that occurs at repetitive periods of time. To model a constraint C_i several parameters are necessary:

$$C_i = \{CP_i, T_i, KC_i, P_i, A_i\}$$

The first parameter CP_i is the constrained point (see figure 5). It is linked to a body segment and its position is defined using a 3D local offset from the root of this segment. Next parameter T_i is the type of the constraint among the following ones:

- Position: the user can impose a desired position for a given point of the skeleton. Since all the computations are performed in real-time, the imposed position can change every time. For example, it is then easy to use the position of an avatar's articulation as the constraint for another character (see section 10);
- Orientation: the user can enforce a segment's orientation. This orientation can be directly applied to the corresponding reference frame;
- Distance: the user can impose a distance between two points of the skeleton. This constraint is also used to ensure contact between two points by imposing a null distance. For example, this constraint makes it possible to hold several different objects with different sizes according to a unique two-hands holding motion provided by the motion database;
- Restricted area: the user can constraint points into a spatial area (see *wristsAboveTable.avi*). Figure 6 shows the use of such constraints applied to the wrists. The avatar in the background talks while placing the wrists on his thighs, as in the original captured motion. In the front, the character plays the same motion while adapting it in order to keep the wrists above the table.

Depending on the constraint, specific parameters must obviously be added such as the desired position of the constraint or the dimensions of the restricted area.

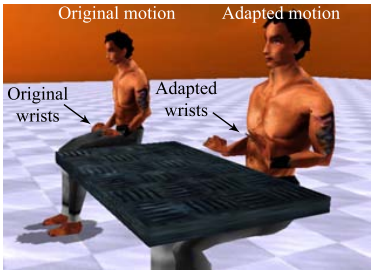


Figure 6: *The wrists are constrained in the area above the table.*

The next parameter KC_i defines the kinematic chain associated to the constraint. It offers the user the possibility to specify the set of usable body segments in order to solve the constraint. For example, in figure 7, three constraints are specified. C_1 is applied on the right hand and acts on all the segments ranging from the hand to the abdomen. The constraint C_2 only involves the arm and the clavicle.

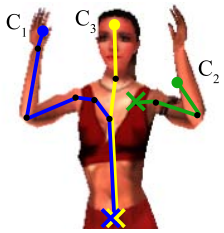


Figure 7: *Constraints are associated to a subset of the complete kinematic chain. The three different constraints are represented with big dots. The lines are their kinematic chains and their root are represented with the big crosses.*

The priority of the constraint, called P_i, indicates the importance of a constraint compared to others. The constraints with low priorities are only verified after those associated to higher priorities. Finally, the user can start and stop constraints using the parameter A_i which is the activation of the constraint (see figure 8). A_i includes the start and stop times of a constraint as well as the duration of activation and deactivation phases.

6. Morphological adaptation

With the representation of posture presented in this paper, morphological adaptation can be easily performed. Indeed, it consists in inverting the segments normalization process (described in equation 1) using the dimensions of the new body segments:

$$newKC = \sum_j newLength(S_j) * \frac{E-R}{\sum_j originalLength(S_j)} \quad (2)$$

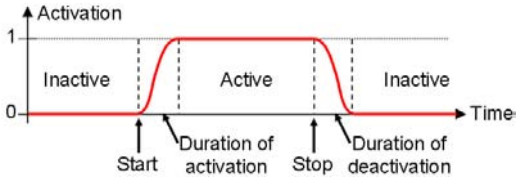


Figure 8: *Activation of a constraint.*

This equation demonstrates that no loss of information occurs by using our representation if *newLength* is equal to *originalLength*. This morphological adaptation can be performed rapidly in the real-time animation module. Consequently, a motion stored with this representation can be shared by all the avatars and adapted to new anthropometric properties. In the video (see *morphoAdaptation.avi*), several characters share the same dancing motion. In this video, the displacement of the young girl in black is smaller than the others. The step length is automatically scaled to her small legs. As morphological adaptation is performed in real-time the system is able to deal with body segments that could increase and decrease at each time step.

At this step of the whole process, intermediate articulations such as the elbows, the knees and the vertebrae are not yet reconstructed. Moreover, since the position of the root is normalized using the leg length, it is also adapted in order to be at the correct height from the floor. But if the ground is different or changes in real-time, ground adaptation is needed.

7. Ground adaptation

The ground adaptation consists in adapting the scaled relative posture according to the footprints placed on the ground. These footprints are calculated according to the foot-contact constraints specified in the motion file. If a function is able to provide the height of the ground (axis Z) for a given 2D position (X,Y), the new footprints can then be placed automatically. This function allows to handle any kind of ground even those that are interactively modified.

Ground adaptation ensures that the new position of the pelvis (and consequently the root) is compatible with the leg lengths while being as near as possible from the position stored in the original motion file. First, the desired positions of the ankles are retrieved according to the shape of the ground. Actually, it corresponds to a translation of the ankles along the Z-axis from their original positions (denoted *desired ankle* in figure 9, part (b)).

Second, the width of the pelvis is suppressed from the ankles positions in order to express them relatively to the root (see (c) in figure 9). The next step consists in calculating the new height of the root (see (d) in figure 9). The root's position must verify two conditions. It should

be close to the original height $h1$ in order to preserve the initial style. It should also ensure that the root is reachable according to the two desired ankles positions (providing two maximum heights $h2$ and $h3$ for the two legs). The height of the root is then the minimum of these three heights: $hRoot = min(h1, h2, h3)$.

Next, the hips are retrieved by adding the pelvis width to the root position resulting from the previous steps (see (e) in figure 9). Our representation then enables to easily adapt the legs in order to respect the resulting hips and ankles. Indeed, the blue dot-lines in the figure represents the limbs with variable length. Adapting the leg is then simply achieved by aligning the Z axis of the limb frame with the new vector $\overrightarrow{ankle - hip}$. The normalized length of the leg is set to $\|\overrightarrow{ankle - hip}\|$.

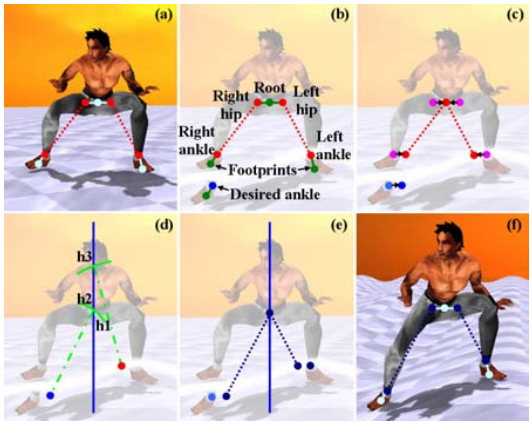


Figure 9: *Our normalized representation is suitable for ground adaptation of posture. (a) Original posture. (b) The blue dot is the desired position for the right ankle. (c) Pelvis width is suppressed to all the positions. (d) Three heights are considered for the root: its actual height and the maximum height the hips can reach keeping the feet on the ground. (e) The lowest height is chosen. (f) Final posture: with the normalized skeleton, legs are reconstructed automatically.*

8. Posture conversion

None of the previous adaptations use intermediate articulations (the elbows, the knees and the vertebrae). However, when the posture is rendered, those articulations are needed. Consequently, a conversion is needed to transform our representation to a more classical one. The calculation of some intermediate articulations can also be useful for complex constraints solving where knees or elbows are involved.

The conversion is performed according to the type of the kinematic chain associated to the constraint: normalized segments, limbs or the spine. To retrieve the position of a vertebrae, the spline representing the spine is simply discretized according to the distances between the vertebrae. For the limbs of variable length, the position of the intermediate articulations can be obtained analytically. To explain this computation, let us consider how the position of the elbow is retrieved, given that the knee is retrieved the same way. The position of the elbow is placed on the intersection of two spheres (resulting in a circle). The first sphere is centered at the shoulder and its radius is equal to the arm's length. The second sphere is centered at the wrist and its radius is equal to the forearm's length. The position of the elbow is then placed at the intersection of the circle and the half-plane which orientation is stored in the motion file. Moreover, since this half-plane is defined in the shoulder reference frame, the local position of the elbow in this frame (see figure 4) is easily computed:

$$z = \frac{armLgth^2 - forearmLgth^2 + limbLgth^2}{2 * limbLgth}$$

$$y = \sqrt{armLgth^2 - z^2}$$

where z and y are the local coordinates of the elbow in the shoulder frame, $armLgth$ (resp. $forearmLgth$) is the length of the arm (resp. of the forearm) and $limbLgth$ is the current distance between the wrist and the shoulder.

9. Specific constraints solver

To deal with all the constraints, we adapted a Cyclic Coordinate Descent algorithm that takes advantages of our representation. By nature, CCD method converges to a unique solution given that no secondary task can be specified contrary to classical approaches [CB04]. With CCD those secondary tasks are replaced by heuristics in the iterative search algorithm. Nevertheless, the choice of those heuristics is a difficult task. Selecting bad heuristics may engender an inhomogeneous repartition of the deformation along the kinematic chain. A commonly used heuristic, called damping, is to threshold the computed rotations applied to the body segments at each iteration. This method enables to improve the homogeneity of the solution but requires more iterations.

Instead of using a recursive search of the solution such as Badler et al. [BMB87] have done, we propose to subdivide the skeleton into six main groups (see figure 10): the two legs, the head, the two arms and the trunk. Each group G_j (where j is the number of the group) has a set of constraints $CS_j = \{C_k\}$ which is a subset of the complete set of constraints CS imposed by the user: $CS_j \subset CS$. All the constraints contained in CS_j are linked to at least one body segment belonging to the group G_j. For example, if a constraint C_i is acting on the right forearm, it is obviously linked to the right arm group G_a and consequently added to CS_a. If the kinematic chain KC_i associated to C_i also involves the trunk (belonging to the trunk group G_t), the constraint is then added to the set of constraints CS_t associated to G_t.

As the trunk group uses constraints shared with other

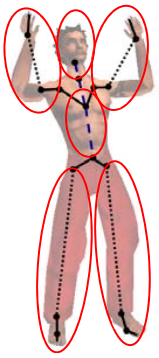


Figure 10: *Six groups used to order the CCD algorithm: the head, the two arms, the two legs and the trunk.*

groups (such as the arms or the head), our constraints solver makes the motion adaptation starts from the legs to the trunk in the following order: the legs, the head, the arms and finally the trunk. The main iteration loop is then:

```
try = 0;
Do
  adaptGroup(RIGHT_LEG)
  adaptGroup(LEFT_LEG)
  adaptGroup(HEAD)
  adaptGroup(RIGHT_ARM)
  adaptGroup(LEFT_ARM)
  adaptGroup(TRUNK)
  adaptRoot()
While ( (try++ < maxTries)
        && (adaptation not completed) )
```

The iteration loop ends when no more adaptation is necessary or when the number of iterations is too high.

For each group, adaptation is performed from the extremity to the root of the kinematic chain. In this adaptation process, the first adapted body segments get larger transformations than the last ones. For the arms example, the algorithm starts from the wrist to the clavicle. For each articulation, either an analytical computation or a rotation is performed. In the following algorithm, CP_i is the current position of the constraint C_i and *Target* is its desired position:

```
For all the articulations A in the group
  If A in limb with variable length
    Analytical computation
  Else
    Apply on A the angle to align
      A->CPi with A->Target
  End of if
End of for
```

For the limbs with variable length, analytical computations are performed, as described in section 7 for the legs. In the analytical computation, the constraint with the highest priority is verified first. The other constraints are then solved while ensuring that the previous constraints with higher priorities are not broken. Part (a) of figure 11 shows the adaptation of a constraint C_1 with a high priority. Since the target is reachable, the constrained point is directly placed to the convenient configuration using an analytical computation. Part (b) shows that the constraint with lower priority is computed using a rotation around the axis that goes through the point constrained with C_1 and the root. It provides a circle of solution in which the point that is the closer to the desired target for constraint C_2 is selected.

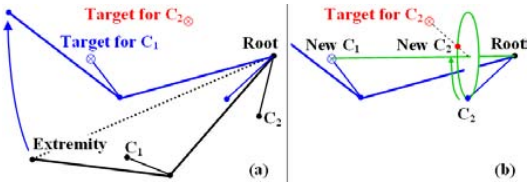


Figure 11: *Analytical solution for two constraints. In part (a), the constraint C_1 with a higher priority is verified. Part (b) is the adaptation of the constraint C_2 of lower priority while preserving C_1 and Root.*

With such a method, only the body segments that are necessary to solve the constraints are used. On the opposite, in the classical CCD methods all the kinematic chain is used. For example, during a grasp, the CCD iterates from the hand to the abdomen in order to verify the constraint. Generally, the constraint is not verified with only one step of the algorithm, the trunk is consequently modified. The resulting motion could be unrealistic because no trunk adaptation is actually required in order to reach targets close to the body. With our method, the arm is moved to the desired constraint in only one step without involving the trunk if only the arm is required. Otherwise, as the arm is displaced to minimize the distance between the current constraint position and the target, the trunk moves with only the minimum required rotation.

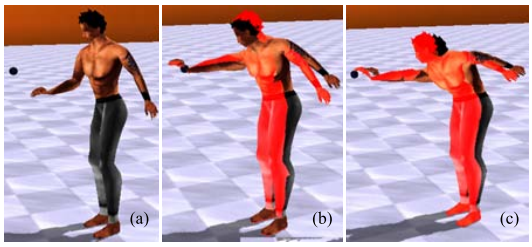


Figure 12: *(a) Original grasping motion. (b) and (c) The character's wrist is constrained while preserving the original COM position. The red character does not preserve the COM. (c) The constraint applied to the wrist cannot be verified.*

When a character is grasping a target that is really far, the arm and the trunk are simply bended in the direction of the target. Since no other limb is used to counterbalance those bending motions, the COM is not placed at its original position. This problem is solved by including the control of the COM in the iteration loop, just after the adaptation of the trunk, inthe function $adaptRoot()$. Hence, at the end of each iteration, the solver translates the root in order to replace the COM to its desired position. This position can simply be the original position of the COM in the original motion or can be obtained using inverse kinetics. Part (a) of figure 12 shows the original grasping motion. In part (b), this motion is adapted in order to verify three constraints while preserving the original COM position: the feet contacts on the ground and a constraint associated to the wrist. The red character, used as a reference, does not perform the control of the COM position while the other one moves its pelvis to counterbalance the upper-body motion. In part (c), the same constraints are used by the target is unreachable.

10. Results

To evaluate our method, we have implemented these algorithms in a more general real-time animation framework [MMKA04]. This framework synchronizes and blends motions in real-time. It also allows to extract the avatar morphology from H-ANIM compliant characters designed in VRML standard format (thanks to Avatar Studio, a product of Canal NuMedia, France). In figure 13 we depict a character that has to deal with four main constraints: two constraints to ensure feet contacts and two constraints to drive the wrist and the elbow of the same arm. The kinematic chain involved in the foot-contact constraints includes only the lower limbs and the pelvis. The constraints on the wrist and the elbow have the same kinematic chain: the corresponding upper-limb and the trunk. The highest priority is given to the wrist constraint whereas a lower one is associated to the elbow. With such priorities, we wish to ensure wrist-contact with the target and, if possible, to verify the constraint on the elbow.

Part (a) of figure 13 represents the original posture without constraint. In part (b), one can see that both the wrist and elbow constraints are verified. In this figure, the targets are represented with red and blue spheres, the yellow sphere locates the position of the COM. In part (c), the wrist reaches its target but the constraint on the elbow is not verified. This result demonstrates task-priority capabilities of our method with very few computation time.

Next, we used eight characters with different morphologies. Those characters are divided into two teams: the red team with red trousers and the white team with white ones (see *fightTraining.avi*). All the characters share the same motion database composed of only three motions: one is for the rest, one is a kick and the last motion allows the avatar to ward off the opponent's actions and then counterattack with a punch. In the sequence of fight training depicted in figure 14, the characters are working in pair. In part (a), white characters kick the red opponents trying to reach the torso (a constraint on the ankle to an absolute position). Concurrently, red characters ward off the kick thanks to the middle of their forearm (a constraint aiming at the opponent's ankle) before performing a punch to their opponents' head (a constraint aiming at the opponent's head or chin). Finally, the white characters bend their trunk and head back in order to avoid the punch, illustrated in part (b) of the figure. This last adaptation is simply done by using a constraint on the head with a target defined as an absolute position. The main complexity of this sequence is that the targets of the constraints change at each time depending on the opponents' gestures and anthropometric data. In order to avoid unwanted gestures at the beginning of the sequence, the constraints are activated after the gesture is initiated. Indeed, as the white character's calf is close to the ground at the beginning of the sequence, activating the constraint sooner would make the red character's forearm follow an unrealistic trajectory, beginning with a down motion before touching the calf at a high position. In the resulting sequence, the COM is constrained to follow the trajectory directly calculated from the captured file. This sequence is calculated in real-time for eight characters at 30Hz (on a P4 2.8GHz with GeForceFX5600) and additional characters could be added before decreasing this frame rate. For example, when the two arms and the trunk are adapted in order to respect two unreachable constraints on the wrists, up to 22 characters can be animated without visualization.

11. Discussion

In this paper, we proposed a representation of motion that simplifies the process of motion retargeting and adaptation in interactive environments where constraints can change at any time. In this representation, the data do not depend on body segments length. Consequently, scaling those data to

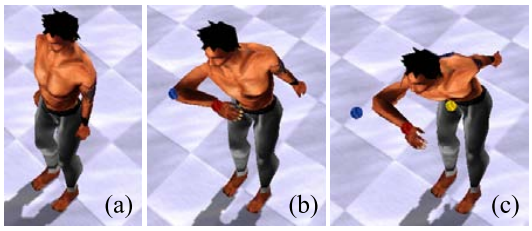


Figure 13: *Control of the wrist (with high priority) and the elbow (with a low priority) positions in order to reach two targets. The yellow sphere is the position of the COM. (a) Original posture without constraint. (b) Posture adaptation if the two targets are reachable. (c) Posture adaptation when only the constraint on the wrist can be verified.*

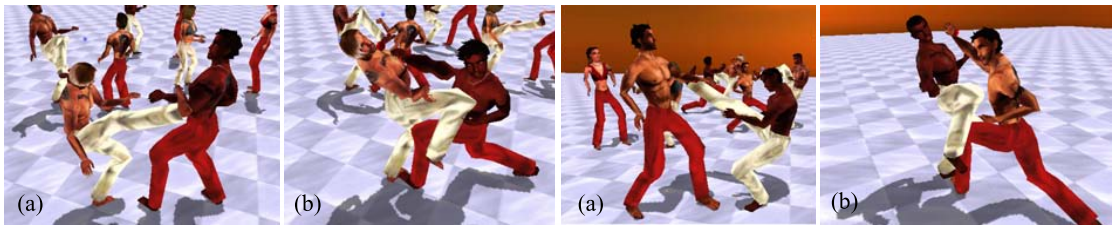


Figure 14: *(a) Red fighters ward off the opponent's kick with the forearm. (b) White fighters avoid the opponent's punch.*

new characters is performed very simply and rapidly. Moreover, the constraints that are intrinsically linked to the motion are also stored ensuring that the main properties of the original motion are preserved (such as foot-contacts). Previous works, such as [MBBT00], proposed to use an intermediate simplified skeleton to perform motion retargeting. The work presented in this paper is based on such an approach but also proposes a global framework that allows to take a better advantage of such a representation by proposing an adaptation of the CCD method.

In our representation of the motion, the root translation is normalized by the leg length which handles most of the motions. It would be interesting to extend this in order to take some specific motions into account. For example, for a handstand, the root translation should be normalized by the arm and trunk lengths.

Although an iterative process is used to adapt the motion at each time step, only few computation time is required to solve many constraints on a complete skeleton. Indeed, instead of modifying all the body segments in an iteration step, we only change groups of body segments. For example, the adaptation of the arm and the forearm orientation becomes the modification of the relative position of the wrist in the shoulder reference frame. Hence, for example, a constraint which is applied on one point of the forearm and which goal is a target in the Cartesian frame, is solved in one step with our method. If the target is not reachable then the arm is placed at the configuration that minimizes the error between the constraint position and the target. Moreover, with this hierarchical process, if the solution is reached by only moving the arm, there is no need to continue the adaptation to other groups of joints (such as moving the clavicle and the trunk). Consequently, this process naturally moves the heavier and proximal segments after the distal ones. This segmental sequence is classically identified in biomechanics for many natural gestures [Put93].

As prioritized inverse kinematics with Jacobian [CB04] does, we can deal with priorities. Consequently, high-priority constraints are verified first (if possible) before trying to verify constraints with lower ones. The user can modify these priorities interactively in order to obtain the desired animation. Finally, it would be interesting to offer the user another level of control of the constraints, making him access to different priorities for each articulation of a unique kinematic chain.

A control of the COM is proposed but is limited to the kinematic chain linked to the activated constraints. In our system, the COM is implicitly associated to the highest priority. Consequently, some other constraints may not be verified if they are not reachable without moving the COM. Future developments will allow to use the COM as the other groups, with its own priority. In the current version of this work, the control of the COM is ensured if a constraint changes the initial posture. Hence, the body segments involved in the control of the COM are limited to those included in the kinematic chain associated to the activated constraint. Consequently, if the constraint only acts on an arm and the trunk, it is not possible to recruit the legs or the other arm to control the COM position. As a perspective, we are working on a new algorithm that would be able to use other body parts in order to completely control the COM, if necessary.

Despite those limitations, our method provides animators with powerful tools for real-time animation of groups of humans in interactive environments. The characters can deal with many geometric constraints and only a small set of captured or edited motions is required. The method presented in this paper was implemented in industrial software environments for applications in multimedia and video games.

Acknowledgments

We would like to gratefully thank Stéphane Donikian for its helpful review of this paper and its suggestions. This work was partially funded by the French research program RIAM within the Semocap project.

References

[BB04] BAERLOCHER P., BOULIC R.: An inverse kinematics architecture enforcing an arbitrary number of strict priority levels. *Visual Computer 20*, 6 (2004), 402–417. 2

[BCHB03] BOULIC R., CALLENNEC B. L., HERREN

M., BAY H.: Motion editing with prioritized constraints. In *Proceedings of RichMedia* (Lausanne, Oct. 2003). 2

[BH00] BRAND M., HERTZMANN A.: Style machines. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (2000), ACM Press/Addison-Wesley Publishing Co., pp. 183–192. 2

[BMB87] BADLER N. I., MANOOCHEHRI K. H., BARAFF D.: Multi-dimensional input techniques and articulated figure positioning by multiple constraints. In *Proceedings of the 1986 workshop on Interactive 3D graphics* (1987), ACM Press, pp. 151–169. 6

[CB04] CALLENNEC B. L., BOULIC R.: Interactive motion deformation with prioritized constraints. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation* (2004), ACM Press, pp. 163–171. 2, 6, 9

[FP03] FANG A. C., POLLARD N. S.: Efficient synthesis of physically valid human motion. *ACM Trans. Graph. 22*, 3 (2003), 417–426. 2

[Gle98] GLEICHER M.: Retargetting motion to new characters. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques* (1998), ACM Press, pp. 33–42. 2

[GMHP04] GROCHOW K., MARTIN S. L., HERTZMANN A., POPOVIĆ Z.: Style-based inverse kinematics. *ACM Trans. Graph. 23*, 3 (2004), 522–531. 2

[GR96] GUO S., ROBERGÉ J.: A high-level control mechanism for human locomotion based on parametric frame space interpolation. In *Proceedings of the Eurographics workshop on Computer animation and simulation '96* (1996), Springer-Verlag New York, Inc., pp. 95–107. 1

[KGP02] KOVAR L., GLEICHER M., PIGHIN F.: Motion graphs. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques* (2002), ACM Press, pp. 473–482. 1

[Lan98] LANDER J.: Making kine more flexible. *Game Developer Magazine 1* (November 1998), 15–22. 2

[LCR*02] LEE J., CHAI J., REITSMA P. S. A., HODGINS J. K., POLLARD N. S.: Interactive control of avatars animated with human motion data. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques* (2002), ACM Press, pp. 491–500. 2

[LL04] LEE J., LEE K. H.: Precomputing avatar behavior from human motion data. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation* (2004), ACM Press, pp. 79–87. 1

[LS99] LEE J., SHIN S. Y.: *A hierarchical approach to interactive motion editing for human-like figures*. Acm press/addison-wesley publishing co., Proceedings of the 26th annual conference on Computer graphics and interactive techniques, 1999. 2

[LWS02] LI Y., WANG T., SHUM H.-Y.: Motion texture: a two-level statistical model for character motion synthesis. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques* (2002), ACM Press, pp. 465–472. 2

[MBBT00] MONZANI J.-S., BAERLOCHER P., BOULIC R., THALMANN D.: Using an intermediate skeleton and inverse kinematics for motion retargeting. *Computer Graphics Forum 19*, 3 (August 2000). ISSN 1067-7055. 2, 9

[MKMA04] MÉNARDAIS S., KULPA R., MULTON F., ARNALDI B.: Synchronization for dynamic blending of motions. In *Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (August 2004), pp. 325–336. 3

[MMKA04] MÉNARDAIS S., MULTON F., KULPA R., ARNALDI B.: Motion blending for real-time animation while accounting for the environment. In *Computer Graphics International* (June 2004). 8

[Put93] PUTNAM C.: Sequential motion of body segments in striking and throwing skills: description and explanation. *Journal of biomechanics 26* (1993), 125–135. 9

[PW99] POPOVIĆ Z., WITKIN A.: Physically based motion transformation. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques* (1999), ACM Press/Addison-Wesley Publishing Co., pp. 11–20. 2

[SHP04] SAFONOVA A., HODGINS J. K., POLLARD N. S.: Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. *ACM Trans. Graph. 23*, 3 (2004), 514–521. 2

[SLSG01] SHIN H. J., LEE J., SHIN S. Y., GLEICHER M.: Computer puppetry: An importance-based approach. *ACM Trans. Graph. 20*, 2 (2001), 67–94. 2

[SP05] SULEJMANPAŠIĆ A., POPOVIĆ J.: Adaptation of performed ballistic motion. *ACM Trans. Graph. 24*, 1 (2005), 165–179. 2

[TGB00] TOLANI D., GOSWAMI A., BADLER N. I.: Real-time inverse kinematics techniques for anthropomorphic limbs. *Graph. Models Image Process. 62*, 5 (2000), 353–388. 2

[TySK02] TAK S., YOUNG SONG O., KO H.-S.: Space-time sweeping: An interactive dynamic constraints solver. In *Proceedings of the Computer Animation* (2002), IEEE Computer Society, p. 261. 2

[WC91] WANG L.-C. T., CHEN C. C.: A combined optimization method for solving the inverse kinematics problem of mechanical manipulators. *IEEE Trans. On Robotics and Applications 7*, 4 (Aug. 1991), 489–499. 2

[Wel93] WELMAN C.: *Inverse Kinematics and Geometric Constraints For Articulated Figure Manipulation*. Master's thesis, Simon Fraser University, 1993. 2

EG:742

EG:743

Computational Photography

Eurographics 2007 Tutorial T7

Organizers & Speakers

Ramesh Raskar (MERL)
Jack Tumblin (Northwestern University)

EG:744

Computational Photography

Ramesh Raskar, Jack Tumblin, Ankit Mohan, Amit Agrawal, Yuanzen Li

MERL and Northwestern University, USA

Abstract

Computational photography combines plentiful computing, digital sensors, modern optics, actuators, probes and smart lights to escape the limitations of traditional film cameras and enables novel imaging applications. Unbounded dynamic range, variable focus, resolution, and depth of field, hints about shape, reflectance, and lighting, and new interactive forms of photos that are partly snapshots and partly videos are just some of the new applications found in Computational Photography. The computational techniques encompass methods from modification of imaging parameters during capture to sophisticated reconstructions from indirect measurements. We provide a practical guide to topics in image capture and manipulation methods for generating compelling pictures for computer graphics and for extracting scene properties for computer vision, with several examples.

Many ideas in computational photography are still relatively new to digital artists and programmers and there is no up-to-date reference text. A larger problem is that a multi-disciplinary field that combines ideas from computational methods and modern digital photography involves a steep learning curve. For example, photographers are not always familiar with advanced algorithms now emerging to capture high dynamic range images, but image processing researchers face difficulty in understanding the capture and noise issues in digital cameras. These topics, however, can be easily learned without extensive background. The goal of this presentation is to present both aspects in a compact form.

The new capture methods include sophisticated sensors, electromechanical actuators and on-board processing. Examples include adaptation to sensed scene depth and illumination, taking multiple pictures by varying camera parameters or actively modifying the flash illumination parameters. A class of modern reconstruction methods is also emerging. The methods can achieve a 'photomontage' by optimally fusing information from multiple images, improve signal to noise ratio and extract scene features such as depth edges. The presentation briefly reviews fundamental topics in digital imaging and then provides a practical guide to underlying techniques beyond image processing such as gradient domain operations, graph cuts, bilateral filters and optimizations.

The participants learn about topics in image capture and manipulation methods for generating compelling pictures for computer graphics and for extracting scene properties for computer vision, with several examples. We hope to provide enough fundamentals to satisfy the technical specialist without intimidating the curious graphics researcher interested in recent advances in photography.

The intended audience is photographers, digital artists, image processing programmers and vision researchers using or building applications for digital cameras or images. They will learn about camera fundamentals and powerful computational tools, along with many real world examples.

1 Introduction

1.1 Film-like Photography

Photography is the process of making pictures by, literally, 'drawing with light' or recording the visually meaningful changes in the light leaving a scene. This goal was established for film photography about 150 years ago.

Currently, 'digital photography' is electronically implemented film photography, refined and polished to achieve the goals of the classic film camera which were governed by chemistry, optics, mechanical shutters. Film-like photography presumes (and often requires) artful human judgment, intervention, and interpretation at every stage to choose viewpoint, framing, timing, lenses, film properties, lighting, developing, printing, display, search, index, and labeling.

In this document, we plan to explore a progression away from film and film-like methods to something more comprehensive that exploits plentiful low-cost computing and memory with sensors, optics, probes, smart lighting and communication.

1.2 What is Computational Photography?

Computational Photography (CP) is an emerging field, just getting started. We don't know where it will end up, we can't yet set its precise, complete definition, nor make a reliably comprehensive classification. But here is the scope of what researchers are currently exploring in this field.

- Computational photography attempts to record a richer visual experience, captures information beyond just a simple set of pixels and makes the recorded scene representation far more machine readable.

- It exploits computing, memory, interaction and communications to overcome long-standing limitations of photographic film and camera mechanics that have persisted in film-style digital photography, such as constraints on dynamic range, depth of field, field of view, resolution and the extent of scene motion during exposure.

- It enables new classes of recording the visual signal such as the 'moment' [Cohen 2005], shape boundaries for non-photorealistic depiction [Raskar et al 2004] , foreground versus background mattes, estimates of 3D structure, 'relightable' photos and interactive displays that permit users to change lighting, viewpoint, focus, and more, capturing some useful, meaningful fraction of the 'light field' of a scene, a 4-D set of viewing rays.

- It enables synthesis of impossible photos that could not have been captured at a single instant with a single camera, such as wrap-around views ('multiple-center-of-projection' images [Rademacher and Bishop 1998]), fusion of time-lapsed events [Raskar et al 2004], the motion-microscope (motion magnification [Liu et al 2005]), video textures and panoramas [Agarwala et al 2005]. They also support seemly impossible camera movements such as the 'bullet time' (Matrix) sequence recorded with multiple cameras with staggered exposure times.

- It encompass previously exotic forms of scientific imaging and data gathering techniques e.g. from astronomy, microscopy, and tomography.

1.3 Elements of Computational Photography

Traditional film-like photography involves (a) a lens, (b) a 2D planar sensor and (c) a processor that converts sensed values into an image. In addition, the photography may involve (d) external illumination from point sources (e.g. flash units) and area sources (e.g. studio lights).

Computational Photography generalizes these four elements.

(a) Generalized Optics: Each optical element is treated as a 4D ray-bender that modifies a light field. The incident 4D light field for a given wavelength is transformed into a new 4D lightfield. The optics may involve more than one optical axis [Georgiev et al 2006]. In some cases, the perspective foreshortening of objects based on distance may be modified using wavefront coded optics [Dowski and Cathey 1995]. In recent lensless imaging methods [Zomet and Nayar 2006] and coded-aperture imaging [Zand 1996] used for gamma-ray and X-ray astronomy, the traditional lens is missing entirely. In some cases optical elements such as mirrors [Nayar et al 2004] outside the camera adjust the linear combinations of ray bundles that reach the sensor pixel to adapt the sensor to the viewed scene.

(b) Generalized Sensors: All light sensors measure some combined fraction of the 4D light field impinging on it, but traditional sensors capture only a 2D projection of this lightfield. Computational photography attempts to capture more; a 3D or 4D ray representation using planar, non-planar or even volumentric sensor assemblies. For example, a traditional out-of-focus 2D image is the result of a capture-time decision: each detector pixel gathers light from its own bundle of rays that do not converge on the focused object. But a Plenoptic Camera [Adelson and Wang 1992, Ren et al 2005] subdivides these bundles into separate measurements. Computing a weighted sum of rays that converge on the objects in the scene creates a digitally refocused image, and even permits multiple focusing distances within a single computed image. Generalizing sensors can extend their dynamic range [Tumblin et al 2005] and wavelength selectivity as well. While traditional sensors trade spatial resolution for color measurement (wavelengths) using a Bayer grid or red, green or blue filters on individual pixels, some modern sensor designs determine photon wavelength by sensor penetration, permitting several spectral estimates at a single pixel location [Foveon 2004].

(c) Generalized Reconstruction: Conversion of raw sensor outputs into picture values can be much more sophisticated. While existing digital cameras perform 'de-mosaicking,' (interpolate the Bayer grid), remove fixed-pattern noise, and hide 'dead' pixel sensors, recent work in computational photography can do more. Reconstruction might combine disparate measurements in novel ways by considering the camera intrinsic parameters used during capture. For example, the processing might construct a high dynamic range scene from multiple photographs from coaxial lenses, from

sensed gradients [Tumblin et al 2005], or compute sharp images a fast moving object from a single image taken by a camera with a 'fluttering' shutter [Raskar et al 2006]. Closed-loop control during photography itself can also be extended, exploiting traditional cameras' exposure control, image stabilizing, and focus, as new opportunities for modulating the scene's optical signal for later decoding.

(d) Computational Illumination: Photographic lighting has changed very little since the 1950's: with digital video projectors, servos, and device-to-device communication, we have new opportunities to control the sources of light with as much sophistication as we use to control our digital sensors. What sorts of spatio-temporal modulations for light might better reveal the visually important contents of a scene? Harold Edgerton showed high-speed strobes offered tremendous new appearance-capturing capabilities; how many new advantages can we realize by replacing 'dumb' the flash units, static spot lights and reflectors with actively controlled spatio-temporal modulators and optics? Already we can capture occluding edges with multiple flashes [Raskar 2004], exchange cameras and projectors by Helmholz reciprocity [Sen et al 2005], gather relightable actor's performances with light stages [Wagner et al 2005] and see through muddy water with coded-mask illumination [Levoy et al 2004]. In every case, better lighting control during capture to builds richer representations of photographed scenes.

2 Sampling Dimensions of Imaging

2.1 Epsilon Photography for Optimizing Film-like Camera

Think of film cameras at their best as defining a 'box' in the multi-dimensional space of imaging parameters. The first, most obvious thing we can do to improve digital cameras is to expand this box in every conceivable dimension. This effort reduces Computational Photography to 'Epsilon Photography', where the scene is recorded via multiple images, each captured by epsilon variation of the camera parameters. For example, successive images (or neighboring pixels) may have different settings for parameters such as exposure, focus, aperture, view, illumination, or the instant of capture. Each setting allows recording of partial information about the scene and the final image is reconstructed from these multiple observations. Epsilon photography is thus concatenation of many such boxes in parameter space; multiple film-style photos computationally merged to make a more complete photo or scene description. While the merged photo is superior, each of the individual photos is still useful and comprehensible on its own, without any of the others. The merged photo contains the best features from all of them.

(a) Field of View: A wide field of view panorama is achieved by stitching and mosaicking pictures taken by panning a camera around a common center of projection or by translating a camera over a near-planar scene.

(b) Dynamic range: A high dynamic range image is captured by merging photos at a series of exposure values [Debevec and Malik 1997, Kang et al 2003]

(c) Depth of field: All-in-focus image is reconstructed from images taken by successively changing the plane of focus [Agrawala et al 2005].

(d) Spatial Resolution: Higher resolution is achieved by tiling multiple cameras (and mosaicing individual images) [Wilburn et al 2005] or by jittering a single camera [Landolt et al 2001].

(e) Wavelength resolution: Traditional cameras sample only 3 basis colors. But multi-spectral (multiple colors in the visible spectrum) or hyper-spectral (wavelengths beyond the visible spectrum) imaging is accomplished by taking pictures while successively changing color filters in front of the camera, using tunable wavelength filters or using diffraction gratings.

(f) Temporal resolution: High speed imaging is achieved by staggering the exposure time of multiple low-framerate cameras. The exposure durations of individual cameras can be non-overlapping [Wilburn et al 2005] or overlapping [Shechtman et al 2002].

Taking multiple images under varying camera parameters can be achieved in several ways. The images can be taken with a single camera over time. The images can be captured simultaneously using 'assorted pixels' where each pixel is a tuned to a different value for a given parameter [Nayar and Narsimhan 2002]. Simultaneous capture of multiple samples can also be recorded using multiple cameras, each camera having different values for a given parameter. Two designs are currently being used for multi-camera solutions: a camera array [Wilburn et al 2005] and single-axis multiple parameter (co-axial) cameras [Mcguire et al 2005].

2.2 Coded Photography

But there is much more beyond the 'best possible film camera'. Instead of increasing the field of view by panning a camera, can we create a wrap-around view of an object? Panning a camera allows us to concatenate and expand the the box in the camera parameter space in the dimension of 'field of view'. But a wrap around view spans multiple disjoint pieces along these dimensions. We can virtualize the notion of the camera itself if we consider it as a device that collects bundles of rays, each ray with its own wavelength spectrum.

Coded Photography is a notion of an 'out-of-the-box' photographic method, in which individual (ray) samples or data sets are not comprehensible as 'images' without further decoding, re-binning or reconstruction. For example, a wrap around view is built from images taken with multiple centers of projection but by taking only a few pixels from each input image. Some other examples include confocal images and coded aperture images.

We may be converging on a new, much more capable 'box' of parameters in computational photography that we don't yet recognize; there is still quite a bit of innovation to come!

In the rest of the document, we survey recent techniques that exploit exposure, focus and active illumination.

3 High Dynamic Range

3.1 Multiple Exposures

One approach of capturing high dynamic range scenes is to capture multiple images using different exposures, and then merge these images. The basic idea is that when high exposures are used, dark regions are well imaged but bright regions are saturated. On the other hand, when low exposures are used, dark regions are too dark but bright regions are well imaged. If exposure varies and multiple pictures are taken of the same scene, value of a pixel can be taken from those images where it's neither too dark nor saturated. This type of approach is often referred to as exposure bracketing, and has been widely adopted [Morimura 1993, Burt and Kolczynski 1993,Madden 1993,Tsai 1994]. Imaging devices usually contain nonlinearities, where pixel values are nonlinearly related to the brightness values in the scene. Some authors have proposed to use images acquired under different exposures to estimate the radiometric response function of an imaging device, and use the estimated response function to process the images before merging them [Mann and Picard 1995, Debevec and Malik 1997, Mitsunaga and Nayar 1999.]

3.2 Sensor Design

At the sensor level, various approaches have also been proposed for high dynamic range imaging. One type of approach is to use multiple sensing elements with different sensitivities within each cell [Street 1998, Handy 1986, Wen 1989, Hamazaki 1996]. Multiple measurements are made from the sensing elements, and they are combined on-chip before a high dynamic range image is read out from the chip. Spatial sampling rate is lowered in these sensing devices, and spatial resolution is sacrificed. Another type of approach is to adjust the well capacity of the sensing elements during photocurrent integration [Knight 1983, Sayag 1990, Decker 1998] but this gives higher noise. A different approach is proposed by [Brajovic and Kanade 1996], where the time it takes to reach saturation is measured, by a computation element attached to each sensing element. This time encodes high dynamic range information, as it is inversely proportional to the brightness at each pixel. Logarithmic sensors [Scheffer et al 2000] have also been proposed to increase the dynamic range. Brightside exploits the interline transfer of a charge coupled device (CCD) based camera to capture two exposures during a single mechanical shutter timing.

High dynamic range sensor design is in progress, but the implementation is usually costly. A rather novel and flexible approach is proposed by [Nayar and Mitsunaga 2000, Narasimhan and Nayar 2005], where exposures vary across space of the imager. A pattern with varying sensitivities is applied to the pixel array. It resembles the Bayer pattern in color imaging, but the sampling is made along the exposure instead of wavelength. The particular form of the sensitivity pattern, and the way of implementing it, are both quite flexible. One way of implementing it is to place a mask with cells of varying optical transparencies in front of the sensing array. Here, just as in Bayer mosaic, spatial resolution is sacrificed to some extent and aliasing can occur. Measurements under different exposures (sensitivities) are spatially interpolated, and combined into a high dynamic range image.

4 Aperture and Focus

Several concepts in exploiting focus and aperture parameters can be understood by considering the 4D lightfields transfer via lens and its 2D, 3D or 4D projection recorded on the image sensor.

Defocus Video Matting

Video matting is the process of recovering a high-quality alpha matte and foreground from a video sequence. Common approaches require either a known background (e.g., a blue screen) or extensive user interaction (e.g., to specify known foreground and background elements). The matting problem is generally under-constrained, unless additional information is recorded at the time of capture. McGuire et. al. have proposed a novel, fully autonomous method for pulling a matte using multiple synchronized video cameras that share the center of projection but differ in their plane of focus [McGuire et. al 2005]. The multi-camera data stream over-constrains the problem and the solution is obtained by directly minimizing the error in filter-based image formation equations. Their system solves the fully dynamic video matting problem without user assistance: both the foreground and background may be high frequency and have dynamic content, the foreground may resemble the background, and the scene may be lit by natural (as opposed to polarized or collimated) illumination. The authors capture 3 synchronized video

streams using a 3 cameras and beam splitters. The first camera has a pinhole sensor has a small aperture that creates a large depth of field. The second and third cameras have large apertures, creating narrower depths of field focused on foreground and background, respectively. The foreground sensor produces sharp images for objects within about 0.5m of depth of the foreground object and defocuses objects farther away. The background sensor produces sharp images for objects from about 5m to infinity and defocuses the foreground object. Given the three video streams, at each frame the optical formation of each of the three images is expressed as the function of the unknowns background, foreground and alpha values.

Plenoptic Camera

Ren et. al. have developed a camera that can capture the 4D light field incident on the image sensor in a single photographic exposure [Ren et al 2005]. This is achieved by inserting a microlens array between the sensor and main lens, creating a plenoptic camera. Each microlens measures not just the total amount of light deposited at that location, but how much light arrives along each ray. By re-sorting the measured rays of light to where they would have terminated in slightly different, synthetic cameras, one can compute sharp photographs focused at different depths. A linear increase in the resolution of images under each microlens results in a linear increase in the sharpness of the refocused photographs. This property allows one to extend the depth of field of the camera without reducing the aperture, enabling shorter exposures and lower image noise.
To the photographer, the plenoptic camera operates exactly like an ordinary hand-held camera. The ability to digitally refocus and extend the depth of field is ideal of portraits, high-speed action and macro close-ups. In a related paper, the authors have derived a Fourier representation of photographic imaging. The Fourier representation is conceptually and computationally simpler than the spatial domain representation. The theory enables one to compute photographs focused at different depths more quickly from the 4D light field data.

Synthetic Aperture Imaging

Synthetic aperture focusing consists of warping and adding together the images in a 4D light field so that objects lying on a specified surface are aligned and thus in focus, while objects lying off this surface are misaligned and hence blurred. This provides the ability to see through partial occluders such as foliage and crowds, making it a potentially powerful tool for surveillance [Vaish et al 2004].

Confocal microscopy is a family of imaging techniques that employ focused patterned illumination and synchronized imaging to create cross-sectional views of 3D biological specimens. Levoy et. al. have adapted confocal imaging to large-scale scenes by replacing the optical apertures used in microscopy with arrays of real or virtual video projectors and cameras [Levoy 2004]. A dense array of projectors allows to simulate a wide aperture (Synthetic Aperture Illumination) projector which can produce a real image with small depth of field. By projecting coded patterns and combining the resulting views using an array of virtual projectors, one can selectively image any plane in a partially occluded environment. These ideas were demonstrated on enhancing visibility in weakly scattering environments, such as murky water, to compute cross-sectional images and to see through partially occluded environments, such as foliage.

5 Motion Blur

Motion Deblurring using Hybrid Imaging

Motion blur due to camera motion can significantly degrade the quality of an image. Since the path of the camera motion can be arbitrary, deblurring of motion blurred images is a hard problem. Previous methods to deal with this problem have included blind restoration of motion blurred images, optical correction using stabilized lenses, and special CMOS sensors that limit the exposure time in the presence of motion. Ben-Ezra et. al. exploit the fundamental trade off between spatial resolution and temporal resolution to construct a hybrid camera that can measure its own motion during image integration [Ben-Ezra and Nayar 2005]. The acquired motion information is used to compute a point spread function (PSF) that represents the path of the camera during integration. This PSF is then used to deblur the image. Results were shown on several indoor and outdoor scenes using long exposure and complex camera motion paths.

The hybrid imaging system proposed by the author consists of a high resolution primary detector and a low resolution secondary detector. The secondary detector is used to compute the motion information and the PSF. The motion between successive frames is limited to a global rigid transformation model which is computed using a multi-resolution iterative algorithm that minimizes the optical flow based error function. The resulting continuous PSF is then used for motion deblurring using the Richardson-Lucy algorithm. The authors used a 3M pixel Nikon still camera as the primary detector and a Sony DV camcoder as the secondary detector. The two detectors were calibrated offline. Results on several real sequences with exposure time ranging from 0.5 seconds to 4 seconds and the blur ranging up to 130 pixels were shown.

Recently, Fergus et al have shown that, in case of camera shake, the point spread function can be estimated from a single image. They exploit the natural image statistics on image gradients and then use the probably blur function to deblur the image [Fergus et al 2006].

Blur due to camera shake is different from blur due to object motion. And so far, there appears to be no good techniques for estimating object motion blur function.

Coded Exposure

In a conventional single-exposure photograph, moving objects or moving cameras cause motion blur. The exposure time defines a temporal box filter that smears the moving object across the image by convolution. This box filter destroys important high-frequency spatial details so that deblurring via deconvolution becomes an ill-posed problem. Raskar et. al. have proposed to flutter the camera's shutter open and closed during the chosen exposure time with a binary pseudo-random sequence, instead of leaving it open as in a traditional camera [Raskar et al 2006]. The flutter changes the box filter to a broad-band filter that preserves high-frequency spatial details in the blurred image and the corresponding deconvolution becomes a well-posed problem.

Results on several challenging cases of motion-blur removal including outdoor scenes, extremely large motions, textured backgrounds and partial occluders were presented. However, the authors assume that PSF is given or is obtained by simple user interaction. Since changing the integration time of conventional CCD cameras is not feasible, an external ferro-electric shutter is placed in front of the lens to code the exposure. The shutter is driven opaque and transparent according to the binary signals generated from PIC using the pseudo-random binary sequence.

6 Computational Illumination

6.1 Flash-no flash

The simplest form of computational illumination is perhaps the ubiquitous camera flash. [DiCarlo et al 2001] first explored the idea of capturing a pair of images for the same camera position - one illuminated with ambient light only, and the other using the camera flash as an additional light source. They use this image pair to estimate object reflectance functions, an the spectral distribution of the ambient illumination. [Hoppe et al.2003] acquire multiple photos under different flash intensities, and allow the user to interpolate between them to simulate intermediate flash intensities.

Concurrent work by [Petschnigg et al. 2004] and [Eisemann et al.2004] proposed very similar techniques of combining the information contained in the flash and no-flash image pair to generate a single nice image. The no-flash photo captures the large-scale illumination effects such as the ambiance of the scene. However, in a low-light situation, the no-flash photo generally has excessive noise. The flash photo in contrast has much lower noise and more high frequency details, but fails to preserve the mood of the scene. The basic idea here is to decouple the high and low frequency components of the images, and then recombine to preserve the desired characteristics (detail from the flash photo, and large scale ambiance from the no-flash photo). This decoupling is done using a modified bilateral filter called joint bilateral filter,

The bilateral filter is basically an edge-preserving blur that gives the low frequency component of the photo. In the joint bilateral filter, the intensity difference in the flash photo is used. Since the flash photo has lower noise, this gives a better results and avoids over or under blurring.

Agrawal et al. [Agrawal et al 20005] use the flash no-flash photo pair to remove reflections and hotspots from flash photos. They rely on the observation that the orientation of image gradients due to reflectance geometry are illumination invariant, while those due to changes in illumination are not. They propose a gradient projection scheme to decompose the illumination effects from the rest of the image. Based on the ratio of the flash and no-flash photos, they compensate for flash intensity falloff due to depth. Finally, they also propose a unified flash-exposure space that contains photos taken by varying the flash intensity and the shutter speed, and a method for adaptively sampling this space to capture a flash-exposure high dynamic range image.

Raskar et al.[Raskar et al 2004] used a multi-flash camera to find the silhouettes in a scene. They take four photos of an object with four different light positions (above, below, left and right of the lens). They detect shadows cast along the depth discontinuities are use them to detect depth discontinuities in the scene. The detected silhouettes are then used for stylizing the photograph and highlighting important features. They also demonstrate silhouette detection in a video using a repeated fast sequence of flashes.

6.2 4D acquisition

Light fields [Levoy 1996] and Lumigraph [Gortler 1996} reduced the more general plenoptic function [Adelson 1991] to a four dimensional function, L(u,v,s,t) that describes the presence of light in free space, ignoring the effect of wavelength and time. Here (u,v) and (s,t) are the parameters on two parallel planes respectively that describe a ray of light in space. A slightly different parameterization can be used to

describe the incident light field on an object. If we think of the object surrounded by a while sphere of imaginary projectors looking inwards, ($theta_i$, phi_i) describes the angular position of the projector on the unit sphere, and (u,v) the pixel position on that projector. Thus, the function L_i(u,v,theta,phi) gives complete control over the incident light on an object in free space. Similarly a sphere of inward looking cameras would capture the entire radiant light field of an object, L_r(u,v,theta,phi). Debevec et al.[Debevec et al 2001] introduced the 8D reflectance field that describes relationship of the incident and radiant light fields of a scene. An additional dimension of time is sometimes added to describe light interaction with an object that changes over time.

While the reflectance field gives a complete description of how light interacts with a scene, acquiring this complete function would require enormous amounts of time and storage. Significant work has been done in trying to acquire lower dimensional subsets of this function, and using it for restricted re-lighting and rendering.

Most image based relighting work relies on the simple observation that light interacts linearly with materials [Nimeroff 1994, Haeberli 1992]. If a fixed camera makes an image I_i from a fixed scene lit only by a light L_i , then the same scene lit by many lights scaled by weights w_i will make an image I_{out}=sum$_i$ ($w_i I_i$). Adjusting weights lets us ``relight'' the image, as if the weights modulate the lights rather than the images.

Debevec et al.[Debevec et al 2001] used a light stage comprising of a light mounted on a rotating robotic arm to acquire the non-local reflectance field of a human face. The point-like light source can be thought of as a simplified projector with a single pixel. Thus, the incident light field is reduced to a 2D function. They acquired images of the face using a small number of cameras with densely sampled lighting directions. They demonstrated generation of novel images from the original viewpoints under arbitrary illumination. This is done by simply adjusting the weights w_i to match the desired illumination intensity from different directions. They also are also able to simulate small changes in the viewpoint using a simple model for the skin reflectance. Hawkins et al.[Hawkins et al 2001] used a similar setup and used it for digitizing cultural artifacts. They argue for the use reflectance field in digital archiving instead of geometric models and reflectance textures. Koudelka et al.[Koudelka et al 2001} acquire a set of images from a single viewpoint as a point light source moved around the object, and estimate the surface geometry by using two set of basis images. They then estimate the apparent BRDF for each pixel in the images, and use this to render the object under arbitrary illumination.

Debevec et al.[Debevec ey al 2002} proposed an enhanced light stage comprising of a large number (156) of inward pointing LEDs distributed on a spherical structure, about two meters in diameter, around the actor. They set each light to an arbitrary color and intensity to simulate the effect of a real world environment around the actor. The images gathered by the light stage, together with a mask of the actor captured using infrared sources and detector, were used to seamlessly composite the actor into a virtual set while maintaining consistent illumination. Malzblender et al. [Malzbender et al 2001] used 50 inward looking flashes placed on a hemispherical dome and a novel scheme for compressing and storing the 4D reflectance field, called the Polynomial Texture Map. They assumed that the color of a pixel changed smoothly as the light moved around the object, and store only the coefficients of a biquadratic polynomial that best models this change for each pixel. This highly compact representation allows for real time rendering of the scene with arbitrary illumination, and works fairly well for diffuse objects; specular highlights are not modeled very nicely by the polynomial model and result in visual artifacts.

The free-form light stage [Masselus 2002] presented a way to acquire a 4D slice of the reflectance field without the use of an extensive light-stage. Instead, they used a handheld, free-moving light source around the object. The light position was estimated automatically from four diffuse spheres placed near the object in the field of view of the camera. The data acquisition time was reported as 25-30 minutes. Winnemoller et al. [Winnemoeller et al 2005] used dimensionality reduction and a slightly constrained light scanning pattern to estimate approximate light source position without the need for any additional fiducials in the scene.

Akers et al. [Akers et al 2003] use spatially varying image weights on images acquired with a light stage similar to [Debevec et al 2001]. They use a painting interface allow an artist to locally modify the relit image as desired. While the spatially varying mask gives greater flexibility, it might also gives results that are not physically realizable and look unrealistic. [Anrys et al.2004] and [Mohan et al.2005] used a similar painting interface to help a novice user in lighting design for photography. The users sketch a target image, and the system finds optimal weights for each basis image to get a physically realizable result that is closest to the target. [Mohan et al.2005] argue that accurate calibration is not necessary for the application photographic relighting, and propose a novel reflector based acquisition system. They place a moving-head gimbaled disco light inside a diffuse enclosure, together with the object to be photographed. The spot from the light on the enclosure acts as an area light source that illuminates the object. The light source is moved by simply rotating the light and capturing images for various light positions. The idea of area light sources was also used in bayesian relighting [Fuchs 2005].

7 Future Directions

7.1 Smart Sensors

Digital camera sensors typically use a color mosaic or a Bayer pattern of R, G, and B filters to sense 3 different spectral bands, forming a basis for color reproduction. So-called 'demosaicing' methods, though widely varied and often proprietary, convert raw, interleaved color sensor values from the Bayer grid into R,G,B estimates for each pixel with as many luminance details and as few chrominance artifacts as possible, but the task itself forces tradeoffs and continued innovation.. Sony's four color CCD uses 'emerald' pixels which allow for correcting for defects in the rendition of red tones at certain frequencies. The Foveon sensor found in some Sigma digital cameras avoids the Bayer filter entirely, and instead detects wavelength bands for color according to photon penetration depths in a novel silicon detector design that stacks three layers of photodetectors, one below the other. This eliminates all the potential errors and artifacts of demosaicking, and reduces post-processing requirements substantially.

By sensing different between neighboring pixels instead of actual intensities, Tumblin et al [Tumblin et al 2005] have shown that a 'Gradient Camera' can record large global variations in intensity. Rather than measure absolute intensity values at each pixel, this proposed sensor measures only forward differences between them, which remain small even for extremely high-dynamic range scenes, and reconstructs the sensed image from these differences using Poisson solver methods. This approach offers several advantages: the sensor is nearly impossible to over- or under-expose, yet offers extremely fine quantization, even with very modest A/D converters (e.g. 8 bits). The thermal and quantization noise occurs in the gradient domain, and appears as low frequency 'cloudy' noise in the reconstruction, rather than uncorrelated high-frequency noise that might obscure the exact position of scene edges.

Several companies now offer '3D cameras' that estimate depth for each pixel of the images they gather. Systems by Canesta and Zcam operate by precise measurement of the 'time-of-flight' (TOF) required for modulated infrared illumination to leave the camera, reflect from the scene and return to fast camera sensors. Several earlier, laser-based TOF systems, e.g. Cyberware, used 'flying spot' scanning to estimate depth sequentially. Without scanning these newer systems apply incoherent light (e.g. IR LEDs) and electronic gating to build whole-frame depth estimates at video rates. Canesta systems integrate the emitters in the same chip substrate as the detector, enabling a compact single-chip sensor unit; the Zcam device augments professional television camera units (ENG) to provide real-time depth keying and 3D reprojection.

Line Scan cameras. Several systems for critically-timed sports (e.g. sprints, horse racing) high-speed narrow-view or line-scan cameras hold more opportunities for capturing visual appearance. The 'FinishLynx' Lynx System Developers Inc. camera views a race finish-line through a narrow vertical slit, and assembles and image whose horizontal axis measures time instead of position. Despite occasionally strange distortions, the camera reliably depicts the first racer's body part to cross the finish line as the right-most feature in the time-space image.

7.2 Smart Optics

Wavefront coded imaging. Geometric aberrations in lenses cause image distortions, but these distortions can be modeled, computed, and in some cases robustly reversed. In 1995, Dowski and Cathey introduced a 'wavefront coded' optical element that forms intentionally distorted images with small, low cost optics [Dowski and Cathey 1995]. These seemingly out-of-focus images are computationally reversible, and allow reconstruction of an image with extended depth of focus, forming images with a focusing range up to 10X the abilities of conventional lenses. What other optical distortions might prove similarly advantageous?

Plenoptic Camera. As early as 1992 [Adelson and Wang 1992] several researchers have recognized the value of sensing the direction of incident light at each point on the focal plane behind a lens. Adelson's 1992 camera system combined a large front lens and a field of micro-lenses behind it, gathering what is now known as a 4D light field estimate, and he used it for single-lens stereo reconstruction. More recently, [Ng et al 2005] refined the idea further with an elegant hand-held digital camera for light-field capture that permits digital re-focussing and slight changes of viewpoint computationally.

Recently [Georgiev et al 2006] modeled the optics of these cameras using ray-matrix formulation, and showed an intriguing alternative. Instead of adding many tiny microlenses directly on top of delicate camera sensors, he builds a bundle large lenses and prisms attached to externally to the camera. The resulting light-field captured allows much larger computational changes in viewpoint in exchange for coarser digital re-focussing.

As these examples indicate, we have scarcely begun to explore the possibilities offered by combining computation, 4D modeling of light transport, and novel

optical systems. Nor have such explorations been limited to photography and computer graphics. Computer vision, microscopy, tomography, astronomy and other optically driven fields already contain some ready-to-use solutions to borrow and extend. For example, N. Ahuja has explored cameras with spinning dispersion plates to allow a single camera to gather images from many virtual viewpoints for robust stereo reconstruction. How might other spinning optical elements help with appearance capture?

Tools for Optics

Until recently, ray-based models of light transport have been entirely adequate for computer graphics and computer vision, sometimes extended with special-case models for diffraction [Stam 1999]. Some early excursions into **wave optics** models by Gershon Elber [Elber 1994] proved computationally intense, and pinhole-camera and ideal-thin-lens models of optics have been entirely adequate for computer graphics use.

As computational photography considers more complex lens systems, ray-only models of light transport begin to fail; to adequately model the spatial frequency response and wavelength dependence of optical systems we can move first to ray-matrix formulations, commonly used for optical fiber models and single-axis multi-lens systems, or move to **Fourier Optics** models to more accurately model the diffraction effects that predict the spatial frequency response of lens systems with adjustable apertures, and include accurate modeling of coherent light as well, including holography. The classic text by Goodman [Goodman 1968] is an elegant introduction to this topic. The computational requirements for Fourier analysis of optics is no longer formidable, especially with GPU assistance, and recent work by [Ng 2005] has already tied lightfields to images by showing it follows the 4D projection-slice theorem [Rosenfeld and Kak, 1987] that became a fundamental tenet of medical tomography.

Beyond Fourier Optics, we can resort to specialized lens-design descriptors such as Zernicke polynomials and remain within the realm of practical computation. Further refinement by resorting to full electromagnetic simulation can model polarization and optical effects due to structures smaller than the wavelength of light. These models can directly predict the optical behavior of superlattice structures such as iridescent butterfly wings, the transparency of finely-fibred structures such as the lens and cornea of the eye, and strange retro-reflectance properties of some classes of diseased cell bodies. While medical researchers and others are actively pursuing such simulations, the computational requirements are still daunting, and appear out of reach for current experiments in computational photography.

7.3 Other Dimensions

As noted in the 'Assorted pixels' paper [Nayar2003], photographic capture gathers optical data along many dimensions, and few are fully exploited. In 4-dimensional ray space we sense and measure more than simple intensity (or more formally, radiance), but also visually assess wavelength, time, materials, illumination direction and more. **Polarization** is also sometimes revealing, and the mapping from polarization direction of the illuminant to the polarization of reflected light is not a simple one: for some biological materials, the mappings are nonlinear and unexplored [Wu et al 2003]. Extended exploration of wavelength dependence is already well advanced. Hyperspectral imaging has already gathered a rich and growing literature for a broad range of applications from astronomy to archival imaging of museum treasures.

Film-style photography relies on an 'instantaneous' ideal: we attempt 'stop time' by capturing any photographed scene quickly enough to ignore any movement that happens during the measurement process. Even 'motion pictures' commit serial attempts at instantaneous capture, rather than direct sensing of the motions themselves. Harold Edgerton pushed the instantaneous ideal to extremes by using ultra-short strobes to illuminate transient phenomena, and ultra-short shutters to measure ultra-bright phenomena quickly, such as his famous high-speed movies of atomic bomb explosions.

Digital sensors offer new opportunities for more direct sensing, and digital displays permit interactive display of the movements we capture. Accordingly, Michael Cohen has proposed that the film-rooted distinction between 'still' cameras and 'video' cameras should gradually disappear. He proposed that we need an intermediate digital entity he calls a '**moment'**; one visually meaningful action we wish to remember—a child's fleeting expression of delighted surprise, a whisper of wind that sways the trees, etc., and it might fit in short video clips [Cohen 2005]. Motion sensing and deblurring itself can improve in the future [BenEzra 2004, Raskar 2006]. Movement also causes difficulties for constructing panoramas. However, if the movement is statistically consistent, it is possible to combine conventional image stitching operations with so-called 'video texturing' [Schödl 2000] methods to create consistent, seamless movement that captures the 'moment' of the panorama quite well. It can be further extended to capture video texture panoramas [Agarwala 2005]

7.4 Scientific Imaging

Scene measurement and representation in 4-D and beyond encompasses previously isolated "Islands" of Ingenious Scientific Imaging & Measuring. What can we learn from them? Can we extend their methods? Particularly promising fields include the following.

(i) Tomography: For any penetrating measurements, attenuation along straight-line paths can be used to construct 3D images of internal structures. This is currently used measuring sound transmission to electrical capacitance, from seismographic disturbances to ultrasonics to X-rays.

(ii) Spectrographic methods: complex interdependencies between wavelengths, reflectance, and transmissions are used for image forming, and broad classes of statistical measurements help decipher or identify useful features for land management, pollution studies, atmospheric patterns, wildlife migration, and geological and mineral features..

(iii) Confocal Methods and Synthetic Aperture methods: As described above, one can achieve very narrow depth-of-field image by collecting a widely divergent rays from each imaged point and these methods can extend to macroscopic scales via multiple cameras and multiple video projectors.

(iv) Fluorescence Methods: Some materials respond to absorbed photons by re-emitting other photons at different wavelengths, a phenomena known as fluorescence While very few materials fluoresce in the narrow range (< 1 octave!) of visible wavelengths, hyperspectral imaging reveals instructive fluorescence phenomena occur over much wider bands of wavelengths. Many organic chemicals have strongly varied fluorescent responses to ultraviolet light, and some living tissues can be chemically or genetically tagged with fluorescent markers that reveal important biological processes. Accordingly, hyperspectral imaging and illuminants can directly reveal chemical or biological features that may be further improved by 4D methods.

7.5 Fantasy Configurations

Beyond what we can do now, what would we like to achieve in computational photography? Freed from practical limits, a few fantasy devices come to mind. If the goal of photography is to capture the visual essence of an object in front of us, then perhaps the ideal photography studio is not a room full of lights and box-like cameras at all, but a flexible cloth we can rub gently over the surface of the object itself. The cloth would hold microscopic, interleaved video projectors and video cameras. It would emit hyperspectrally colorful patterns of light in all possible directions from all possible points on the cloth (a flexible 4D light source), while simultaneously making coordinated hyperspectral measurements in all possible directions from all possible points on the cloth (a flexible 4D camera). Wiping the cloth over a surface would illuminate and photograph inside even the tiniest crack or vent hole of the object, banishing occlusion from the data set; a quick wipe would characterize any rigid object thoroughly.

Suppose we wish to capture the appearance of a soft object, without touching it? Then perhaps a notebook-like device made of two plates hinged together would help. Each panel would consist of interleaved cameras and projectors in a sheet-like arrangement; simply placing it around the object would provide sufficient optical coupling between the embedded 4D illuminators and 4D cameras to assess the object thoroughly.

Yet even these are not the whole answer. If the goal of photography is to capture, reproduce, and manipulate a meaningful visual experience, then the ‘camera cloth’ is not sufficient to capture even the most rudimentary birthday party. The human experience and our personal viewpoint is missing. Ted Adelson suggested ‘camera wallpaper’ or the ‘balloon camera’, ubiquitous sensors that would enable us to compute arbitrary viewpoints at arbitrary times. Thad Starner and other ‘cybernauts’ who began personally instrumenting themselves in the 1990s have experimented with ‘always-on’ video cameras, and projects at Microsoft and the MIT Media Lab have explored gathering ‘video memories’ of every waking moment. So called ‘smart dust’ sensors and other unstructured ubiquitous sensors might gather views, sounds, and appearance from anywhere in a large city. What makes these moments special? What parts of this video will become keepsakes or evidence? How do we find what we care about in this flood of video? Computational Photography can supply us with visual experiences, but can't decide which one’s matter most to humans.

Bibliography

Fusion of Images Taken by Varying Camera and Scene Parameters

General

AKERS, D., LOSASSO, F., KLINGNER, J., AGRAWALA, M., RICK, J., AND HANRAHAN, P. 2003. Conveying shape and features with image-based relighting. In IEEE Visualization, 349–354.

BURT, P., AND KOLCZYNSKI, R. 1993. Enhanced image capture through fusion. In International Conference on Computer Vision (ICCV 93), 173–182.

LEVIN, A., ZOMET, A., PELEG, S., AND WEISS, Y. 2004. Seamless image stitching in the gradient domain. In European Conference on Computer Vision (ECCV 04).

MASSEY, M., AND BENDER, W. 1996. Salient stills: Process and practice. IBM Systems Journal 35, 3&4, 557–574.

MUTTER, S., AND KRAUSE, M. 1992. Surrational Images: Photomontages. University of Illinois Press.

ROBINSON, H. P. 1869. Pictorial Effect in Photography: Being Hints on Composition and Chiaroscuro for Photographers. Piper & Carter.

MUYBRIDGE, E. 1955. The human figure in motion. Dover Publications, Inc.

AGARWALA, A., DONTCHEVA, AGRAWALA, M., DRUCKER, COLBURN, CURLESS, SALESIN AND COHEN, M. Interactive Digital Photomontage. ACM Transactions on Graphics (Proceedings of SIGGRAPH 2004), 2004.

S.K. Nayar and S.G. Narasimhan , "Assorted Pixels: Multi-Sampled Imaging With Structural Models," Europian Conference on Computer Vision (ECCV), Vol.IV, pp.636-652, May, 2002.

Time

BRAUN, M. 1992. Picturing Time: The Work of Etienne-Jules Marey. The University of Chicago Press.

FREEMAN, W. T., AND ZHANG, H. 2003. Shape-time photography. In Conference on Computer Vision and Pattern Recognition (CVPR 03), 151–157.

Exposure

FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. ACM Transactions on Graphics 21, 3, 249–256.

REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. ACM Transactions on Graphics 21, 3, 267– 276.

DEBEVEC, AND MALIK. 1997. Recovering high dynamic range radiance maps from photographs. In Proc. SIGGRAPH.

DURAND, AND DORSEY. 2002. Fast bilateral filtering for the display of high-dynamic-range images. ACMTrans. on Graphics 21, 3.

MANN, AND PICARD. 1995. Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In Proc. IS&T 46th ann. conference.

TUMBLIN, AND TURK. 1999. LCIS: A boundary hierarchy for detail-preserving contrast reduction. In Proc. SIGGRAPH.

DICARLO, J., AND WANDELL, B. 2000. Rendering high dynamic range images. Proc. SPIE: Image Sensors 3965, 392–401.

Kang, S. B., Uyttendaele, M., Winder, S., and Szeliski, R. 2003. High dynamic range video. ACM Trans. Graph. 22, 3 (Jul. 2003), 319-325.

Focus

HAEBERLI, P. 1994. Grafica Obscura web site. http://www.sgi.com/grafica/.

MORGAN MCGUIRE, MATUSIK, PFISTER, HUGHES, AND DURAND, Defocus Video Matting, ACM Transactions on Graphics, Vol 24, No 3, July 2005, (Proceedings of ACM SIGGRAPH 2005).

Illumination

Frederik Anrys and Philip Dutr´e. Image based lighting design. In The 4th IASTED International Conference on Visualization, Imaging, and Image Processing, 2004.

David Akers, Frank Losasso, Jeff Klingner, Maneesh Agrawala, John Rick, and Pat Hanrahan. Conveying shape and features with image-based relighting. In IEEE Visualization, 2003.

Amit Agrawal, Ramesh Raskar, Shree K. Nayar, and Yuanzhen Li. Removing photography artifacts using gradient projection and flash-exposure sampling. In SIGGRAPH, pages 828–835, 2005.

Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In SIGGRAPH, pages 145–156, 2000.

Paul Debevec, AndreasWenger, Chris Tchou, Andrew Gardner, JamieWaese, and Tim Hawkins. A lighting reproduction approach to live-action compositing. In SIGGRAPH,pages 547–556, 2002.

Passive Illumination

RASKAR, R., ILIE, A., AND YU, J. 2004. Image fusion for context enhancement and video surrealism. In NPAR 2004: Third International Symposium on Non- Photorealistic Rendering.

WEISS, Y. 2001. Deriving intrinsic images from image sequences. In International Conference On Computer Vision (ICCV 01), 68–75.

Polorization

Y. Y. SCHECHNER, S. G. NARASIMHAN and S. K. NAYAR, Instant Dehazing of Images Using Polarization, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, December 2001.

S. K. NAYAR, X. FANG, and T. E. BOULT, Removal of Specularities using Color and Polarization, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.

Wu, P. Walsh, J.T. “Tissue Polarization Imaging of Multiply-Scattered Reflected Light Lasers in Surgery and Medicine, Supplement 15, 2003.

Wavelength

D. A. SOCOLINSKY, “Dynamic range constraints in image fusion and realization.” Proc. IASTED Int. Conf. Signal and
Image Process, 349-354 (2000).

Y. Y. SCHECHNER and S. K. NAYAR , Uncontrolled Modulation Imaging, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Washington DC, June 2004.

Location

Aseem Agarwala, Maneesh Agrawala, Michael Cohen, David Salesin, Rick Szeliski. ``Photographing long scenes with multi-viewpoint panoramas," ACM Transactions on Graphics (Proceedings of SIGGRAPH 2006), To appear, 2006.

B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, M. Levoy, High-Performance Imaging Using Large Camera Arrays.. ACM Transactions on Graphics, Vol 24, No 3, July 2005, pp 765-776 (Proceedings of ACM SIGGRAPH 2005).

Matting

CHUANG, Y.-Y., CURLESS, B., SALESIN, D., AND SZELISKI, R. 2001. A Bayesian approach to digital matting. In Proceedings of Computer Vision and Pattern Recognition (CVPR 2001), vol. II, 264 – 271.

PORTER, T., AND DUFF, T. 1984. Compositing digital images. In Computer Graphics (Proceedings of ACM SIGGRAPH 84), vol. 18, 253–259.

SMITH, A. R., AND BLINN, J. F. 1996. Blue screen matting. In Proceedings of ACM SIGGRAPH 96, 259–268.

Jian SUN, Jiaya JIA, Chi-Keung TANG and Heung-Yeung SHUM, Poisson Matting, ACM Transactions on Graphics, also in SIGGRAPH 2004, vol. 23, no. 3, July 2004, pages 315-321.

Techniques

General

DANIELSSON, P.-E. 1980. Euclidean distance mapping. Computer Graphics and Image Processing 14, 227–248.

LUCAS, B. D., AND KANADE, T. 1981. An iterative image registration technique with an application to stereo vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI ’81), 674–679.

MORTENSEN, E. N., AND BARRETT, W. A. 1995. Intelligent scissors for image composition. In Proceedings of SIGGRAPH 95, Computer Graphics Proceedings, Annual Conference Series, 191–198.

Graph Cuts

BOYKOV, Y., VEKSLER, O., AND ZABIH, R. 2001. Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 11, 1222–1239.

KWATRA, V., SCH ¨ODL, A., ESSA, I., TURK, G., AND BOBICK, A. 2003. Graphcut textures: Image and video synthesis using graph cuts. ACM Transactions on Graphics 22, 3, 277–286.

SHI, J., AND MALIK, J. Normalized Cuts and Image Segmentation. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), June 1997, Puerto Rico

Gradient Domain

PEREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. ACM Transactions on Graphics 22, 3, 313–318.

T. Georgiev. Covariant Derivatives and Vision. Proceedings of ECCV 2006.

Smoothing, Bilateral and Trilateral Filter

C. TOMASI, AND R. MANDUCHI, Bilateral Filtering of gray and colored images, Proc. IEEE Intl. Conference on Computer Vision, pp. 836-846, 1998.

CHOUDHURY, P., TUMBLIN, J., "The Trilateral Filter for High Contrast Images and Meshes", Proc. of the Eurographics Symposium on Rendering, Per. H. Christensen and Daniel Cohen eds., pp. 186-196, 2003

Video Textures

Schödl, A., Szeliski, R., Salesin, D. H., and Essa, I. 2000. Video textures. In Proceedings of the 27th Annual Conference on Computer Graphics and interactive Techniques International Conference on Computer Graphics and Interactive Techniques. ACM Press/Addison-Wesley Publishing Co., New York, NY, 489-498.

Agarwala, A., Zheng, K. C., Pal, C., Agrawala, M., Cohen, M., Curless, B., Salesin, D., and Szeliski, R. 2005. Panoramic video textures. ACM Trans. Graph. 24, 3 (Jul. 2005), 821-827.

Feature Extraction and Scene Understanding

Shape/Material/Illumination, Surface normals

BASRI, R. JACOBS, D. Photometric stereo with general, unknown lighting, Computer Vision and Pattern Recognition, 2001

B. K. P. HORN, "Shape from shading: A method for obtaining the shape of a smooth opaque object from one view," MIT Project MAC Int. Rep. TR-79 and MIT AI Lab, Tech. Rep. 232, Nov. 1970.

TODD ZICKLER, PETER N. BELHUMEUR, AND DAVID J. KRIEGMAN, "Helmholtz Stereopsis: Exploiting Reciprocity for Surface Reconstruction." Proc. 7th European Conference on Computer Vision, May 2002. Vol. III, pp 869-884.

ANDREAS WENGER, A GARDNER, CHRIS TCHOU, J UNGER, T HAWKINS, P DEBEVEC, Postproduction Relighting and Reflectance Transformation With Time-Multiplexed Illumination, SIGGRAPH 2005

Depth edges

Ramesh RASKAR , Karhan TAN, Rogerio FERIS, Jingyi YU, Matthew TURK, Non-photorealistic Camera: Depth Edge Detection and Stylized Rendering Using a Multi-Flash Camera, SIGGRAPH 2004

Depth

S. K. NAYAR and Y. NAKAGAWA,, Shape from Focus, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No. 8, pp. 824-831,

Motion

M. Ben-Ezra and S.K. Nayar, "Motion-based Motion Deblurring,", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.26, No.6, pp.689-698, Jun, 2004.

Raskar, R., Agrawal, A., and Tumblin, J. 2006. Coded exposure photography: motion deblurring using fluttered shutter. ACM Trans. Graph. 25, 3 (Jul. 2006), 795-804.

Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, William T. Freeman.,"Removing Camera Shake from a Single Photograph", ACM Trans. on Graphics (Proc. SIGGRAPH 2006).

Liu, C., Torralba, A., Freeman, W. T., Durand, F., and Adelson, E. H. 2005. Motion magnification. ACM Trans. Graph. 24, 3 (Jul. 2005), 519-526.

Transfer and denoising

Flash to no-flash

Elmar EISEMANN and Frédo DURAND, Flash Photography Enhancement Via Intrinsic Relighting, SIGGRAPH 2004

Georg PETSCHNIGG, Maneesh AGRAWALA, Hugues HOPPE, Richard SZELISKI, Michael COHEN, Kentaro TOYAMA. Digital Photography with Flash and No-Flash Image Pairs.ACM Transactions on Graphics (Proceedings of SIGGRAPH 2004), 2004.

DICARLO, J. M., XIAO, F., AND WANDELL, B. A. 2001. Illuminating illumination. In 9th Color Imaging Conference, 27–34.

Noise

P. MEER, J. JOLION, AND A. ROSENFELD, "A Fast Parallel Algorithm For Blind Estimation of Noise Variance," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 2, pp. 216-223, 1990.

Eric P. BENNETT and Leonard McMILLAN "Video Enhancement Using Per-Pixel Virtual Exposures" SIGGRAPH 2005

Geometric Operations

Panorama

DAVIS, J. 1998. Mosaics of scenes with moving objects. In Computer Vision and Pattern Recognition (CVPR 98), 354–360.

UYTTENDAELE, M., EDEN, A., AND SZELISKI, R. 2001. Eliminating ghosting and exposure artifacts in image mosaics. In Conference on Computer Vision and Pattern
Recognition (CVPR 01), 509–516.

SZELISKI, R., AND SHUM, H.-Y. 1997. Creating full view panoramic mosaics and environment maps. In Proceedings of SIGGRAPH 97, Computer Graphics Proceedings,
Annual Conference Series, 251–258.

Synthetic Aperture and Multi-views

E. H. Adelson and J. Y.A. Wang, “Single Lens Stereo with a Plenoptic Camera” in IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, Feb 1992.

Marc LEVOY, Billy CHEN, Vaibhav VAISH, Mark HOROWITZ, Ian MCDOWALL, Mark BOLAS, Synthetic Aperture Confocal Imaging. ACM SIGGRAPH 2004.

REN NG, Fourier Slice Photography, SIGGRAPH 2005

Georgiev, T., Zheng, C., Nayar, Sh., Curless, B., Salasin, D., Intwala, Ch., Spatio-angular Resolution Trade-offs in Integral Photography, EGSR 2006

A. STERN and B. JAVIDI, "3-D computational synthetic aperture integral imaging (COMPSAII)," Opt. Express 11, 2446-2451 (2003),

C. OLIVER and S. QUEGAN, Understanding Synthetic Aperture Radar Images. London: Artech House, 1998.

Vaibhav Vaish, Gaurav Garg, Eino-Ville (Eddy) Talvala, Emilio Antunez, Bennett Wilburn, Mark Horowitz, Marc Levoy, “Synthetic Aperture Focusing using a Shear-Warp Factorization of the Viewing Transform”, Proc. Workshop on Advanced 3D Imaging for Safety and Security (A3DISS) 2005 (in conjunction with CVPR 2005)

Deblurring and Superresolution

M. BEN-EZRA AND S. K. NAYAR , Motion Deblurring using Hybrid Imaging, In Proc. IEEE Computer Vision and Pattern Recognition (CVPR), Wisconsin, June 2003.

ZHOUCHEN LIN, HEUNG-YEUNG SHUM Fundamental Limits of Reconstruction-Based Superresolution Algorithms under Local Translation PAMI, January 2004 - (Vol. 26, No. 1) pp. 83-9

O. LANDOLT, A. MITROS, AND C. KOCH, “Visual Sensor with Resolution Enhancement by Mechanical Vibrations,” Proc. 2001 Conf. Advanced Research in VLSI, pp. 249-264, 2001.

Dynamic Range

A. Morimura. Imaging method for a wide dynamic range and an imaging device for a wide dynamic range. U.S. Patent 5455621, October 1993.

P. Burt and R. J. Kolczynski. Enhanced Image Capture Through Fusion. Proc. Of International Conference on Computer Vision (ICCV), pages 173-182, 1993.

B. Madden. Extended Intensity Range Imaging. Technical Report MS-CIS-93-96, Grasp Laboratory, University of Pennsylvania, 1993.

Y. T. Tsai. Method and apparatus for extending the dynamic range of an electronic imaging system. U.S. Patent 5309243, May 1994.

T. Mitsunaga and S. K. Nayar. Radiometric self calibration. In Proc CVPR, volume 2, pages 374--380, June 1999.

R. A. Street. High dynamic range segmented pixel sensor array. U.S. Patent 5789737, August 1998.

R. J. Handy. High dynamic range ccd detector imager. U.S. Patent 4623928, November 1986.

D. D.Wen. High dynamic range charge coupled device. U.S. Patent 4873561, October 1989.

M. Hamazaki. Driving method for solid-state image pickup device. Japanese Patent 08-331461, December 1996.

Knight, T. (1983) "Design of an Integrated Optical Sensor with On-Chip Pre-Processing," Ph.D. Thesis, Department of Electrical Engineering and Computer Science, MIT.

M. Sayag, "Non-linear Photosite Response in CCD Imagers." U.S Patent No. 5,055,667, 1991.

S. J. Decker, R. D. McGrath, K. Brehmer, and C. G. Sodini, "A 256x256 CMOS imaging array with wide dynamic range pixels and column-parallel digital output." IEEE J. of Solid State Circuits, Vol. 33, pp. 2081-2091, Dec. 1998

V. Brajovic and T. Kanade. A Sorting Image Sensor: An Example of Massively Parallel Intensity-to-Time Processing for Low-Latency Computational Sensors. Proc. of IEEE Conference on Robotics and Automation, pages 1638-1643, April 1996.

D. Scheffer S. Kavadias, B. Dierickx, A. Alaerts, D. Uwaerts, and J. Bogaerts. A Logarithmic Response CMOS Image Sensor with On-Chip Calibration. IEEE JSSC, 35(8):1146--52, August 2000

Shree K. Nayar, Tomoo Mitsunaga: High Dynamic Range Imaging: Spatially Varying Pixel Exposures. CVPR 2000: 1472-1479

Active Illumination

Edward H. Adelson and James R. Bergen. The plenoptic function and the elements of early vision. M. Landy and J. A. Movshon, (eds) Computational Models of Visual Processing, 1991.

Martin Fuchs, Volker Blanz, and Hans-Peter Seidel. Bayesian relighting. In Rendering Techniques, pages 157–164, 2005.

Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. The lumigraph. In SIGGRAPH, pages 43–54, 1996.

Paul Haeberli. Graphics obscura.

Tim Hawkins, Jonathan Cohen, and Paul Debevec. A photometric approach to digitizing cultural artifacts. In Proceedings of conference on Virtual reality, archeology, and cultural heritage, pages 333–342, 2001.

Tim Hawkins, Per Einarsson, and Paul E. Debevec. A dual light stage. In Rendering Techniques, pages 91–98, 2005.

Hugues Hoppe and Kentaro Toyama. Continuous flash. Technical Report 63, Microsoft Research, 2003.

Andrew Jones, Andrew Gardner, Mark Bolas, Ian McDowall, and Paul Debevec. Performance geometry capture for spatially varying relighting. SIGGRAPH Sketch, 2005.

Feng Xiao Jeffrey M. DiCarlo and Brian A Wandell. Illuminating illumination. In Ninth Color Imaging Conference, pages 27–34, 2001.

Melissa L Koudelka, Peter N Belhumeur, Sebastian Magda, and David J. Kriegman. Image-based modeling and rendering of surfaces with arbitrary brdfs. In IEEE CVPR, pages 568–575, 2001.

Marc Levoy and Pat Hanrahan. Light field rendering. In SIGGRAPH, pages 31–42, 1996.

Tom Malzbender, Dan Gelb, and Hans Wolters. Polynomial texture maps. In SIGGRAPH, pages 519–528. ACM Press, 2001.

Vincent Masselus, Pieter Peers, Philip Dutr´e, and Yves D. Willems. Relighting with 4d incident light fields. In SIGGRAPH, volume 22, pages 613–620, 2003.

Ankit Mohan, Jack Tumblin, Bobby Bodenheimer, Cindy Grimm, and Reynold J. Bailey. Table-top computed lighting for practical digital photography. In Rendering Techniques, pages 165–172, 2005.

Jeffry S. Nimeroff, Eero Simoncelli, and Julie Dorsey. Efficient re-rendering of naturally illuminated environments. In Proceedings of the Fifth Eurographics Workshop on Rendering, pages 359–373, 1994.

AndreasWenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. Performance relighting and reflectance transformation with time-multiplexed illumination. In SIGGRAPH, volume 24, pages 756–764, 2005.

Holger Winnem¨oeller, Ankit Mohan, Jack Tumblin, and Bruce Gooch. Light waving: Estimating light positions from photographs alone. Computer Graphics Forum, 24(3):to appear, 2005.

Smart, Unconventional Cameras

MEMS Technology

S. K. NAYAR, V. BRANZOI, AND T. BOULT. Programmable Imaging using a Digital Micromirror Array, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Washington DC, June 2004.

High Speed Imaging

B. WANDELL, P. CATRYSSE, J. DICARLO, D. YANG AND A. EL GAMAL Multiple Capture Single Image Architecture with a CMOS Sensor , In Proceedings of the International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives, pp. 11-17, Chiba, Japan, October 21-22 1999. (Society of Multispectral Imaging of Japan.)

S. KLEINFELDER, S.H. LIM, X.Q. LIU AND A. EL GAMAL A 10,000 Frames/s CMOS Digital Pixel Sensor, In IEEE Journal of Solid State Circuits, Vol.36, No.12, Pages 2049-2059, December 2001

X.Q. LIU AND ABBAS EL GAMAL, Synthesis of High Dynamic Range Motion Blur Free Image From Multiple Captures, In IEEE Transactions on circuits and systems (TCASI), VOL. 50, NO. 4, pp 530-539, APRIL 2003

Ramesh RASKAR, Amit AGRAWAL, Jack TUMBLIN, Coded Exposure Photography: Motion Deblurring using Fluttered Shutter, ACM SIGGRAPH 2006.

SHECHTMAN, E., CASPI, Y., AND IRANI, M. 2002. Increasing space-time resolution in video. In ECCV, Springer-Verlag, London, UK, 753–768.

Programmable SIMD

JOHANSSON, R., LINDGREN, L., MELANDER, J., AND MOLLER, B. 2003. A multi-resolution 100 gops 4 gpixels/s programmable cmos image sensor for machine vision. In Proceedings of the 2003 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors, IEEE.

Advanced, Programmable, Demodulating Cameras and Temporal Correlation

CANESTA Inc, 2004

PIXIM Inc, 2004

FOVEON Inc, 2004

JENOPTIK Inc, 2004

IVP Inc, Ranger Camera, 2004

F. XIAO, J. DICARLO, P. CATRYSSE AND B. WANDELL, Image Analysis using Modulated Light Sources, In Proceedings of the SPIE Electronic Imaging '2001 conference, Vol. 4306, San Jose, CA, January 2001.

ANDO, S., K. NAKAMURA, AND T. SAKAGUCHI. Ultrafast Correlation Image Sensor: Concept, Design, and Applications,. in Proc. IEEE CCD/AIS Workshop. 1997. Bruges, Belgium: IEEE.

ANDO, S. AND A. KIMACHI. Time-Domain Correlation Image Sensor: First CMOS Realization of Demodulator Pixels Array. in Proc. '99 IEEE CCD/AIS Workshop. 1999.

Others

Michael Cohen. “Capturing the Moment”, Symposium on Computational Photography and Video, Boston in May, 2005.

Pilu, Maurizio, “Casual Capture Project at HP labs”, http://www.hpl.hp.com/news/2004/jan-mar/casualcapture.html

TUMBLIN, J., AGRAWAL, A. AND RASKAR, R. Why I want a Gradient Camera, IEEE CVPR 2005

Optics

Jos Stam, "Diffraction Shaders", In SIGGRAPH 99 Conference Proceedings, Annual Conference Series, August 1999, 101-110.

Elber, G. 1994. Low cost illumination computation using an approximation of light wavefronts. In Proceedings of the 21st Annual Conference on Computer Graphics and interactive Techniques SIGGRAPH '94. ACM Press, New York, NY, 335-342.

Rosenfeld A. and Kak A.C., Digital Image Processing, New York; Academic Press, Cap 11, 1987.

J.W. Goodman, Introduction to Fourier Optics, McGraw-Hill Book Co.,New York, N.Y., (1968).

Dowski, E.R.,Jr., Cathey, W.T., Extended depth of field through wave-front coding, Applied Optics, Vol. 34, No. 11, 10 April 1995, pp. 1859-1866.

A. Zomet and S.K. Nayar, "Lensless Imaging with a Controllable Aperture," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun, 2006.

Zand, J., Coded aperture imaging in high energy astronomy, NASA Laboratory for High Energy Astrophysics (LHEA) at NASA's GSFC, 1996.

Light field optics

E. Hecht and A. Zajac. Optics. Addison-Wesley, 1979.

A. Gerrard and J. M. Burch. Introduction to Matrix Methods in Optics. Dover Publications, 1994.

V. Guillemin and S. Sternberg. Symplectic techniques in physics. Cambridge University Press.

Frédo Durand, Nicolas Holzschuch, Cyril Soler, Eric Chan & François Sillion. A Frequency analysis of Light Transport. Proceedings of SIGGRAPH 2005.

Todor Georgiev, Chintan Intwala. Integral View Photography. Adobe Tech Report 2006:

Todor Georgiev, Colin Zheng, Shree Nayar, David Salesin, Brian Curless, Chintan Intwala. Spatio-angular Resolution Trade-Offs in Integral Photography. In proceedings, EGSR 2006.
http://www.tgeorgiev.net/Spatioangular.pdf

Fourier optics

J. Goodman, Introduction to Fourier Optics, McGraw-Hill, 1968.

M. Born, E. Wolf, Principles of Optics, Pergamon Press, 1964.

Computational Photography

Organisers

Ramesh Raskar
Mitsubishi Electric Research Labs

Jack Tumblin
Northwestern University

Course WebPage :
http://www.merl.com/people/raskar/photo

Computational Photography

Course WebPage

http://www.merl.com/people/raskar/photo

Source Code, Slides, Bibliography, Links and Updates

Welcome

- Understanding Film-like Photography
  - Parameters, Nonlinearities, Ray-based concepts
- Image Processing and Reconstruction Tools
  - Multi-image Fusion, Gradient domain, Graph Cuts
- Improving Camera Performance
  - Better dynamic range, focus, frame rate, resolution
- Future Directions
  - HDR cameras, Gradient sensing, Smart optics/lighting

Goals

- Capture-time Techniques
  - Manipulating optics, illumination and sensors
- Fusion and Reconstruction
  - Beyond digital darkroom experience
- Improving Camera Performance
  - Better dynamic range, focus, frame rate, resolution
  - Hint of shape, reflectance, motion and illumination
  - Computational Imaging in Sciences
- Applications
  - Graphics, Special Effects, Scene Comprehension, Art

Speaker: Ramesh Raskar

Senior Research Scientist at MERL.

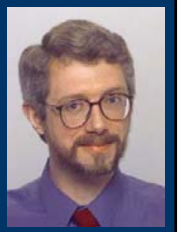
His research interests include projector-based graphics, computational photography and non-photorealistic rendering. He has published several articles on imaging and photography including multi-flash photography for depth edge detection, image fusion, gradient-domain imaging and projector-camera systems. His papers have appeared in SIGGRAPH, EuroGraphics, IEEE Visualization, CVPR and many other graphics and vision conferences. He was a course organizer at Siggraph 2002 through 2005. He was the panel organizer at the Symposium on Computational Photography and Video in Cambridge, MA in May 2005 and taught a graduate level class on Computational Photography at Northeastern University, Fall 2005. He is a member of the ACM and IEEE.

http://www.merl.com/people/raskar/raskar.html

Speaker: Jack Tumblin

Assistant Professor of Computer Science at Northwestern Univ.

His interests include novel photographic sensors to assist museum curators in historical preservation, computer graphics and visual appearance, and image-based modeling and rendering. During his doctoral studies at Georgia Tech and post-doc at Cornell, he investigated tone-mapping methods to depict high-contrast scenes. His MS in Electrical Engineering (December 1990) and BSEE (1978), also from Georgia Tech, bracketed his work as co-founder of IVEX Corp., (>45 people as of 1990) where his flight simulator design work was granted 5 US Patents. He was an Associate Editor of ACM Transactions on Graphics (2000-2006), a member of the SIGGRAPH Papers Committee (2003, 2004), and in 2001 was a Guest Editor of IEEE Computer Graphics and Applications.



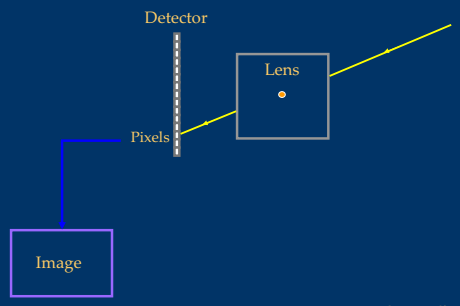
http://www.cs.northwestern.edu/~jet

Opportunities

- Unlocking Photography
  - How to expand camera capabilities
  - Digital photography that goes beyond film-like photography
- Think beyond post-capture image processing
  - Computation well before image processing and editing
- Learn how to build your own camera-toys
- Review of 30+ recent papers

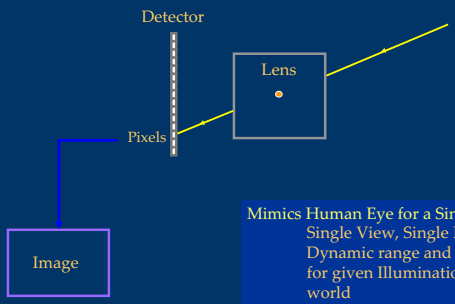
- What we will not cover
  - Film Cameras, Novel view rendering (IBR), Color issues, Traditional image processing/editing

Traditional Photography



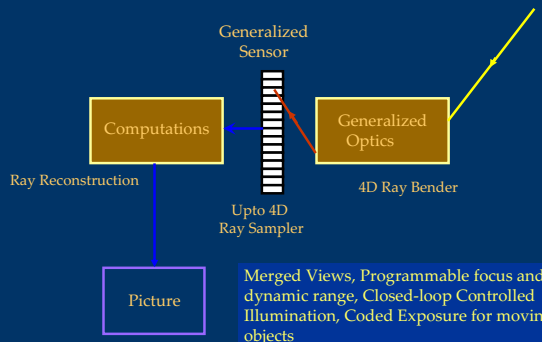
Courtesy: Shree Nayar

Traditional Photography



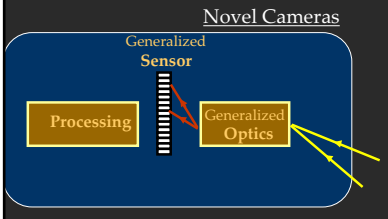
Mimics Human Eye for a Single Snapshot: Single View, Single Instant, Fixed Dynamic range and Depth of field for given Illumination in a Static world

Computational Photography: Optics, Sensors and Computations

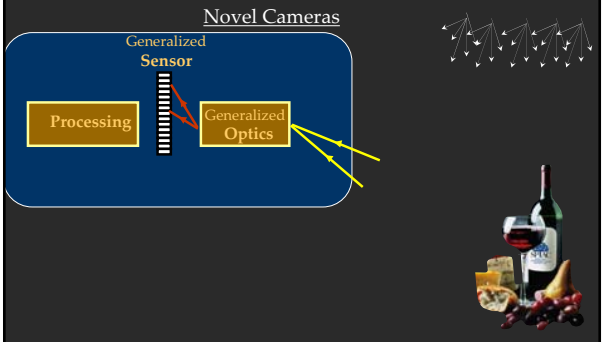


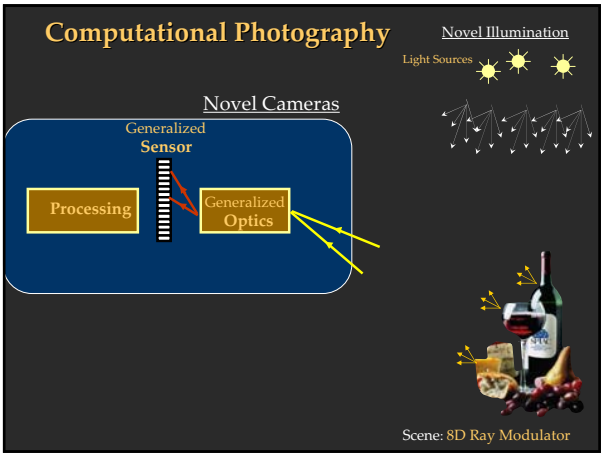
Merged Views, Programmable focus and dynamic range, Closed-loop Controlled Illumination, Coded Exposure for moving objects

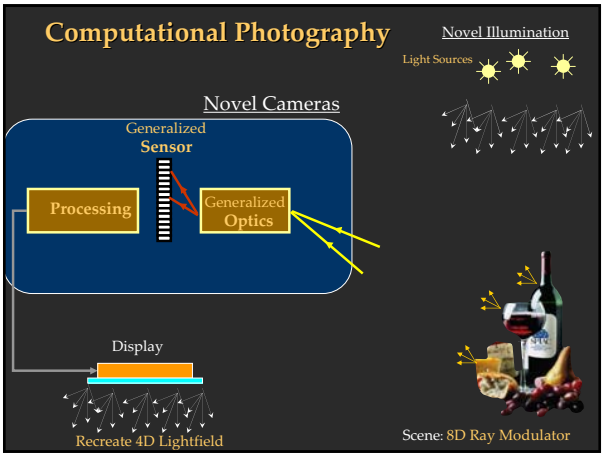
Computational Photography

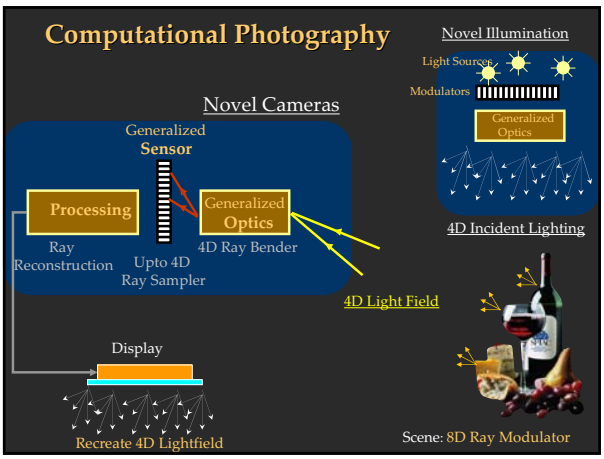


Computational Photography

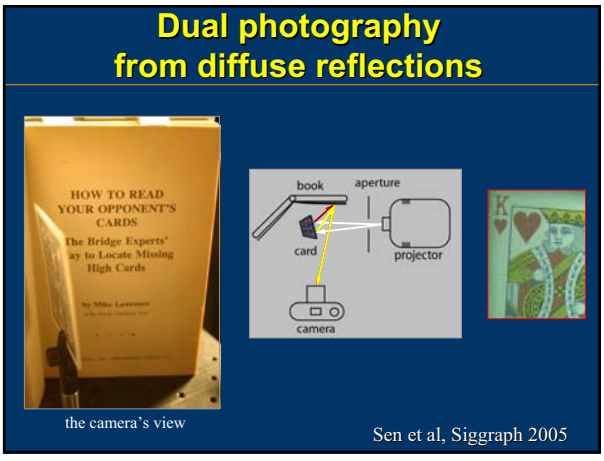


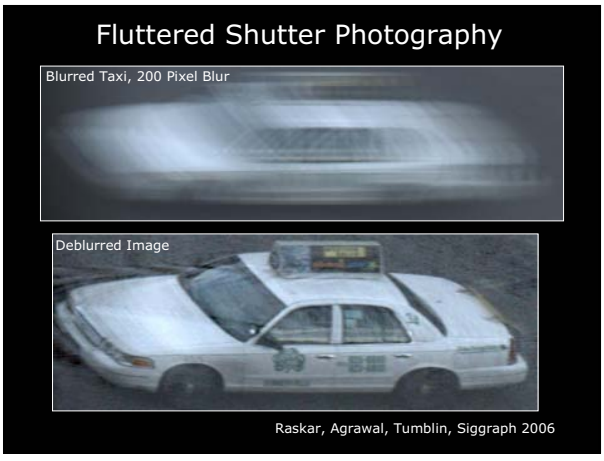

Computational Photography
Novel Illumination
Light Sources
Novel Cameras
Generalized
Sensor
Processing
Generalized
Optics
Scene: 8D Ray Modulator


Computational Photography
Novel Illumination
Light Sources
Novel Cameras
Generalized
Sensor
Processing
Generalized
Optics
Display
Recreate 4D Lightfield
Scene: 8D Ray Modulator


Computational Photography
Novel Illumination
Light Sources
Modulators
Generalized
Optics
4D Incident Lighting
Novel Cameras
Generalized
Sensor
Processing
Generalized
Optics
Ray
Reconstruction
Upto 4D
Ray Sampler
4D Ray Bender
4D Light Field
Display
Recreate 4D Lightfield
Scene: 8D Ray Modulator


Radial Stereoscopic Imaging
(Kuthirummal, Nayar SIGGRAPH 06)
© Shree Nayar, Columbia University


Dual photography
from diffuse reflections
HOW TO READ
YOUR OPPONENT'S
CARDS
book
aperture
card
projector
camera
the camera's view
Sen et al, Siggraph 2005


Fluttered Shutter Photography
Blurred Taxi, 200 Pixel Blur
Deblurred Image
Raskar, Agrawal, Tumblin, Siggraph 2006

Computational Photography

Mastering New Techniques for Lenses, Lighting and Sensors

- Ramesh Raskar and Jack Tumblin

- Book Publishers: A K Peters
- Siggraph 2006 booth: 20% off
- Coupons 25% Off

Computational Photography

Core Concepts

Jack Tumblin
Northwestern University

OUTLINE

- What is Photography?
- What is 'The Photographic Signal'?
- Perfecting Film-Like Photography:
  Old Problems, New Approaches
- Photography Beyond Film:
  New Goals, Methods, Expressions

'Film-Like' Photography

Film Camera design assumptions:

- 'Instantaneous' light measurement…
- Of focal plane image behind a lens.
- Reproduce those amounts of light.

Implied:

"What we see is ≅ focal-plane intensities."

well, no…we see *much* more!

(seeing is deeply cognitive)

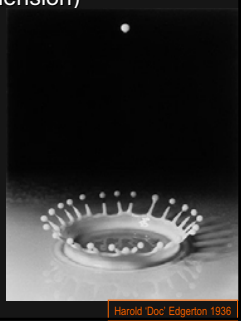
Our Definitions

- **'Film-like' Photography:**
  - Static 'instantaneous' record of the 2D image formed by a lens

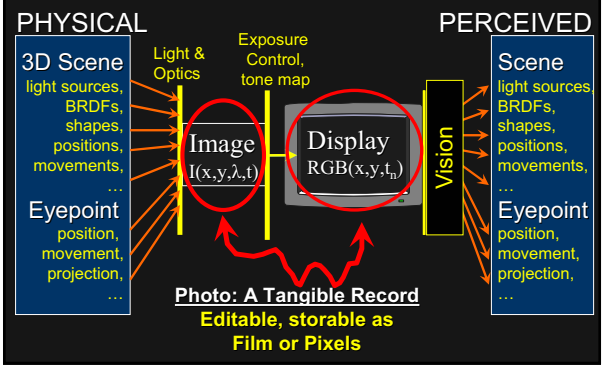
  Display image ≅ sensor image
- **'Computational' Photography:**
  - displayed image ≠ sensor image
  - A more expressive, controllable displayed result, from transformed, merged, decoded sensor data

What *is* Photography?

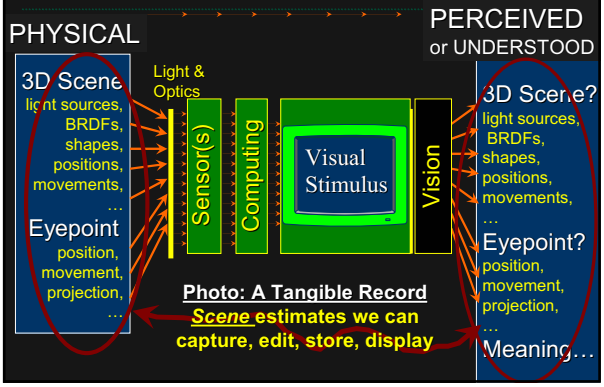
- A 'bucket' word: a neat container for messy notions (e.g. aviation, music, comprehension)
- A record of what we see, or would like to see, in tangible form.
- **Does 'film' photography always capture it?** no.
- **So, what do we see?**



What *is* Photography?

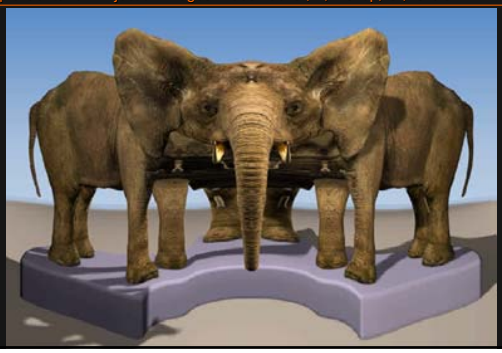


Ultimate Photographic Goals



Missing: Viewpoint Freedom

"Multiple-Center-of-Projection Images" Rademacher, P, Bishop, G., SIGGRAPH '98



Missing: Reliable Visual Boundaries

5 ray sets → explicit geometric occlusion boundaries

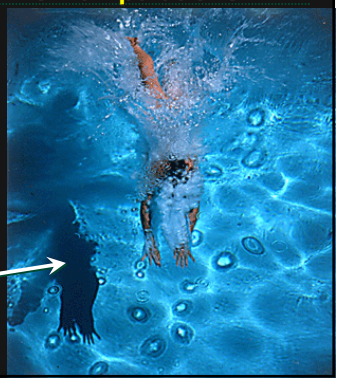
Ramesh Raskar, MERL, 2004

Missing: Expressive Time Manipulations

What other ways better reveal appearance to human viewers?

(Without direct shape measurement?)

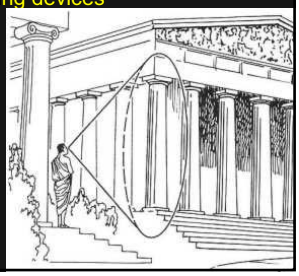
Can you understand this shape better?



Time for space wiggle. Gasparini, 1998.

Photographic Signal: ~~Pixels~~ Rays

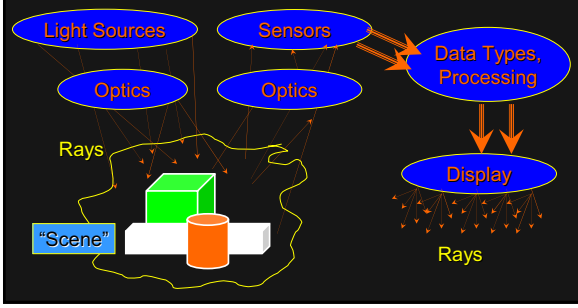
- Core ideas are ancient, simple, seem obvious:
  - **Lighting:** ray sources
  - **Optics:** ray bending/folding devices
  - **Sensor:** measure light
  - **Processing:** assess it
  - **Display:** reproduce it
- Ancient Greeks: 'eye rays' wipe the world to feel its contents…



http://www.mlahanas.de/Greeks/Optics.htm

The Photographic Signal Path

Computing can improve every component:



Review: How many Rays in a 3-D Scene?

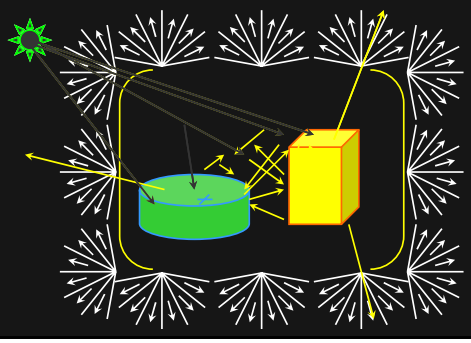
A 4-D set of infinitesimal members.

Imagine: (Levoy et al. SIGG'96) (Gortler et al. '96)

- Convex Enclosure of a 3D scene
- Inward-facing ray camera at every surface point
- Pick the rays you need for ANY camera outside.
- ~~2~~2D surface of cameras,
  2D ray set for each camera,
  → 4D set of rays.

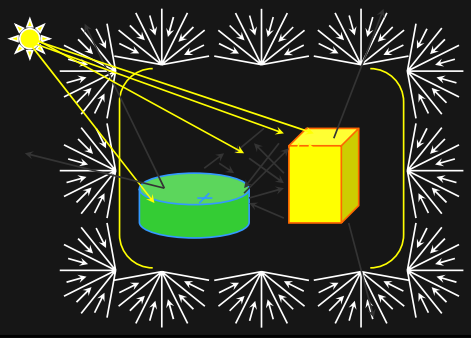
4-D Light Field / Lumigraph

Measure all the ***outgoing*** light rays.



4-D Illumination Field

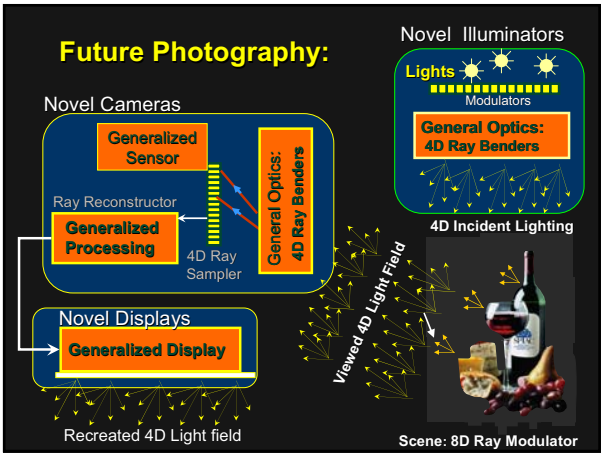
Same Idea: Measure all the ***incoming*** light rays



4D x 4D = 8-D Reflectance Field

Ratio: R_{ij} = (outgoing ray_i) / (incoming ray_j)

Future Photography:



Expand Optics Into Software

Programmable Optical Devices
enable new forms of:

[Nayar]

- **Omni-Directional Lens Systems**
  (Hi-Def 360° video…)
- **'Assorted Pixels' Sensors**
  (Robust HDR, multispectral…)
- **Lensless Adaptive-Aperture Cameras**
  (tracking without panning…)

Expand Scientific / Medical Imaging

4D light sources + 4D cameras
enable new forms of:

[Levoy]

- **Synthetic Aperture Imaging**
  (see through trees…)
- **Tomography**
  (3-D volumetric imaging…)
- **Confocal Scanning**
  (look inside muddy water…)

'The Ideal Photographic Signal'

I CLAIM IT IS:

~~All Rays? Some Rays?~~ ***Changes*** **in Some Rays**

Photographic ray space is vast and redundant
>8 dimensions: 4D view, 4D light, time, λ,

? Gather only 'visually significant' ray changes ?

? What rays should we measure ?
? How should we combine them ?
? How should we display them ?

Beyond 'Film-Like' Photography

Call it **'Computational Photography':**
To make 'meaningful ray changes' tangible,

- **Optics** can do more…
- **Sensors** can do more…
- **Light Sources** can do more…
- **Processing** can do more…

by applying low-cost storage,
computation, and control.

Computational Photography

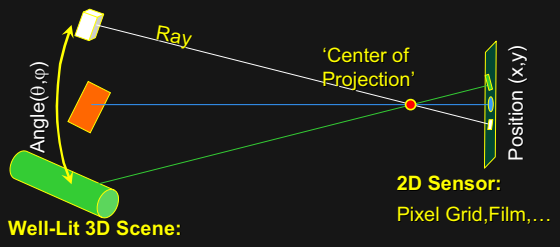
Understanding Film-Like Photography

or 'from 2D Pixels to 4D Rays'

Jack Tumblin
Northwestern University

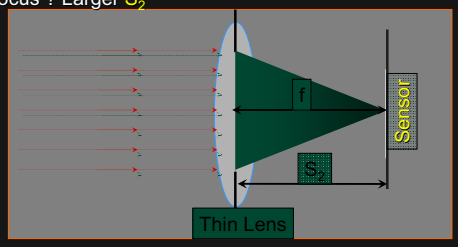
Naïve, Ideal Film-like Photography

Sensor: a film emulsion,
: or a grid of light meters (pixels)



Rays and the 'Thin Lens Law'

- Focal length f: where parallel rays converge
- Focus at infinity: Adjust for S_2=f
- Closer Focus ? Larger S_2

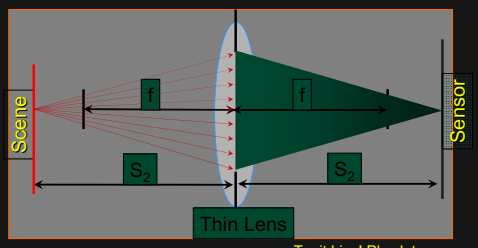


http://webphysics.davidson.edu/Applets/Optics/intro.html Try it Live! Physlets…

Rays and the 'Thin Lens Law'

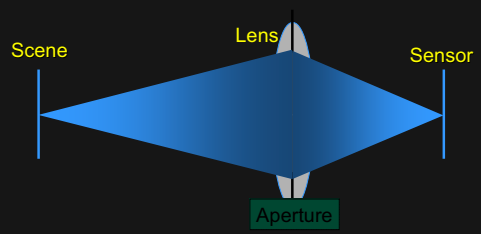
- Focal length f: where parallel rays converge
- Focus at infinity: Adjust for S_2=f
- Closer Focus ? Larger S_2

$$\frac{1}{S_1}+\frac{1}{S_2}=\frac{1}{f}$$



http://webphysics.davidson.edu/Applets/Optics/intro.html Try it Live! Physlets…

Not *One* Ray, but a *Bundle* of Rays

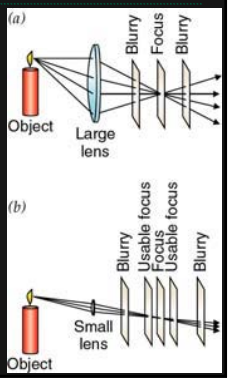


- BUT Ray model isn't perfect: ignores diffraction
- Lens, aperture set the point-spread-function (PSF)
  (How? See: Goodman,J.W. 'An Introduction to Fourier Optics' 1968)

Basic Ray Optics: Lens Aperture

For the *same* focal length:

- **Larger lens**
  - Gathers a wider ray bundle:
  - More light: brighter image
  - Narrower depth-of-focus
- **Smaller lens**
  - dimmer image
  - focus becomes less critical



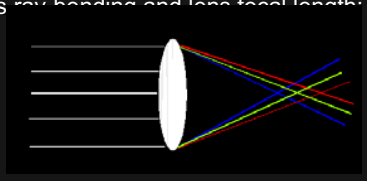
Film-like Optics: Thin Lens *Flaws*

- Aberrations:
  Real lenses don't converge rays perfectly
- **Spherical:** edge rays ≠ center rays
- **Coma:** diagonal rays focus deeper at edge

http://www.nationmaster.com/encyclopedia/Lens-(optics)

Lens Flaws: Chromatic Aberration

- Dispersion: wavelength-dependent refractive index
  - (enables prism to spread white light beam into rainbow)
- Modifies ray bending and lens focal length: f(λ)



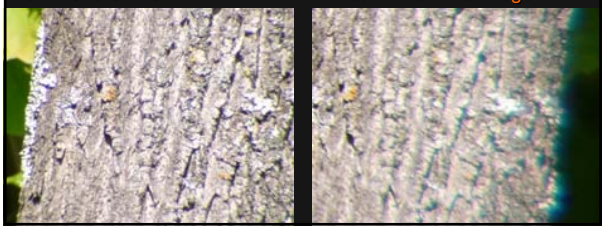
- color fringes near edges of image

http://www.swgc.mun.ca/physics/physlets/opticalbench.html

Chromatic Aberration

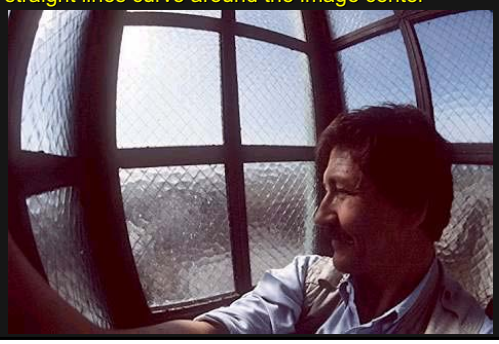
- Lens Design Fix: Multi-element lenses
  Complex, expensive, many tradeoffs!
- Computed Fix: Geometric warp for R,G,B.

Near Lens Center | Near Lens Outer Edge



Radial Distortion (*e.g.* 'Barrel' and 'pin-cushion')

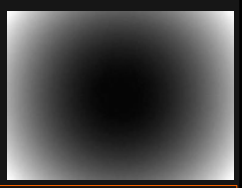
straight lines curve around the image center



Vignette Effects

Bright at center, dark at edges.
Several causes compounded:

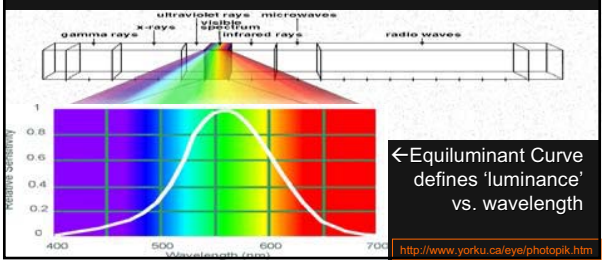
- Edge pixels span smaller angle and center pixels
- Ray path length is longer off-axis
- Internal shadowing
- Compensation:
  - Use anti-vignetting filters, (darkest at center)
  - OR Position-dependent pixel-detector sensitivity.



http://homepage.ntlworld.com/j.houghton/vignette.htm

Film-like Color Sensing

- Visible Light: narrow band of e'mag. spectrum
- λ ≈ **400-700 nm** (nm = 10-9 meter wavelength)
- (humans:<1 octave ←→honey bees: 3-4 'octaves
  do honey bees sense harmonics, see color 'chords' ?

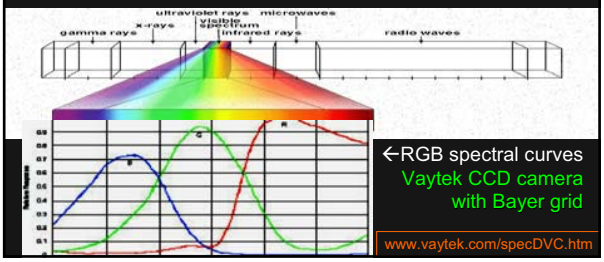


←Equiluminant Curve defines 'luminance' vs. wavelength

http://www.yorku.ca/eye/photopik.htm

Film-like Color Sensing

- Visible Light: narrow band of emag spectrum
- λ ≈ **400-700 nm** (nm = 10-9 meter wavelength)
- At least 3 spectral bands required (e.g. R,G,B)

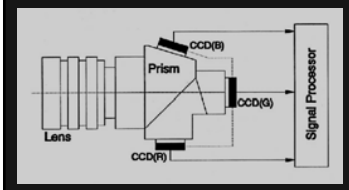


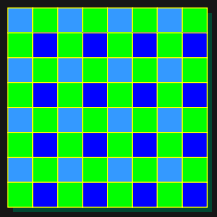
←RGB spectral curves
Vaytek CCD camera with Bayer grid

www.vaytek.com/specDVC.htm

Color Sensing

- 3-chip: vs. 1-chip: quality vs. cost



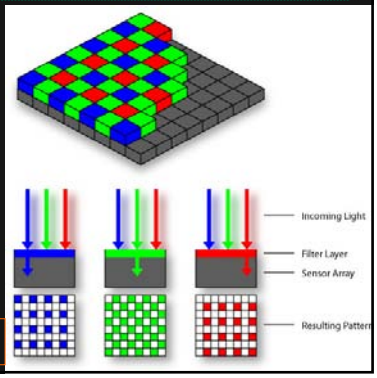


http://www.cooldictionary.com/words/Bayer-filter.wikipedia

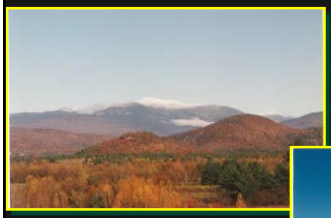
1-Chip Color Sensing: Bayer Grid

- Estimate RGB at 'G' cels from neighboring values

http://www.cooldictionary.com/words/Bayer-filter.wikipedia



Polarization



Sunlit haze is often strongly polarized. Polarization filter yields much richer sky colors

RAYS and PROCESSING

- ONE Ray carries doubly infinitesimal power:

  Ray bundles with finite, measurable power will:
  - Span a non-zero area
  - Fill a non-zero solid angle
- Everything is Linear: (HUGE win!)

  Ray reflectance, transmission, absorption, scatter*…
- Rays are REVERSIBLE. Helmholtz reciprocity

  Ray bundles? Not so much: falls quickly with angle,area growth…

Film-like Photography: Many Limitations

- Optics:

  Single focus distance, limited depth-of-field, limited field-of-view, internal reflections/flare/glare
- Lighting:

  Camera has no knowledge of ray source strength, position, direction; little control (e.g. flash)
- Sensor:

  Exposure setting, motion blur, noise, response time,
- Processing:
  - Quantization/color depth, camera shake, scene movement…

Conclusions

- Film-like photography methods limit digital photography to film-like results or less.
- Broaden, unlock our views of photography:
- 4-D, 8-D, even 10-D Ray Space holds the photographic signal. Look for new solutions by creating, gathering, processing RAYS, not focal-plane intensities.
- Choose the best, most expressive sets of rays,
  THEN find the best way to measure them.

Useful links:

Interactive Thin Lens Demo
(or search 'physlet optical bench')

www.swgc.mun.ca/physics/physlets/opticalbench.html

For more about color:

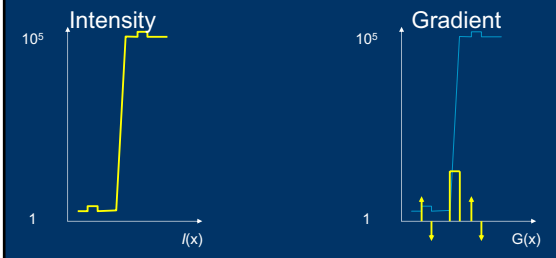
- Prev. SIGGRAPH courses (Stone et al.)
- Good: www.cs.rit.edu/~ncs/color/a_spectr.html
- Good: www.colourware.co.uk/cpfaq.htm
- Good: www.yorku.ca/eye/toc.htm

Image Processing and Reconstructions Tools

Image Tools

- Gradient domain operations,
  - Tone mapping, fusion and matting
- Graph cuts,
  - Segmentation and mosaicing
- Bilateral and Trilateral filters,
  - Denoising, image enhancement

Intensity Gradient in 1D

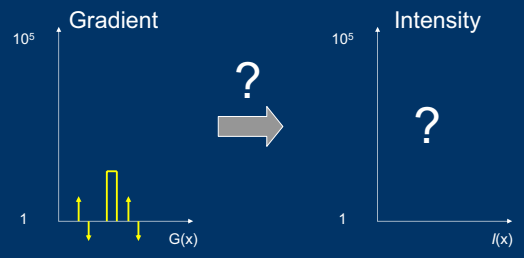


Gradient at x,

$G(x) = I(x+1) - I(x)$

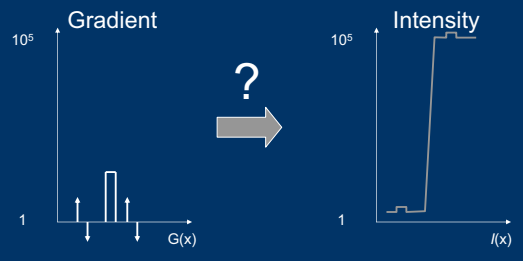
Forward Difference

Reconstruction from Gradients



For n intensity values, about n gradients

Reconstruction from Gradients



1D Integration

$I(x) = I(x-1) + G(x)$

Cumulative sum

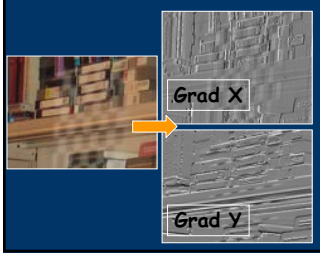
Intensity Gradient in 2D

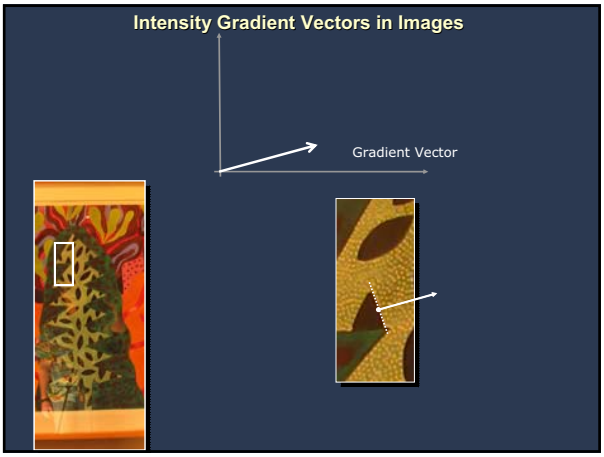
Gradient at x,y as Forward Differences

$G_x(x,y) = I(x+1\ ,\ y) - I(x,y)$

$G_y(x,y) = I(x\ ,\ y+1) - I(x,y)$

$G(x,y) = (G_x\ ,\ G_y)$




Intensity Gradient Vectors in Images
Gradient Vector

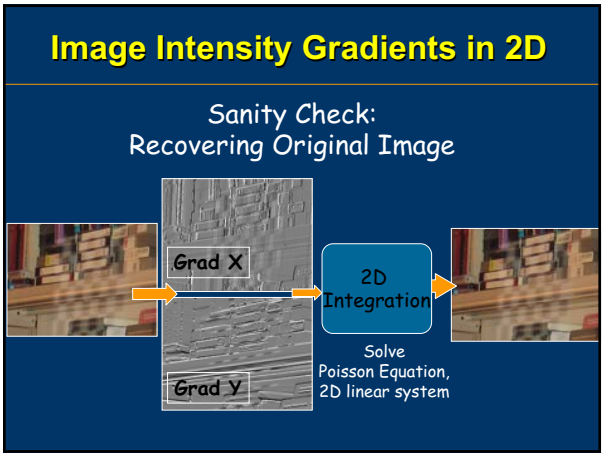
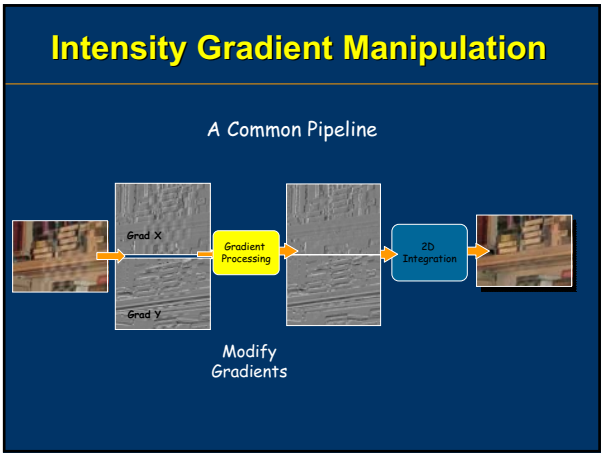
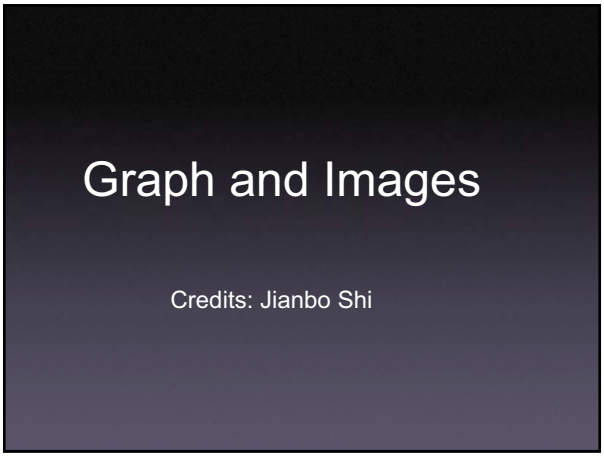
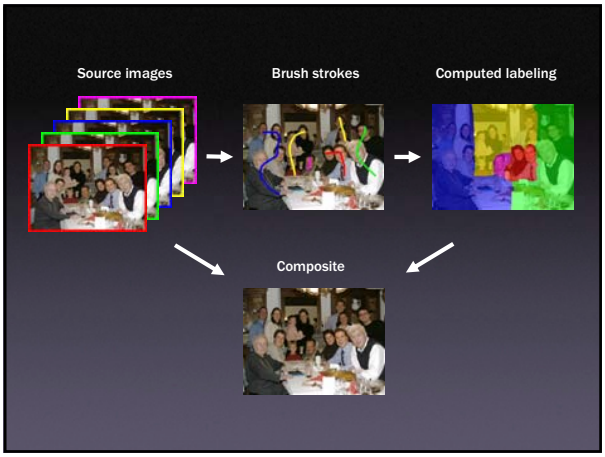


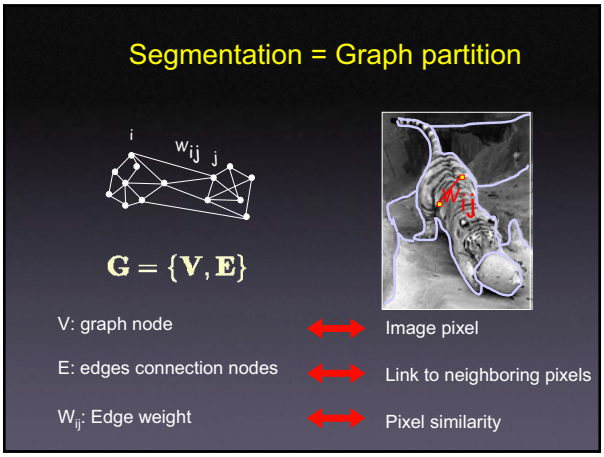
Image Intensity Gradients in 2D
Sanity Check:
Recovering Original Image
Grad X
Grad Y
2D Integration
Solve
Poisson Equation,
2D linear system

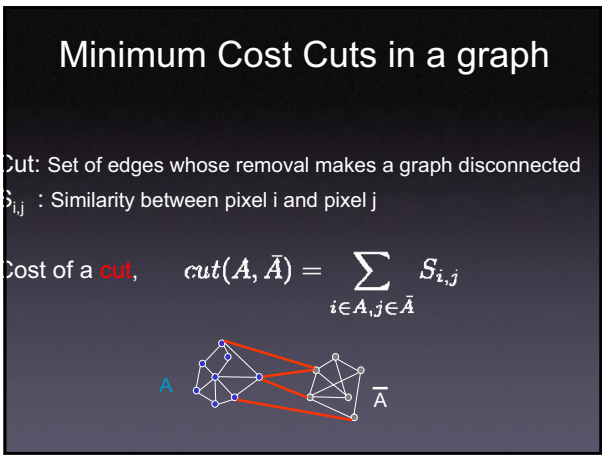

Intensity Gradient Manipulation
A Common Pipeline
Grad X
Grad Y
Gradient Processing
2D Integration
Modify
Gradients

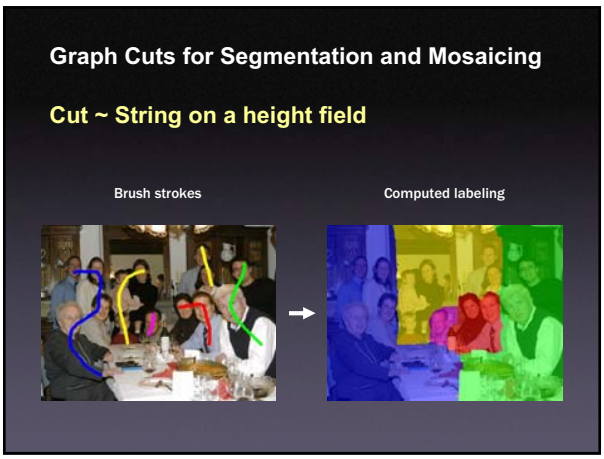

Graph and Images
Credits: Jianbo Shi

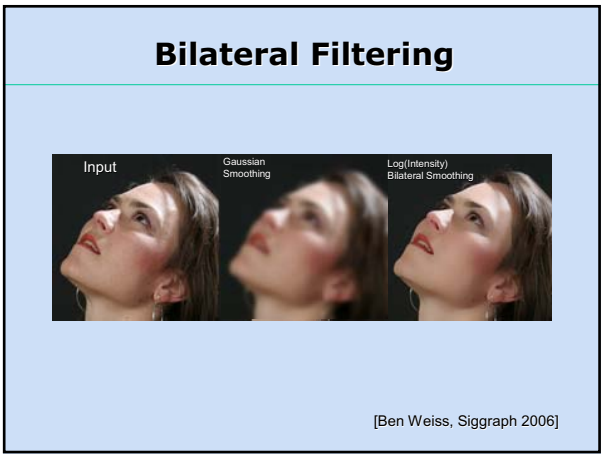

Agrawala et al, Digital Photomontage, Siggraph 2004

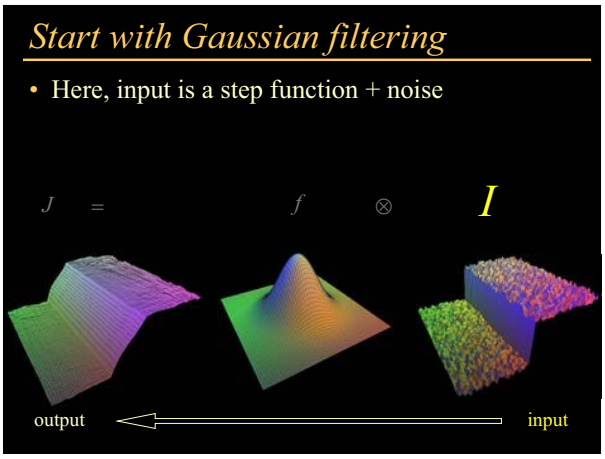

Source images
Brush strokes
Computed labeling
Composite


Segmentation = Graph partition
i
w_{ij}
j
$\mathbf{G} = \{\mathbf{V}, \mathbf{E}\}$
V: graph node
Image pixel
E: edges connection nodes
Link to neighboring pixels
W_{ij}: Edge weight
Pixel similarity


Minimum Cost Cuts in a graph
Cut: Set of edges whose removal makes a graph disconnected
$S_{i,j}$: Similarity between pixel i and pixel j
Cost of a cut, $cut(A, \bar{A}) = \sum_{i \in A, j \in \bar{A}} S_{i,j}$
A
$\bar{A}$


Graph Cuts for Segmentation and Mosaicing
Cut ~ String on a height field
Brush strokes
Computed labeling

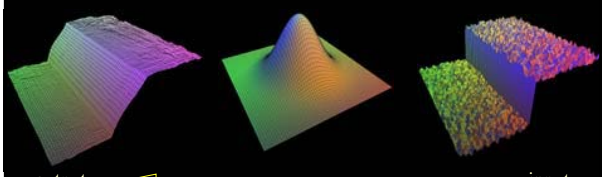

Bilateral Filtering
Input
Gaussian Smoothing
Log(Intensity) Bilateral Smoothing
[Ben Weiss, Siggraph 2006]


Start with Gaussian filtering
• Here, input is a step function + noise
J = f ⊗ I
output
input

Start with Gaussian filtering

- Spatial Gaussian f

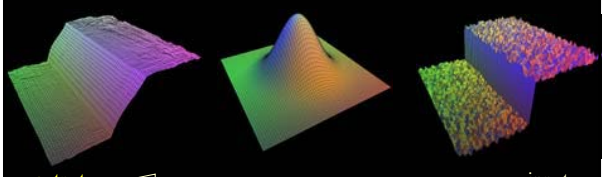
$J = f \otimes I$



Start with Gaussian filtering

- Output is blurred

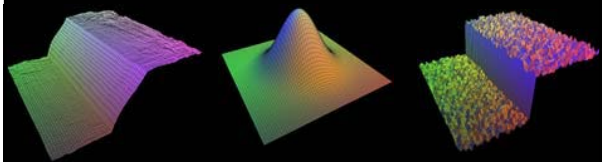
$J = f \otimes I$



Gaussian filter as weighted average

- Weight of ξ depends on distance to x

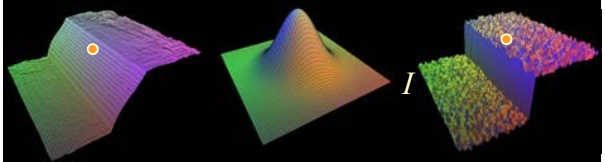
$$J(x) = \sum_{\xi} f(x,\xi) \; I(\xi)$$



The problem of edges

- Here, $I(\xi)$ "pollutes" our estimate J(x)
- It is too different

$$J(x) = \sum_{\xi} f(x,\xi) \; I(\xi)$$

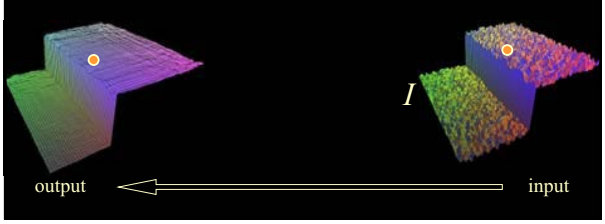


Principle of Bilateral filtering

[Tomasi and Manduchi 1998]

- Penalty g on the intensity difference

$$J(x) = \frac{1}{k(x)} \sum_{\xi} f(x,\xi) \; g(I(\xi)-I(x)) \; I(\xi)$$

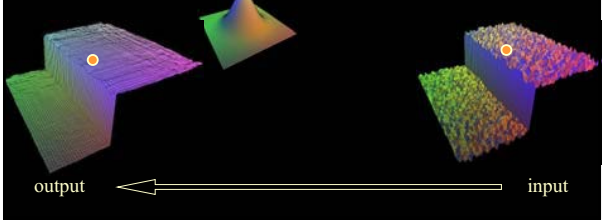


Bilateral filtering

[Tomasi and Manduchi 1998]

- Spatial Gaussian f

$$J(x) = \frac{1}{k(x)} \sum_{\xi} f(x,\xi) \; g(I(\xi)-I(x)) \; I(\xi)$$

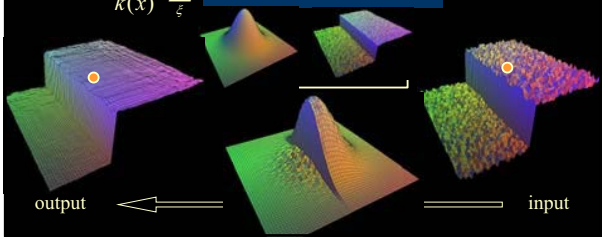


Bilateral filtering

[Tomasi and Manduchi 1998]

- Spatial Gaussian f
- Gaussian g on the intensity difference

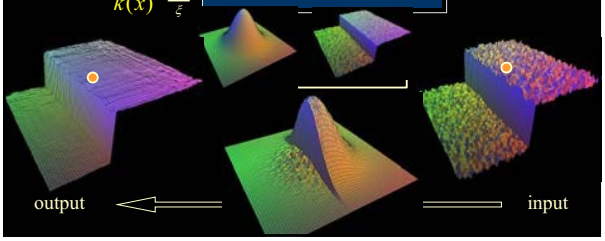
$$J(x) = \frac{1}{k(x)} \sum_{\xi} f(x,\xi) \quad g(I(\xi) - I(x)) \quad I(\xi)$$



[Tomasi and Manduchi 1998]

- The weights are different for each output pixel

$$J(x) = \frac{1}{k(x)} \sum_{\xi} f(x,\xi) \quad g(I(\xi) - I(x)) \quad I(\xi)$$

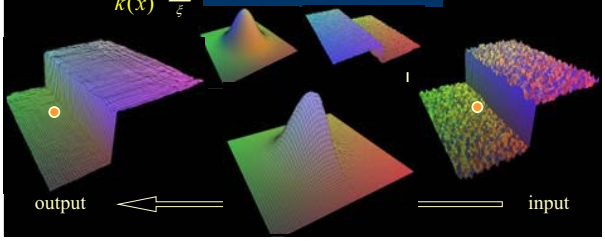


Bilateral filtering is non-linear

[Tomasi and Manduchi 1998]

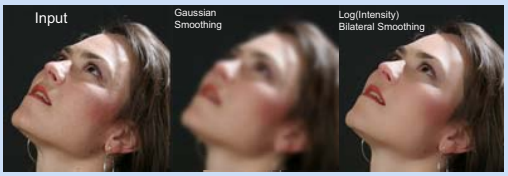
- The weights are different for each output pixel

$$J(x) = \frac{1}{k(x)} \sum_{\xi} f(x,\xi) \quad g(I(\xi) - I(x)) \quad I(\xi)$$



Bilateral Filtering

- Unilateral filtering
  - Smoothing using filtering
- Bilateral filtering
  - Edge-preserving smoothing



[Ben Weiss, Siggraph 2006]

Computational Photography

Image Fusion & Reconstruction

Photography: Full of Tradeoffs...

- **No Flash:** Candle warmth, but high noise
- **Flash:** low noise, but no candle warmth

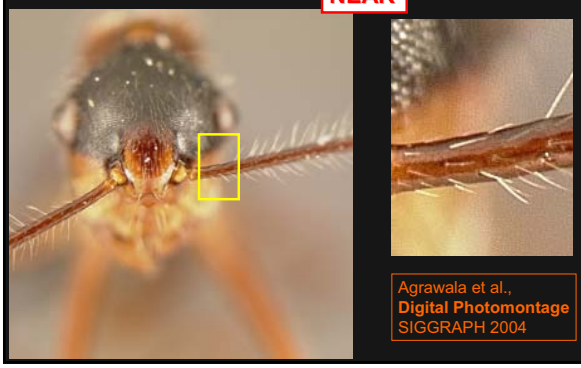
"...a-Hew Cain't, allwezz Gitt, Whut chew wawnt..." –1969,Jagger, M. et al

Image Fusion & Reconstruction

- Single photo: forces narrow tradeoffs:
  - Focus, Exposure, aperture, time, sensitivity, noise,
  - Usual result: Incomplete visual appearance.
- →Multiple photos, assorted settings for Optics, Sensor, Lighting, Processing
- **Fusion:**
  'Merge the best parts' or
- **Reconstruction:**
  Detect changes photo changes, compute scene invariants

FUSION: Best-Focus Distance

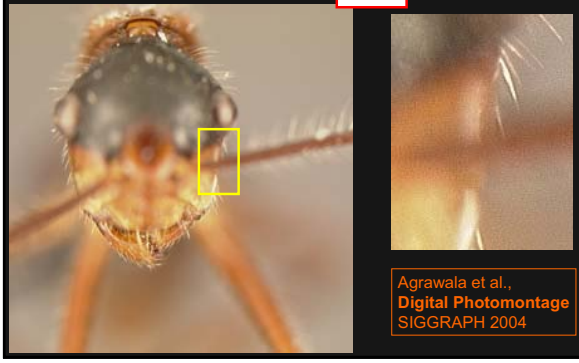
NEAR



Agrawala et al., **Digital Photomontage** SIGGRAPH 2004

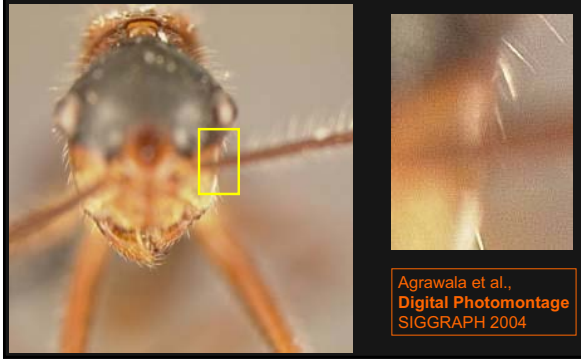
FUSION: Best-Focus Distance

FAR

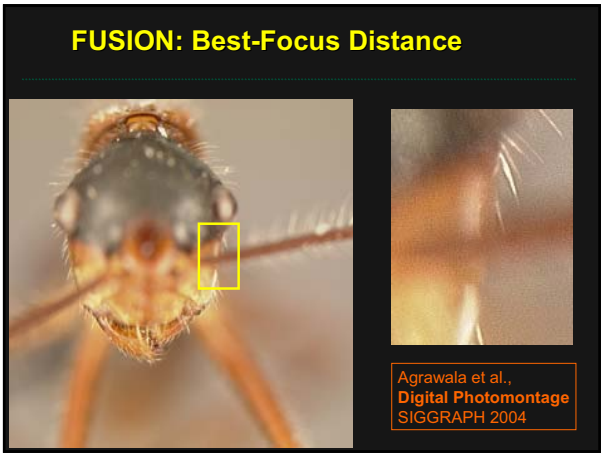


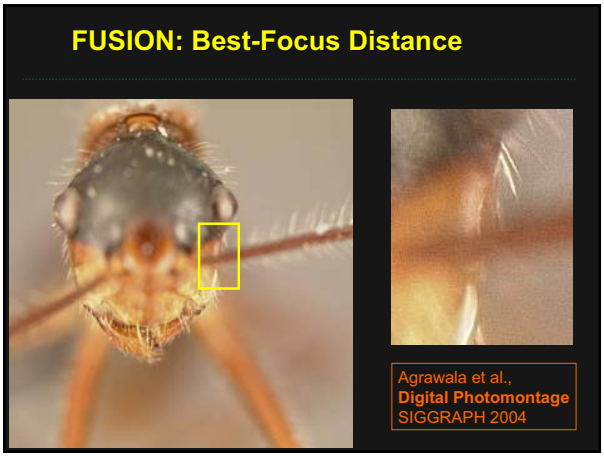
Agrawala et al., **Digital Photomontage** SIGGRAPH 2004

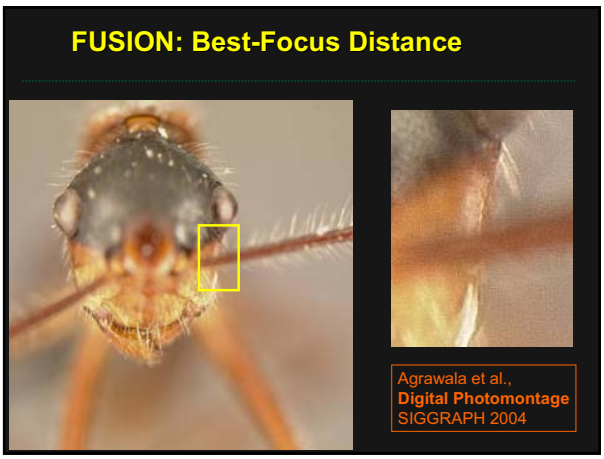
FUSION: Best-Focus Distance

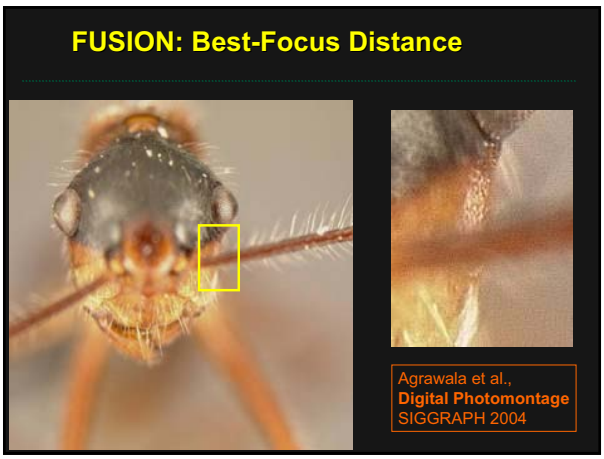


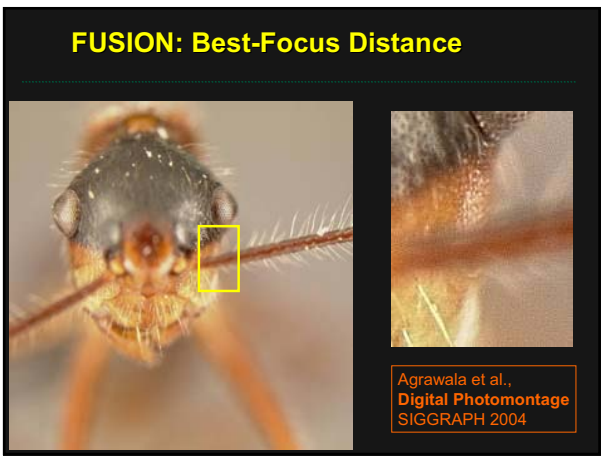
Agrawala et al., **Digital Photomontage** SIGGRAPH 2004

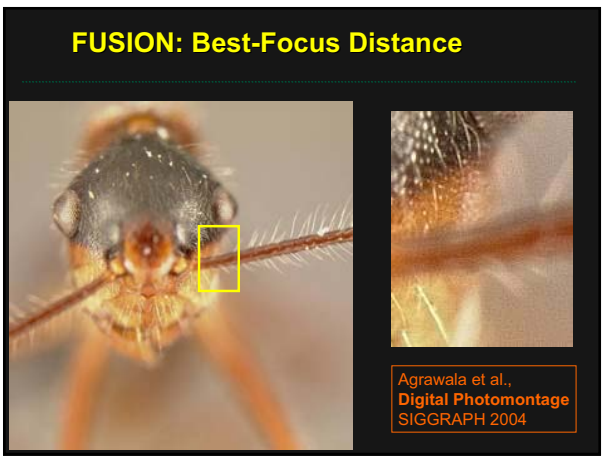

FUSION: Best-Focus Distance
Agrawala et al.,
Digital Photomontage
SIGGRAPH 2004

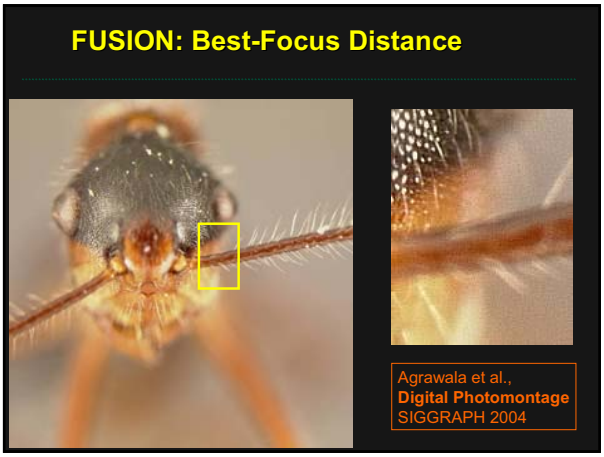

FUSION: Best-Focus Distance
Agrawala et al.,
Digital Photomontage
SIGGRAPH 2004

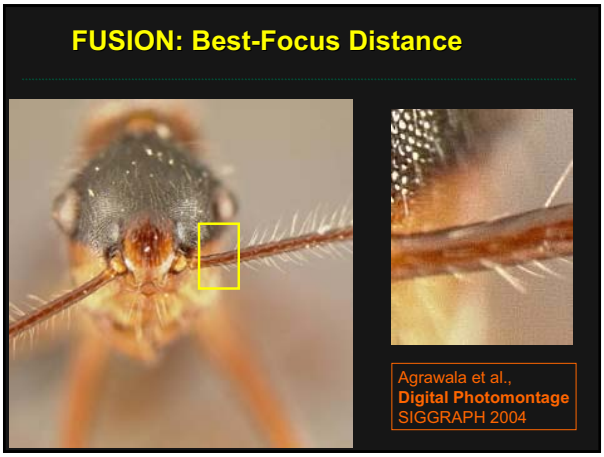

FUSION: Best-Focus Distance
Agrawala et al.,
Digital Photomontage
SIGGRAPH 2004

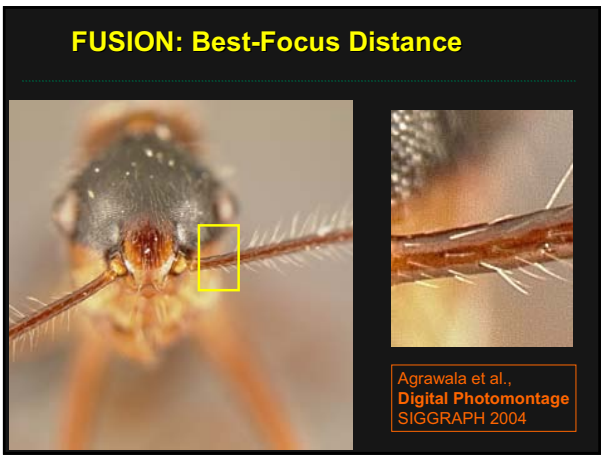

FUSION: Best-Focus Distance
Agrawala et al.,
Digital Photomontage
SIGGRAPH 2004

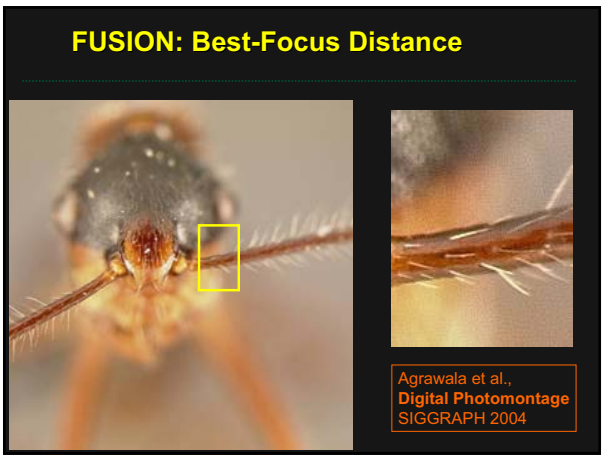

FUSION: Best-Focus Distance
Agrawala et al.,
Digital Photomontage
SIGGRAPH 2004

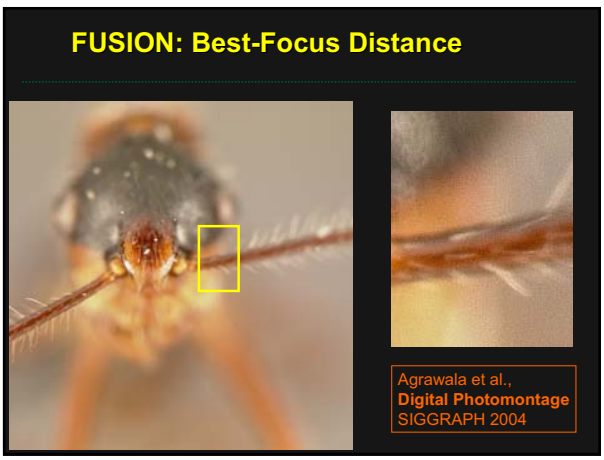

FUSION: Best-Focus Distance
Agrawala et al.,
Digital Photomontage
SIGGRAPH 2004

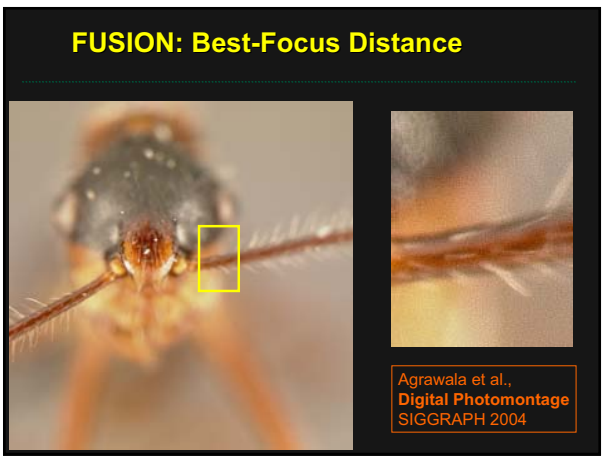

FUSION: Best-Focus Distance
Agrawala et al.,
Digital Photomontage
SIGGRAPH 2004

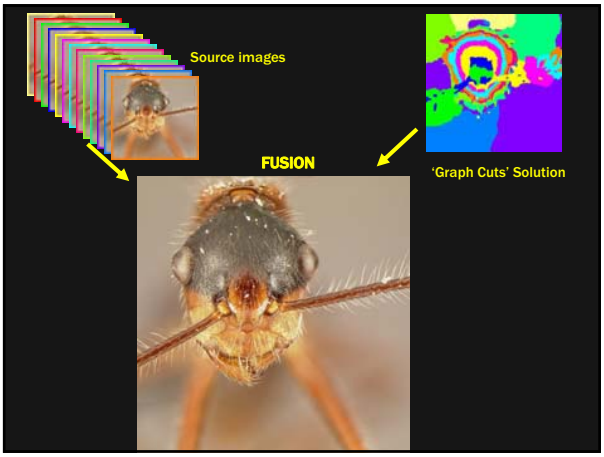

FUSION: Best-Focus Distance
Agrawala et al.,
Digital Photomontage
SIGGRAPH 2004


FUSION: Best-Focus Distance
Agrawala et al.,
Digital Photomontage
SIGGRAPH 2004


FUSION: Best-Focus Distance
Agrawala et al.,
Digital Photomontage
SIGGRAPH 2004


FUSION: Best-Focus Distance
Agrawala et al.,
Digital Photomontage
SIGGRAPH 2004


FUSION: Best-Focus Distance
Agrawala et al.,
Digital Photomontage
SIGGRAPH 2004



What else can we extend?

Film-Like Camera Parameters:

- Field of View: image stitching for panoramas
- Dynamic Range: Radiance Maps
- Frame Rate: Interleaved Video [Levoy]
- Resolution: 'Super-resolution' methods [Nayar]

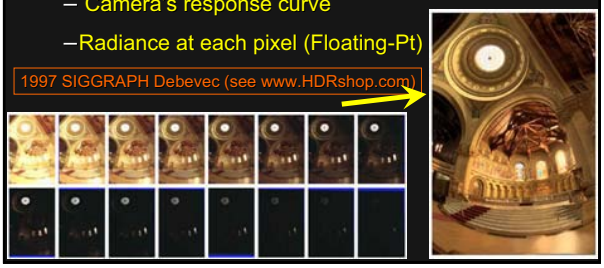
Visual Appearance & Content:

- Tone Map: Detail in every shadow and highlight
- Color2grey: Keep *all* color changes in grayscale
- Temporal Continuity: Space-time fusion
- Viewpoint Constraints: Multiple COP images and more…

High Dynamic Range Capture

- Series of Photos, progressive exposure time
- Solve least-squares matrix problem to get:
  - Camera's response curve
  - Radiance at each pixel (Floating-Pt)

1997 SIGGRAPH Debevec (see www.HDRshop.com)

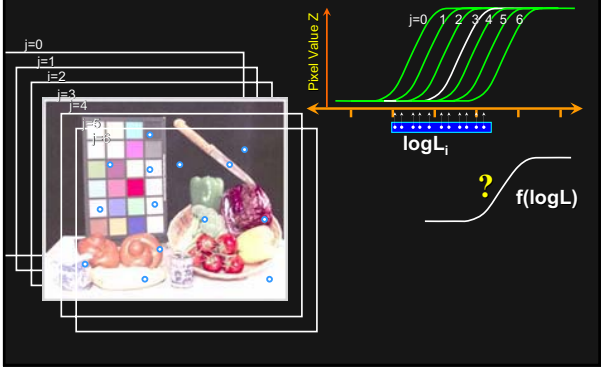


Film-Style Sensor: Dynamic Range Limits

Under-Exposure
- Highlight details: Captured
- Shadow details: **Lost**

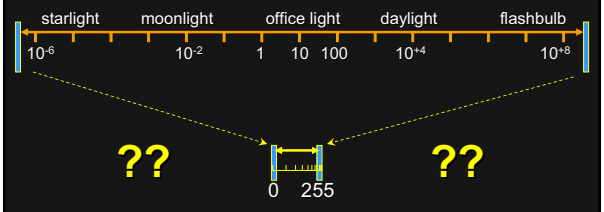
Over-Exposure
- Highlight details: **Lost**
- Shadow details: Captured

Debevec'97 (see www.HDRshop.com)



'Tone Map' Problem: HDR Scene ←→ Limited Displays

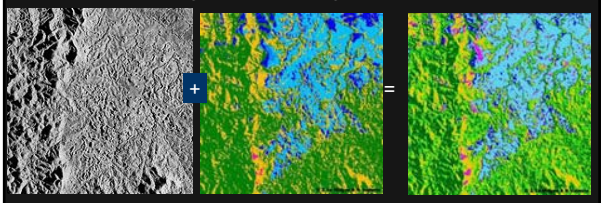
Domain of Human Vision:
from ~10^{-6} to ~10^{+8} cd/m^2



Range of Typical Displays:
from ~1 to ~100 cd/m^2

FUSION: Multispectral Wavelengths

Vegetation Mapping of the Forest

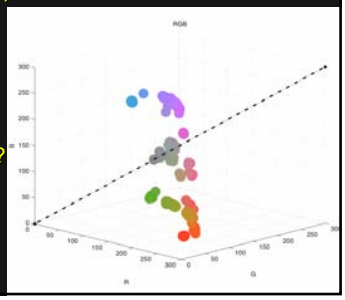


Color2Gray: Salience-Preserving Color Removal

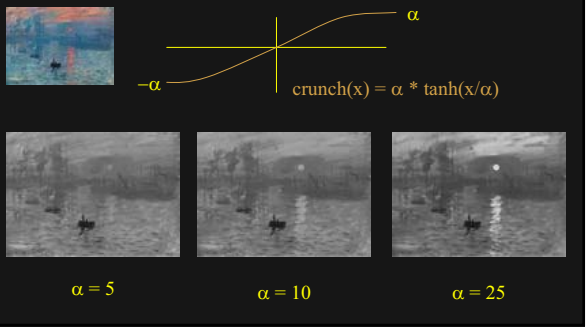
SIGGRAPH 2006
Gooch, Olsen, Tumblin, Gooch

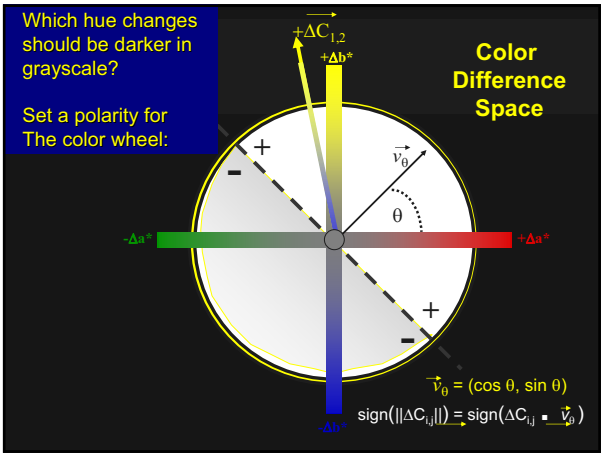
Iso-Luminant Colors

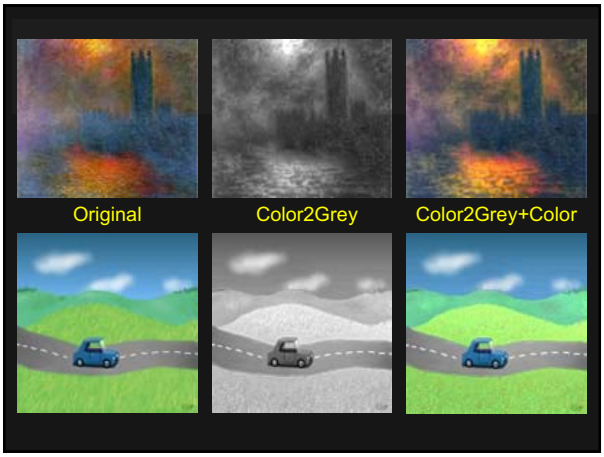
- Iso-luminant color changes are visually important;
- What transfer to luminance will
  - Keep them visible?
  - Preserve image appearance?

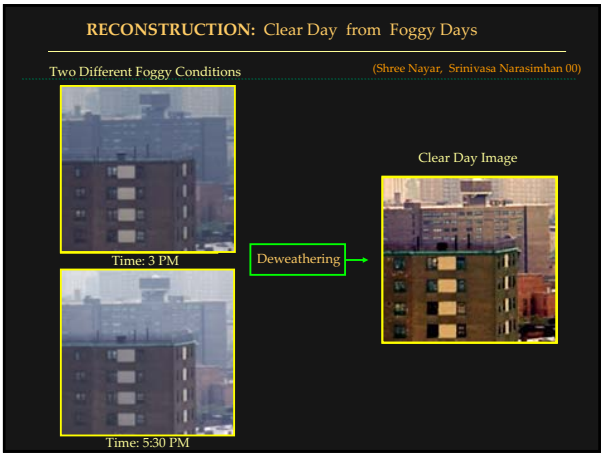


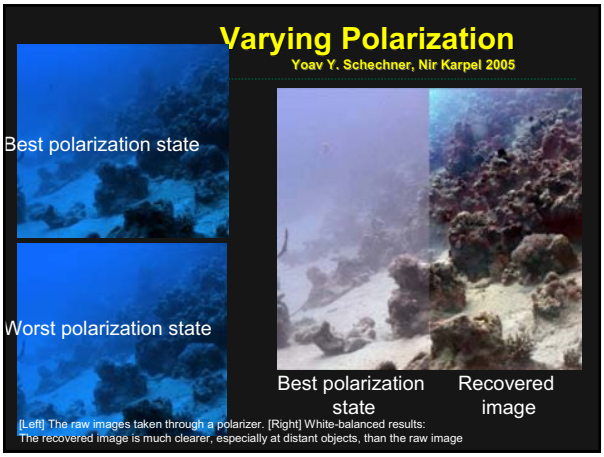
Control color *change's* ability to cause luminance *change*

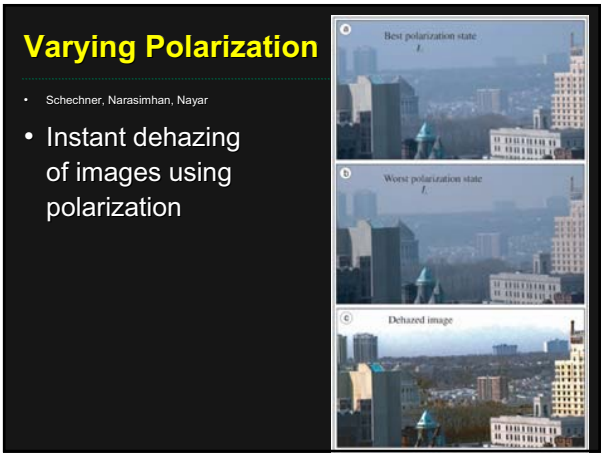


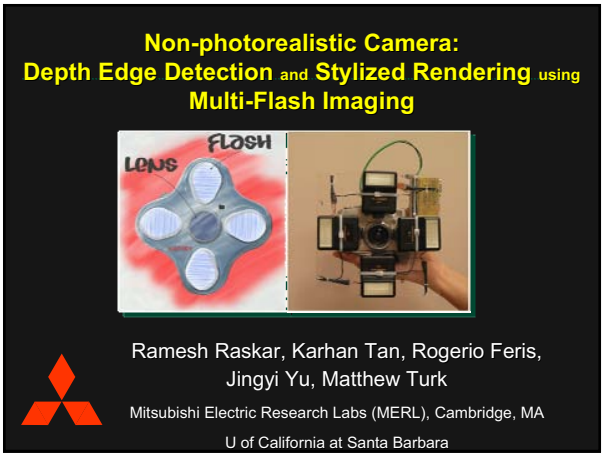

Which hue changes should be darker in grayscale?
Set a polarity for The color wheel:
Color Difference Space
+ΔC1,2
+Δb*
-Δb*
-Δa*
+Δa*
θ
+
-
+
-
sign(||ΔCi,j||) = sign(ΔCi,j • v̂θ)


Original
Color2Grey
Color2Grey+Color

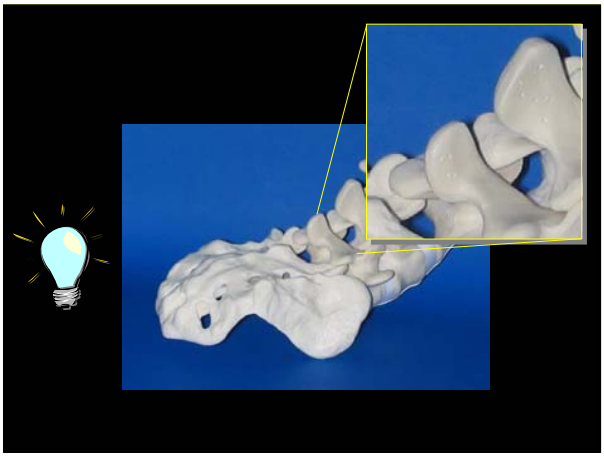

RECONSTRUCTION: Clear Day from Foggy Days
Two Different Foggy Conditions
(Shree Nayar, Srinivasa Narasimhan 00)
Time: 3 PM
Time: 5:30 PM
Deweathering
Clear Day Image

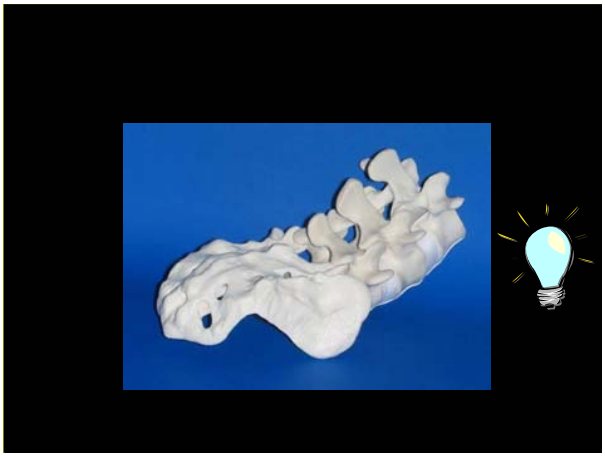

Varying Polarization
Yoav Y. Schechner, Nir Karpel 2005
Best polarization state
Worst polarization state
Best polarization state
Recovered image
[Left] The raw images taken through a polarizer. [Right] White-balanced results:
The recovered image is much clearer, especially at distant objects, than the raw image

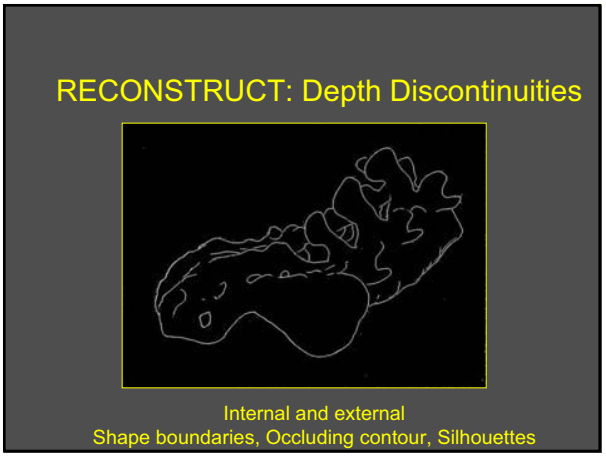

Varying Polarization
Schechner, Narasimhan, Nayar
Instant dehazing of images using polarization
Best polarization state
Worst polarization state
Dehazed image

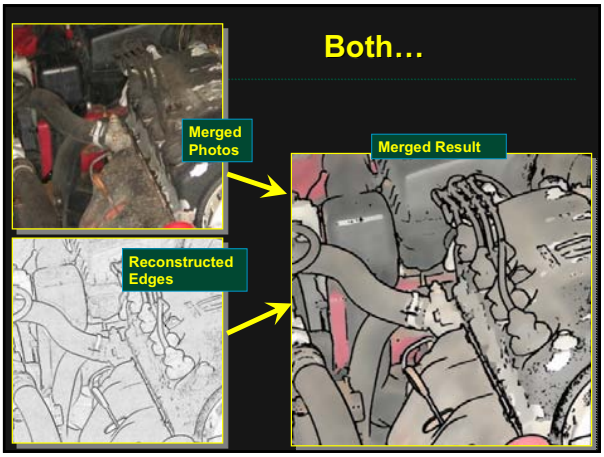

Non-photorealistic Camera:
Depth Edge Detection and Stylized Rendering using Multi-Flash Imaging
Flash
Lens
Ramesh Raskar, Karhan Tan, Rogerio Feris, Jingyi Yu, Matthew Turk
Mitsubishi Electric Research Labs (MERL), Cambridge, MA
U of California at Santa Barbara

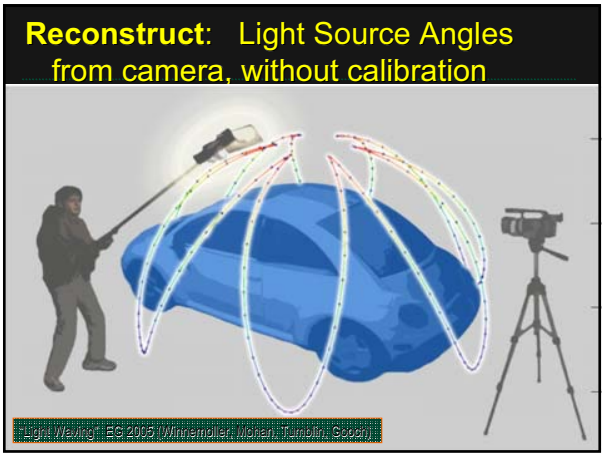

Canny
Our Method

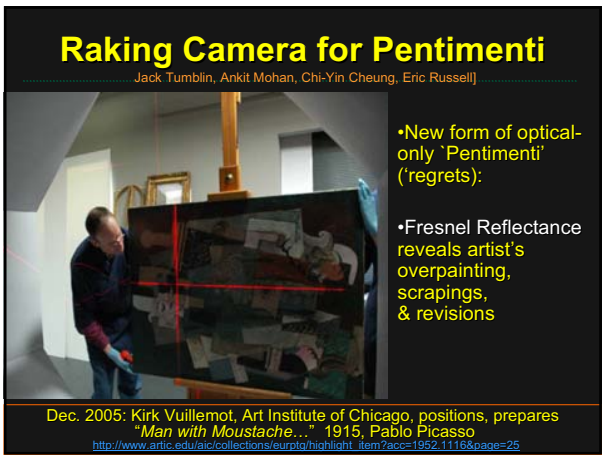


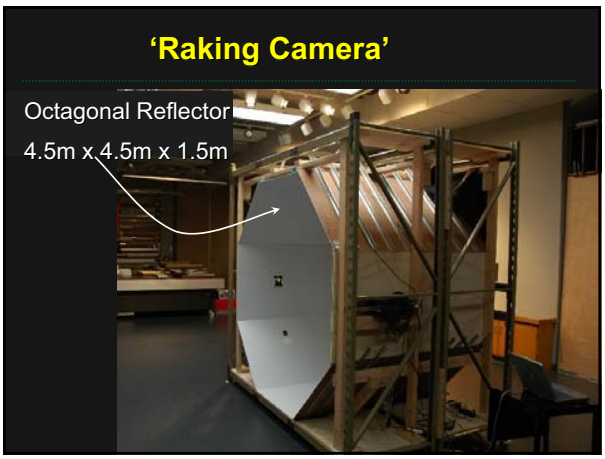



RECONSTRUCT: Depth Discontinuities
Internal and external
Shape boundaries, Occluding contour, Silhouettes


Both…
Merged Photos
Merged Result
Reconstructed Edges


Reconstruct: Light Source Angles from camera, without calibration
Light Waving! EG 2003 (Winnemöller, Mohan, Tumblin, Gooch)


Raking Camera for Pentimenti
Jack Tumblin, Ankit Mohan, Chi-Yin Cheung, Eric Russell]
•New form of optical-only `Pentimenti' ('regrets):
•Fresnel Reflectance reveals artist's overpainting, scrapings, & revisions
Dec. 2005: Kirk Vuillemot, Art Institute of Chicago, positions, prepares "Man with Moustache…" 1915, Pablo Picasso
http://www.artic.edu/aic/collections/eurptg/highlight_item?acc=1952.1116&page=25

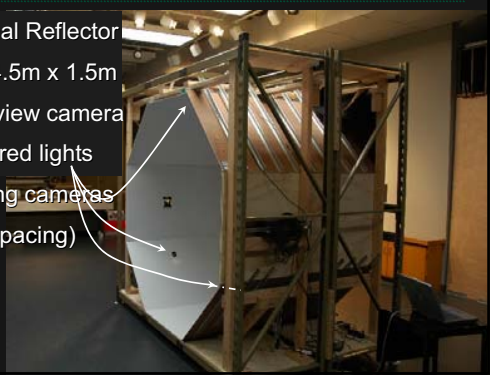

'Raking Camera'
Octagonal Reflector
4.5m x 4.5m x 1.5m


'Raking Camera' Art Inst. Chicago
Octagonal Reflector
4.5m x 4.5m x 1.5m
• 1 top view camera (not visible)


'Raking Camera' Art Inst. Chicago
Octagonal Reflector
4.5m x 4.5m x 1.5m
• 1 top view camera
• 2 steered lights

‘Raking Camera’ Art Inst. Chicago

Octagonal Reflector
4.5m x 4.5m x 1.5m
- 1 top view camera
- 2 steered lights
- 3 raking cameras
(120° spacing)



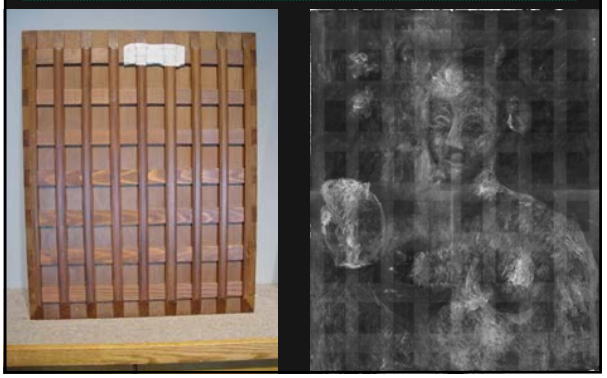
Reconstruction: Polynomial Texture Maps

Store just ***6*** coefficients at each pixel, get Interactive re-lighting...

A Mostly 2-D Method

2001 SIGGRAPH
Malzbender, HPlabs

X-ray Study: we cancel the backing?



Pentimenti: Can we see, follow the revisions?

Shape-Time Photography

Freeman et al 2003

Visually Expressive ‘Time Fragments’

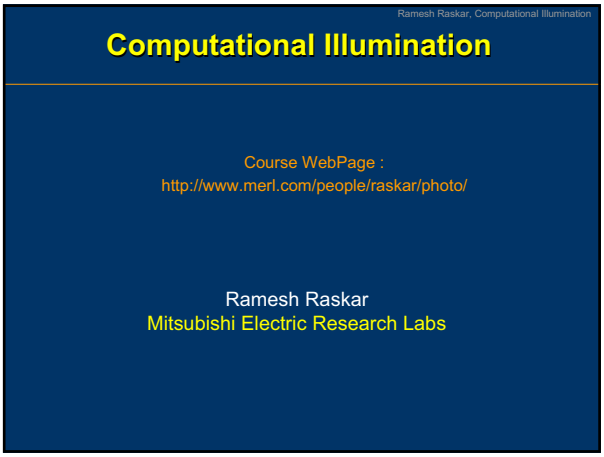
- Duchamp
  - Nude Descending a Staircase

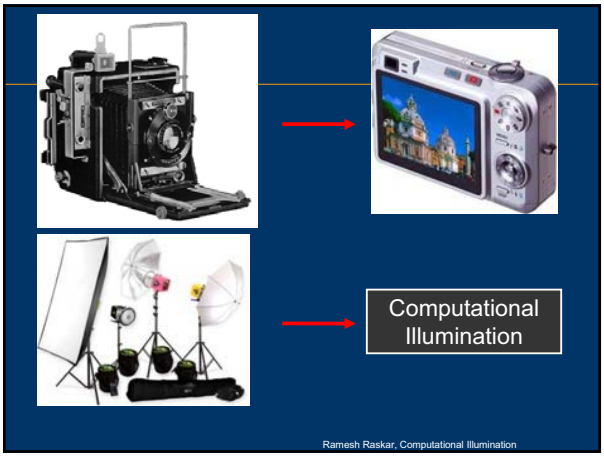
FUSION: Time/Space Mixtures

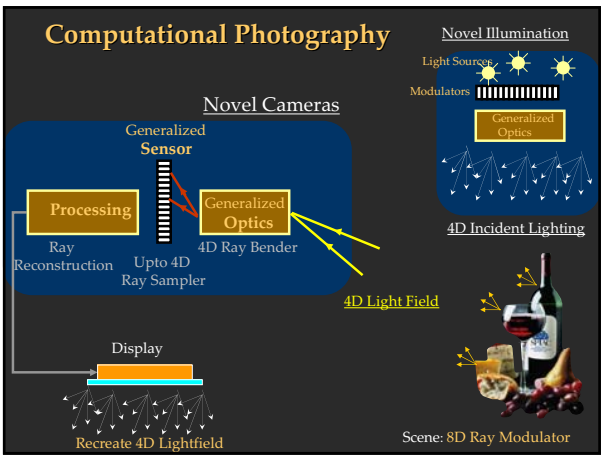
Richard Hundley 2001

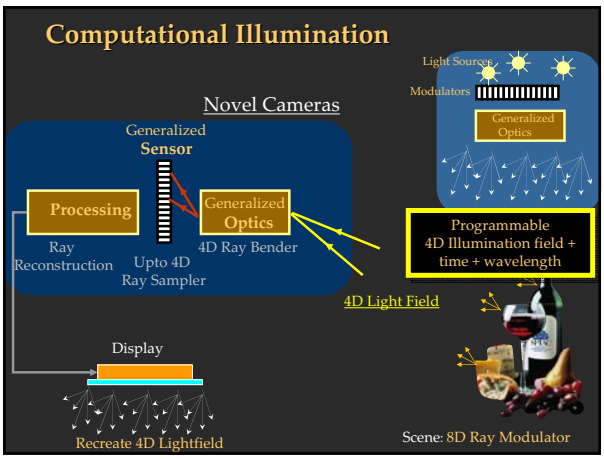
Concluding Questions

- How should we modify 'film-like' photography to better gather the 'visually essential' (not just the 'optically essential') contents of a scene?

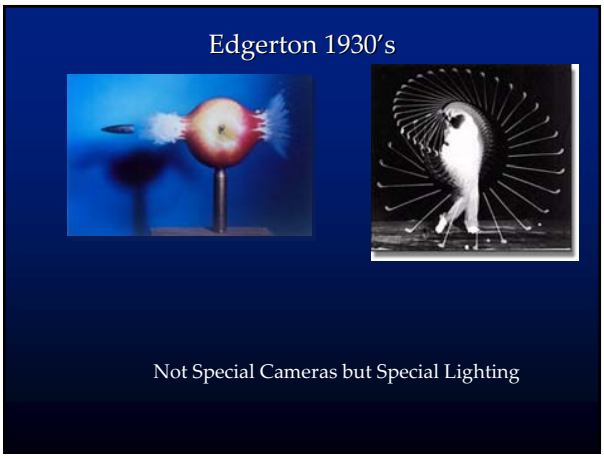

Ramesh Raskar, Computational Illumination
Computational Illumination
Course WebPage :
http://www.merl.com/people/raskar/photo/
Ramesh Raskar
Mitsubishi Electric Research Labs

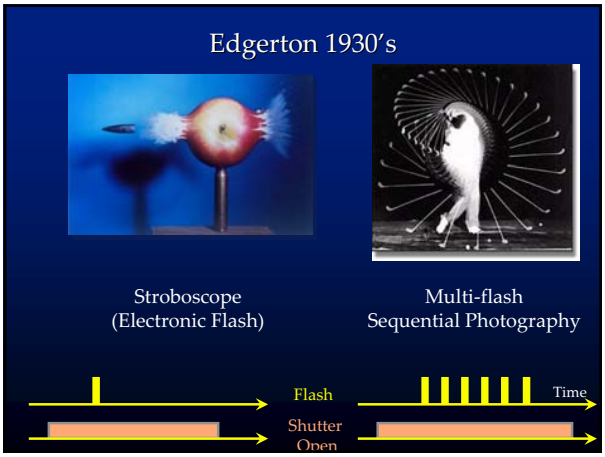

Computational Illumination
Ramesh Raskar, Computational Illumination


Computational Photography
Novel Illumination
Light Sources
Modulators
Generalized Optics
Novel Cameras
Generalized Sensor
Processing
Generalized Optics
Ray Reconstruction
Upto 4D Ray Sampler
4D Ray Bender
4D Incident Lighting
4D Light Field
Display
Recreate 4D Lightfield
Scene: 8D Ray Modulator


Computational Illumination
Light Sources
Modulators
Generalized Optics
Novel Cameras
Generalized Sensor
Processing
Generalized Optics
Ray Reconstruction
Upto 4D Ray Sampler
4D Ray Bender
Programmable 4D Illumination field + time + wavelength
4D Light Field
Display
Recreate 4D Lightfield
Scene: 8D Ray Modulator


'Smarter' Lighting Equipment
What Parameters Can We Change ?


Edgerton 1930's
Not Special Cameras but Special Lighting



Ramesh Raskar, Computational Illumination

Computational Illumination:

Programmable 4D Illumination Field + Time + Wavelength

- Presence or Absence, Duration, Brightness
  - Flash/No-flash
- Light position
  - Multi-flash for depth edges
  - Programmable dome (image re-lighting and matting)
- Light color/wavelength
- Spatial Modulation
  - Synthetic Aperture Illumination
- Temporal Modulation
  - TV remote, Motion Tracking, Sony ID-cam, RFIG
- Exploiting (uncontrolled) natural lighting condition
  - Day/Night Fusion

Ramesh Raskar, Computational Illumination

Computational Illumination

- Presence or Absence, Duration, Brightness
  - Flash/No-flash
- Light position
  - Multi-flash for depth edges
  - Programmable dome (image re-lighting and matting)
- Light color/wavelength
- Spatial Modulation
  - Synthetic Aperture Illumination
- Temporal Modulation
  - TV remote, Motion Tracking, Sony ID-cam, RFIG
- General lighting condition
  - Day/Night

Denoising Challenging Images

Available light:

+ nice lighting

- noise/blurriness
- color

Flash:

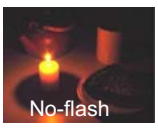
+ details
+ color

- flat/artificial

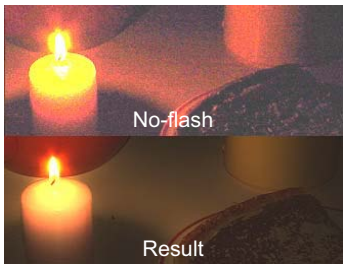
Elmar Eisemann and Frédo Durand , Flash Photography Enhancement via Intrinsic Relighting

Georg Petschnigg, Maneesh Agrawala, Hugues Hoppe, Richard Szeliski, Michael Cohen, Kentaro Toyama. Digital Photography with Flash and No-Flash Image Pairs

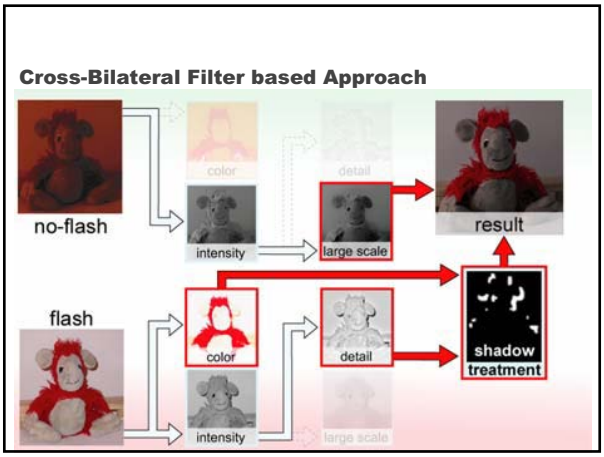
Denoise no-flash image using flash image

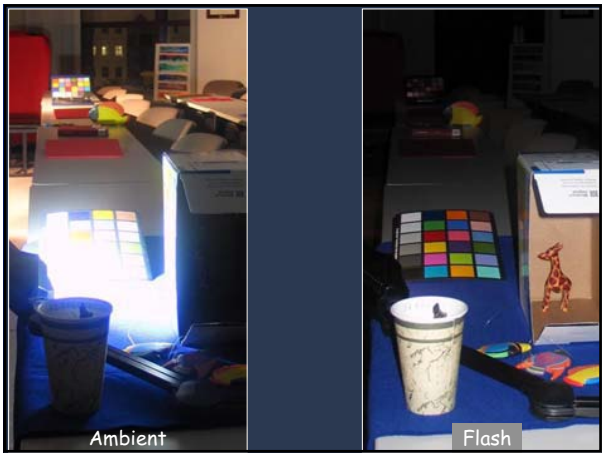


Transfer detail from flash image to no-flash image



+ original lighting
+ details/sharpness
+ color

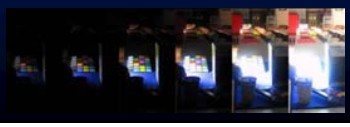


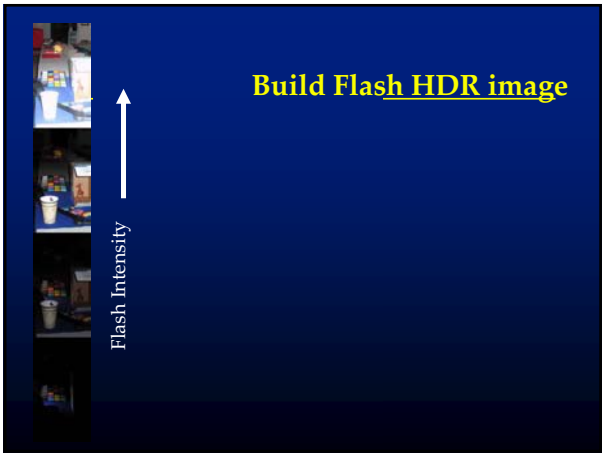


Build Exposure HDR image

- Multiple images with different exposure
  - Debevec & Malik, Siggraph 97
  - Nayar & Mitsunaga, CVPR 00

Increasing Exposure →

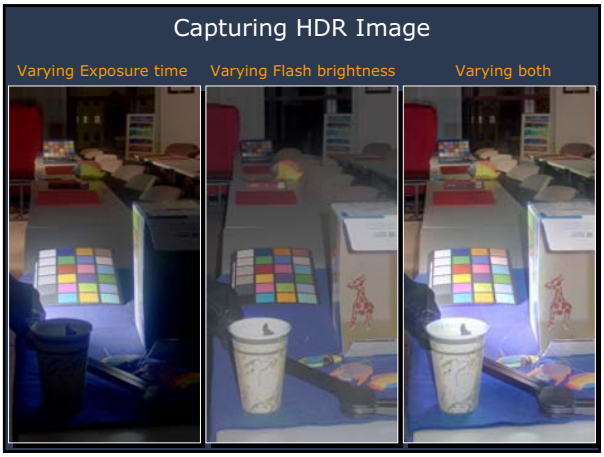


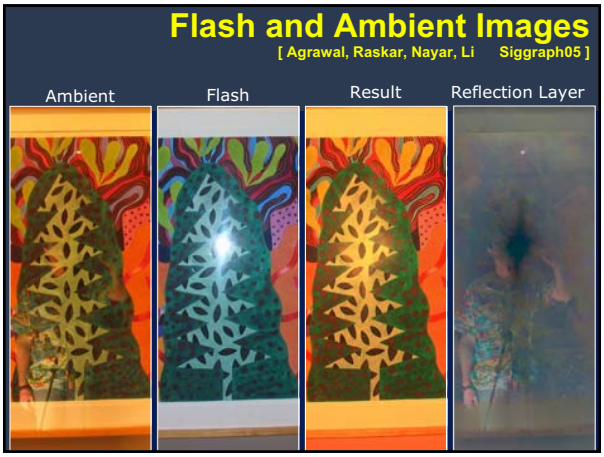


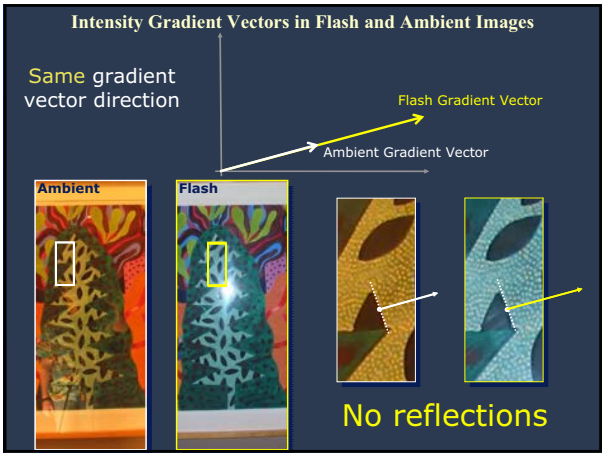
Build Flash HDR image

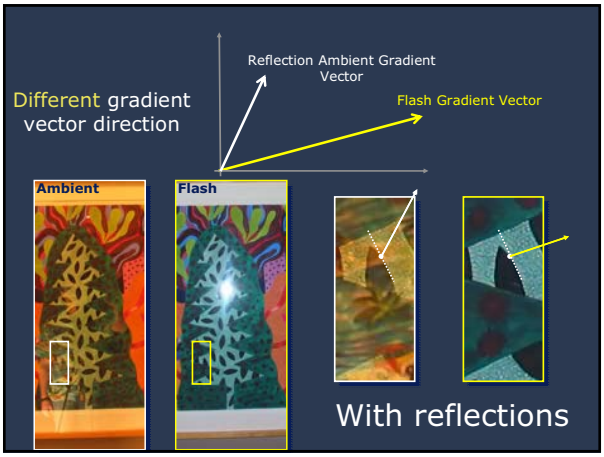
Build Flash-Exposure HDR image

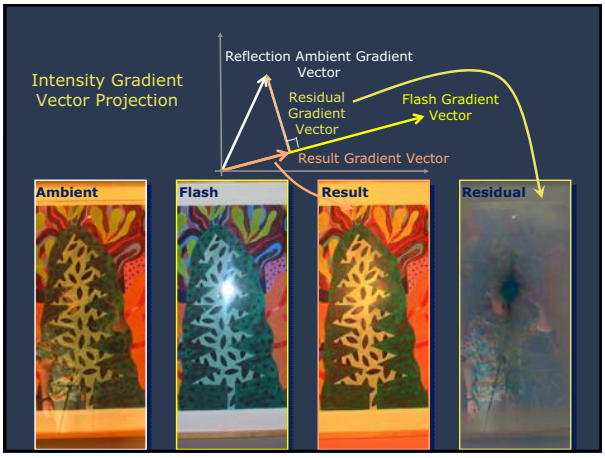
Agrawal, Raskar, Nayar, Li
Siggraph05

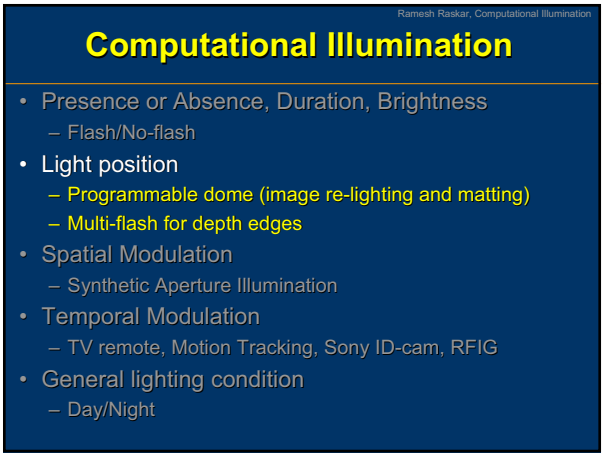

Capturing HDR Image
Varying Exposure time
Varying Flash brightness
Varying both


Flash and Ambient Images
[Agrawal, Raskar, Nayar, Li Siggraph05]
Ambient
Flash
Result
Reflection Layer


Intensity Gradient Vectors in Flash and Ambient Images
Same gradient vector direction
Flash Gradient Vector
Ambient Gradient Vector
Ambient
Flash
No reflections


Different gradient vector direction
Reflection Ambient Gradient Vector
Flash Gradient Vector
Ambient
Flash
With reflections


Intensity Gradient Vector Projection
Reflection Ambient Gradient Vector
Residual Gradient Vector
Flash Gradient Vector
Result Gradient Vector
Ambient
Flash
Result
Residual


Ramesh Raskar, Computational Illumination
Computational Illumination
• Presence or Absence, Duration, Brightness
– Flash/No-flash
• Light position
– Programmable dome (image re-lighting and matting)
– Multi-flash for depth edges
• Spatial Modulation
– Synthetic Aperture Illumination
• Temporal Modulation
– TV remote, Motion Tracking, Sony ID-cam, RFIG
• General lighting condition
– Day/Night

Synthetic Lighting

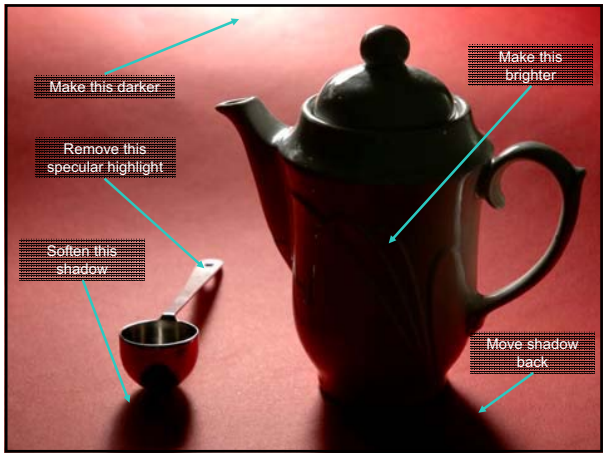
Paul Haeberli, Jan 1992

Ramesh Raskar, Computational Illumination

Table-top Computed Lighting for Practical Digital Photography

Ankit Mohan, Jack Tumblin
Northwestern University
Bobby Bodenheimer
Vanderbilt University
Cindy Grimm, Reynold Bailey
Washington University in St. Louis

Ramesh Raskar, Computational Illumination

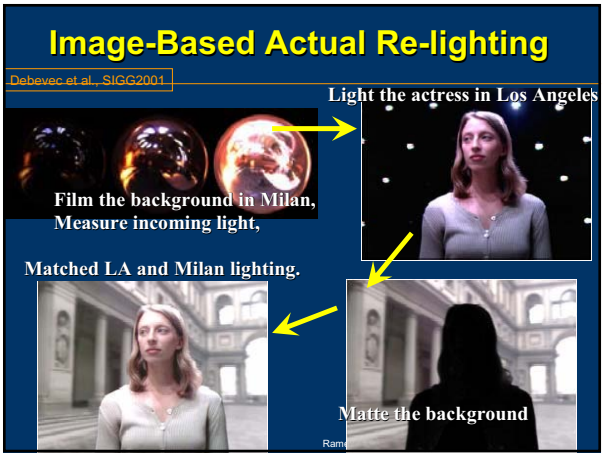


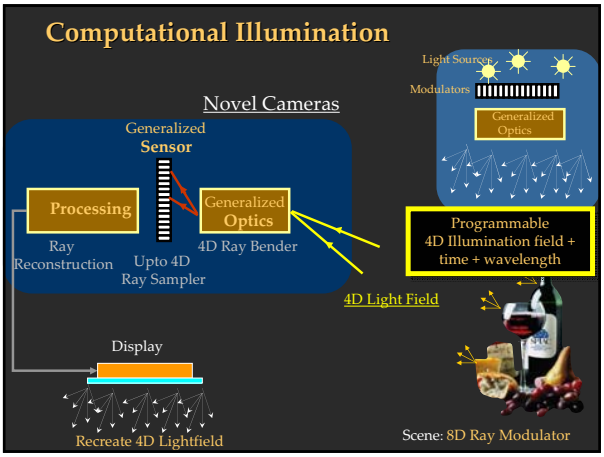
Debevec et al. 2002: 'Light Stage 3'

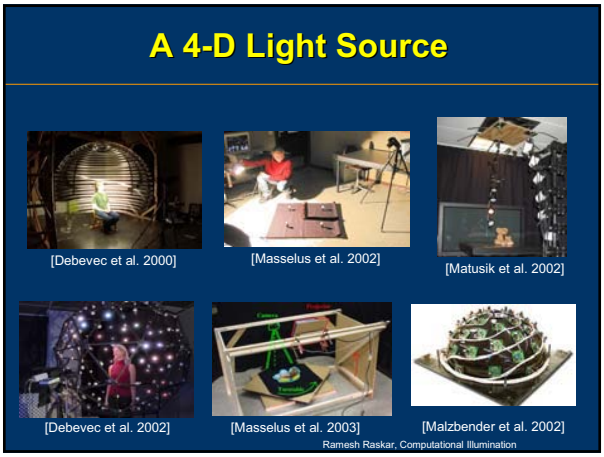
Ramesh Raskar, Computational Illumination

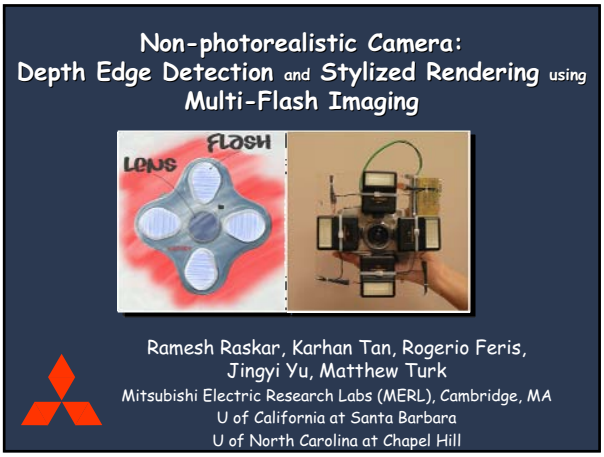
Image-Based Actual Re-lighting

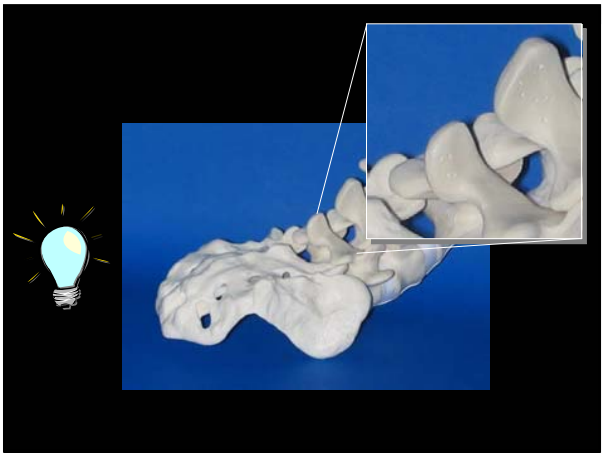
Debevec et al., SIGG2001

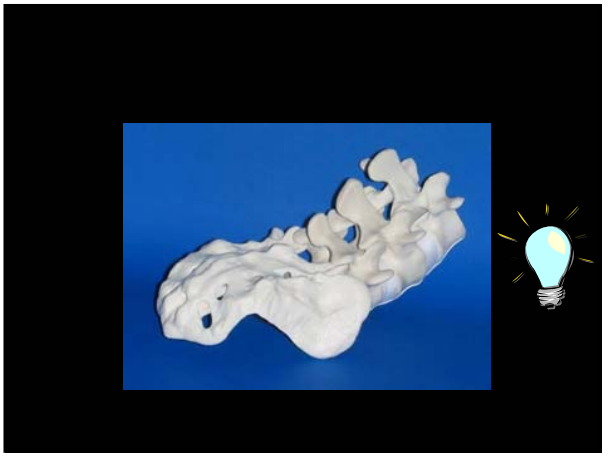


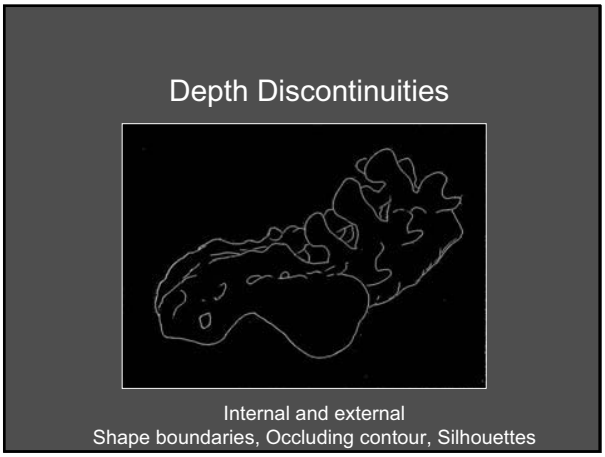

Computational Illumination
Light Sources
Modulators
Generalized Optics
Novel Cameras
Generalized Sensor
Processing
Generalized Optics
Ray Reconstruction
Upto 4D Ray Sampler
4D Ray Bender
Programmable 4D Illumination field + time + wavelength
4D Light Field
Display
Recreate 4D Lightfield
Scene: 8D Ray Modulator

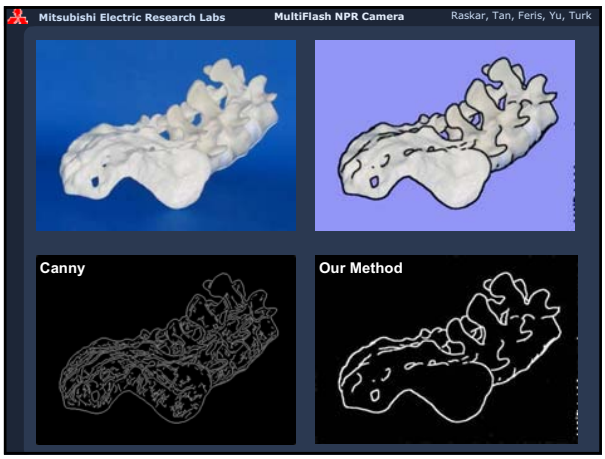

A 4-D Light Source
[Debevec et al. 2000]
[Masselus et al. 2002]
[Matusik et al. 2002]
[Debevec et al. 2002]
[Masselus et al. 2003]
[Malzbender et al. 2002]
Ramesh Raskar, Computational Illumination

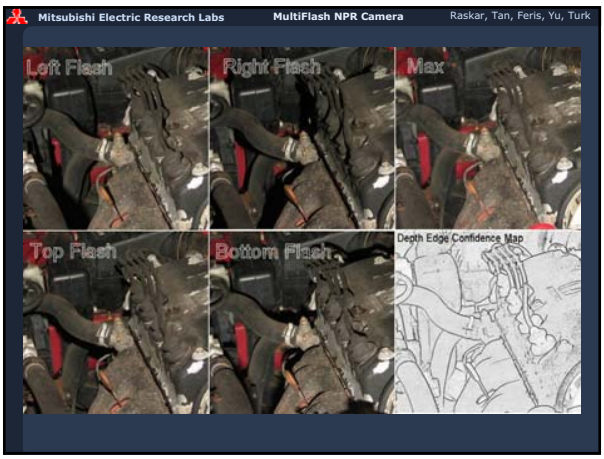

Non-photorealistic Camera:
Depth Edge Detection and Stylized Rendering using Multi-Flash Imaging
Flash
Lens
Ramesh Raskar, Karhan Tan, Rogerio Feris,
Jingyi Yu, Matthew Turk
Mitsubishi Electric Research Labs (MERL), Cambridge, MA
U of California at Santa Barbara
U of North Carolina at Chapel Hill

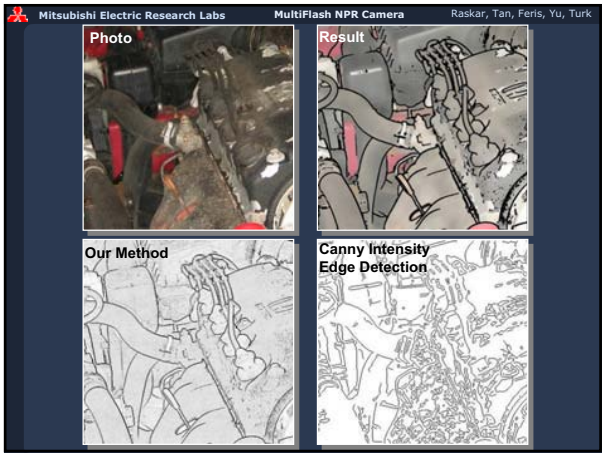


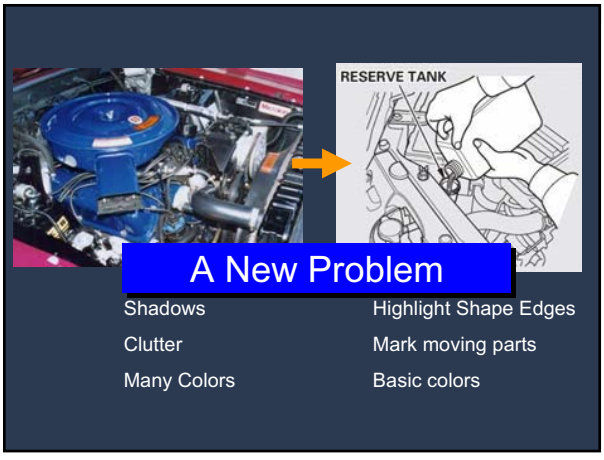



Depth Discontinuities
Internal and external
Shape boundaries, Occluding contour, Silhouettes


Mitsubishi Electric Research Labs
MultiFlash NPR Camera
Raskar, Tan, Feris, Yu, Turk
Canny
Our Method


Mitsubishi Electric Research Labs
MultiFlash NPR Camera
Raskar, Tan, Feris, Yu, Turk
Left Flash
Right Flash
Max
Top Flash
Bottom Flash
Depth Edge Confidence Map


Mitsubishi Electric Research Labs
MultiFlash NPR Camera
Raskar, Tan, Feris, Yu, Turk
Photo
Result
Our Method
Canny Intensity
Edge Detection


RESERVE TANK
A New Problem
Shadows
Clutter
Many Colors
Highlight Shape Edges
Mark moving parts
Basic colors

Ramesh Raskar, Computational Illumination

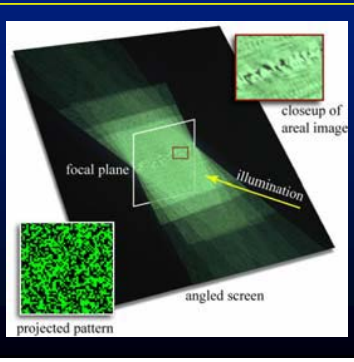
Computational Illumination

- Presence or Absence
  - Flash/No-flash
- Light position
  - Multi-flash for depth edges
  - Programmable dome (image re-lighting and matting)
- Spatial Modulation (Intra-flash 2D Modulation)
  - Camera flash = Projector
  - Synthetic Aperture Illumination
  - Dual Photography
- Temporal Modulation
  - TV remote, Motion Tracking, Sony ID-cam, RFIG
- General lighting condition
  - Day/Night

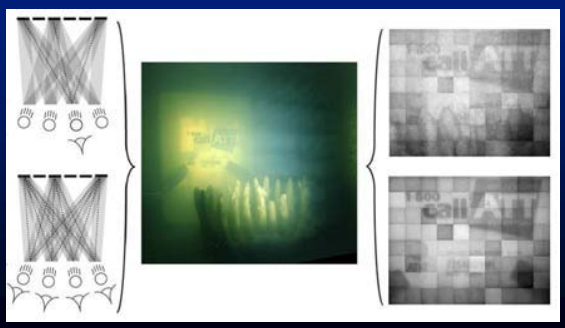
Synthetic aperture imaging + illumination

Levoy et al 2004

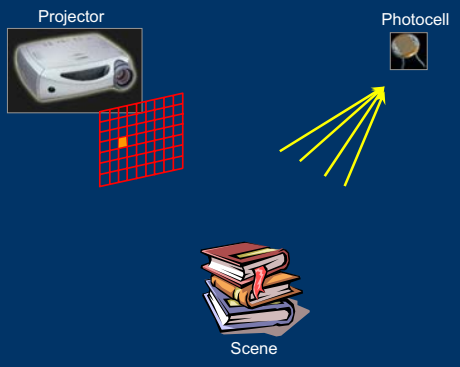
What does synthetic aperture illumination look like?



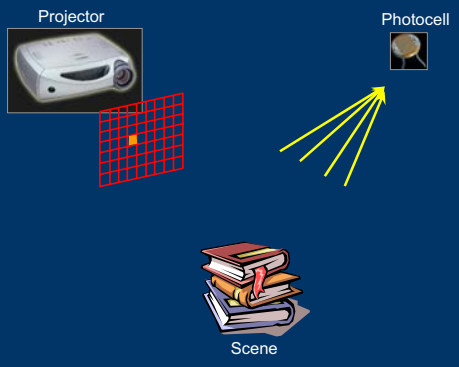
Underwater confocal imaging with and without SAP

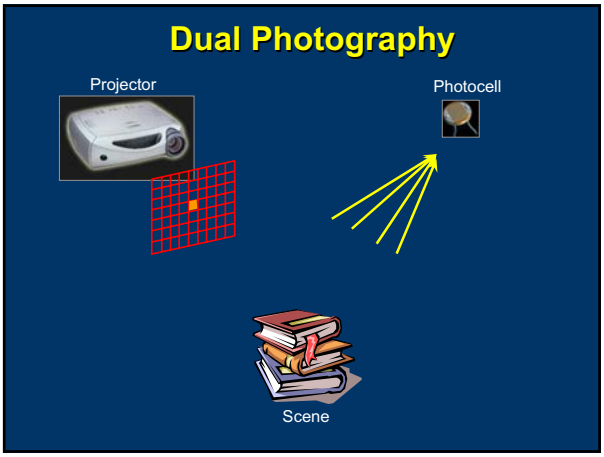


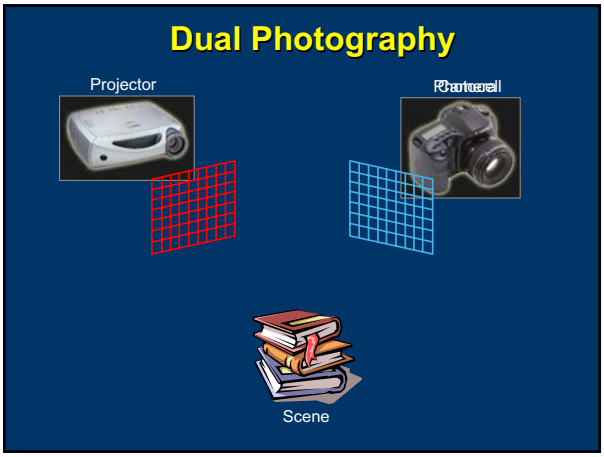
Dual Photography

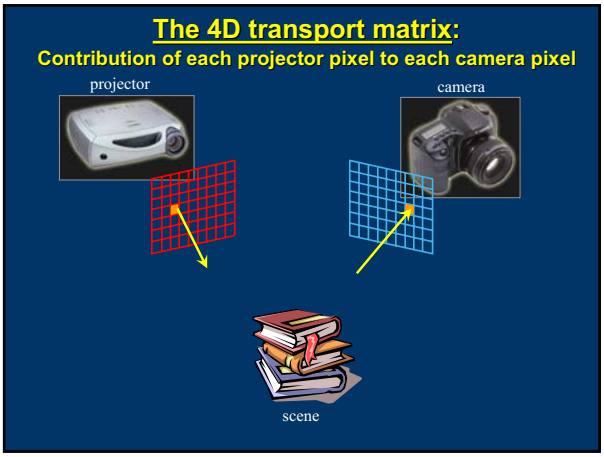


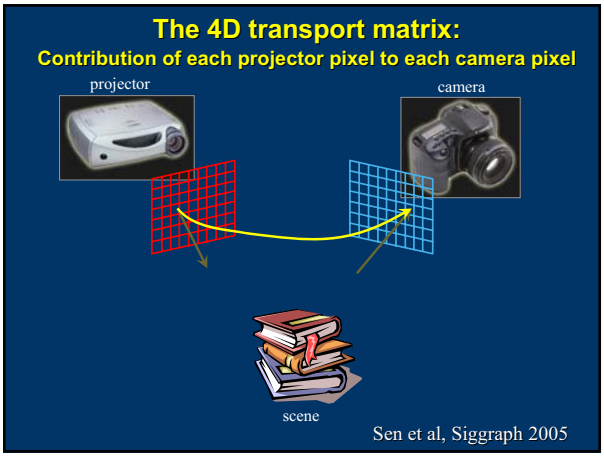
Dual Photography

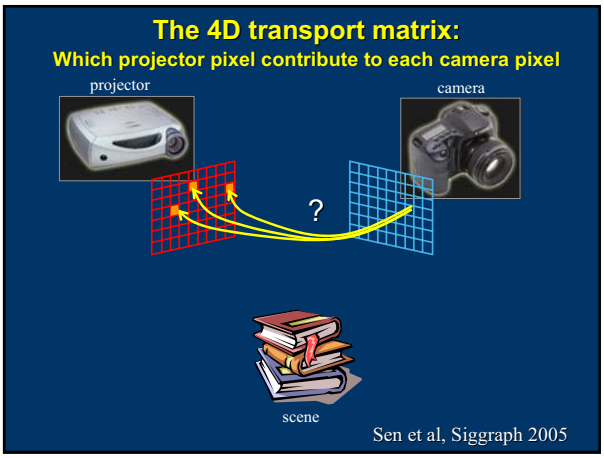


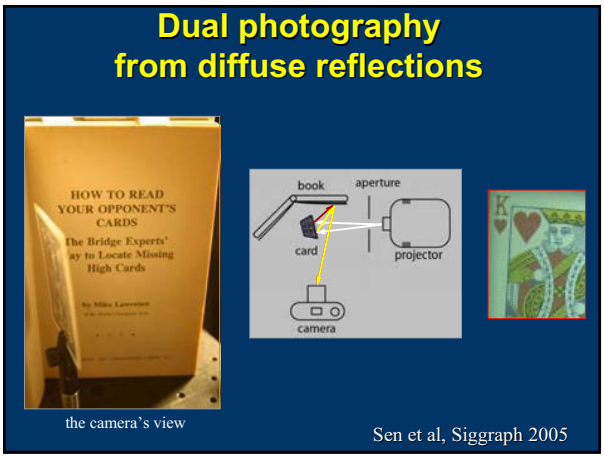

Dual Photography
Projector
Photocell
Scene


Dual Photography
Projector
Scene


The 4D transport matrix:
Contribution of each projector pixel to each camera pixel
projector
camera
scene


The 4D transport matrix:
Contribution of each projector pixel to each camera pixel
projector
camera
scene
Sen et al, Siggraph 2005


The 4D transport matrix:
Which projector pixel contribute to each camera pixel
projector
camera
?
scene
Sen et al, Siggraph 2005


Dual photography
from diffuse reflections
HOW TO READ YOUR OPPONENT'S CARDS
The Bridge Experts' ay to Locate Missing High Cards
book
aperture
card
projector
camera
the camera's view
Sen et al, Siggraph 2005

Ramesh Raskar, Computational Illumination

Computational Illumination

- Presence or Absence
  - Flash/No-flash
- Light position
  - Multi-flash for depth edges
  - Programmable dome (image re-lighting and matting)
- Light color/wavelength

- Spatial Modulation
  - Synthetic Aperture Illumination
- Temporal Modulation
  - TV remote, Motion Tracking, Sony ID-cam, RFIG
- General lighting condition
  - Day/Night

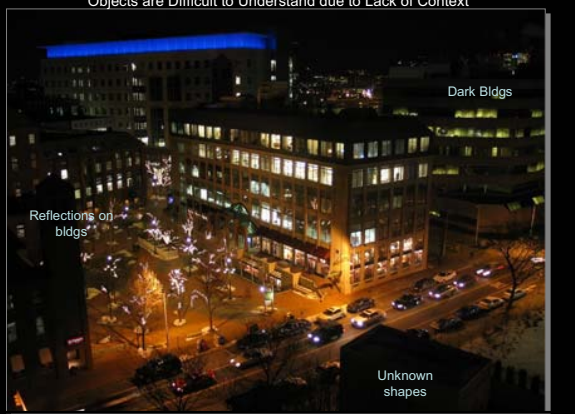
Ramesh Raskar, Computational Illumination

Computational Illumination

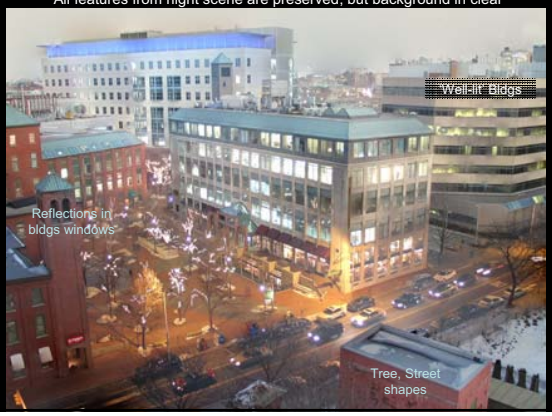
- Presence or Absence
  - Flash/No-flash
- Light position
  - Multi-flash for depth edges
  - Programmable dome (image re-lighting and matting)
- Light color/wavelength

- Spatial Modulation
  - Synthetic Aperture Illumination
- Temporal Modulation
  - TV remote, Motion Tracking, Sony ID-cam, RFIG
- Natural lighting condition
  - Day/Night Fusion

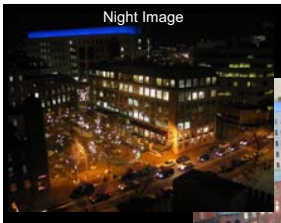
A Night Time Scene:
Objects are Difficult to Understand due to Lack of Context



Enhanced Context :
All features from night scene are preserved, but background in clear

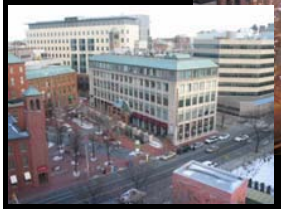


Night Image



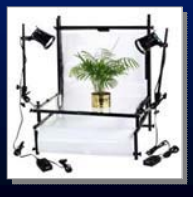
Background is captured from day-time scene using the same fixed camera

Result: Enhanced Image

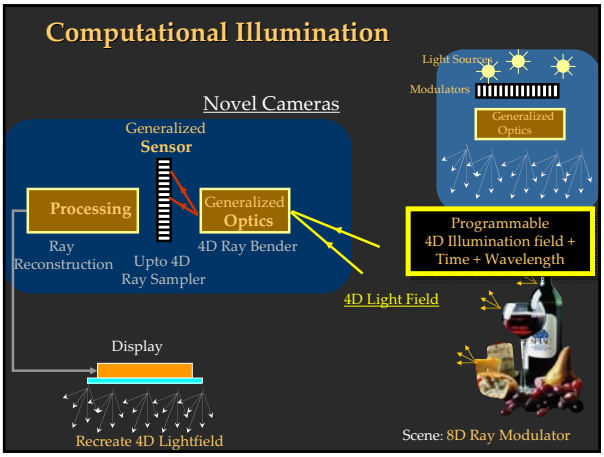


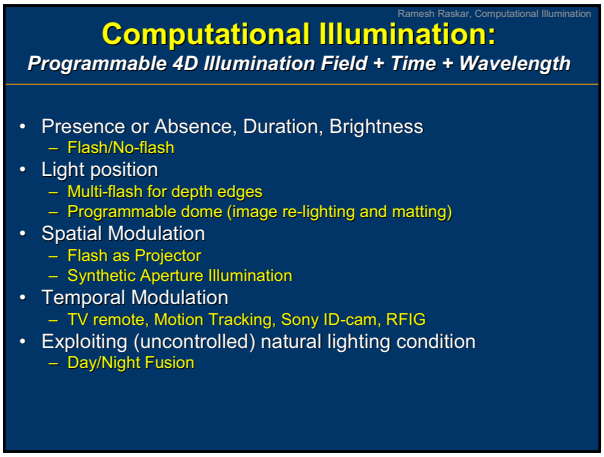
Day Image

'Smarter' Lighting Equipment



Programmable Parameters

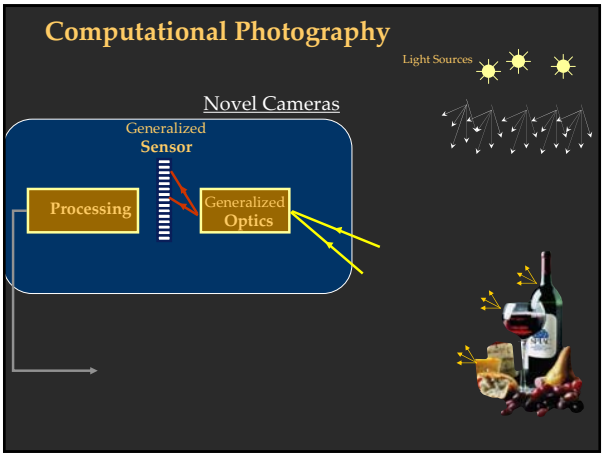

Computational Illumination
Novel Cameras
Generalized
Sensor
Processing
Generalized
Optics
Ray
Reconstruction
Upto 4D
Ray Sampler
4D Ray Bender
Light Sources
Modulators
Generalized
Optics
Programmable
4D Illumination field +
Time + Wavelength
4D Light Field
Display
Recreate 4D Lightfield
Scene: 8D Ray Modulator


Ramesh Raskar, Computational Illumination
Computational Illumination:
Programmable 4D Illumination Field + Time + Wavelength
• Presence or Absence, Duration, Brightness
– Flash/No-flash
• Light position
– Multi-flash for depth edges
– Programmable dome (image re-lighting and matting)
• Spatial Modulation
– Flash as Projector
– Synthetic Aperture Illumination
• Temporal Modulation
– TV remote, Motion Tracking, Sony ID-cam, RFIG
• Exploiting (uncontrolled) natural lighting condition
– Day/Night Fusion

Smart Optics, Modern Sensors and Future Cameras

Course WebPage :
http://www.merl.com/people/raskar/photo

Computational Photography

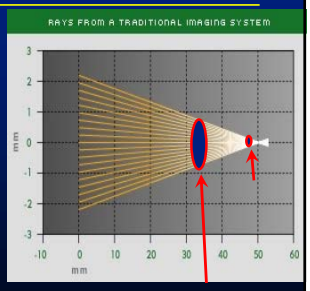


Future Directions

- Scientific Imaging
  - Tomography, Deconvolution, Coded Aperture Imaging
- Computational Illumination
  - Light stages, Domes, Light waving, Towards 8D
- Smart Optics
  - Handheld Light field camera, Programmable imaging/aperture
- Smart Sensors
  - HDR Cameras, Gradient Sensing, Line-scan Cameras, Demodulators
- Speculations

Wavefront Coding: 10X Depth of Field

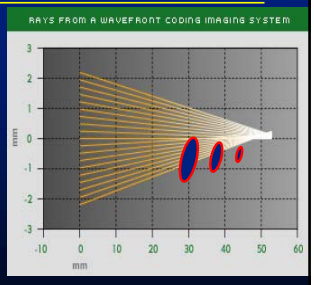
- Traditional Lens:
  - Defocus ('circle of confusion) dependent on distance from plane of focus



http://www.cdm-optics.com/site/extended_dof.php

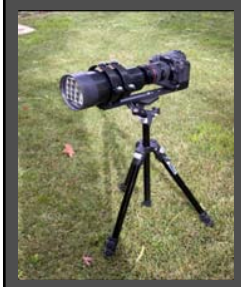
Wavefront Coding: 10X Depth of Field

- Traditional Lens:
  - Defocus dependent on distance from plane of focus
- Cubic Phase Plate
  - Defocus nearly independent of distance
  - All points 'blurred'
  - Deconvolve to get sharper image



http://www.cdm-optics.com/site/extended_dof.php

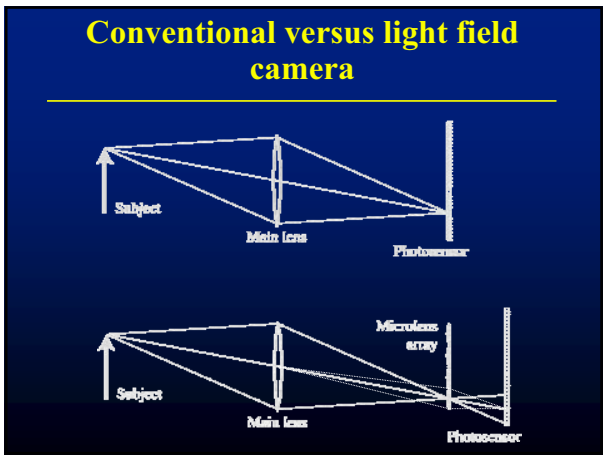
Integral Photography

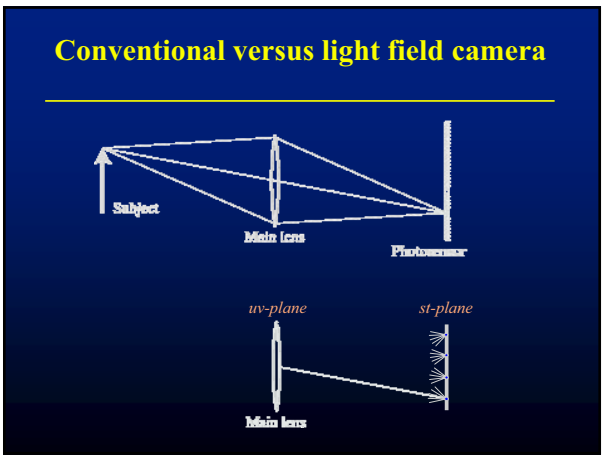


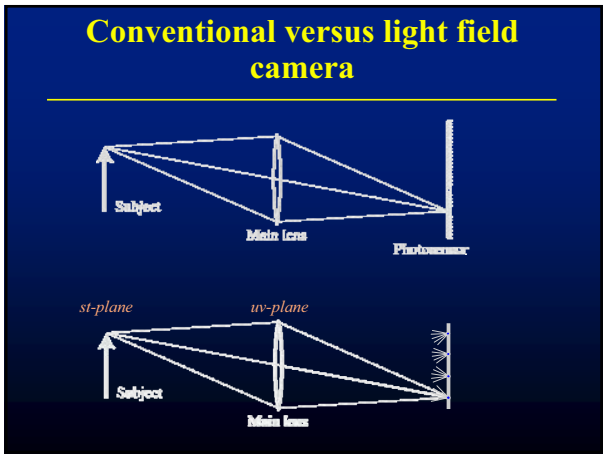
Todor Georgeiv et al 2006


Georgeiv et al 2006
(b) Synthetic aperture results with the focal plane moving from near to far.
(c) Synthetic aperture results with varying depth of field. (Left image demonstrates sparse sampling.)


Light field photography using a handheld plenoptic camera
Ren Ng, Marc Levoy, Mathieu Brédif,
Gene Duval, Mark Horowitz and Pat Hanrahan

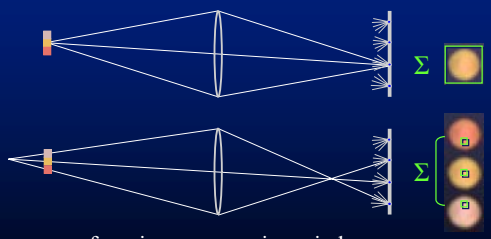

Conventional versus light field camera
Subject
Main lens
Photosensor
Subject
Microlens array
Main lens
Photosensor


Conventional versus light field camera
Subject
Main lens
uv-plane
st-plane
Main lens


Conventional versus light field camera
Subject
Main lens
st-plane
uv-plane
Subject
Main lens


Prototype camera
Contax medium format camera
Kodak 16-megapixel sensor
Adaptive Optics microlens array
125µ square-sided microlenses
4000 × 4000 pixels ÷ 292 × 292 lenses = 14 × 14 pixels

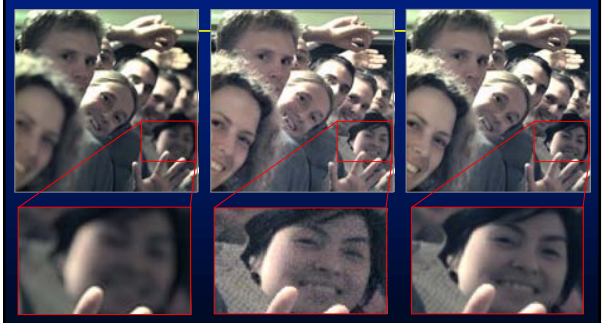
Digital refocusing



refocusing = summing windows
extracted from several microlenses

Example of digital refocusing

Extending the depth of field



conventional photograph, main lens at $f/4$

conventional photograph, main lens at $f/22$

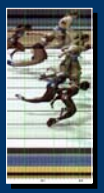
light field, main lens at $f/4$, after all-focus algorithm [Agarwala 2004]

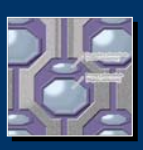
Future Directions

- Scientific Imaging
  - Tomography, Deconvolution, Coded Aperture Imaging
- Computational Illumination
  - Light stages, Domes, Light waving, Towards 8D
- Smart Optics
  - Handheld Light field camera, Programmable imaging/aperture
- Smart Sensors
  - HDR Cameras, Gradient Sensing, Line-scan Cameras, Demodulators
- Speculations

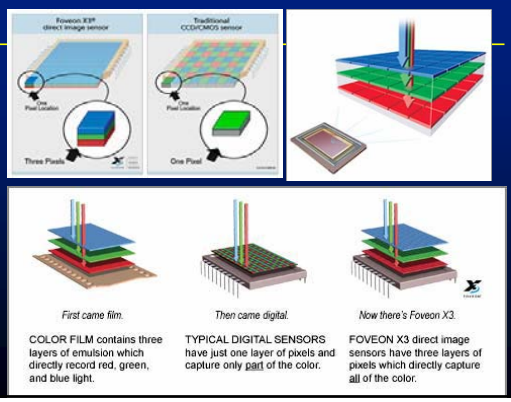
Novel Sensors

- Color
  - Foveon
- Dynamic Range
  - HDR Camera, Log sensing
  - Gradient sensing
- Identity
  - Demodulation
- 3D
  - ZCam, Canesta
- Motion
  - Line scan Camera
  - Flutter Shutter





Foveon: All Colors at a Single Pixel



High Dynamic Range

http://www.cybergrain.com/tech/hdr/

Fuji's SuperCCD S3 Pro camera has a chip with high and low sensitivity sensors per pixel location to increase dynamic range

Gradient Camera

Sensing Pixel Intensity Difference with Locally Adaptive Gain

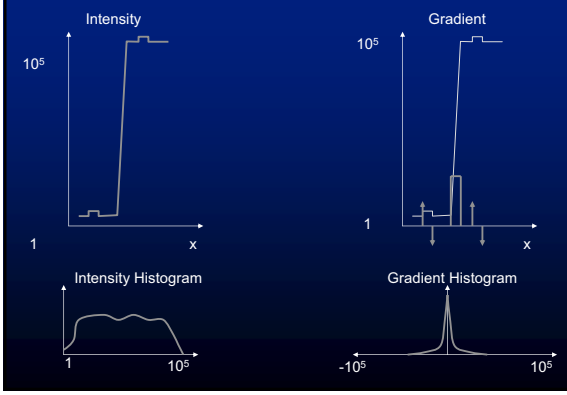
Ramesh Raskar, MERL
Work with Jack Tumblin, Northwestern U,
Amit Agrawal, U of Maryland

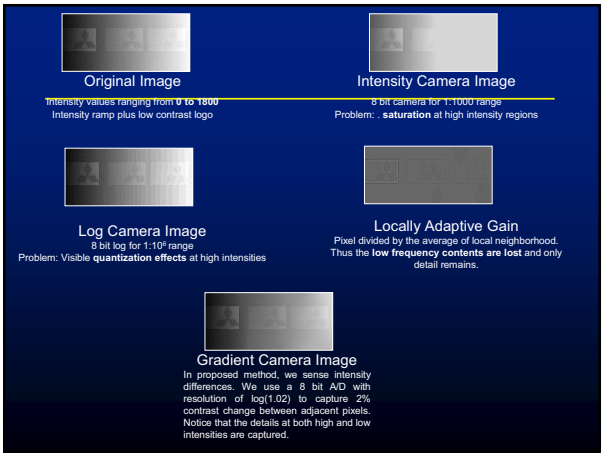
High Dynamic Range Images

Scene | Intensity camera saturation map | Gradient camera saturation map

Intensity camera fail to capture range
Gradients saturate at very few isolated pixels

Natural Scene Properties

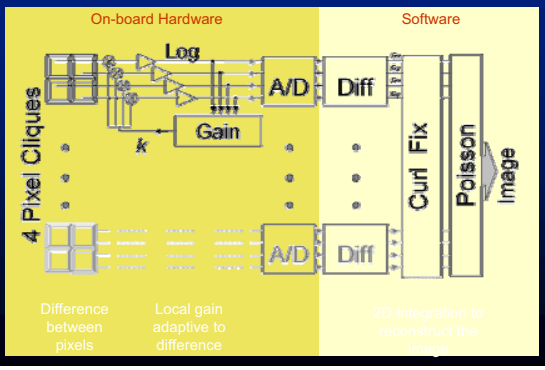




Gradient Camera

- Two main features
  1. Sense difference between neighboring pixel intensity
     At each pixel, measure (∇_x, ∇_y), $\nabla_x = I_{x+1,y} - I_{x,y}$, $\nabla_y = I_{x,y+1} - I_{x,y}$
  2. With locally adaptive gain
- Gradient camera is very similar to locally adaptive gain camera
- Locally Adaptive Gain Camera
  - Gain is **different** for each pixel
  - Problem: Loses low frequency detail and preserves only high frequency features (edges)
- Gradient Camera
  - The gain is **same** for four adjacent pixels
  - Difference between two pixels is measured with same gain on both pixels
  - Reconstruct original image in software from pixel differences by solving a linear system (solving Poisson Equation)

Camera Pipeline



Detail Preserving

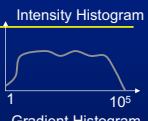
Intensity Camera | Log Intensity Camera | Gradient Camera

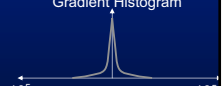
Intensity cameras capture detail but lose range
Log cameras capture range but lose detail

Quantization


Original Image


Uniform quantization 3 bits






Log Uniform quantization 3 bits

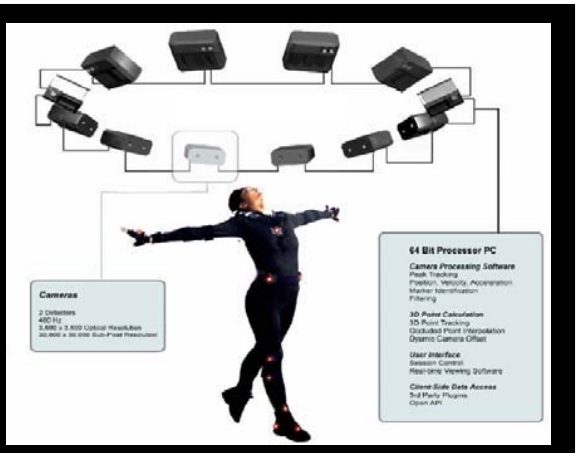

Log Uniform gradients quantization 3 bits

GradCam requires fewer bits

In the reconstructed image, error is pushed to high gradient pixel positions which is visually imperceptible

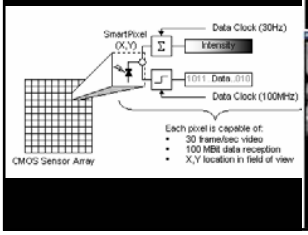
Demodulating Cameras

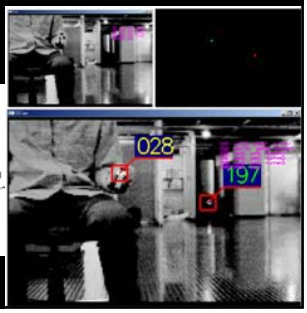
- Visualeyez™ VZ4000 Tracking System
- PhaseSpace motion digitizer



Demodulating Cameras

- Decode signals from blinking LEDs + image
  - Sony ID Cam
  - Phoci



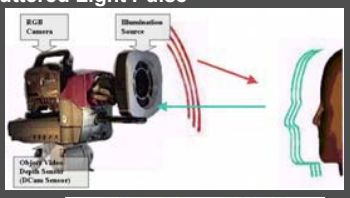


3D Cameras

- Time of flight
  - ZCam (Shuttered Light Pulse)
- Phase Decoding of modulated illumination
  - Canesta (Phase comparison)
  - Phase difference = depth
  - Magnitude = reflectance
- Structured Light
  - Binary coded light and triangulation

ZCam (3Dvsystems),

Shuttered Light Pulse



Resolution :
1cm for 2-7 meters

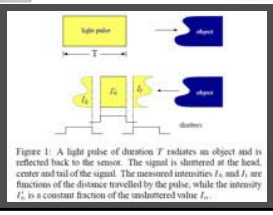


Figure 1: A light pulse of duration T radiates an object and is reflected back to the sensor. The signal is shuttered at the head, center and tail of the signal. The measured intensities I_0 and I_1 are functions of the distance travelled by the pulse, while the intensity I_0 is a constant fraction of the unshuttered value I_{in}.

Graphics can inserted behind and between characters

Canesta: Modulated Emitter

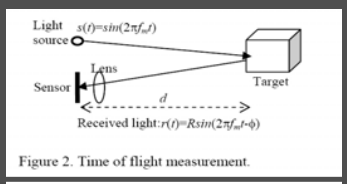


Figure 2. Time of flight measurement.

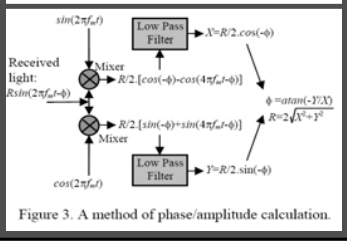
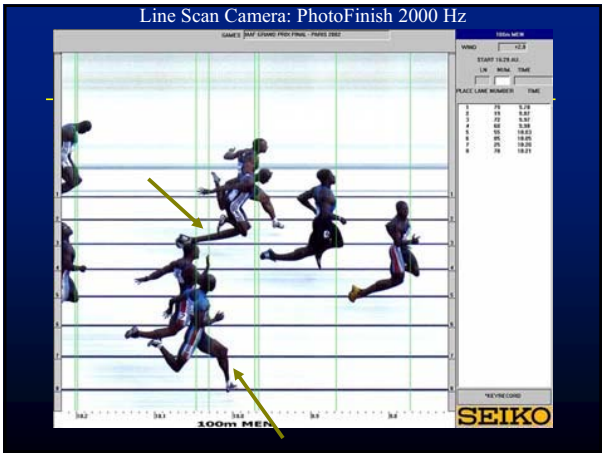
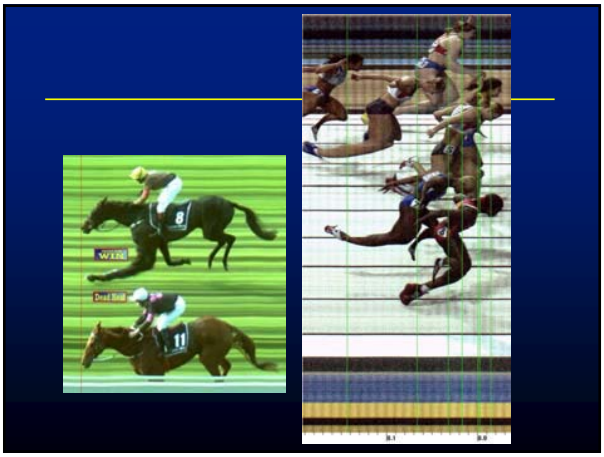


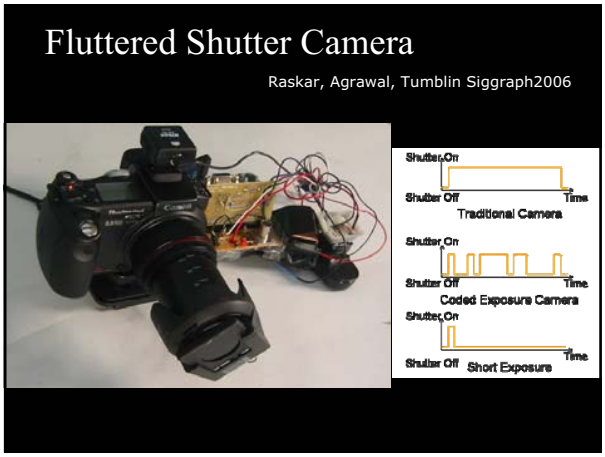
Figure 3. A method of phase/amplitude calculation.

Phase ~ distance
Amplitude ~ reflectance

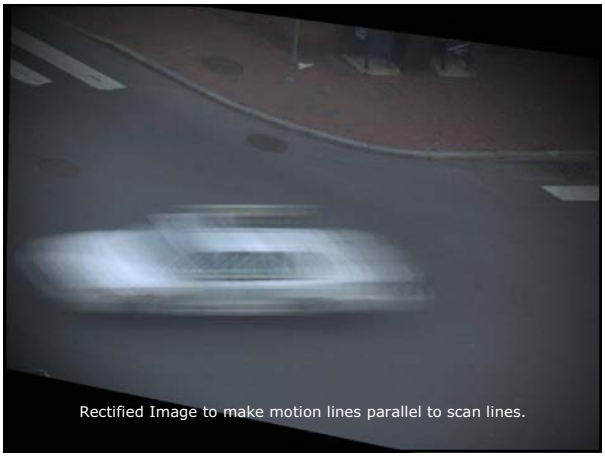
Motion _ _

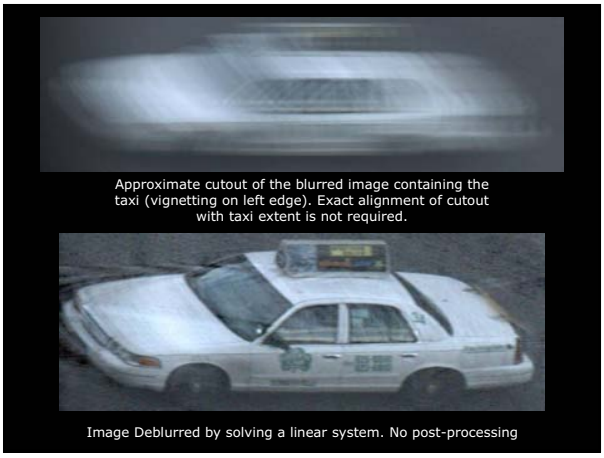

Line Scan Camera: PhotoFinish 2000 Hz
SEIKO



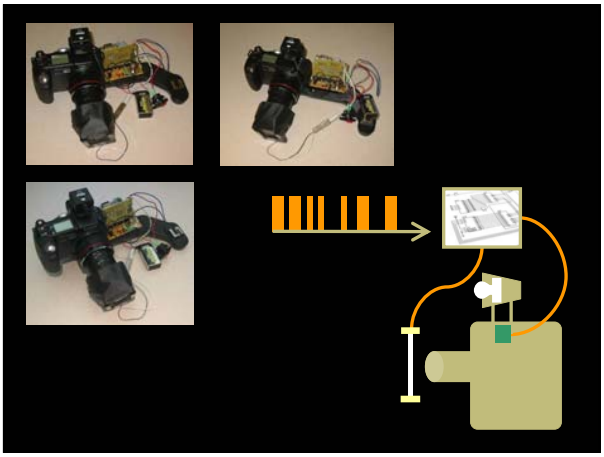

Fluttered Shutter Camera
Raskar, Agrawal, Tumblin Siggraph2006
Shutter On
Shutter Off
Time
Traditional Camera
Shutter On
Shutter Off
Time
Coded Exposure Camera
Shutter On
Shutter Off
Time
Short Exposure

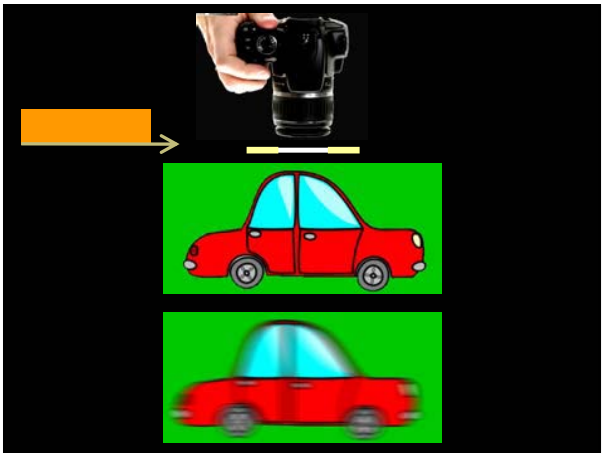

Input Image

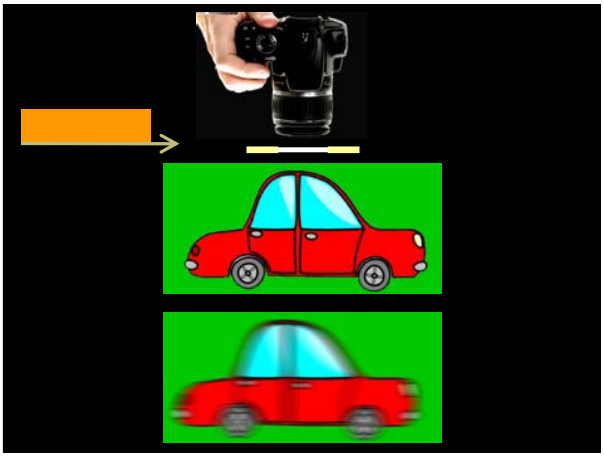

Rectified Image to make motion lines parallel to scan lines.

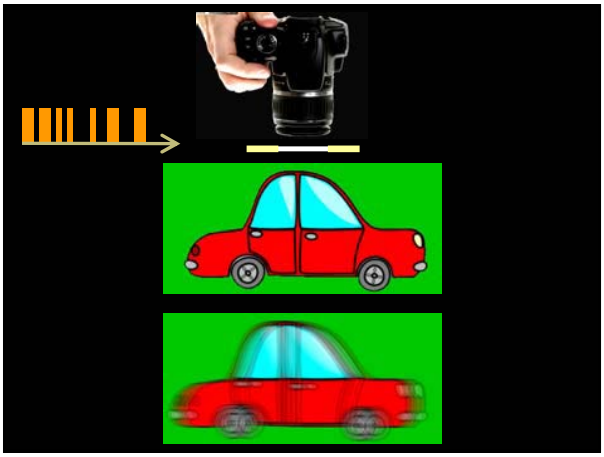

Approximate cutout of the blurred image containing the taxi (vignetting on left edge). Exact alignment of cutout with taxi extent is not required.
Image Deblurred by solving a linear system. No post-processing

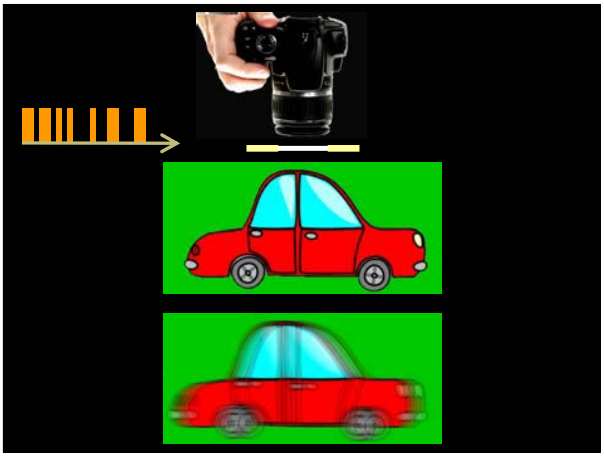
EG:810

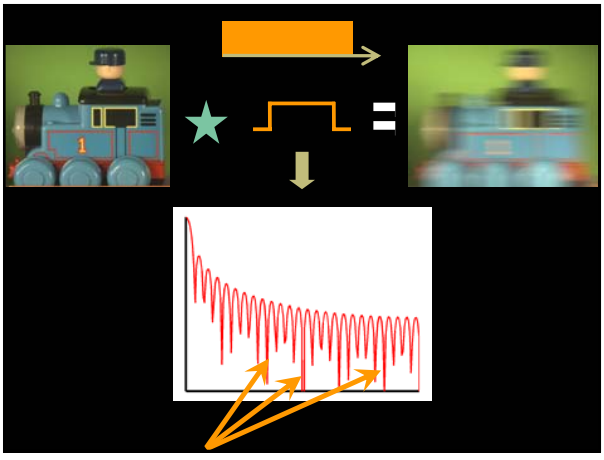


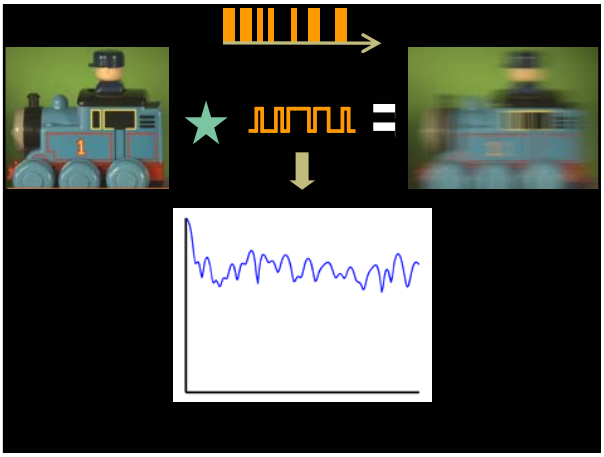


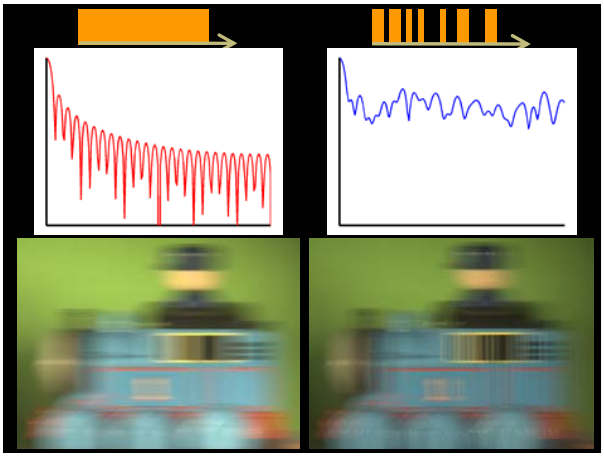


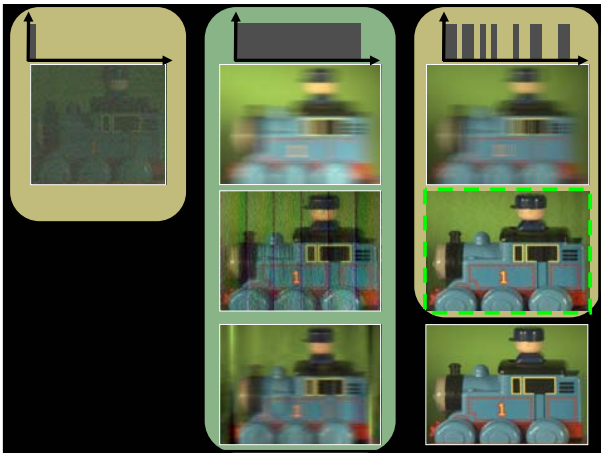








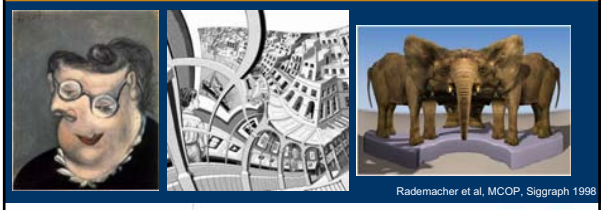




Novel Sensors

- Color
  - Foveon
- Dynamic Range
  - HDR Camera, Log sensing
  - Gradient sensing
- Identity
  - Demodulation
- 3D
  - ZCam, Canesta
- Motion
  - Line scan Camera
  - Flutter Shutter

Perspective? Or Not?



Rademacher et al, MCOP, Siggraph 1998

Agrawala et al, Long Scene Panoramas, Siggraph 2006

Multiperspective Camera?

[Jingyi Yu' 2004]

Fantasy Configurations

- 'Cloth-cam': 'Wallpaper-cam'
  - Fusion of 4D light emission and 4D capture in the surface of a cloth…
  - Invisible cloak
- Floating Cam:
  - Ad-hoc wireless networks form camera arrays in environment…
- Other ray sets:
  - Multilinear cameras (linear combination of 8 types) [Yu, McMillan'04, '05]

Goals

- Capture-time Techniques
  - Manipulating optics, illumination and sensors
- Fusion and Reconstruction
  - Beyond digital darkroom experience
- Improving Camera Performance
  - Better dynamic range, focus, frame rate, resolution
  - Hint of shape, reflectance, motion and illumination
  - Computational Imaging in Sciences
- Applications
  - Graphics, Special Effects, Scene Comprehension, Art

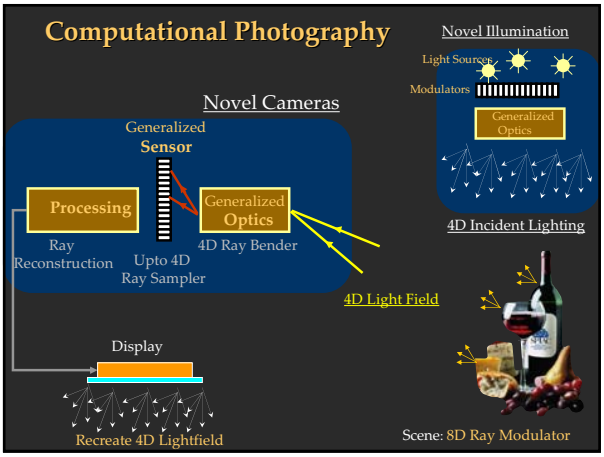
Acknowledgements

- MERL, Northwestern Graphics Group
- Amit Agrawal
- Shree Nayar
- Marc Levoy
- Jinbo Shi
- Ankit Mohan, Holger Winnemoller
- Image Credits
  - Ren Ng, Vaibhav Vaish, William Bennet
  - Fredo Durand, Aseem Agrawala
  - Morgan McGuire, Paul Debevec
  - And more

Computational Photography

Mastering New Techniques for Lenses, Lighting and Sensors

- Ramesh Raskar and Jack Tumblin
- Book Publishers: A K Peters
- Coupons 25% Off



EG:813

Applications of Information Theory to Computer Graphics

Eurographics 2007 Tutorial T8

Organizers

Mateu Sbert (University of Girona, Spain)
Miquel Feixas (University of Girona, Spain)

Speakers

Mateu Sbert (University of Girona, Spain)
Miquel Feixas (University of Girona, Spain)
Jaume Rigau (University of Girona, Spain)
Ivan Viola (University of Bergen, Norway)
Miguel Chover (Jaume I University, Spain)

EG:814

Applications of Information Theory to Computer Graphics Part I: Introduction

Mateu Sbert*, Miquel Feixas*, Jaume Rigau*, Ivan Viola** and Miguel Chover***

* University of Girona, Spain
** University of Bergen, Norway
*** Jaume I University, Spain

Abstract

We present different applications of information theory to computer graphics, based on the use of the measures of entropy, mutual information, f-divergences and generalized entropies. The application areas are hierarchical radiosity, adaptive ray-tracing, selection of best viewpoints, object and scene exploration, mesh saliency, mesh simplification and scientific visualization. We also give some hints on information-theoretic applications to object recognition and image processing.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computing Methodologies]: Computer Graphics

1. Presentation

These are the notes for the Eurographics 2007 tutorial *Applications of Information theory to Computer Graphics*. After an introduction to the basics concepts of Information Theory, different applications are presented to the fields of radiosity, adaptive ray-tracing, selection of best viewpoints, object and scene exploration, mesh saliency, mesh simplification and scientific visualization. These applications are based on the use of the measures of entropy, mutual information, f-divergences and generalized entropies. Other applications, such as image-based rendering, object recognition, image processing and light positioning, do not appear in the current version of these notes.

This document is organized in the following way. After this introductory part, second part deals with the basics of Information Theory (IT). The concept of information channel is introduced, and the quantities of entropy and mutual information are defined together with important relationships such as the Jensen-Shannon inequality. Finally the information bottleneck method is explained.

Third part deals with applications to radiosity. Radiosity is a viewpoint independent global illumination technique that discretizes the scene into small polygons or patches to solve a transport system of equations. The way the scene is discretized is critical for the efficiency of the result. We define first a scene information channel, which allows us to study the interchange of information between the patches. From the study of this channel several refinement oracles, i.e., criteria for subdividing the geometry, are obtained, aimed at maximizing the transport of information. Both classic and generalized information-theoretic quantities are used for this.

Fourth part is about adaptive ray-tracing. This technique is aimed at tracing more rays only where they are needed. For instance, smoothly illuminated regions of the scene with low variation do not need as much effort as rapidly varying illumination or also geometric discontinuities. The information theory quantities will be used again to define adaptive refinement oracles. New oracles are also defined for radiosity and adaptive ray-tracing using the following f-divergences: Kullback-Leibler, Chi-square, and Hellinger distances.

In the fifth part we define a viewpoint information channel between the points of view around an object and the polygons of the object. Several quantities associated to this channel, such as mutual information and entropy, are interpreted in terms of viewpoint quality measures. Viewpoint similarity and stability are defined, as well as methods for the selection of best n-views and for exploring the object. Mesh saliency is interpreted in terms of the viewpoint channel and polygonal mutual information as an ambient occlusion quantity. Impor-

tance is introduced into the scheme and saliency is used as importance to guide the viewpoint selection.

Sixth part deals with view selection in scientific visualization. The problem and context of the visualization of volumetric data sets is presented, together with the different viewpoint quality measures used. The framework presented in part sixth is then applied to this context. Guided navigation using higher-level semantics is also studied.

Finally, seventh part is about viewpoint-driven simplification. The several simplification algorithms are based on the variations perceived in image space, measured in our case through information-theoretic metrics: entropy, Kullback-Leibler distance and mutual information. These techniques are shown to give a better simplified mesh than object-based approaches, although at the cost of an increased processing time.

2. Authors

- Mateu Sbert
  University of Girona, Spain
  e-mail: mateu@ima.udg.edu
  Mateu Sbert is a professor in Computer Science at the University of Girona, Spain. He received a M.Sc. in Theoretical Physics (1977) at the university of Valencia, a M.Sc. in Mathematics (Statistics and Operations Research, 1983) at U.N.E.D. university (Madrid) and a Ph.D. in Computer Science at the U.P.C.(Universitat Politècnica de Catalunya, 1997, Best PhD Award). Mateu Sbert's research interests include the application of Monte Carlo, Integral Geometry and Information Theory techniques to Radiosity, Global Illumination and Image Based Rendering. He has authored or co-authored more than 100 papers in his areas of research, participated in several Eurographics tutorials and served as a member of program committee in Spanish and international conferences. He has participated in several European and Spanish research projects, leading the VIth European Framework Gametools project. Mateu Sbert coorganized the Dagsthul Seminars number 01242, Stochastic Methods in Rendering and number 06221, Computational Aesthetics in Graphics, Visualization and Imaging.
  http://ima.udg.edu/~mateu/
- Miquel Feixas
  University of Girona, Spain
  e-mail: feixas@ima.udg.edu
  http://ima.udg.edu/~feixas/
  Miquel Feixas is an associate professor in Computer Science at the University of Girona, Spain. He received a M.Sc. in Theoretical Physics (1979) at the UAB (Universitat Autònoma de Barcelona) and a Ph.D. in Computer Science at the UPC (Universitat Politècnica de Catalunya, 2002). His research is focused on the application of Information Theory techniques to Radiosity, Global Illumination, Viewpoint Selection and Image Processing. He has co-authored several papers in his area of research. He acted as a reviewer for conferences in the field of computer graphics and image processing. He has participated in Spanish research projects and joint actions with several European universities.
- Jaume Rigau
  University of Girona, Spain
  e-mail: rigau@ima.udg.edu
  http://ima.udg.edu/~rigau/
  Jaume Rigau is an associate professor in Computer Science at the University of Girona, Spain. He received a M.Sc. in Computer Science (1993) at the UPC (Universitat Politècnica de Catalunya) and a Ph.D. by the UPC (2006). His research is focused on the application of Information-Theory techniques to Radiosity, Global Illumination and Image Processing. He has co-authored several papers in his area of research. He has participated in several Spanish research projects and joint actions with several European universities.
- Ivan Viola
  University of Bergen, Norway
  e-mail: ivan@ii.uib.no
  http://www.ii.uib.no/~ivan/
  Ivan Viola is a Post Doctor research associate at University of Bergen, Norway. He was formerly associated with Vienna University of Technology, Austria, where he received M.Sc. in 2002 and Ph.D. in 2005. His research is focused on development of novel methods for automatically generating expressive visualizations of complex data. Viola co-authored several scientific works published on international conferences such as IEEE Visualization, EuroVis, and Vision Modeling and Visualization and acted as a reviewer and program committee member for conferences in the field of computer graphics and visualization. Recently he co-organized series of tutorials on Illustrative Visualization.
- Miguel Chover
  Jaume I University, Spain
  e-mail: chover@uji.es
  http://www3.uji.es/~chover/
  Miguel Chover is an associate professor at the Universitat Jaume I de Castelló, Spain. His research focused on interactive computer graphics, computer games and Web3D. His current work includes level of detail modelling, simplification algorithms, rendering natural phenomena and texturing techniques. He received his MS degree in Computer Science in 1992 and a PhD in Computer Science in 1996 from the Universidad Politécnica de Valencia, Spain. He is member of Eurographics.

3. Syllabus

1. Introduction (5 min)
   Speaker: Mateu Sbert
2. Information Theory Basics (35 min)
   Speaker: Miquel Feixas

- Information channel: entropy and mutual information
- Important inequalities
- Information bottleneck method

3. Refinement Criteria for Radiosity (20 min)
   Speaker: Jaume Rigau
   - Scene information channel
   - Refinement criteria for hierarchical radiosity
   - Mutual-information-based oracle
   - f-divergence-based oracles
4. Adaptive Refinement for Ray-tracing (15 min)
   Speaker: Jaume Rigau
   - Refinement criteria for ray-tracing
   - Entropy-based refinement criteria
   - f-divergence-based refinement criteria
5. Viewpoint Selection and Mesh Saliency (30 min)
   Speaker: Mateu Sbert
   - Viewpoint information-theoretic measures
   - Viewpoint information channel: mutual information, similarity and stability
   - Selection of best views and object exploration
   - Polygonal mutual information and information-theoretic ambient occlusion
   - Mesh saliency
   - Importance-driven viewpoint selection
6. View Selection in Scientific Visualization (30 min)
   Speaker: Ivan Viola
   - View Selection for Volumes and Iso-Surfaces
   - Importance-Driven Focus of Attention
   - Guided Navigation using Higher-Level Semantics
7. Viewpoint-driven Simplification (30 min)
   Speaker: Miguel Chover
   - Recent work on simplification
   - Information-theoretic metrics: entropy, Kullback-Leibler distance and mutual information
   - Simplification algorithms
8. Other Applications (15 min)
   Speaker: Miquel Feixas and Mateu Sbert
   - Image processing: registration and segmentation
   - Object recognition: shape descriptors

Aknowledgements

This project has been funded in part with grant numbers TIN2004-07451-C03-01 of the Spanish Government and IST-2-004363 (GameTools: Advanced Tools for Developing Highly Realistic Computer Games) from the VIth European Framework.

Applications of Information Theory to Computer Graphics Part II: Information Theory Basics

Miquel Feixas, Jaume Rigau, and Mateu Sbert

University of Girona, Spain

1. Introduction

In 1948, Claude Shannon published "A mathematical theory of communication" [Sha48] which marks the beginning of information theory. In this paper, he defined measures such as entropy and mutual information (called rate of transmission), and introduced the fundamental laws of data compression and transmission.

Information theory deals with the transmission, storage and processing of information and is used in fields such as physics, computer science, mathematics, statistics, economics, biology, linguistics, neurology, learning, etc. It is applied successfully in areas such as medical image processing, computer vision, robot motion and computer graphics.

In this part, we present some basic concepts of information theory. A very good reference is the text by Cover and Thomas [CT91]. Other references are Blahut [Bla87] and Lubbe [vdL97].

2. Entropy

In [Sha48], after representing a discrete information source as a Markov process, Shannon asks himself: "Can we define a quantity which will measure, in some sense, how much *information* is "produced" by such a process, or better, at what rate information is produced?".

His answer is: "Suppose we have a *set of possible* events whose *probabilities of occurrence* are $p_1, p_2, \ldots, p_n$. These probabilities are known but that is all we know concerning which event will occur. Can we find a measure of how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome?

If there is such a measure, say $H(p_1, p_2, \ldots, p_n)$, it is reasonable to require of it the following properties:

1. H would be continuous in the p_i.
2. If all the p_i are equal, $p_i = \frac{1}{n}$, then H should be a monotonic increasing function of n. With equally likely events there is more choice, or uncertainty, when there are more possible events.
3. If a choice is broken down into two successive choices, the original H should be the weighted sum of the individual values of H. The meaning of this is illustrated in Figure 1.

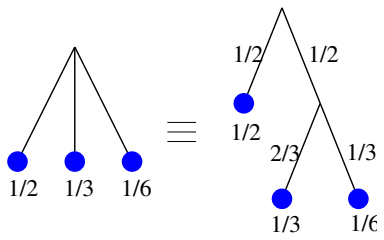


Figure 1: *Grouping property of the entropy.*

On the left, we have three possibilities $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{6}$. On the right, we first choose between two possibilities each with probability $\frac{1}{2}$, and if the second occurs, we make another choice with probabilities $\frac{2}{3}$, $\frac{1}{3}$. The final results have the same probabilities as before. We require, in this special case, that $H(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) = H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2}H(\frac{2}{3}, \frac{1}{3})$. The coefficient $\frac{1}{2}$ is because this second choice only occurs half the time."

After these requirements, he introduces the following *theorem*: "The only H satisfying the three above assumptions is of the form:

$$H = -K \sum_{i=1}^{n} p_i \log p_i \qquad (1)$$

where K is a positive constant". When $K = 1$ and the logarithm is $\log_2$, information is measured in bits.

Shannon calls this quantity *entropy*, as "the form of H will be recognized as that of entropy as defined in certain formulations of statistical mechanics where p_i is the probability of

a system being in cell i of its phase space". There are other axiomatic formulations which involve the same definition of entropy [CT91].

The Shannon entropy is the classical measure of *information*, where information is simply *the outcome of a selection among a finite number of possibilities*. Entropy also measures *uncertainty* or *ignorance*.

Thus, the *Shannon entropy* $H(X)$ of a discrete random variable X with values in the set $S = \{x_1, x_2, \ldots, x_n\}$ is defined as

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \tag{2}$$

where $n = |\mathcal{X}|$, $p_i = p(x_i) = Pr[X = x_i]$ for $i \in \{1, \ldots, n\}$, the logarithms are taken in base 2 (entropy is expressed in bits), and we use the convention that $0 \log 0 = 0$, which is justified by continuity. We can use interchangeably the notation $H(X)$ or $H(p)$ for the entropy, where p is the probability distribution $\{p_1, p_2, \ldots, p_n\}$, also represented by p_i. As $-\log p_i$ represents the *information* associated with the result x_i, the entropy gives us the *average information* or *uncertainty* of a random variable. Information and uncertainty are opposite. Uncertainty is considered before the event, information after. So, information reduces uncertainty. Note that the entropy depends only on the probabilities.

Some other relevant properties [Sha48] of the entropy are

1. $0 \leq H(X) \leq \log n$
   - $H(X) = 0$ if and only if all the probabilities except one are zero, this one having the unit value, i.e., when we are certain of the outcome.
   - $H(X) = \log n$ when all the probabilities are equal. This is the most uncertain situation.
2. If we equalize the probabilities, entropy increases.

If we consider another random variable Y with probability distribution q_i corresponding to values in the set $\mathcal{Y} = \{y_1, y_2, \ldots, y_m\}$, the *joint entropy* of X and Y is defined as

$$H(X,Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log p_{ij} \tag{3}$$

where $m = |S'|$ and $p_{ij} = P(x_i, y_j) = Pr[X = x_i, Y = y_j]$ is the joint probability.

When $n = 2$, the *binary* entropy (Figure 2) is given by

$$H(X) = -p \log p - (1-p) \log (1-p) \tag{4}$$

where $p = \{p, 1-p\}$.

Also, the *conditional entropy* is defined as

$$H(X|Y) = -\sum_{j=1}^{m} \sum_{i=1}^{n} p_{ij} \log p_{i|j} \tag{5}$$

where $p_{i|j} = p_{x_i|y_j} = Pr[X = x_i | Y = y_j]$ is the conditional probability.

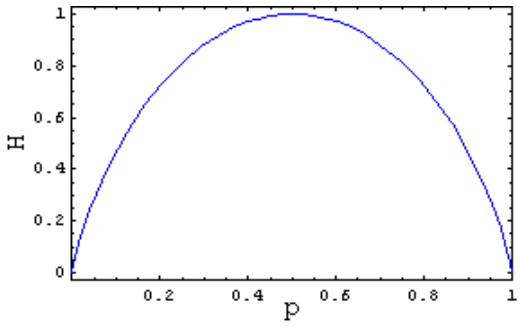


Figure 2: *Binary entropy.*

The Bayes theorem expresses the relation between the different probabilities:

$$p_{ij} = p_i p_{j|i} = q_j p_{i|j} \tag{6}$$

If X and Y are *independent*, then $p_{ij} = p_i q_j$.

The conditional entropy can be thought of in terms of a *channel* whose input is the random variable X and whose output is the random variable Y. $H(X|Y)$ corresponds to the uncertainty in the channel input from the receiver's point of view, and vice versa for $H(Y|X)$. Note that in general $H(X|Y) \neq H(Y|X)$.

The following properties are also met:

1. $H(X,Y) \leq H(X) + H(Y)$
2. $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$
3. $H(X) \geq H(X|Y) \geq 0$

3. Mutual Information

The *mutual information* between two random variables X and Y is defined as

$$\begin{aligned} I(X,Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= -\sum_{i=1}^{n} p_i \log p_i + \sum_{j=1}^{m} \sum_{i=1}^{n} p_{ij} \log p_{i|j} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log \frac{p_{ij}}{p_i q_j} \end{aligned} \tag{7}$$

Mutual information represents the amount of information that one random variable, the output of the channel, gives (or contains) about a second random variable, the input of the channel, and vice versa, i.e., how much the knowledge of X decreases the uncertainty of Y and vice versa. Therefore, $I(X,Y)$ is a measure of the shared information between X and Y.

Mutual information $I(X,Y)$ has the following properties:

1. $I(X,Y) \geq 0$ with equality if, and only if, X and Y are independent.

2. $I(X,Y) = I(Y,X)$
3. $I(X,Y) = H(X) + H(Y) - H(X,Y)$
4. $I(X,Y) \leq H(X)$

The relationship between all the above measures can be expressed by the Venn diagram, as shown in Figure 3.

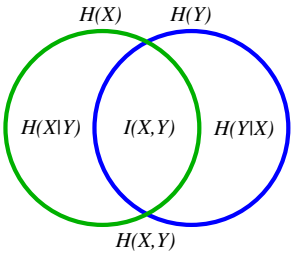


Figure 3: *Venn diagram of a discrete channel.*

The *relative entropy* or *Kullback-Leibler distance* between two probability distributions $p = \{p_i\}$ and $q = \{q_i\}$, that are defined over the set S, is defined as

$$D_{KL}(p\|q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \tag{8}$$

where, from continuity, we use the convention that $0 \log 0 = 0$, $p_i \log \frac{p_i}{0} = \infty$ if $a > 0$ and $0 \log \frac{0}{0} = 0$.

The relative entropy is "a measure of the inefficiency of assuming that the distribution is q when the true distribution is p" [CT91].

The relative entropy satisfies the *information inequality* $D_{KL}(p\|q) \geq 0$, with equality only if $p = q$. The relative entropy is also called *discrimination* and it is not strictly a distance, since it is not symmetric and does not satisfy the triangle inequality. Moreover, we have to emphasize that the mutual information can be expressed as

$$I(X,Y) = D_{KL}(\{p_{ij}\}\|\{p_i q_j\}) \tag{9}$$

4. Entropy Rate of a Markov Chain

The joint entropy of a collection of n random variables is given by

$$H(X_1,\ldots,X_n) = H(X_1) + H(X_2|X_1) + \ldots + H(X_n|X_{n-1},\ldots,X_1) \tag{10}$$

The *entropy rate* or *entropy density* of a stochastic process $\{X_i\}$ is defined by

$$\begin{aligned} h &= \lim_{n\to\infty} \frac{1}{n} H(X_1,X_2,\ldots,X_n) \\ &= \lim_{n\to\infty} H(X_n|X_{n-1},\ldots,X_1) \end{aligned} \tag{11}$$

representing the *average information content* per output symbol † [CT91]. It is the "uncertainty associated with a given symbol if all the preceding symbols are known" and can be viewed as "the intrinsic *unpredictability*" or "the irreducible *randomness*" associated with the chain [FC98].

In particular, a Markov chain can be considered as a chain of random variables complying with

$$H(X_n|X_1,X_2,\ldots,X_{n-1}) = H(X_n|X_{n-1}) \tag{12}$$

An important result is the following theorem: For a stationary Markov chain, with stationary distribution w_i, the entropy rate or information content is given by

$$\begin{aligned} h &= \lim_{n\to\infty} \frac{1}{n} H(X_1,X_2,\ldots,X_n) \\ &= \lim_{n\to\infty} H(X_n|X_{n-1}) \\ &= H(X_2|X_1) = -\sum_{i=1}^{n} w_i \sum_{j=1}^{n} P_{ij} \log P_{ij} \end{aligned} \tag{13}$$

where w_i is the equilibrium distribution and P_{ij} is the transition probability from state i to state j.

Finally, the *excess entropy* or *effective measure complexity* [CP83, Gra86, Sha84, SG86] of an infinite chain is defined by

$$E = \lim_{n\to\infty} (H(X_1,X_2,\ldots,X_n) - nh) \tag{14}$$

where h is the entropy rate of the chain and n is the length of this chain. The excess entropy can be interpreted as the mutual information between two semi-infinite halves of the chain. "Another way of viewing this, is that excess entropy is the *cost of amnesia* – the excess entropy measures how much more random the system would become if we suddenly forgot all information about the left half of the string" [Fel97].

5. Important Inequalities

Some of the above properties can be deduced from the following inequalities [CT91].

Jensen's inequality

A function $f(x)$ is *convex* over an interval (a,b) (the graph of the function lies below any chord) if for every $x_1, x_2 \in (a,b)$ and $0 \leq \lambda \leq 1$,

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \tag{15}$$

A function is strictly convex if equality holds, only if $\lambda = 0$ or $\lambda = 1$. A function $f(x)$ is *concave* (the graph of the function lies above any chord) if $-f(x)$ is convex.

For instance, $x \log x$ for $x \geq 0$ is a strictly convex function, and $\log x$ for $x \geq 0$ is a strictly concave function [CT91].

Jensen's inequality: If f is convex on the range of a random variable X, then

$$f(E[X]) \leq E[f(X)] \tag{16}$$

where E denotes expectation. Moreover, if $f(x)$ is strictly

† At least, h exists for all stationary stochastic processes.

convex, the equality implies that $X = E[X]$ with probability 1, i.e., X is a deterministic random variable with $Pr[X = x_0] = 1$ for some x_0.

One of the most important consequences of Jensen's inequality is the information inequality $D_{KL}(p\|q) \geq 0$. Other previous properties can also be derived from this inequality.

Observe that if $f(x) = x^2$ (convex function), then $E[X^2] - (E[X])^2 \geq 0$. So, the variance is invariably positive.

If f is given by the Shannon entropy, which is a concave function, we obtain the *Jensen-Shannon inequality* [BR82]:

$$JS(\pi_1, \pi_2, \ldots, \pi_N; p_1, p_2, \ldots, p_N) \equiv H\left(\sum_{i=1}^{N} \pi_i p_i\right) - \sum_{i=1}^{N} \pi_i H(p_i) \geq 0, \quad (17)$$

where $JS(\pi_1, \pi_2, \ldots, \pi_N; p_1, p_2, \ldots, p_N)$ is the *Jensen-Shannon divergence* of probability distributions $p_1, p_2, \ldots, p_N$ with prior probabilities or weights $\pi_1, \pi_2, \ldots, \pi_N$, fulfilling $\sum_{i=1}^{N} \pi_i = 1$. The JS-divergence measures how 'far' are the probabilities p_i from their likely joint source $\sum_{i=1}^{N} \pi_i p_i$ and equals zero if and only if all the p_i are equal. It is important to note that the JS-divergence is identical to $I(X,Y)$ when $\pi_i = p(x_i)$ and $p_i = p(Y|x_i)$ for each $x_i \in \mathcal{X}$, where $p(X) = \{p(x_i)\}$ is the input distribution, $p(Y|x_i) = \{p(y_1|x_i), p(y_2|x_i), \ldots, p(y_M|x_i)\}$, $N = |\mathcal{X}|$, and $M = |\mathcal{Y}|$ [BR82, ST00]. This notation is followed in some parts of the tutorial.

The log-sum inequality

Log-sum inequality: If $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ are non-negative numbers, then

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i\right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \quad (18)$$

with equality if and only if $\frac{a_i}{b_i} = constant$.

Note that the conditions in this inequality are much weaker than for Jensen's inequality.

From this inequality, certain results can be derived:

1. $D_{KL}(p\|q)$ is convex in the pair (p,q)
2. $H(X)$ is a concave function of p
3. If X and Y have the joint pdf $p(x,y) = p(x)p(y|x)$, then $I(X,Y)$ is a concave function of $p(x)$ for fixed $p(y|x)$ and a convex function of $p(y|x)$ for fixed $p(x)$.

Data processing inequality

Data processing inequality: If $X \rightarrow Y \rightarrow Z$ is a Markov chain, then

$$I(X,Y) \geq I(X,Z) \quad (19)$$

This result demonstrates that no processing of Y, deterministic or random, can increase the information that Y contains about X.

Fano's inequality

Suppose we have two correlated random variables X and Y and we wish to measure the probability of error in guessing X from the knowledge of Y. Fano's inequality gives us a tight lower bound on this error probability in terms of the conditional entropy $H(X|Y)$ [CT91, FM94]. As $H(X|Y)$ is zero if and only if X is a function of Y, we can estimate X from Y with zero probability of error if and only if $H(X|Y) = 0$. Intuitively, we expect to be able to estimate X with a low probability of error if and only if $H(X|Y)$ is small [CT91].

If X and Y have the joint pdf $p(x,y) = p(x)p(y|x)$, from Y we calculate a function $g(Y) = \widehat{X}$ which is an estimate of X. Observe that $X \rightarrow Y \rightarrow \widehat{X}$ is a Markov chain. The probability of error is defined by

$$P_e = Pr[\widehat{X} \neq X] \quad (20)$$

Fano's inequality:

$$H(P_e) + P_e \log n \geq H(X|Y) \quad (21)$$

where $H(P_e)$ is the binary entropy from $\{P_e, 1 - P_e\}$.

This inequality can be weakened to

$$P_e \geq \frac{H(X|Y) - 1}{\log n} \quad (22)$$

Thus, Fano's inequality bounds the probability that $\widehat{X} \neq X$.

6. Entropy and Coding

Other ways of interpreting the Shannon entropy are possible:

- As we have seen in section 2, $-\log p_i$ represents the *information* associated with the result x_i. But $-\log p_i$ can also be interpreted as the *surprise* associated with the outcome x_i. If p_i is small, the surprise is large; if p_i is large, the surprise is small. Thus, the entropy

  $$H(X) = -\sum_{i=1}^{n} p_i \log p_i$$

  is the expectation value of the surprise [Fel97].
- Entropy is also related to the difficulty in guessing the outcome of a random variable. Thus, it can be seen [CT91, Fel97] that

  $$H(X) \leq \overline{questions} < H(X) + 1 \quad (23)$$

  where $\overline{questions}$ is the average minimum number of binary questions to determine X. This idea agrees with the interpretation of entropy as a measure of uncertainty and also with the next interpretation.
- A fundamental result of information theory is the Shannon source coding theorem, which deals with the encoding of an object in order to store or transmit it efficiently [CT91, Fel97]. "Data compression can be achieved by assigning short descriptions to the most frequent outcomes

of the data source and necessarily longer descriptions to the less frequent outcomes" [CT91]. For instance, Huffman instantaneous coding [‡] is optimal and fulfils the following theorems:

– Similarly to (23), we have

$$H(X) \leq \overline{\ell} < H(X) + 1 \tag{24}$$

where $\overline{\ell}$ is the expected length of the optimal binary code for X.

– If we encode n identically distributed random variables X with a binary code, the Shannon source coding theorem can be enunciated in the following way:

$$H(X) \leq \overline{\ell_n} < H(X) + \frac{1}{n} \tag{25}$$

where $\overline{\ell_n}$ is the expected codeword length per unit symbol. Thus, by using large block lengths, we can achieve an expected codelength per symbol arbitrarily close to the entropy [CT91].

– For a stationary stochastic process, we have

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} \leq \overline{\ell_n} < \frac{H(X_1, X_2, \ldots, X_n)}{n} + 1 \tag{26}$$

and thus, by definition of entropy rate h (11),

$$\lim_{n\to\infty} \overline{\ell_n} \to h \tag{27}$$

Thus, the entropy rate is the expected number of bits per symbol required to describe the stochastic process.

In conclusion, the entropy of a random variable is a measure of the amount of information required on average to describe it.

7. Continuous Channel

In this section, entropy and mutual information are defined for continuous sources of information. For a continuous source X, messages are taken from a continuous set S.

The entropy of a discrete set of probabilities p has been defined (2) as

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \tag{28}$$

Similarly, the *continuous entropy* of a continuous random variable X with a probability density function $p(x)$ is defined by

$$H^c(X) = -\int_S p(x) \log p(x) dx \tag{29}$$

In the same way, for two continuous random variables X and Y, the continuous conditional entropy is defined as

$$H^c(X|Y) = -\int_S \int_S p(x,y) \log p(x|y) dxdy \tag{30}$$

and the continuous mutual information is defined as

$$I^c(X,Y) = \int_S \int_S p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy \tag{31}$$

where $p(x|y)$ and $p(x,y)$ are, respectively, the conditional density function and the joint density function associated with X and Y.

If we divide the range of the continuous random variable X into n bins of length Δ, and we consider its discretised version X^Δ (see [CT91]), it can be seen that the entropy of a continuous random variable does not equal the entropy of the discretised random variable in the limit of a finer discretisation [Sha48, CT91, Fel97]:

$$\lim_{\Delta\to 0} H(X^\Delta) = H^c(X) - \log \Delta \tag{32}$$

On the other hand, the mutual information between two continuous random variables X and Y is the limit of the mutual information between their discretised versions. Thus, when the number of bins tends to infinity:

$$\lim_{\Delta\to 0} I(X^\Delta, Y^\Delta) = I^c(X,Y) \tag{33}$$

In addition, Kolmogorov [Kol56] and Pinsker [Pin60] defined mutual information as $I(X,Y) = sup_{P,Q} I([X]_P, [Y]_Q)$, where the supremum is over all finite partitions P and Q. From this definition, two important properties can be deduced: *the continuous mutual information is the least upper bound for the discrete mutual information* and *refinement can never decrease the discrete mutual information* [Gra90]. This last property can also be deduced from the data processing inequality (19) [Gra90].

[‡] A code is called a prefix or instantaneous code if no codeword is a prefix of any other codeword.

References

[Bla87] BLAHUT R. E.: *Principles and Practice of Information Theory*. Addison-Wesley, Reading (MA), USA, 1987.

[BR82] BURBEA J., RAO C. R.: On the convexity of some divergence measures based on entropy functions. *IEEE Transactions on Information Theory 28*, 3 (May 1982), 489–495.

[CP83] CRUTCHFIELD J. P., PACKARD N.: Symbolic dynamics of noisy chaos. *Physica 7D* (1983), 201–223.

[CT91] COVER T. M., THOMAS J. A.: *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.

[FC98] FELDMAN D. P., CRUTCHFIELD J. P.: *Discovering Noncritical Organization: Statistical Mechanical, Information Theoreticand Computational Views of Patterns in One-Dimensional Spin Systems*. Working Paper 98–04–026, Santa Fe Institute, Santa Fe (NM), USA, April 1998.

[Fel97] FELDMAN D. P.: A brief introduction to: Information theory, excess entropy and computational mechanics, 1997.

[FM94] FEDER M., MERHAV N.: Relations between entropy and error probability. *IEEE Transactions on Information Theory 40*, 1 (January 1994), 259–266.

[Gra86] GRASSBERGER P.: Toward a quantitative theory of self-generated complexity. *International Journal of Theoretical Physics 25*, 9 (1986), 907–938.

[Gra90] GRAY R. M.: *Entropy and Information Theory*. Springer-Verlag Vienna-New York, New York (NY), USA, 1990.

[Kol56] KOLMOGOROV A. N.: On the shannon theory of information transmission in the case of continuous signals. *IRE Transactions on Information Theory 2* (1956), 102–108.

[Pin60] PINSKER M. S.: *Information and Stability of Random Variables and Processes*. Izdatel'stvo Akademii Nauk SSSR, Moscow, Russia, 1960. Translated by A. Feinstein, 1964.

[SG86] SZÉPFALUSY P., GYÖRGYI G.: Entropy decay as a measure of stochasticity in chaotic systems. *Physical Review A 33*, 4 (1986), 2852.

[Sha48] SHANNON C. E.: A mathematical theory of communication. *The Bell System Technical Journal 27* (July, October 1948), 379–423, 623–656.

[Sha84] SHAW R.: *The Dripping Faucet as a Model Chaotic System*. Aerial Press, Santa Cruz (CA), USA, 1984.

[ST00] SLONIM N., TISHBY N.: Document clustering using word clusters via the information bottleneck method. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (2000), ACM Press, pp. 208–215. Held in Athens, Greece.

[vdL97] VAN DER LUBBE J. C.: *Information Theory*. Cambridge University Press, Cambridge, UK, 1997.

Applications of Information Theory to Computer Graphics Part III: Refinement Criteria for Radiosity

Miquel Feixas, Jaume Rigau, and Mateu Sbert

University of Girona, Spain

1. Radiosity Method

The radiosity method, first introduced in [GTGB84, NN85, CG85], solves the problem of illumination in an environment of diffuse surfaces. In this section, we look at the radiosity equation, the form factor computation and some refinement criteria for hierarchical radiosity.

1.1. Rendering Equation

The light transport in a virtual closed environment or scene (Figure 1) is described by the *rendering equation* [Kaj86], which is a second-order Fredholm integral equation. This equation, which describes all energy exchanges between surfaces, gives us the distribution of light at every point of a scene:

$$L(x,\omega_x) = L_e(x,\omega_x) + \int_S \rho_{bd}(x,\omega_x,-\omega_{y\to x})L(y,\omega_y)G(x,y)dA_y \quad (1)$$

where

- S is the set of surfaces that form the environment
- x and y are points on S
- dA_y is a differential area at point y
- ω_x is a given outgoing direction at point x and $\omega_{y\to x}$ is the outgoing direction at point y towards point x ($\omega_{y\to x}$ can also be seen as an incoming direction at point x coming from point y) (Figure 2)
- $L(x,\omega_x)$ is the *radiance* at point x in direction ω_x (radiance can be defined as the power arriving at or leaving from a surface per unit solid angle and per unit projected area, $\frac{W}{sr.m^2}$) and $L(y,\omega_{y\to x})$ is the radiance at point y in direction $\omega_{y\to x}$
- $L_e(x,\omega_x)$ is the *emitted radiance* at point x in direction ω_x
- $\rho_{bd}(x,\omega_x,-\omega_{y\to x})$, with units sr^{-1}, is the *bidirectional reflectance distribution function* (BRDF) at point x, which is the ratio between the outgoing radiance at x in direction ω_x and the incident radiant flux density (irradiance, $\frac{W}{m^2}$) at x from direction $\omega_{y\to x}$ (Figure 3a)

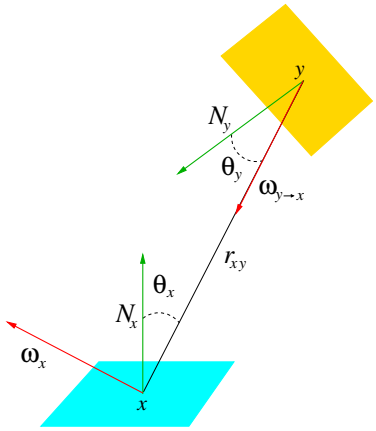


Figure 2: *Outgoing and incoming directions at point x.*

- $G(x,y)$ is the *geometric kernel*, equal to $\frac{\cos\theta_x \cos\theta_y}{r_{xy}^2}V(x,y)$, where θ_x and θ_y are the angles that the line joining x and y form with the normals at x and y respectively, r_{xy} is the distance between x and y, and $V(x,y)$ is a visibility function which is equal to 1 if x and y are mutually visible and 0 if not

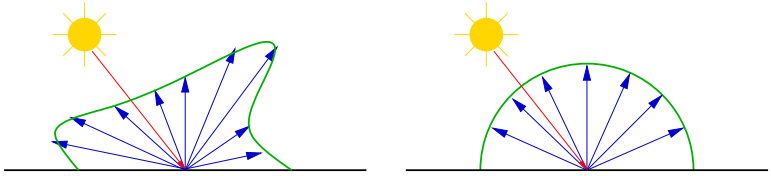
This equation can be presented in slightly different forms (global illumination equation, radiance equation) [SP94, Gla95]. Observe that "the radiance distribution L is described implicitly, so we know what conditions it must satisfy, but we don't know what it actually is" [Gla95].

1.2. Continuous Radiosity Equation

For diffuse surfaces, the BRDF does not depend on the outgoing and incoming directions. Thus, the outgoing radiance $L(x,w_x)$ and the self-emitted radiance $L_e(x,w_x)$ are also independent of the outgoing direction (Figure 3b). From this simplification, the *rendering equation for diffuse surfaces*

(a) (b)

Figure 1: *Two different illuminated scenes.*



(a) (b)

Figure 3: *(a) Bidirectional reflectance distribution function. (b) Diffuse reflectance.*

can be expressed as

$$L(x) = L_e(x) + \int_S \rho_{bd}(x) L(y) G(x,y) dA_y \tag{2}$$

If we integrate $L(x)$ on the whole hemisphere Ω_x of the outgoing directions w_x at point x, we obtain the total outgoing flux over the hemisphere per unit area, called the *radiosity* at point x (power per unit area) [SP94, Gla95]:

$$B(x) = \int_{\Omega_x} L(x) \cos\theta_x d\omega_x = \pi L(x) \tag{3}$$

where $d\omega_x$ is the differential solid angle containing the direction ω_x and θ_x is the angle that the direction ω_x forms with the normal at x. In addition, the total self-emitted flux per unit area is expressed by $E(x) = \pi L_e(x)$, and is called the *emittance* at point x.

Note that $\rho_{bd}(x)$ is the ratio of outgoing radiance to incoming flux density. A more convenient quantity is the ratio of reflected to incoming total flux, which must be between 0 and 1 according to the energy conservation law (the energy reflected must be a fraction of the energy received, the other fraction is absorbed). This ratio is the diffuse reflectance, or simply *reflectance*, and is given by $\rho(x) = \pi \rho_{bd}(x)$ (Figure 3b).

The *radiosity equation* is then obtained by multiplying both sides of equation (2) by π:

$$B(x) = E(x) + \frac{\rho(x)}{\pi} \int_S B(y) G(x,y) dA_y \tag{4}$$

where

- $B(x)$ and $B(y)$ are, respectively, the radiosities at points x and y ($\frac{W}{m^2}$)
- $E(x)$ is the emittance or emitted flux of energy per unit area at point x ($\frac{W}{m^2}$)
- $\rho(x)$ is the diffuse reflectance at point x (dimensionless)

The radiosity equation can also be written in a directional form [SP94]:

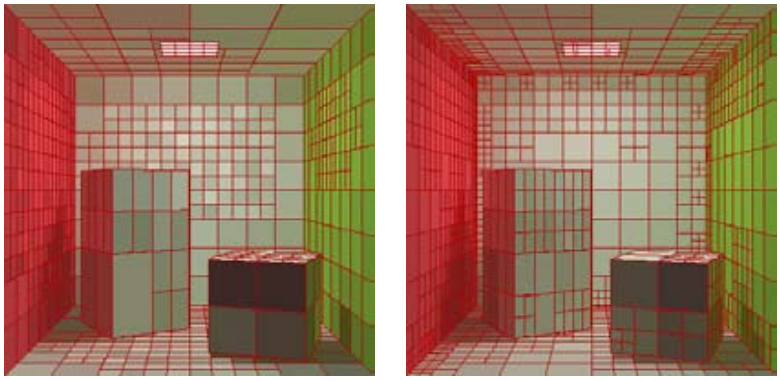
$$B(x) = E(x) + \frac{\rho(x)}{\pi} \int_{\Omega_x} B(y) \cos\theta_x d\omega_x \tag{5}$$

In this conversion, $d\omega_x = \frac{\cos\theta_y}{r_{xy}^2} dA_y$ has been used.

1.3. Discrete Radiosity Equation and Form Factors

To solve the radiosity equation we can use a finite element approach, discretising the environment into n_p patches and considering the radiosities, emissivities and reflectances constant over the patches (Figure 4).

With these assumptions, the integral equation (4) becomes



(a) (b)

Figure 4: *A scene with two different discretisations.*

the system of radiosity equations [GTGB84]:

$$B_i = E_i + \rho_i \sum_{j=1}^{n_p} F_{ij} B_j \tag{6}$$

where

- B_i, E_i, and ρ_i are respectively the radiosity, emittance (or emissivity), and reflectance of patch i, and B_j is the radiosity of patch j
- F_{ij} is the *patch-to-patch form factor*, only dependent on the geometry of the scene:

$$\begin{aligned} F_{ij} &= \frac{1}{A_i}\int_{A_i}\int_{A_j}\frac{G(x,y)}{\pi}dA_y dA_x \\ &= \frac{1}{A_i}\int_{A_i}\int_{A_j}\frac{\cos\theta_x\cos\theta_y}{\pi r_{xy}^2}V(x,y)dA_y dA_x \\ &= \frac{1}{A_i}\int_{A_i}\int_{A_j}F(x,y)dA_y dA_x \\ &= \frac{1}{A_i}\int_{A_i}\int_{\Omega_{x\to j}}\frac{\cos\theta_x}{\pi}V(x,y)d\omega_x dA_x \end{aligned} \tag{7}$$

  where A_i and A_j represent, respectively, the surfaces and also the areas of patches i and j, x and y are, respectively, points on A_i and A_j, $F(x,y) = \frac{\cos\theta_x\cos\theta_y}{\pi r_{xy}^2}V(x,y)$ is the *point-to-point form factor*, and $\Omega_{x\to j}$ represents the set of directions going from x to patch j

Form factor properties

Form factors have the following properties:

- Reciprocity

$$A_i F_{ij} = A_j F_{ji} \quad \forall i,j \tag{8}$$

- Energy conservation

$$\sum_{j=1}^{n_p} F_{ij} = 1 \quad \forall i \tag{9}$$

- Additivity

$$F_{i(k\cup l)} = F_{ik} + F_{il} \tag{10}$$

  where i, k, and l are three disjoint patches. In general the reverse is not true

$$F_{(k\cup l)i} \neq F_{ki} + F_{li} \tag{11}$$

  In fact, if the patch i is divided into n_i subpatches, we obtain

$$\sum_{k=1}^{n_i} A_{i_k} F_{i_k j} = A_i F_{ij} \tag{12}$$

  As a direct consequence of this equation, if patch i is divided into n_i subpatches of equal area, we have

$$n_i F_{ij} = \sum_{k=1}^{n_i} F_{i_k j} \tag{13}$$

  or

$$F_{ij} = \frac{1}{n_i}\sum_{k=1}^{n_i} F_{i_k j} \tag{14}$$

  In this case, F_{ij} is the average of the form factors between the subpatches of i and patch j.

Differential-area-to-area form factor

The form factor integral (7) can be considered as an average over the area of patch i of the inner integral. Thus, we have

$$F_{ij} = \frac{1}{A_i}\int_{A_i} F_{dA_x,A_j} dA_x \tag{15}$$

where F_{dA_x,A_j} is the *differential-area-to-area form factor* and is equal to

$$\begin{aligned} F_{dA_x,A_j} &= \int_{A_j}\frac{\cos\theta_x\cos\theta_y}{\pi r_{xy}^2}V(x,y)dA_y \\ &= \int_{\Omega_{x\to j}}\frac{\cos\theta_x}{\pi}V(x,y)d\omega_x \end{aligned} \tag{16}$$

If patches i and j are very distant from each other, then we can assume that F_{dA_x,A_j} is constant over patch i. So, we can evaluate this integral only at a point c (usually the center) of patch i, obtaining

$$\begin{aligned} F_{ij} &= \frac{1}{A_i}\int_{x\in A_i} F_{dA_x,A_j}\,dA_x \\ &\approx F_{dA_c,A_j}\frac{1}{A_i}\int_{x\in A_i} dA_x = F_{dA_c,A_j} \end{aligned} \tag{17}$$

The differential-area-to-area form factor is also called *point-to-patch form factor* and can be thought of as the limit of the patch-to-patch form factor when the area of one of the patches decreases to zero [SP94]. We define $F_j(x) \equiv F_{dA_x,A_j}$.

1.4. Form Factor Computation

The form factor computation is the most costly step of the radiosity method. More specifically, its cost is mainly due to the presence of the visibility term in the geometric kernel.

Analytical and deterministic numerical solutions

No analytical closed-form solution exists except for very simple shapes without occlusions. Schroeder and Hanrahan [SH93] solved the polygon-to-polygon case. In [SH92, Gla95] there is an extensive list of formulae for simple shapes. Here we only review one of them: the form factor between two patches of the interior of a sphere without occlusion, which is a paradigmatic case in our work.

From Figure 5, it can be easily obtained that the form factor F_{ij} between two spherical patches i and j is equal to $\frac{A_j}{A_S}$, where A_S is the area of the sphere. As θ_x and θ_y are equal and $\frac{\cos\theta_x}{r_{xy}} = \frac{\cos\theta_y}{r_{xy}} = \frac{1}{2R}$, where R is the radius of the sphere, the expression for the form factor becomes

$$\begin{aligned} F_{ij} &= \frac{1}{A_i}\int_{A_i}\int_{A_j} \frac{\cos\theta_x\cos\theta_y}{\pi r_{xy}^2} V(x,y)\,dA_x dA_y \\ &= \frac{1}{\pi A_i}\int_{A_i}\int_{A_j} \frac{\cos\theta_x}{r_{xy}}\frac{\cos\theta_y}{r_{xy}} V(x,y)\,dA_x dA_y \\ &= \frac{1}{4\pi R^2 A_i}\int_{A_i}\int_{A_j} dA_x dA_y = \frac{A_j}{A_S} \end{aligned} \tag{18}$$

Note also that $F(x,y) = \frac{\cos\theta_x\cos\theta_y}{\pi r_{xy}^2} = \frac{1}{A_S}$ in a spherical scene.

When occlusions between patches exist, we can use deterministic numerical approximations to compute the form factors. Different methods can be found in the literature [SP94, Gla95]: Wallace's method, Nusselt's analogy, hemicube method.

Monte Carlo integration

In this section, the form factor integral (7) will be evaluated by the Monte Carlo method. So, we will give a brief overview of this technique. For a more detailed description,

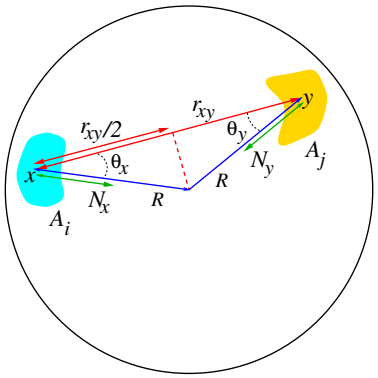


Figure 5: *Geometry for the form factor between two spherical patches.*

see [KW86]. Also [SP94, Gla95, Bek99] review it and give different techniques to sample a random variable.

Monte Carlo integration enables us to estimate integrals using random techniques. Let us suppose we want to solve the integral of a function $g(x)$. This can be written as

$$I = \int_D g(x)dx = \int_D \frac{g(x)}{f(x)} f(x)dx \tag{19}$$

If $f(x) > 0$ ($\forall x \in D$) and $\int_D f(x)dx = 1$, then $f(x)$ can be considered as a *probability density function* (pdf) of a random variable X and the integral (19) can be read as the expected value of the random variable $\frac{g(X)}{f(X)}$ with respect to the pdf $f(x)$:

$$I = E_f\left[\frac{g(X)}{f(X)}\right] \tag{20}$$

The term $\frac{g(x_1)}{f(x_1)}$, where x_1 is obtained by sampling from the pdf $f(x)$, is a *primary* estimator for the integral I:

$$I \approx \widehat{I} = \frac{g(x_1)}{f(x_1)} \tag{21}$$

This estimator is unbiased, i.e., the expected value of this estimator is the value of the integral: $E[\widehat{I}] = I$. The variance of this estimator is given by

$$V[\widehat{I}] = E\left[\left(\frac{g(X)}{f(X)}\right)^2\right] - \left(E\left[\frac{g(X)}{f(X)}\right]\right)^2 = \int_D \frac{g(x)^2}{f(x)}dx - I^2 \tag{22}$$

Averaging N independent primary estimators (obtained by sampling N independent values $x_1, x_2, \ldots, x_N$ from $f(x)$), we obtain the unbiased *secondary* estimator $\widehat{I}_N$

$$I \approx \widehat{I}_N = \frac{1}{N}\sum_{k=1}^{N}\frac{g(x_k)}{f(x_k)} \tag{23}$$

with variance

$$V[\widehat{I_N}] = \frac{1}{N} V[\widehat{I}] = \frac{1}{N}\left(\int_D \frac{g(x)^2}{f(x)} dx - I^2\right) \qquad (24)$$

So, we obtain better estimators as the number of samples increases. This result is according to the *weak law of large numbers*, which states that, for identically independent distributed (i.i.d.) random variables, $\frac{1}{N}\sum_{k=1}^{N} X_k$ is close to its expected value $E[X]$ for large numbers of N. Obviously the variance depends on the pdf chosen. When we use a pdf that resembles the integrand we are doing *importance sampling*, which can reduce dramatically the variance of our estimator [KW86].

With respect to the variance, let us remember that the standard deviation of X, which represents the error, is given by $\sigma = \sqrt{V[X]}$. It can be observed from (24) that σ decreases at a rate of $\frac{1}{\sqrt{N}}$ as the number of samples increases.

The mean square error MSE of an estimator $\widehat{\theta}$ of θ is given by

$$MSE(\widehat{\theta}) = E\left[\left(\widehat{\theta} - \theta\right)^2\right] \qquad (25)$$

and it is equal to the variance when the estimator is unbiased.

Monte Carlo form factor computation

Three different ways of computing the patch-to-patch form factor F_{ij} are here reviewed (for a brief survey, see [Bek99]). Uniform area sampling and uniformly distributed local and global lines can be used to estimate the form factors:

- **Uniform area sampling**
  To calculate the patch-to-patch form factor (7)

$$F_{ij} = \frac{1}{A_i}\int_{A_i}\int_{A_j} \frac{\cos\theta_x \cos\theta_y}{\pi r_{xy}^2} V(x,y) dA_x dA_y$$

  we take random points x and y on patches i and j respectively (Figure 6). This means taking as pdf $f(x,y) = \frac{1}{A_i A_j}$, which is a uniform distribution.
  A primary estimator $\widehat{F_{ij}^1}$ is given by

$$\widehat{F_{ij}^1} = \frac{1}{A_i}\frac{F(x,y)}{f(x,y)} = \frac{1}{A_i}\frac{F(x,y)}{\frac{1}{A_i A_j}} = A_j F(x,y) \qquad (26)$$

  where (x,y) is a pair of random points. It is easy to see that this estimator is unbiased ($E[\widehat{F_{ij}^1}] = F_{ij}$) and its variance is given by

$$V[\widehat{F_{ij}^1}] = \int_{A_i}\int_{A_j} (A_j F(x,y))^2 \frac{1}{A_i A_j} dA_x dA_y - F_{ij}^2 \qquad (27)$$

  We can see that a strong singularity for abutting patches is produced due to the term r_{xy}^4 in the denominator of the integrand. In this case, this technique is not satisfactory, as the variance can be very high [Bek99].

For N samples of pairs (x,y), the form factor integral is approximated by the secondary estimator:

$$\widehat{F_{ij}^1} = A_j \frac{1}{N}\sum_{k=1}^{N} F(x_k, y_k) \qquad (28)$$

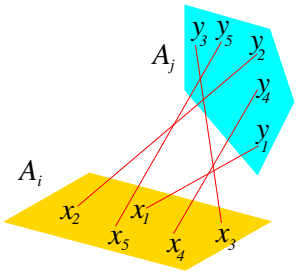


Figure 6: *Form factor F_{ij} can be computed by taking random points on patches i and j.*

- **Uniformly distributed lines**
  - **Local lines** A local line is a ray with its origin uniformly distributed on the surface of i and its direction distributed according to the cosine with respect to the normal at the origin. So, we estimate the integral (7)

$$F_{ij} = \frac{1}{A_i}\int_{A_i}\int_{\Omega_{x\to j}} \frac{\cos\theta_x}{\pi} V(x,y) dA_x d\omega_x$$

    taking as pdf $f(x,w_x) = \frac{1}{A_i}\frac{\cos\theta_x}{\pi}$.
    An unbiased primary estimator $\widehat{F_{ij}^2}$ takes the value 1 if the local line hits the patch j directly and 0 if not. Let us recall that if a random variable X takes the values 1 and 0 with probabilities p and $1-p$, its variance is given by $V[X] = p(1-p)$ [Pap84]. Thus,

$$V[\widehat{F_{ij}^2}] = F_{ij}(1 - F_{ij}) \qquad (29)$$

    A secondary estimator for F_{ij} is given by

$$\widehat{F_{ij}^2} = \frac{N_{ij}}{N_i} \qquad (30)$$

    where N_i is the number of local lines with origin on i and N_{ij} is the number of local lines with origin on i that hit j. It shows clearly that the form factor F_{ij} can be interpreted as the fraction of local lines with origin on i that have j as the nearest patch intersected (Figure 7a).
  - **Global lines**
    Global lines [Sbe93] can be generated by putting the scene within a sphere and selecting pairs of random points on the surface of this sphere. The lines connecting each pair of points are uniformly distributed throughout the scene. So, the form factor F_{ij} can also be considered as the probability of a global line that, crossing i, hits j (Figure 7b). It can be shown that each

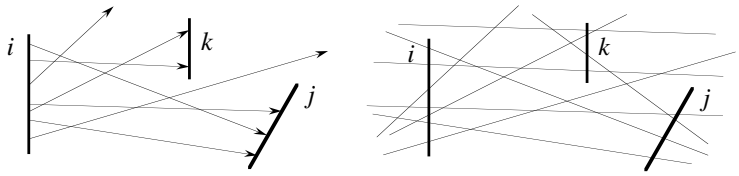


Figure 7: *Form factors can be computed with local (a) and global lines (b).*

line can contribute to the computation of several form factors (Figure 8).

Also, it is important to note that, from integral geometry [San76, Sbe96], the probability that, for a planar patch, a global line intersects patch i is proportional to A_i.

A secondary estimator for F_{ij} is given by

$$\widehat{F_{ij}^3} = \frac{N_{ij}}{N_i} \tag{31}$$

where N_i is the number of global lines which cross i and N_{ij} is the number of global lines that, crossing i, hit j. Its variance is

$$V[\widehat{F_{ij}^3}] = \frac{1}{N_i} F_{ij}(1 - F_{ij}) \tag{32}$$

To sample with global lines is equivalent to casting, for each patch, a number of local lines proportional to its area.

Observe that the variance will be higher for smaller patches as N_i is proportional to A_i. If we identify the lines connecting two patches with visibility, the form factor gives us the *visibility* between patches [Sbe96].

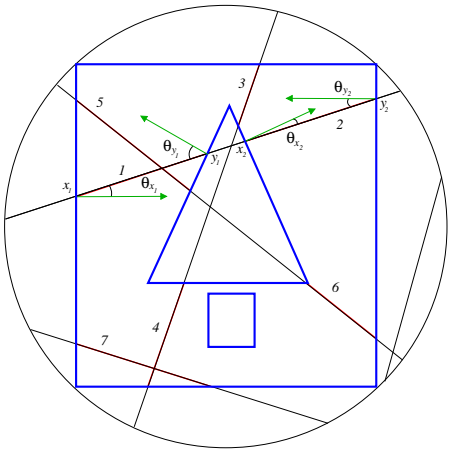


Figure 8: *Each segment of a global line contributes to the computation of two form factors. Thus, the depicted line is used in four form factor computations.*

1.5. Solution to the Radiosity Equation

The classic radiosity method consists of the following steps [CW93, SP94, Gla95]:

- Discretise the environment into patches
- Compute the form factors F_{ij} for each pair of patches i and j (form factor matrix)
- Solve the system of linear equations
- Display the solution

In this method, the input data are the geometric information about the scene (for the form factors), the physical properties of the materials (for the emissivities and reflectances), and viewing conditions [SP94].

The radiosity equation (6), which refers to a single patch, can be expressed as a system of n_p linear equations. This linear system can be written in the form

$$\mathbf{B} = \mathbf{E} + \mathbf{RB} \tag{33}$$

where $\mathbf{B}$ and $\mathbf{E}$ are, respectively, the vectors of radiosities and emittances, and $\mathbf{R}$ is the $n_p \times n_p$ matrix of the terms $\rho_i F_{ij}$. The solution $\mathbf{B}$ of such a system can be written as a Neumann series. As ρ_i is strictly less than 1, the matrix $\mathbf{R}$ has a norm strictly less than 1. In this case, the Neumann series converges and we can write the radiosity vector as a sum of an infinite series:

$$\mathbf{B} = \mathbf{E} + \mathbf{RE} + \mathbf{R}^2\mathbf{E} + \cdots + \mathbf{R}^n\mathbf{E} + \cdots \tag{34}$$

Since $\mathbf{R}$ represents the effect of one reflection on all the surfaces of the scene, $\mathbf{R}^n\mathbf{E}$ can be interpreted as the radiosity obtained after n rebounds of the emitted light through the scene.

In the literature, different iterative solution methods [SP94, Gla95] are available to solve the radiosity and power systems: Jacobi relaxation, Gauss Seidel relaxation, Southwell relaxation, and also their respective stochastic versions [Shi90, SP94, NFP95, Neu95, NNB97, Bek99].

1.6. Random Walks and Markov Chains

A discrete *random walk* [KW86, Rub81] is a Monte Carlo technique used to solve linear systems of equations. A *random walk* in a scene can be considered as a *Markov chain* [Col74, CT91, MR95]. This is a discrete stochastic process defined over a set of states $S = 1, 2, \ldots, n$ which is described by a *transition probability matrix* $\mathbf{P}$. In each step, the imaginary particle (or ray) makes a transition from its current state i to a new state j with *transition probability* P_{ij}. The transition probabilities only depend on the current state. A Markov chain can also be seen as a sequence of random variables $X_k, k = 0 \ldots \infty$, in which each $X_k, k \geq 1$, depends only on the previous X_{k-1} and not on the ones before. The random variables X_k indicate the probability of finding an imaginary particle in each state i after k steps from an initial distribution given by X_0.

Thus, for all $i, j \in S$, we have $\sum_{j=1}^{n} P_{ij} = 1$. Also, if we are in state i, the probability of being in state j after n steps is $(\mathbf{P}^n)_{ij}$ or P_{ij}^n. Under certain conditions, which are met in the context of this dissertation, the probabilities of finding the particle in each state i converge to a *stationary distribution* $\mathbf{w} = \{w_1, \ldots, w_n\}$ after a number of steps. The stationary or equilibrium probabilities w_i fulfil the relation $w_i = \sum_{j=1}^{n} w_j P_{ji}$ and also the reciprocity relation $w_i P_{ij} = w_j P_{ji}$.

A transition probability matrix is said to be aperiodic if it has no periodic state [Col74]. A periodic state is a state that can be visited back by a path starting from it only at multiples of a given period. A probability matrix is said to be irreducible if there is always a path between any two states.

The form factor matrix $\mathbf{F}$ meets all the required conditions to be a valid transition matrix of a random walk. The states of the random walk correspond to the patches of a scene and n_p denotes the number of patches. In order to determine the next state of a random walk, the form factors of the current patch need to be sampled. Such sampling can be carried out easily without having to compute the form factors explicitely [Shi90, PM92, FP93, Sbe97].

For the purpose of our work we are mainly interested in the following two properties [Sbe96]:

1. If the form factor matrix $\mathbf{F}$ is irreducible and aperiodic, then we have

$$\lim_{m \to \infty} (\mathbf{F}^m)_{ij} \to \frac{A_j}{A_T} = a_j \tag{35}$$

   for all the patches of a scene, where A_j is the area of patch j, $A_T = \sum_{i=1}^{n_p} A_i$, and a_j is the relative area of patch j. Thus, as the stationary or equilibrium distribution for a Markov chain is defined as the limit of the mth power of the transition matrix when m grows to infinity, if we know the areas of the patches, we also know the stationary distribution $\mathbf{a} = \{a_i\}$ of the random walk.
2. When the length of a random walk with transition matrix $\mathbf{F}$ grows to infinity, the number of hits on any patch i becomes proportional to a_i, independently of where the random walk started its trajectory.

A Markov chain with a stationary distribution is called ergodic. Thus, the form factors correspond to an ergodic Markov chain.

When the states form a countable set, as stated before, the Markov chain is called a *discrete* chain. When the states are not countable, the chain is called *continuous*. For instance, when taking infinitesimal areas dx at each point x on the surfaces S of the scene as the states and differential form factors $F(x, y)$, with $x, y \in S$ as transition probabilities, a continuous Markov chain with stationary distribution $w(x) = \frac{1}{A_T}$ results.

It can be shown that in flatland the stationary probabilities of the resulting discrete Markov chain are given by $w_i = \frac{L_i}{L_T} = l_i$, where L_T is the total length of all segments of the scene and L_i is the relative length of segment i. When taking infinitesimal lengths dx at each point x on the set of segments L of the scene as the states and differential form factors $F(x, y)$, with $x, y \in L$ as transition probabilities, a continuous Markov chain with stationary distribution $w(x) = \frac{1}{L_T}$ results.

1.7. Hierarchical Radiosity

The classic radiosity method means the entire matrix of form factors has to be computed before a solution can be obtained. This is the most costly step in all the process. To manage the complexity of this problem, different strategies can be used to reduce the number of form factors that need to be computed. Obviously, we also have to take into account the accuracy of the solution. A good strategy has to balance the reduction of the number of patches and the precision of the illumination.

Some techniques have been introduced in order to reduce the computational cost: progressive refinement, substructuring, adaptive refinement and hierarchical refinement. Other techniques try to reduce the number of form factors arriving at a solution within a given error bound [SP94].

The hierarchical refinement algorithm was first introduced in [HSA91] by Hanrahan and Salzman. Additional information can be found in [SP94, Gla95]. This algorithm is based on the objective of reducing the number of form factors needed to propagate the light through the environment: "At the first this may hardly seem possible; after all, the form factors describe the interaction of light between pairs of surfaces. How can we delete any of them and still hope to get an accurate solution?" [Gla95]. Hanrahan and Salzman observed that the N-body problem and the form factor problem were very similar. It is worth noting that both problems are based on the interaction between all pairs of objects and also that the gravitational force and the form factor have a similar mathematical expression. The idea behind both problems is that "small details don't matter when we are far away from

something" [Gla95]. Thus, the clustering algorithms of the N-body problem were applied to radiosity, resulting in the *hierarchical radiosity algorithm*.

If each of the N particles exerts interactions on the other $N-1$ particles, thus there exist $\mathcal{O}(N^2)$ interactions to account for. But two distant groups of particles can be considered, in terms of interaction, as two single particles. In hierarchical radiosity the particles are substituted by patches and these are subdivided into smaller elements if necessary, in order to achieve an accurate light transport between them. The main objective is to obtain an accurate piecewise constant approximation of the radiosity on all the elements. To do this, the mesh is generated adaptively: when a constant radiosity assumption on patch i is not valid for the radiosity due to another patch, the refinement algorithm will refine i in a set of subpatches or elements. Finally, a multiresolution element mesh will enable us to accurately represent the energy transport between patches [Gla95].

An oracle or refinement criterion, based on an error estimation, informs us if a subdivision of the surfaces is needed. The oracle takes geometrical and visibility information about the patches and also the source radiosity and receiver reflectance, and returns whether or not the interaction is valid. Some of them will need further refinement, until a certain level where no further refinement is needed or a previously imposed bound on the area of the patches is reached. Its cost should not make the method prohibitive.

Bekaert et al. [BNN*98] have incorporated hierarchical refinement in Monte Carlo radiosity (more specifically in stochastic Jacobi radiosity) by means of *per-ray* refinement.

1.8. Refinement Criteria

In this section, we review some refinement criteria for hierarchical radiosity [†]. The cheapest and most widely-used oracle has been the power-based oracle [HSA91]. However, it leads to unnecessary subdivisions in smoothly illuminated unoccluded regions receiving a lot of power. As an alternative, oracles based on the smoothness of the geometrical kernel and the received radiosity have been proposed [SAS92, GSCH93, LTG93, PB95, LSG94, BW96, SSS97, HS98]. Nevertheless, oracles based on kernel smoothness also have the problem of unnecessary subdivisions where the kernel is unbounded, and the ones based on received radiosity rely on a costly accurate computation of form factors. All in all, the additional cost invested in both smoothness-based oracles, mainly through visibility computations, may outweigh the improvements obtained [Bek99].

The application of a good refinement criterion and strategy is fundamental for the efficiency of the hierarchical refinement algorithm. Next we review some oracles proposed in the past.

[†] This section follows closely the discussion in [Bek99]

Oracle based on transported power

Hierarchical refinement radiosity was initially presented for constant radiosity approximations by Hanrahan et al. [HSA91]. A cheap form factor estimate F_{ij} which ignores visibility was used to measure the accuracy of an interaction from an element j to an element i. If $\max(F_{ij}, F_{ji})$ exceeds a given threshold ε, the larger of the two elements i and j is subdivided using regular quadtree subdivision. Otherwise, the candidate link is considered admissible.

Hanrahan et al. [HSA91] also observed that the number of form factors can be reduced considerably without affecting image quality by weighting the link error estimates F_{ij} with the source element radiosity B_j and receiver element area A_i. Weighting with receiver reflectance ρ_i also further reduces the number of links without deteriorating image quality. Thus, the power-based criterion to stop refinement can be given by

$$\rho_i A_i F_{ij} B_j < \varepsilon \tag{36}$$

Other strategies [TH93, FH96] can also be used to reduce the number of form factors by taking visibility information about candidate interactions into account. We can see that power-based refinement criterion uses no information about the variation of the received radiosity across the receiver element. This results, for instance, in sub-optimal shadow boundaries and excessively fine refinement in smooth areas. The main advantage of criterion (36) is its very low computational cost while yielding a fair image quality.

Oracle based on kernel smoothness

In order to improve on power-based refinement, the variation of the radiosity kernel between a pair of elements is taken into account. In [SAS92], the refinement criterion is given by

$$\rho_i (F_{ij}^{max} - F_{ij}^{min}) A_j B_j < \varepsilon \tag{37}$$

where $F_{ij}^{max} = max_{x \in A_i, y \in A_j} F(x,y)$ and $F_{ij}^{min} = min_{x \in A_i, y \in A_j} F(x,y)$ are the maximum and minimum radiosity kernel values and are estimated by taking the maximum and minimum value computed between pairs of random points on both elements (A_i and A_j represent the surfaces of the elements), ε is a predefined threshold, B_j is the source radiosity and ρ_i the receiver reflectivity.

A similar approach was used in [GSCH93] in order to drive hierarchical refinement with higher-order approximations. When applied to constant approximations, the refinement criterion is given by

$$\rho_i max(F_{ij}^{max} - F_{ij}^{av}, F_{ij}^{av} - F_{ij}^{min}) A_j B_j < \varepsilon \tag{38}$$

where $F_{ij}^{av} = F_{ij}/A_j$ is the average radiosity kernel value. Kernel variation is a sufficient condition for received radiosity variation, but not a necessary condition [Bek99].

Oracle based on smoothness of received radiosity

Because bounding kernel variation is not a necessary condition for bounding received radiosity variation, we can expect that hierarchical refinement based on kernel smoothness will yield hierarchical meshes with more elements and links than required. Optimal refinement can be expected by directly estimating how well the radiosity $B_j(x)$, received at $x \in A_i$ from A_j, is approximated by a linear combination of the basis functions on A_i, i.e., by estimating the discretisation error directly.

This approach was first proposed by Lischinski et al. [LTG93] for constant approximations. Pattanaik and Bouatouch [PB95] proposed a similar strategy for linear basis functions. Other approaches are given in [LSG94, BW96, SSS97, HS98]. The computation cost of kernel and radiosity smoothness-based oracles has not yet been found to compensate for the gain in mesh quality [Bek99].

2. The Scene as a Discrete Channel

In order to apply information theory tools to a scene, we *model* the scene in two equivalent ways:

- *A random walk*
  A discrete random walk (section 1.6) in a discretised scene is a discrete Markov chain where the transition probabilities are the form factors and the stationary distribution is given by the relative area of patches (Figure 9).
- *An information channel*
  A scene can be interpreted as a discrete information channel where the input and output variables take values over the set of patches (the relative area of patches is the probability distribution of these variables) and the channel transition matrix is the form factor matrix.

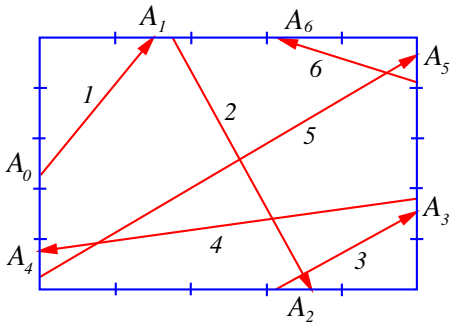


Figure 9: *Discrete random walk in a scene.*

In previous sections (section 1.6 and Part II), we reviewed the most basic concepts about a Markov chain and an information channel. Now, to work with a scene, the following mapping or transformation is done from those general definitions:

- For a discrete Markov chain:
  - number of states: $n \Longrightarrow$ number of patches: n_p
  - transition probability: $P_{ij} \Longrightarrow$ form factor: F_{ij}
  - stationary probability of state i: $w_i \Longrightarrow$ relative area of patch i: $a_i = \frac{A_i}{A_T}$
- For a discrete information channel with random variables X and Y:
  - probability distributions of X and Y: $\mathbf{p}$ and $\mathbf{q} \Longrightarrow$ relative area of patches: $\mathbf{a} = \{a_i\} = \{\frac{A_i}{A_T}\}$
  - conditional probability: $p_{j|i} \Longrightarrow$ form factor: F_{ij}

2.1. Discrete Scene Visibility Entropy

From the above assumptions, we define the *discrete scene visibility entropy rate*, or simply *scene visibility entropy*, as

$$H_S = H(Y|X) = -\sum_{i=1}^{n_p} a_i \sum_{j=1}^{n_p} F_{ij} \log F_{ij} \tag{39}$$

The scene entropy can be interpreted as the average uncertainty that remains about the destination patch of a random walk (or ray) when the source patch is known. It also expresses the average information content of a random walk in a scene. Moreover, H_S can also be seen as the average of the entropies of all patches, where the *entropy of a patch i* is defined by

$$H_i = H(Y|X=i) = -\sum_{j=1}^{n_p} F_{ij} \log F_{ij} \tag{40}$$

and thus (39) can be written as

$$H_S = \sum_{i=1}^{n_p} a_i H_i \tag{41}$$

The entropy of patch i expresses *the uncertainty (or ignorance) of a ray exiting from i about the destination patch*. In fact, it is the Shannon entropy of the form factors of patch i.

The Bayes theorem can be now expressed by the *reciprocity* property of the form factors (8)

$$p_{ij} = a_i F_{ij} = a_j F_{ji} \quad \forall i, j \tag{42}$$

Also, we define the *scene visibility positional entropy* as

$$H_P = H(X) = H(Y) = -\sum_{i=1}^{n_p} a_i \log a_i \tag{43}$$

which may be interpreted as *the uncertainty on the position (patch) of a ray traveling an infinite random walk*. It is the Shannon entropy of the stationary distribution.

The *joint entropy of a scene* is given by

$$H_J = H(X,Y) = -\sum_{i=1}^{n_p} \sum_{j=1}^{n_p} a_i F_{ij} \log a_i F_{ij} \tag{44}$$

This entropy can be interpreted as *the uncertainty about the pair position-target of a ray in an infinite random walk*. It is

the Shannon entropy of a random variable with probability distribution $\{a_i F_{ij}\}$.

2.2. Discrete Scene Visibility Mutual Information

The *discrete scene visibility mutual information* is defined as

$$\begin{aligned} I_S &= I(X,Y) = H(Y) - H(Y|X) \\ &= -\sum_{i=1}^{n_p} a_i \log a_i + \sum_{i=1}^{n_p} a_i \sum_{j=1}^{n_p} F_{ij} \log F_{ij} \end{aligned} \tag{45}$$

and can be interpreted as *the amount of information that the destination patch conveys about the source patch*, and vice versa. I_S *is a measure of the average information transfer in a scene*.

Let us remember that mutual information can be defined as a Kullback-Leibler distance (see Part II): $I(X,Y) = D_{KL}(\{p_{ij}\} \| \{p_i q_j\})$. Thus, scene mutual information is the *distance* or discrimination between the scene probability distribution $\{p_{ij}\} = \{\frac{A_i}{A_T} F_{ij}\}$ and the *independence* distribution of a scene $\{p_i q_j\} = \{a_i a_j\}$. It can also be expressed as

$$\begin{aligned} I_S &= \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} a_i F_{ij} \log \frac{a_i F_{ij}}{a_i a_j} \\ &= \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} a_i F_{ij} \log \frac{F_{ij}}{a_j} \end{aligned} \tag{46}$$

2.3. Properties

In a discretised scene the following properties are met:

- From the reciprocity property of the form factors (8), the reversibility of the channel (and also of the Markov chain) can be obtained: $H_S = H(Y|X) = H(X|Y)$.
- $H_J = H_P + H_S = 2H_P - I_S = 2H_S + I_S$ (Figure 10)

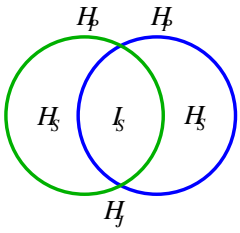


Figure 10: *Venn diagram of a scene.*

- If all the patches have the same area, then $a_i = \frac{1}{n_p}$ and $F_{ij} = F_{ji}$, for all i and j, and therefore

$$H_S = -\frac{1}{n_p} \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} F_{ij} \log F_{ij} = \frac{1}{n_p} \sum_{i=1}^{n_p} H_i \tag{47}$$

and

$$H_P = \log n_p \tag{48}$$

If all the F_{ij} are also equal, then H_S reaches its maximum value: $H_S = H_P = \log n_p$. The minimum value for H_S will be reached when all the form factors from any patch are near zero except one with value near 1: $H_S \approx 0$.

3. Randomness versus Correlation

As we have seen in the previous section, scene visibility entropy H_S is a general measure of the uncertainty or information content associated with a scene: "The entropy density provides an answer to the question: in a Markov chain, given the knowledge of the previous symbol, how uncertain are you, on average, about the next symbol?" [Fel97]. Thus, H_S can be seen as the intrinsic *unpredictability* or the irreducible *randomness* associated with the chain. H_S is also the expected minimum number of bits per symbol required to code a random walk in a scene.

On the other hand, scene mutual information I_S, which expresses the average information transfer, is a measure of the *dependence* or *correlation* between the different parts of a scene. According to W.Li [Li90], "it is now well understood that mutual information measures the general dependence, while the correlation function measures the linear dependence, and mutual information is a better quantity than the correlation function to measure dependence".

3.1. Maximum and Minimum Scene Entropy

It is especially interesting to ask about the extremal cases of maximum and minimum visibility entropy, which correspond to the maximum *disorder* (unpredictability or randomness in the ray path) and the maximum *order* (predictability), respectively. We must remark here that the concepts of order and disorder are not directly referred to the collocation of objects in space, but to visibility criteria. Maximum unpredictability can be obtained in scenes with no privileged visibility directions, and maximum predictability in the contrary case.

Both cases can be illustrated with the following two examples:

- The maximum entropy is exemplified by the interior of an empty sphere divided into equal area patches. Here all the form factors are equal and the uncertainty of the destination patch is maximum:

$$H_S = H_P = \log n_p \tag{49}$$

This means that no visibility direction is privileged and the information transfer is zero: $I_S = 0$.
The sphere represents independence, equilibrium, homogeneity. It is modeled by a channel where the variables X and Y are independent, because in a sphere $F_{ij} = a_j$ and, thus, $p_{ij} = a_i F_{ij} = a_i a_j$. This is the expression of *independence* in a scene. Thus, if the independence is represented by a sphere, the discrete scene mutual information

expresses the *distance* between a given scene and a sphere discretised with the same number and area distribution of the patches.
Note that if the number of patches is doubled, the information content (entropy) of an empty sphere with equal area patches increases by one bit.

- The minimum entropy can be represented by a scene with almost touching objects, as for instance two near concentric spheres with a suitable discretisation. In this case there are strongly privileged visibility directions. This system is highly correlated and the information transfer is large.

3.2. Empirical Results

In this section we show the behaviour of the entropy and mutual information in simple scenes. In the following experiments, form factors have been computed using global lines (see section 1.4).

3.2.1. Different scenes but the same discretisation of the environment

In scenes with the same discretisation (as in Figure 11/Table 1, where we have a cubical enclosure with 512 interior cubes), and consequently with the same H_P, where the interior objects have simply been displaced, we can observe that the increase of entropy remains compensated by a mutual information decrease, and vice versa: more randomness means less correlation (Figure 11a), less randomness means more correlation (Figure 11c). The Venn diagrams in Figure 12 illustrate this behaviour.

Scene	H_S	I_S
Fig.11a	6.761	4.779
Fig.11b	5.271	6.270
Fig.11c	4.849	6.692

Table 1: *Entropy and mutual information for Figures 11a, 11b and 11c. For each scene, $H_P = 11.541$ and 10^7 global lines have been used to compute the form factors.*

3.2.2. The same scene but different discretisations of the environment

How does scene entropy behave with an increase in the number of patches? According to information theory, when the number of patches goes to infinity, the scene entropy also goes to infinity, but scene mutual information tends to a finite value (see the next chapter, section 7). So, in general, the increase in H_S has to be greater than the increase in I_S. This fact is partially illustrated in Table 2, corresponding to Figure 13, where we have a cubical enclosure with three different regular discretisations of its surfaces (600, 2400, and 5400 patches, respectively). We can see that, for each scene, $H_P = \log n_p$, as all the patches have the same area. The Venn diagrams in Figure 14 illustrate the behaviour of the entropy and mutual information of these scenes.

Scene	H_S	I_S	H_P
Fig.13a	7.838	1.391	9.229
Fig.13b	9.731	1.498	11.229
Fig.13c	10.852	1.547	12.399

Table 2: *Results for a cubical enclosure with different discretisations of its surfaces (Figure 13). 10^9 global lines have been used to compute the form factors.*

3.2.3. Normalized measures

In order to account for changes in the proportion of randomness (H_S) and correlation (I_S) in a scene, these can be normalized by dividing them by the positional entropy H_P. So, they range from 0 to 1.

Normalized scene entropy can be defined as

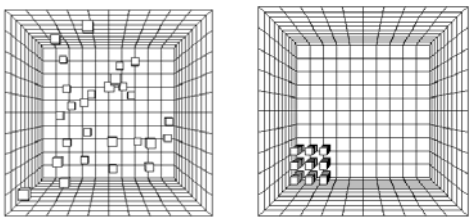
$$\overline{H_S} = \frac{H_S}{H_P} \tag{50}$$

and *normalized scene mutual information* as

$$\overline{I_S} = \frac{I_S}{H_P} = 1 - \frac{H_S}{H_P} \tag{51}$$

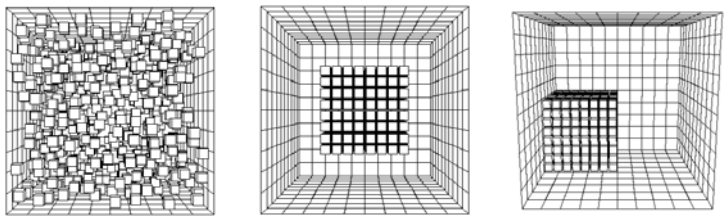
In the literature, normalized mutual information is considered as a measure of *correlation* [CT91]. Also, it can be useful to normalize H_S and I_S with respect to the joint entropy H_J: $\widehat{H_S} = \frac{H_S}{H_J}$ and $\widehat{I_S} = \frac{I_S}{H_J}$. Obviously, they also range from 0 to 1. A similar approach has been used as a measure of 3D medical image alignment [Stu97].

In the sequence of results of Table 3, where we start with an empty cubical enclosure (Figure 13a) and then we add small interior cubes (Figures 15, 16, 11a, and 11c), we can observe how the normalized entropy decreases when we introduce progressively more cubes. This fact increases the correlation in the scene, to the detriment of its randomness, in spite of the fact that H_P also increases.



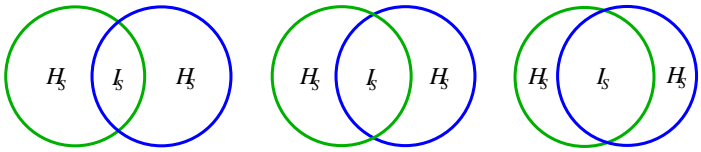
(a) (b)

Figure 15: *(a) Random and (b) clustered configurations with 27 cubes.*



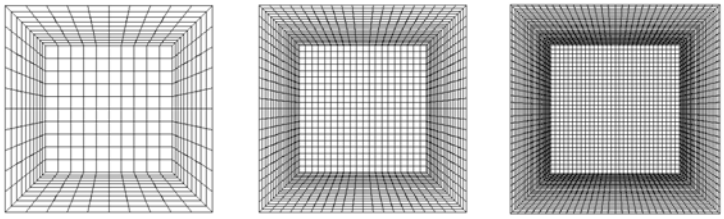
(a) (b) (c)

Figure 11: *A random configuration and two clustered configurations with* 512 *cubes.*



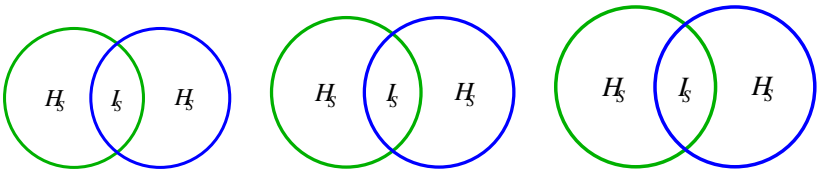
(a) (b) (c)

Figure 12: *Venn diagrams corresponding to different scenes with the same discretisation. The size of the circles (H_P) remains the same in all the diagrams.*



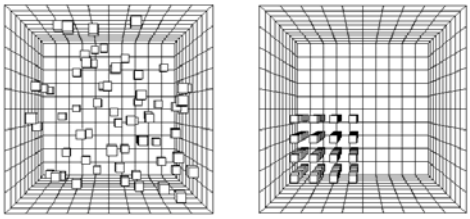
(a) (b) (c)

Figure 13: *Three empty cubical enclosures with their surfaces regularly discretised into (a)* 10×10*, (b)* 20×20*, and (c)* 30×30 *patches, respectively. The total number of patches is, respectively,* 600*,* 2400*, and* 5400*.*



(a) (b) (c)

Figure 14: *Venn diagrams corresponding to the scenes of Figure 13 where we have a cubical enclosure with successive refinements. The size of the circles (H_P) increases with the number of patches.*



(a) (b)

Figure 16: *(a) Random and (b) clustered configurations with 64 cubes.*

Scene	H_S	I_S	$\overline{H_S}$
Fig.13a	7.821	1.408	0.847
Fig.15a	7.780	1.669	0.823
Fig.15b	7.589	1.861	0.803
Fig.16a	7.709	2.009	0.793
Fig.16b	7.606	2.112	0.783
Fig.11a	6.761	4.779	0.586
Fig.11c	4.849	6.692	0.420

Table 3: *Results for the empty scene of Figure 13a and the scenes with 27, 64 and 512 small cubes of Figures 13a, 15, 16, 11a, and 11c. For each scene,* 10^7 *global lines have been cast.*

From the previous results, we see that scene entropy (randomness) tends to increase with the *number of patches* and scene mutual information (correlation) tends to increase with the *number of objects* within a enclosure. So, the increase in the number of patches and the increase in the number of objects work in different (but complementary) directions.

4. Complexity of a Scene

Scene complexity has often been expressed as the number of patches into which a scene is subdivided. But, what do we really expect scene complexity to measure? In our context, scene complexity has to answer the question of how difficult it is to compute the visibility and radiosity of a scene with sufficient accuracy. Studying scene complexity will help to improve our knowledge about the behaviour of the visibility and radiosity of a scene.

To solve the illumination in a diffuse environment, we need to simulate the interreflection of light between all the surfaces. This simulation presents typical characteristics of complex behaviour. The *difficulty in obtaining a precise illumination solution* depends on

- the degree of dependence between all the surfaces
- how the interaction between these surfaces changes in dependence when the system is subdivided
- the degree of unpredictability

The two first considerations can be represented by a *statistical complexity measure*, which quantifies *correlation*, *structure*, or *interdependence* between the parts of a system, and the third one by the *entropy*, which measures *randomness* or *unpredictability*. In this work, the word *complexity* is reserved for a measure of *statistical complexity* and *entropy* is referred to as *randomness*.

The most representative measures of statistical complexity are excess entropy and mutual information. For scene visibility, the following proposition is fulfilled:

Proposition 1 From the point of view of scene visibility, the excess entropy (see Part II) becomes the mutual information.

Proof. From definitions in Part II and section 2, we have

$$\begin{aligned} H(X_1,\ldots,X_n) &= H(X_1)+\ldots+H(X_n|X_1,\ldots,X_{n-1}) \\ &= H(X_1)+H(X_2|X_1)+\ldots+H(X_n|X_{n-1}) \\ &= H_P+(n-1)H_S \end{aligned}$$

Thus, $E=\lim_{n\to\infty}(H_P+(n-1)H_S-nH_S)=H_P-H_S=I_S$. □

So, *we propose taking the mutual information as a measure of scene complexity*.

According to Feldman, Rohilla, and Crutchfield:

- "It has become (in our sense) more broadly understood that a system's randomness and unpredictability fail to capture its patterns and correlational structure." [FC98]
- Entropy and mutual information are orthogonal or complementary: "Complexity and randomness each capture a useful property to describe how a process manipulates information" [SC99].

Entropy and mutual information express two basic aspects of scene complexity. But, we are sure that other measures could capture other perspectives of the scene complexity. Remember that one of the most basic formulae of information theory (see part II) relates entropy to mutual information:

$$H(X)=H(X|Y)+I(X,Y) \tag{52}$$

Thus, complexity and randomness are combined in the same expression.

A very simple approach would be to consider that the complexity can be represented by the number of patches. It is true that scene entropy is strongly dependent on n_p and increases with it, with a maximum value of $\log n_p$. However, as we will see in this chapter, scene mutual information presents a very different behaviour with respect to n_p.

5. Continuous Scene Visibility Mutual Information

A scene is a continuous system. Thus, by discretising a scene into patches, a *distortion* or *error* is introduced. In a way, to discretise means to make it uniform, and consequently *some information is lost*. Obviously, the maximum accuracy of the discretisation is accomplished when the number of patches tends to infinity. Since the *continuous mutual information expresses with maximum precision the information transfer or correlation in a scene*, it will be considered as the main measure of the *scene complexity*. On the other hand, *discrete mutual information* will represent the *complexity of a discretised scene*.

Similarly to the previous chapter, now the scene is modeled by a continuous random walk (Figure 17) or by a continuous information channel. As we have seen in Part II, the mutual information between two continuous random variables X and Y is the limit of the mutual information between their discretised versions. On the contrary, the entropy of a continuous random variable does not equal the entropy of its discretised version in the limit of a finer discretisation.

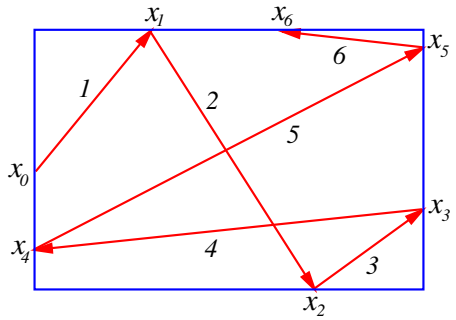


Figure 17: *Continuous random walk in a scene.*

Thus, in a scene, discrete mutual information I_S converges to continuous mutual information I_S^c when the number of patches tends to infinity:

$$I_S^c = \lim_{n_p \to \infty} I_S \tag{53}$$

In this chapter we will see that this result is very important for this dissertation because it will enable us to calculate the distance to the ideal discretisation, represented by the continuous mutual information.

Scene visibility entropy however tends to infinity when the number of patches tends to infinity:

$$\lim_{n_p \to \infty} H_S = \infty \tag{54}$$

As we have seen (section 1.6), when the states form an uncountable set, we deal with a continuous Markov chain. We can obtain the continuous formulae for the entropy and mutual information of a scene from the respective discrete definitions using the following substitutions:

- Each state by an infinitesimal area and each summatory by an integral.
- $w_i = \frac{A_i}{A_T} \Longrightarrow \frac{1}{A_T}$. This means substituting the discrete probability of taking patch i by the continuous probability of selecting any point.
- $F_{ij} \Longrightarrow F(x,y)$. This means substituting the patch-to-patch form factor by the point-to-point form factor. Remember that the value of $F(x,y)$ is $\frac{cos\theta_x cos\theta_y}{\pi r_{xy}^2}$ for mutually visible points, or zero otherwise, θ_x and θ_y being the angles which the normals at x and y form with the segment joining x and y, and r_{xy} the distance between x and y (see section 1.3).

In the same way, the continuous formulae for a scene can also be obtained from the continuous formulae of the entropy and mutual information (see Part II) by applying the following substitutions:

- $dx \Longrightarrow dA_x$, $dy \Longrightarrow dA_y$
- $p(x) \Longrightarrow \frac{1}{A_T}$
- $p(y|x) \Longrightarrow F(x,y)$
- $p(x,y) \Longrightarrow \frac{1}{A_T} F(x,y)$

Thus, we obtain

- *Continuous positional entropy*

$$H_P^c = -\int_S \frac{1}{A_T} \log \frac{1}{A_T} dA_x = \log A_T \tag{55}$$

- *Continuous scene visibility entropy*

$$H_S^c = -\int_S \int_S \frac{1}{A_T} F(x,y) \log F(x,y) dA_x dA_y \tag{56}$$

- *Continuous scene visibility mutual information*

$$\begin{aligned} I_S^c &= \log A_T + \int_S \int_S \frac{1}{A_T} F(x,y) \log F(x,y) dA_x dA_y \\ &= \int_S \int_S \frac{1}{A_T} F(x,y) \log (A_T F(x,y)) dA_x dA_y \end{aligned} \tag{57}$$

For example, in the interior of an empty sphere, as any pair (x,y) fulfills $F(x,y) = \frac{1}{A_T}$, the result obtained is, as expected, $I_S^c = 0$. Remember that, in a sphere, $I_S = 0$ and thus $\lim_{n_p \to \infty} I_S = I_S^c = 0$.

Note that H_P^c and H_S^c are not invariant to a change in the scale of a scene. For our objectives, we are only interested in their discrete versions, which are always invariant to a change in the scale. I_S and I_S^c also have this desirable property.

6. Monte Carlo Computation of the Scene Visibility Complexity

Now we will show how the continuous mutual information can be computed using local or global lines.

6.1. Monte Carlo Integration

The continuous mutual information integral can be solved by Monte Carlo integration. Reparametrizing the integral, we have

$$\begin{aligned} I_S^c &= \int_S \int_S \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y \\ &= \int_S \int_{\Omega_x} \frac{1}{A_T} \frac{\cos\theta_x}{\pi} \log(A_T F(x, y(x,\omega_x))) dA_x d\omega_x \end{aligned} \tag{58}$$

where $y(x,\omega_x)$ is the point visible from x in the ω_x direction. Now we can use $\frac{\cos\theta_x}{\pi A_T}$ as probability density function ($\int_S \int_{\Omega_x} \frac{\cos\theta_x}{\pi A_T} dA_x d\omega_x = 1$). Drawing samples according to this distribution means simply selecting first a random point in the scene upon the area and a direction upon the form factor distribution. This can be achieved with *local* lines or *global* lines. The result obtained is

$$\begin{aligned} I_S^c &\approx \frac{1}{N} \sum_{k=1}^{N} \log(A_T F(x_k, y_k(x_k, \omega_{x_k})) \\ &= \frac{1}{N} \sum_{k=1}^{N} \log\left(\frac{A_T \cos\theta_{x_k} \cos\theta_{y_k}}{\pi r_{x_k y_k}^2}\right) \end{aligned} \tag{59}$$

In the global line case, N stands for the total number of segments of the global lines or the number of pairs of points considered, which is the total number of intersections divided by two (see Figure 8). In the local line case, N represents the total number of local lines used in a scene and the quantity of lines cast from each patch, N_i, proportional to its area ($N_i = \frac{A_i}{A_T} N$). In this chapter, the scene complexity has been calculated using both global and local lines.

6.2. Empirical Results

We begin by computing the complexity of platonic solids and the Cornell box (Figure 1a). In Table 4, we can observe that the minimum complexity corresponds to a sphere and the maximum complexity to a tetrahedron. As we expected, the polyhedra that are nearer to the sphere are less complex, i.e., they have less correlation. Thus, complexity appears to be inversely proportional to the number of faces. The complexity of the Cornell box is clearly greater than the one for the empty cube, as we have increased the cube complexity by introducing objects in its interior.

In addition, in Table 5, we show the complexity for the scenes of Figure 18. In Figure 18a, an object, made up of a table and four chairs, is situated in the middle of a room. In Figures 18b and 18c, arrays of 4 or 16 objects have been situated in the middle of the same room. In Figures 18d, 18e and 18f, the same 16 objects have been distributed in different ways. We can see that the introduction of objects increases the complexity and that the scenes with the same objects (18c, 18d, 18e and 18f) show similar complexities. In this case, the increase in complexity is produced when there are objects near the walls because this fact increases the correlation in the scene.

Scene	I_S^c
sphere	0
icosahedron	0.543
dodecahedron	0.825
octahedron	1.258
cube	1.609
tetrahedron	2.626
Cornell box	3.274

Table 4: *Complexity of platonic solids and the Cornell box. For each scene,* 10^6 *global lines have been cast.*

7. Complexity and Discretisation

Now, we will try to show that the scene complexity I_S^c is closely related to the difficulty in obtaining an accurate discretisation. In a way, to discretise a scene is to model it. "A system is not complex by some abstract criterion but because it is intrinsically hard to model" [BP97]. This point of view is compatible with W.Li's comment that: "An intuitively satisfactory definition of complexity should measure the amount of effort put in that generates *correlations* in a sequence. Of course, one cannot be sure that all the effort is spent on generating correlations. As a result, a measure of correlation typically provides a lower bound of a measure of complexity, and might be a reasonable estimate of the complexity" [Li91].

7.1. Continuous versus Discrete Mutual Information

From Part II, we know that

- The mutual information between two continuous random variables is the limit of the mutual information between their discretised versions.
- Refinement can never decrease the discrete mutual information.
- The continuous mutual information is the least upper bound for the discrete mutual information.

Now, if we apply these results to scene visibility, we find that:

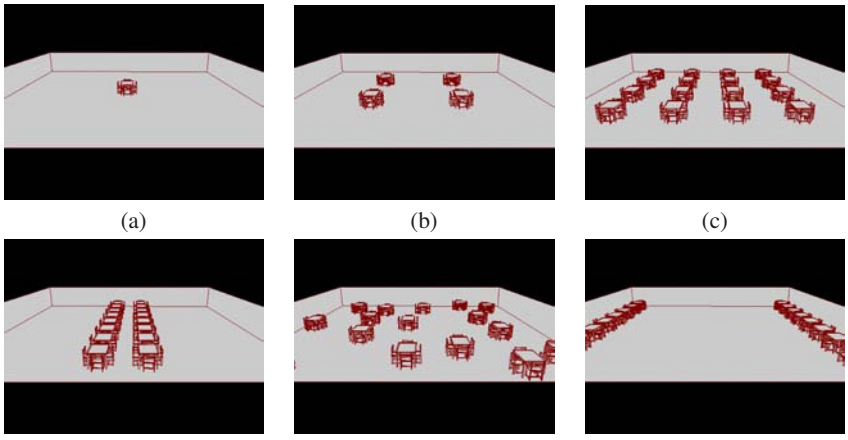


Figure 18: *(a) An object, composed of a table and four chairs, and (b) an array of* 4 *objects with the same composition, have been situated in the middle of a room. (c, d, e, f) The same* 16 *objects have been distributed in four different ways.*

Scenes	Fig.18a	Fig.18b	Fig.18c	Fig.18d	Fig.18e	Fig.18f
I_S^c	3.837	4.102	5.023	5.043	5.044	5.089

Table 5: *Complexity of the scenes of Figure 18. For each scene,* 10^6 *global lines have been cast.*

- If any patch is divided into two or more patches, the discrete mutual information I_S of the new scene increases or remains the same.
- The continuous scene visibility mutual information is the least upper bound to the discrete scene visibility mutual information.

Thus, a scene fulfils:

$$I_S^c - I_S \geq 0 \tag{60}$$

Initially, these results were proved in [FS98, FdABS99]. In the next chapter, we give a general proposition proving these properties for visibility, radiosity and importance.

As we can see in Tables 6 and 7, corresponding to Figures 13 and 11(a), respectively, the computational cost of I_S^c is much lower than the cost of I_S: with few lines I_S^c can be computed with enough precision, unlike I_S which needs a lot of lines to get a precise measurement. Observe that I_S increases with the number of patches but is always less than I_S^c. We can also see that, due to the Monte Carlo error, the value of the discrete mutual information decreases (until convergence is achieved) with the increase in the number of lines cast. With few lines per patch, the values of the form factors give us an erroneous high correlation.

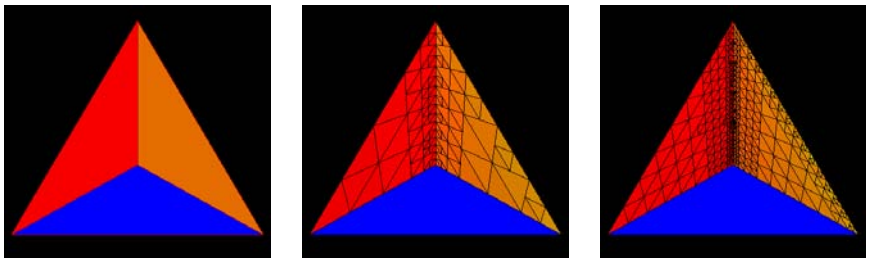
In Tables 8 and 9, corresponding to Figures 19 and 20, we also show how discrete mutual information I_S increases with the mesh refinement.

Scene	*Lines* (10^6)	H_S	I_S	I_S^c
Fig.13a	0.1	6.482	2.747	1.610
Fig.13a	10	7.821	1.408	1.612
Fig.13a	1000	7.838	1.391	1.610
Fig.13b	0.1	5.342	5.887	1.608
Fig.13b	10	9.420	1.809	1.611
Fig.13b	1000	9.731	1.498	1.610
Fig.13c	0.1	4.313	8.086	1.610
Fig.13c	10	9.684	2.715	1.611
Fig.13c	1000	10.852	1.547	1.610

Table 6: *Results for the cubical enclosure of Figure 13 with different discretisations of its surfaces. For each scene,* 10^5*,* 10^7*, and* 10^9 *global lines have been cast.*

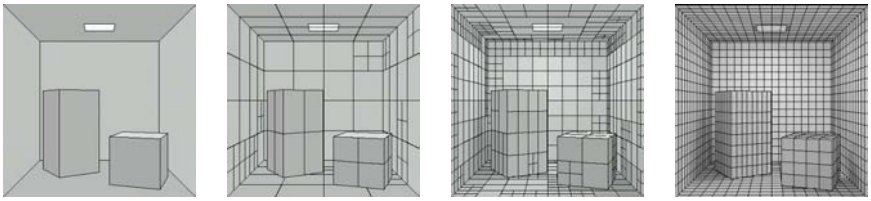
Lines (10^6)	0.01	0.1	1	10
I_S	8.773	6.398	5.171	4.779
I_S^c	5.650	5.636	5.631	5.632

Table 7: *Discrete and continuous mutual information for a scene with 512 cubes (Figure 11a).*



(a) (b) (c)

Figure 19: *Three different discretisations for a tetrahedron. The total number of patches is, respectively,* 4*,* 151*, and* 793*.*



(a) (b) (c) (d)

Figure 20: *Four different discretisations for the Cornell box. The total number of patches is, respectively,* 19*,* 196*,* 826*, and* 1924*.*

Scene	*Patches*	I_S	I_S^c
Fig.19a	4	0.415	2.626
Fig.19b	151	1.217	2.626
Fig.19c	793	1.445	2.626

Table 8: I_S^c *and* I_S *for the scenes in Figure 19. For each scene,* 10^7 *local lines have been cast.*

Scene	*Patches*	*Lines* (10^6)	I_S	I_S^c
Fig.20a	19	10	0.690	3.273
Fig.20b	196	10	2.199	3.273
Fig.20c	826	10	2.558	3.273
Fig.20d	1924	100	2.752	3.273

Table 9: I_S^c *and* I_S *for the scenes in Figure 20.*

7.2. Discretisation Accuracy

We know that the difference $I_S^c - I_S$ (60), always positive, expresses the *loss of information transfer* due to the discretisation. From this assumption, we can now take a leap forward with two fundamental proposals. The first appears naturally in an information-theory context and the second will be experimentally checked:

1. From an information-theory point of view, the ideal discretisation is the one that captures all the information transfer in a scene. Thus, between different discretisations of the same scene, the most precise will be the one that has a higher discrete mutual information I_S, i.e., the one that best captures the information transfer. With this in mind, we can express the *discretisation error* as the difference

$$\delta^v = I_S^c - I_S \tag{61}$$

and the *relative discretisation error* as the quotient

$$\overline{\delta}^v = \frac{I_S^c - I_S}{I_S^c} \tag{62}$$

The *relative discretisation accuracy* is given by $\frac{I_S}{I_S^c}$.

2. Continuous mutual information I_S^c expresses the difficulty in obtaining an accurate discretisation. The higher the I_S^c (i.e., when there is more information transfer in a scene), the more difficult it is to obtain an accurate discretisation, and probably more refinements will be necessary to achieve a given precision. From this point of view, the difficulty in discretising the interior of an empty sphere is null (the discretisation error is always equal to zero). The polyhedra that are "nearer" to the sphere are less complex than the others, and so easier to discretise.

8. Loss of Information Transfer due to the Discretisation

To obtain a refinement criterion based on the discretisation error between two patches, we need to consider both continuous and discrete patch-to-patch information transfers.

8.1. Mutual Information Matrix

As we have seen, discrete scene mutual information is given by

$$I_S = \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} a_i F_{ij} \log\left(\frac{F_{ij}}{a_j}\right) \tag{63}$$

From this formula, the term

$$I_{ij} = a_i F_{ij} \log\left(\frac{F_{ij}}{a_j}\right) \tag{64}$$

can be considered as an element of a mutual information matrix, and it is easy to see that $I_{ij} = I_{ji}$. Each element represents the information transfer between patches i and j. Also, we can consider that

$$I_i = \sum_{j=1}^{n_p} a_i F_{ij} \log\left(\frac{F_{ij}}{a_j}\right) \tag{65}$$

expresses the information transfer from patch i. Thus, we can write

$$I_S = \sum_{i=1}^{n_p} I_i = \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} I_{ij} \tag{66}$$

If we analyze the terms I_{ij}, we observe that negative values appear when $F_{ij} < a_j$. This situation reflects a very low interaction between the two patches involved. On the other hand, using the concavity property of the logarithm function (see Part II), it is easy to see that $I_i \geq 0$ (substituting a_k, b_k, and n by F_{ij}, a_j, and n_p, respectively).

The information transfer between two patches can be obtained more accurately if we consider the continuous mutual information between them. Thus, from the continuous mutual information, we obtain the following results.

Continuous information transfer:

$$\begin{aligned} I_S^c &= \int_S \int_S \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y \\ &= \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} \int_{A_i} \int_{A_j} \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y \end{aligned} \tag{67}$$

Continuous information transfer due to patch i:

$$I_i^c = \sum_{j=1}^{n_p} \int_{A_i} \int_{A_j} \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y \tag{68}$$

Continuous information transfer between patches i and j:

$$I_{ij}^c = \int_{A_i} \int_{A_j} \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y \tag{69}$$

This continuous measure expresses with maximum precision the visibility information transfer between two elements.

8.2. Discretisation Error Between Two Patches

As we have seen (section 7.2), a general discretisation error for a scene can be given by

$$\delta = I_S^c - I_S \geq 0 \tag{70}$$

In order to propose a refinement oracle for hierarchical radiosity, we are interested in the contribution to this discretisation error of the patch-to-patch interaction (i.e., discretisation error between two patches). So, we calculate, respectively, the difference between continuous and discrete patch-to-patch mutual information. For each one, three different Monte Carlo techniques can be used: patch-to-patch random lines, local lines and global lines.

1. **Patch-to-patch random lines:** The computation of (69)

$$I_{ij}^c = \int_{A_i} \int_{A_j} \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y$$

can be done with an area-to-area sampling (section 1.4), i.e., using random lines joining both elements i and j (the pdf is $\frac{1}{A_i A_j}$). For N_{ij} lines, we have

$$I_{ij}^c \approx \frac{A_i A_j}{A_T} \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} F(x_k, y_k) \log\Big(F(x_k, y_k) A_T\Big) \tag{71}$$

where x_k and y_k are, respectively, the end-points on patches i and j of the k-th line.
From (64), I_{ij} can be expressed as

$$I_{ij} = \frac{A_i F_{ij}}{A_T} \log\left(\frac{F_{ij} A_T}{A_j}\right) = \frac{A_i A_j}{A_T} \frac{F_{ij}}{A_j} \log\left(\frac{F_{ij}}{A_j} A_T\right) \tag{72}$$

Now, taking $\frac{F_{ij}}{A_j} \approx \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} F(x_k, y_k)$ (28), we obtain the

visibility discretisation error between patches i and j:

$$\begin{aligned}\delta_{ij}^v &= I_{ij}^c - I_{ij} \\ &\approx \frac{A_i A_j}{A_T}\left(\frac{1}{N_{ij}}\left(\sum_{k=1}^{N_{ij}} F(x_k,y_k)\log\Big(F(x_k,y_k)A_T\Big)\right)\right. \\ &\left. - \frac{F_{ij}}{A_j}\log\Big(\frac{F_{ij}}{A_j}A_T\Big)\right) \\ &= \frac{A_i A_j}{A_T}\left(\frac{1}{N_{ij}}\left(\sum_{k=1}^{N_{ij}} F(x_k,y_k)\log(F(x_k,y_k))\right)\right. \\ &\left. - \frac{F_{ij}}{A_j}\log\Big(\frac{F_{ij}}{A_j}\Big)\right) \\ &= \frac{A_i A_j}{A_T}\left(\frac{1}{N_{ij}}\left(\sum_{k=1}^{N_{ij}} F(x_k,y_k)\log(F(x_k,y_k))\right)\right. \\ &\left. - \Big(\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}} F(x_k,y_k)\Big)\log\Big(\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}} F(x_k,y_k)\Big)\right) \geq 0\end{aligned} \tag{73}$$

where we have used the log-sum inequality (see Part II). This difference gives us the discretisation error between two elements and it is used as the basis for our mutual-information-based (MI-based) oracle. Observe also that δ_{ij}^v is symmetric: $\delta_{ij}^v = \delta_{ji}^v$.

2. **Local or global lines:**
The computation of I_{ij}^c can also be done with uniformly distributed local or global lines (section 1.4). From (59), we obtain

$$I_{ij}^c \approx \frac{A_i F_{ij}}{A_T}\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}}\log(A_T F(x_k,y_k)) \tag{74}$$

where N_{ij} is the number of local lines or segments of global lines which connect patches i and j. Hence, we find

$$\begin{aligned}\delta_{ij}^v &= I_{ij}^c - I_{ij} \\ &\approx \frac{A_i F_{ij}}{A_T}\left(\frac{1}{N_{ij}}\left(\sum_{k=1}^{N_{ij}}\log(F(x_k,y_k))\right) - \log\frac{F_{ij}}{A_j}\right)\end{aligned} \tag{75}$$

As we expected, it is easy to see that the discretisation error between two spherical patches is equal to zero.

9. Mutual-Information-Based Oracle for Hierarchical Radiosity

We introduce in this section an information-theory oracle based on the radiosity kernel smoothness to be used in the hierarchical refinement algorithm. As the refinement strategy in hierarchical radiosity deals with one pair of elements at a time, we have to look for a similar interaction in our information theory framework.

The fundamental idea in our approach is the following: the difference between continuous and discrete patch-to-patch (or element-to-element) mutual information, i.e., discretisation error, gives us the *loss* of information transfer or the *maximum potential gain* of information transfer between two elements. Hence this difference can be interpreted as the *benefit to be gained by refining* and can be used as a decision criterion.

9.1. An Oracle Based on the Discretisation Error between Two Patches

To obtain a mutual-information-based oracle, we take a similar approach to the classic smoothness-based oracles, which multiplies $\rho_i B_j$ (from the radiosity equation (6)) by an expression of the visibility gradient between the two patches involved. In our case, the visibility gradient is given by the discretisation error $\delta_{ij}^v = I_{ij}^c - I_{ij}$, which also represents the *variation of the radiosity kernel*.

Our oracle will be based on the following considerations:

- In the radiosity equation (6)
$$B_i = E_i + \rho_i \sum_{j=1}^{n_p} F_{ij} B_j$$
the contribution of patch j to the radiosity of patch i is given by $\rho_i F_{ij} B_j$. Thus, the geometric factor, i.e., the radiosity kernel, is weighted by $\rho_i B_j$.
- The kernel-smoothness-based oracles reviewed in section 1.8, such as
$$\rho_i (F_{ij}^{max} - F_{ij}^{min}) A_j B_j < \varepsilon$$
and
$$\rho_i max(F_{ij}^{max} - F_{ij}^{av}, F_{ij}^{av} - F_{ij}^{min}) A_j B_j < \varepsilon$$
, try to capture the *variation of the radiosity kernel* using the maximum and minimum kernel values.

Our oracle proposal takes these two facts on board, weighting the variation of the radiosity kernel (expressed by the visibility discretisation error δ_{ij}^v between two patches) by $\rho_i B_j$. So, we find that the *mutual-information-based (MI-based) oracle* is given by

$$\rho_i \delta_{ij}^v B_j < \varepsilon \tag{76}$$

which can be computed with N_{ij} element-to-element random lines between elements i and j:

$$\begin{aligned}&\frac{\rho_i A_i A_j B_j}{A_T}\left(\frac{1}{N_{ij}}\left(\sum_{k=1}^{N_{ij}} F(x_k,y_k)\log(F(x_k,y_k))\right)\right. \\ &\left. - \Big(\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}} F(x_k,y_k)\Big)\log\Big(\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}} F(x_k,y_k)\Big)\right) < \varepsilon\end{aligned} \tag{77}$$

Observe that in this expression the receiver area appears weighting the oracle and thus avoiding an excessively small receiver subdivision.

It is important to note that δ_{ij}^{v} can be used as an oracle for visibility. This only takes into account the variation of the radiosity kernel and the areas of the patches involved.

9.2. Empirical Results

To check the performance of the MI-based oracle, we have implemented a power-based oracle (36), a classic kernel-smoothness-based (KS-based) oracle (38) and our MI-based oracle (77) in the hierarchical Monte Carlo radiosity [BNN*98] method of the RenderPark [Com00] system (`www.renderpark.be`). It should be noted that our oracle can be used with any hierarchical radiosity method.

In our experiments, we use two scenes: the Cornell box (Figures 21 and 24) and the cube room (Figures 22, 23 and 25). Six different discretisations were generated for the Cornell box: three coarse (Figure 21I) and three finer ones (Figure 21II). These discretisations have been obtained from three meshing strategies based, respectively, on transported power (36) (Figures 21a.I and 21a.II), classic kernel smoothness (38) (Figures 21b.I and 21b.II), and mutual information (77) (Figures 21c.I and 21c.II). In a similar way, we compared our strategy with the KS-based strategy using two different views of the cube room scene (Figures 22 and 23). Both KS-based and MI-based oracles were evaluated for each discretisation decision with 10 element-to-element random lines (except in Figures 24 and 25, where only 4 rays were used). For the power-based oracle we used a cheap form factor estimate (see section 1.8).

In Figures 21I and 21II we see the behaviour of the three oracles for two different levels of discretisation. Using the power-based and KS-based oracles, the shadow of the small cube gets an accurate representation only at the finer level of discretisation, whereas the MI-based oracle already produces a good representation in the coarse mesh. The power-based oracle overdiscretises the rear wall and the top of the prism, as expected, while the smoothness-based oracles correct this effect. However, the MI-based oracle supports the change from a coarse to a finer mesh much better (see again the rear wall).

Figures 22 and 23 show the behaviour of the classic KS-based and MI-based oracle for the cube room scene. Observe the accurate representation of the shadow of the chair near the right wall (Figure 22b) and front wall (Figure 23b) obtained by the MI-based oracle. Observe also the much better discrimination in the mesh, seen for instance on the floor and walls, and how the shadows on the table are represented more accurately in Figure 23b.

In Figures 24 and 25, the robustness of the classic KS-based and MI-based oracle are tested against a decrease from 10 to 4 point-to-point form factor computations for each oracle evaluation. The performance of the classic KS-based oracle degenerates to a degree similar to the power-based oracle, see for instance the rear wall in Figure 24a (compare with Figures 21b.I) and the same happens in Figure 25a (compare with Figures 22a). On the other hand, the MI-based oracle maintains most of its good performance (compare Figure 24b with Figures 21c.II). See also the shadow of the chair near the right wall in Figures 25b and 22b.

In Figure 26 we show a more accurate solution computed with the MI-based oracle, 10 element-to-element random lines for each oracle evaluation and 2684260 rays for radiosity computation.

References

[Bek99] BEKAERT P.: *Hierarchical and Stochastic Algorithms for Radiosity*. PhD thesis, Katholieke Universiteit Leuven, Leuven, Belgium, December 1999.

[BNN*98] BEKAERT P., NEUMANN L., NEUMANN A., SBERT M., WILLEMS Y. D.: Hierarchical Monte Carlo radiosity. In *Rendering Techniques'98 (Proceedings of the 9th Eurographics Workshop on Rendering)* (New York (NY), USA, June 1998), Drettakis G., Max N., (Eds.), Springer-Verlag Vienna-New York, pp. 259–268. Held in Vienna, Austria.

[BP97] BADII R., POLITI A.: *Complexity. Hierarchical Structures and Scaling in Physics*. Cambridge University Press, 1997.

[BW96] BEKAERT P., WILLEMS Y. D.: Error control for radiosity. In *Rendering Techniques'96 (Proceedings of the 7th Eurographics Workshop on Rendering)* (New York (NY), USA, June 1996), Pueyo X., Schröder P., (Eds.), Springer-Verlag Vienna-New York, pp. 153–164. Held in Vienna, Austria.

[CG85] COHEN M. F., GREENBERG D. P.: The hemicube: A radiosity solution for complex environments. *Computer Graphics (Proceedings of SIGGRAPH'85) 19*, 3 (July 1985), 31–40. Held in San Francisco (CA), USA.

[Col74] COLEMAN R.: *Stochastic Processes*. George Allen & Unwin Ltd., London, UK, 1974.

[Com00] COMPUTER GRAPHICS RESEARCH GROUP: *RenderPark: A Photorealistic Rendering Tool*. Katholieke Universiteit Leuven, Leuven, Belgium, November 2000.

[CT91] COVER T. M., THOMAS J. A.: *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.

[CW93] COHEN M. F., WALLACE J. R.: *Radiosity and Realistic Image Synthesis*. Academic Press Professional, Boston (MA), USA, 1993.

[FC98] FELDMAN D. P., CRUTCHFIELD J. P.: *Discovering Noncritical Organization: Statistical Mechanical, Information Theoreticand Computational Views of Patterns in One-Dimensional Spin Systems*. Working Paper 98–04–026, Santa Fe Institute, Santa Fe (NM), USA, April 1998.

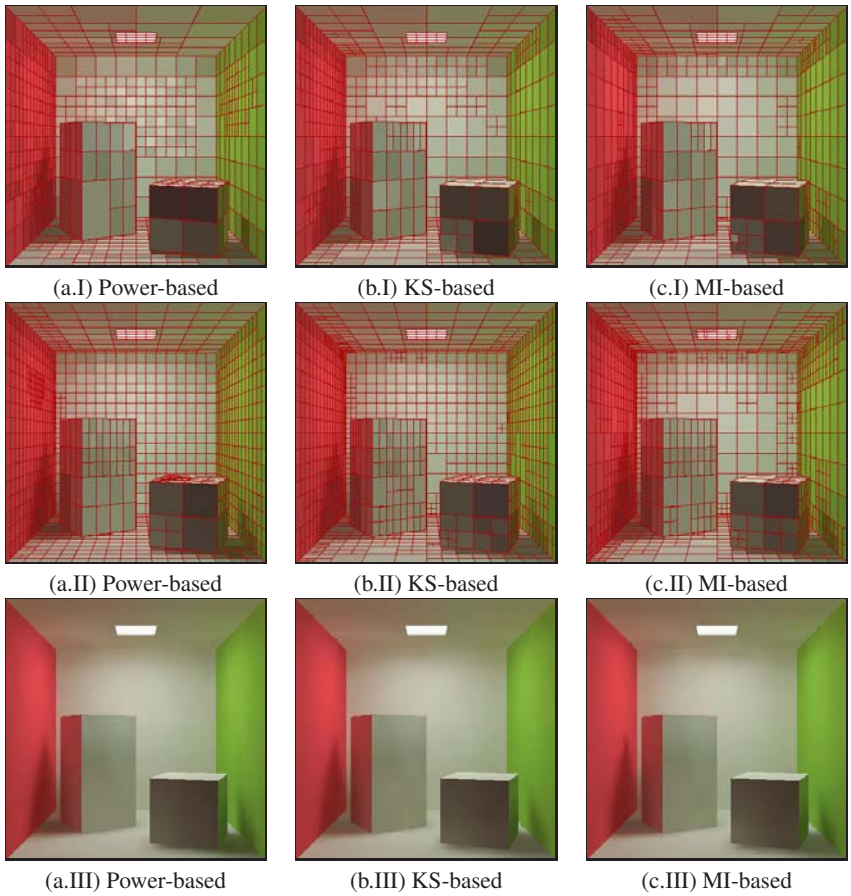


Figure 21: *Power-based (a), KS-based (b) and MI-based (c) methods with the Cornell box scene. A coarse mesh is shown in (I) with 1051 (a.I), 1039 (b.I), and 1047 (c.I) patches, with 19472 rays for the radiosity computation. A fine mesh is shown in (II) with 1979 (a.II), 1955 (b.II), and 1995 (c.II) patches, with 116780 rays for the radiosity computation. The Gouraud shaded solution for (II) is shown in (III). For images (b) and (c), 10 rays are cast for each oracle evaluation.*

[FdABS99] FEIXAS M., DEL ACEBO E., BEKAERT P., SBERT M.: An information theory framework for the analysis of scene complexity. *Computer Graphics Forum (Proceedings of Eurographics'99) 18*, 3 (September 1999), 95–106.

[Fel97] FELDMAN D. P.: A brief introduction to: Information theory, excess entropy and computational mechanics, 1997.

[FH96] FRISKEN S. F., HUBBOLD R. J.: Efficient hierarchical refinement and clustering for radiosity in complex environments. *Computer Graphics Forum 15*, 5 (1996), 297–310.

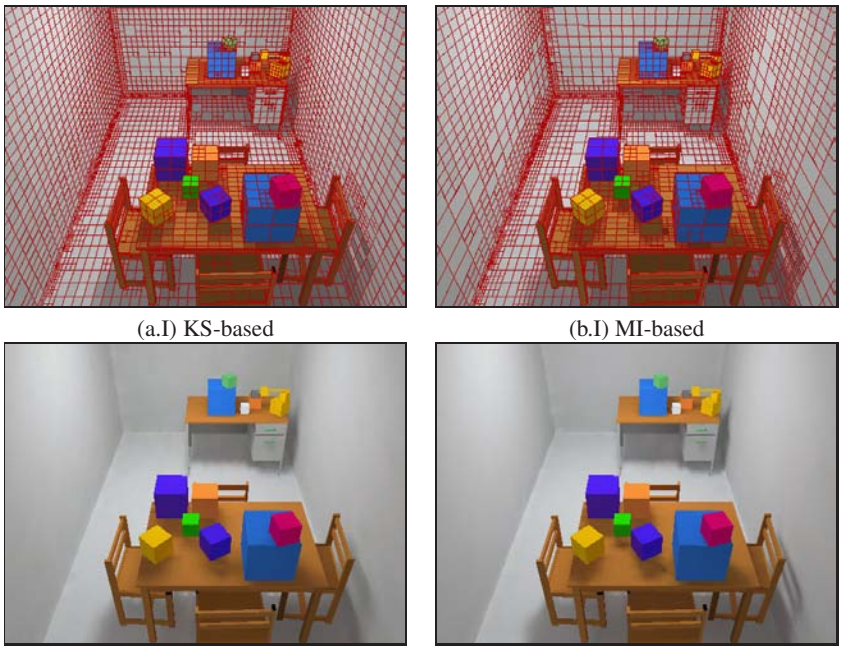
[FP93] FEDA M., PURGATHOFER W.: Progressive ray refinement for Monte Carlo radiosity. In *Eurographics Rendering Workshop 1993* (June 1993), Cohen M. F., Puech C., Sillion F. X., (Eds.), Eurographics Association, pp. 15–26. Held in Paris, France.

[FS98] FEIXAS M., SBERT M.: *Scene Continuous Mutual Information as Least Upper Bound of Discrete One*. Research Report IIiA–98–27–RR, Institut d'Informàtica i Aplicacions, Universitat de Girona, Girona, Spain, 1998.

[Gla95] GLASSNER A. S.: *Principles of Digital Image Synthesis*. Morgan Kaufmann Publishers, San Francisco (CA), USA, 1995.

[GSCH93] GORTLER S. J., SCHRÖDER P., COHEN M. F., HANRAHAN P.: Wavelet radiosity. *Computer Graphics (Proceedings of SIGGRAPH'93) 27* (August 1993), 221–230. Held in Anaheim (CA), USA.

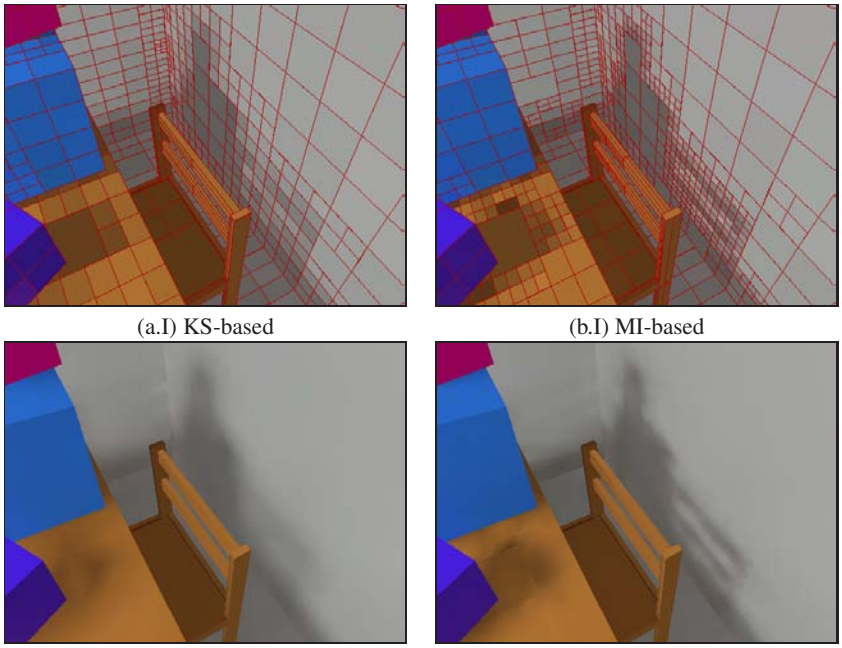
[GTGB84] GORAL C. M., TORRANCE K. E., GREENBERG D. P., BATTAILE B.: Modelling the interaction of light between diffuse surfaces. *Computer Graphics (Proceedings of SIGGRAPH'84) 18*, 3 (July 1984), 213–222. Held in Minneapolis (MN), USA.



(a.I) KS-based (b.I) MI-based

(a.II) KS-based (b.II) MI-based

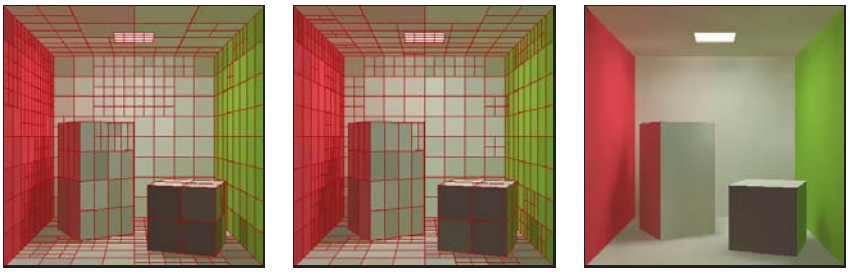
Figure 22: *KS-based (a) and MI-based (b) methods with the cube room scene showing the mesh (I) and Gouraud shaded solution (II). The number of patches is 13902 and 13878, respectively. For each scene, we cast 402650 rays for radiosity computation and 10 rays for each oracle evaluation.*



(a.I) KS-based (b.I) MI-based

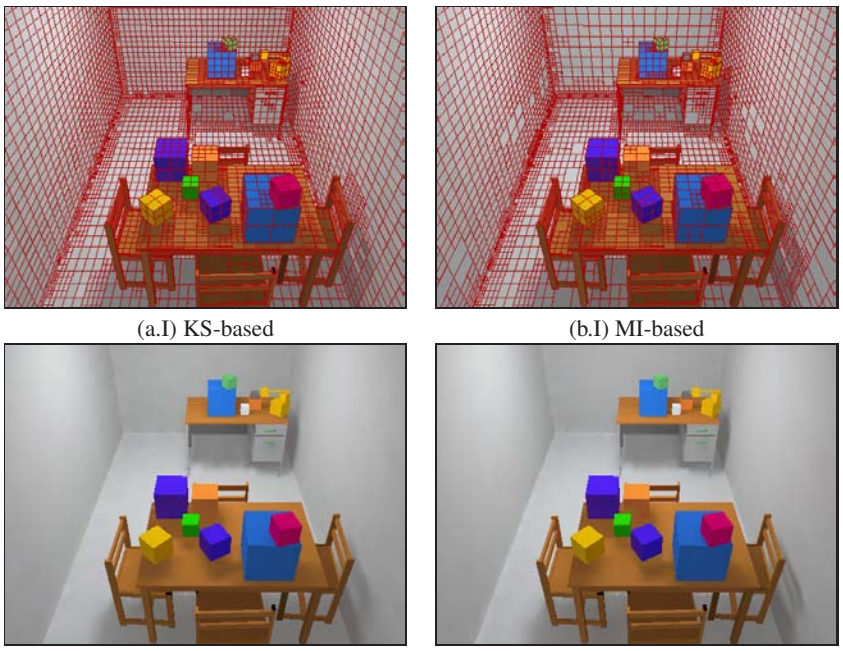
(a.II) KS-based (b.II) MI-based

Figure 23: *A different view of the scene shown in Figure 22.*



(a) KS-based (b) MI-based (c) MI-based

Figure 24: *KS-based (a) and MI-based (b) methods with the Cornell box scene showing the mesh. The number of patches is 875 and 891, respectively. For each scene, we cast 19458 rays for radiosity computation and 4 rays for each oracle evaluation. The Gouraud shaded solution for (b) is shown in (c).*



(a.I) KS-based (b.I) MI-based

(a.II) KS-based (b.II) MI-based

Figure 25: *KS-based (a) and MI-based (b) methods with the cube room scene showing the mesh (I) and Gouraud shaded solution (II). The number of patches is* 13690 *and* 13758*, respectively. For each scene, we cast* 402565 *rays for radiosity computation and* 4 *rays for each oracle evaluation.*

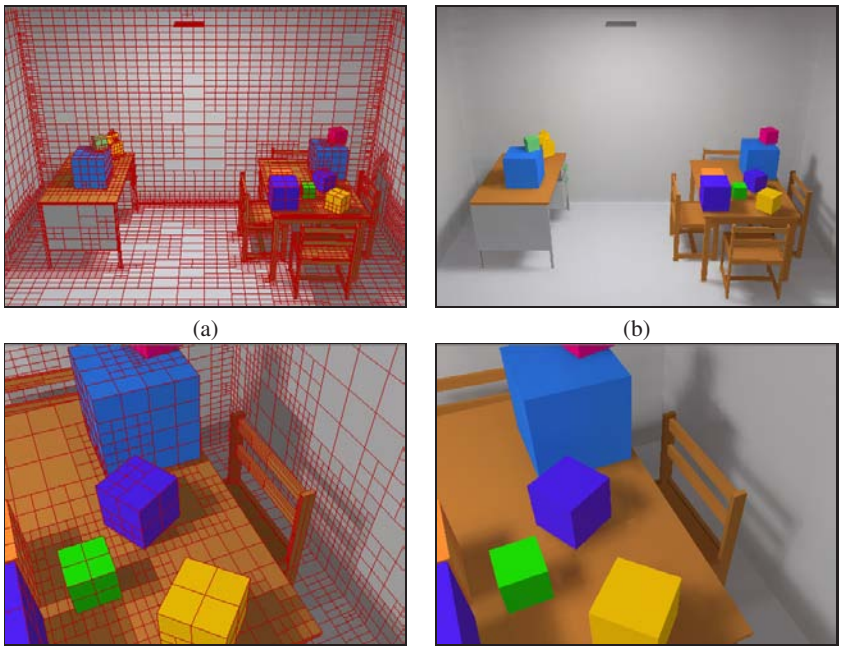


Figure 26: *MI-based method with the scene shown in Figure 22. The number of patches is* 18338*. We cast* 2684260 *rays for radiosity computation and* 10 *rays for each oracle evaluation.*

[HS98] HOLZSCHUCH N., SILLION F.: An exhaustive error-bounding algorithm for hierarchical radiosity. *Computer Graphics Forum 17*, 4 (1998), 197–218.

[HSA91] HANRAHAN P., SALZMAN D., AUPPERLE L.: A rapid hierarchical radiosity algorithm. *Computer Graphics (Proceedings of SIGGRAPH'91) 25*, 4 (July 1991), 197–206. Held in Las Vegas (NV), USA.

[Kaj86] KAJIYA J. T.: The rendering equation. *Computer Graphics (Proceedings of SIGGRAPH'86) 20*, 4 (August 1986), 143–150. Held in Dallas (TX), USA.

[KW86] KALOS M. H., WHITLOCK P. A.: *The Monte Carlo Method*. John Wiley & Sons, 1986.

[Li90] LI W.: Mutual information functions versus correlation functions. *Journal of Statistical Physics 60*, 5/6 (1990), 823–837.

[Li91] LI W.: On the relationship between complexity and entropy for markov chains and regular languages. *Complex Systems 5*, 4 (1991), 381–399.

[LSG94] LISCHINSKI D., SMITS B., GREENBERG D. P.: Bounds and error estimates for radiosity. *Computer Graphics (Proceedings of SIGGRAPH'94) 28* (July 1994), 67–74. Held in Orlando (FL), USA.

[LTG93] LISCHINSKI D., TAMPIERI F., GREENBERG D. P.: Combining hierarchical radiosity and discontinuity meshing. *Computer Graphics (Proceedings of SIGGRAPH'93) 27* (August 1993), 199–208. Held in Anaheim (CA), USA.

[MR95] MOTWANI R., RAGHAVAN P.: *Randomized Algorithms*. Cambridge University Press, New York (NY), USA, 1995.

[Neu95] NEUMANN L.: Monte Carlo radiosity. *Computing 55*, 1 (1995), 23–42.

[NFP95] NEUMANN L., FEDA M., PURGATHOFER W.: A new stochastic radiosity method for highly complex scenes. In *Photorealistic Rendering Techniques (Proceedings of the 5th Eurographics Workshop on Rendering)* (New York (NY), USA, 1995), Haas S., Mueller S., Sakas G., Shirley P., (Eds.), Springer-Verlag Vienna-New York, pp. 201–213. Held in Darmstad, Germany (June 1994).

[NN85] NISHITA T., NAKAME E.: Continuous tone representation of 3-d objects taking account of shadows and interreflection. *Computer Graphics (Proceedings of SIGGRAPH'85) 19*, 3 (July 1985), 23–30. Held in San Francisco (CA), USA.

[NNB97] NEUMANN L., NEUMANN A., BEKAERT P.: Radiosity with well distributed ray sets. Fellner D., Szirmay-Kalos L., (Eds.), vol. 16, Blackwell Publishers, pp. 261–270. Held in Budapest, Hungary.

[Pap84] PAPOULIS A.: *Probability, Random Variables,*

and Stochastic Processes, 2 ed. McGraw-Hill, New York (NY), USA, 1984.

[PB95] PATTANAIK S. N., BOUATOUCH K.: Linear radiosity with error estimation. In *Rendering Techniques'95 (Proceedings of the 6th Eurographics Workshop on Rendering)* (New York (NY), USA, June 1995), Hanrahan P., Purgathofer W., (Eds.), Springer-Verlag Vienna-New York. Held in Dublin, Eire.

[PM92] PATTANAIK S. N., MUDUR S. P.: Computation of global illumination by Monte Carlo simulation of the particle model of light. In *Proceedings of the 3th Eurographics Workshop on Rendering* (May 1992), pp. 71–83. Held in Bristol, UK.

[Rub81] RUBINSTEIN R. Y.: *Simulation and the Monte Carlo Method*. John Wiley & Sons, New York (NY), USA, 1981.

[San76] SANTALÓ L. A.: *Integral Geometry and Geometric Probability*. Addison-Wesley, Reading (MA), USA, 1976.

[SAS92] SMITS B. E., ARVO J., SALESIN D.: An importance-driven radiosity algorithm. *Computer Graphics (Proceedings of SIGGRAPH'92) 26*, 2 (July 1992), 273–282. Held in Chicago (IL), USA.

[Sbe93] SBERT M.: An integral geometry based method for fast form-factor computation. *Computer Graphics Forum (Proceedings of Eurographics'93) 12*, 3 (1993), 409–420. Held in Barcelona, Spain.

[Sbe96] SBERT M.: *The Use of Global Random Directions to Compute Radiosity. Global Monte Carlo Methods*. PhD thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, November 1996.

[Sbe97] SBERT M.: Error and complexity of random walk Monte Carlo radiosity. *IEEE Transactions on Visualization and Computer Graphics 3*, 1 (March 1997), 23–38.

[SC99] SHALIZI C. R., CRUTCHFIELD J. P.: *Computational Mechanics: Pattern and Prediction, Structure and Simplicity*. Working Paper 99–07–044, Santa Fe Insitute, Santa Fe (NM), USA, July 1999.

[SH92] SIEGEL R., HOWELL J. R.: *Thermal Radiation Heat Transfer*, 3 ed. Hemisphere Publishing Corporation, New York (NY), USA, 1992.

[SH93] SCHRÖDER P., HANRAHAN P.: On the form factor between two polygons. *Computer Graphics Proceedings (Proceedings of SIGGRAPH'93) 27* (August 1993), 163–164. Held in Anaheim (CA), USA.

[Shi90] SHIRLEY P.: A ray tracing method for illumination calculation in diffuse–specular scenes. In *Proceedings of Graphics Interface'90* (May 1990), pp. 205–212. Held in Halifax, Canada.

[SP94] SILLION F. X., PUECH C.: *Radiosity and Global Illumination*. Morgan Kaufmann Publishers, San Francisco (CA), USA, 1994.

[SSS97] STAMMINGER M., SLUSALLEK P., SEIDEL H.-P.: Bounded radiosity-illumination on general surfaces and clusters. *Computer Graphics Forum (Proceedings of Eurographics'97) 16*, 3 (1997), 300–317. Held in Budapest, Hungary.

[Stu97] STUDHOLME C.: *Measures of 3D Medical Image Alignment*. PhD thesis, University of London, London, UK, August 1997.

[TH93] TELLER S. J., HANRAHAN P.: Global visibility algorithms for illumination computation. *Computer Graphics (Proceedings of SIGGRAPH'93) 27* (August 1993), 239–246. Held in Anaheim (CA), USA.

Applications of Information Theory to Computer Graphics Part III-IV: Refinement Criteria Based on *f*-Divergences

Jaume Rigau, Miquel Feixas, and Mateu Sbert

University of Girona, Spain

1. Introduction

Just as we have seen in the previous part, when sampling a signal we need a criterion to decide whether to take additional samples, albeit within the original domain or within a hierarchical subdivision. The refinement criteria are mainly based on the homogeneity encountered in the samples. Heterogeneity should lead to further sampling, possibly with an adaptive subdivision of the domain. Oracles are then built based on these criteria.

In this part, we introduce new refinement criteria based on *f-divergence*s. The introduction of these measures is motivated by the observation that the mutual information-based oracle (see Part III) can be rewritten as an *f*-divergence. *f*-Divergences are a family of convex functions that possess very remarkable properties. They were introduced by Csiszár [Csi63] and Ali and Silvey [AS66] as measures of discrimination or distance between probability distributions and have been successfully used in image processing and several engineering areas [OBS98, LeG99, HMMG01, Plu01].

Our purpose is to demonstrate the usefulness of *f*-divergences in computer graphics by applying them to defining new refinement criteria for the techniques of the previous chapters: hierarchical radiosity (§4) and adaptive sampling in ray-tracing (§5). We consider that some divergences are perfectly fitted as homogeneity measures, when we consider how distant the distribution of the samples is with respect to the uniform distribution. We will see how, compared with classic refinement criteria, the *f*-divergence-based ones give significantly better results. Previously, we give a brief introduction to divergence measures (§2) and in particular to the *f*-divergences (§3).

2. Divergence Measures

In this section, we establish the semantics of divergence measures (§2.1) and we present three specific types of them (§2.2).

2.1. Concept

What does "divergence" mean? One brief definition for divergence is *"a deviation from a course or standard"* [Mer05]. In general, the difference in shades of meaning between words such as difference, dissimilarity, distance, and divergence are so subtle that we end up considering them practically synonyms in every day language. In a statistical context, the objective is to measure the level of separation between two elements of a sampling. Depending on the properties that make up the measure, it can be qualified in one sense or another [Gow85, BB97, Mar01].

Let X be a nonempty set and $d : X^2 \to \mathbb{R}$ a function. Then, d is a measure of

Difference If it fulfils

- *Symmetry*: $d(x,y) = d(y,x) \quad \forall x,y \in X$
- *Minimum difference*: $d(x,y) \geq d(x,x) \quad \forall x,y \in X$

Dissimilarity If it is a difference measure which fulfils

- *Non-negativity*: $d(x,y) \geq 0 \quad \forall x,y \in X$
- *Self-similarity*: $d(x,x) = 0 \quad \forall x \in X$

Distance Also called a *metric*, if it is a dissimilarity measure which fulfils ($\forall x,y,z \in X$)

- *Defined*: $d(x,y) = 0 \Rightarrow x = y$
- *Triangle inequality*: $d(x,y) + d(y,z) \geq d(x,z)$

When the objective of the measure is to reflect the discrepancy or difference between two probability distributions, it is called divergence (without loss of generality, divergences are limited to discrete probability distributions) [Bur83]:

Definition 1 Let $\mathcal{X}$ be a countable observation space with $n > 1$ elements and $\mathcal{P}$ the set of all the possible probability distributions of $\mathcal{X}$. Then, $D : \mathcal{P}^2 \to \mathbb{R}^+ \cup \{0\}$ is a divergence if, for all $(p,q) \in \mathcal{P}^2$, it fulfils:

- $D(\{p_1,\ldots,p_n\},\{q_1,\ldots,q_n\})$ is a continuous function of its $2 \cdot n$ variables.
- $D(p,q)$ is invariant under the permutations of the pairs (p_i,q_i) for $i \in \{1,\ldots,n\}$.

- $D(\{p_1,\ldots,p_n,0\},\{q_1,\ldots,q_n,0\}) = D(p,q)$.
- $D(p,q) \geq 0$.
- $D(p,q) = 0 \Leftrightarrow p = q$.

Note that if a divergence were symmetric it would be equivalent to a defined dissimilarity which only lacks the triangle inequality to attain the category of metric.

2.2. Divergence Classes

With the previous definition of divergence, it is possible to obtain a large set of divergence measures $D(p,q)$. In general, the problem consists in discerning the suitable measures for every specific case. Convexity is a desirable property. A generalisation of it, Jensen's inequality (see Part II), is widely used in mathematics, information theory, and different engineering areas as a kernel of divergence measures. For example, it has been successfully applied to image registration [HHK03] and DNA segmentation [BOR99].

We now see particular examples of divergences (following Pardo [PV03]). From the perspective of information theory, the importance of the information divergence, or Kullback-Leibler distance, is objectively accepted. In mathematical statistics, the same role is played by the chi-square divergence. Also, in convex analysis, arithmetic and geometric means are used in the arithmetic-geometric divergence. In the probability theory, the Vasershtein-Ornstein divergence plays an important role together with L_α-norm divergence, which is also very useful in statistics and other mathematical areas (see Table 1).

These examples and many others are special cases of a kind of divergence measure which obeys the scheme (in this context, we keep the usual functional notation for the probability distributions).

$$D(p,q) = \sum_{x\in\mathcal{X}} \varphi(p(x),q(x)) \tag{1}$$

for a given real function $\varphi(u,v)$ of positive variables u,v. This function is assumed to be extended to $[0,\infty)\times[0,\infty)$ taking:

$$\varphi(0,0) = 0 \quad \varphi(0,v) = \lim_{u\to 0^+} \varphi(u,v) \quad \varphi(u,0) = \lim_{v\to 0^+} \varphi(u,v) \tag{2}$$

for $u,v > 0$ where the limits may be infinite.

This scheme has been introduced in information theory for some classes of functions $\varphi(u,v)$ where $(u,v) \in [0,1]^2$. An important case is the kind of divergences that can be generated with the help of a convex function $f : (0,\infty) \to (-\infty,\infty)$, extended to $[0,\infty] \to (-\infty,\infty]$ by the continuity rules:

$$f(0) = \lim_{x\to 0^+} f(x) \qquad f(\infty) = \lim_{x\to\infty} f(x). \tag{3}$$

Let $\mathcal{F}$ be the set of these functions f which are twice differentiable with continuous derivatives satisfying $f(1) = 0$ and $f''(1) > 0$. Using $f \in \mathcal{F}$, the following divergences are defined in accordance with (1):

Csiszár divergences $\varphi(u,v) = vf\left(\frac{u}{v}\right)$
Bregman divergences $\varphi(u,v) = f(u) - f(v) - f'(v)(u-v)$
Burbea-Rao divergences $\varphi(u,v) = \frac{f(u)+f(v)}{2} - f\left(\frac{u+v}{2}\right)$

These three kinds are partially overlapping [PV97]. We focus our attention on Csiszár divergences [Csi63, Csi67], also called *f-divergences* and denoted by $D_f(p,q)$.

3. *f*-Divergences

f-Divergences are based on convex functions and were independently introduced by Csiszár [Csi63], and Ali and Silvey [AS66]. These measures have been applied to different areas, such as medical image registration [Plu01] and classification and retrieval [HMMG01], among others. We introduce, in this section, the definition, properties, and particular instances which we will use in the following sections [LV87, Ö02a]. Accordingly §2.2, we take $\varphi(u,v) = vf\left(\frac{u}{v}\right)$ and

Definition 2 Let $(p,q) \in \mathcal{P}^2$ and $f \in \mathcal{F}$. The f-divergence of the probability distributions p and q is given by

$$D_f(p,q) = \sum_{x\in\mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right). \tag{4}$$

By extension rules (2), for $p,q > 0$

$$0f\left(\frac{0}{0}\right) = 0 \quad qf\left(\frac{0}{q}\right) = qf(0) \quad 0f\left(\frac{p}{0}\right) = p\lim_{y\to\infty}\frac{f(y)}{y}. \tag{5}$$

Then, $D_f(p,q)$ is well defined as a divergence measure [Vaj89]. Two important properties are:

- $D_f(p,q)$ is convex on (p,q).
  If (p,q) and (p',q') are two pairs of probability distributions, then
  $$\lambda D_f(p,q) + (1-\lambda) D_f(p',q') \geq D_f(\lambda p + (1-\lambda)p', \lambda q + (1-\lambda)q'). \tag{6}$$
- Uniqueness.
  If $f,g \in \mathcal{F}$, then $D_f(p,q) = D_g(p,q) \Leftrightarrow \exists r \in \mathbb{R}.\ f(u) - g(u) = r(u-1)$.

f-Divergences have been studied in depth. The Research Group in Mathematical Inequalities and Applications (Victoria University, Melbourne, Australia) deserves a special mention since over recent years its members have made many contributions to this area [OV93, Dra00, BCDS02, Ö02a]. *f*-Divergences can be grouped together in terms of their convex functions. Considering the classification of Österreicher [Ö02a], we have the following types: χ^α-divergences, (symmetrised) dichotomy, Matusita's divergences, elementary divergences, Puri-Vincze Divergences, and Divergences of Arimoto-type. Within each type, other families of *f*-divergences can be created. We should mention particularly the subtype of f^α-divergences (dichotomy class) presented by Liese and Vajda [LV87].

divergence	*field*	*definition*
information	information theory	$I(p,q) = \sum_{x\in\mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$
chi-squared	statistics	$\chi^2(p,q) = \sum_{x\in\mathcal{X}} \frac{(p(x)-q(x))^2}{q(x)}$
arithmetic-geometric	convex analysis	$AG(p,q) = \sum_{x\in\mathcal{X}} \ln \frac{(p(x)+q(x))/2}{\sqrt{p(x)q(x)}}$
Vasershtein-Ornstein	probability theory	$VO(p,q) = 1 - \sum_{x\in\mathcal{X}} \min\{p(x),q(x)\}$
L_α-norm	mathematics	$L_\alpha(p,q) = \sum_{x\in\mathcal{X}} \mid p(x)-q(x)\mid^\alpha \quad \alpha \in \{1,2\}$

Table 1: *A subset of useful divergences with its most common fields of application.*

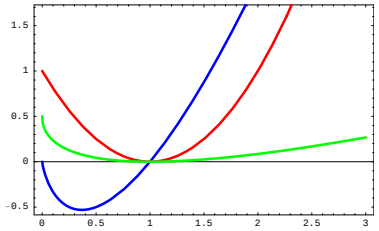


Figure 1: *Plot for $x \in [0,3]$ of three strictly convex functions: $u\log u$ (blue), $(u-1)^2$ (red), and $\frac{1}{2}(\sqrt{u}-1)^2$ (green). The shared intersection is $f(1) = 0$. From these functions, the Kullback-Leibler, chi-square, and the Hellinger f-divergences are obtained, respectively.*

Next, we select three of the most important f-divergences [Dra00, GS01, Ö02a], called "distances" in the literature. They are built up from the convex functions in Fig. 1:

Kullback-Leibler $f(u) = u\log u$:

$$D_{\mathrm{KL}}(p,q) = \sum_{x\in\mathcal{X}} p(x) \log \frac{p(x)}{q(x)}. \tag{7}$$

Introduced by Kullback and Leibler [KL51], it corresponds to the relative entropy or Kullback-Leibler distance (see Part II). Based on continuity arguments, $0\log\frac{0}{q(x)} = 0$ for all $q(x)$, and $p(x)\log\frac{p(x)}{0} = \infty$ for all $p(x) > 0$ (2). Hence, the measure takes values in $[0,\infty]$. It is not a metric, since it is not symmetric and does not satisfy triangle inequality but, despite of this, it has many useful properties [Rei89, CT91, GS01]. A square root version of Kullback-Leibler divergence has been used by Yang and Barron [YB99]. In Fig. 2.*a* we show the behaviour of this divergence by means of the contribution of a pair $(p(x),q(x))$. The maximum contribution is ∞ in $(p(x),0)$, and the minimum is $-(e\ln 2)^{-1} \approx -0.531$ in $(\frac{1}{e},1)$. The contribution is null for any pair where $p(x) = 0$ or $p(x) = q(x)$. Note the relevance which the divergence takes with respect to p.

Chi-square $f(u) = (u-1)^2$:

$$D_{\chi^2}(p,q) = \sum_{x\in\mathcal{X}} \frac{(p(x)-q(x))^2}{q(x)}. \tag{8}$$

Defined by Pearson [Pea00] (the history of this measure can be found in Liese and Vajda [LV87]), this measure takes values in $[0,\infty]$ due to the limit when $q(x) = 0$ (2). It is not symmetric. Reiss [Rei89] defined a divergence using the square root of D_{χ^2}. In Fig. 2.*b* we show the same representation as in the previous divergence. The maximum contribution is also ∞ in $(p(x),0)$, but the minimum is 0 and it is attained in all the pairs where $p(x) = q(x)$. Observe that, even though a strong relevance with respect to p is maintained, the values of q take on more importance than in the case of D_{KL}.

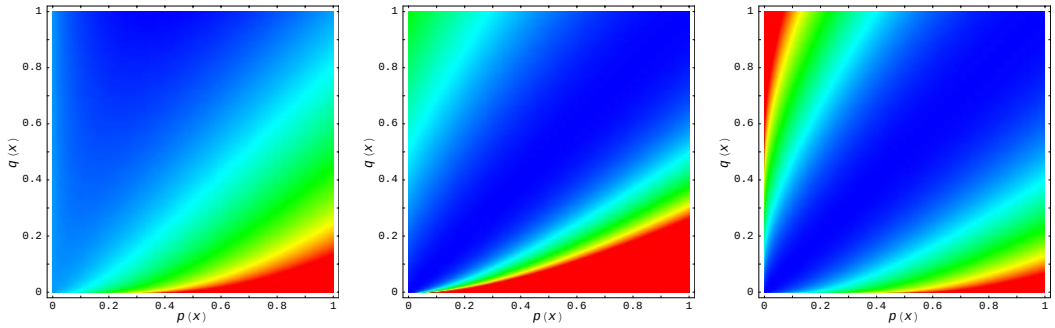
Hellinger $f(u) = \frac{1}{2}(\sqrt{u}-1)^2$:

$$D_{h^2}(p,q) = \frac{1}{2}\sum_{x\in\mathcal{X}} \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2. \tag{9}$$

The origins are in Hellinger [Hel09] (historical references can be found in Liese and Vajda [LV87] and Le Cam and Yang [LY90]). This symmetric measure takes values in $[0,1]$ due to the normalisation factor of $\frac{1}{2}$. If it is omitted in f, we obtain the general Hellinger form $2D_{h^2}$. The quantity $1 - D_{h^2}$ is called the Hellinger affinity, a measure popularised by Kakutani [Kak48] who also applied the square root to the general form of Hellinger obtaining a metric [Top00] (the normalisation factor is $\frac{1}{\sqrt{2}}$). In Fig. 2.*c*, the contribution of each pair is shown. The maximum contribution is $\frac{1}{2}$ in $(1,0)$ and $(0,1)$ and the minimum is 0 when $p(x) = q(x)$. Note how the relevance between p and q has balanced out due to the symmetry.

However, none of the above f-divergences are true distances. In Österreicher [Ö02b] there is a discussion about which f-divergences have a metric behaviour. Gibbs and Su [GS01] provide a summary of bounds between probability metrics and distances. Three relationships between the f-divergences presented are (Fig. 3):

- $D_{h^2}(p,q) \le \frac{1}{2} D_{\mathrm{KL}}(p,q)$ [Rei89]
- $D_{h^2}(p,q) \le \sqrt{D_{\chi^2}(p,q)}$ [Rei89]
- $D_{\mathrm{KL}}(p,q) \le \log(1 + D_{\chi^2}(p,q))$ [GS01]



(a) D_{KL} (b) D_{χ^2} (c) D_{h^2}

Figure 2: *Density maps of the contribution of a pair $(p(x),q(x))$ for all $x \in \mathcal{X}$: (a) D_{KL}, (b) D_{χ^2}, and (c) D_{h^2}.*

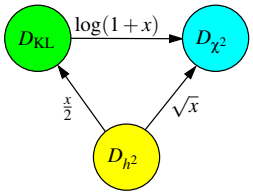


Figure 3: *Bounds between D_{KL}, D_{χ^2}, and D_{h^2}, where $D_f \xrightarrow{h} D_g$ means $D_f(p,q) \leq h(D_f(p,q))$.*
Credit: Adapted from Gibbs and Su [GS01].

4. *f*-Divergences in Radiosity

In this section, new refinement criteria based on *f*-divergences are introduced for hierarchical radiosity (object-space approach).

4.1. Method

Analysing the mutual information-based oracle $\rho_i \delta_{ij} B_j < \varepsilon$ (see Part III) [FRBS02] we observe that it can be rewritten from a Kullback-Leibler distance. In fact, the kernel of the oracle is based on the mutual information, which at the same time is defined as a the Kullback-Leibler distance and which, in accordance with (7), belongs to the *f*-divergences family. In order to obtain the new expression, we need to make the following considerations:

- Let $|\mathcal{S}_{i\times j}| = N_s$ be the number of samples of the area sampling form factor computation.
- Let $\widehat{F} = \sum_{k=1}^{N_s} F_{x_k \leftrightarrow y_k}$ be the form factor estimation where $(x_k, y_k) \in \mathcal{S}_{i\times j}$ ($\widehat{F} \approx \frac{F_{ji}}{A_i} = \frac{F_{ij}}{A_j}$).
- Let $p = \{p_k = \frac{F_{x_k \leftrightarrow y_k}}{\widehat{F}} \mid 1 \leq k \leq N_s\}$ be the probability distribution given for the contribution of every sample to $\widehat{F}$.

Note that $\mathrm{avg}_{1\leq k\leq N_s}\{p_k\} = \frac{1}{N_s}$. Then, from the discretisation error (see Part III)

$$\delta_{ij} \approx \frac{A_i A_j}{A_T} \left(\frac{1}{|\mathcal{S}_{i\times j}|} \left(\sum_{(x,y)\in\mathcal{S}_{i\times j}} F_{x\leftrightarrow y} \log F_{x\leftrightarrow y} \right) - \left(\frac{1}{|\mathcal{S}_{i\times j}|} \sum_{(x,y)\in\mathcal{S}_{i\times j}} F_{x\leftrightarrow y} \right) \log \left(\frac{1}{|\mathcal{S}_{i\times j}|} \sum_{(x,y)\in\mathcal{S}_{i\times j}} F_{x\leftrightarrow y} \right) \right), \tag{10}$$

we can rewrite

$$\begin{aligned} \delta_{ij} &\approx \frac{A_i A_j}{A_T} (\mathrm{avg}_{1\leq k\leq N_s}\{F_{x_k\leftrightarrow y_k} \log F_{x_k\leftrightarrow y_k}\} \\ &\quad - \mathrm{avg}_{1\leq k\leq N_s}\{F_{x_k\leftrightarrow y_k}\} \log \mathrm{avg}_{1\leq k\leq N_s}\{F_{x_k\leftrightarrow y_k}\}) \qquad (11) \\ &= \frac{A_i A_j}{A_T} \widehat{F} (\mathrm{avg}_{1\leq k\leq N_s}\{p_k \log p_k\} \\ &\quad - \mathrm{avg}_{1\leq k\leq N_s}\{p_k\} \log \mathrm{avg}_{1\leq k\leq N_s}\{p_k\}) \qquad (12) \\ &= \frac{A_i A_j}{A_T} \widehat{F} \left(\mathrm{avg}_{1\leq k\leq N_s} p_k \log p_k - \frac{1}{N_s} \log \frac{1}{N_s} \right) \\ &= \frac{A_i A_j}{A_T N_s} \widehat{F} \left(\left(\sum_{1\leq k\leq N_s} p_k \log p_k \right) - \log \frac{1}{N_s} \right) \\ &= \frac{A_i A_j}{A_T N_s} \widehat{F}\, D_{\mathrm{KL}}(p,q), \qquad (13) \end{aligned}$$

where $q = \{q_k = \frac{1}{N_s} \mid 1 \leq k \leq N_s\}$ is the uniform distribution.

This fact suggests that we try other *f*-divergences in the kernel of the refinement oracle. These measures will give us the distance of the distribution of the point-to-point form factors, p, with respect to the uniform distribution, q. Thus, the Kullback-Leibler (7), chi-square (8), and Hellinger (9) distances have been tested. The Kullback-Leibler-based oracle has already been studied in [FRBS02, Fei02] from an information-theoretic perspective.

Definition 3 Three oracles for hierarchical radiosity, based on their respective f-divergences, are given by

- Kullback-Leibler divergence

$$\rho_i A_i A_j \widehat{F} D_{KL}(p,q) B_j < \varepsilon \qquad \text{(KL)}$$

- Chi-square divergence

$$\rho_i A_i A_j \widehat{F} \chi^2(p,q) B_j < \varepsilon \qquad \text{(CS)}$$

- Hellinger divergence

$$\rho_i A_i A_j \widehat{F} D^2_{h^2}(p,q) B_j < \varepsilon \qquad \text{(HL)}$$

Observe that the constants $\frac{1}{A_T}$ and $\frac{1}{N_s}$ have been removed since they are specific constants for each scene and are implicit in the threshold.

It is important to note that the expression between parenthesis in (11) corresponds to Jensen's inequality with $f(x) = x \log x$ and $x \in p$. Moreover, we can also see that expression (12) is equal to the first term of the log-sum inequality (see Part II), taking $a_i = p_k$, $b_i = 1$, and $n = N_s$. Thus, $\delta_{ij} \geq 0$.

4.2. Results

For comparative effects, the kernel-smoothness-based oracle,

$$\rho_i \max\{F_{ij}^{\max} - F_{ij}^{\text{avg}}, F_{ij}^{\text{avg}} - F_{ij}^{\min}\} A_j B_j < \varepsilon, \qquad \text{(KS)}$$

is chosen as a representative of the oracles which work evaluating the variation of the radiosity kernel between a pair of elements (see Part III). This oracle and the *f*-divergence-based oracles have been implemented on top of the hierarchical Monte Carlo radiosity method. It should be noted that our oracles can be used with any hierarchical radiosity method.

In Fig. 4 we show a general view of the test scene obtained with the KL oracle. The left image (*i*) shows the Gouraud shaded solution, while the right one (*ii*) corresponds to the subdivision obtained. Each oracle has been evaluated with: $N_s = 10$ random lines between the corresponding pair of elements; an average of 2,684,000 rays to distribute the power have been cast for each solution; and the ε parameter has been tuned so that the meshes obtained have approximately 19,000 patches in all the methods.

For another view of the test scene, we present the results obtained with the *f*-divergence-based oracles KL, CS, and HL (Figs. 5.*a*–*c*, respectively) and the KS oracle (Fig. 6.*a*). We can see how the *f*-divergence-based oracles outperform the KS one, working the more complex light zones better and obtaining an improved sharpness in the objects. The meshes created are of higher quality and their precision in the corners and in the transitions of light show this feature. On the other hand, comparing our three *f*-divergence oracles we conclude that, although they exhibit a similar quality, the KL one is slightly better. For instance, observe that the shadows on the table are more defined. A possible explanation for this better behaviour could be that the KL oracle, unlike the other ones, meets the conditions of Jensen's inequality. This confers a distinct theoretical advantage on this oracle.

From the above, one could be tempted to use Jensen's inequality alone as a kernel for a refinement oracle. We have experimented with the function $f(x) = x^2$, which when substituted in Jensen's inequality, corresponds to the variance. Thus, substituting $F_{x_k \leftrightarrow y_k} \log F_{x_k \leftrightarrow y_k}$ by $F^2_{x_k \leftrightarrow y_k}$ in (11), the variance-based oracle is given by

$$\rho_i A_i A_j \widehat{F}^2 V(p,q) B_j < \varepsilon, \qquad \text{(VT)}$$

where $V(p,q) = \text{avg}_{1 \leq k \leq N_s}\{p_k^2\} - \left(\frac{1}{N_s}\right)^2$. The results obtained are presented in Fig. 6.*b*, showing the inadequacy of this function.

In Table 2, the results of the previous methods are evaluated with the RMSE and PSNR measures. The improvement with regard to the KS and VT is reflected and the results of the KL oracle are noteworthy.

5. *f*-Divergences in Adaptive Sampling for Ray-Tracing

In this section, we apply the *f*-divergences to the refinement criteria based on ray-tracing (pixel-driven approach). To do this, we incorporate the divergences into the adaptive sampling scheme using the same basic idea as in hierarchical radiosity (§4) but considering the luminance information instead of the geometric information of the form factors. Therefore, we evaluate the homogeneity of a region of the image plane in accordance with the divergence between its luminance distribution and the uniform distribution. To make the comparison easier, we use the same framework as in the entropy-based refinement criteria for ray-tracing (see Part IV).

5.1. Method

The *f*-divergences defined in §3 will be used to evaluate the heterogeneity of a set of samples in a region. The scheme used is the following:

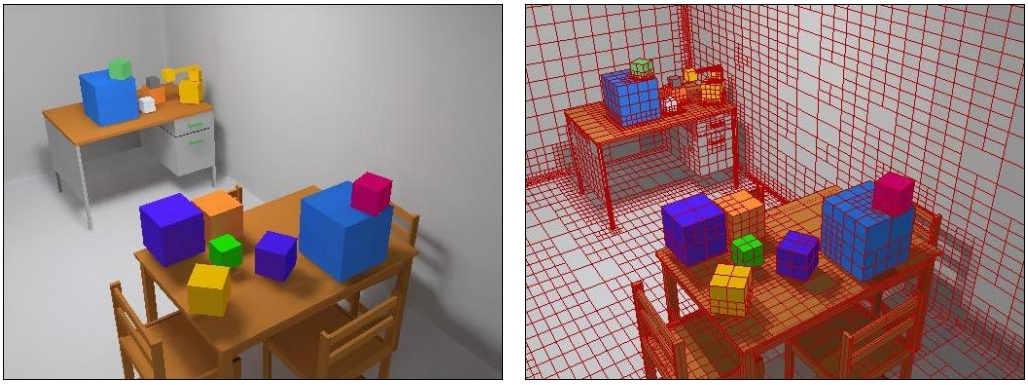
1. A first batch of N_s^p samples is cast through a pixel and the corresponding luminances $L_{i \in \{1,\ldots,N_s^p\}}$ are obtained. For an sRGB colour system, the luminance corresponds to the value of Y in the corresponding XYZ system [Com98].
2. The *f*-divergences $D_f(p,q)$ are taken between the normalised distribution of the obtained luminances,

$$p = \{p_i = \frac{L_i}{\sum_{j=1}^{N_s^p} L_j} \mid 1 \leq i \leq N_s^p\}, \qquad (14)$$

   and the uniform distribution $q = \{q_i = \frac{1}{N_s^p} \mid 1 \leq i \leq N_s^p\}$.
3. The refinement criterion, given by

$$\frac{1}{N_s^p} \overline{L} D_f(p,q) < \varepsilon \qquad (15)$$

   is evaluated, where D_f represents the Kullback-Leibler,



(*i*) KL (*ii*) KL

Figure 4: *KL oracle for the test scene: (i) Gouraud shaded solution and (ii) its final adaptive mesh. The oracle has been evaluated with 10 random lines between elements.*

oracle	*view*$_1$				*view*$_2$			
	RMSE_a	RMSE_p	PSNR_a	PSNR_p	RMSE_a	RMSE_p	PSNR_a	PSNR_p
KL	9.475	8.698	28.599	29.342	9.712	8.956	28.385	29.088
CS	10.097	9.355	28.047	28.710	10.556	9.825	27.661	28.284
HL	9.990	9.217	28.139	28.839	10.404	9.687	27.787	28.407
KS	13.791	13.128	25.339	25.767	15.167	14.354	24.513	24.991
VT	16.414	15.898	23.826	24.104	17.829	17.378	23.108	23.331

Table 2: *The RMSE and PSNR measures of the* f*-divergence-based, KS, and VT oracles applied to the test scene. A set of images are shown in Fig. 4 (KL), Fig. 5 (KL, CS, and HL), and Fig. 6 (KS and VT). The oracles have been evaluated with 10 random lines between elements.*

chi-square, or Hellinger divergences, $\overline{L}$ is the average luminance

$$\overline{L} = \frac{1}{N_s^p} \sum_{i=1}^{N_s^p} L_i, \tag{16}$$

and ε is a predefined threshold for the refinement test. The divergence measure $D_f(p,q)$ in the kernel plays the role of a contrast. Note that to assign an importance to this value, we weight it with the average luminance (16), as in Glassner's version of classic contrast [Gla95], used also in the method CC (see Part IV). Division by the number of samples N_s^p in (15) ensures that the refinement process stops.

4. Successive batches of N_s^p rays are cast until the result of the test is true and no more refinement is necessary.

The new criteria give good visual results, but the RMSE obtained in our tests (see Table 3), although better than for the classic contrast, is higher than with the confidence test criterion (see Part IV). Our next logical step was to try the square root of Hellinger divergence, as it is a true metric. The results obtained were very encouraging and, by analogy, we extended the experimentation to the square root of the other divergences (also used in other fields [Rei89, YB99, Top00]). The results also improved the previous ones and were also better than in the confidence test case. The square root versions of this set of *f*-divergences have already been used previously in statistics. Thus,

Definition 4 Three refinement criteria for adaptive ray-tracing, based on their respective f-divergences, are given by

- Square root of Kullback-Leibler divergence

$$\frac{1}{N_s^p} \overline{L} D_{KL}^{\frac{1}{2}}(p,q) < \varepsilon \qquad (\text{KL}^{\frac{1}{2}})$$

- Square root of chi-square divergence

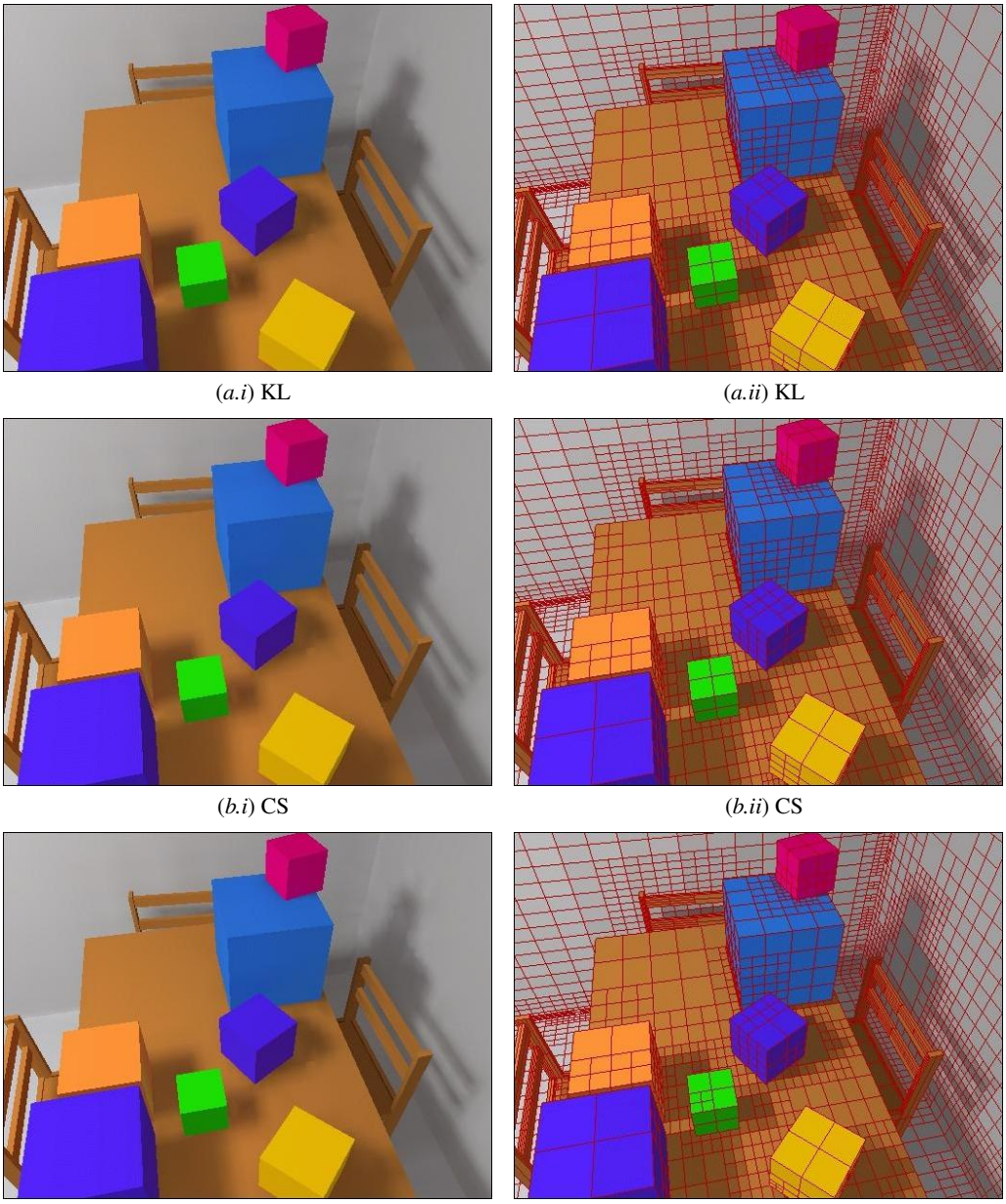
$$\frac{1}{N_s^p} \overline{L} D_{\chi^2}^{\frac{1}{2}}(p,q) < \varepsilon \qquad (\text{CS}^{\frac{1}{2}})$$

- Square root of Hellinger divergence

$$\frac{1}{N_s^p} \overline{L} D_{h^2}^{\frac{1}{2}}(p,q) < \varepsilon \qquad (\text{HL}^{\frac{1}{2}})$$

5.2. Results

In Fig. 8 and Fig. 9 we present comparative results with different techniques for the test scene in Fig. 7. The following two methods are compared with the three *f*-divergence-based criteria ($\text{KL}^{\frac{1}{2}}$, $\text{CS}^{\frac{1}{2}}$, and $\text{HL}^{\frac{1}{2}}$ of Def. 4):

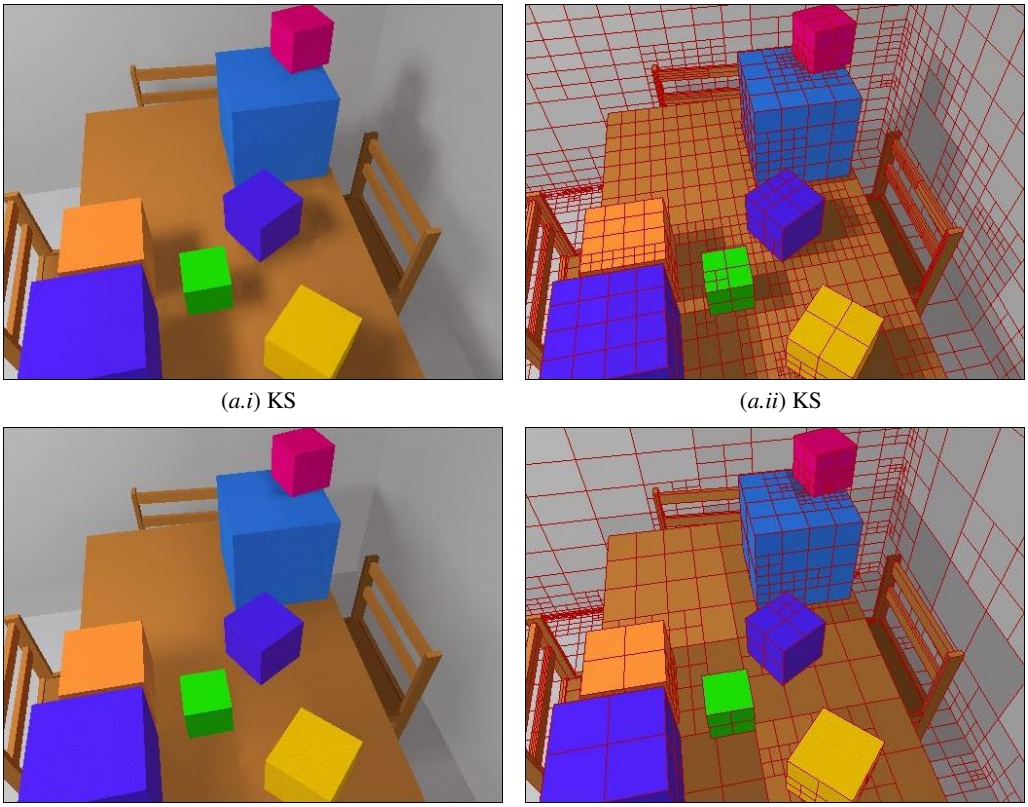


(*a.i*) KL (*a.ii*) KL

(*b.i*) CS (*b.ii*) CS

(*c.i*) HL (*c.ii*) HL

Figure 5: *The view$_2$ of the scene for comparison of* f*-divergence-based oracles, (a) Kullback-Leibler (KL), (b) chi-square (CS), and (c) Hellinger (HL), versus KS and VT ones (Fig. 6). By columns, (i) Gouraud shaded solution and (ii) its final adaptive mesh. The oracles have been evaluated with 10 random lines between elements.*



(*a.i*) KS (*a.ii*) KS

(*b.i*) VT (*b.ii*) VT

Figure 6: *The view$_2$ of the scene for comparison of (a) kernel-smoothness-based (KS) and (b) variance-based (VT) oracles versus* f*-divergence-based ones (Fig. 5). By columns, (i) Gouraud shaded solution and (ii) its final adaptive mesh. The oracles have been evaluated with 10 random lines between elements.*

- CC: Classic contrast (see Part IV) of the luminance weighted with the respective importance $\overline{L}$.
- CT: Confidence test with a confidence level of $\alpha = 0.1$ and a tolerance $t = 0.025$ (see Part IV).

In order to evaluate their behaviour, the images are generated by a similar process to that of adaptive sampling ray-tracing in Part IV. Clearly, all the methods are directly applicable to adaptive sampling schemes such as that presented in Part IV. In all the methods, 8 initial rays are cast in a stratified way at each pixel to compute the contrast measures for the refinement decision, and 8 additional rays are successively added until the condition of the criterion is met. An implementation of classic path-tracing with next event estimator was used to compute all images. The parameters were tuned so that all five test images were obtained with a similar average number of rays per pixel ($N_s^p = 60$) and a similar computational cost. The reconstruction method applied is the piecewise-continuous image with box filter. Finally, the pixel value is the reconstructed signal average at pixel domain.

The resulting images are shown in Figs. 8.∗.*i* (CC and CT) and Figs. 9.∗.*i* ($KL^{\frac{1}{2}}$, $CS^{\frac{1}{2}}$, and $HL^{\frac{1}{2}}$), with the sampling density maps in Figs. 8.∗.*ii* and Figs. 9.∗.*ii*, respectively. The analysis of the critical points of the images shows how our sampling scheme performs the best. Observe, for instance, the reduced noise in the shadows cast by the objects. Observe also the detail of the shadow of the sphere reflected on the pyramid.

Comparison of the SDMs shows a better discrimination of complex regions of the scene in the three divergence cases against the classic contrast and confidence test cases. This explains the better results obtained by our approach. On the other hand, the confidence test approach also performs better than the classic contrast-based method. Its SDM also explains why it performs better than the contrast-based. However, it is unable to suitably render the reflected shadows under the mirrored pyramid and sphere with precision.

In Table 3, we show the RMSE and PSNR of the images obtained with classic (Figs. 8.∗.*i*), *f*-divergence, and square root of *f*-divergence (Figs. 9.∗.*i*) methods respective to the

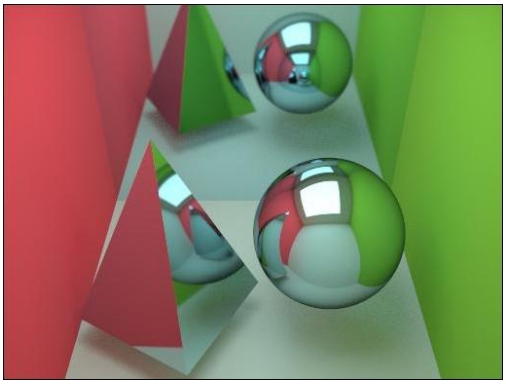
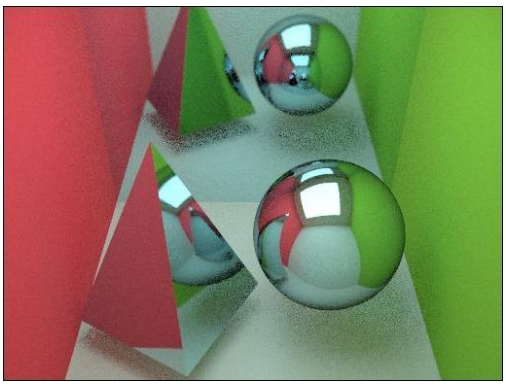
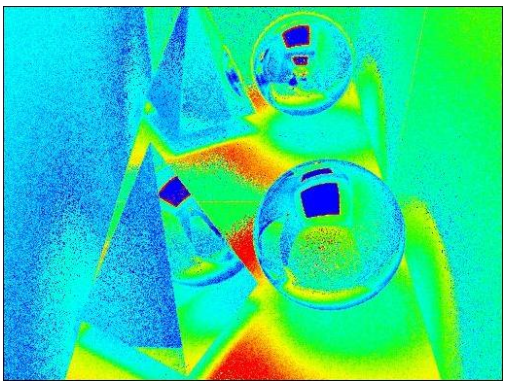


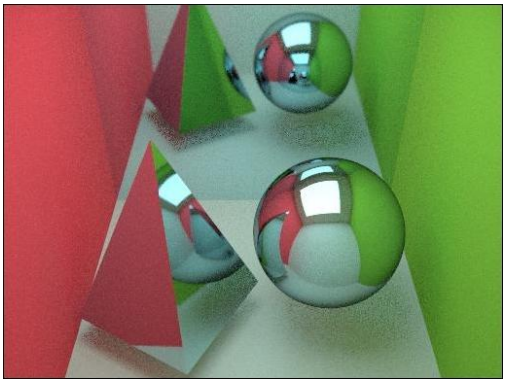
Figure 7: *Test scene for the ray-tracing comparison in Fig. 8 and Fig. 9, obtained with a path-tracing algorithm with 1,024 samples per pixel in a stratified way.*



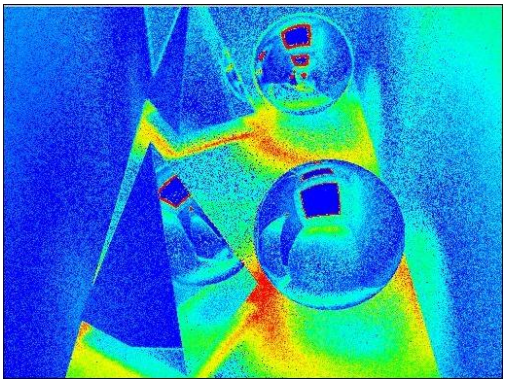
(*a.i*) CC



(*a.ii*) SDM of (*a.i*)



(*b.i*) CT



(*b.ii*) SDM of (*b.i*)

Figure 8: *Images of the test scene (Fig. 7) obtained with an adaptive sampling scheme based on (a) classic contrast (CC) and (b) confidence test (CT) methods. By columns, (i) shows the resulting images and (ii) the sampling density maps of (i). The average number of rays per pixel is 60 in all the methods.Compare with the images in Fig. 9.*

test image in Fig. 7. Visual comparison is in concordance with numerical data. The f-divergence-based criteria used in our experiments ($KL^{\frac{1}{2}}$, $CS^{\frac{1}{2}}$, and $HL^{\frac{1}{2}}$) outperform both classic contrast and confidence test experiments. Finally, the better results of the $HL^{\frac{1}{2}}$ criterion could be explained by the fact that it is a true distance.

References

[AS66] ALI M. S., SILVEY S. D.: A general class of coefficient of divergence of one distribution from another. *Journal of Royal Statistical Society (Serie B) 28*, 1 (1966), 131–142.

[BB97] BREN M., BATAGELJ V.: *The Metric Index*. Tech. Rep. 35-561, Institute of mathematics, physics, and mechanics, Department of mathematics, University of Ljubljana, Ljubljana, Slovenia, 1997.

[BCDS02] BARNETT N. S., CERONE P., DRAGOMIR S. S., SOFO A.: Approximating Csiszár f-divergence by the use of Taylor's formula with integral remainder. *Mathematical Inequalities and Applications 5*, 3 (2002), 417–434.

[BOR99] BERNAOLA P., OLIVER J. L., ROMÁN R.: Decomposition of DNA sequence complexity. *Physical Review Letters 83*, 16 (October 1999), 3336–3339.

[Bur83] BURBEA J.: J-divergences and related concepts. In *Encyclopedia of Statistical Sciences*, Kotz S., Johnson N. L., Read C. B., (Eds.), vol. 4. Wiley Interscience, New York (NY), USA, 1983, pp. 290–296.

[Com98] COMMISSION I. E.: Default RGB colour space - sRGB. In *Colour Management in Multimedia Systems*. IEC, Geneva, Switzerland, 1998, ch. Part 2.1. IEC/PT61966(PL)34.

[Csi63] CSISZÁR I.: Eine Informationsheoretische Ungleichung und ihre Anwendungen auf den Beweis der Ergodizität von Markoffschen Ketten (An information-theoretic inequality and its application to the proof of the ergodicity of Markov chains). *Magyar Tudományos Akadémia Közleményei 8* (1963), 85–108.

[Csi67] CSISZÁR I.: Information-type measure of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica 2* (1967), 299–318.

[CT91] COVER T. M., THOMAS J. A.: *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.

[Dra00] DRAGOMIR S. S.: Some inequalities for the Csiszár f-divergence. In *Inequalities for Csiszár f-Divergence in Information Theory*. Research Group in Mathematical Inequalities and Applications, Melbourne, Australia, 2000. http://rgmia.vu.edu.au/monographs/csiszar_list.html.

[Fei02] FEIXAS M.: *An Information-Theory Framework for the Study of the Complexity of Visibility and Radiosity in a Scene*. PhD thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, Desember 2002.

[FRBS02] FEIXAS M., RIGAU J., BEKAERT P., SBERT M.: Information-theoretic oracle based on kernel smoothness for hierarchical radiosity. In *Short Presentations (Eurographics '02)* (September 2002), pp. 325–333.

[Gla95] GLASSNER A. S.: *Principles of Digital Image Synthesis*. Morgan Kaufmann Publishers, San Francisco (CA), USA, 1995.

[Gow85] GOWER J. C.: Measures of similarity, dissimilarity, and distance. In *Encyclopedia of Statistical Sciences*, Kotz S., Johnson N. L., Read C. B., (Eds.), vol. 5. Wiley Interscience, New York (NY), USA, 1985, pp. 397–405.

[GS01] GIBBS A. L., SU F. E.: On choosing and bounding probability metrics. *International Statistical Review 70* (2001), 419–435.

[Hel09] HELLINGER E. D.: Neue Begründung der Theorie der Quadratischen Formen von Unendlichen Vielen Veränderlichen (A new foundation of the theory of quadratic forms of infinite many variables). *Journal für Reine und Angewandte Mathematik 136* (1909), 210–271.

[HHK03] HE Y., HAMZA A., KRIM H.: A generalized divergence measure for robust image registration. *IEEE Transactions on Signal Processing 51*, 5 (May 2003), 1211–1220.

[HMMG01] HERO A. O., MA B., MICHEL O., GORMAN J.: *Alpha-Divergence for Classification, Indexing and Retrieval*. Tech. Rep. CSPL-328, Communications and Signal Processing Laboratory, Ann Arbor (MI), USA, May 2001.

[Kak48] KAKUTANI S.: On equivalence of infinite product measures. *Annals of Mathematics 49* (1948), 214–224.

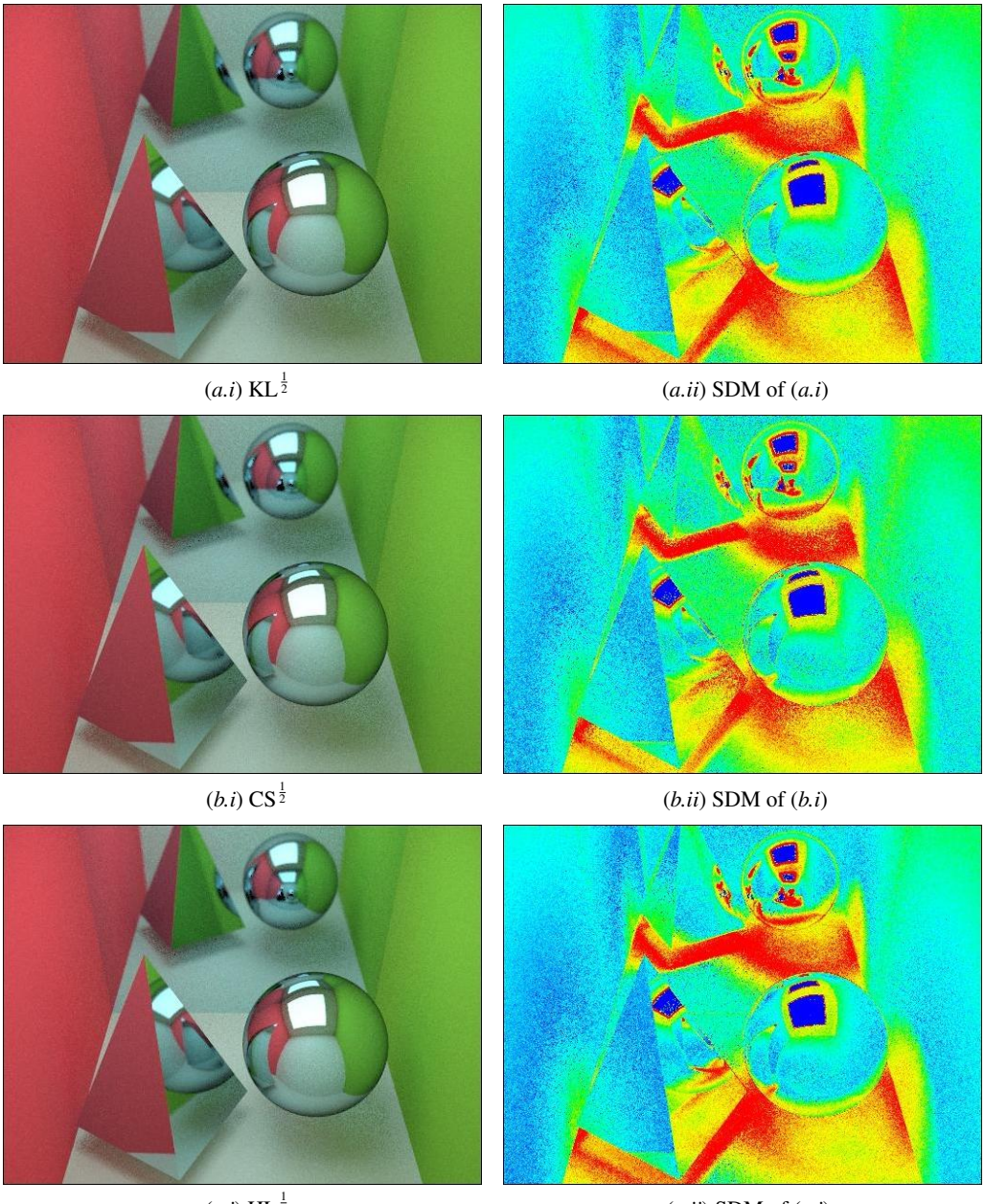
[KL51] KULLBACK S., LEIBLER R. A.: On information and sufficiency. *Annals of Mathematical Statistics 22* (1951), 76–86.

[LeG99] LEGLAND F.: Stability and approximation of nonlinear filters: an information theoretic approach. In *Proceedings of the 38th IEEE Conference on Decision and Control* (December 1999), pp. 1889–1894.

[LV87] LIESE F., VAJDA I.: *Convex Statistical Distances*. Teubner Verlagsgesellschaft, Leipzig, Germany, 1987.

[LY90] LE CAMP L. M., YANG G. L.: *Asymptotics in Statistics: Some Basic Concepts*. Springer-Verlag, New York (NY), USA, 1990.

[Mar01] MARTÍN J. A.: *Medidas de Diferencia y Clasificación no Paramétrica de Datos Composicionales (Measures of difference and non-parametric clasification of*



(*a.i*) $\mathrm{KL}^{\frac{1}{2}}$ (*a.ii*) SDM of (*a.i*)

(*b.i*) $\mathrm{CS}^{\frac{1}{2}}$ (*b.ii*) SDM of (*b.i*)

(*c.i*) $\mathrm{HL}^{\frac{1}{2}}$ (*c.ii*) SDM of (*c.i*)

Figure 9: *Images of the test scene (Fig. 7) obtained with an adaptive sampling scheme based on square root of (a) Kullback-Leibler ($KL^{\frac{1}{2}}$), (b) chi-square ($CS^{\frac{1}{2}}$), and (c) Hellinger ($HL^{\frac{1}{2}}$) f-divergences. By columns, (i) shows the resulting images and (ii) the sampling density maps of (i). The average number of samples per pixel is 60 in all the methods. Compare with the images in Fig. 8.*

method	*criterion*	$RMSE_a$	$RMSE_p$	$PSNR_a$	$PSNR_p$
Classic	Contrast (CC)	6.157	6.126	32.344	32.387
	Confidence test (CT)	5.194	5.174	33.822	33.855
f-divergences	Kullback-Leibler (KL)	5.508	5.553	33.311	33.241
	chi-square (CS)	5.414	5.452	33.461	33.400
	Hellinger (HL)	5.807	5.862	32.852	32.770
Square root of f-divergences	Kullback-Leibler ($KL^{\frac{1}{2}}$)	4.824	4.793	34.463	34.519
	chi-square ($CS^{\frac{1}{2}}$)	4.772	4.736	34.557	34.623
	Hellinger ($HL^{\frac{1}{2}}$)	4.595	4.560	34.884	34.951

Table 3: *The RMSE and PSNR measures of the CC, CT, and* f*-divergence-based refinement criteria applied to Fig. 7. The images for the CC and CT methods are shown in Fig. 8, and for the* f*-divergence-based ones, in Fig. 9. The average number of samples per pixel is 60 in all the methods.*

compositional data). PhD thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, 2001.

[Mer05] MERRIAM-WEBSTER, INC.: *Merriam-Webster Online Dictionary*. Springfield (MA), USA, 2005. http://www.m-w.com © 2005 by Merriam-Webster, Inc.

[Ö02a] ÖSTERREICHER F.: Csiszár's f-divergences - Basic properties. In *Workshop of the Research Group in Mathematical Inequalities and Applications* (Melbourne, Australia, November 2002).

[Ö02b] ÖSTERREICHER F.: f-Divergences - Representation theorem and metrizability. In *Workshop of the Research Group in Mathematical Inequalities and Applications* (Melbourne, Australia, November 2002).

[OBS98] O'SULLIVAN J. A., BLAHUT R. E., SNYDER D. L.: Information-theoretic image formation. *IEEE Transactions on Information Theory 44*, 6 (October 1998), 2094–2123.

[OV93] ÖSTERREICHER F., VAJDA I.: Statistical information and discrimination. *IEEE Transactions on Information Theory 39*, 3 (May 1993), 1036–1039.

[Pea00] PEARSON K.: On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine V*, 1 (1900), 157–175.

[Plu01] PLUIM J. P. W.: *Mutual Information Based Registration of Medical Images*. PhD thesis, Utrecht University, Utrecht, Holland, 2001.

[PV97] PARDO M., VAJDA I.: About distances of discrete distributions satisfying the data processing theorem of information theory. *IEEE Transactions on Information Theory 43*, 7 (1997), 1288–1293.

[PV03] PARDO M., VAJDA I.: On asymptotic properties of information-theoretic divergences. *IEEE Transactions on Information Theory 49*, 7 (2003), 1860–1868.

[Rei89] REISS R.-D.: *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics*. Springer-Verlag, New York (NY), USA, 1989.

[Top00] TOPSØE F.: Some inequalities for information divergence and related measures of discrimination. *IEEE Transactions on Information Theory 46*, 4 (2000), 1602–1609.

[Vaj89] VAJDA I.: *Theory of Statistical Inference and Information*. Kluwer Academic Publishers, Boston (MA), USA, 1989.

[YB99] YANG Y., BARRON A.: Information theoretic determination of minimax rates of convergence. *Annals of Statistics 27* (1999), 1546–1599.

Applications of Information Theory to Computer Graphics Part IV: Adaptive Refinement for Ray-tracing

Jaume Rigau, Miquel Feixas, and Mateu Sbert

University of Girona, Spain

1. Introduction

Although ray-tracing is a straightforward and powerful image synthesis technique, it usually requires many rays per pixel to eliminate the aliasing or noise in the final image. However, not all the pixels in the image require the same number of rays. The edge of an object, the contour of a shadow, and a high illumination gradient will require a much better treatment than a region with almost uniform illumination. To this effect, many pixel supersampling refinement criteria have been defined in the literature.

The measures used in these criteria are based on intensities (image space) and also on geometry (object space). They are also useful for an adaptive subdivision of image space for progressive refinement [PS89]. Some of them have been applied in the image based rendering field for weighting pixel colour for reconstruction [PCD*97] and adaptive sampling strategies [DC96, DCV97], and creating a priority scheme for sampling in interactive rendering [SS00]. The final objective is always to find the best final-image quality with a reasonable cost. In order to do this, we have to sample each pixel of the image plane carefully. It is essential to have a quantitative measure in order to evaluate when there is sufficient information about the pixel.

The data of a sample set through the pixel can be used to calculate a pixel homogeneity measure from two different points of view: *radiance* and *visibility*. The information which we will manipulate will be exclusively colour (radiance) and geometry (visibility), essential parameters for deciding on the "quality" of a pixel. In this context, the Shannon entropy will be interpreted as a measure of the degree of homogeneity of a pixel in the sense that the more heterogeneous the pixel, the more difficult it is to obtain its actual value. From it, we associate homogeneity with quality, so that the need for pixel refinement is proportional to the lack of quality (i.e., heterogeneity of the samples). The idea behind the new scheme is to obtain sufficient information in the refinement algorithm in order to find out the sampling needs.

Consequently, in this part we present a framework for entropy-based sampling applied to ray-tracing methods. First, definitions of new measures of *pixel quality* based on entropy are presented (§3). Next, we present the pixel quality as a measure of *pixel contrast* (§4). Then, this contrast is applied to classic *supersampling* ray-tracing (§5) and *adaptive sampling* (§6). This framework is easily adaptable to other stochastic processes which require measures of quality in order to reach decisions.

2. Previous Work

Three principal subproblems make up the process of obtaining a good quality image: efficient sample generation, adaptive control of the sampling rate, and filtering for image reconstruction [PS89]. Many approaches are to be found to deal with them:

1. Different pixel sampling methods have been introduced, among them: jittered sampling [CPC84, DW85], Poisson disk sampling [DW85, Mit87, MF92], hierarchical sampling [Kaj86], complete stratification at each refinement level [Sch91], importance sampling [Shi90], and quasi-Monte Carlo sampling [KH94, OA96].
2. Diverse refinement criteria for adaptive sampling, based on colour intensities and/or scene geometry, can be found to control the sampling rate: Dippé and Wold [DW85] present an error estimator based on the RMS signal to noise ratio and also consider its variance as a function of the number of samples; Mitchell [Mit87] proposes a contrast [Cae81] based on the characteristics of the human eye; Lee et al. [LRU85], Purgathofer [Pur86], and Tamstorf and Jensen [TJ97] develop different methods based on the variance of the samples with their respective confidence intervals.
3. Samples are filtered to produce the final pixel values. Different filter shapes have been used in image reconstruc-

tion: box filter, triangular filter, Gaussian filter, multi-stage filter, etc. (see [Gla95]).

We review here three commonly used refinement criteria: contrast, depth difference, and variance of the samples.

Mitchell, in [Mit87], uses a contrast measure [Cae81] for each RGB channel defined by

$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \tag{1}$$

where $I_{\min}$ and $I_{\max}$ are, respectively, the minimum and maximum light intensities of the channel. Supersampling is done if any contrast is higher than a given threshold. Mitchell proposes RGB threshold values (0.4, 0.3 and 0.6, respectively) based on the relative sensitivity of the visual system.

In [SS00], within an interactive rendering context, Simmons uses a *priority* value p_c based on the above concepts (contrast and perception) [Mit87, Gla95] defined by

$$p_c = .4\frac{r_{\max} - r_{\min}}{r_{\max} + r_{\min}}\overline{r} + .3\frac{g_{\max} - g_{\min}}{g_{\max} + g_{\min}}\overline{g} + .6\frac{b_{\max} - b_{\min}}{b_{\max} + b_{\min}}\overline{b} \tag{2}$$

where max, min, and the overline represent, respectively, the maximum, minimum, and average values for r, g, and b colour channels.

On the other hand, a useful and simple geometric measure for refinement is *depth difference*, used recently in image based rendering [DC96, DCV97, PCD*97] and interactive rendering [SS00]. Depth difference is given by

$$p_d = 1 - \frac{d_{\min}}{d_{\max}} \tag{3}$$

where $d_{\max}$ and $d_{\min}$ represent maximum and minimum distance. In [SS00], p_c and p_d measures are combined in

$$p_v = \delta p_c + (1-\delta)p_d \tag{4}$$

with $\delta = 0.9$.

The basic idea of variance-based methods [LRU85, Pur86, TJ97] is to continue sampling until the confidence level or probability that the true value L is within a given tolerance d of the estimate value $\widehat{L}$ is $1-\alpha$:

$$\Pr[L \in (\widehat{L}-d, \widehat{L}+d)] = 1-\alpha. \tag{5}$$

Mitchell considers that variance is a poor measure of visual perception of local variation [Mit87]. Kirk and Arvo showed that these methods are biased and proposed a simple correction scheme [KA91].

Refinement criteria have also been applied in the image-based rendering field to weight pixel colour for reconstruction purposes [PCD*97] and adaptive sampling strategies [DC96, DCV97]. Also Bolin and Meyer [BM98] have developed a perceptually-based approach using statistical and vision models.

3. Pixel Quality

In this section we introduce a new pixel quality measure, the *pixel entropy*. This measure will be defined from the information provided by set of samples on the image plane. We use the following sets:

- Let P be the set of pixels of the image plane with $|\mathsf{P}| = N_{\mathsf{p}} > 0$.
- Let S_{p} be the set of samples of a pixel $\mathsf{p} \in \mathsf{P}$ with $|S_{\mathsf{p}}| = N_s^{\mathsf{p}} > 1$.
- Let S_{P} be the set of samples of the image plane where $S_{\mathsf{P}} = \cup_{\mathsf{p}\in\mathsf{P}} S_{\mathsf{p}}$ with $|S_{\mathsf{P}}| = N_s^{\mathsf{P}} = \sum_{\mathsf{p}\in\mathsf{P}} N_s^{\mathsf{p}}$.

The implementation of a sample consists in casting a ray r_v^{Θ} from a scene viewpoint v through the image plane and, in particular, through a pixel: $\Theta \in \Omega_{v\to\mathsf{P}}$. Let us consider that each sample $s \in S_{\mathsf{P}}$ that hits a scene surface gives us information about the colour, distance and orientation of the hit point with respect to the viewpoint.

The definition of entropy

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i, \tag{6}$$

measures the expectation of the surprise of the distribution p and it can be considered also to be a measure of its homogeneity. From the sample set and from the entropy, two different quality measures are defined, *pixel colour entropy* and *pixel geometry entropy*, based on the colour and geometry respectively.

3.1. Pixel Colour Entropy

Our first objective is to define the pixel colour entropy. We start with a global definition of entropy concerning all the samples passing through the image plane. We consider that the colour belongs to a colour system $\mathbf{c}$ structured in components called *colour channels*. Without loss of generality, in the majority of cases our colour measures will refer to a single channel $c \in \mathbf{c}$, $c(s)$ being the colour channel data of a sample $s \in S$ (e.g., radiance, luminance, and RGB values).

Let us consider the probability of each image plane sample as its colour channel contribution relative to the whole of the image plane sample set. Thus,

Definition 1 The image plane channel entropy of a channel c is given by

$$H^c(\mathsf{P}) = -\sum_{i=1}^{N_s^{\mathsf{P}}} r_i \log r_i \qquad r_i = \frac{c(s_i)}{\sum_{j=1}^{N_s^{\mathsf{P}}} c(s_j)}, \tag{7}$$

where r_i represents the channel colour fraction of sample $s_i \in S_{\mathsf{P}}$.

This measure can be interpreted as the colour channel homogeneity of the samples passing through the image plane. Analogously, at the pixel level, we consider the probability of each pixel sample as its colour channel contribution relative to the whole of the pixel sample set. Then,

Definition 2 The pixel channel entropy of a channel c is given by

$$H^c(\mathsf{p}) = -\sum_{i=1}^{N_s^{\mathsf{p}}} p_i \log p_i \qquad p_i = \frac{c(s_i)}{\sum_{j=1}^{N_s^{\mathsf{p}}} c(s_j)}, \tag{8}$$

where p_i represents the channel colour fraction of sample $s_i \in S_{\mathsf{p}}$.

From the properties of the entropy, the image plane channel entropy ranges from 0 to $\log N_s^{\mathsf{P}}$ and the pixel channel from 0 to $\log N_s^{\mathsf{p}}$. The maximum values are obtained when the channel colour of all the samples is the same (i.e., we have an uniform probability distribution). Using the grouping property of entropy (see Part II) it is easy to see that image plane and pixel channel entropies can be related in the following way:

$$H^c(\mathsf{P}) = \sum_{i=1}^{N_{\mathsf{p}}} q_i H^c(\mathsf{p}_i) - \sum_{i=1}^{N_{\mathsf{p}}} q_i \log q_i = \sum_{i=1}^{N_{\mathsf{p}}} q_i H^c(\mathsf{p}_i) + H_{\mathrm{I}}^c(\mathsf{P}), \tag{9}$$

where $q_i = \sum_{j=1}^{N_s^{\mathsf{p}i}} r_j$ is the *importance* (sum of probabilities) of pixel p_i, $H^c(\mathsf{p}_i)$ is the channel entropy of pixel p_i, and $H_{\mathrm{I}}^c(\mathsf{P}) = -\sum_{i=1}^{N_{\mathsf{p}}} q_i \log q_i$ is the *importance entropy* of the image plane calculated from the importance of each pixel. Thus, the global entropy of the image plane is the sum of all the pixel entropies, weighted by the importance of each pixel, and the importance entropy obtained from the importance of each pixel.

The image plane and pixel entropies can be interpreted as the colour *homogeneity* or *uniformity* measured by its sample set and thus can be considered measures of the *quality* of the colour channel (i.e., lack of heterogeneity and noise). We can also observe that the entropy increases with the number of samples. In order to give a pixel quality measure between 0 and 1, the pixel channel entropy can be normalised with $\log N_s^{\mathsf{p}}$. Thus,

Definition 3 The pixel channel quality of a channel c is given by

$$Q^c(\mathsf{p}) = \frac{H^c(\mathsf{p})}{\log N_s^{\mathsf{p}}}. \tag{10}$$

If we want to consider the global quality of a pixel, we need only mix its set of channels. Then,

Definition 4 The pixel colour quality of a colour system $\mathbf{c}$ is given by the weighting of its pixel channel qualities:

$$\mathbf{Q^c}(\mathsf{p}) = \frac{\sum_{c\in\mathbf{c}} w^c Q^c(\mathsf{p})}{\sum_{c\in\mathbf{c}} w^c}, \tag{11}$$

where w^c is the weight of channel c.

The weighted values depend on each colour system. Without a priori information, the same weight per channel can be considered, otherwise a weight based on human perception (for an sRGB system, $w^{\mathrm{R}} = 0.2126$, $w^{\mathrm{G}} = 0.7152$, and $w^{\mathrm{B}} = 0.0722$). This measure will enable us to define a new colour contrast measure for pixel sampling (§4.1). Note that the larger the number of samples the more accurate the quality measure.

In Fig. 1.*b*, we present a *colour quality map* to show the colour quality of all the pixels from Fig. 1.*a* using an sRGB colour system with the same weight by channel. With respect to the colour scale used, the minimum quality corresponds to the blue and the maximum to the red (in order to observe more details in the colour quality maps, the *outliers* are reduced to the borders of the interval $[-k\sigma, k\sigma]$ where σ is the standard deviation of the results and k is a parameter that modulates the width of the interval). A low quality in shadow areas and edges can be observed.

3.2. Pixel Geometry Entropy

Similar concepts introduced in the above section can be defined in this one with respect to a geometric measure. If $x = \Lambda(v,\Theta)$ is the hitpoint of a sample ray $s = r_v^{\Theta}$, the geometric information of each sample is given by $\theta_{n_x}^{-\Theta}$ (i.e., the angle of the normal at the hit point) and by the distance r_{vx} between this point and the origin of the ray (i.e., ray length). We take

$$g(s) = \frac{\cos\theta_{n_x}^{-\Theta}}{r_{vx}^2} \tag{12}$$

as a geometry factor of a sample. This value provides a quality measure of visibility of a scene point from the observer's point of view.

Let us define the probability of each image plane sample as its relative geometric contribution to the whole image plane sample set.

Definition 5 The image plane geometry entropy is given by

$$H^g(\mathsf{P}) = -\sum_{i=1}^{N_s^{\mathsf{P}}} r_i \log r_i \qquad r_i = \frac{g(s_i)}{\sum_{j=1}^{N_s^{\mathsf{P}}} g(s_j)}, \tag{13}$$

where r_i represents the geometry fraction of sample $s_i \in S_{\mathsf{P}}$.

Considering the probability of each pixel sample as its relative geometrical contribution to the whole of the pixel sample set we have

Definition 6 The pixel geometry entropy is given by

$$H^g(\mathsf{p}) = -\sum_{i=1}^{N_s^{\mathsf{p}}} p_i \log p_i \qquad p_i = \frac{g(s_i)}{\sum_{j=1}^{N_s^{\mathsf{p}}} g(s_j)}, \tag{14}$$

where p_i represents the geometry fraction of sample $s_i \in S_{\mathsf{p}}$.

Analogously to the pixel colour entropy (9), an identical relation can be established between the geometric entropies of the image plane and the pixel. We can also normalise the pixel geometry entropy and therefore,

Definition 7 The pixel geometry quality is given by

$$Q^g(\mathsf{p}) = \frac{H^g(\mathsf{p})}{\log N_s^{\mathsf{p}}}. \tag{15}$$

In Fig. 1.*c* we show the *geometry quality map* from Fig. 1.*a* based on a grey scale. The lowest entropy corresponds to the darkest part, the highest entropy to the lightest (the outliers have the same treatment as in the colour quality maps). Observe that the edges have a very low entropy and are very clearly emphasised.

4. Pixel Contrast

In this section we present new pixel contrast measures based on pixel entropy (§3). As the entropy represents the homogeneity of the information brought back by the samples (i.e., rays crossing a pixel), we can define a simple measure which expresses the diversity or contrast of a pixel.

4.1. Pixel Colour Contrast

In the colour theory, the *colour contrast* is the phenomenon that alters the observation of the colours depending on their surroundings (the origin of the colour contrast is the way in which the information is transmitted from the retinal photoreceptors to the brain and the name of this study is the *colour opponency* theory [Gla95]). We use the same words to express the degree of heterogeneity of the colour in the region defined by a pixel given that this value depends directly on the colours that are around it. As we have seen, $H^c(\mathsf{p})$ represents the entropy or the degree of colour homogeneity of pixel p. From this measure,

Definition 8 The pixel channel contrast of a channel c is given by

$$C^c(\mathsf{p}) = 1 - Q^c(\mathsf{p}) = 1 - \frac{H^c(\mathsf{p})}{\log N_s^{\mathsf{p}}}. \tag{16}$$

It represents the colour channel heterogeneity or contrast of a pixel with a range of $[0,1]$. We can also introduce the pixel binary contrast from minimum and maximum colour channel probabilities captured by this pixel. This measure is obtained from the binary entropy of these values. Thus,

Definition 9 The pixel channel binary contrast of a channel c is given by

$$C_b^c(\mathsf{p}) = 1 - H_b^c(\mathsf{p}) H_b^c(\mathsf{p}) = H(\{\frac{p_{\min}}{p_{\min}+p_{\max}}, \frac{p_{\max}}{p_{\min}+p_{\max}}\}), \tag{17}$$

where $H_b^c(\mathsf{p})$ is the binary entropy of the minimum and maximum channel colour probabilities, $p_{\min}$ and $p_{\max}$, respectively.

Both measures, $H_{\mathrm{b}}^c(\mathsf{p})$ and $C_{\mathrm{b}}^c(\mathsf{p})$, range also between 0 and 1 due to the fact that, in this case, only two values are taken into account. As we will see in our experiments, this binary measure yields more radical contrast than $C^c(\mathsf{p})$.

Similarly to previous works [Mit87, Gla95, SS00], we can obtain the colour contrast of a pixel by averaging all the colour channel contrasts weighted by their respective importances (colour channel average). This avoids oversampling on the areas with small colour values. Then, considering all the colour channels,

Definition 10 The pixel colour contrast of a colour system $\mathbf{c}$ is given by the weighting of its pixel channel contrasts:

$$\mathbf{C}^{\mathbf{c}}(\mathsf{p}) = \frac{\sum_{c\in\mathbf{c}} w^c \overline{c} C^c(\mathsf{p})}{\sum_{c\in\mathbf{c}} w^c \overline{c}} \qquad \overline{c} = \frac{1}{N_s^{\mathsf{p}}} \sum_{i=1}^{N_s^{\mathsf{p}}} c(s_i), \tag{18}$$

where the channel contrasts are weighted by perceptual coefficients w^c, and $\overline{c}$ is the colour average in channel c of all $s \in S_{\mathsf{p}}$.

Definition 11 The pixel colour binary contrast of a colour system $\mathbf{c}$ is given by

$$\mathbf{C}_b^{\mathbf{c}}(\mathsf{p}) = \frac{\sum_{c\in\mathbf{c}} w^c \overline{c} C_b^c(\mathsf{p})}{\sum_{c\in\mathbf{c}} w^c \overline{c}}. \tag{19}$$

In an sRGB system, the colour contrast measures ($\mathbf{C}^{\mathbf{sRGB}}$ and $\mathbf{C}_{\mathrm{b}}^{\mathbf{sRGB}}$) have three channels with coefficients w^{R}, w^{G}, and w^{B}. These values depend on the specific use of contrast, but in general they can take the values proposed in the pixel colour quality (11), or also, for a perceptual balance of the channels, they can take those of the thresholds proposed in [Mit87, SS00]: 0.4, 0.3, and 0.6, respectively.

This last option is chosen in the next examples where a *colour contrast map* is used as a visual representation of the contrast measures in the same way that the quality map is used for the quality measures. Thus, in Fig. 2 we show different colour contrast maps to compare the heuristic (2), p_c (Fig. 2.*b*), with measures $\mathbf{C}^{\mathbf{c}}$ (Fig. 2.*c*), and $\mathbf{C}_{\mathrm{b}}^{\mathbf{c}}$ (Fig. 2.*d*). Another comparison is shown in Fig. 3 with a more complex scene. We can observe how these measures present a very good behaviour in critical areas (represented by warm colours) such as object edges and shadow contours. With respect to Fig. 2.*b* and Fig. 3.*b*, our measures are more discriminating, especially the binary contrast.

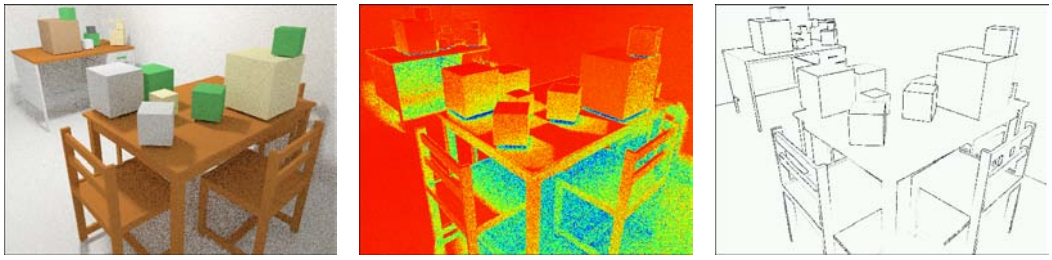
4.2. Pixel Geometry Contrast

As we have seen in (14), H^{g} represents the entropy or the degree of geometric homogeneity of a pixel. From this measure,

Definition 12 The pixel geometry contrast is given by

$$C^g(\mathsf{p}) = 1 - Q^g(\mathsf{p}) = 1 - \frac{H^g(\mathsf{p})}{\log N_s^{\mathsf{p}}}. \tag{20}$$

Similarly to the above section, we introduce the pixel binary contrast from minimum and maximum geometry factor probabilities of this pixel. Thus,



(*a*) Reference (*b*) $\mathbf{Q^c}$ (*c*) Q^{g}

Figure 1: *Colour and geometry quality maps. (a) Reference image obtained with* $N_s^{\mathrm{p}} = 8$. *(b) Pixel colour quality* $\mathbf{Q^{sRGB}}$ *with the same weight per channel. (c) Pixel geometry quality* Q^g.

Definition 13 The pixel geometry binary contrast is given by

$$C_b^g(\mathsf{p}) = 1 - H_b^g(\mathsf{p}) H_b^g(\mathsf{p}) = H(\{\frac{p_{\min}}{p_{\min}+p_{\max}}, \frac{p_{\max}}{p_{\min}+p_{\max}}\}), \tag{21}$$

where $H_b^g(\mathsf{p})$ is the binary entropy of the minimum and maximum geometry factor probabilities, $p_{\min}$ and $p_{\max}$ respectively.

A third case can also be considered:

Definition 14 The pixel logarithmic-difference contrast is given by

$$C_{\log}^g(\mathsf{p}) = \log p_{\max} - \log p_{\min} = \log \frac{p_{\max}}{p_{\min}}. \tag{22}$$

This measure, introduced in Rigau et al. [RFBS01], is based on the gradient between the minimum and maximum complexity segments. As we will see, $C_{\log}^{\mathrm{g}}(\mathsf{p})$ also shows a good behaviour.

For the previous reference scene in Fig. 2.*a*, we now show the geometry contrast measures using the corresponding maps in Figs. 4.*a–c*. These maps are compared with the map in Fig. 4.*d*, created using the depth difference heuristic p_d (3). The same comparison is carried out for the reference scene Fig. 3.*a* in Fig. 5. It can be seen that our measures capture the majority of edges because we take into account two components: distance and orientation. These geometry contrast maps have been generated by using the representation scale of the colour contrast maps in order to be able to compare, visually with each other, how the two types of contrast, colour (Fig. 3) and geometry (Fig. 5), work. The specialisation of each of the contrasts is evident: colour maps show the heterogeneity of regions while geometric maps identify edges.

4.3. Pixel Colour-Geometry Contrast

Finally, a combination of colour and geometry contrasts is considered. This combination enables us to graduate, with a coefficient δ between 0 and 1, the influence of both measures. Then,

Definition 15 The pixel contrast of colour system $\mathbf{c}$ is given by

$$\mathcal{C}^{\mathbf{c}}(\mathsf{p}) = \delta \mathbf{C}^{\mathbf{c}}(\mathsf{p}) + (1-\delta) C^g(\mathsf{p}). \tag{23}$$

This combination can be made with any type of pixel colour contrast and geometry contrast. In general, good behaviour has been shown with binary contrasts (colour and geometry), and $\delta \in [0.8, 0.95]$.

We show for another scene, Fig. 6, two different lineal combinations. On the one hand, in Fig. 6.*a* we use the priority-value combination p_v (4), made up also of colour, p_c (2), and geometry, p_d (3). And, on the other hand, in Fig. 6.*b* we combine our measures $\mathbf{C}_{\mathrm{b}}^{\mathbf{c}}$ (19) and C^{g} (20). The same values $N_s^{\mathrm{p}} = 4$ and $\delta = 0.9$ are used in both cases. A significant difference is observed: our combination tends to obtain more radical contrasts (highly or less complex cases) as opposed to the other option which takes values in a far more homogeneous interval. The explanation lies in the behaviour of the binary colour contrast which works exclusively with the extreme data.

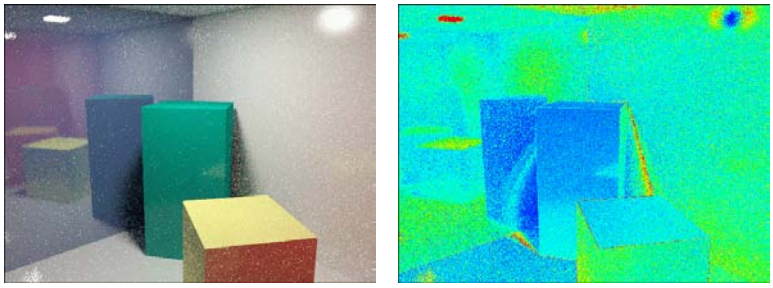
5. Entropy-Based Supersampling

In this section, we apply the newly defined contrast measures to supersampling in a stochastic ray-tracing implementation.

5.1. Method

Ray-tracing is a point-sampling-based technique for image synthesis. Rays are traced from the camera through a pixel to sample radiance at the hitpoint in the scene, where radiance is usually computed by a random walk method. Since a finite set of samples is used, some of the information in the scene is lost. Thus, aliasing errors are unavoidable [DW85]. These errors can be reduced by using extra sampling, called *supersampling*, in regions where the sample values vary most.

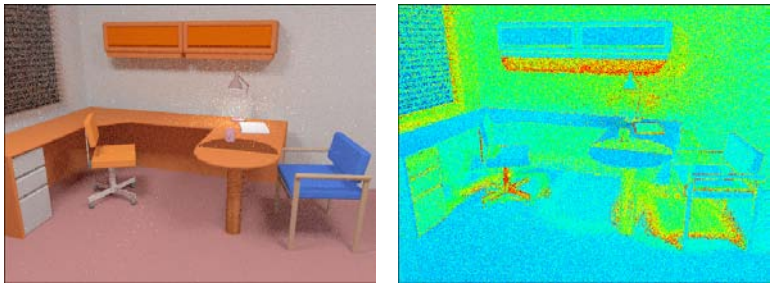
In order to obtain reliable data to achieve photo-realistic effects (e.g., diffuse and specular interreflections, shadow



(*a*) Reference (*b*) p_c

(*c*) $\mathbf{C^c}$ (*d*) $\mathbf{C^c_b}$

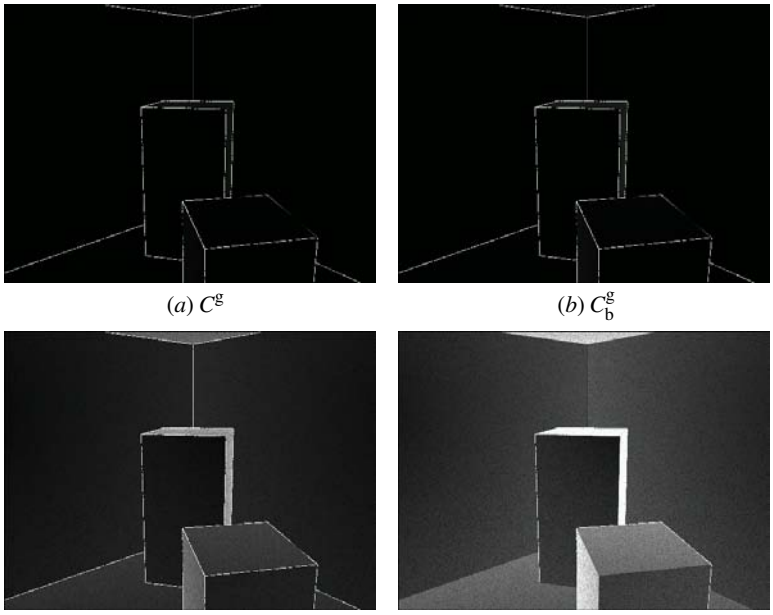
Figure 2: *Colour contrast maps. (a) Reference image obtained with* $N_s^p = 8$. *(b) Pixel colour contrast* p_c *(2). (c) Pixel colour contrast* $\mathbf{C^c}$. *(d) Pixel colour binary contrast* $\mathbf{C^c_b}$.



(*a*) Reference (*b*) p_c

(*c*) $\mathbf{C^c}$ (*d*) $\mathbf{C^c_b}$

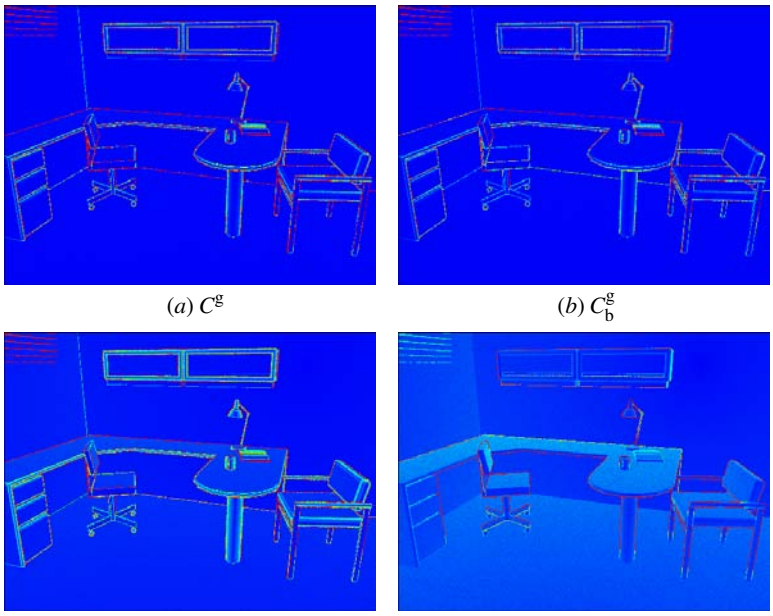
Figure 3: *Colour contrast maps. (a) Reference image obtained with* $N_s^p = 8$. *(b) Pixel colour contrast* p_c *(2). (c) Pixel colour contrast* $\mathbf{C^c}$. *(d) Pixel colour binary contrast* $\mathbf{C^c_b}$.
Credit: Modelled by Gregory J. Ward, Albany (CA), USA.



(*a*) C^{g} (*b*) $C^{\mathrm{g}}_{\mathrm{b}}$

(*c*) $C^{\mathrm{g}}_{\log}$ (*d*) p_d

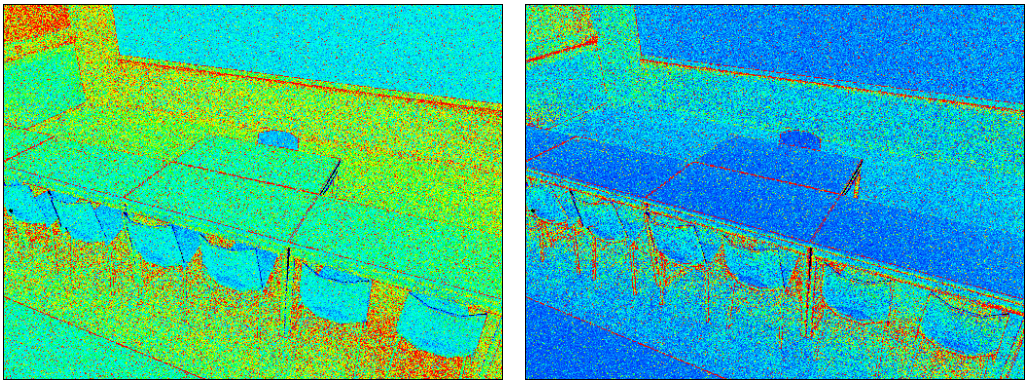
Figure 4: *Geometry contrast maps from Fig. 2.*a *obtained with* $N_s^{\mathrm{p}} = 8$. *(a) Pixel geometry contrast* C^g. *(b) Pixel geometry binary contrast* C^g_b. *(c) Pixel logarithmic-difference contrast* $C^g_{\log}$. *(d) Pixel depth difference* p_d.



(*a*) C^{g} (*b*) $C^{\mathrm{g}}_{\mathrm{b}}$

(*c*) $C^{\mathrm{g}}_{\log}$ (*d*) p_d

Figure 5: *Geometry contrast maps from Fig. 3.*a *obtained with* $N_s^{\mathrm{p}} = 8$. *For a visual comparison between the geometry and colour contrast in Fig. 3, the thermic scale is used. (a) Pixel geometry contrast* C^g. *(b) Pixel geometry binary contrast* C^g_b. *(c) Pixel logarithmic-difference contrast* $C^g_{\log}$. *(d) Pixel depth difference* p_d.
Credit: Modelled by Gregory J. Ward, Albany (CA), USA.



(*a*) p_v (*b*) $\mathcal{C}^{\mathbf{c}}$

Figure 6: *Pixel contrast obtained with $N_s^{\mathsf{p}} = 4$ and a lineal combination with $\delta = 0.9$. (a) Pixel priority-value p_v (4). (b) Pixel contrast $\mathcal{C}^{\mathbf{c}}$ using $\mathbf{C}_b^{\mathbf{c}}$ and C^g.*
Credit: Model included in RenderPark [Com00], Computer Graphics Research Group, Department of Computer Science, Katholieke Universiteit Leuven, Leuven, Belgium.

and penumbra, depth of field, motion blur, and translucency), the regions of the scene with the most complex illumination would need a more intensive treatment than a region with almost uniform illumination. This way of supersampling is called *adaptive sampling* [DW85, PS89]. A pixel is first sampled at a relatively low density. From this set of samples, a *refinement criterion* is used to decide whether more sampling is required or not. Finally, all the samples are used to obtain the final pixel colour values. We can consider two kinds of adaptive sampling: first, when the refinement criterion plays the role of an oracle which decides the place and the quantity of supersampling necessary in one evaluation of the initial sampling only and, second, when the refinement criterion constantly evaluates the information received because of a supersampling and acts in consequence until it becomes satisfied. In this work we use the term *supersampling* exclusively for the first case and *adaptive sampling* for the second (§6).

We implement a simple supersampling technique: the sample set S will be proportionally distributed over the image plane with respect to the contrast $\mathcal{C}^{\mathbf{c}}$ estimated in each $\mathsf{p} \in \mathsf{P}$. This is equivalent to the use of the *pixel contrast* as an oracle. For definition of the measure itself (§4), the cost in samples is controlled by the diversity of colour and geometry in the pixel (i.e., low quality). Given that a high contrast is synonymous of low pixel quality and low contrast of high quality pixel, the measure adapts the densities of sampling to the necessity of improvement in pixel quality. Remembering the importance of each one of the samples and without more prior information, this system will improve, on average, the pixel quality in particular and the image in general. The generic procedure is made up of three sequentially quite different phases:

Oracle A pixel contrast is selected as oracle and a first estimate of actual contrast per pixel is obtained using an initial stratified sampling against the image plane. Usual values are 2, 4, and 8 (n_s^{p}). If we consider that the total number of samples N_s^{P} destined for the image is prefixed, the final value of the average of samples per pixel is $N_s^{\mathsf{p}} = \frac{N_s^{\mathsf{P}}}{N_{\mathsf{p}}}$ and then, $1 < n_s^{\mathsf{p}} \ll \lfloor N_s^{\mathsf{p}} \rfloor$ where the left side inequality is due to the definition of the measure of contrast and the right side inequality to being able to carry out supersampling. The result of this phase is the answer from the oracle: the contrast map.

Sampling The unused samples in the calculation of the contrast, $N_{\mathsf{p}}(N_s^{\mathsf{p}} - n_s^{\mathsf{p}})$, are proportionally distributed to the values of contrast obtained per pixel. In each one of them, the sampling is also carried out with stratification. The distribution of the new set of samples gathers information from the scene in the regions of more diversity, with a proportional effort on this. The result of this phase is a supersampling directed exclusively by the contrast map generated in the previous phase.

Reconstruction The colour information gathered in the previous phase is put together with that obtained in the initial phase. Its evaluation allows us to achieve a more precise vision of the contents of the pixel and as a result an improvement in its quality. It only remains for the signal to be recuperated and to carry out the resampling process for each one of the pixels $\mathsf{p} \in \mathsf{P}$ with any of the reconstruction methods applicable to the sampling system used. The result of this phase is the solution for the image plane thanks to the assignment of the final colour to all of its pixels.

This proceeding is adaptable in any of its phases (e.g.,

stochastic ray-tracing method, pattern of initial sampling, supersampling method, and filters).

5.2. Results

Here, we show an example of our contrast measure $\mathcal{C}^{\mathbf{c}}$ (23) used as a supersampling oracle in path-tracing.

In Fig. 7.*a.i* we show a supersampling image obtained with $N_s^{\mathsf{p}} = 32$ in the following way. First, a uniform stratified sampling with $n_s^{\mathsf{p}} = 8$ has been made in order to obtain the contrast map in Fig. 7.*a.ii*. Secondly, this map has been used in the supersampling process with an average of 24 rays per pixel. And thirdly, in order to analyse the behaviour of the contrast, the signal reconstruction in the last phase is carried out by a piece-wise continuous reconstruction using a box filter. The final pixel value corresponds to its signal average. The contrast measure used is a colour and geometry combination with $\delta = 0.5$ based on binary contrasts $\mathbf{C}_b^{\mathbf{c}}$ (19) and C_b^g (21). This means that the more critical the area, the more supersampled it is (warm colours), and the less critical, the more undersampled it is (cool colours, with a minimum of 8 rays per pixel). Two detailed regions are compared from the supersampling image (Figs. 7.*b–c.i*) and a similar image obtained by uniform stratified sampling with $N_s^{\mathsf{p}} = 32$ (Figs. 7.*b–c.ii*). We can observe a decrease in noise in the supersampled regions, and a better representation of shadow contour and edges.

6. Entropy-Based Adaptive Sampling

In (9) we have seen that image plane and pixel channel entropies are related thanks to the grouping property of entropy. It is important to note that this kind of decomposition can be applied recursively if the pixels are recursively subdivided. We will show in this section that this recursive decomposition provides us with a natural method of dealing with an adaptive sampling technique. Our scheme, valid for any pixel sampling and ray-tracing method, is applied to stochastic ray-tracing and compared with other options.

6.1. Adaptive Sampling

In order to obtain a realistic image, the aliasing has to be reduced by adapting the density sampling to the complexity of the region aiming at a good balance between cost and quality. We consider three phases in order to describe a generic process of adaptive sampling [Gla95] (Fig. 8), for which the scheme used in the supersampling procedure of §5.1 is a particular case:

Initial sampling An *initial sampling* pattern at a predetermined density is established. Normally, in order to choose its density we assume that the signal has a Nyquist rate similar to the frequency of the reconstruction samples (e.g., one sample per pixel (the minimum number of samples per pixel fulfils $N_s^{\mathsf{p}} > 1$ (§3)). It is also usual for this density to be constant across the space, even though in the next phase, it should be increased locally if necessary.

Refinement tree The image space is divided up into regions (e.g., pixels). For each one of them, a *refinement test geometry* selects a subset of samples for evaluation. A *refinement test* is a criterion used for the evaluation of one or more characteristics which estimate the good quality of the current density. If the result is negative, a new set of samples are generated at the points indicated by the new *sampling geometry* and the process goes back to the refinement test geometry until the refinement criterion decides that the density of sampling in the region is accurate enough. The result of this process is a *refinement tree* of the image space where every node is a region with a density of sampling adapted to its own signal. In order to control extreme cases, it is usual to dispose of other criteria to finish the recursion (e.g., minimum area of the regions and/or maximum depth of tree).

Reconstruction The information of the signal collected at every region is unified by a *reconstruction* process and, if necessary, sent to a *filtering* process. Finally, a *resampling* process (e.g., centre of pixel) determines the final values for each of the pixels on the image plane.

Note that however much we increase the density of sampling locally, given that the signal is not usually band-limited, the sampling theory tell us that we can never capture it correctly. Thus, fine details of edges, shadings, textures, and others will hardly have enough quality in the final image. The approximation done by the method consists in looking for the minimum set of samples which estimates the signal locally accurately enough. A critical subproblem appears in each of its phases [PS89] and many approaches are found to deal with them (§2).

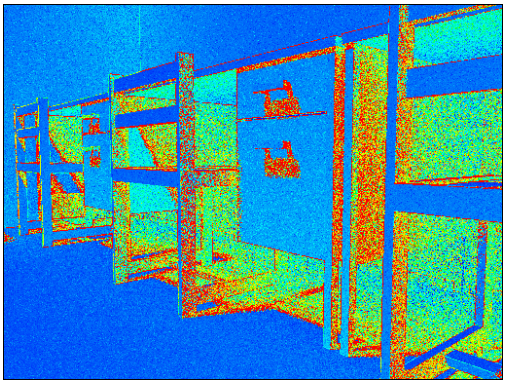
We focus our attention on obtaining an adaptive algorithm centred mainly on the refinement tree phase bringing a new perspective to the subproblem of controlling the sampling rate (new refinement criterion).

6.2. Recursive Entropy Tree

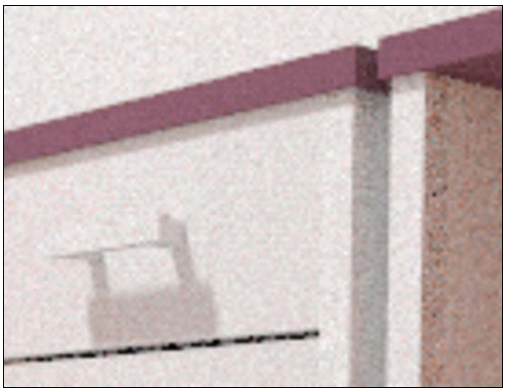
The natural way to represent information is by entropy, which in our context is interpreted as a measure of the degree of homogeneity of a region. Thus, using an entropy criterion means to evaluate the homogeneity (§3) or heterogeneity (§4) on a pixel. The fundamental idea behind our scheme is to capture the information in the refinement tree which results from the recursive decomposition of the entropy.

Generalising the grouping property of entropy, the entropy can be recursively decomposed in the following way: Let X be a discrete random variable over the set $\mathcal{X} = \{x_1, \ldots, x_n\}$ with probability distribution $p = \{p_1, \ldots, p_n\}$. Let us consider a partition of the set $\mathcal{X}$ in m-disjoint sets $\mathcal{Y} = \{\mathcal{Y}_1, \ldots, \mathcal{Y}_m\}$ where $|\mathcal{Y}_j| = n_j$. Let us associate the discrete random variable Y to $\mathcal{Y}$ with probability distribution

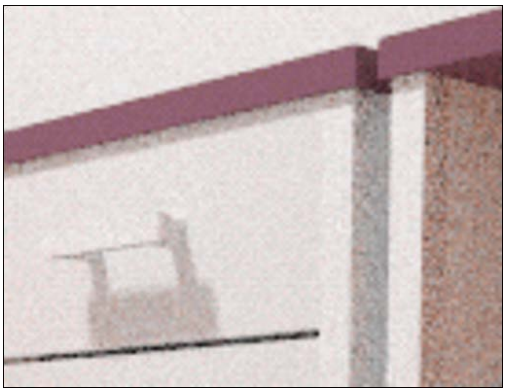
(*a.i*) Supersampling image



(*a.ii*) Oracle $\mathcal{C}^{\mathbf{c}}$ for (*a.i*)



(*b.ii*) Close-up from (*a.i*)



(*b.ii*) Close-up from uniform sampling

(*c.i*) Close-up from (*a.i*)

(*c.ii*) Close-up from uniform sampling

Figure 7: *Entropy-based supersampling versus uniform sampling. (a.i) Supersampling with an* $N_s^{\mathbf{p}} = 32$*. (a.ii) Binary contrast map* $\mathcal{C}^{\mathbf{c}}$ *used as oracle to obtain (a.i). It has been calculated with* $n_s^{\mathbf{p}} = 8$*,* $\mathbf{C}_b^{\mathbf{c}}$*,* C_b^g*, and* $\delta = 0.5$*. Close-up details from (a.i) are shown in (b-c.i). They are compared with the same regions, (b-c.ii) respectively, taken from a uniform stratified sampling image with* $N_s^{\mathbf{p}} = 32$*.*

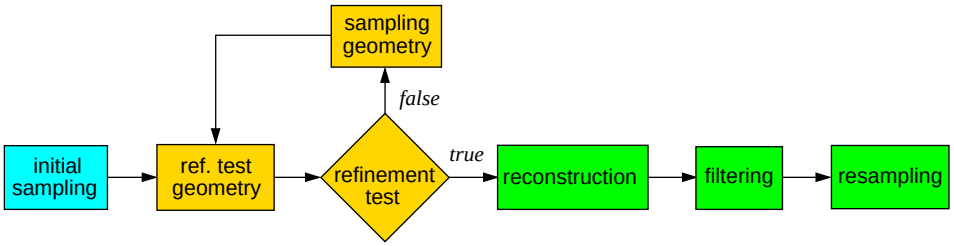


Figure 8: *Adaptive sampling process with three phases: initial sampling (blue), refinement tree (gold), and image reconstruction (green).*

$q = \{q_1, \ldots, q_m\}$ where $q_j = \sum_{k=1}^{n_j} p_{j_k}$ ($j_k \in \{1, \ldots, n\}$), and a new discrete random variable Y_j to each set $\mathcal{Y}_j$ with probability distribution $r_j = \{r_{j_1}, \ldots, r_{j_{n_j}}\}$ where $r_{j_k} = \frac{p_{j_k}}{q_j}$. Then

$$H(X) = \sum_{j=1}^{m} q_j H(Y_j) - \sum_{j=1}^{m} q_j \log q_j. \qquad (24)$$

This formula can be written as $H(X) = H_{\text{in}}(\mathcal{Y}) + H_{\text{out}}(\mathcal{Y})$ where $H_{\text{in}}(\mathcal{Y}) = \sum_{j=1}^{m} q_j H(Y_j)$ and $H_{\text{out}}(\mathcal{Y}) = H(Y) = -\sum_{j=1}^{m} q_j \log q_j$ represent, respectively, the hidden information (pending to be discovered) and the information already acquired in the descent of the tree created from an $\mathcal{Y}$ partition (Fig. 9).

In our case, (24) can also be interpreted taking into account only one colour channel (8) in the following way:

- $H(X)$ represents the entropy of the image plane.
- $H(Y_j)$ represents the entropy of each root pixel.
- Probability q_j is the proportion between the channel colour of pixel j and the sum of the channel colour of all pixels. It can be considered the "importance" of pixel j.

The decomposition of entropy can be recursively extended to the subpixels. This interpretation can also be applied to geometry entropy (14).

In our approach, probabilities are obtained by stochastic sampling. From the definition of entropy, we can see that when the number of samples tends to infinity, entropy also goes to infinity. In fact, we can consider that the original continuous scene contains infinite information. The following sampling algorithm will extract more information from the regions with more sample variation.

6.3. Algorithm

We present a new adaptive scheme for adaptive sampling, complementary to the entropy-based supersampling method (§5.1), with the important feature that it is based on the recursive expression of the Shannon entropy (i.e., the entropy tree). For the sake of simplicity, in the following analysis we only consider the colour information of one channel, although in the final algorithm we will take the combination of colour and geometry contrasts into account (23).

A general description of our algorithm is as follows: On the image plane we sample each pixel to capture the colour of hitpoints and thus evaluate the information content (entropy) from the colour probability distribution. If the information of a pixel is high enough (i.e., the rays provide us with sufficient colour homogeneity on that pixel), refinement is not made, and the colour reconstruction of this pixel is done. When the information is not high enough, this pixel is subdivided into regions and we proceed in the same way for each region (subpixel). *"The approach will be to make sure that all the samples in a given region are similar in some specified way, so we can feel that we have captured what is happening in a region of the signal"* [Gla95].

This recursive process defines a tree with two well-separated phases for a pixel:

- *Pixel refinement.* Until enough information is extracted (tree descent).
- *Pixel colour.* Computation of the final colour (tree ascent).

The descent in the refinement tree can be interpreted as a progressive gain in information. The information acquired at each level is added together so that, at the end of the refinement process, the total information from the tree is the sum of the information obtained over all the branches (24). The measure used to capture the information will be the pixel contrast (§4).

Before introducing the algorithm we will give the definitions of the data used in it. Concerning the tree data structure, n represents the tree level where

- $n = 0$ is the image level (root).
- $n = 1$ is the pixel level (composed of N_{p} pixels of the image).
- $n > 1$ is the subpixel level.

We consider an n-node at any node of the tree with a level of $n > 0$ (i.e., no root). The set of data is described in Table 1. To compute the final colour of a pixel, we follow a path through the tree (Fig. 10). In the analysis below, we focus our

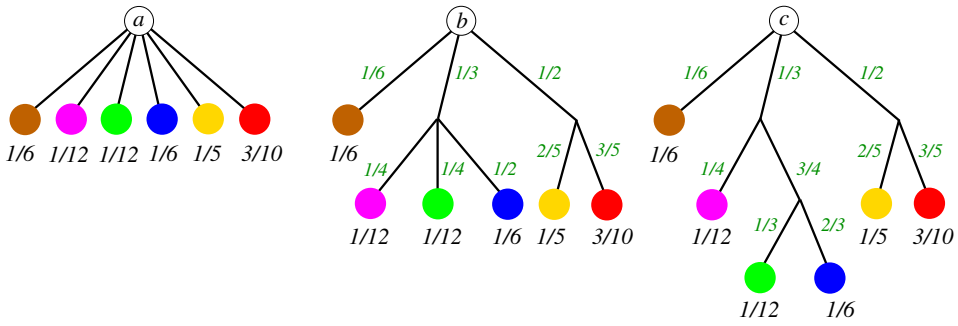


Figure 9: *Grouping property of entropy. The entropy of probability distributions of ⓐ is* $H(\{\frac{1}{6},\frac{1}{12},\frac{1}{12},\frac{1}{6},\frac{1}{5},\frac{3}{10}\})$*, of ⓑ is* $H(\{\frac{1}{6},\frac{1}{3},\frac{1}{2}\})+\frac{1}{3}H(\{\frac{1}{4},\frac{1}{4},\frac{1}{2}\})+\frac{1}{2}H(\{\frac{2}{5},\frac{3}{5}\})$*, and of ⓒ is* $H(\{\frac{1}{6},\frac{1}{3},\frac{1}{2}\})+\frac{1}{3}(H(\{\frac{1}{4},\frac{3}{4}\})+\frac{3}{4}H(\{\frac{1}{3},\frac{2}{3}\}))+\frac{1}{2}H(\{\frac{2}{5},\frac{3}{5}\})$*. Accordingly to (24), they all have the same value:* $H(a)=H(b)=H(c)=2.445$.

attention on the tree-path k of length m going from pixel k_0 to subpixel k_{m-1}. In this path, p_n represents the probability of the tree-branch at level n and q_n the *importance* of the n-node. In our algorithm, this quantity appears naturally due to recursive decomposition of the entropy (see (24), Fig. 9, and Fig. 10). The value of importance is given by the probability of the n-node:

$$q_n = \begin{cases} 1, & \text{if } n=0, \\ p_0\cdots p_{n-1} = \frac{\overline{c}_{0,k_0}}{\sum_{i\in R_0}\overline{c}_{0,i}}\prod_{\ell=1}^{n-1} p_\ell, & \text{if } n>0. \end{cases} \tag{25}$$

For our purposes, q_n does not need to be normalised, thus we omit the normalisation constant $\sum_{i\in R_0}\overline{c}_{0,i}$ and we take $q_n = \overline{c}_{0,k_0}\prod_{\ell=1}^{n-1} p_\ell$.

Proposition 1 The computation of q_n can be simplified to (in an abuse of notation, all the superindexes corresponding to arithmetic expressions must be interpreted as a power)

$$q_n \approx \frac{\overline{c}_n}{N_r^{n-1}}. \tag{26}$$

Observe first that for a given path and $n>0$, the colour $\overline{c}_n$ of an n-node is more accurate than the colour average of its respective region, k_{n-1}, in the preceding level. Thus, the accuracy of p_n, and at the same time of q_n, can be increased by substituting $\overline{c}_{n-1,k_{n-1}}$ for $\overline{c}_n$. Let us prove now (26) by induction.

Proof: For $n=1$,

$$q_n = \overline{c}_{0,k_0} \approx \overline{c}_1 = \frac{\overline{c}_1}{N_\mathrm{r}^0} = \frac{\overline{c}_n}{N_\mathrm{r}^{n-1}}.$$

Hypothesis: $\forall_{0<\ell<n}.\, q_\ell = \frac{\overline{c}_\ell}{N_\mathrm{r}^{\ell-1}}$. Then, for $n>1$

$$\begin{aligned} q_n &= \overline{c}_{0,k_0}\prod_{\ell=1}^{n-1} p_\ell = q_{n-1}p_{n-1} \\ &= \frac{\overline{c}_{n-1}}{N_\mathrm{r}^{n-2}}\frac{\overline{c}_{n-1,k_{n-1}}}{\sum_{i\in R_{n-1}}\overline{c}_{n-1,i}} \\ &\approx \frac{\overline{c}_{n-1}}{N_\mathrm{r}^{n-2}}\frac{\overline{c}_n}{\overline{c}_{n-1}N_\mathrm{r}} = \frac{\overline{c}_n}{N_\mathrm{r}^{n-1}}. \quad \square \end{aligned}$$

Now we can proceed to explain the algorithm. In the descent phase we sample an n-node and compute the contrast using expression $\mathcal{C}^{\mathbf{c}}$ (23). In (18) we must substitute the channel importance $\overline{c}$ by q_n and, according to §4, for a sRGB colour system we can take the perceptual coefficients $w^\mathrm{R}=0.2126$, $w^\mathrm{G}=0.7152$ and $w^\mathrm{B}=0.0722$ which capture the sensitivity of human colour perception [Com98].

Thus, for each n-node, the colour contrast (18) converts into

$$\mathbf{C}_n^{\mathbf{c}} = \sum_{c\in\mathbf{c}} w^c C_n^c q_n^c \tag{27}$$

and the colour and geometry combination (23) will be

$$\mathcal{C}_n^{\mathbf{c}} = \delta\mathbf{C}_n^{\mathbf{c}} + (1-\delta)C_n^{\mathrm{g}}. \tag{28}$$

Note that this expression could also be calculated from the respective binary versions of colour and geometry contrasts (§4).

In the algorithm, we subdivide the pixel or subpixel when the contrast of an n-node is not less than a given threshold ($\mathcal{C}_n^{\mathbf{c}} \geq \varepsilon$). Thus, the phase of ascent begins when the test fails ($\mathcal{C}_n^{\mathbf{c}} < \varepsilon$). This happens because either the contrast (which represents the colour heterogeneity) or the importance ($q_n \to 0$ for growing n) are low. In this phase, each n-node in the path provides its colour estimation $\widehat{c}_n$ from the signal reconstructed for each $c(s)$ where $s\in S_n$.

id	*description*	*asserts*
N_r	Number of regions in which an n-node can potentially be subdivided.	$N_\mathrm{r} > 1 \wedge A_{1\leq i\leq N_\mathrm{r}} = \frac{A_{n-\mathrm{node}}}{N_\mathrm{r}}$
N_s^r	Number of samples of an n-node.	$N_\mathrm{s}^\mathrm{r} \geq N_\mathrm{r} \wedge N_\mathrm{s}^\mathrm{r} \in N_\mathrm{r}\mathbb{N}^+$
R_n	Set of regions of an n-node.	$\lvert R_0\rvert = N_\mathrm{p} \wedge \forall_{n>0}.\, \lvert R_n\rvert = N_\mathrm{r}$
S_n	Set of samples of an n-node.	$\lvert S_0\rvert = N_\mathrm{s}^\mathrm{r} N_\mathrm{p} \wedge \forall_{n>0}.\, \lvert S_n\rvert = N_\mathrm{s}^\mathrm{r}$
$S_{n,i}$	Set of samples of an n-node region $i \in R_n$.	$S_n = \bigcup_{i\in R_n} S_{n,i}$ $\lvert S_{n,i}\rvert = \frac{\lvert S_n\rvert}{\lvert R_n\rvert} = \frac{N_\mathrm{s}^\mathrm{r}}{N_\mathrm{r}}$
k	Path-tree $k = (k_0,\ldots,k_{m-1})$ where k_n is the region chosen at level n.	$m > 0 \wedge \forall_{n<m}.\, k_n \in R_n$
$\bar{c}_n$	Average colour channel data in an n-node.	$\bar{c}_n = \frac{1}{\lvert S_n\rvert}\sum_{s\in S_n} c(s)$
$\bar{c}_{n,i}$	Average colour in an n-node region $i \in R_n$.	$\bar{c}_{n,i} = \frac{1}{\lvert S_{n,i}\rvert}\sum_{s\in S_{n,i}} c(s)$ $\bar{c}_n = \frac{1}{\lvert R_n\rvert}\sum_{i\in R_n} \bar{c}_{n,i}$
p_n	Probability of region k_n of an n-node in a path k.	$p_n = \frac{\sum_{s\in S_{n,k_n}} c(s)}{\sum_{s\in S_n} c(s)} = \frac{\bar{c}_{n,k_n}}{\sum_{i\in R_n}\bar{c}_{n,i}}$
q_n	Probability of an n-node in a path k.	$q_n = \prod_{\ell=0}^{n-1} p_\ell$

Table 1: *Description of the data set of the refinement phase of entropy-based adaptive sampling in an image plane of N_p pixels. An n-node is a node of level $n > 0$ in the refinement-tree.*

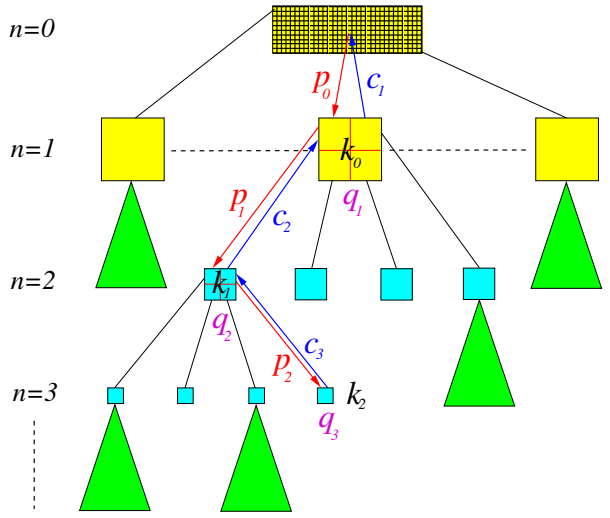


Figure 10: *A refinement-tree-path $k = (k_0, k_1, k_2)$ of length $m = 3$ of entropy-based sampling. The number of regions of an n-node is $N_r = 4$. We show the computation of the k_0-pixel colour: $c_{0,k_0} = c_1$ from the refinement (red) and reconstruction (blue) phases. The probabilities p_n and importances q_n are computed in the refinement phase to evaluate the entropy contrast (23).*

The final colour of an n-node is given by

$$c_n = \begin{cases} \widehat{c}_n, & \text{if } \mathcal{C}_n^{\mathbf{c}} < \varepsilon, \\ \sum_{i \in R_n} c_{n,i}, & \text{otherwise}, \end{cases} \tag{29}$$

where $c_{n,i}$ is the final colour of i-region of the n-node. Finally, we get c_1 for the colour of the pixels (or equivalently c_{0,k_0} in the path considered). An example of this process is shown in Fig. 10.

Observe that importance sampling is naturally integrated in the algorithm. Following importance sampling criteria a function should be sampled proportionally to its value which is what we obtain with our adaptive descent.

6.4. Results

For the purpose of comparison, in Fig. 12 we present the results for different techniques for the test scene in Fig. 11. We compare the following methods:

- *Classic contrast* (CC): A recursive adaptive sampling scheme based on contrast by channel (1) (with thresholds proportional to the visual system) weighted by its respective channel colour average [Gla95, SS00]. The maximum recursive level has been limited to 4 (Fig. 12.*a*).
- *Importance-weighted contrast* (IC): The same as in CC but each channel contrast is weighted with the respective importance q (26), as in our approach (Fig. 12.*b*).
- *Confidence test* (CT): Statistical approach based on a confidence interval (5) with a confidence level of $\alpha = 0.1$ and a tolerance $t = 0.025$ (Fig. 12.*c*).
- *Entropy-based contrast* (EC): Our approach (29) taking only colour contrast, $\delta = 1$ in (28) (Fig. 12.*d*).

Observe that the EC approach can be easily implemented on any standard hierarchical algorithm, using importance (26) and the new refinement criterion (28), with negligible additional cost.

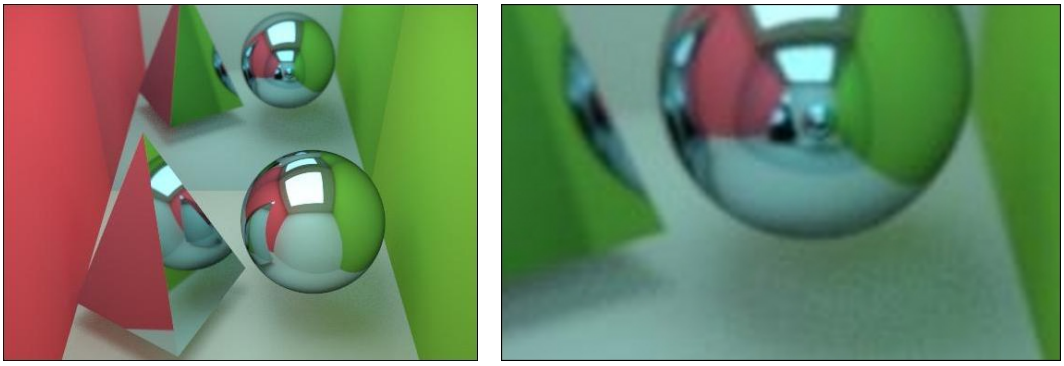
In CC, IC, and EC, the number of subdivisions is $N_{\mathrm{r}} = 4$ and the number of samples is $N_{\mathrm{s}}^{\mathrm{r}} = 8$. To compute the contrast measures for the refinement decision, the samples have been cast in a stratified way at each n-node (i.e., pixel or subpixel) and re-used at the next levels in the tree. In CT, groups of 8 samples were added in a stratified way until meeting the condition of the criterion. An implementation of classic path-tracing with next event estimator was used to compute all images. The parameters were tuned so that all four test images were obtained with a similar average number of rays per pixel ($N_{\mathrm{s}}^{\mathrm{p}} = 60$) and computational cost. The resulting images are shown in Figs. 12.$*$.*i* with close-ups in Figs. 12.$*$.*ii*. A sampling density map (SDM) for each one is given in Figs. 12.$*$.*iii* (generated under the same conditions as the quality (§3) and contrast (§4) maps; warm colours correspond to the highest sampling rate and cold colours to the lowest)

The overall aspect of the images in Figs. 12.$*$.*i* shows that our supersampling scheme performs best. Observe, for instance, the reduced noise in the shadows cast by the objects. This is further checked in the close-up images in Figs. 12.$*$.*ii*. Observe also the detail of the sphere shadow reflected on the pyramid. It is important to note that we managed to improve the classic contrast approach in CC greatly by including the importance used in our scheme (compare results in Fig. 12.*a* with Fig. 12.*b*). A comparison of the SDMs shows a better discrimination of complex regions of the scene in the entropy case against the classic contrast case. This explains the better results obtained by our approach. Moreover, the confidence test approach CT (Fig. 12.*c*) also performs better than the classic contrast-based methods CC (Fig. 12.*a*) and IC (Fig. 12.*b*). The SDM of CT also explains why it performs better. However, it is unable to render the reflected shadows under the mirrored pyramid and sphere with precision (see close-up in Fig. 12.*c*.*ii*).

In Table 2 we show two measures (error and quality) obtained in Figs. 12.$*$.*i–ii* with respect to the test scene in Figs. 11.*i–ii*, respectively. We select the root of the mean square error (it is calculated from the MSE of each colour channel) (RMSE) and the peak signal to noise ratio (PSNR) to evaluate the results (measure of the quality of a reconstructed image compared with an original image computing the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation). Because many signals have a very wide dynamic range, PSNR is usually expressed in terms of the logarithmic decibel scale: $10\log_{10}(I_{\max}^2/MSE)$ dB. For each one, we consider a weight balanced by every colour channel ($\mathrm{RMSE_a}$ and $\mathrm{PSNR_a}$) and a perceptual one ($\mathrm{RMSE_p}$ and $\mathrm{PSNR_p}$) in accordance with the sRGB system ($w^{\mathrm{R}} = 0.2126$, $w^{\mathrm{G}} = 0.7152$, and $w^{\mathrm{B}} = 0.0722$. These measures reflect the good behaviour mentioned in CT and EC oracles (i.e., low RMSEs and high PSNRs). Although the error obtained using our approach is bigger than that with CT method, the visual results are better in the EC case (observe Figs. 12.*c–d*). This is due to the fact that the measures do not manage to reflect exactly the perceptual quality of the image. The EC images look better because the oracle distributes the samples in the perceptual critical regions more accurately (see SDMs).

Now, we present a test using the geometry component with $1 - \delta = 0.1$ in (28) and, at the same time, the binary contrast in colour and geometry. To do this, our approach is compared with the priority-value combination (4) made up of a colour contrast of the CC type and also by an usual geometry factor. Perceptual coefficients are taken equal as in our approach in both cases (§6.3). The tree depth level is set to 4 and the $N_{\mathrm{s}}^{\mathrm{r}}$ is reduced by half (i.e., 4 samples) but maintaining the average per pixel (i.e., $N_{\mathrm{s}}^{\mathrm{p}} = 60$).

The images obtained are shown in Fig. 13. In Fig. 13.*a*, the entropy-contrast $\mathcal{C}_n^{\mathbf{c}}$ (28) with $\mathbf{C}^{\mathbf{c}}$ (18) and C^{g} (20). In Fig. 13.*b*, the binary-entropy contrast: $\mathcal{C}_n^{\mathbf{c}}$ using $\mathbf{C}_{\mathrm{b}}^{\mathbf{c}}$ (19)



(*i*) General view (*ii*) Close-up of (*i*)

Figure 11: *Reference image used in the test in Fig. 12: (i) general view and (ii) close-up of (i). The image has been obtained with a path-tracing algorithm with 1,024 samples per pixel in a stratified way.*

oracle	*general view*				*close-up*			
	$RMSE_a$	$RMSE_p$	$PSNR_a$	$PSNR_p$	$RMSE_a$	$RMSE_p$	$PSNR_a$	$PSNR_p$
CC	13,727	13,599	25,379	25,461	20,276	20,024	21,991	22,100
IC	8,124	8,110	29,935	29,951	13,751	13,568	25,364	25,481
CT	5,194	5,174	33,822	33,855	8,407	8,338	29,638	29,710
EC	6,937	7,018	31,308	31,207	9,886	9,933	28,231	28,189

Table 2: *The RMSE and PSNR of CC, IC, CT, and EC oracles applied to the general view (Fig. 11.*i*) and close-up (Fig. 11.*ii*) of the test scene. The average number of rays per pixel is $N_s^p = 60$ in all methods.*

and C_b^g (21). And, in Fig. 13.*c*, the priority-value approach p_v (4) with p_c (2) and p_d (3). The respective SDMs from Figs. 13.∗.*i* are shown in Figs. 13.∗.*ii*.

We see from comparing the images that the entropy contrast is much better than the classic contrast used here. Observe for instance the ceiling, the shadows and the mirroring wall. A drawback of our approach is the peaks of high radiance that we observe on the right wall because this region is undersampled in our method. However, this effect can be easily solved by a filtering technique. The comparison of the SDMs shows a better discrimination of complex regions of the scene in the entropy case (Figs. 13.*a–b.ii*) against the priority-value contrast case (Fig. 13.*c.ii*). This explains the better results obtained with our approach.

Finally, in Fig. 14.*a* we show another scene obtained with our approach using an average of $N_s^p = 200$ and $\delta = 0.95$. Observe, in Fig. 14.*b*, how well the SDM works out both the geometry and colour details as in the shadow contours on the walls (the remaining spiked noise could easily be eliminated by filtering with an image smoothing method).

References

[BM98] BOLIN M. R., MEYER G. W.: A perceptually based adaptive sampling algorithm. In *SIGGRAPH '98 Conference Proceedings* (New York (NY), USA, July 1998), Cohen M., (Ed.), Annual Conference Series, ACM SIGGRAPH, ACM Press, pp. 299–309.

[Cae81] CAELLI T. M.: *Visual Perception: Theory and Practice*. Pergamon Press, Oxford, UK, 1981.

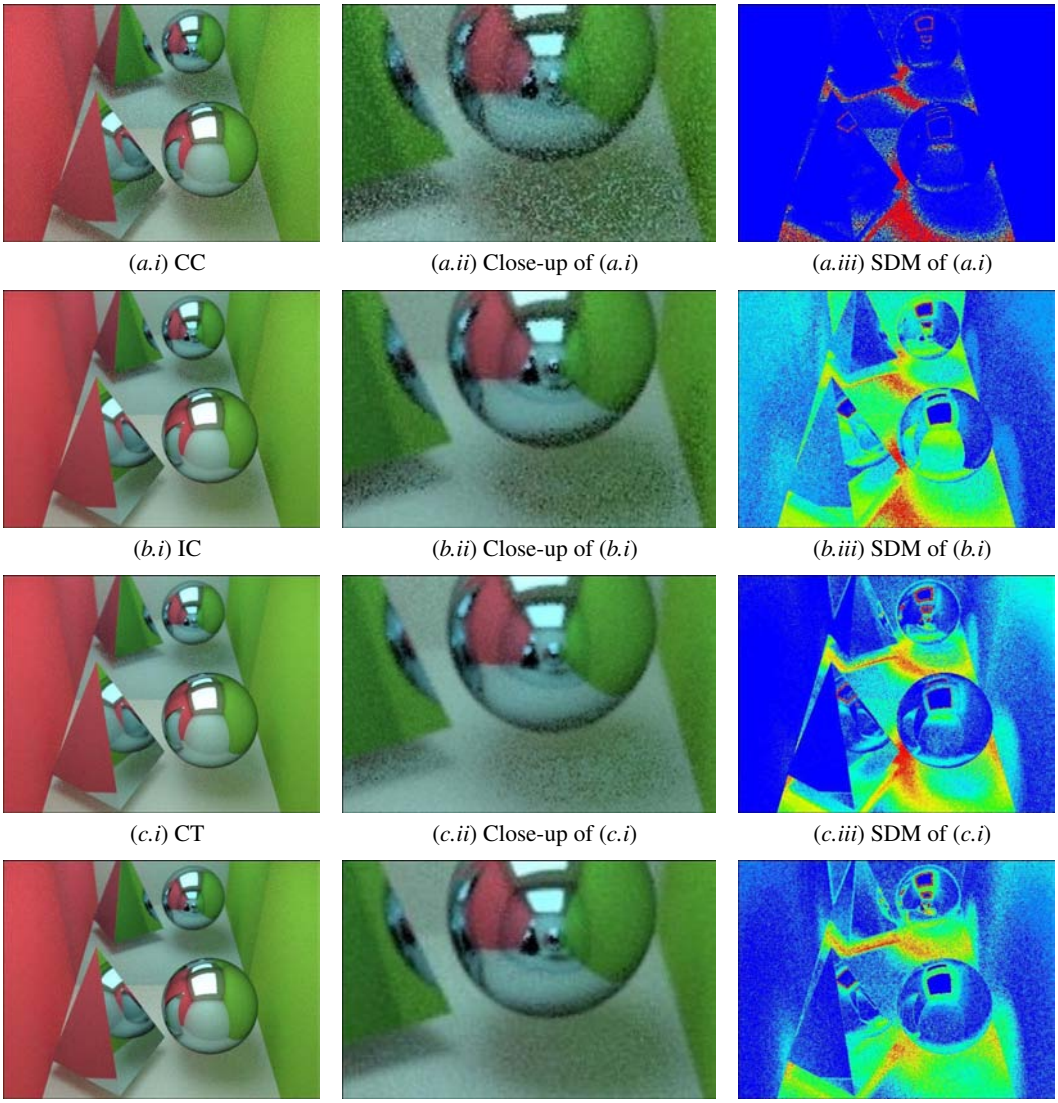
[Com98] COMMISSION I. E.: Default RGB colour space - sRGB. In *Colour Management in Multimedia Systems*. IEC, Geneva, Switzerland, 1998, ch. Part 2.1. IEC/PT61966(PL)34.

[Com00] COMPUTER GRAPHICS RESEARCH GROUP: *RenderPark: A Photorealistic Rendering Tool*. Software, Department of Computer Science, Katholieke Universiteit Leuven, Leuven, Belgium, 2000. http://www.renderpark.be.

[CPC84] COOK R. L., PORTER T., CARPENTER L.: Distributed ray tracing. *Computer Graphics (Proceedings of SIGGRAPH '84) 18*, 3 (July 1984), 137–145.

[DC96] DARSA L., COSTA B.: Multi-resolution representation and reconstruction of adaptively sampled images. In *Proceedings of IX Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI '96)* (October 1996), pp. 321–328.

[DCV97] DARSA L., COSTA B., VARSHNEY A.: Navigating static environments using image-space simplification and morphing. In *1997 Symposium on Interactive*



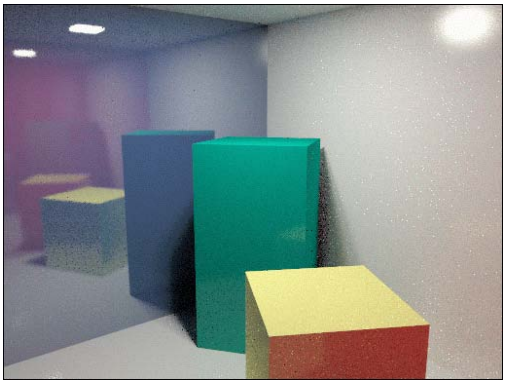
(*a.i*) CC (*a.ii*) Close-up of (*a.i*) (*a.iii*) SDM of (*a.i*)

(*b.i*) IC (*b.ii*) Close-up of (*b.i*) (*b.iii*) SDM of (*b.i*)

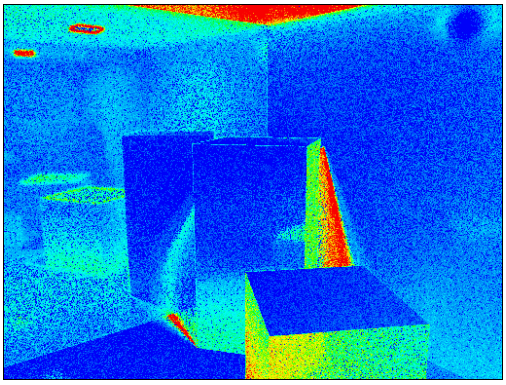
(*c.i*) CT (*c.ii*) Close-up of (*c.i*) (*c.iii*) SDM of (*c.i*)

(*d.i*) EC (*d.ii*) Close-up of (*d.i*) (*d.iii*) SDM of (*d.i*)

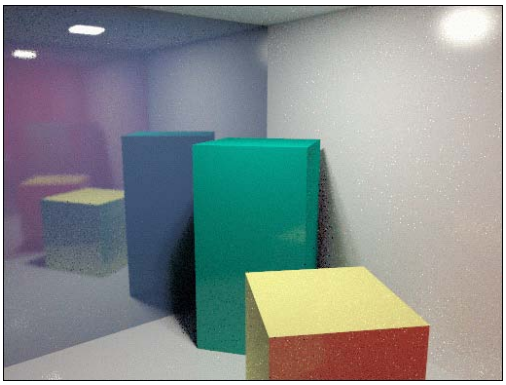
Figure 12: *Results of comparisons: (a) adaptive sampling scheme based on classic contrast (CC), (b) importance-weighted contrast (IC), same as in (a) but weighting with importance q (26), (c) confidence test method (CT), and (d) entropy-based method (EC) with colour contrast only ($\delta = 1$). By columns: (i) shows the resulting images, (ii) close-up of regions of (i), and (iii) the sampling density maps of (i). The average number of rays per pixel is $N_s^p = 60$ in all methods, with a similar computational cost.*



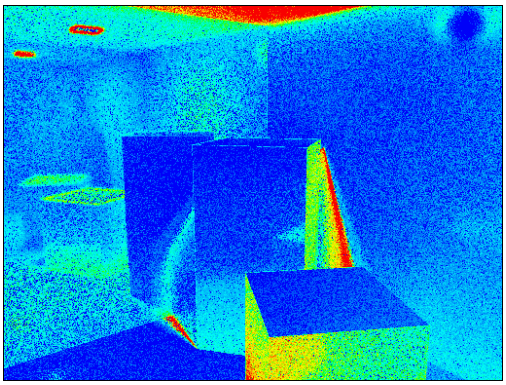
(*a.i*) Entropy contrast



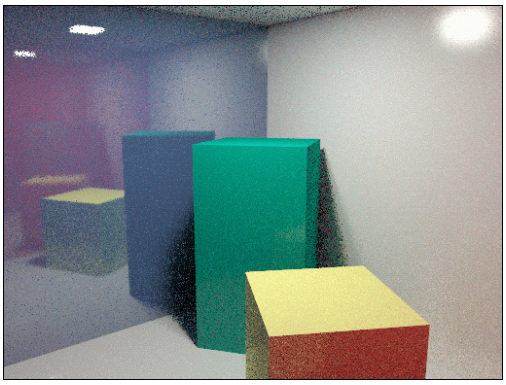
(*a.ii*) SDM of (*a.i*)



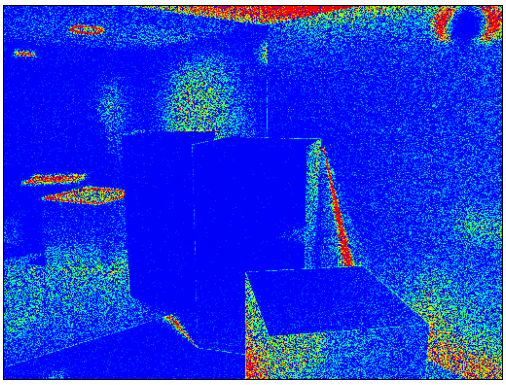
(*b.i*) Binary-entropy contrast



(*b.ii*) SDM of (*b.i*)

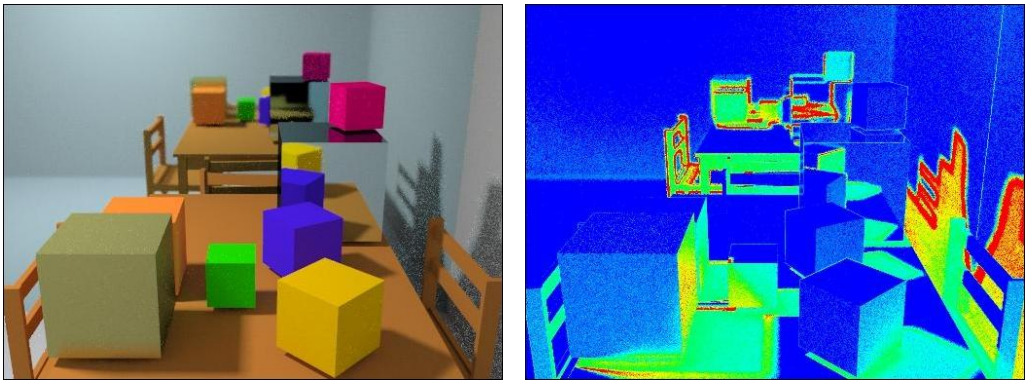


(*c.i*) Priority-value contrast



(*c.ii*) SDM of (*c.i*)

Figure 13: *Images obtained with adaptive sampling where* $N_{\bar{s}}^{\mathsf{p}} = 60$ *and* $N_s^r = 4$*: (a) entropy contrast, (b) binary-entropy contrast, and (c) priority-value contrast. By columns: (i) Image sampled and (ii) sampling density map of (i).*



(*a*) Entropy-based contrast (*b*) SDM of (*a*)

Figure 14: *Image obtained with entropy-based adaptive sampling where $N_s^p = 200$ and $\delta = 0.95$. (a) Sampled image. (b) Sampling density map of (a).*

3D Graphics (April 1997), Cohen M., Zeltzer D., (Eds.), ACM SIGGRAPH, pp. 25–34.

[DW85] DIPPÉ M. A. Z., WOLD E. H.: Antialiasing through stochastic sampling. *Computer Graphics (Proceedings of SIGGRAPH '85) 19*, 3 (July 1985), 69–78.

[Gla95] GLASSNER A. S.: *Principles of Digital Image Synthesis*. Morgan Kaufmann Publishers, San Francisco (CA), USA, 1995.

[KA91] KIRK D., ARVO J.: Unbiased variance reduction for global illumination. In *Proceedings of the 2nd Eurographics Workshop on Rendering* (May 1991), pp. 153–156.

[Kaj86] KAJIYA J. T.: The rendering equation. *Computer Graphics (Proceedings of SIGGRAPH '86) 20*, 4 (August 1986), 143–150.

[KH94] KELLER A., HEINRICH S.: *Quasi-Monte Carlo Methods in Computer Graphics, Part I: The QMC-Buffer*. Tech. Rep. 242/94, University of Kaiserslautern, Kaiserslautern, Germany, 1994.

[LRU85] LEE M. E., REDNER R. A., USELTON S. P.: Statiscally optimized sampling for distributed ray tracing. *Computer Graphics (Proceedings of SIGGRAPH '85) 19*, 3 (July 1985), 61–67.

[MF92] MCCOOL M., FIUME E.: Hierarchical Poisson disk sampling distributions. In *Proceedings of Graphics Interface '92* (May 1992), pp. 94–105.

[Mit87] MITCHELL D. P.: Generating antialiased images at low sampling densities. *Computer Graphics (Proceedings of SIGGRAPH '87) 21*, 4 (July 1987), 65–72.

[OA96] OHBUCHI R., AONO M.: *Quasi–Monte Carlo Rendering with Adaptive Sampling*. Tech. rep., Tokyo Research Laboratory, IBM Japan Ltd., Tokyo, Japan, 1996.

[PCD*97] PULLI K., COHEN M. F., DUCHAMP T., HOPPE H., SHAPIRO L., STUETZLE W.: View-based rendering: Visualizing real objects from scanned range and color data. In *Rendering Techniques '97 (Proceedings of the 8th Eurographics Workshop on Rendering)* (New York (NY), USA, June 1997), Dorsey J., Slusallek P., (Eds.), Springer-Verlag, pp. 23–34.

[PS89] PAINTER J., SLOAN K.: Antialiased ray tracing by adaptive progressive refinement. *Computer Graphics (Proceedings of SIGGRAPH '89) 23*, 3 (July 1989), 281–288.

[Pur86] PURGATHOFER W.: A statistical method for adaptive stochastic sampling. *Eurographics '86: Proceedings of the European Conference and Exhibition 11*, 2 (August 1986), 157–162.

[RFBS01] RIGAU J., FEIXAS M., BEKAERT P., SBERT M.: View-dependent information theory measures for pixel sampling and scene discretization in flatland. In *Proceedings of Spring Conference on Computer Graphics '01* (Los Alamitos (CA), USA, April 2001), IEEE Computer Society, pp. 173–180.

[Sch91] SCHLICK C.: An adaptive sampling technique for multidimensional ray tracing. In *Proceedings of the 2nd Eurographics Workshop on Rendering* (May 1991), pp. 48–56.

[Shi90] SHIRLEY P.: A ray tracing method for illumination calculation in diffuse-specular scenes. In *Proceedings of Graphics Interface '90* (Toronto (Ontario), Canada, May 1990), Canadian Information Processing Society, pp. 205–212.

[SS00] SIMMONS M., SÉQUIN C. H.: Tapestry: A dynamic mesh-based display representation for interactive rendering. In *Rendering Techniques 2000 (Proceedings of the 11th Eurographics Workshop on Rendering)* (New

York (NY), USA, June 2000), Péroche B., Rushmeier H., (Eds.), Springer-Verlag, pp. 329–340.

[TJ97] TAMSTORF R., JENSEN H. W.: Adaptive sampling and bias estimation in path tracing. In *Rendering Techniques '97 (Proceedings of the 8th Eurographics Workshop on Rendering)* (New York (NY), USA, June 1997), Dorsey J., Slusallek P., (Eds.), Springer-Verlag, pp. 285–295.

Applications of Information Theory to Computer Graphics Part V: Viewpoint Selection and Mesh Saliency

Miquel Feixas, Mateu Sbert, and Francisco González García

University of Girona, Spain

1. Introduction

In computer graphics, several *viewpoint quality* measures have been applied in areas such as scene understanding [PB96, VFSH01, PPB*05], scene exploration [AVF04, SPT06], image-based modeling [VFSH03], and volume visualization [BS05, TFTN05, VFSG06, JS06]. In other areas, such as object recognition and mobile robotics, best view selection is also a fundamental task. Many works have demonstrated that the recognition process is view-dependent [PRC81, BET95, TBZB97, BTB99]. In [TBZB97], the authors found that "visual recognition may be explained by a view-based theory in which viewpoint-specific representations encode both quantitative and qualitative features". In robotics, the simultaneous localization and mapping problem (SLAM) requires that the robot decides on its own the necessary motions to construct the most accurate map possible. In [GBL02], an algorithm is proposed to guide the robot through a series of good positions, where 'good' refers to the expected amount and quality of the information that will be revealed at each new location.

The basic question underlying the viewpoint selection study and application is "what is a 'good' scene viewpoint?" Obviously, this question does not have a unique answer. Depending on our objective, the best viewpoint can be, for instance, the most *representative* one or the most *unstable* one, i.e., the one that maximally changes when it is moved within its close neighborhood [BS05]. [PRC81] and [BTB99] have presented different experiments demonstrating that observers prefer views (called *canonical views*) that avoid occlusions and that are off-axis (such as a three-quarter viewpoint), salient (the most significant characteristics of an object are visible), stable and with a large number of visible surfaces.

Extending the work initiated in [VFSH01, SPFG05], in this tutorial we will present a unified and robust framework to deal with viewpoint selection and mesh saliency. Given a set of viewpoints surrounding the object, we define an *information channel* between the viewpoints and the polygons of the object. From this channel, the *viewpoint mutual information* is used to obtain the best views of an object, to calculate the stability of a viewpoint, and to guide the object exploration. Then, we reverse the channel and we compute both the information and the saliency associated with each polygon. Finally, this polygonal saliency is used to calculate how salient is a viewpoint and is incorporated to viewpoint mutual information to drive the viewpoint selection. Our framework is also applicable to any set of viewpoints in a closed scene and, although only the geometric properties of an object have been considered, other aspects such as lighting could be incorporated.

In these notes, we only present some related work (Section 2) and the viewpoint channel (Section 3). Other aspects of the viewpoint framework (viewpoint similarity and unstability, object exploration, saliency, information-theoretic ambient occlusion, best view selection, etc.) can be found in the slides corresponding to this Part V.

2. Related Work

We review now some viewpoint quality measures for polygonal models. In [PB96], the quality of a viewpoint v of a scene is computed using the *heuristic measure* (HM) given by

$$C(v) = \frac{\sum_{i=1}^{n} \lceil \frac{P_i(v)}{P_i(v)+1} \rceil}{n} + \frac{\sum_{i=1}^{n} P_i(v)}{r} \quad , \qquad (1)$$

where $P_i(v)$ is the number of pixels corresponding to the polygon i in the image obtained from the viewpoint v, r is the total number of pixels of the image (resolution of the image), and n is the total number of polygons of the scene. In this formula, $\lceil x \rceil$ denotes the smallest integer, greater than or equal to x. The first term in (1) gives the fraction of visible surfaces with respect to the total number of surfaces, while the second term is the ratio between the projected area of the

scene (or object) and the screen area (thus, its value is 1 for a closed scene).

From the definition of entropy (see Part II), the *viewpoint entropy* (VE) [VFSH01] has been defined from the relative area of the projected polygons over the sphere of directions centered at viewpoint v. Thus, the viewpoint entropy was defined by

$$H_v = -\sum_{i=0}^{N_f} \frac{a_i}{a_t} \log \frac{a_i}{a_t}, \tag{2}$$

where N_f is the number of polygons of the scene, a_i is the projected area of polygon i over the sphere, a_0 represents the projected area of background in open scenes, and $a_t = \sum_{i=0}^{N_f} a_i$ is the total area of the sphere. The maximum entropy is obtained when a certain viewpoint can see all the polygons with the same projected area. The best viewpoint is defined as the one that has maximum entropy. In molecular visualization, both maximum and minimum entropy views show relevant characteristics of a molecule [VFSL06].

A new viewpoint quality measure, called *viewpoint Kullback-Leibler distance* (VKL) [SPFG05], has been defined by

$$KL_v = \sum_{i=1}^{N_f} \frac{a_i}{a_t} \log \frac{\frac{a_i}{a_t}}{\frac{A_i}{A_T}}, \tag{3}$$

where a_i is the projected area of polygon i, $a_t = \sum_{i=1}^{N_f} a_i$, A_i is the actual area of polygon i and $A_T = \sum_{i=1}^{N_f} A_i$ is the total area of the scene or object. The VKL measure is interpreted as the distance between the normalized distribution of projected areas and the 'ideal' projection, given by the normalized distribution of the actual areas. In this case, the background can not be taken into account. The minimum value 0 is obtained when the normalized distribution of projected areas is equal to the normalized distribution of actual areas. Thus, to select views of high quality means to minimize KL_v.

Apart from the previous references on viewpoint quality measures, [PPB*05] describe a number of different ways to measure the goodness of a view of an object. After analyzing different view descriptors, they conclude that no single descriptor does a perfect job and possibly a combination of them would amplify the advantage that each one has. Given a sphere of viewpoints, [YSY*06] compute the similarity between each two disjoint views using Zernike moments analysis and obtain a similarity weighted spherical graph. A view is considered to be stable if all edges incident on its viewpoint in the spherical graph have high similarity weights. [AVF04] and [SPT06] present two different exploration algorithms guided by viewpoint entropy and the total curvature of a visible surface, respectively. In the volume rendering field, [BS05], [TFTN05] and [JS06] use an extended version of viewpoint entropy and [VFSG06] introduce the viewpoint mutual information. [CSCF07] use viewpoint entropy as a perceptual measure for mesh simplification.

Based on the investigation on canonical views, [GRMS01] present a new method for constructing images, where the viewpoint is chosen to be both off-axis and 'natural', and [LME06] obtain the viewing direction from the combination of factors such as saliency, occlusion, stability and familiarity. [LVJ05] have introduced the saliency as a measure for regional importance for graphics meshes and [KV06] presented a visual-saliency-based operator to enhance selected regions of a volume. [GCO06] introduced a method for partial matching of surfaces by using the abstraction of salient geometric features and a method to construct them.

3. Viewpoint Channel

In this section, we introduce an information channel between a set of viewpoints and the set of polygons of an object to deal with viewpoint selection. Then we define the viewpoint mutual information to select the most representative views of an object.

3.1. Viewpoint Mutual Information

Our viewpoint selection framework is constructed from an information channel $V \rightarrow O$ between the random variables V (input) and O (output), which represent, respectively, a set of viewpoints and the set of polygons of an object (see Figure 1(a)). This channel, which we call *viewpoint channel*, is defined by a conditional probability matrix obtained from the projected areas of polygons at each viewpoint. Viewpoints will be indexed by v and polygons by o. Throughout the development of this viewpoint framework, the capital letters V and O as arguments of $p()$ will be used to denote probability distributions. For instance, while $p(v)$ will denote the probability of a single viewpoint v, $p(V)$ will represent the input distribution of the set of viewpoints.

The viewpoint channel can be interpreted as an *observation channel* where the conditional probabilities represent the probability of *seeing* a determined polygon from a given viewpoint (see Figure 1(b)). The three basic elements of this channel are:

- Conditional probability matrix $p(O|V)$, where each element $p(o|v) = \frac{a_o}{a_t}$ is defined by the normalized projected area of polygon o over the sphere of directions centered at viewpoint v. Conditional probabilities fulfil $\sum_{o \in \mathcal{O}} p(o|v) = 1$. In our approach, background is not taken into account but it could be considered as another polygon.
- Input distribution $p(V)$, which represents the probability of selecting a viewpoint. In our experiments, $p(V)$ will be obtained from the normalization of the projected area of the object at each viewpoint. This can be interpreted as the

probability that a random ray originated at v hits (*sees*) the object. This assignation is consistent with the objective of selecting the viewpoints which *see* more projected area. Let us remember that this is a characteristic of a canonical view (see Section 1). The input distribution can also be interpreted as the *importance* assigned to each viewpoint v. For instance, the input distribution could also be defined by $p(v) = \frac{1}{N_v}$, where N_v is the number of viewpoints.

- Output distribution $p(O)$, defined by

$$p(o) = \sum_{v \in \mathcal{V}} p(v) p(o|v), \tag{4}$$

which represents the average projected area of polygon o, i.e., the probability of polygon o to be hit (*seen*) by a random ray cast from the viewpoint sphere.

From the previous definitions, the *conditional entropy* is given by the average of all viewpoint entropies:

$$\begin{aligned} H(O|V) &= -\sum_{v \in \mathcal{V}} p(v) \sum_{o \in \mathcal{O}} p(o|v) \log p(o|v) \\ &= \sum_{v \in \mathcal{V}} p(v) H(O|v), \end{aligned} \tag{5}$$

where $H(O|v) = -\sum_{o \in \mathcal{O}} p(o|v) \log p(o|v)$ is the *viewpoint entropy* H_v (2) and measures the degree of uniformity of the projected area distribution at viewpoint v. Let us observe that H_v has been now rewritten in a different form. Both entropies $H(O|v)$ and $H(O|V)$ tend to infinity when polygons are infinitely refined. This makes these measures very sensitive to the discretisation of the object and in general not appropriate to evaluate the quality of a viewpoint.

We now devote our attention to the *mutual information* between V and O, that expresses the degree of *dependence* or *correlation* between the set of viewpoints and the object. Mutual information is given by

$$\begin{aligned} I(V,O) &= \sum_{v \in \mathcal{V}} p(v) \sum_{o \in \mathcal{O}} p(o|v) \log \frac{p(o|v)}{p(o)} \\ &= \sum_{v \in \mathcal{V}} p(v) I(v,O), \end{aligned} \tag{6}$$

where we define

$$I(v,O) = \sum_{o \in \mathcal{O}} p(o|v) \log \frac{p(o|v)}{p(o)} \tag{7}$$

as the *viewpoint mutual information* (VMI), which gives us the degree of dependence between the viewpoint v and the set of polygons, and it is a measure of the *quality* of viewpoint v. Consequently, mutual information $I(V,O)$ can be interpreted as the average viewpoint quality. *Quality* is considered here equivalent to *representativeness*. It is also important to indicate that the level of resolution of the viewpoint sphere will determine the accuracy of the measures.

In our framework, the best viewpoint is defined as the one that has *minimum* VMI. High values of the measure mean a high dependence between viewpoint v and the object, indicating a highly *coupled* view (for instance, between the viewpoint and a small number of polygons with low average visibility). On the other hand, the lowest values correspond to the most *representative* or *relevant* views, showing the maximum possible number of polygons in a balanced way.

3.2. Discussion

Note that $I(v,O) = KL(p(O|v)|p(O))$, where $p(O|v)$ is the conditional probability distribution between v and the object and $p(O)$ is the marginal probability distribution of O, which in our case corresponds to the distribution of the average of projected areas. It is worth observing that $p(O)$ plays the role of the *target* distribution in the KL distance and also the role of the *optimal* distribution since our objective is that $p(O|v)$ becomes similar to $p(O)$ to obtain the best views. On the other hand, this role agrees with intuition since $p(O)$ is the average visibility of polygon o over all viewpoints, i.e., the *mixed distribution* of all views, and we can think of $p(O)$ as representing, with a single distribution, the knowledge about the scene. Note the difference between VMI (7) and VKL (3), due to the fact that in the last case the distance is taken with respect to the actual areas.

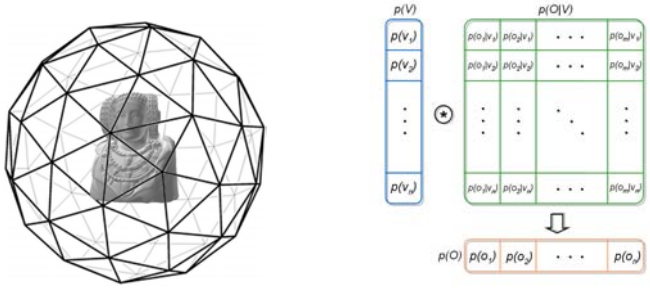
In [VFSG06], it has been shown that the main advantage of VMI over VE is its robustness to deal with any type of discretisation or resolution of the volumetric dataset. The same advantage can be observed for polygonal data. Thus, while a highly refined mesh will attract the attention of VE, VMI will be almost insensitive to changes in the mesh resolution. This behavior of both measures with respect to the discretization can be deduced from the mathematical analysis of VE and VMI. For instance, let us assume that a regular polygon o of the object is subdivided into two equal parts o_1 and o_2 such that $p(o_1|v) = p(o_2|v)$, $p(o_1) = p(o_2)$, $p(o|v) = p(o_1|v) + p(o_2|v)$ and $p(o) = p(o_1) + p(o_2)$. Assuming that only the term referred to polygon o can change in the formulas for VE (2) and VMI (7), we analyze their variation after the subdivision of o. The variation of VE is given by

$$\begin{aligned} \delta H(O|v) = -p(o_1|v) \log p(o_1|v) - p(o_2|v) \log p(o_2|v) - \\ (-p(o|v) \log p(o|v)) = p(o|v). \end{aligned}$$

Therefore, VE increases with a value $p(o|v)$ after the subdivision. On the other hand, the variation of VMI is given by

$$\begin{aligned} \delta I(v,O) = p(o_1|v) \log \frac{p(o_1|v)}{p(o_1)} + p(o_2|v) \log \frac{p(o_2|v)}{p(o_2)} - \\ p(o|v) \log \frac{p(o|v)}{p(o)} = 0. \end{aligned}$$

Thus, VMI remains invariant to the proposed subdivision. In general, if we compare both measures for different discretisations, mutual information will give similar results and VE will show an erratic behavior. Note that HM is also highly dependent on the discretisation, since the first term in (1) is



(*a*) Viewpoint sphere. (*b*) Probability distributions of channel $V \rightarrow O$.

Figure 1: *Viewpoint information channel.*

given by the quotient between the number of visible polygons and the total number of polygons. The behavior of all these measures with respect to the discretisation will be experimentally shown in the next section.

References

[AVF04] ANDÚJAR C., VÁZQUEZ P. P., FAIRÉN M.: Way-finder: guided tours through complex walthrough models. *Computer Graphics Forum (Eurographics 2004)* (2004).

[BET95] BÜLTHOFF H., EDELMAN S., TARR M.: How are three-dimensional objects represented in the brain? *Cerebral Cortex 5* (1995), 247–260.

[BS05] BORDOLOI U. D., SHEN H.-W.: Viewpoint evaluation for volume rendering. In *Visualization, IEEE 2005* (May 2005), pp. 62–62.

[BTB99] BLANZ V., TARR M., BÜLTHOFF H.: What object attributes determine canonical views? *Perception 28* (1999), 575–599.

[CSCF07] CASTELLÓ P., SBERT M., CHOVER M., FEIXAS M.: Viewpoint entropy-driven simplification. In *Proceedings of WSCG 2007* (2007).

[GBL02] GONZÁLEZ-BAÑOS H. H., LATOMBE J.-C.: Navigation strategies for exploring indoor environments. *I. J. Robotic Res. 21*, 10-11 (2002), 829–848.

[GCO06] GAL R., COHEN-OR D.: Salient geometric features for partial shape matching and similarity. *ACM Trans. Graph. 25*, 1 (2006), 130–150.

[GRMS01] GOOCH B., REINHARD E., MOULDING C., SHIRLEY P.: Artistic composition for image creation. In *Rendering Techniques* (2001), pp. 83–88.

[JS06] JI G., SHEN H.-W.: Dynamic view selection for time-varying volumes. *IEEE Trans. Vis. Comput. Graph. 12*, 5 (2006), 1109–1116.

[KV06] KIM Y., VARSHNEY A.: Saliency-guided enhancement for volume visualization. *IEEE Trans. Vis. Comput. Graph. 12*, 5 (2006), 925–932.

[LME06] LU A., MACIEJEWSKI R., EBERT D. S.: Volume composition using eye tracking data. In *Eurographics/ IEEE-VGTC Symposium on Visualization* (2006), Ertl T., Joy K., Santos B., (Eds.).

[LVJ05] LEE C. H., VARSHNEY A., JACOBS D. W.: Mesh saliency. *Computer Graphics (Proceedings of SIGGRAPH'05)* (July - August 2005). Held in Los Angeles, USA.

[PB96] PLEMENOS D., BENEYADA M.: Intelligent display techniques in scene modelling. new techniques to automatically compute good views. In *International Conference GraphiCon'96* (July 1996). Held in St. Petersburg, Russia.

[PPB*05] POLONSKY O., PATANÈ G., BIASOTTI S., GOTSMAN C., SPAGNUOLO M.: What's in an image? *The Visual Computer 21*, 8-10 (2005), 840–847.

[PRC81] PALMER S., ROSCH E., CHASE P.: Canonical perspective and the perception of objects. *Attention and Performance IX* (1981), 135–151.

[SPFG05] SBERT M., PLEMENOS D., FEIXAS M., GONZÁLEZ F.: Viewpoint quality: Measures and applications. In *Proceedings of 1st Computational Aesthetics in Graphics, Visualization and Imaging* (May 2005). Held in Girona, Spain.

[SPT06] SOKOLOV D., PLEMENOS D., TAMINE K.: Methods and data structures for virtual world exploration. *The Visual Computer 22*, 7 (2006), 506–516.

[TBZB97] TARR M., BÜLTHOFF H., ZABINSKI M., BLANZ V.: To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science 8*, 4 (1997), 282–289.

[TFTN05] TAKAHASHI S., FUJISHIRO I., TAKESHIMA Y., NISHITA T.: Locating optimal viewpoints for volume visualization. In *Visualization, IEEE 2005* (May 2005).

[VFSG06] VIOLA I., FEIXAS M., SBERT M., GRÖLLER M. E.: Importance-driven focus of attention. *IEEE Trans. Vis. Comput. Graph. 12*, 5 (2006), 933–940.

[VFSH01] VÁZQUEZ P. P., FEIXAS M., SBERT M., HEIDRICH W.: Viewpoint selection using viewpoint entropy. In *Proceedings of Vision, Modeling, and Visualization 2001* (Stuttgart, Germany, November 2001), Ertl T., Girod B., Greiner G., Niemann H., Seidel H.-P., (Eds.), pp. 273–280. Held in Stuttgart, Germany.

[VFSH03] VÁZQUEZ P. P., FEIXAS M., SBERT M., HEIDRICH W.: Automatic view selection using viewpoint entropy and its application to image-based modeling. *Computer Graphics Forum* (Desember 2003).

[VFSL06] VÁZQUEZ P. P., FEIXAS M., SBERT M., LLOBET A.: Realtime automatic selection of good molecular views. *Computers & Graphics 30*, 1 (February 2006), 98–110.

[YSY*06] YAMAUCHI H., SALEEM W., YOSHIZAWA S., KARNI Z., BELYAEV A. G., SEIDEL H.-P.: Towards stable and salient multi-view representation of 3d shapes. In *SMI* (2006), p. 40.

Applications of Information Theory to Computer Graphics Part VI: View Selection in Scientific Visualization

Ivan Viola

University of Bergen, Norway

1. Introduction

Concepts for optimal viewpoint estimation in computer graphics have been intensively studied in the past decade. These concepts (reviewed in Part V) accommodate information-theoretic measures into viewpoint characteristics such as viewpoint entropy, which is based on projected area and number of visible polygons. These accommodation of information-theory measures have been used for estimating the most *informative* viewpoint or a minimal set of viewpoints that characterize given polygonal scene in a most expressive way.

Viewpoint selection is of increasing interest for the scientific visualization research. Approaches for estimation of most informative viewpoints for scientific data sets are somehow similar to those, developed for polygonal data, however there are differences that arise due to different focus of scientific visualization. One difference, as opposed to polygonal graphics, is that the underlying data is larger and more complex. Scientific data is originating from measurements and simulations that have very heterogeneous resulting output. Some data types can be aligned to a structure as a grid, other types are missing such a structure and are represented as sparse points without any connectivity information. Medical data sets are often represented as scalar values per sample point. In other areas, such as meteorological or flow simulations, several attributes per sample point are usual. Many natural phenomena are studied through their development in time. These studies are represented through time-varying data sets, of which visualization is often effective way of providing insights. Moreover scientific data often differ in the level of semantics. Some data contain solely measurements without any a priori knowledge about structures, other data sets contain information about most relevant structures (e.g., critical points in flow data). Summarizing, the heterogeneity among scientific data types, as opposed to *simple* polygonal scene, is aligned with heterogeneity in visualization approaches of viewpoint quality computation.

Second aspect of scientific visualization, as opposed to standard computer graphics, is that it serves gaining insights of complex underlying data. There is always a purpose that drives visualization, usually in form of needs abstracted from industry, sciences, or medicine. Here visualization can serve two purposes, exploration of new, partly unknown data by way of visual analysis by a group of domain users and dissemination of knowledge gained during the exploration process towards another spectrum of audience. Computer-guided viewpoint selection seems to be promising tool for addressing both scenarios, knowledge gain as well as knowledge dissemination.

In this part of the tutorial we are reviewing state-of-the-art view selection strategies for volume visualization. We start with approaches where viewpoint selection for polygonal data has been adapted for view selection on density volumes. Next we review approaches for viewpoint quality measures for time-varying volumes, in terms of estimation of static informative viewpoints for the entire time span or a dynamic viewpoint showing whole time series by continuously changing the viewpoint position to show most of the information.

After describing how information-theoretic measures can be applied for unclassified volumetric data, we give more space to approaches that incorporate higher level semantics into viewpoint selection. In particular we focus on guided navigation technique denoted as *importance-driven focus of attention* where information-theoretic measures work on volumetric data organized into *objects* of varying importance. We give an outlook towards semantics-driven viewpoint selection and review early work in this promising direction.

2. Optimal Views for Volume Data

Techniques to measure *informativeness* of a viewpoint reflect how much of the overall scene is *visible* through particular viewpoint. These metrics differ according to infor-

mation provided in the data as well as according to the aim of visualization. In the following we will review view selection techniques where knowledge about the data is limited to voxel values and visual mapping specification.

2.1. Viewpoint Estimation for Interval Volumes

One of the first approaches on view selection for volume data has been done on evaluation of viewpoint entropy on extracted iso-surfaces or interval volumes respectively [TFTN05]. This approach represents a transition between surface-based visibility estimation working on polygons and purely voxel-based visibility estimation. The viewpoint quality measure here is viewpoint entropy (see Part V), applied to faces of iso-surfaces or interval volumes. These interval volumes or iso-surfaces represent set of features directly derived from the original data values that can be optionally weighted by average opacity specified by transfer function as importance distribution among features. Viewpoint quality considers feature arrangement in order to avoid feature overlapping.

This technique performs quite well during the visibility estimation phase and the viewpoint selection calculations are nearly real-time. This might be of considerable advantage, however, this approach does not compute visibilities of the volume but of surfaces surrounding volumes instead. Therefore very little portion of information contained in the volume is processed. The only way how the volume information is integrated into the viewpoint estimation process is the importance computed by the average opacity of the interval volume. Figure 1 shows best and worst views at set of interval volumes extracted from the hydrogen data set including bounding sphere plot indicating viewpoints quality.

2.2. View Selection for Static and Temporal Data

In parallel to the interval volumes approach discussed above, the viewpoint selection algorithm [BS05] that evaluates visibility of the volume directly, has been proposed. Viewpoint selection for volumes is an adaptation of the viewpoint entropy approach for polygonal data to compute the visibility from volume, i.e., voxels, by substituting the area visibility distribution of faces by the distribution obtained from the quotient between the voxel visibility and the voxel importance (*noteworthiness* factor). Importance distribution is defined per-voxel and is equal to the opacity value specified in the transfer function. This means more opaque voxels will get more prominence than the low opacity ones. This work has additionally suggested information-theoretic measures for clustering views according to similarity using the Jensen-Shannon divergence (see parts II and V). In addition, this work proposes scheme for estimation of static viewpoint on time-varying data. This is realized through conditional entropy where the different random variable distributions are obtained from neighboring time-steps.

This work proposes to perform visibility calculations directly on the volume and thus conveying more information about the volume as compared to the interval volume approach. In addition, the partitioning approach enables to capture entire scene by small set of snapshots. The viewpoint estimation is considering the temporal aspect in the evaluation instead of evaluating viewpoint for every time step separately. The importance distribution, however, is on the per-voxel level meaning that the data is not organized into higher semantics other than original data elements. Figure 2 shows tooth data set from several viewpoints that together capture the most of information about the scene and a static viewpoint on time-varying shockwave data set.

Previously mentioned approach on view selection for volumes has been recently extended to support dynamic viewpoint change for static and time-varying data [JS06] enabling guided flythrough over most interesting viewpoints. The new approach in addition modifies the voxel relevance function by incorporating shape characteristics and color in addition to opacity value.

3. View Selection Using Higher-Level Semantics

Above mentioned approaches are focusing on how to estimate good viewpoints driven by values in the data or assigned visual representation. They do not incorporate additional gained knowledge about the data and do not specifically address particular application scenario. In this section we will focus on view selection techniques that incorporate higher-level knowledge. Voxels or polygons are organized into higher-level semantics, i.e., *objects*. Sufficient knowledge about the data is gained beforehand in a user-steered classification process. View selection explicitly targets gained knowledge dissemination in form of guided navigation with enhancement of features in focus.

First we will discuss focus of attention approach based on varying importance distribution among objects. The second discussed approach targets guided navigation for knowledge dissemination for medical intervention planning.

3.1. Importance-Driven Focus of attention

Importance-driven focus of attention [VFSG06] provides a guided navigation through pre-classified features in the volumetric data set. Object of interest (focus) is directly specified by user. A characteristic viewpoint for this object is selected in combination with a visually pleasing discrimination of the focus from the remaining context information. By changing the object of interest, both viewpoint settings and visual parameters are smoothly modified to put emphasis on the newly selected object of interest. Characteristic viewpoints are estimated in an information-theoretic framework using viewpoint mutual information, a measure adapted from information theory (see part V). Both stages, i.e., the interactive focusing approach and the identification of character-

istic viewpoints, are controlled by an intuitive importance distribution among structures within the volumetric data.

Focus of attention needs visual discrimination of interesting objects from other elements in the scene. It is realized through a visual emphasis of the object of interest, while other objects presented as context are visually suppressed. Changes in opacity and color saturation are used to control the visual emphasis.

In addition to visual discrimination, objects in focus have to be shown from a good and characteristic view where most of the focus structures are perceivable. The most interesting object must not be occluded by less relevant parts. If possible the focus should be in front of other features. In case where the feature of interest is always occluded by other features, cut-away views or other concepts from illustration are be included into the visualization. Furthermore a proper orientation of the up-vector of the viewpoint and a proper positioning of the focus are important to consider in the viewpoint specification.

Finding a viewpoint where the characteristics of a specific feature are clearly visible naturally requires the visibility estimation of the feature under specific viewing settings. It requires ray casting of the whole data set from various viewpoints similarly to view selection for volumes discussed above (see Section 2.2). Visibility computation is based on the opacity contribution of each voxel and object visibility is computed as the sum of voxel visibilities belonging to the object. Voxel membership to a particular object is a fuzzy value, i.e., voxel may belong to several objects at the same time. Additionally two weights influence the visibility of an object, i.e., image-space weight and object-space weight. Image-space weight penalizes the visibility of objects when they are located outside the image center. Object-space weight assigns higher visibility to objects which are closer to the viewing plane and penalizes those that are more far away.

Object visibility is then mapped to a conditional probability of the object for a given viewpoint. These values are used for computation of good viewpoints for a given object by using information-theoretic framework based on viewpoint mutual information and combined with object importance information.

After selecting visual representations of objects and by identifying representative viewpoints, the crucial information to perform interactive focus of attention is available. The importance distribution is in the interactive part a direct mapping of user's interest. Importance is directly mapped to focus discrimination and level of ghosting in cut-away views. The viewpoint transformation is also controlled by importance distribution, smoothly changing to characteristic viewpoint of selected object.

Guided navigation drives user's focus to the object of interest, although still allowing to manipulate the viewpoint manually for custom inspection. In this scenario the object of interest is represented densely and clearly visible by using interactive cut-away views. After the manual viewpoint manipulation is finished, the viewpoint smoothly changes back to the characteristic viewpoint of the object in focus. User is then free to select another object to be focused onto. Focusing on a specific feature in the human hand dataset is shown in Figure 3. Viewpoint smoothly changes from characteristic viewpoint for the entire volume to viewpoint emphasizing the object of interest. As shown, parallel to the viewpoint change, the focus is discriminated from neighboring structures and the level of ghosting is continuously suppressing visual representation of occluding structures.

3.2. Semantics-Driven Guided Navigation

Viewpoint selection for intervention planning [KMP07] estimates visibility of objects from extracted iso-surfaces. Good viewpoints are estimated using many parameters with adjustable influence: object entropy, importance of occluders, size of unoccluded surface, preferred view region by surgeons, distance to viewpoint, and viewpoint stability. As example how tightly is viewpoint estimation bound to specific application is demonstrated by one of parameters: the distance to important feature defines importance of other features (e.g., neck muscles close to lymph node which is in focus). Furthermore guided navigation supports close zooming to the object of interest. Guided navigation through lymph nodes in the neck data set is shown in Figure 4.

Strong contribution of the domain knowledge parameters in the viewpoint selection computation shows that semantic-driven techniques outperform other existing automatic view selection techniques in a specific visualization scenario. This statement has been supported by evaluation of a user study that significantly included medical domain users.

4. Conclusion

Selection of a preferred viewpoint without targeting specific application is a very subjective choice. High knowledge about the data and the purpose of selecting particular viewpoint for a very specific application is moving viewpoint selection from subjective *taste* to more objective characteristics. Current trend in visualization research indicates that this is the direction for a good good viewpoint estimation. The work on viewpoint selection for intervention planning nicely demonstrates this trend.

References

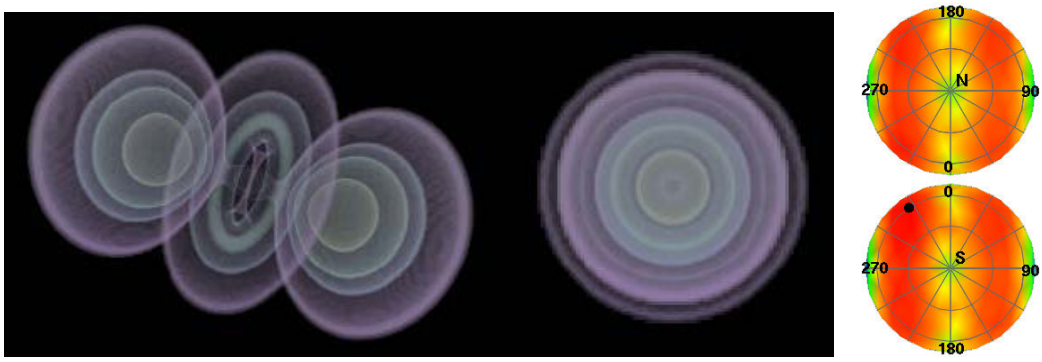
[BS05] BORDOLOI U. D., SHEN H.-W.: Viewpoint evaluation for volume rendering. In *Visualization, IEEE 2005* (May 2005), pp. 62–62.

[JS06] JI G., SHEN H.-W.: Dynamic view selection for time-varying volumes. *IEEE Trans. Vis. Comput. Graph. 12*, 5 (2006), 1109–1116.

[KMP07] KONRAD MÜHLER MATHIAS NEUGEBAUER C. T., PREIM B.: Viewpoint selection for intervention planning. In *Proceedingss of Eurographics/ IEEE-VGTC Symposium on Visualization* (2007), pp. 267–274.

[TFTN05] TAKAHASHI S., FUJISHIRO I., TAKESHIMA Y., NISHITA T.: Locating optimal viewpoints for volume visualization. In *Visualization, IEEE 2005* (May 2005).

[VFSG06] VIOLA I., FEIXAS M., SBERT M., GRÖLLER M. E.: Importance-driven focus of attention. *IEEE Trans. Vis. Comput. Graph. 12*, 5 (2006), 933–940.



(*a*) (*b*)

Figure 1: *Viewpoint estimation for interval volumes [TFTN05]: (a) best and worst views at interval volumes extracted from data set containing simulated electron density distribution in a hydrogen atom; (b) the sphere plot shows the viewpoint quality distribution where the green color encodes low quality viewpoints and the red encodes good and* informative *viewpoints. Best and worst views are indicated by a black and white dot respectively. Used by permission.*

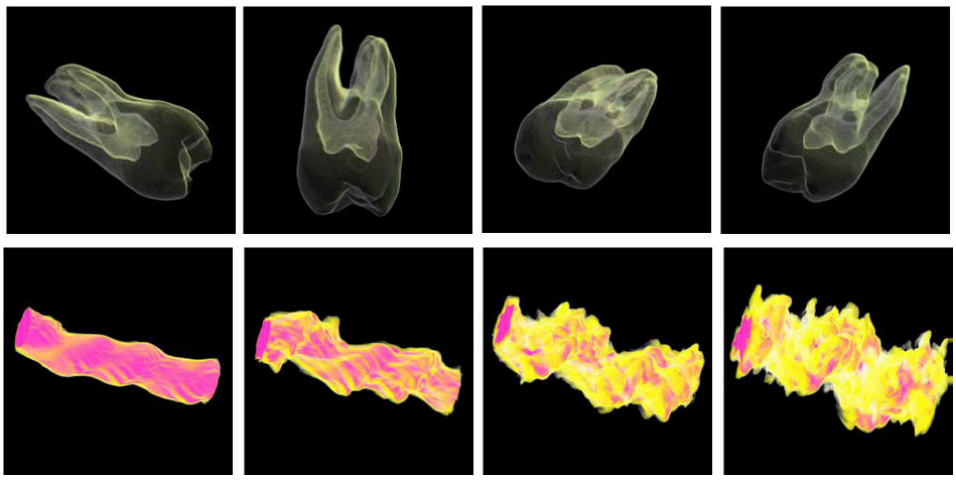


Figure 2: *View selection for static and time-varying volumes [BS05]: (upper row) four selected views on the tooth data set from four bounding sphere partitions; (bottom row) time-series of the shockwave data set from best temporal domain preserving viewpoint. Used by permission.*

Figure 3: *Importance-driven focusing [VFSG06]: Smooth guided navigation from overall characteristic viewpoint to focus on object of interest.*

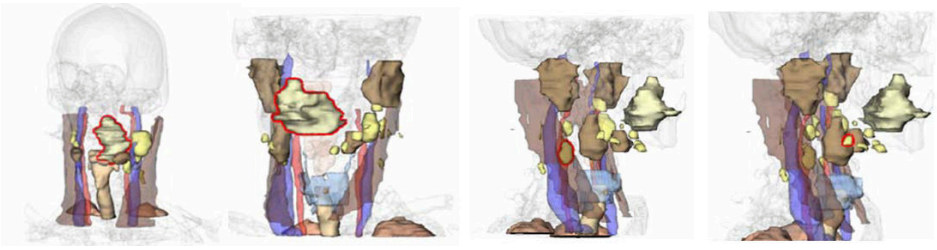


Figure 4: *Semantics-driven view selection [KMP07]: Guided navigation through features in the human neck data set assists studying the correspondence between focus objects, i.e., lymph nodes and surrounding tissue such as neck muscle. Used by permission.*

Applications of Information Theory to Computer Graphics Part VII: Viewpoint-driven Simplification

Pascual Castelló[*], Miguel Chover[*], Mateu Sbert[**], and Miquel Feixas[**]

[*] Jaume I University, Spain
[**] University of Girona, Spain

1. Introduction

Currently, polygonal models dominate interactive computer graphics. Polygons are the simplest primitive and lead to regular rendering algorithms that fit well in hardware. Unfortunately, the complexity of these models seems to grow faster than the ability of graphics hardware to render them interactively. Polygonal simplification offers one solution to this problem.

Most common polygonal simplification methods use some technique based on a geometric distance as a measure of quality between an original mesh and the one obtained from simplification. With these methods we can achieve meshes that are very similar to the original. In general, one of the most important advantages of geometry-oriented methods is their low temporal cost. This fact makes them suitable for scanned models, since these models are composed of thousands or even millions of polygons. In addition, geometric methods are very useful in applications that require exact geometric tolerances with regard to the original model. Examples of such applications include collision detection and path planning for part insertion and removal.

In contrast, image-based simplification methods carry out a simplification guided by differences between images more than by geometric distances. In other words, their goal is to create simplified meshes that appear similar according to visual criteria. These methods present a high temporal cost compared to geometric ones. The applications that can benefit from image-based methods are those in which the main requirement is visual similarity. Examples of such applications are video games, vehicle simulations and walkthroughs.

2. Recent work on simplification

The most important improvement in geometry-oriented simplification methods in recent years was the incorporation of mesh attributes such as color, normals and textures. For example, Hoppe extended his initial work [Hop96] by proposing a new quadric metric that includes colors and texture coordinates [Hop99], and the QSlim algorithm [GH97] was also extended with those attributes [GH98]. Cohen et al. [COM98] developed an algorithm based on edge collapses that samples the vertex position, normal and color attributes of the original mesh and then converts them to normal and texture maps. This algorithm is based on a texture deviation metric. Lately, a general method to incorporate texture information for edge collapse-based simplification algorithms has been proposed in [GCC07].

Lindstrom et al. [LT00] was the first to address the problem of *visual similarity* by developing a pure image-based metric. Basically, their method determines the cost of an edge collapse operation by rendering the model from several viewpoints. The algorithm compares the rendered images to the original ones and adds the mean-square error in luminance across all the pixels of all the images. Then, all edges are sorted by the total error induced in the images and after that the edge collapse that produces the least error is chosen. Lindstrom et al. used 20 viewpoints in their implementation to compute that error. The main advantage of this method is that the metric provides a natural way to balance the geometric and shading properties without requiring the user to perform an arbitrary weighting of them. On the other hand, its main disadvantage is the high temporal cost.

Karni et al. [KG00] proposed a metric to capture the visual difference between two approximations, the average of the norm of the geometric distance between models and the norm of the Laplacian difference. By introducing the Laplacian component, some visual properties the human eye appreciates such as smoothness are captured better.

Luebke et al. [LH01] presented a method to perform a view-dependent polygonal simplification using perceptual metrics. These metrics derive from a measure of low-level

perceptibility of visual stimuli in humans. Later on, Williams et al. [WLC*03] extended this work for lit and textured meshes.

Zhang et al. [ZT02] proposed a new algorithm that takes visibility into account. This approach defines a visibility function between the surfaces of a model and a surrounding sphere of cameras. The number of cameras increases both accuracy and calculation time. Zhang et al. used up to 258 cameras. In order to guide the simplification process, they combined their visibility measure with the quadric measure introduced by Garland et al. [GH97].

Recently, Lee et al. [LVJ05] introduced the idea of mesh saliency as a measure of regional importance for graphics meshes. This measure was incorporated into mesh simplification. Briefly, this approach consists of generating a saliency map, and then simplifying by using this map in the QSlim algorithm [ZT02]. The new edge collapse cost is that of the quadric multiplied by the saliency of this edge.

3. Information-theoretic metrics for viewpoint selection

Information-theoretic-based viewpoint selection metrics have been successfully applied in different areas of computer graphics, such as scene understanding, virtual exploration [VFSH01, Váz03] and volume visualization [BS05, TFTN05]. In this section, we review *viewpoint entropy* [VFSH01, Váz03] and *viewpoint Kullback-Leibler distance* [SPFG05] that have been used to compute the best viewpoints in a scene, and *viewpoint mutual information* that has been introduced to select the best views in volume rendering [VFSG06] and for polygonal meshes [FSG06].

3.0.1. Viewpoint Entropy

Viewpoint entropy, based on the Shannon entropy definition, was introduced in [VFSH01, Váz03] and can be used as a measure of the information provided by a point of view. The Shannon entropy equation (see Part II) is used as a basis to define viewpoint entropy. The relative area of the projected polygons over a sphere of directions centered in the viewpoint is taken as a probability distribution. Thus, given a viewpoint v, the entropy of v is defined as

$$H_v = -\sum_{i=0}^{N_f} \frac{a_i}{a_t} \log \frac{a_i}{a_t}, \tag{1}$$

where N_f is the number of polygons in the scene, a_i is an approximation on a plane of the projected area of polygon i over the sphere, a_0 represents the projected area of background in open scenes, and a_t is the total area of the sphere. In a closed scene, the whole sphere is covered by the projected polygons, and thus $a_0 = 0$. The maximum entropy is obtained when a certain point can see all the polygons with the same relative area. So, in an open scene, the maximum entropy is $\log(N_f + 1)$ and, in a closed scene it is equal to $\log N_f$. Among all the possible viewpoints, the best is the one that has maximum entropy, i.e. maximum information captured.

3.1. Viewpoint Mutual Information

In [VFSG06, FSG06], *viewpoint mutual information* was introduced to select the best views. An *information channel* $V \rightarrow O$, called a *viewpoint information channel*, between the random variables V and O was defined. This channel represents, respectively, a set of viewpoints and the set of polygons of an object. Viewpoints will be indexed by v and polygons by o. The marginal probability distribution of V is given by $p(v) = \frac{1}{N_v}$, where N_v is the number of viewpoints. That is, the same probability is assigned to each viewpoint, although other distributions could be used. The conditional probability $p(o|v) = \frac{a_o}{a_t}$ is defined by the normalized projected area of polygon o over the sphere of directions centered at viewpoint v. Finally, the marginal probability distribution of O is given by $p(o) = \sum_{v \in \mathcal{V}} p(v)p(o|v) = \frac{1}{N_v} \sum_{v \in \mathcal{V}} p(o|v)$.

From this channel, the *mutual information* (see Part II) between V and O, which expresses the degree of *dependence* or *correlation* between a set of viewpoints and the polygons of the object, is given by

$$\begin{aligned} I(V,O) &= \sum_{v \in \mathcal{V}} p(v) \sum_{o \in \mathcal{O}} p(o|v) \log \frac{p(o|v)}{p(o)} \\ &= \frac{1}{N_v} \sum_{v \in \mathcal{V}} I(v,O), \end{aligned} \tag{2}$$

where

$$I(v,O) = \sum_{o \in \mathcal{O}} p(o|v) \log \frac{p(o|v)}{p(o)}, \tag{3}$$

called *viewpoint mutual information* (VMI), represents the degree of correlation between the viewpoint v and the set of polygons O. High values of the measure mean a high degree of dependence between the viewpoint and a set of polygons, thus indicating a very *highly coupled* view. On the other hand, low values correspond to low dependence, which allows for more *representative* views of the object. Observe that $I(v,O) = KL(p(O|v)|p(O))$, where upper case indicates that $p(O|v)$ is the conditional probability distribution between v and the object and $p(O)$ is the marginal probability distribution of O.

3.2. Viewpoint Kullback-Leiber distance

In [SPFG05], a viewpoint quality measure based on the *Kullback-Leibler distance* (see Part II) was proposed. This measure minimizes the distance between the projected and actual area distributions of the polygons in the object and is defined by

$$KL_v = \sum_{i=1}^{N_f} \frac{a_i}{a_t} \log \frac{\frac{a_i}{a_t}}{\frac{A_i}{A_T}}, \tag{4}$$

where a_i is the projected area of polygon i, $a_t = \sum_{i=1}^{N_f} a_i$, A_i is the actual area of polygon i and $A_T = \sum_{i=1}^{N_f} A_i$ is the total area of the scene or object. In this case, the background is not taken into account.

4. Viewpoint-Based Error Metric

In this section, a new error metric based on viewpoint selection measures is presented. This metric can be used to evaluate the cost of a decimation operation. The edge collapse is chosen as the decimation operation, although any other simplification operation could be performed such as removing a vertex, replacing a cluster of vertices by a single one and contracting an edge.

Viewpoint selection measures express the accessible information about an object from a particular viewpoint. Given a particular viewpoint, we can consider the following: if the simplification is produced near the silhouette, probably it will change the shape of the object. Therefore, if the goal is to preserve the silhouette of the model we should try to reduce this change. In addition, in order to preserve the global appearance of the model, several equidistant viewpoints surrounding the model are required, so that the whole model will be fully covered. This distribution guaranties a uniform simplification.

Taking into account the facts mentioned above, the variation of a viewpoint selection measure for each viewpoint can provide us with an error metric to guide the simplification process. Thus the *simplification error deviation* for an edge collapse e from all viewpoints V is defined by:

$$C_e = \sum_{v \in \mathcal{V}} |I_v - I'_v|, \tag{5}$$

where I_v represents the viewpoint selection measure before the edge collapse e and I'_v afterwards.

4.1. Discussion

We choose some information-theoretic viewpoint selection measures to test our simplification method: *viewpoint entropy* (H_v), *viewpoint mutual information* (VMI) and *viewpoint Kullback-Leibler distance* (KL_v). These measures express the accessible information about the object from a given viewpoint as mentioned at the beginning of this section. However, viewpoint mutual information presents an important feature, it can be used as a shape descriptor for object recognition [RFS05], which is suitable for capturing the shape variation.

Both viewpoint entropy and mutual information are based on the distribution of areas of polygons seen from a viewpoint. The area of the background is also included as the polygon 0. These facts allow viewpoint entropy and mutual information to preserve the silhouette better. But maybe the main implication of considering the projected areas is that the hidden geometry will be initially removed, because if a polygon is not seen from any point of view, it will not introduce error.

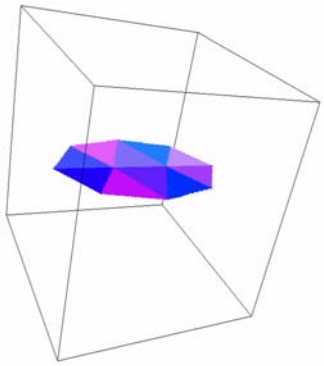


Figure 1: *Example of 8 camera positions surrounding the Test Model. These camera positions correspond to the 8 vertices of the cube. This allows covering the whole object because the viewpoints are equidistant from each other.*

Figure 1 shows the original Test model and how the viewpoints are distributed around it. These viewpoints are associated with the vertices of the Cube in which the object is inscribed. Figure 2 shows the VMI for the original Test model using the 8 viewpoints shown in Figure 1. As can be seen, the different viewpoints have the same VMI. This is because the object is wholly seen from every viewpoint and each viewpoint sees the same as any other. Note that this is a very special situation because the object is quite simple and regular. Normally, in more complex models every viewpoint will have a different VMI.

Figures 3 and 4 illustrate how VMI can be employed to conduct the simplification. Figure 3 shows the Test model after performing the best edge collapse e and Figure 4 after performing the worst edge collapse e'. The best edge collapse belongs to the lowest simplification error C_e (5) and the worst to the highest. As can be observed all the VMI values for every viewpoint decreased after an edge collapse (see for instance Figure 2(a) compared to Figure 3(a) or 4(a)). This is because the visible area did not increase in both cases and also the complexity is always reduced during the simplification process. But in a more general case, it is possible that after an edge collapse some previously hidden parts of the mesh may now appear, thus increasing the visible area. If we pay attention to Figure 2(b) and compare this same viewpoint after the best edge collapse (see Figure 3(b)), it can be appreciated that although the number of triangles is reduced (T=8), the visible area remains the same. The simplification error for this viewpoint using VMI is C_e=0.004097-0.003651=0.000446. If we analyze the same

(a) (b) (c) (d)

Figure 2: *Original Test model. T=10. $I(v,O)$=0.004097 where $v = \{1,..,8\}$. Only 4 viewpoints are shown because the rest are symmetric.*

(a) I(0,O)=0.003830 (b) I(1,O)=0.003651 (c) I(2,O)=0.003651 (d) I(3,O)=0.003830

Figure 3: *Test model after performing the best edge collapse e using VMI. T=8. C_e=0.002573.*

viewpoint in the worst edge collapse operation (see Figure 4(b)), it can be seen that although the number of triangles is less reduced, the total visible area is somewhat decreased. The simplification error for this viewpoint is $C_{e'}$=0.004097-0.003372=0.000725, which is higher than the error committed in the best edge collapse.

Nevertheless, due to the fact that KL_v considers the actual area of polygons, after an edge collapse, normally one or two polygons will be removed, thus decreasing the total actual area. This will change the value for KL_v after an edge collapse. Therefore the error committed will be distinct from zero. The consequence is that even hidden polygons will have error when simplifying and will not be completely removed during the initial steps of the algorithm. Hidden polygons will be removed according their actual area. Thus, the smallest polygons will be simplified before, preserving the main features of the object in its internal parts.

5. Simplification algorithm

The simplification process, like many other simplification algorithms, is based on the edge collapse operation. However, we use the *half-edge collapse* operation. According to this, the remaining vertex for an edge collapse $e(u,v)$ is vertex u or v (see Figure 5). By using half-edge collapses it is possible to reuse the simplification process in order to generate multiresolution models. These models can use the current hardware in a more efficient way because no new vertices are added to the original model. Furthermore, the half-edge representation is useful for progressive transmission. The main disadvantage is a slight loss of quality of the final mesh, although the complexity of the simplification algorithm is reduced because we do not have to compute the position of the new vertex v' resulting from the edge collapse. In any case, the general edge collapse operation can be applied to our algorithm. However, a strategy is required to compute the position of the resulting vertex.

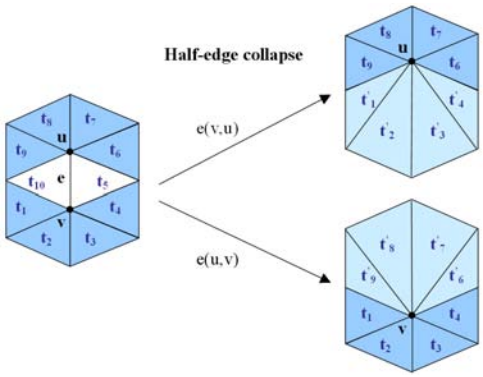


Figure 5: *The half-edge collapse operation. In this example edge e is collapsed into vertex u (see $e(v,u)$), but it is also collapsed into v (see $e(u,v)$). The triangles t_{10} and t_5 are removed.*

Brute force selection of edges can introduce mesh inconsistencies. In order to avoid these artifacts, we only take into account the edges which have at most two adjacent polygons, that is 2-manifold edges. And we also consider bound-

(a) I(0,O)=0.003238 (b) I(1,O)=0.003372 (c) I(2,O)=0.003372 (d) I(3,O)=0.003238

Figure 4: *Test model after performing the worst edge collapse e' using VMI. T=9. $C_{e'}$=0.006228.*

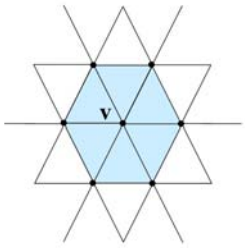


Figure 6: *Edges adjacent to vertices adjacent to vertex v.*

ary edges, i.e. edges which have one single adjacent polygon.

The best half-edge collapse is the decimation operation chosen in our algorithm. Note that the cost of collapsing vertex u to v may be different than the cost of collapsing v to u. In our strategy in order to determine the best orientation of an edge collapse, we would have to render the two possibilities and compute that error. However, this would increase considerably the number of renderings and consequently the number of framebuffer readings. Therefore the temporal cost would be penalized. To avoid that, we used the approach developed by Melax [Mel98] that takes into account polygon normals. Within this approach, the two orientations $e(u,v)$ and $e(v,u)$ are calculated and finally the orientation that produces a minor change in the curvature of the local region around the edge collapse is applied. Hence the simplification error deviation is only computed for that orientation. In Figure 7 we show a summary of the simplification algorithm.

At each iteration the edge cost has to be evaluated for the entire set of remaining edges. An edge collapse in our algorithm could, in principle, affect the cost of any remaining edge. But this does not always happen to each edge. At each step we only choose a small group of edges that are affected by an edge collapse and then the cost is recalculated for these edges. These edges are the ones that are adjacent to the vertices adjacent to the vertex v resulting from a half-edge collapse (see Figure 6).

```
// Compute initial viewpoint selection
// measure for mesh M
Compute I_v, where v = {1,..,n}

// Build initial heap of edge collapses
for (e ∈ M )
   Choose the best orientation of e
   Perform collapse e
   Compute I_v, where v = {1,..,n}
   Compute collapse cost C_e = Σ_{v=1}^{n} |I_v − I'_v|
   Insert the duple (e,C_e) in heap h
   Undo collapse e
end for

// Update mesh M
while (heap h not empty)
   Delete from heap h edge e with lowest C_e
   Perform collapse e
   Recalculate cost for the neighborhood of e
   and update their location in heap h
end while
```

Figure 7: *Pseudo-code of the viewpoint-based simplification algorithm.*

6. Experiments

We carried out our tests with low complexity models from CAD programs. All models were simplified on a Pentium Xeon 2GHz with 1GB RAM and an NVIDIA 7800 GTX 512MB graphics card from 20 viewpoints. The results obtained with the viewpoint-driven simplification method were compared to the results with QSlim v2.0 [GH97], using the best half-edge collapse, at the same level of simplification. We chose QSlim because it is a well-known pure geometric-based algorithm, freely available, and produces high quality simplifications. The images shown were obtained using different viewpoints from those used during the simplification process.

We have implemented the root mean square error (*RMSE*) of the pixel-to-pixel image difference defined in [LT00] to measure the mean visual error between the original and the simplified model. This error was taken using 24 viewpoints

and 512x512 resolution images. We must emphasize that each viewpoint was different from the one used during the simplification and the resolution was higher.

We measured the geometric error using the mesh comparison tool called *Metro* v4.06 [CRS98]. This tool measures the Hausdorff distance between two meshes.

6.0.1. Experiments for viewpoint entropy

Model	Triangles		RMSE		Metro	
	Original	Final	H_v	QSlim	H_v	QSlim
Fish	815	100	11.40	22.83	0.05	0.09
Galleon	4698	500	17.74	36.84	0.11	0.22
Fracttree	4806	1200	30.19	34.10	0.08	0.12
Galo	6592	500	9.03	12.40	0.03	0.05
Octopus	8468	500	17.35	25.84	0.09	0.16
Porsche	10 474	1000	7.48	8.28	0.04	0.10
Unicycle	13 810	1000	10.32	11.06	0.03	0.07

Table 1: *Errors measured for all models.*

Table 1 presents the error committed in our experiments using H_v. It analyzes the visual error and the geometric error. Our results are rather better than the geometric method used for comparison purposes. This makes us highly confident about our approach. For example, the geometric error committed in the Galleon, Porsche and Unicycle models using H_v is 50% less than with QSlim.

Figure 8 shows the results for some models analyzed. The H_v achieves much better simplification than QSlim. For example, in the Fish model the tail and the mouth shape is kept better, and in the Galleon model the same can be said for the sails and the masts.

Figure 9 shows how H_v acts at several degrees of simplification for the Galleon model. We have measured the RMSE and the geometric error. As shown in 9(a), if the level of simplification is increased the difference between H_v and QSlim becomes larger and the visual quality of H_v is even much higher. The geometric error of H_v is also lower than QSlim, except during the very first stages as can be observed in 9(b). This can be accounted for by the fact that H_v is a global measure, and it is possible that, in these stages, QSlim could often be better because it evaluates the error locally.

6.1. Experiments for viewpoint mutual information

First of all, in this section, we perform a comparison between viewpoint entropy and mutual information for all models, the results of which appear in Table 2. As shown in this table the visual error is lower in VMI. The temporal cost is almost the same. This difference between the VMI and the H_v lies in the calculation performed to obtain the mean projected area of the polygons. In fact, this is not necessary in H_v. Therefore the temporal cost of H_v is just a bit lower.

Table 3 depicts the visual and geometric error introduced in our experiments and the simplification time. Clearly, the visual error committed in our method is quite low compare to QSlim, and can even be 50% lower, as shown in case of the Shark and the Elephant models.

Our results for the geometric error are rather better than the geometric method used for comparison purposes. This makes us highly confident about our approach. For example, the geometric error committed in the Galo and Greekship models using VMI is 50% less than with QSlim and even 75% in the Shark model. However, it is possible that in particular models this error will be slightly higher than QSlim. The reason for this is because the model has lots of hidden interiors and these are completely removed, thus increasing the geometric error but not the visual one.

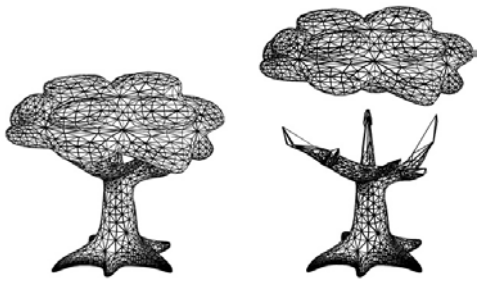
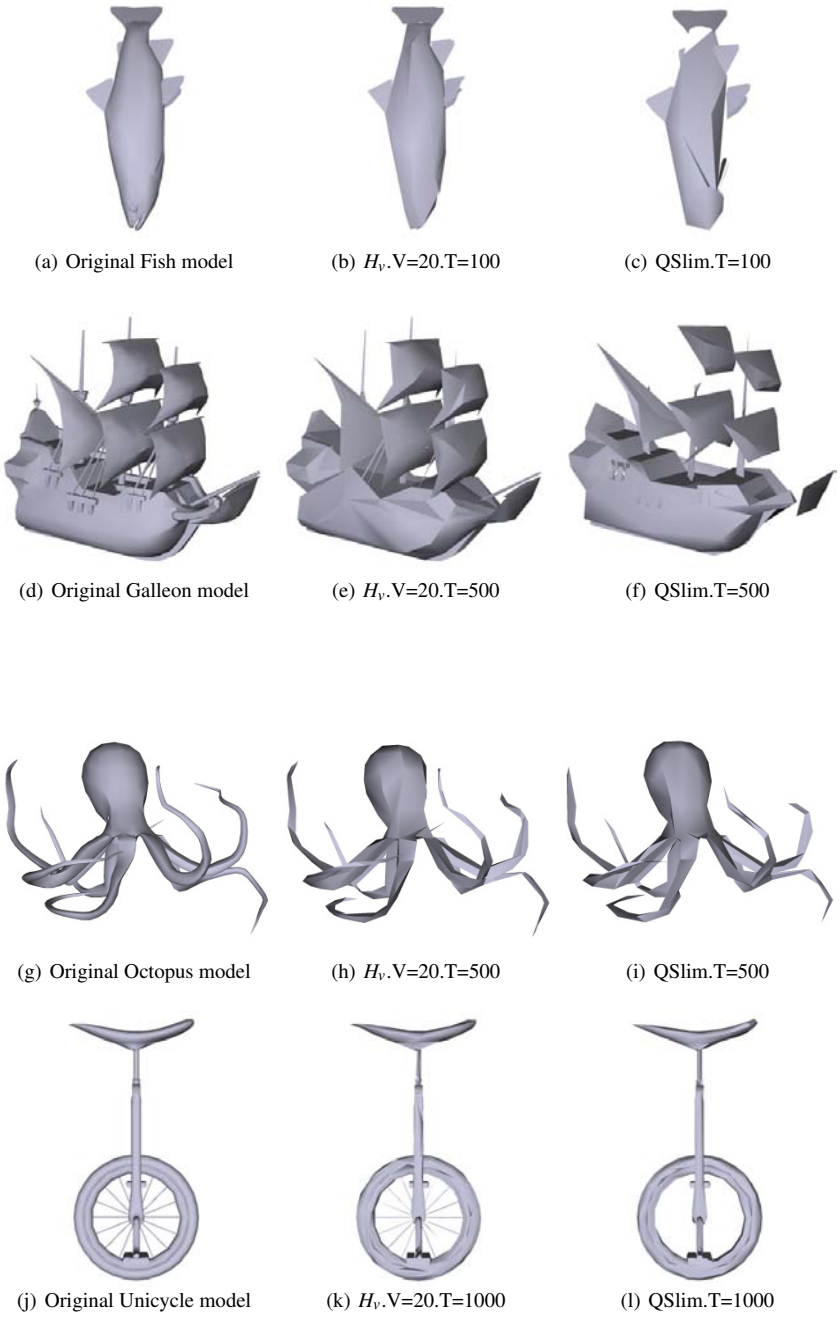


Figure 14: *VMI operating at the first stages of simplification (T=7400) for the Simpletree model. VMI is able to remove all the hidden interiors.*

An analysis of the temporal cost is also shown in this table. This cost is proportional to the complexity of the model and the final number of triangles demanded. However, the QSlim algorithm is extremely fast. Its times for these models are less than a second. In any case, our method produces high quality simplifications according to visual similarity. So it is valid, despite the high temporal cost compared to geometry-based methods.

Figure 10 shows the results for the Shark model. VMI achieves much better simplification than QSlim. The fins, the head and the tail are kept better in VMI than QSlim. Figure 11 shows the Galo model. The tail and the crest are kept better in VMI. Figure 12 shows the Greekship model. The oars are kept much better in VMI, as well as the rope that comes from the mast. In QSlim it is removed completely. Lastly, Figure 13 shows the results for the Elephant model. VMI keeps the shape of the ears and the tusks much better than QSlim. In summary, VMI attains better simplification than the geometric method QSlim. The difference between VMI and QSlim is bigger if the model presents lots of hidden interiors, in which case VMI can accomplish much better simplifications.

Figure 14 shows how VMI works at very early simplification levels. In this case we analyze the Simpletree model since it presents hidden interiors in the branches which are in contact with the tree top. The Simpletree model was simplified by around 66% (T=7400). As shown in this figure, VMI



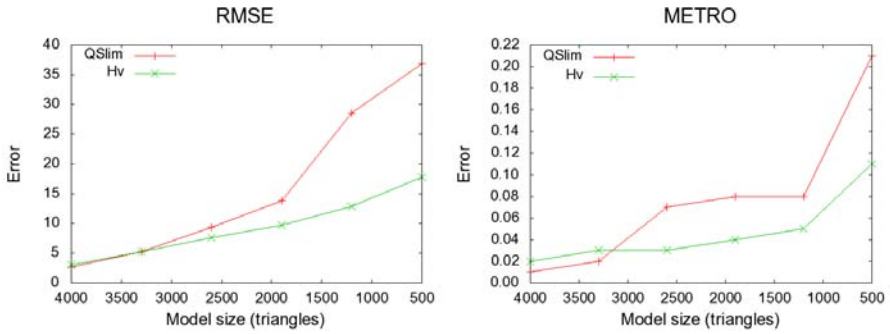
(a) Original Fish model (b) H_v.V=20.T=100 (c) QSlim.T=100

(d) Original Galleon model (e) H_v.V=20.T=500 (f) QSlim.T=500

(g) Original Octopus model (h) H_v.V=20.T=500 (i) QSlim.T=500

(j) Original Unicycle model (k) H_v.V=20.T=1000 (l) QSlim.T=1000

Figure 8: *Results for all models. V indicates the number of viewpoints and T the number of triangles.*



(a) (b)

Figure 9: *Errors measured for the Galleon model at different levels of simplification using H_v.*

(a) Original model (b) QSlim.T=80 (c) VMI.T=80

Figure 10: *Results for the Shark model. Image (a) shows the original model (T=734), (b) the model simplified with QSlim and (c) with VMI.*

(a) Original model (b) QSlim.T=500 (c) VMI.T=500

Figure 11: *Results for the Galo model. Image (a) shows the original model (T=6592), (b) the model simplified with QSlim and (c) with VMI.*

(a) Original model (b) QSlim.T=600 (c) VMI.T=600

Figure 12: *Results for the Greekship model. Image (a) shows the original model (T=9510), (b) the model simplified with QSlim and (c) with VMI.*

Model	Triangles		RMSE		Time	
	Original	Final	H_v	VMI	H_v	VMI
Shark	734	80	14.78	14.65	10.24	10.23
Galo	6592	500	9.05	8.38	141.75	142.24
Greekship	9510	600	13.37	12.85	241.78	246.72
Simpletree	11 136	600	17.23	16.60	321.06	332.49
Hammer	13 380	500	8.13	7.43	404.33	423.05
Elephant	31 548	900	13.75	11.60	2197.67	2309.79

Table 2: *Results for H_v and VMI measuring visual error (RMSE) and simplification time.*

Model	Triangles		RMSE		Metro		Time	
	Original	Final	QSlim	VMI	QSlim	VMI	QSlim	VMI
Shark	734	80	33.41	14.65	0.20	0.04	0.02	10.20
Galo	6592	500	12.40	8.38	0.05	0.01	0.08	142.24
Greekship	9510	600	17.20	12.85	0.21	0.09	0.11	246.72
Simpletree	11 136	600	20.73	16.60	0.11	0.13	0.20	332.49
Hammer	13 380	500	8.99	7.43	0.03	0.04	0.20	423.05
Elephant	31 548	900	25.32	11.60	0.08	0.03	0.52	2309.79

Table 3: *Results for QSlim and VMI measuring visual error (RMSE), geometric error (Metro) and simplification time.*

accomplishes a great level of simplification in this region. At this level, most simplifications focused on hidden interiors.

6.2. Experiments for viewpoint Kullback-Leibler distance

Table 4 depicts the visual and geometric error using KL_v. Clearly, the visual error committed with KL_v is quite low compare to QSlim, and can even be 50% lower, as shown in case of the Fish and the Elephant model. However, the visual error is slightly improved in H_v. This is due to the fact that H_v removes completely the hidden interiors and non-visible regions of the model increasing the visual quality.

The geometric error for KL_v is better in all the models than the QSlim method and H_v. For example, the geometric error committed in the Galo, the Simpletree and the Big_atc models using KL_v is at least 75% less than with QSlim. Figure 19 show the Al model rendered with transparency. This model has lots of hidden interiors. For instance, the hidden joints of the arms and the hip are preserved better in KL_v (see Figure 19(d)) whereas in H_v (see Figure 19(c)) these interpenetrating parts are partially removed. Therefore, in summary, the approximations produced with KL_v preserve the interior regions better than H_v and consequently present a lower geometric error.

As shown in Table 5 the temporal cost is proportional to the complexity of the model and the desired number of triangles. However, the QSlim algorithm is extremely fast. Its times for these models are less than a second. H_v is faster than KL_v. The reason is that H_v does not have to compute the actual areas for the polygons and its initial value can be updated at every simplification step because the total projected area is the image resolution. KL_v considers the total actual area and this area changes at every simplification. Accordingly, KL_v must be recomputed. In any case, KL_v is valid, despite the high temporal cost because it improves the geometric error without decreasing significantly the visual quality. This can be useful in many applications as mentioned earlier.

Figure 15 shows the results for the Shark model. The fins, the head and the tail are kept better in KL_v and H_v than QSlim. Figure 16 shows the Simpletree model. The roots at the base of the trunk are kept much better in KL_v and H_v and even both are able to keep the tree top better than QSlim. To sum up, KL_v and H_v attain better simplification than the quadric method QSlim. The difference between KL_v and H_v is noticeable if the model presents hidden interiors. In such case H_v can accomplish a slightly better visual simplifications at the expense of increasing the geometric error, since it may remove completely those hidden interiors (see Figure 19).

Some experiments with more viewpoints for some of our test models were conducted as shown in Figure 18. These different viewpoint configurations correspond to the vertices of Platonic solids. This guarantees that the viewpoints are distributed uniformly. The last configuration (42 viewpoints) was obtained by subdivision. The visual and the geometric error is analyzed in this figure together with the temporal cost. Both visual and geometric errors hardly improve when the number of cameras increases from 20 to 42. Nevertheless, the temporal cost is about twice as high (see Figure 18(c)). Therefore, we believe that 20 viewpoints is a good compromise between quality and efficiency. The two other measures tested in this section (H_v and VMI) have the same behavior when different number of viewpoints are considered.

References

[BS05] BORDOLOI U., SHEN H.-W.: View selection for volume rendering. In *IEEE Visualization* (2005), p. 62.

[COM98] COHEN J., OLANO M., MANOCHA D.: Appearance-preserving simplification. In *SIGGRAPH '98: Proceedings of the 25th annual conference on Com-*

(a) Original model (b) QSlim.T=900 (c) VMI.T=900

Figure 13: *Results for the Elephant model. Image (a) shows the original model (T=31 548), (b) the model simplified with QSlim and (c) with VMI.*

(a) Original model (b) QSlim.T=80 (c) H_v.T=80 (d) KL_v.T=80

Figure 15: *Results for the Shark model. Image (a) shows the original model ($T = 734$), (b) the model simplified with QSlim, (c) with H_v and d) with KL_v.*

(a) Original model (b) QSlim.T=1000 (c) H_v.T=1000 (d) KL_v.T=1000

Figure 16: *Results for the Big_atc model. Image (a) shows the original model ($T = 13594$), (b) the model simplified with QSlim, (c) with H_v and (d) with KL_v.*

(a) Original model (b) QSlim.T=600 (c) H_v.T=600 (d) KL_v.T=600

Figure 17: *Results for the Simpletree model. Image (a) shows the original model ($T = 11136$), (b) the model simplified with QSlim, (c) with H_v and (d) with KL_v.*

Model	Triangles		RMSE			Metro		
	Original	Final	QSlim	H_v	KL_v	QSlim	H_v	KL_v
Fish	815	100	22.83	11.57	12.98	0.09	0.03	0.03
Galo	6592	500	12.40	9.34	10.48	0.05	0.03	0.01
Al	7124	1000	17.66	11.47	12.07	0.03	0.08	0.03
Simpletree	11 136	600	20.73	16.98	18.04	0.11	0.13	0.04
Big_atc	13 594	1000	16.50	15.97	15.44	0.08	0.05	0.03
Elephant	31 548	900	25.32	13.18	13.40	0.08	0.14	0.05

Table 4: *Errors for all models using QSlim, KL_v and H_v.*

Model	Triangles		Time		
	Original	Final	QSlim	H_v	KL_v
Fish	815	100	0.03	10.01	11.31
Galo	6592	500	0.08	141.75	237.30
Al	7124	1000	0.08	150.90	273.18
Simpletree	11 136	600	0.20	332.49	605.49
Big_atc	13 594	1000	0.27	535.23	835.88
Elephant	31 548	900	0.52	2197.67	4016.78

Table 5: *Simplification time (seconds) for all models using QSlim, KL_v and H_v.*

puter graphics and interactive techniques (New York, NY, USA, 1998), ACM Press, pp. 115–122.

[CRS98] CIGNONI P., ROCCHINI C., SCOPIGNO R.: Metro: Measuring error on simplified surfaces. *Computer Graphics Forum 17*, 2 (1998), 167–174.

[FSG06] FEIXAS M., SBERT M., GONZÁLEZ F.: *A Unified Information-Theoretic Framework for Viewpoint Selection and Mesh Saliency*. Tech. Rep. IIiA-06-06-RR, Institut d'Informàtica i Aplicacions, Universitat de Girona (Girona, Spain), 2006.

[GCC07] GONZÁLEZ C., CASTELLÓ P., CHOVER M.: A texture-based metric extension for simplification methods. In *Proc. of GRAPP 2007, Barcelona, Spain* (2007), pp. 69–77.

[GH97] GARLAND M., HECKBERT P.: Surface simplification using quadric error metrics. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (1997), ACM Press/Addison-Wesley Publishing Co., pp. 209–216.

[GH98] GARLAND M., HECKBERT P. S.: Simplifying surfaces with color and texture using quadric error metrics. In *VIS '98: Proceedings of the conference on Visualization 1998* (Los Alamitos, CA, USA, 1998), IEEE Computer Society Press, pp. 263–269.

[Hop96] HOPPE H.: Progressive meshes. *Proceedings of SIGGRAPH 96* (August 1996), 99–108. ISBN 0-201-94800-1. Held in New Orleans, Louisiana.

[Hop99] HOPPE H.: New quadric metric for simplifying meshes with appearance attributes. In *VIS '99: Proceedings of the 10th IEEE Visualization 1999 Conference* (Washington, DC, USA, 1999), IEEE Computer Society.

[KG00] KARNI Z., GOTSMAN C.: Spectral compression of mesh geometry. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 2000), ACM Press/Addison-Wesley Publishing Co., pp. 279–286.

[LH01] LUEBKE D. P., HALLEN B.: Perceptually-driven simplification for interactive rendering. In *Proceedings of the 12th Eurographics Workshop on Rendering Techniques* (London, UK, 2001), Springer-Verlag, pp. 223–234.

[LT00] LINDSTROM P., TURK G.: Image-driven simplification. *ACM Transaction Graphics 19*, 3 (2000), 204–241.

[LVJ05] LEE C. H., VARSHNEY A., JACOBS D. W.: Mesh saliency. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers* (New York, NY, USA, 2005), ACM Press, pp. 659–666.

[Mel98] MELAX S.: A simple, fast, and effective polygon reduction algorithm. *Game Developer* (Nov. 1998), 44–48.

[RFS05] RIGAU J., FEIXAS M., SBERT M.: Shape complexity based on mutual information. In *SMI '05: Proceedings of the International Conference on Shape Modeling and Applications* (Washington, DC, USA, 2005), IEEE Computer Society, pp. 357–362.

[SPFG05] SBERT M., PLEMENOS D., FEIXAS M., GONZÁLEZ F.: Viewpoint quality: Measures and applications. In *Computational Aesthetics in Graphics, Visualization and Imaging* (2005).

[TFTN05] TAKAHASHI S., FUJISHIRO I., TAKESHIMA Y., NISHITA T.: A feature-driven approach to locating optimal viewpoints for volume visualization. In *IEEE Visualization* (2005), p. 63.

[Váz03] VÁZQUEZ P. P.: *On the Selection of Good Views and its Application to Computer Graphics*. PhD thesis, Universitat Politècnica de Catalunya, April 2003.

[VFSG06] VIOLA I., FEIXAS M., SBERT M., GRÖLLER M. E.: Importance-driven focus of attention. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (Oct. 2006), 933–940.

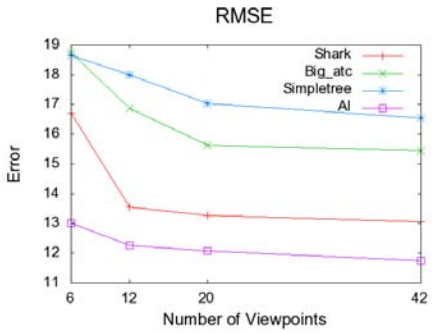
(a) Original model (b) QSlim.T=1000 (c) H_v.T=1000 (d) KL_v.T=1000

Figure 19: *The Al model rendered with transparency. Image (a) shows the original model, (b) the model simplified with QSlim, (c) with H_v and (d) with KL_v.*

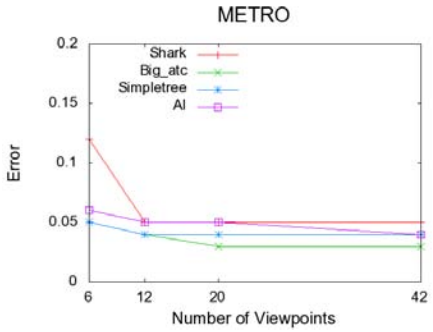
[VFSH01] VÁZQUEZ P. P., FEIXAS M., SBERT M., HEIDRICH W.: Viewpoint selection using viewpoint entropy. In *VMV '01: Proceedings of the Vision Modeling and Visualization Conference* (2001), Aka GmbH, pp. 273–280.

[WLC*03] WILLIAMS N., LUEBKE D., COHEN J. D., KELLEY M., SCHUBERT B.: Perceptually guided simplification of lit, textured meshes. In *SI3D '03: Proceedings of the 2003 symposium on Interactive 3D graphics* (New York, NY, USA, 2003), ACM Press, pp. 113–121.

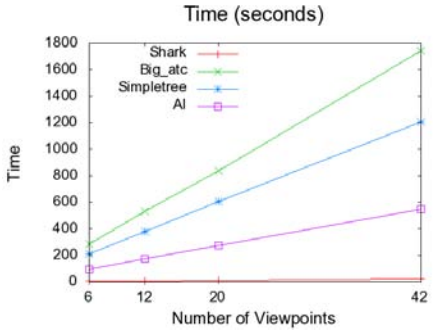
[ZT02] ZHANG E., TURK G.: Visibility-guided simplification. In *VIS '02: Proceedings of the conference on Visualization 2002* (Washington, DC, USA, 2002), IEEE Computer Society, pp. 267–274.



(a)



(b)



(c)

Figure 18: *Errors and times measured for some models simplified with KL_v using different number of viewpoints.*

Applications of Information Theory to Computer Graphics

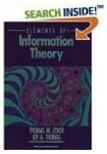
Mateu Sbert
Miquel Feixas
Jaume Rigau

Ivan Viola

Miguel Chover

07 EUROGRAPHICS Prague Czech Republic

(2) Information Theory Basics



Miquel Feixas
Jaume Rigau
Mateu Sbert

07 EUROGRAPHICS Prague Czech Republic

Information Theory

- Claude Elwood **Shannon**, 1916-2001
- **A mathematical theory of communication**. The Bell System Technical Journal, July, October 1948
- Transmission, storage and processing of information
- Applications:
  - Physics, computer science, mathematics, statistics, economics, biology, linguistics, neurology, learning, etc
  - Medical image processing, computer vision, robot motion, etc
- **Shannon entropy** measures the information content or uncertainty of a random variable
- **Mutual information** measures the information transfer in a communication channel

Shannon Entropy

- Discrete random variable X

  $X : \{x_1, x_2, \ldots, x_n\}$, $p_i = p(x_i) = Pr\{X = x_i\}$

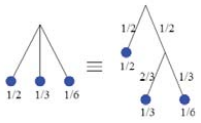
  Shannon entropy of X : **uncertainty**, **information**

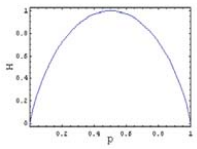
$$H(X) = -\sum_{i=1}^{n} p_i \log p_i$$

- How difficult it is to guess the values of a random variable
- Homogeneity or uniformity of a probability distribution

Shannon Entropy

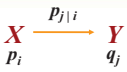
- Properties
  - $0 \le H(X) = \log n$
  - $H(X) = \sum_{i=1}^{m} q_i H(Y_i) - \sum_{i=1}^{m} q_i \log q_i$
  - Binary entropy





Information Channel

- Information channel



$$p_{ij} = p_i p_{j|i}$$

- Conditional entropy

$$H(Y|X) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log p_{j|i}$$

- Joint entropy

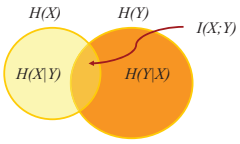
$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log p_{ij}$$

- Mutual information: **dependence**, **correlation**, **shared information**

$$I(X,Y) = H(X) - H(X|Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log \frac{p_{ij}}{p_i q_j}$$

Information Channel

- Properties
  - $0 \le H(X \mid Y) \le H(X)$
  - $H(X,Y) = H(X) + H(Y \mid X)$
  - $H(X,Y) = H(X) + H(Y) - I(X,Y)$
  - $I(X,Y) = I(Y,X) \ge 0$
  - $I(X,Y) \le H(X)$



Inequalities

- Jensen's inequality: if ***f(x)*** is a convex function

$$f(E[X]) \le E[f(X)]$$

- Log-sum inequality

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \ge \left(\sum_{i=1}^{n} a_i\right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}$$

- Data processing inequality : if $X \to Y \to Z$ is a Markov chain, then $I(X,Y) \ge I(X,Z)$

Relative Entropy

- Kullback-Leibler distance

$$D_{KL}(p \,\|\, q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_j}$$

- Properties
  - $D_{KL}(p \,\|\, q) \ge 0$
  - $I(X,Y) = D_{KL}(\{p_{ij}\} \,\|\, \{p_i q_j\})$

Jensen-Shannon Divergence

- Jensen-Shannon divergence

$$JS(\pi_1,\ldots,\pi_N; p_1,\ldots,p_N) = H\left(\sum_{i=1}^{N} \pi_i p_i\right) - \sum_{i=1}^{N} \pi_i H(p_i)$$

$$JS(\pi_1,\ldots,\pi_N; p_1,\ldots,p_N) = \sum_{i=1}^{N} \pi_i D_{KL}\left(p_i \,\|\, \sum_{i=1}^{N} \pi_i p_i\right)$$

- Properties
  - Concavity of entropy: $JS(\pi_1,\ldots,\pi_N; p_1,\ldots,p_N) \ge 0$
  - $JS(p(x_1),\ldots,p(x_n); p(y \mid x_1),\ldots,p(y \mid x_n)) = I(X,Y)$

f-Divergences

- Family of convex functions based on a convex function ***f***

$$D_f(p,q) = \sum_{x \in X} q(x) f\left(\frac{p(x)}{q(x)}\right)$$

- $D_f(p,q)$ is convex on (p,q)
- $D_f(p,q) \ge 0$
- $D_f(p,q) = 0 \Leftrightarrow p = q$

- Kullback-Leibler distance $D_{KL}(p \,\|\, q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$
- Chi-square distance $D_{\chi^2}(p,q) = \sum_{x \in X} \frac{(p(x) - q(x))^2}{q(x)}$
- Hellinger distance $D_{h^2}(p,q) = \frac{1}{2} \sum_{x \in X} (\sqrt{p(x)} - \sqrt{q(x)})^2$

Continuous Channel

- Continuous entropy

$$H^c(X) = -\int_S p(x) \log p(x) dx \qquad \lim_{\Delta \to 0} H(X^\Delta) \ne H^c(X)$$

- Continuous mutual information

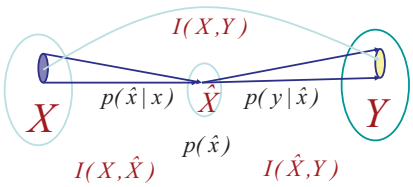
$$I^c(X,Y) = \int_S \int_S p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dx dy$$

$$\lim_{\Delta \to 0} I(X^\Delta, Y^\Delta) = I^c(X,Y)$$

- $I^c(X,Y)$ is the least upper bound for $I(X,Y)$
- Refinement can never decrease $I(X,Y)$

Information Bottleneck Method (IBM)

- Tishby, Pereira and Bialek, 1999
- **Find a compressed signal $\hat{X}$ that needs short encoding (small $I(X,\hat{X})$) while preserving as much as possible the information on the relevant signal ($I(\hat{X},Y)$)**



Agglomerative IBM

- Goal: find a clustering that minimizes the loss of mutual information
- Clustering or merging: loss of mutual information

$$I(X,Y) - I(\tilde{X},Y) =$$
$$p(\tilde{x})JS(p(\tilde{x}_1)/p(\tilde{x}),\ldots,p(\tilde{x}_m)/p(\tilde{x});p(y\mid\tilde{x}_1),\ldots,p(y\mid\tilde{x}_m))$$
$$p(\tilde{x}) = \sum_{k=1}^{m} p(\tilde{x}_k)$$

- The quality of each cluster $\hat{x}$ is measured by the Jensen-Shannon divergence between the individual distributions in the cluster

Generalised Entropy

- Harvda-Charvát-Tsallis entropy (HCT)

$$H_\alpha(X) = k\frac{1-\sum_{i=1}^{n} p_i^\alpha}{\alpha-1} \qquad k>0, \alpha \in R\setminus\{1\}$$

$$H_1(X) \equiv \lim_{\alpha\to 1} H_\alpha(X) = -k\sum_{i=1}^{n} p_i \ln p_i$$

- Generalised mutual information

$$I_\alpha(X,Y) = \frac{1}{1-\alpha}\left(1-\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{p_{ij}^\alpha}{p_i^{\alpha-1}q_j^{\alpha-1}}\right)$$

(3) Refinement Criteria for Radiosity

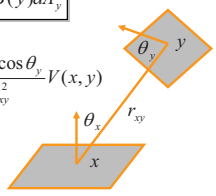
Jaume Rigau
Miquel Feixas
Mateu Sbert

Radiosity Method

- The radiosity method solves the problem of illumination in an environment of diffuse surfaces
- Continuous radiosity equation

$$B(x) = E(x) + \rho(x)\int_S F(x,y)B(y)dA_y$$

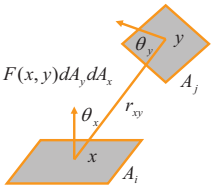
$$F(x,y) = \frac{\cos\theta_x \cos\theta_y}{\pi r_{xy}^2} V(x,y)$$



Radiosity Method

- Discrete radiosity equation

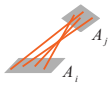
$$B_i = E_i + \rho_i\sum_{j=1}^{n_p} F_{ij}B_j \qquad F_{ij} = \frac{1}{A_i}\int_{A_i}\int_{A_j} F(x,y)dA_y dA_x$$



- Form factor properties
  - Reciprocity $A_iF_{ij} = A_jF_{ji}$
  - Energy conservation $\sum_{j=1}^{n_p} F_{ij} = 1$

Form Factor Computation

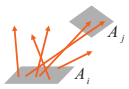
- Analytical solutions
  - Between two spherical patches $F_{ij} = \frac{A_j}{A_S}$
- Monte Carlo computation
  - Uniform area sampling



$$\hat{F}_{ij} = A_j \frac{1}{N}\sum_{k=1}^{N} F(x_k, y_k)$$

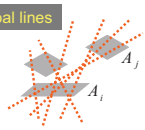
  - Uniformly distributed lines

Local lines



$$\hat{F}_{ij} = \frac{N_{ij}}{N_i}$$

Global lines



Refinement Criteria for HR

- In hierarchical radiosity (HR), the mesh is generated adaptively
- Oracles based on
  - Transported power

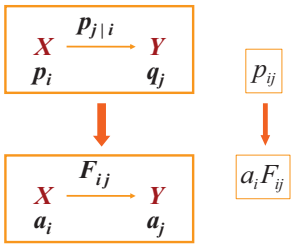
$$\rho_i A_i F_{ij} B_j < \varepsilon$$

  - Kernel smoothness

$$\rho_i \max(F_{ij}^{\max} - F_{ij}^{av}, F_{ij}^{av} - F_{ij}^{\min}) A_j B_j < \varepsilon$$

Scene Information Channel

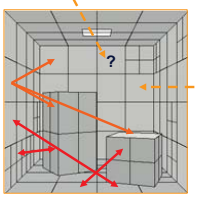
- The scene is modelled as an information channel



Scene Information Channel

Positional entropy $H_P = -\sum_{i=1}^{n_P} a_i \log a_i$

Scene entropy $H_S = -\sum_{i=1}^{n_P} a_i \sum_{j=1}^{n_P} F_{ij} \log F_{ij}$



$a_i = \frac{A_i}{A_T}$

Scene mutual information $I_S = H_P - H_S = \sum_{i=1}^{n_P}\sum_{j=1}^{n_P} a_i F_{ij} \log \frac{F_{ij}}{a_j}$

Continuous Mutual Information

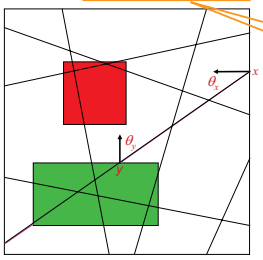
- By discretising a scene, a distortion or error is introduced: information loss
- **From** discrete **to** continuous
  - $\Sigma \rightarrow \int$
  - $F_{ij} \rightarrow F(x,y)$
  - $a_i = A_i / A_T \rightarrow 1 / A_T$

$$I_s^c = \int_{x \in S} \int_{y \in S} \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dx dy$$

Monte Carlo Computation

$$I_S^c \approx \frac{1}{N}\sum_{1}^{N} \log\left(\frac{A_T \cos\theta_x \cos\theta_y}{\pi r_{xy}^2}\right)$$

contribution of each segment



Total area = A_T

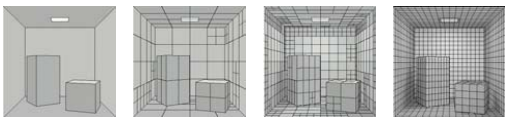
Lines cast = K

Line segments = N

Dicretisation Error

- Two basic results
  - If any patch is subdivided, I_S increases or remains the same
  - I_S^c is the least upper bound to I_S
- Discretisation error $\boxed{I_S^c - I_S \geq 0}$

$I_S^c = 3.273$



$I_S = 0.690$ $I_S = 2.199$ $I_S = 2.558$ $I_S = 2.752$

Information Transfer

- Mutual information matrix

$$I_S = \sum_{i=1}^{n_p}\sum_{j=1}^{n_p} \boxed{a_i F_{ij} \log \frac{F_{ij}}{a_j}} \Leftarrow I_{ij}$$

information transfer between patches ***i*** and ***j***

$\Uparrow I_i$

information transfer from patch ***i***

$$I_S^c = \sum_{i=1}^{n_p}\sum_{j=1}^{n_p} \boxed{\int_{A_i}\int_{A_j} \frac{1}{A_T} F(x,y)\log(A_T F(x,y))dxdy}$$

Discretisation Error Between Two Patches

- Discretisation error between two elements: loss of information transfer

$$\boxed{\delta_{ij} = I_{ij}^c - I_{ij}}$$

$\Downarrow$

Monte Carlo integration

log-sum inequality

$$\delta_{ij} \approx \frac{A_i A_j}{A_T}\left[\frac{1}{N_{ij}}\left(\sum_{k=1}^{N_{ij}} F(x_k,y_k)\log F(x_k,y_k)\right) - \left(\left(\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}} F(x_k,y_k)\right)\log\left(\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}} F(x_k,y_k)\right)\right)\right] \geq 0$$

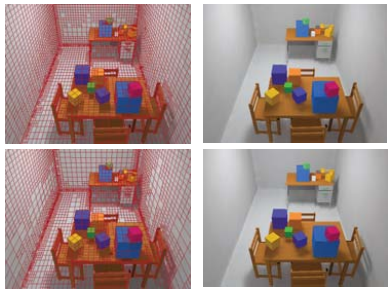
MI-based Oracle

- **From** radiosity equation and kernel-smoothness-based oracle
  - $B_i = E_i + \sum_{j=1}^{n_p} \boxed{\rho_i F_{ij} B_j}$
  - $\rho_i \boxed{\max(F_{ij}^{\max} - F_{ij}^{av}, F_{ij}^{av} - F_{ij}^{\min})A_j} B_j < \varepsilon$
- **to** MI-based oracle
  - $\boxed{\rho_i (I_{ij}^c - I_{ij}) B_j = \rho_i \delta_{ij} B_j < \varepsilon}$

Oracles for HR

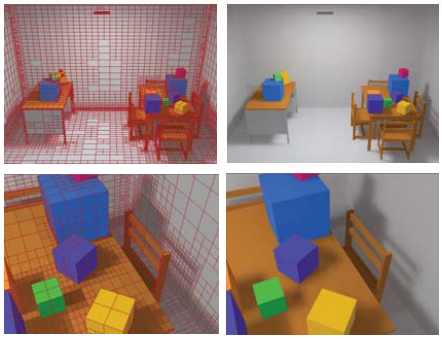
Kernel-smoothnes-based

MI-based



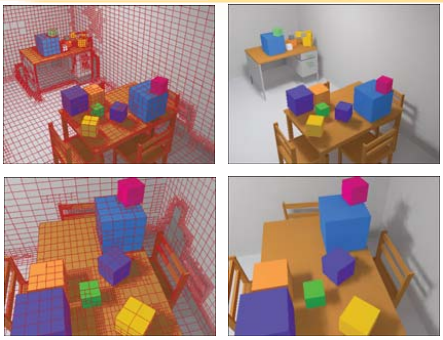
2684000 rays - 19000 patches - 10 lines FF

MI-based Oracle for HR



2684000 rays - 19000 patches - 10 lines FF

Generalised MI-based Oracle



2684000 rays - 19000 patches α=0.50 - 10 lines FF

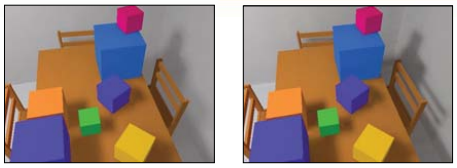
Generalised MI-based Oracle

α=0.50 - 10 lines FF - 9268000 rays - 10000 patches

f-Divergence-based Oracles

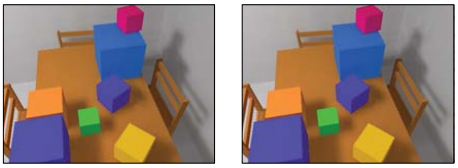
Kernel-Smoothness

Kullback-Leibler



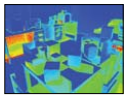
Chi-Square

Hellinger



10 lines FF - 2684000 rays - 19000 patches

(4) Refinement Criteria for Ray-Tracing

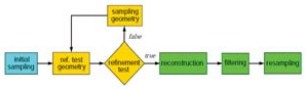


Jaume Rigau
Miquel Feixas
Mateu Sbert

07 EUROGRAPHICS Prague Czech Republic

Adaptive Sampling

- Adaptive control of the sampling rate



- Image-Space
  - Intensity Comparison
  - Intensity Statistics

$$C(S)=\frac{S_{\max}-S_{\min}}{S_{\max}+S_{\min}}$$ [Mitchell, 87]

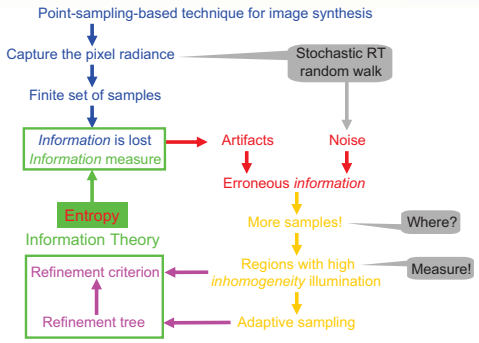
$$\Pr\{\overline{S}_T \in [\overline{S}-t,\overline{S}+t]\}=1-\alpha$$
[Purgathofer, 87]
[Tamstorf and Jensen, 97]

- Object-Space

$$p_g=1-\frac{d_{\min}}{d_{\max}}$$
[Simmons and Séquin, 00]

- Hybrid (image+object spaces)

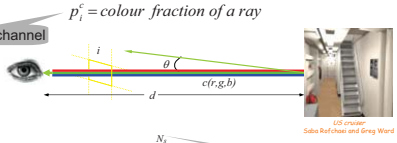
Pixel Measures



Pixel Colour Quality

$p_i^c = colour\ fraction\ of\ a\ ray$

For each channel



US cruiser
Soba Rafchaei and Greg Ward

pixel channel entropy $H^c = -\sum_{i=1}^{N_s} p_i^c \log p_i^c$

Number of samples

pixel channel quality $Q^c = \frac{H^c}{\log N_s}$

Channel perception coefficient

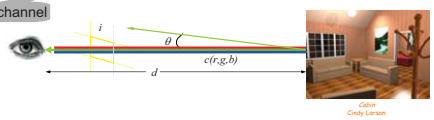
pixel colour quality $\mathbf{Q}^{\mathbf{c}} = \frac{\sum_{c \in \mathbf{c}} w^c Q^c}{\sum_{c \in \mathbf{c}} w^c}$

Colour system

Pixel Colour Contrast

$p_i^c = colour\ fraction\ of\ a\ ray$

For each channel



Cabin
Cindy Larson

pixel channel contrast $C^c = 1 - Q^c$

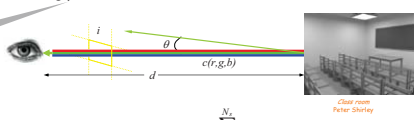
Pixel channel colour average

pixel colour contrast $\mathbf{C}^{\mathbf{c}} = \frac{\sum_{c \in \mathbf{c}} w^c \bar{c}\, C^c}{\sum_{c \in \mathbf{c}} w^c \bar{c}}$

Pixel Geometry Contrast

$p_i^g = geometric\ fraction\ of\ a\ ray$

$\frac{\cos\theta}{d^2}$



Clock room
Peter Shirley

pixel geometric entropy $H^g = -\sum_{i=1}^{N_s} p_i^g \log p_i^g$

pixel geometric quality $Q^g = \frac{H^g}{\log N_s}$

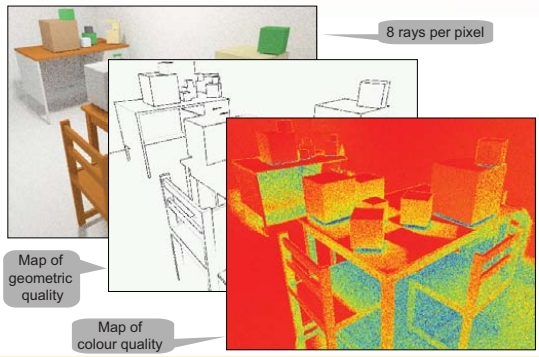
pixel geometric contrast $C^g = 1 - Q^g$

Combination coefficient

pixel contrast $C^{\mathbf{c}} = \delta \mathbf{C}^{\mathbf{c}} + (1-\delta) C^g$

Combination of colour and geometry

Quality Map

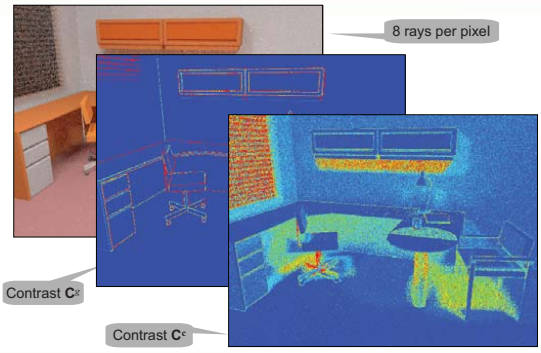


8 rays per pixel

Map of geometric quality

Map of colour quality

Contrast Map

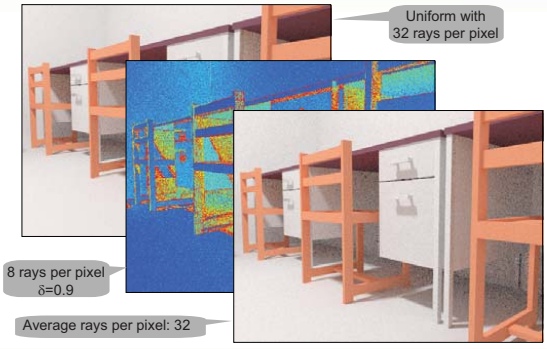


8 rays per pixel

Contrast $\mathbf{C}^g$

Contrast $\mathbf{C}^c$

Supersampling



Uniform with 32 rays per pixel

8 rays per pixel
δ=0.9

Average rays per pixel: 32

Entropy-based Adaptive Sampling

Grouping property of Entropy

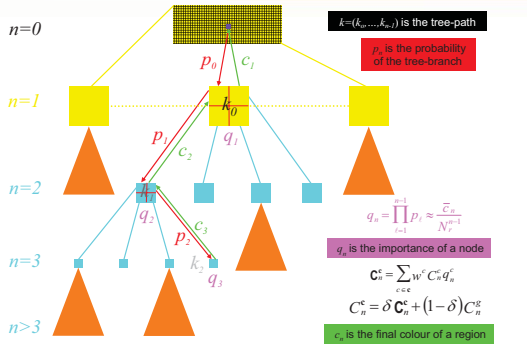
$$H(X) = \underbrace{-\sum_{i=1}^{m} q_i \log q_i}_{\text{information acquired}} + \underbrace{\sum_{i=1}^{m} q_i H(Y_i)}_{\text{hidden information}}$$

image *information*

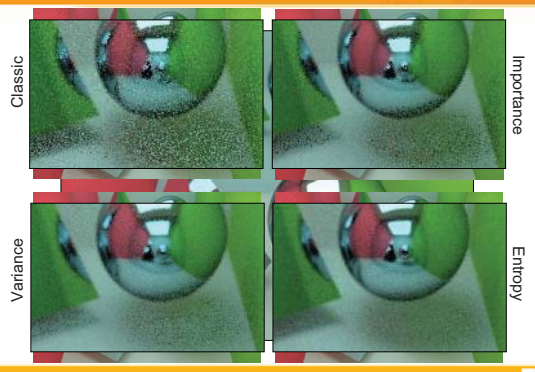
- $H(X)$ ≡ entropy of the whole image
- $H(Y_i)$ ≡ entropy of each root pixel
- q_i ≡ colour probability of pixel i

The decomposition of H can be recursively extended to the subpixels

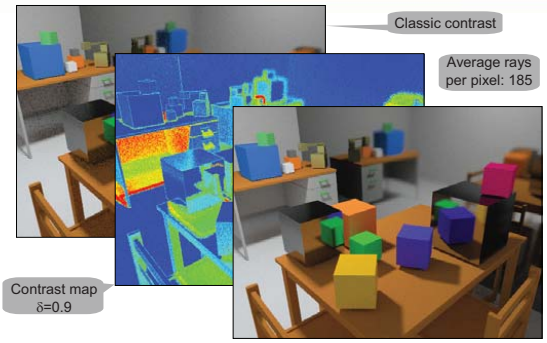
Contrast Tree



Results



Results

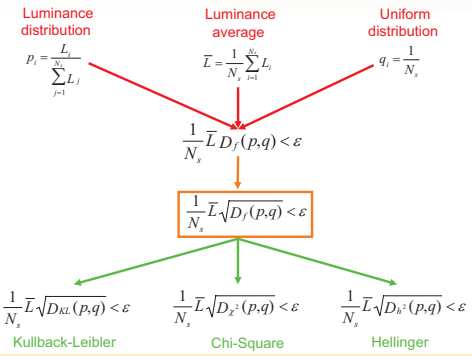


f-Divergences

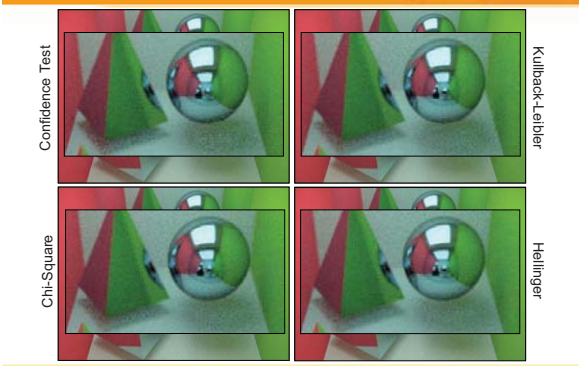
- *f*-Divergences as refinement criteria in RT ?
- Distributions
  - $\{p\}$ = Luminance L of N_S-samples
  - $\{q\}$ = Uniform $1/N_S$
- Homogeneity: $D_f(p,q)$
- Weights for D_f
  - Importance: $avg(L_i)$
  - Convergence: $1/N_S$

$$D_f(p,q) = \sum_{x \in X} q(x) f\left(\frac{p(x)}{q(x)}\right)$$

f-Divergence-based Adaptive Sampling



Results



49

Introduction

- Viewpoint selection is an emerging area in computer graphics with applications in fields such as scene understanding, volume visualization, image-based modeling, and molecular visualization
- We present a unified framework for viewpoint selection and mesh visibility / saliency / simplification based on an information channel between a set of viewpoints and the polygons of an object
- Tools: entropy, mutual information, Jensen-Shannon divergence
- This framework is based on the geometric characteristics of the object, but it can be extended to other characteristics
- It is also valid for any set of viewpoints in a closed scene
- What is a good viewpoint? Depending on our objective, the best viewpoint can be the most representative one or the most unstable one (maximally changes when it is moved within its close neighborhood) or …
  - Representative views can help us to understand the object
  - Unstable views enable us to obtain critical viewpoints to capture the structure of the object

51

Background and Related Work

- **Information Theory**
  - Discrete random variable X

$$X : \{x_1, x_2, \ldots, x_n\}, \; p(x_i) = Pr\{X = x_i\}$$

  - Shannon entropy of X : uncertainty, ignorance

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x).$$

52

Background and Related Work

- **Information Theory**
  - Information Channel

$$X \xrightarrow{\{p(y|x)\}} Y$$
$$\{p(x)\} \qquad \{p(y)\}$$

  - Conditional Entropy

$$H(Y|X) = -\sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x).$$

  - Mutual Information

$$I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = \sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log \frac{p(y|x)}{p(y)}.$$

  - Jensen-Shannon inequality

$$JS(p_1, p_2, \ldots, p_N) = H(\sum_{i=1}^{N} \pi_i p_i) - \sum_{i=1}^{N} \pi_i H(p_i) \geq 0.$$

53

Background and Related Work

- **Related Work**
  - Heuristic measure **Plemenos et al. [1996]**

$$C(v) = \frac{\sum_{i=1}^{n} \lceil \frac{P_i(v)}{P_i(v)+1} \rceil}{n} + \frac{\sum_{i=1}^{n} P_i(v)}{r},$$

  - Viewpoint Entropy

$$H(v) = -\sum_{i=0}^{N_f} \frac{a_i}{a_t} \log \frac{a_i}{a_t},$$

  - Kullback-Leibler distance

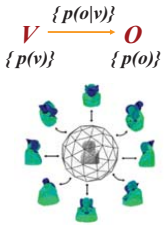
$$KL(v) = \sum_{i=1}^{N_f} \frac{a_i}{a_t} \log \frac{\frac{a_i}{a_t}}{\frac{A_i}{A_T}},$$

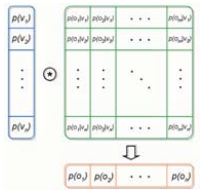
  - **Origins Rigau et. al [2000], Vázquez et al. [2001-2006], Sbert. Et al [2005]**

54

Viewpoint Information Channel

- We formalize the viewpoint selection using an information channel





$$p(o) = \sum_{v \in \mathcal{V}} p(v)p(o|v)$$

- This framework is based on geometric characteristics

Viewpoint Information Channel

- **Viewpoint Mutual Information**
  - Conditional Entropy

$$H(O|V) = -\sum_{v \in \mathcal{V}} p(v) \sum_{o \in \mathcal{O}} p(o|v) \log p(o|v)$$
$$= \frac{1}{N_v} \sum_{v \in \mathcal{V}} H(v),$$

- *H(v)* depends on the polygonal discretization
- *MI* converges to a finite value when the mesh is infinitely refined

  - Mutual Information: degree of correlation, dependence

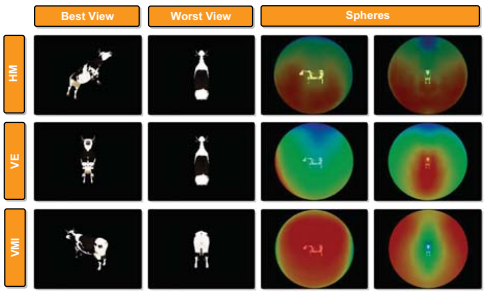
$$I(V,O) = \sum_{v \in \mathcal{V}} p(v) \sum_{o \in \mathcal{O}} p(o|v) \log \frac{p(o|v)}{p(o)}$$
$$= \sum_{v \in \mathcal{V}} p(v) I(v,O)$$

- **Low values:** representative views
- **High values:** highly coupled views

$$I(v,O) = \sum_{o \in \mathcal{O}} p(o|v) \log \frac{p(o|v)}{p(o)}$$

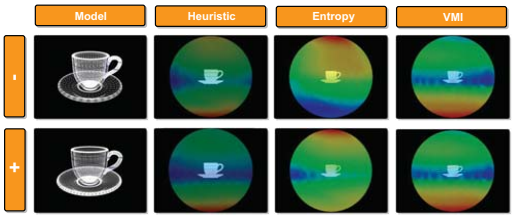
Viewpoint Information Channel

- **Viewpoint Mutual Information evaluation (I)**



Viewpoint Information Channel

- **Viewpoint Mutual Information evaluation (II)**



Viewpoint Information Channel

- Viewpoint Similarity and Unstability
  - Viewpoint Similarity
    - Any clustering over $V \rightarrow \hat{V}$ or $O \rightarrow \hat{O}$ reduce *I(V,O)*

$$\begin{aligned}\delta I(v_i, v_j) &= I(V,O) - I(\hat{V},O) \\ &= (p(v_i)I(v_i,O) + p(v_j)I(v_j,O)) - p(\hat{v})I(\hat{v},O) \\ &= p(\hat{v})\left(\frac{p(v_i)}{p(\hat{v})}I(v_i,O) + \frac{p(v_j)}{p(\hat{v})}I(v_j,O) - I(\hat{v},O)\right) \\ &= p(\hat{v})D(v_i,v_j),\end{aligned}$$

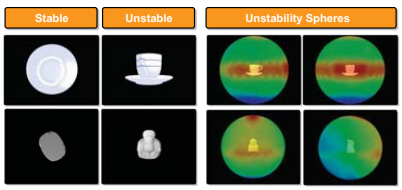
$$D(v_i,v_j) = JS\left(\frac{p(v_i)}{p(\hat{v})}, \frac{p(v_j)}{p(\hat{v})}; p(O|v_i), p(O|v_j)\right)$$

Viewpoint Information Channel

- Viewpoint Similarity and Unstability
  - Viewpoint Unstability

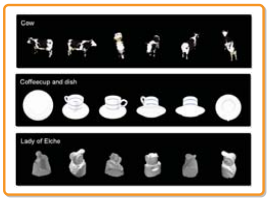
$$U(v_i) = \frac{1}{N_n} \sum_{j=1}^{N_n} D(v_i, v_j)$$

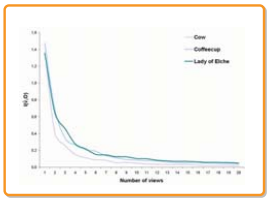
- The maximum change in view that occur when the camera position is shifted within a small neighborhood



Viewpoint Information Channel

- **Selection of *n* Best Views**
  - **Objective:** to select the minimal set of representative views
  - **Ideal proposal:** ***n*** views that maximize their ***JS*** (to capture the maximum information of the object)
  - **Greedy strategy:** to select successive views that maximize ***JS***

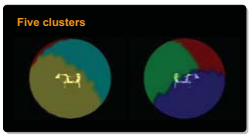




61

Viewpoint Information Channel

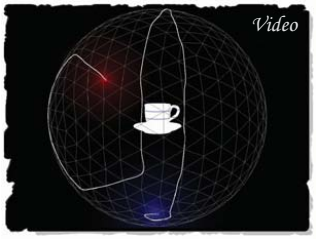
- Viewpoint Clustering
  - Clustering algorithm
    - Select the ***n*** best views
    - Assign each viewpoint to the ***nearest*** best viewpoint

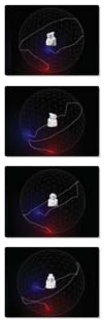


62

Scene Exploration

- **Exploratory Tour**

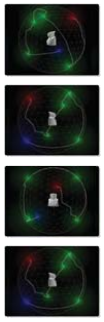




63

Scene Exploration

- **Guided Tour**



64

Mesh Visibility

- **Reversion of the Channel**
  - Channel is reversed using the Bayes theorem

$$p(v,o) = p(v)p(o|v) = p(o)p(v|o)$$

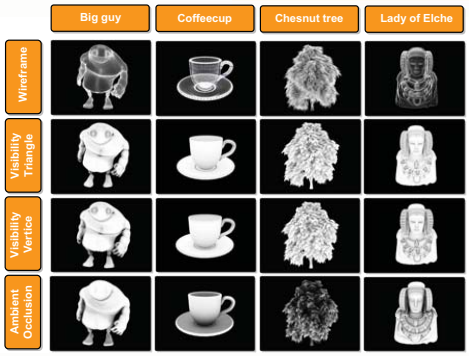
$$I(V,O) = \sum_{v\in\mathcal{V}} p(v) \sum_{o\in\mathcal{O}} p(o|v) \log \frac{p(o|v)}{p(o)}$$

$$I(V,O) = \sum_{o\in\mathcal{O}} p(o) \sum_{v\in\mathcal{V}} p(v|o) \log \frac{p(v|o)}{p(v)} = \sum_{o\in\mathcal{O}} p(o) I(V,o),$$

- ***I(V,o)*** is the polygonal mutual information
- Degree of correlation between the polygon o and the set of viewpoints

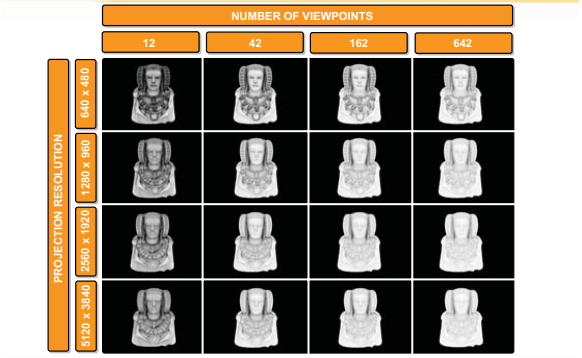
65

Mesh Visibility



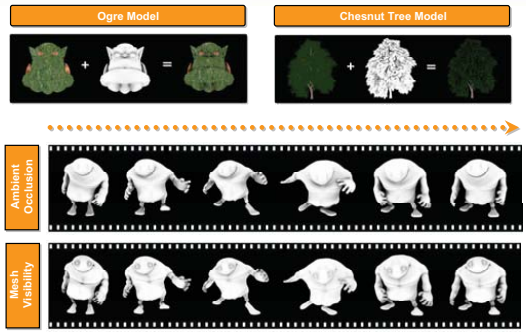
66

Mesh Visibility



67

Mesh Visibility



68

Mesh Visibility

69

Mesh Visibility

70

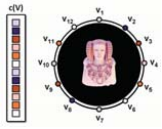
Mesh Visibility

- **Applications**
  - Important viewpoints
    - Importance at the viewpoint space
    - Selection according to geometry and saliency

71

Mesh Visibility

- **Applications**
  - Relighting for *Non-Photorealistic* Rendering
    - Warping a color palette texture to the viewpoint sphere



    - Color ambient occlusion + *NPR* technique

72

Mesh Visibility

- **Applications**
  - Relighting *NPR* + *Coloroid Palettes*

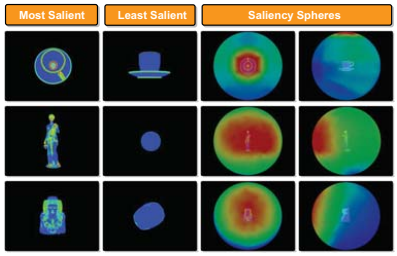
Mesh Visibility

Mesh Saliency

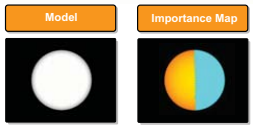
$$S(o_i) = \frac{1}{N_o} \sum_{j=1}^{N_o} JS(p(V|o_i), p(V|o_j)) \geq 0,$$

Viewpoint Saliency

$$S(v) = \sum_{o \in \mathcal{O}} S(o) p(v|o)$$

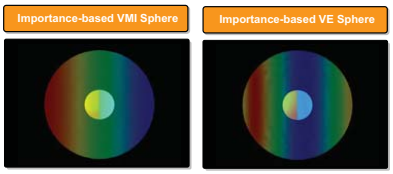


Importance-based Viewpoint Mutual Information



$$I'(v,O) = \sum_{o \in \mathcal{O}} p(o|v) \log \frac{p(o|v)}{p'(o)},$$

$$p'(o) = \frac{p(o)i(o)}{\sum_{o \in \mathcal{O}} p(o)i(o)}$$



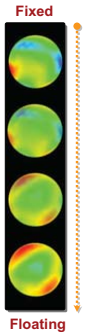
Importance-based Viewpoint Mutual Information

View-based Object Recognition

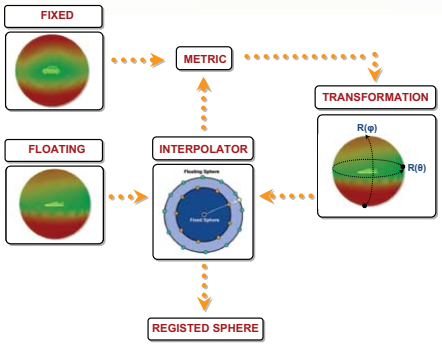
- **System features**
  - VMI Sphere ⟹ View-based Shape descriptor
  - Rigid registration system ⟹ Rotations (θ, φ)
  - 642 viewpoints
  - Fixed & Floating Sphere
  - Metric

$$MSE(A,B)=\sqrt{\sum_{i=1}^{N}(a_i-b_i)^2}$$

  - Interpolator ⟹ Nearest Neighbour

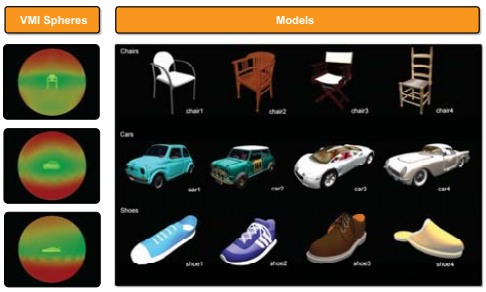


View-based Object Recognition



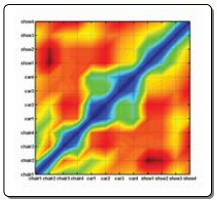
View-based Object Recognition

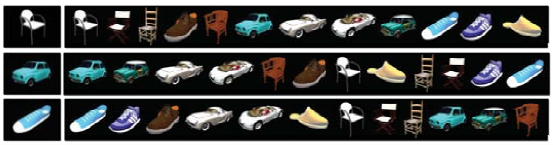
- **Results**



View-based Object Recognition

- **Results**





Ivan Viola
University of Bergen
Norway

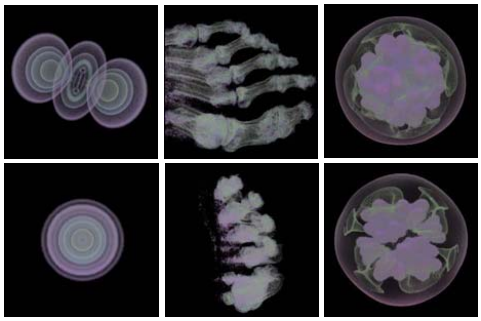
07 EUROGRAPHICS Prague Czech Republic

View Selection for Volume Data

- Viewpoint quality = visibility of data
- Visibility computation
- Information-theoretic measures for characteristic viewpoint estimation
  - Viewpoint entropy
  - Mutual information
- View selection approaches for
  - 3D scalar fields
  - 3D + time scalar fields
  - Objects in volume data

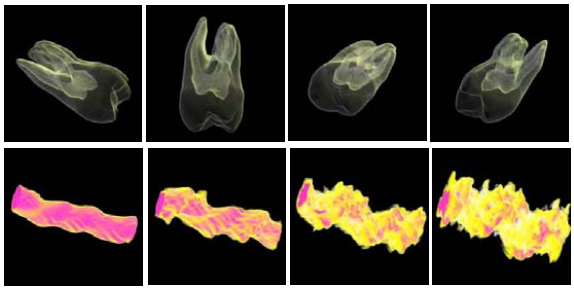
View Selection for Set of Iso-Surfaces

[Takahashi et al. Vis05]



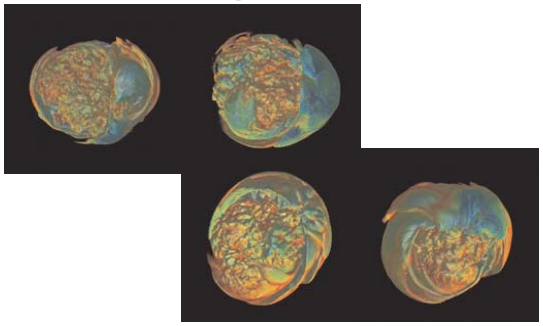
View Selection for Scalar Volumes (+ Time)

[Bordoloi and Shen Vis05]



Dynamic Views for Time-Varying Volumes

[Ji06 and Shen Vis06]



Focus of Attention

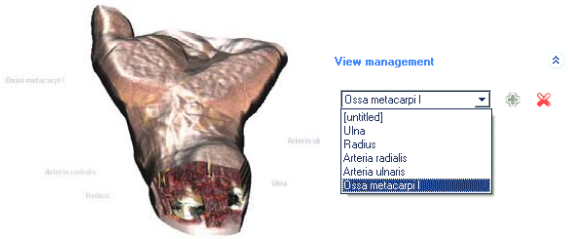
View Selection for Volumetric Objects

Focus of Attention

- Importance distribution among objects controls:
  - Characteristic view computation
  - Interactive focusing
- Characteristic view computation
  - View rating image and object weights
  - For every object + context
- Interactive focusing
  - Visual emphasis and cutaways
  - Changing the focus among objects

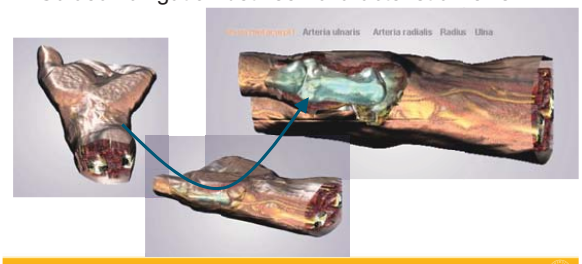
Goal

- Input: known and classified volumetric data
- High level request: show me object *X*
- Output: guided navigation to object *X*



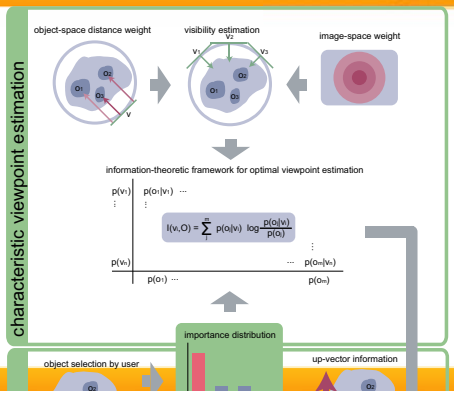
Focusing Considerations

- Characteristic view
- Emphasis of focus object
- Guided navigation between characteristic views

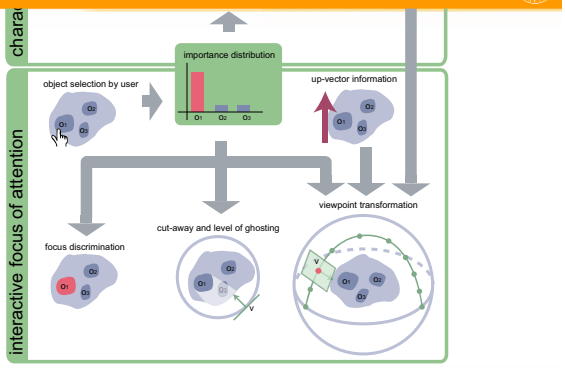


91

Framework

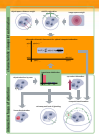


Framework



93

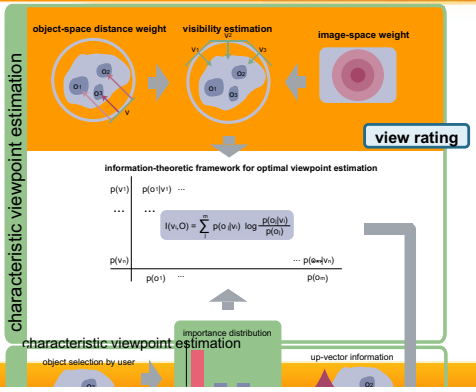
Characteristic Views



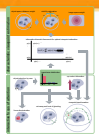
- Overview
  - All objects are visible
  - Visibility of objects is balanced
- Characteristic view of focus object
  - High visibility for focus object
  - If possible other objects also visible

94

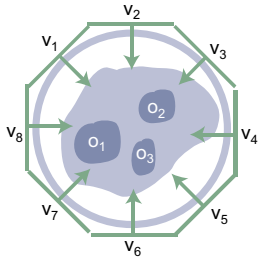
Characteristic View Estimation



View rating



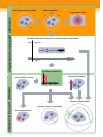
- For every view
  - For every object



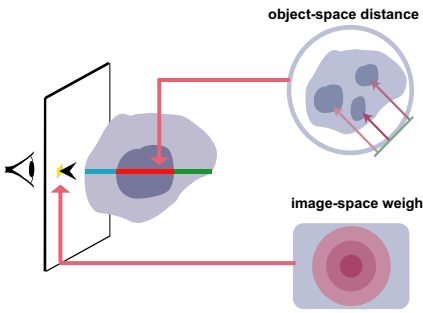
96

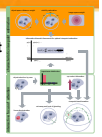
View Rating

- Visibility
  - High
  - Low
- Location in image
  - In image center
  - Outside center
- Distance to the viewer
  - Object close to the viewer
  - Far from the viewer

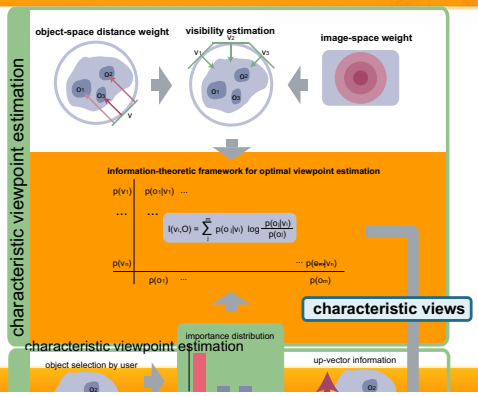


View Rating Weights



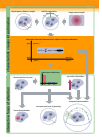


Characteristic Viewpoint Estimation



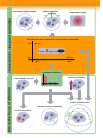
Characteristic Views

- Overview
  - All objects are visible
  - Visibility of objects is balanced
- Characteristic view of focus object
  - High view rating (visibility) for focus object
  - If possible other objects also visible



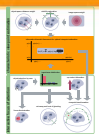
Obtaining Characteristic Views

- Sets of views and objects are random variables
  - Views $V=(v_1, v_2, v_3, \dots, v_n)$
  - Objects $O=(o_1, o_2, o_3, \dots, o_m)$
- View rating (visibility, weights)
  - Information channel between $V \rightarrow O$
  - Conditional probability $p(o_j|v_i)$
- Mutual information between V and O expresses degree of dependance

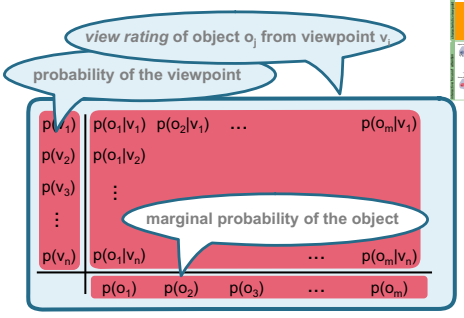


Obtaining Characteristic Views

- Viewpoint mutual information is dependance between v_i and O
  - High values = high dependance
    - Small number of objects
    - Low average visibility
  - Low values = low dependance
    - Maximum objects visible
    - Object visibility is balanced
- Minimal VMI determines the best view



Probability Transition Matrix

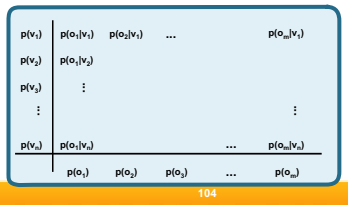


103

Viewpoint Mutual Information

- Degree of correlation $v_j \leftrightarrow O$

$$I(v_i, O) = \sum_j p(o_j \mid v_i) \log \frac{p(o_j \mid v_i)}{p(o_j)}$$



104

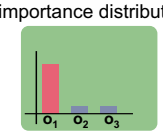
Characteristic Views

- Overview
  - All objects are visible
  - Visibility of objects is balanced
- Characteristic view at focus object
  - High view rating for focus object
  - If possible other objects also visible

105

Incorporating Importance

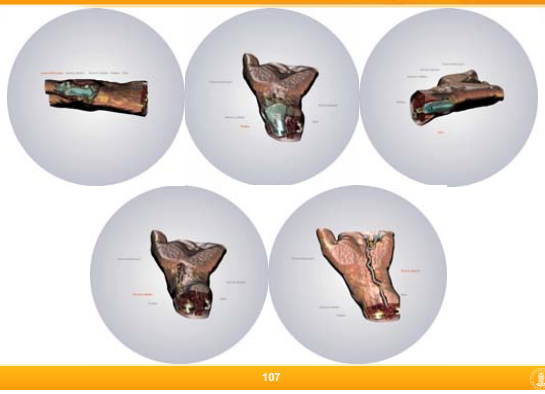
importance distribution



$$I(v_i, O) = \sum_j p(o_j \mid v_i) \log \frac{p(o_j \mid v_i)}{\frac{p(o_j)im(o_j)}{\sum_k p(o_k)im(o_k)}}$$

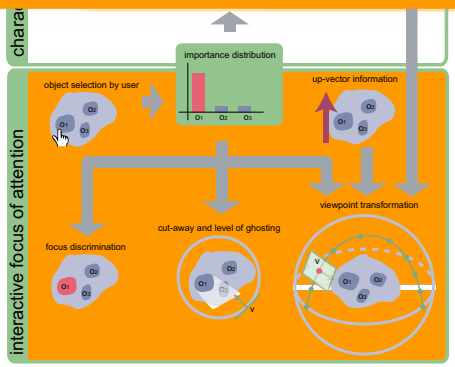
106

Resulting Characteristic Viewpoints



107

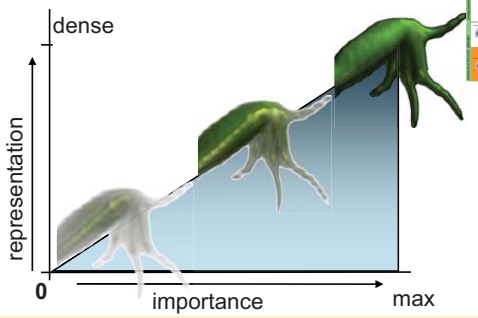
Interactive Focus of Attention



108

Emphasis of Focus Object

- Levels of sparseness



109

Emphasis of Focus Object

- Cut-aways to unveil internal features
- Labeling to add textual information

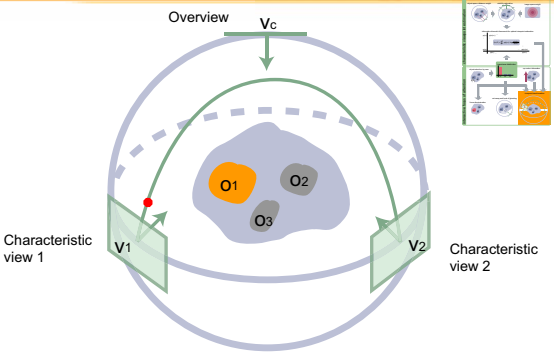
110

Guided Navigation Between Objects

- Decreasing importance of Object *X*
  - De-emphasis of Object *X*
  - Change to *overview*
- Increasing importance of Object *Y*
  - Emphasis of Object *Y*
  - Change to characteristic view of *Y*

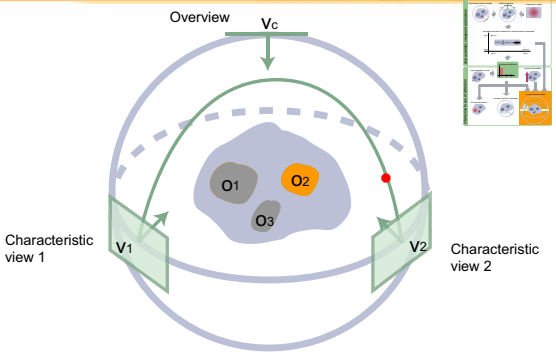
111

Refocusing



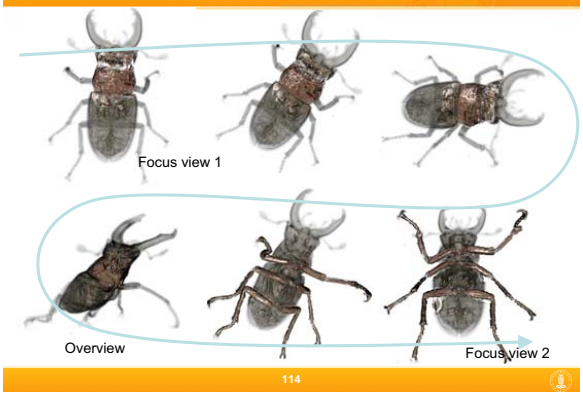
112

Refocusing



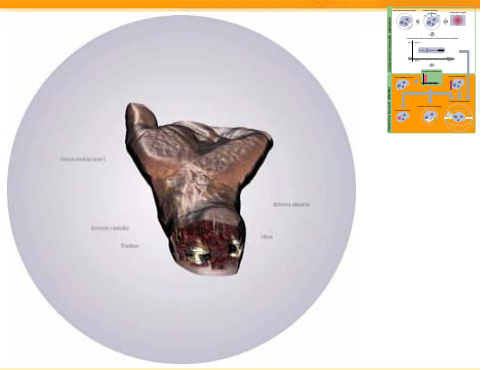
113

Example - Stagbeetle



114

Example - Hand



Application-Driven View Selection

[Mühler et al. EuroVis07]

Introduction

- Most simplification methods use some geometric distance to guide the simplification process
- Recently, some works have developed methods guided by visual error metrics
- In some real-time applications like computer games the main requirement is visual similarity
- We propose new simplification metrics which produce closer approximations to the original model based on Information Theory

Recent Work on Simplification

- Geometry-Based
  - Appearance-Preserving Simplification [Cohen98]
  - Simplifying Surfaces with Color and Texture using Quadric Error Metrics [GH98]
  - New quadric metric for simplifying meshes with appearance attributes [Hoppe99]
  - Mesh Saliency [LVJ05]
- Viewpoint-Based
  - Image-Driven Simplification [LT00]
  - Perceptual-Driven Simplification for Interactive Rendering [LH01]
  - Visibility-Guided simplification [ZT02]
  - Viewpoint Entropy-driven Simplification [CSCF07]

Pros and Cons

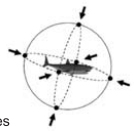
- Geometry- Based
  - The algorithm runs faster
  - Manage complex meshes
    - CAD, Scanned
  - Adjust geometric tolerance
- Viewpoint-Based
  - The algorithm runs slower
  - Deal with simple meshes
    - Games, Virtual Reality
  - Remove interior parts and preserves silhouette

"Current Game Artists make the simplifications by hand"

Viewpoint Entropy

- Definition
  - The Viewpoint Entropy gives a measure of the information provided by a point of view
  - We take as a probability distribution the relative area of the projected polygons over the sphere of directions centered in the viewpoint v

$$H_v = -\sum_{i=0}^{N_f} \frac{a_i}{a_t} \log \frac{a_i}{a_t}$$



  - Where:
    - N_f: number of polygons in the scene
    - a_i: projected area of polygon i over the sphere
      - a_0: projected area of background in open scenes
      - a_t: total area of the sphere

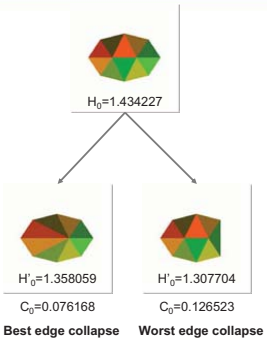
The best viewpoint is the one that has maximum entropy, i.e., maximum information captured

Viewpoint Entropy

- The error metric
  - Defined as the sum of variations of viewpoint entropy for all viewpoints V

$$\mathcal{C} = \sum_{v \in V} |H_v - H'_v|$$

  - Where:
    - H_v is the viewpoint entropy before an edge collapse
    - H'_v is the viewpoint entropy after an edge collapse



Simplification algorithm

```
/* Compute I_v for the original mesh M */
Compute I_v where v={1,..,n}
/* Build initial heap of edge collapses */
for ( e ∈ M)
  Perform collapse e
  Compute I'_v where v={1,..,n}
  Compute collapse cost C_e
  Insert (e, C_e) in heap h
  Undo collapse e
end for
/* Update the mesh */
while ( heap h not empty )
  Remove from heap h the edge e with lowest C_e
  Perform collapse e
  for ( each e' in neighborhood )
    Compute collapse cost C_e'
    Update (e', C_e')  location in heap h
  end for
end while
```


Experiments

- Comparison
  - Algorithm
    - QSLIM v2.0 [Gar97] Well-Know geometric simplification algorithm
  - Tools
    - Geometric error: METRO v4.06 [Cig98]
    - Visual error: RMSE [Lin00]
- 20 viewpoints regularly distributed over a sphere
- Resolution: 256x256 images
- PC, Xeon 2.4 GHz, 1GB RAM, NVIDIA 7800 GTX 512MB
- C++ implementation with OpenGL
  - Vertex Buffer Objects & Frame Buffer Objects

Experiments H_V

Original Fish T=815 | QSlim T=100 | H_v C=20 T=100

Original Galleon T=4.698 | QSlim T=500 | H_v C=20 T=500

Experiments H_V

Original Galo T=6.592 | QSlim T=500 | H_v C=20 T=500

Original Octopus T=8.468 | QSlim T=1.000 | H_v C=20 T=1.000

Experiments H_V


Original Porsche T=10.474


QSlim T=1.000


H_v C=20 T=1.000


Original Unicycle T=13.810


QSlim T=1.000

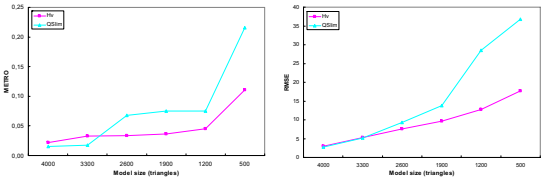

H_v C=20 T=1.000

Experiments H_V

Model	Triangles		RMSE Error		METRO Error		Time	
	Original	Final	QSlim	H_v	QSlim	H_v	QSlim	H_v
Fish	815	100	22,83	**11,40**	0,09	**0,05**	0.02	11.16
Galleon	4.698	500	36,84	**17,74**	0,22	**0,11**	0.06	92.64
Galo	6.592	500	12,40	**9,03**	0,12	**0,08**	0.08	152.29
Octopus	8.468	500	25,84	**17,35**	0,05	**0,03**	0.09	224.89
Porsche	10.474	1.000	8,28	**7,48**	0,16	**0,09**	0.13	299.47
Unicycle	13.810	1.000	11,06	**10,32**	0,10	**0,04**	0.20	451.76

Experiments H_V

Geometric Error **Visual Error**



Comparison at several degrees of simplification of the **Galleon model**

Mutual Information

- Definition
  - The Viewpoint Mutual Information defines an information channel between V and O

$$p(v) = \frac{1}{N_f} \qquad p(o) = \sum_{v \in V} p(v)p(o \mid v) = \frac{1}{N_f}\sum_{v \in V} p(o \mid v)$$

  - The conditional probabilities of $p(o|v)$ are given by the relative area of the projected polygons over the sphere of directions centred at viewpoint v

$$I(V,O) = \sum_{v \in V} p(v) \sum_{o \in O} p(o \mid v) \log \frac{p(o \mid v)}{p(o)} = \frac{1}{N_v}\sum_{v \in V} I(v,O)$$

Mutual Information

- The mutual information for a given viewpoint

$$I(v,O) = \sum_{o \in O} p(o \mid v) \log \frac{p(o \mid v)}{p(o)}$$

- High values mean high degree of dependence “highly coupled view”
- Low values correspond to low dependence “more representative view”
- Observe that

$$I(v,O) = KL\big((p(O \mid v) \mid p(O))\big)$$

Mutual Information

- The error metric
  - Defined as the sum of variations of viewpoint mutual information for all viewpoints V

$$c = \sum_{v \in V} \left| I_v - I'_v \right|$$

Model	Triangles		RMSE Error		Time	
	Original	Final	H_v	VMI	H_v	VMI
Shark	734	80	14,78	**14,65**	10,24	10,23
Galo	6592	500	9,05	**8,38**	141,75	142,24
Greekship	9510	600	13,37	**12,85**	241,78	246,72
Tree	11.136	600	17,23	**16,60**	321,06	332,49
Hammer	13.380	500	8,13	**7,43**	404,33	423,05
Elephant	31.548	900	13,75	**11,60**	2197,67	2309,79

Experiments VMI

Original Shark T=734 | QSlim T=80 | VMI C=20 T=80

Original Galo T=6.592 | QSlim T=500 | VMI C=20 T=500

Experiments VMI

Original Greekship T=9.510 | QSlim T=600 | VMI C=20 T=600

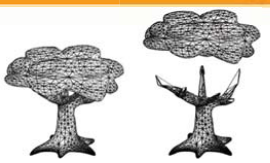
Original Tree T=11.136 | QSlim T=600 | VMI C=20 T=600

Experiments VMI

Original Hammer T=13.380 | QSlim T=500 | VMI C=20 T=500

Original Elephant T=31.548 | QSlim T=900 | VMI C=20 T=900

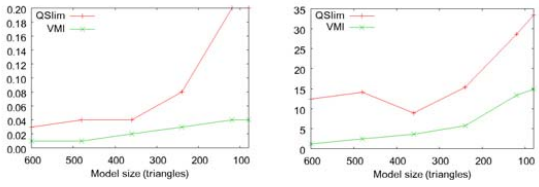
Experiments VMI



Model	Triangles		RMSE		Metro		Time	
	Original	Final	QSlim	VMI	QSlim	VMI	QSlim	VMI
Shark	734	80	33,41	**14,65**	0,20	**0,04**	0,02	10,20
Galo	6592	500	12,40	**8,38**	0,05	**0,01**	0,08	142,24
Greekship	9510	600	17,20	**12,85**	0,21	**0,09**	0,11	246,72
Tree	11.136	600	20,73	**16,60**	**0,11**	0,13	0,20	332,49
Hammer	13.380	500	8,99	**7,43**	**0,03**	0,04	0,20	423,05
Elephant	31.548	900	25,32	**11,60**	0,08	**0,03**	0,52	2309,79

Experiments VMI

Geometric Error | **Visual Error**



Comparison at several degrees of simplification of the **Shark model**

Kullback-Leibler

- Definition
  - The f-divergences quantifying the degree of discrimination between two probability distributions
  - Kullback-Leibler distance

$$KL(p \mid q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$$

  - Viewpoint Kullback-Leibler distance

$$KL_v = \sum_{i=1}^{N_f} \frac{a_i}{a_t} \log \left(\frac{\frac{a_i}{a_t}}{\frac{A_i}{A_T}} \right) \qquad A_T = \sum_{i=1}^{N_f} A_i \qquad a_t = \sum_{i=1}^{N_f} a_i$$

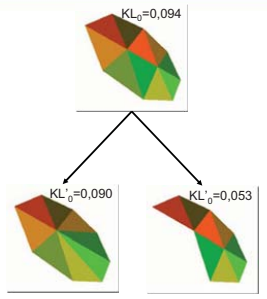
Where a_i is the projected area of the polygon i, A_i is the actual area of the polygon i and A_T is the total area of the object

Kullback-Leibler

- The error metric
  - Defined as the sum of variations of Kullback-Leibler distance for all viewpoints V

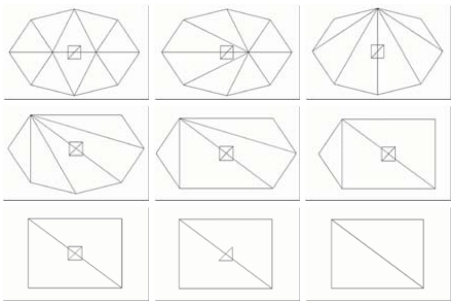
$$c = \sum_{v \in V} \left| KL_v - KL'_v \right|$$

  - The cost of the algorithm is higher than Entropy or Mutual Information due to the A_T computation
  - Hidden polygons will be removed according with their actual area



Kullback-Leibler

- Simplification example using KL_v

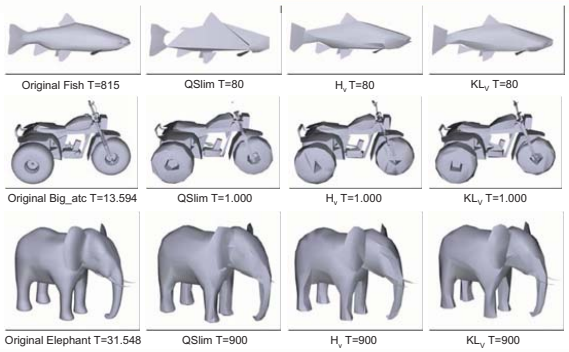


Experiments KL_v

Model	Triangles		RMSE			Metro		
	Original	Final	QSlim	H_v	KL_v	QSlim	H_v	KL_v
Fish	815	100	22.83	**11.57**	12.98	0.09	**0.03**	**0.03**
Galo	6.592	500	12.40	**9.34**	10.48	0.05	0.03	**0.01**
Al Capone	7.124	1.000	17.66	**11.47**	12.07	0.03	0.08	**0.03**
Tree	11.136	600	20.73	**16.98**	18.04	0.11	0.13	**0.04**
Big_atc	13.594	1000	16.50	15.97	**15.44**	0.08	0.05	**0.03**
Elephant	31.548	900	25.32	**13.18**	13.40	0.08	0.14	**0.05**

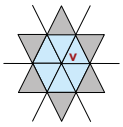
Model	Time		
Original	QSlim	H_v	KL_v
Fish	0.03	10.01	11.31
Galo	0.08	141.75	237.30
Al	0.08	150.90	273.18
Simpletree	0.20	332.49	605.49
Big_atc	0.27	535.23	835.88
Elephant	0.52	2197.67	4016.78

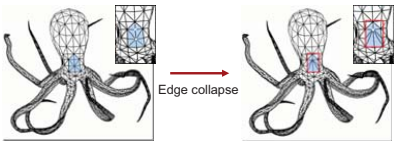
Experiments KL_v



Simplification Algorithm

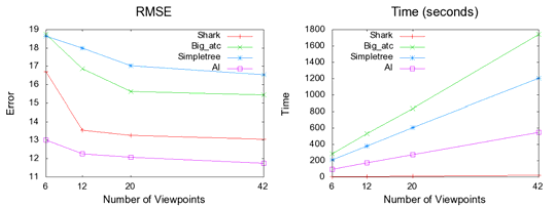
```
/* Update the mesh */
while ( heap h not empty )
  Remove from heap h the edge e with lowest Ce
  Perform collapse e
  for ( each e' in neighborhood )
    Compute collapse cost Ce'
    Update (e', Ce')  location in heap h
  end for
end while
```





Simplification Algorithm

- Analysis on the number of cameras using Mutual Information



Conclusions and future work

- New viewpoint-driven simplification metrics based on Information Theory has been proposed
- The metrics will be improved incorporating attributes (textures)
- We are working to reduce the computation time, although the simplification is an off-line process

EG:929

Visual Mining of Text Collections

Eurographics 2007 Tutorial T11

Organizers & Speakers

Rosane Minghim (University of Sao Paulo, Brazil)
Haim Levkowitz (University of Massachusetts Lowell, USA)

EG:930

Visual Mining of Text Collections

Rosane Minghim[1] and Haim Levkowitz[2]

[1]ICMC - Instituto de Ciências Matemáticas e de Computação
University of São Paulo
Brazil
[2]Institute for Visualization and Perception Research
University of Massachusetts Lowell
USA

Abstract

What happens if you have to examine and reach conclusions from a considerable number of textual documents? If you are faced with this task or with developing tools for completing this task, this tutorial is for you.
Examining text is crucial for many different types of applications. Even applications that rely on additional types of data (such as images, signals, simulations) usually have complementary or alternative text based output. The challenge of interpreting content and extracting useful information from a document collection is the target of efforts in various areas of computer science. Fields such as Text Mining try to extract knowledge automatically or semi-automatically from collected textual information; however, due to the multi-dimensional characteristics of text it is paramount to couple these algorithms with meaningful visual representations in order to improve performance and allow the discovery of relevant information within a text data set.
Since it is not feasible to go through the entire documents' content in detail due to data sizes and time constraints, Visual Text Mining (VTM) — the combination of Text Mining and Visualization — is focused on developing tools to help users extract meaning from text collections without extensive reading.
In this tutorial we introduce the necessary background and the graphical techniques involved in Visual Text Mining of document collections.

Contents

1. Overview, motivation, goals

Text collections are generally considered data sets with a high number of dimensions, where a dimension is a term or expression (an n-gram) of importance within the domain of the document collection. In a conventional vector representation of a document data set, the number of dimensions for a few hundred documents of moderate size can reach a few thousand.

Regardless of these difficulties and their implications for the analysis of text collections, there is a growing number of applications that can benefit from tools to effectively support analysis of document sets and from that analysis, allow the user to reach conclusions and to make decisions. The range of potential applications vary from health studies and diagnosis based on medical records to investigation of unlawful activities. The nature of the textual source is also quite varied. Documents can be snippets from web searches, RSS feeds of various kinds, scientific papers, reports, newspapers articles, automatically generated health test reports, patents, and so on. Every day new applications start to rely on text solely or combined with other data sources (such as table data and images). There is, therefore a compelling need for tools that combine user driven and automatically extracted displays to support the analysis of text collections and relationships amongst texts.

Finding ways to support analysts in quickly and meaningfully extracting meaning from text collections is also a strategic issue for the area of Visual Analytics, defined as "the science of analytical reasoning facilitated by interactive visual interfaces" (from http://nvac.pnl.gov/). The fields of Text Mining and Visual Text Mining aim at combining visualization and mining approaches to achieve solutions for the exploration of text collections.

Handling high-dimensional data poses many problems to researchers and data analysts in general. Visualization techniques seek to bridge the gap between users' visual perception and reasoning capabilities and analytical techniques. Nonetheless, finding intuitive ways to visualize large high-dimensional data sets is a difficult problem. Traditional multidimensional visualizations, such as scatter plot matrices, parallel coordinate plots or pixel-based techniques [OL03] operate by mapping each data attribute into a corresponding visual axis or other representation. As such, they can only handle well a very limited number of attributes and, therefore, are not directly applicable to complex data consisting of many attributes (dimensions) [MPL06].

A widely explored alternative to handle this type of data is to reduce dimensionality prior to visualization, e.g., by projecting the high-dimensional data points into a lower dimensional space (2D, 3D) that is more amenable to user interpretation. Various techniques exist to achieve such a mapping, mostly based on dimensionality reduction, dimensionality clustering, or point placement strategies. The resulting projection can be displayed through a suitable visual representation — e.g., points placed on a plane, graphs, surfaces or volumes — that can be navigated and explored by a data analyst. Projections may be created based on different criteria, but typically they strive to preserve distance relationships amongst data points as defined in the original space. Information loss is inevitable and the extent to which this distance-preserving goal is achieved depends on the precision of the projection.

Figure 1 depicts the overall process in generating interactive visual representations of document data sets via projections (or, in fact, other point placement strategies too). The user can feed the system with (1) a document collection, (2) a structured data table, or (3) a distance matrix. In the first case, documents are either converted into a vector representation or compared using a similarity measure that handles text against text. From vector representations, distance calculations can also be performed. From distance relationships the 2D projection is generated and the user can interact with its visual representation to gain insight to the data set.

This tutorial treats the problem of mapping and exploring, through visual representations, textual data sets and results of information extraction from textual data sets of various different natures. We describe the basic principles for generating such visualizations and make an effort to convey the benefits of achieving a solid graphical view of automatically extracted relationships. The problem of facilitating text analysis is by far an open issue. We hope to identify, suggest and explain the current technologies available and the challenges ahead for professionals dedicated to both using and developing techniques for that purpose. In this tutorial, we offer an overview of the existing techniques underlying techniques to allow the construction of visual document maps from unstructured data sets, that is, without relying on previously extracted meta data. The resulting map should offer insight into the contents of the document data set to support examination and relief the user from extensive reading otherwise necessary.

We begin in the next Section by illustrating the creation and exploration of visual maps of example text collections of very different natures in two quite diverse areas of application, by use of projection techniques.

Following that we describe, in Section 3, some of the most important basic concepts and techniques involved in this area.

Section 4 reviews the techniques published in the literature to achieve some of the goals stated here.

Projection based visualizations and their use for visualization of text collections are detailed in Section 5. Finally, we discuss other underlying mining issues that go together with text analysis, such as Topic Extraction.

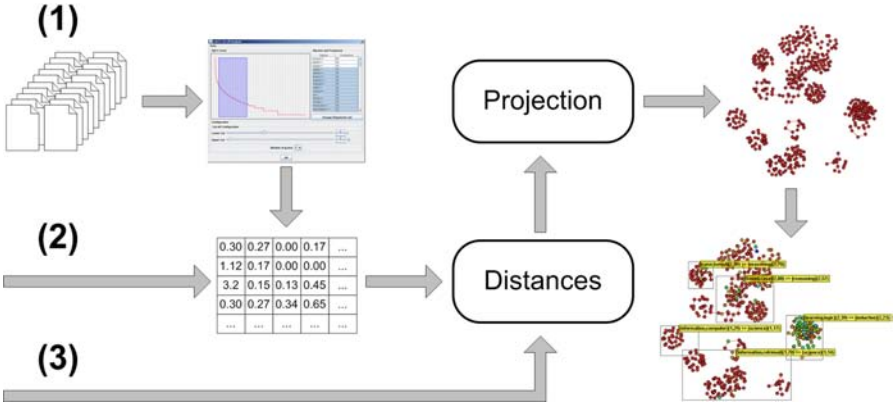


Figure 1: *Generating projection maps of multidimensional data sets.*

2. Two test cases: Visual maps of flash news and scientific papers

Visualizations of text documents and their meta-data can take many forms, whether or not they are coupled with data mining techniques. Since our focus is on non-structured data, here we illustrate the mapping based on content of two document collections bearing very different characteristics. The goal would be, through an analysis process involving such maps, to be able to extract and map features of the data set in a way that can help locate patterns. In each applications, however, what needs to be located is different. In the first case, the mapping of news flashes, the target is to have an overview of the main events that have been reported on Web sites of some relevant news agencies. The second case illustrated here, the mapping of scientific papers, highlights the possibilities of identifying general subjects as well as specific issues and finding relevant material to explore further in academic and scientific applications.

Visually, the resulting representation used here is a graph, whose nodes represent points of the data set (individual textual documents) and whose edges represent some type of relationship between the connected nodes. The graph displays and interactions presented in this Section were generated using PEx, the Projection Explorer, a freely available tool for multi-dimensional visualization via projections and point placement strategies with text processing capabilities (see http://infoserver.lcad.icmc.usp.br/infovis2/PEx).

An alternative view of the same text maps is the landscape or surface view, where vertices are points and meshes codify additional attributes by mapping scalar fields to geometric or visual attributes (height, color, glyphs, etc.). Surface views in the following text are generated using the Super Spider tool, a recent evolution of its precursor, the Spider Cursor [MLN*05], that can be used to explore landscape plots.

The visual representation employed here, whose construction will be described along the text and slides of this Tutorial, can also be the result of various other tools for multidimensional data projection (see Section 4.2, except, maybe, for particularly visual attribute mappins and for the flying quadric in Figure 5.

2.1. Visual mapping of news flashes

The first data set we shall explore was built by gathering RSS feeds of news flashes from the WWW sites of four news agencies: Associated Press, BBC, CNN and Reuters. They represent all the news made available as RSS during approximately 24 hours in April 2006. After some elimination of duplicates, the gathering process resulted in 2,684 news files. For the visual mapping, we have used title and content only, with the goal that similar news (supposedly news treating the same subjects) are mapped to the same neighborhood. For the exploration after mapping, the full file can be used, which contains also the date and the source of the news (the name of the news agency).

The size of the text in an RSS feed (which, for many applications, synthesizes a larger file) is usually quite small, typically 25 to 80 words only. This test case bears resemblance to many applications available today that wish to analyze RSS feeds of all sorts of information (for instance, discussion groups, support groups, article repositories, patent repositories, dictionaries, libraries, and thousands of other such repositories).

Figure 2 shows pictures of two different maps from this data set. In Figure 2(a), which employs a multi-dimensional projection technique called Least Square Projection, or LSP [PNML07, PNML06], the region labeled A is a group of news articles on the Vioxx dispute. Region B centralizes all news on an immigration bill undergoing discussion at the US senate. Group C concentrates the news on a trial related to the Enron case (a particularly important deposition was

about to take place). By coloring nodes according to the source of the news, it is also possible to identify the degree of importance a particular news agency gave to a subject by estimating the number of points with its color. Conversely, the same observation supports identification of the news agencies that dedicated a lot of — or just a little — effort to report a particular event.

Figure 2(b) shows the same map as Figure 2(a), slightly zoomed in, now presenting a set of labels that support identification of the subject mostly tackled by a particular user-selected group of documents. The set of labels in that picture, calculated by a term covariance approach (see Section 5.3.1), identify some of the central events of those two days.

Interpretation of groups in this type of projection is very similar to interpretation of clusters, that is, highly similar data is identified by proximity in the 2D display. In the very middle of the display remain the groups that have been 'unresolved' by the current context, which means that according to the strategy employed for similarity in this domain, it was not possible to distinguish them from the others. Coloring in Figure 2(b) was carried out using a hierarchical clustering method in 2D (in this case k-means with average linkage weighting). It helps distinguish regions of higher density of points (in the 'blue' end of the spectrum) from regions of lower densities (moving towards red as regions become more sparse).

Figure 2(c) shows the visual result of another map of the same data set, built using a new technique for point placement of this type of similarity-based data, called Neighbor-joining Tree (or N-J tree) [CPMT07]. An N-J tree is a type of similarity tree visualization that employs the same principles of phylogeny tree reconstruction [SN87] to reflect the organization of data points (in this case documents) according to their similarity. The interpretation of that type of visualization is slightly different than straight projections. In an N-J tree, very similar (according to the similarity measure employed) data points are gathered in the same and nearby branches of the tree. Proximity on the 2D plane itself does not imply similarity. Interpretation of similarity trees needs, therefore, the branching information to be effective. Branches guide selection and focus also; the neighborhood of a point is extracted searching up and down the tree. Groups of documents with content that is consistent (that is, that are best resolved by the similarity measure employed) are placed at the outer branches. Figure 2(c) uses covariance labels to illustrate the placement of some of the main news in the data set.

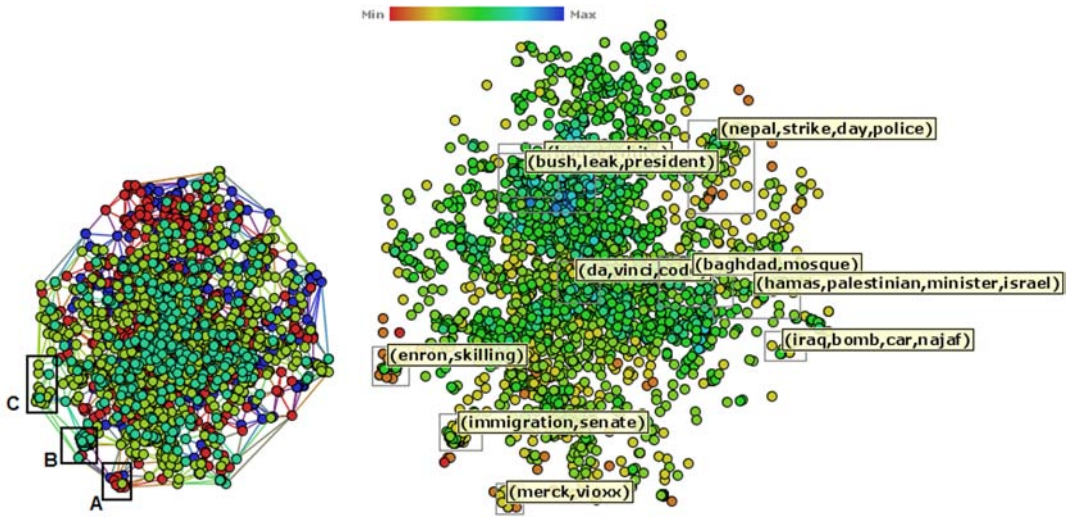
Closer examination of the flash news visualizations reveals that many of them are just headlines followed by a short text of the sort 'read full story for details'. Figure 3(a) highlights, in blue, the points where that occurs. Removing those can help unclutter the view, which was done by building a new map after removing those blue points (see Figure 3(b)). Similarly, other searches on the map can help highlight, select or eliminate specific files and groups of files.

The news data set, after the reduction illustrated in Figure 3, has 2450 news items remaining. The similarity tree by neighbor-joining of that reduced data set is presented in Figure 4(a). That picture also shows a selection of a branch for further examination. The region is zoomed in on in Figure 4(b), and further levels of topic extraction can be observed. We can see in the central part of that picture that the similarity calculations gathered news on various linked Palestinian issues in neighboring branches. These issues varied from formation of the Hamas government to Israeli strikes in Gaza to halting of financial aid from Europe and the United States. The lighter label in Figure 4(b) is the title of one of the news files in that map.

A surface view of a projection-based document map can be built by creating a mesh from the projected points and then using scalar fields attributed to vertices to reflect properties. The result of such a process is illustrated in Figure 5. Various attributes can be mapped to scalar fields, from search results (for instance the number of appearances of a particular expression) to the results of clustering, classification, and categorization of data points. In Figures 5(a) and 5(b) color reflects a 2D clustering (also by k-means) of the points in the projection, employed to help locate pockets of closely projected documents. Figure 5(c) illustrates the capabilities of simultaneous presentation of multiple scalar fields using visual and geometrical attributes. In that picture, height represents the count of the number of times the word 'Bush' appears in the document (higher points are higher counts, naturally). Each of the news sources (four in all) is reflected by a different color. That same information is redundantly reflected on the color of a 'flying' super-quadric that changes color as the user browses through the map. The visual and geometric mappings of the super-quadric can reflect properties of the currently explored point (or focus point). In the case of Figure 5(c), only its color changes, mapping the news agency that published the document in the focus point. Label by title clarifies what the news is about.

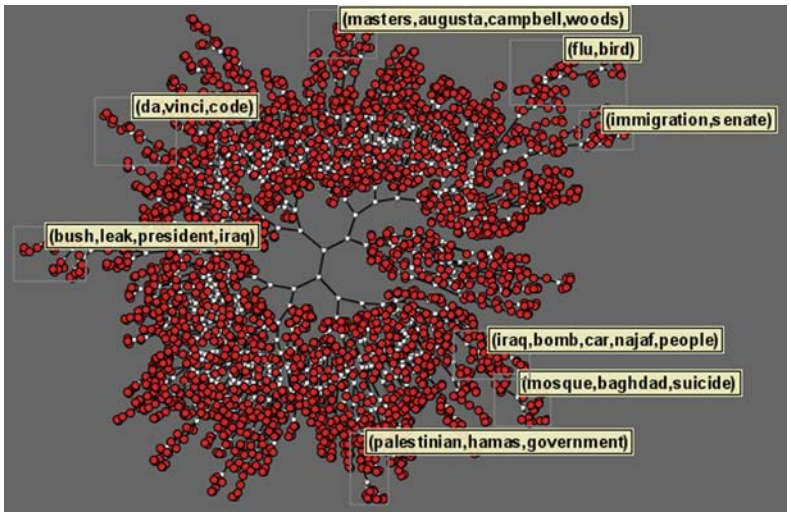
Finally, another type of projection, called ProjClus [PM06] is used to map the same news data set explored thus far (see Figure 6). Figure 6(a) shows the graph view with 2D k-means clustering for grouping location and Figure 6(b) shows the mesh view of the same map. In Figure 6(b) color is hierarchical clustering, that is, darker blue regions are regions of largest concentrations of points, down to red for sparser regions. ProjClus tends to spread the clusters around more, which is confirmed by larger regions with the same top density. Although that property sometimes confuses the boundaries of clusters, ProjClus still bears good 'proximity by similarity' precision and can help estimate the local density of a region since it diminishes cluttering compared to LSP.

Both projections thus far employed in the illustrations are



(a) LSP map. Colors represent different news sources (4 in all).

(b) Labeling of regions of map in Figure 2(a) for general topic identification.



(c) Tree-like map visualization by neighbor-joining with labels.

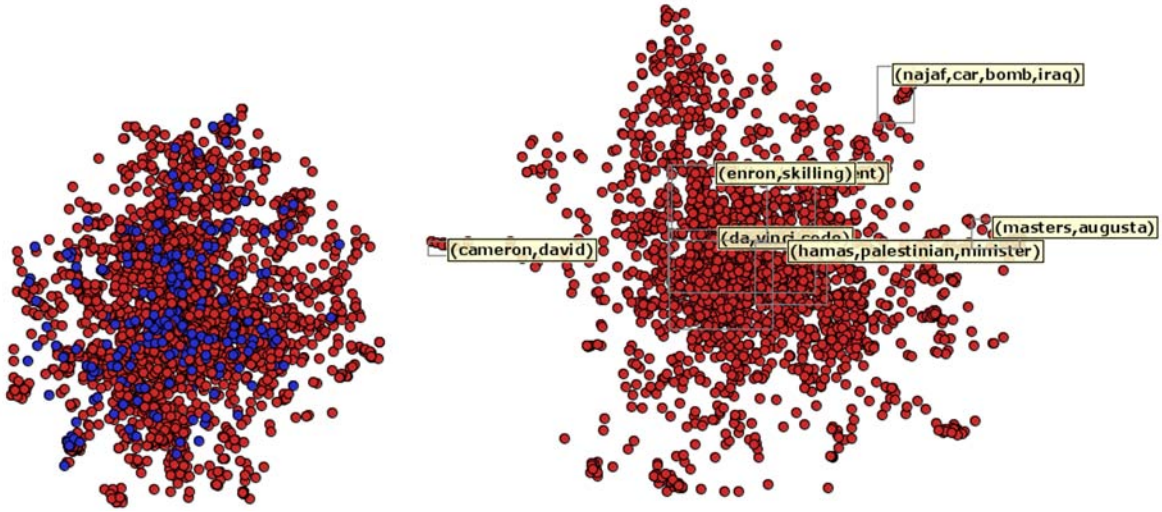
Figure 2: *Document maps of 2,684 news flashes collected for approximately 24 hours in April 2006.*

reasonably fast, leading to an environment were simultaneous examination and coordination of various maps is feasible. Similarity trees have higher computational cost but it is still manageable for data sets whose size is in the hundreds of points, even in the thousands, if similarity is pre-calculated and stored. Neighbor-joining algorithms are typically $O(n^3)$ or $O(n^2)$ at best.

2.2. Visual mapping of scientific papers

The second data set to be illustrated here is composed of 676 papers in four areas of knowledge: Case-based Reasoning (CBR), Inductive Logic Programming (ILP), Information Retrieval (IR) and Sonification (SON). CBR and ILP documents were extracted from Lecture Notes in Computer Science (LNCS) in those subjects. IR and SON papers were retrieved from internet repositories. They are pseudo-classified according to their source. The maps were built using papers' titles, authors, affiliations and references.

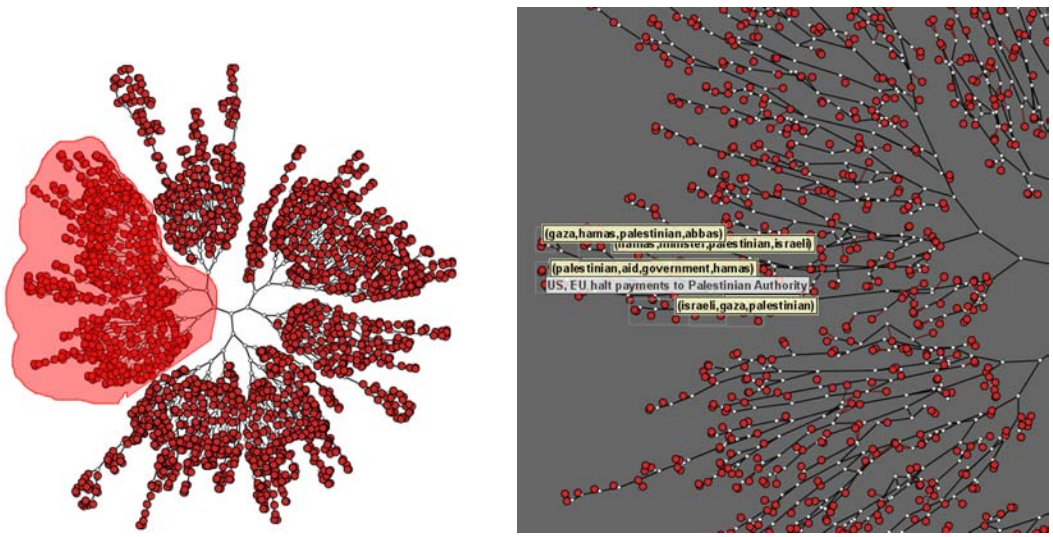
In addition to those documents, six others were added to the data set. These were papers of work published by members present and past of our research group; the goal was to evaluate their local and global positioning on the map. From those, five papers were highly related. They refer to



(a) Highlighting news with words 'full' and 'story' on map of Figure 2(a)

(b) New map without news highlighted in Figure 3(a). 2450 news files remained

Figure 3: *Using search and corpus manipulation to eliminate unwanted files*



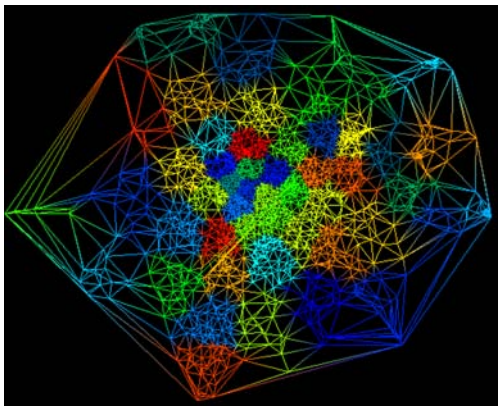
(a) N-J tree of the reduced news data. Selection of focus region.

(b) Scaled presentation of focus region and its surroundings .

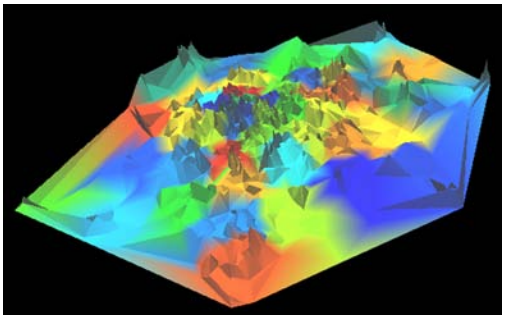
Figure 4: *Focusing on a part of a Map for further exploration*

the evolution of a sonification system and its corresponding user evaluation. Two of them refer to the first version of the system and the other three to its newest version. The last paper was also published having as authors members of our research team, but on a completely different subject (a technique used for image segmentation). Let us call those six additional papers the 'intruders'.

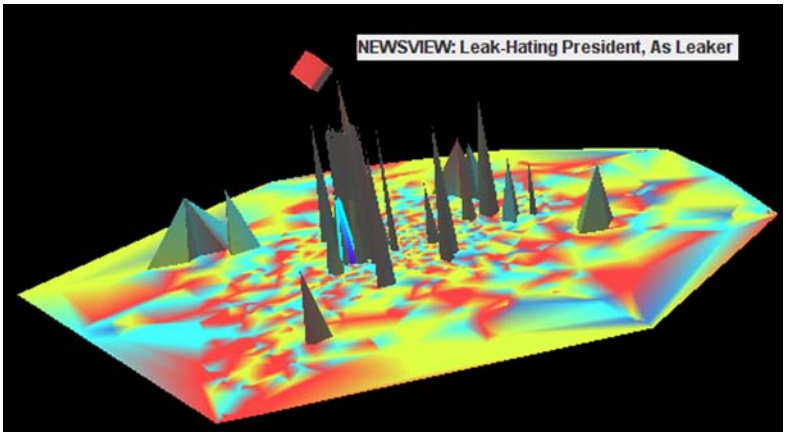
Figure 7 shows general and specific views of the papers data set, mapped by LSP. It can be seen in Figure 7(a), both by coloring and by labeling (each color here represents a pseudo-class), that the projection technique was capable of separating well this heterogeneous document collection by their subject areas. Moreover, subjects with an established and more focused body of techniques, such as CBR and ILP



(a) Mesh generated from a Delaunay triangulation of news jected points.



(b) Surface view of news set. Color and height codify the same scalar field of 3(b).



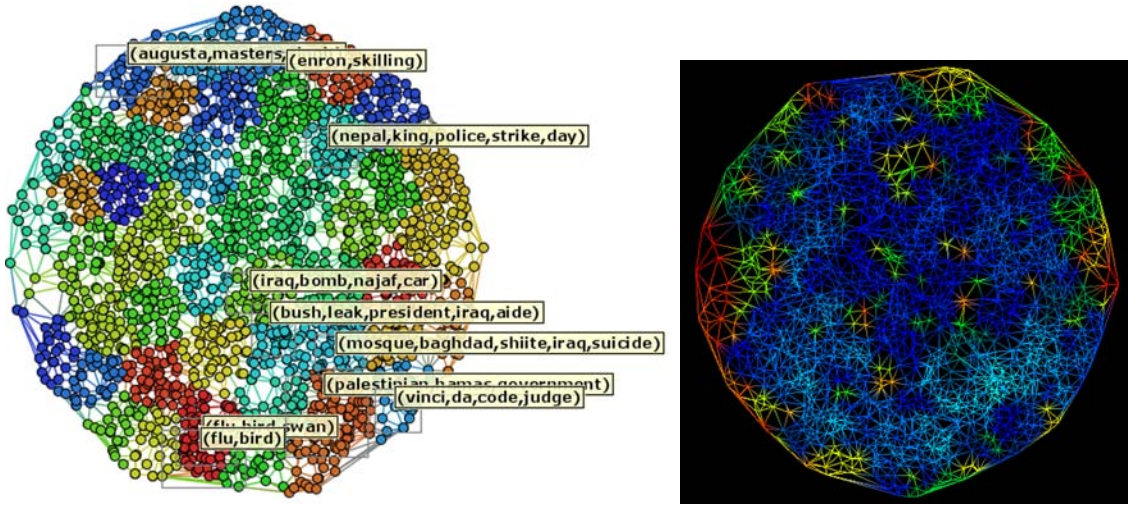
(c) Surface View. Color reflects the source of the news; height codifies the number of appearances of the word 'Bush' in that particular text; the color of 'flying' geometry on the display confirms the source of the news (i.e., the news agency here represented in red) of the focus point; the focus is highlighted by its incident edges colored in blue; label is title of the focus point.

Figure 5: *Surface representation and exploration of content by visual attributes.*

tend to be displayed less sparsely than a newer more diverse area such as Information Retrieval. In the various regions where points of different colors appear close to one another, further examination reveals that there was a high content correlation that lead to that proximity. Most of the time they are applications of techniques in those areas (CBR, ILP, IR, SON) for particular goals (such as modeling, manipulation, or study of particular data sets such as biological sequences).

To check that capability further, it can be seen in Figure 7(b) that all except one of the intruder papers were mapped in the same group. The five sonification papers share a very close neighborhood within the general sonification group, as would be desired. The sixth, unrelated paper, joined another separate group, that seemed to gather together papers from different subject areas. Closer examination by focussing on that particular cluster and loading the documents represented there (see Figure 7(c)) shows that these papers deal with various techniques (in all subject areas) designed to treat imaging tasks (such as retrieval, manipulation and segmentation). In fact all papers in the paper data set dealing with imaging techniques were placed in that group.

Connections amongst nodes on the graph can carry various meanings. For instance, they may reflect neighborhoods within the data set. Figure 8 illustrates the display of similarity relationships also as edges on the graph. In Figure 8(a) documents are connected according to the similarity measure employed for the creation of the map. Each point is connected to its next neighbor in the similarity matrix. It can be



(a) ProjClus map, which spreads groups on the 2D plane; color reflects 2D clustering

(b) Mesh view of ProjClus news map; color reflects levels of concentration of points.

Figure 6: *Map by ProjClus of the reduced news set.*

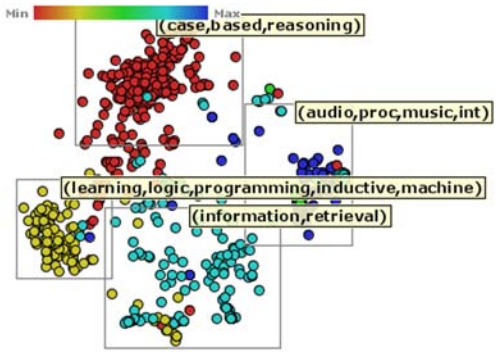
seen that there are a few points whose next neighbors are not within the same general group. That is to be expected, since the main groups are not at all disjunct as far as content similarity is concerned. Still, the small number of 'crossings' outside the clusters indicates the good degree of grouping presented by the projection. New connections can be added. For instance, each point can be connected to its two nearest neighbors in the original data space (see Figure 8(b)); in this case more connections across groups appear, and they can reveal further associations apart from those already revealed by grouping and first neighbors, adding to the possibilities of exploration.

Many other relationships could similarly be mapped to edges on a map. Focussing on parts of the map shows in further detail the subjects within groups. Detail of two groups of the map (see Figure 8(d)) reveals that, within the general sonification group, there are separate groups for audio processing (including retrieval), computer music, sound in user interfaces, data sonification, and sonification for the visually impaired.

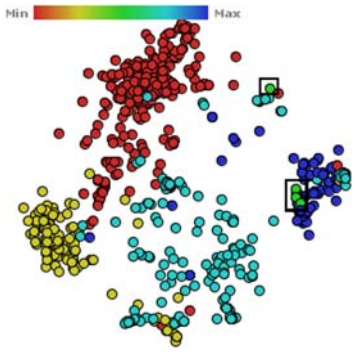
Display on the 2D plane of multidimensional data can be done using a number of techniques. As well as projections, other point placement strategies can be employed. Sometimes it is useful to arrange (or rearrange) a projection according to the connections amongst points. That is illustrated in Figure 9(a) and 9(b). Other important issues in document map exploration are topic extraction and visualization, which help the user find interesting groups of documents to explore further. Figure 9(c) illustrates that using a proper technique (selective generation of association rules, in this case), further detail of topics approached by documents can be viewed.

Finally, we map the scientific paper data set using N-J trees. Since the tree reflects an organization of the similarity measure, it reveals subtleties not easily identifiable in projection layouts. For instance, in Figure 10(a) and 10(b), just by having two different approaches to display the same tree, it is possible to locate different levels of grouping and separation of groups. Additionally, in the same pictures it can be seen that the positioning of the 'intruder' papers in the same neighborhood happens in two levels for those dealing with sonification. The two papers referring to the first version of the sonification system are gathered in the same branch while the others are together in another branch nearby. Figure 10(c) shows an N-J tree map built from another similarity measure, in this case Normalized Compressed Distance (NCD) [TMP07, CV05], see below for more details. Overall NCD gives similar separation of greater groups for this same data set. However, close examination shows that not all relationships are maintained. For instance, in that map all the intruders were, according to that NCD, deemed more similar than using vector representation. This type of presentation gives a good reflection of how a particular similarity measure behaves with a data set.

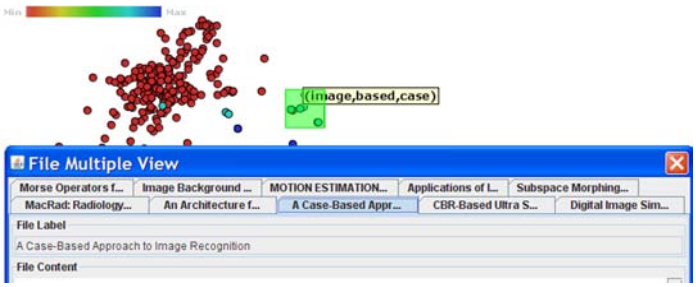
The cases shown here illustrate the use of projection-based displays together with mining techniques (such as clustering, covariance term extraction and association rules extraction) in an integrated manner, with combined available technology to build maps for exploration of text collections. These and other tools put together can support improvement



(a) LSP map with main topic labels, automatically extracted.



(b) 'Intruders' papers, five on sonification and one on a segmentation technique, are displayed in green in the highlighted regions.



(c) File content view of papers in group that gathers techniques for imaging.

Figure 7: *Map of scientific papers in four classes, CBR, ILP, IR and SON, using LSP.*

in examination of larger text collections without extensive reading.

The next section starts detailing concepts and techniques related to the construction and exploration of such maps.

3. Basic Concepts

3.1. Text processing and information retrieval

3.1.1. Introduction and topics

We summarize the fundamentals of text processing and information retrieval here. The main topics to be explored are:

- Modeling for information retrieval
- Retrieval evaluation methods
- Query languages and operations
- Text and multimedia languages and properties
- Text operations
- Indexing and searching

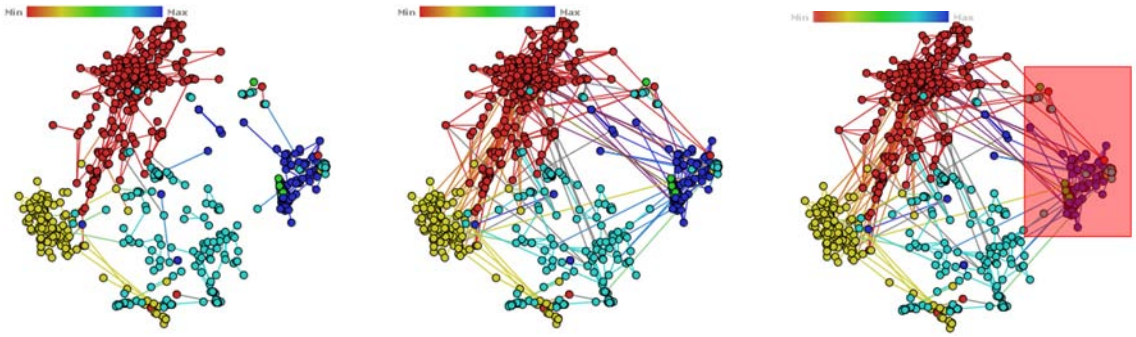
Whereas this include non-text aspects, such as multimedia languages and properties, we leave those here to provide a more comprehensive picture of the subject.

3.1.2. Motivation

Information Retrieval (IR) has become a major factor in today's information-centered world. Whereas in the past, only very technical people used to interact with data sources — actually, mostly databases — today lay people used data sources on a daily basis. Every Web search, every Web-based purchase, and many other application on- and off-line, all focus on data sources and data bases, whether the user is or is not aware of it.

3.1.3. Information versus data retrieval

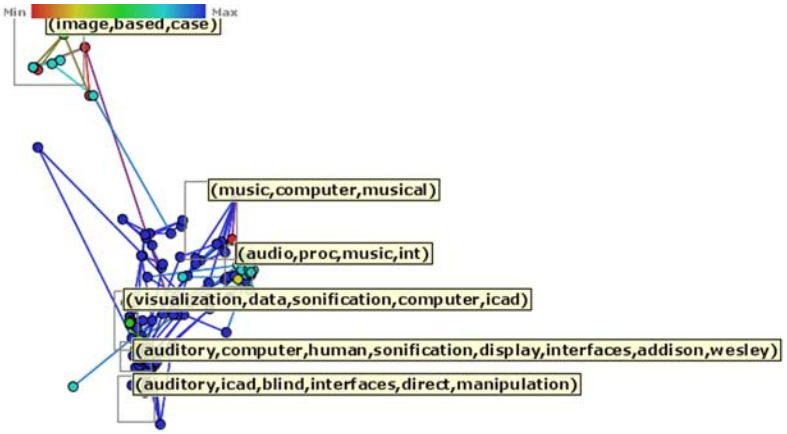
3.1.3.1. Data retrieval : Data retrieval interacts with databases, which are sources of structured data. Databases do not usually contain broad information about a specific "subject or topic", but, rather, store specific data in structured tables, such as catalogs, populations data, and the like. Retrieval from a database requires some knowledge of a query language — the most prevalent variations on SQL, the Standard Query Language. Data retrieval will only succeed if the database includes specific instances of precise query terms. A slight misspelling can make the difference between hit and miss, between success and failure.



(a) Connection by 1-neighboring points according to similarity measure.

(b) Connection by 2-neighboring points according to similarity measure.

(c) Selection of part of the map



(d) Focus on the part of the Map selected in Figure 8(c), with labeling by covariance

Figure 8: *Maps of scientific papers with connection representing similarity in the original space.*

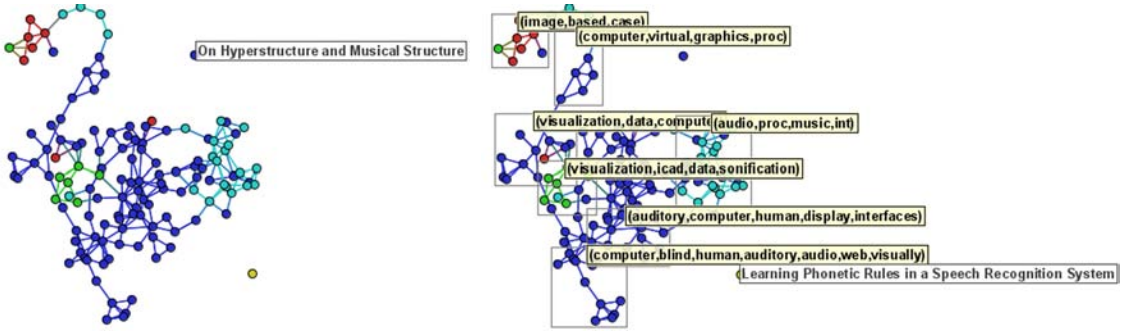
3.1.3.2. Information Retrieval (IR) : IR, on the other hand, focuses on retrieval from unstructured of documents that are related by topic or subject. Query terms are not necessarily expected to be present in their exact form in a document for it to be returned in the result list. Instead, IR systems attempt to interpret the contents of information items — mostly, text documents in a collection. This interpretation involves a comparison between the entered query and the documents in the collection, computation of a "difference" or "distance" measure, and ranking of the returned documents according to their relevance to the user's query.

3.1.4. Information retrieval at center of stage

Information Retrieval has been utilized by librarians and other information professionals for close to three decades. However, with the penetration of the World-Wide Web, it is now one of the most important tools a wide variety of users utilize in their day-to-day interactions with information. The Web has become the single largest repository of, as well as the unified interface to most of the information people have to interact with. However, the Web presents a significant obstacle to effective and efficient information harvesting, namely, the absence of a well-defined underlying data model for Web-based information. This typically leads to information definition and structure that is frequently of low quality.

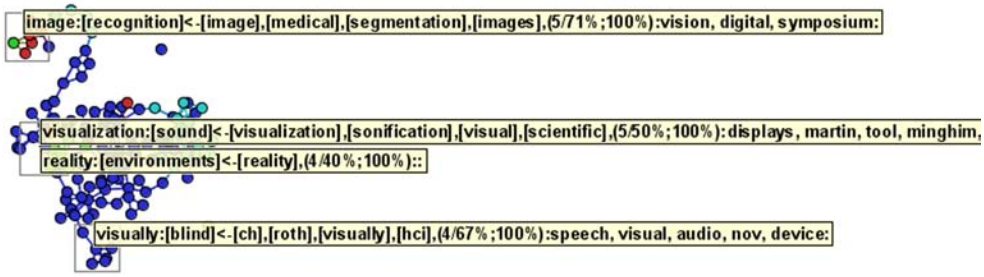
3.1.5. Basic concepts of IR

The IR process starts with a user's need. The user's initial task is to translate her information need into a query, specified in a language that is provided and understood by the system. We have all grown accustomed to entering multiple daily queries into our favorite search engine. If the user's query is entered into an IR system, she typically would specify a set of "key words", which are expected to convey the semantics of the information needed. Using a more rigid data retrieval system, a much more rigorous query expression is required, such as a regular express ion, which contains constraints to be satisfied by objects in the answer set. In both cases, the user searches for some useful information executing a retrieval task.



(a) Rearranging the position of the subgraph of Figure 8(c) by using a force-directed placement strategy.

(b) Labeling by covariance of main topics of detail map.



(c) Detailed topic extraction by association rules [LPMP07].

Figure 9: *Exploration of a sub-graph with rearrangement and topic extraction and visualization.*

A schematic overview of the basic iterative retrieval process, the logical view of a document, and a more detailed picture of the retrieval process, can all be found in the accompanying slides.

3.1.6. A brief history of Information Retrieval

Here, we summarize the history of Information Retrieval, covering the following topics:

1. Early developments;
2. Information Retrieval in the library;
3. the Web and digital libraries.

3.1.7. Early developments

For roughly 4000 years, people have been organizing information to be able to later retrieve and use it as they need. One of the most common examples is the familiar Table of Contents found in many books and other text documents. After Gutenberg's invention of the moveable printing press in 1455, which has made books available and affordable to the general population, rather than just to the wealthiest few who could afford them, the volume of information people have acquired grew beyond just a few books. This, in turn, triggered the development of specialized data structures for faster access to stored information. For example, an old popular data structure for faster IR still found in many books is the index.

An index is a collection of selected words and concepts that serve as associated pointers to related documents and other information sources. Indices are at the core of any modern IR system. They provide faster access to data and they speed the task of query processing. For centuries, indices (aka indexes) were created manually, according to categorization hierarchies. Today, libraries still use categorical hierarchy — which have usually been conceived by human subjects from the library sciences field — to classify volumes and documents.

Computers have lead to the automatic construction of large indexes, which has lead to another view of the retrieval problem, one that is much more related to the system itself than to the user's needs.

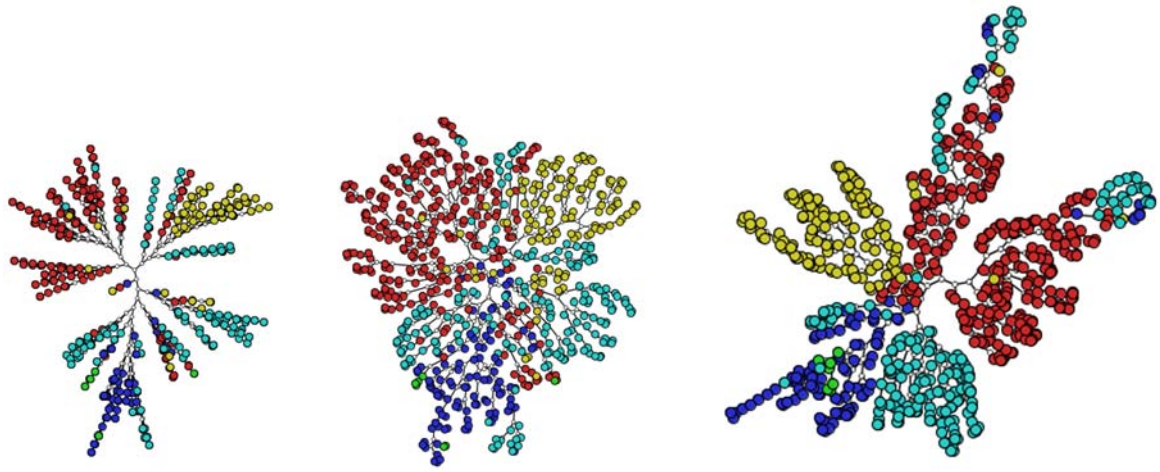
We, thus, face two views of the IR problem, a Computer-centered view, and a human-centered one.

3.1.8. A computer-centered view of the IR problem

As far as computer systems are concerned, the IR problem has the following components:

- build efficient indexes;
- process user queries with high performance; and
- develop ranking efficient algorithms.

All, with one goal in mind: to improve the 'quality' of the retrieved answer set .



(a) N-J tree based on vector representation using cosine distance; initial position

(b) Vector based similarity tree, after repositioning the nodes.

(c) N-J tree from NCD similarities.

Figure 10: *Similarity trees for the scientific papers data set.*

3.1.9. A human-centered view of the IR problem

From the human user's point of view, the IR problem can be summarized as:

- Study the typical behavior of users;
- understand their main needs; and
- determine how such understanding affects the organization and operation of retrieval systems.

From this point of view, a keyword-based query processing might be seen as unlikely to yield a good solution in the long run, leading to the quest for other, more promising alternatives.

3.1.10. Information Retrieval in the library

Libraries were among the first to adopt IR systems. Those systems were developed initially by academic institutions but later by commercial vendors started offering more refined tools and systems. We observe first, second and third generation features in those systems.

3.1.10.1. First generation . The first generation of library IR systems was basically an automation of previous technologies, mostly card catalogs. They allowed searches based on author name and title only.

3.1.10.2. Second generation . Here, increased search functionality added search by subject headings, keywords, and some more complex query facilities.

3.1.10.3. Third generation . More recently deployed, these systems focus on improved graphical interfaces, electronic forms, hypertext features, and open system architectures.

3.1.11. The Web and digital libraries

The Web has become the primary source of all kinds of information. With the spiral growth of the amounts of information available on the Web, search engines have become among the most important Web tools — and assets. Today's search engines continue to use indexes that are very similar to those used by librarians a century ago. So what has changed?

We observe three dramatic and fundamental changes: First, access to various information sources has become a lot cheaper, reaching wider audiences than ever possible before. Second, advances in digital communications have lead to easier, cheaper, and thus greater access to networks, allowing distant, quick access to vast amounts of information throughout the world. And, third, the freedom to post whatever one would like to, has caused the Web to become ever more popular. For the first time in history, most people have an almost free access to a large publishing medium, making the Web (and modern digital libraries) a highly interactive medium, facilitating the exchange of messages, photos, documents, software, video; making interactive chatting convenient and at low (or no) cost, cause overall a fundamental shift in the current communication and information flow paradigm.

3.1.12. The future: three main questions

We are still not through; we do not have all the knowledge, all the answers. The following still need to be addressed:

First, despite the high interactivity, people still find it difficult — impossible at times — to retrieve the information that would be relevant to their information needs, leading to the question "which techniques have the potential to yield retrieval results of higher quality?"

Second, ever increasing demand for access is causing the need for quicker response to be more and more a pressing factor. We therefore need to find out which techniques can support faster indexes and smaller query response times? And,

Third, the quality of the retrieval task is greatly affected by the user's interaction with system, so "how can better understanding of the user's behavior affect the design and deployment of new information retrieval strategies? "

3.1.13. Practical issues

The following issues have also great influence on the entire IR problem and solution set.

- Electronic commerce, a major trend on Web;
- security;
- privacy;
- intellectual property rights and publisher responsibilities;
- internationalization; and how to deal with multimedia (images, sound, video, etc.)

3.1.14. Overview of the rest of IR topics

There are many components to IR. We list here the major topics; these remain, however, outside the scope (and time and space limitations) of this tutorial. The reader is encourage to explore these further. A very comprehensive treatment — and, thus, a good place to start — would be [BYRN99].

1. IR modeling;
2. retrieval evaluation;
3. query languages and operations;
4. text and multimedia languages and properties;
5. text operations;
6. indexing and searching;
7. parallel and distributed IR;
8. user interfaces and visualization; and
9. searching the Web

3.1.15. Text preprocessing for IR, mining, and visualization

In order to reduce the space dimension and to allow refining the text model, text is processed prior to extraction of the vector representation. This process typically involves three steps: (i) removing stopwords, i.e., non-informative words, such as articles, prepositions and such, plus any words known to lack relevance to the context; (ii) stemming, which reduces words to their radicals (e.g., 'motivation', 'motivate', 'motivating', would be all reduced to 'motivat'); and (iii) frequency counting, so as to remove terms that occur too sparsely or too often and hence have little differential capability. Setting suitable frequency thresholds for discarding terms typically demands user knowledge and interaction. Corpus vector representation enables handling the map generation problem as one of mapping objects defined in a high dimensional space into a 2D (or 3D) visual representation space. Clustering and projections are typical approaches to handle the problem. Determining vector similarity — a necessary step for both processes — requires defining a metric in the high dimensional vector space that allows computing distances between vectors. The distance between any pair of document vectors is a measure of their proximity, or semantic similarity. Distances for all pairs can be stored in a triangular distance matrix of dimensions *nxn*, where *n* is the number of documents.

An alternative to the vector space model has been proposed based on Kolmogorov complexity approximation [TMP07], called Normalized Compressed Distance (NCD). NCD computes a distance between a pair of documents straight from their textual content, rather than from the intermediate vector representation. The Kolmogorov complexity of a string x, denoted $K(x)$, is the size of a description of x produced by an optimal specification method. The conditional Kolmogorov complexity $K(x|y)$ of x with respect to another string y can also be defined as the amount of information that x does not have about y.

An information distance between two given strings can be specified based on their conditional Kolmogorov complexity. Though the problem of evaluating a string's Kolmogorov complexity is non-computable, a solution can be approximated through compression; this yields the basis for the definition of NCD. We include a technical report in Appendix A on a formulation of NCD for text collections at the end of the Tutorial notes.

Other alternatives for pairwise similarity calculation, particularly those based on information theory, have also being studied recently [AF03].

3.2. Data and text mining

3.2.1. The knowledge discovery process

Data Exploration is the process of searching and analyzing databases to discover implicit but potentially useful information, mostly for the purpose of supporting the decision making process. The goals of data exploration are:

- Convey information;
- discover new knowledge; and
- identify structure, patterns, anomalies, trends, and relationships.

Major Data Mining Tasks, such as, summarization, association, classification, prediction, clustering, and time-series analysis use major techniques, including Linear Regression Trees, Non-Linear Regression, MARS, Naive Bayes, K-Means and K-Median, Neural Networks, Association Rules,

Decision Trees, Principal Component Analysis, Support Vector Machines, and Genetic Algorithms. These, in turn, are based on statistical tools, such as, Missing Value Imputation, Normalization techniques, Error & Variational Analysis, and Confidence Estimates, to list just a few.

3.3. Projection techniques

Data sources have increased substantially both in size and complexity, but extracting useful information from them is still a challenge. One measure of data complexity is the number of attributes associated with each instance of data. Consider, for example, data from a demographic census: a data instance records attributes such as age, sex, education, occupation, income, and so forth. Considering each data attribute as a data dimension, if we have m such attributes each data instance can be interpreted as an m-dimensional vector placed in an m-dimensional definition space.

In traditional statistical analysis, data instances with three or more dimensions are known as multivariate or hypervariate data. In Information Visualization such data is usually referred to as multidimensional. Conventional methods, such as scatter plots or bar charts, normally employed to assist data interpretation, are not directly applicable to multidimensional data. Moreover, identification of patterns and models grows more difficult as dimensionality increases, and lack of proper representations can severely impair interpretation.

A common way to handle dimensionality is to reduce the number of dimensions, so that strategies that are known to work well with low-dimensional data can be applied. *Multidimensional Projection* techniques are one example of such a strategy. A multidimensional projection technique typically maps the data into a p-dimensional space with $p = \{1,2,3\}$, while retaining, in the target space, some information about distance relationships among the data items in their original definition space. In this way, a graphical representation can be created to take advantage of the human visual ability to recognize structures or patterns based on similarity, such as clusters of elements.

Formally, let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of m-dimensional data, with $\delta(x_i, x_j)$ a dissimilarity (distance) measure between two m-dimensional data instances, and let $Y = \{y_1, y_2, \ldots, y_n\}$ be a set of points in a p-dimensional target space, with $p = \{1,2,3\}$ and $d(y_i, y_j)$ a (Euclidean) distance between two points of the target space. A multidimensional projection technique can be described as a bijective function $f : X \rightarrow Y$ that seeks to make $|\delta(x_i, x_j) - d(f(x_i), f(x_j))|$ as close to zero as possible, $\forall x_i, x_j \in X$.

3.4. Visual representations: graphs, surfaces, volumes, triangulations

Many visual representations have being utilized to represent data. Among the most commonly used ones one finds maps, graphs and networks, surface- and volume representations. The use of maps is so widespread that it is not necessary to discuss them, as every one knows their utilization for geographical applications. The concept has been used also to represent data in which some form of "proximity", whether in physical space or as determined by other measures, is needed to be presented.

Whereas maps are particularly useful to represent physical proximity or its conceptual equivalent, graphs and networks have been primarily utilized where the relationships among participants are to be emphasized. In these types of situations, measurable physical is not important; what is important is how nodes are related to each other. For example, the communication patterns among members of an organization can be nicely visualized by a graph, in which a link between two nodes represents the communication between them. A directed link can show who initiated the communication. A network of such communications can quickly show influential figures, such as "authorities" (those who are considered knowledgeable about a variety of subjects) and "hubs" (those are know who to seek when a question needs to be answered). The mapping of authorities and hubs have been the cornerstones of the original "page rank", the algorithm that provided the initial implementation of Google's search engine. In addition, graphs and networks are common representations for processes in which transitions occur between states, under certain conditions.

Spatial data — and, at times, data that can be mapped to a spatial representation — have been often represented by volumes, or their surfaces. Volume representations preserve the values of data throughout the volume, whereas surface representations discard internal values and only preserve values of the external surfaces selected to be displayed.

The two most crucial aspects of either are the data structures utilized to represent the data, and rendering techniques to display the data on a screen. Multiple data structures have been proposed for storing data of three (or higher) dimensions. Among them, in addition to the straightforward cartesian coordinates, are various hierarchical representations, including trees of various type. The discussion of such data structures are beyond the scope of this tutorial; the reader is referred to the standard data structure literature for details.

Surface representations have included various triangulation techniques, as well as cuberille approaches. Triangulation techniques approximate the surface with a mesh of triangles that attempt to follow the surface as closely as possible. Cuberille techniques try to approximate the surface with a mesh of squares rather than rectangles. Some use graph concepts to implement optimal coverage.

The display of three- or higher-dimensional data on a two-dimensional display requires projection first. Surface and volume rendering techniques have been covered extensively in the computer graphics and visualization literature. The

reader is referred to this literature to familiarize her/himself with the technical details entailed.

The accompanying slides provide some examples of various of these techniques.

4. From Visualization to Visual Text Mining

4.1. Visualization techniques for multidimensional data

"One picture is worth a thousand words" is an old cliche, but it is based in the fundamentals of human perception and processing of information. Sixty percent of all input for decision making comes into the brain of a normal-visioned person from the visual system. It has been demonstrated that, in the event of contradicting information between the visual input and that of another sense (e.g., tactile), the visual stimulus with "win" over the other one, even if the other sense is providing correct information and the visual input is erroneous!

Graphics and Visualization are meant to help the user see (understand), remember, compute, analyze, discover, enjoy, and much much more.

Multiple goals are served by different types of visualization tasks, and the appropriate techniques. The most common visualizations used are for the purpose of presentation of known information. Here, facts to be presented are known (though they may not represent the truth), the visualization process is to choose and tune the appropriate visualization technique, and the result is a high-quality visualization of the data and analysis to present facts (often without the author's presence). Confirmatory analysis visualization starts with some hypotheses about the data, proceeds through a goal-oriented examination of the hypotheses, and results in a visualization of the data to confirm, accept, or reject the hypotheses. Finally, exploratory analysis starts with no hypotheses about the data, proceeds with an interactive, usually undirected search for structures, trends, patterns or anomalies, and yields a visualization of the data to lead to some hypotheses about the data.

All of these utilize a wide variety of technique, some are pure visualizations, some are integrated with analysis, all utilizing to one degree or another an assortment of interaction tools, which help the user control the process and interact with it to yield optimal results.

Some pure visualization techniques include 2D and 3D Scatterplots, Matrix of Scatterplots, Statistical Charts, Line and Multi-line Graphs, Parallel Coordinates, Circle Segments, Polar Charts, Survey Plots, Heatmaps, Height Maps, Iconographic Displays, RadViz, PolyViz, and many more.

Among those that are integrated with analysis one finds Projection Pursuit, Dimensional Stacking, Sammon Plots, Multi-Dimensional Scaling, Principal Component Analysis (PCA) and Principal Curves, and Self Organizing Maps.

Interactions include Selection, Probing, Querying, Grand Tours, and Non-linear Zooms.

It would require hundreds (if not thousands) of pages and hours to discuss all of these in even a moderate amount of details, so we will limit our discussions here to only the very basic concepts, and to a few methods that are particularly relevant to our main focus, Visual Text Mining.

4.1.1. The visualization problem

The visualization problem can be summarized as: Massive amounts of data from various sources, including databases, simulations, sensors, decision systems, and more; limited screen space; and little knowledge about the human perceptual system and the process of information transfer.

> Visualization is a method of computing. It transforms the symbolic into the geometric, enabling researchers to observe their simulations and computations. Visualization offers a method for seeing the unseen. [MDE87]

Visualization now includes other data representations, such as auditory, haptic and tactile, potentially more in the future.

4.1.2. Classifications of visualization techniques

We are primarily interested in those visualization techniques that are suitable for multi-dimensional data sets, as visual text mining deals in high dimensional data.

Many classifications have been proposed for multi-dimensional data visualization techniques.

4.1.2.1. Point-based Point-based techniques map the dimensions of a data record to the attributes of a point on the display, including its size, shape, color, motion, and — occasionally — sound at the point. The most recognized point-based technique is the scatterplot. Scatterplot-based visualization techniques include OmniViz Galaxy (OmniViz, Inc.), Temple MVV Graphics [MGTS90, MTS91], [CC92], and Splat Visualizer and Scatter Visualizer (from Purple Insight).

Because the traditional scatterplot is essentially a two-dimensional data technique, various approaches have been propose to extend scatterplots to higher dimensions. These approaches include layout extensions, either via matrices or radially, and shape extension, utilizing icons and glyphs.

Scatterplot Matrices provide an xy layout of k-dimensional data in a total of $k(k-1)/2$ scatterplots, each showing two of the k dimensions against each other. Clearly, as k grows, the number of scatterplots grow too. E.g., for $k = 100$ the number of scatterplot will be 4,950. The concept has even been extended to scatterplot cubes, following the same idea.

RadVis [Hof99] is a point-based technique that organizes

data dimensions around a circle. Data dimension values function like springs, exerting forces over a data record in various directions. The result is a collection of point clouds, clustered based on dimension values.

Hyperslice is a matrix of k^2 slices through a k-dimensional data set. Slices can be determined interactively [vWvL93].

4.1.2.2. Line-based Line-based techniques include bar charts, line graphs, parallel coordinates.

4.1.2.3. Hierarchical, graphs, trees These methods utilize hierarchical structure, networks, graphs, and trees to present data that is inherently suitable for such structures. Among the best known technique one finds Eick's SeeNet [BEW95], which utilizes spoke, helix, and sphere layouts to plot networks of nodes and their relationships; SeeNet 3D [Eic96], which is an extension of SeeNet to three dimensional rendering; SeeNet ArchView, a member of the SeeNet family, which uses arches as the visual primitives; MBone [MB94]; and DBMiner [HCC*97].

4.1.2.4. Iconographic displays Iconographic displays are perceptually driven displays. They generalize the notion of a pixel to that of an icon (or a glyph). By doing so, they increase the number of parameters "displayed". Icons' visual, color, and auditory attributes are mapped to data parameters, thus making the icon "data-driven". The icons are displayed en-masse and thus harness the perceptual powers of the human early visual system. When presented, icons are available for user interaction. These methods have been successfully used in data fusion of multiple parameters / dimensions.

Among the best known iconographics/glyph systems are Chernoff faces [Che73], Andrews' graphs, Stars or circular plots (Ward et al., various), Stick-figure icons [PG88], Color and Texture [H.91], Beddows' Embedded blocks, Smith's Sound (1989-90), and numerous other examples dated during the nineties.

The accompanying slides provide many examples of these various visualizations.

4.2. Visualization techniques and systems for handling document collections

Several different techniques for visualization of textual results from Web and other searches have been proposed ([ABY03, LA00, SCL*99, BY96]). While these techniques are capable of displaying large document bodies, they tend to make the location of specific relevant reading material difficult. We focus here on complementary tools to support mapping of documents in a way that helps locate neighboring similarities between individual text documents and groups of documents. We therefore assume a pre-filtering task that reduces the universe of targeted documents to a few hundreds (maybe thousands) in a few areas of interest (not necessarily pre-determined).

Many techniques for text document analysis and visualization exist. They usually search for a representation of the content of an individual document or text fragment (e.g., [MWBF98], [RES98], [RES99]), of document collections (e.g., [LA00], [BCG*99], [Wei01]), or of themes approached in text documents (e.g., [HHWN02], [Wis99], [WTP*95]) in order to meet these goals.

The process of text document analysis and visualization usually involves three phases: (i) pre-processing; (ii) dimension reduction; and (iii) attribute mapping to (at least) a visual representation and presentation.

The pre-processing phase takes as input the document collection to be analyzed, and produces an intermediate form, usually based on the vector space model [Sal91] whereby documents are represented as points in a vector space. In this representation each document is represented by a vector whose dimensions are terms (n-grams). The vector coordinates are the weights of the terms based on their frequency of occurrence in the document. Typically, dimensions reach the thousands even for small to medium-sized databases. Transformation of a document collection into a vector space is preceded by elimination of non-influential words (such as stopwords), reduction of words to their radicals (stemming), and some sort of frequency counting (various exist).

The second phase, dimension reduction, typically involves removal of words that are either too frequent or too rare in the collection, and clustering dimensions to generate new 'combined' attributes.

The most common way to extract structure from a text document collection is by applying some dimensional reduction technique over the resulting vector representation. Systems that implement such approaches include those based on Multi-dimensional Scaling (MDS), Principal Component Analysis (PCA) or Latent Semantic Analysis (LSA), which work with statistical measures for subspace reduction, as well as Self-Organizing Maps (SOM), which employ neural computation ([Wei01], [Wis99], [BCG*99], [KHLK98], [WTP*95]).

In the third phase, such techniques can be used to reduce the dimension to two and then plot the data onto 2D space.

However, multidimensional reduction techniques can cause some difficulties, such as [HWR03]: high information loss when applied directly to two dimensions (for display); reduction in input dimensions do not seem to affect greatly the outcome; and there is an inherent discretization problem associated with techniques such as SOM, by which individual documents in groups are not distinguishable. For our goal here, dimension reduction can pose an additional problem: when used to display the results in 2D, the mappings to subspaces may define groups of 'similar

documents', but it is not possible to locally relate neighboring documents. Previously [LMMP06], LSA has been successfully applied to generate document maps with high local content relationships, but the high computational cost remains a problem as well as the handling of vector transformations. Minghim et al. [MPL06] have proposed faster mapping approaches that posses the ability to associate documents by similarity, based on dimensional reduction by fast projections (such as Fastmap [FL95] and Nearest Neighbor Projection (NNP) [TMN03]. The gain in processing time is attained by using these projection techniques, which provide an initial point placement that is prone to speed up force-based improvement schemes. Those too are based on the vector representations. Those projections provide an initial placement on the visualization plane, but also require further use of point placement strategies [TMN03] in order to recover information lost during projections.

Point placement strategies, including force-based point placement, have been used before to generate document displays [Cha96, AKS*02a, GSAK04]. They can avoid, partially or completely, the extensive calculations needed in dimensional reduction techniques by starting with a semi-random point placement and re-adjusting their position based on attraction by similarity. However, our experience trying to find relations in their application to text content has been that the precision of the placement is limited as applied to our goals for these maps.

Other major techniques to visualize inter-document similarities are document networks ([FFW91], [TC89]), spring embedings ([CC92], [SA98]), and document clustering ([AOL93], [HP96], [LC96], [ZE98]). Of those, only document clustering is sufficiently fast and produces intuitive results to warrant consideration ([ZE00], [HP96], [Zam98], [ZK05], [CHR03], [RK04]). It is often employed in combination with dimensional reduction and SOM ([IR01], [Wei01], [LA00], [Koh97], [Lin91], [Mer97]).

Document clustering techniques provide a way to relate documents to each other, as well as to a set of topics. A document is assigned to a cluster if it contains the contents of the cluster's topic, and is generally similar in contents to that topic and to the other documents in the cluster. Documents in the same cluster are expected to be more similar to each other than to documents in other clusters [ZK05]. Since documents may contain multiple topics, some approaches allow a document to be assigned to more than one cluster, potentially resulting in overlapping clusters. Usually, intra-cluster relationships are not provided as part of the results. However, they are very useful to provide general overviews of large collections, although they usually have to be interpreted by users with certain level of expertise.

Some approaches completely avoid the high dimensionality problem by simply ordering the most used terms in the document and employing the first N terms [RES98], [RES99]. These strategies work well for single document representation and for association of a limited number of documents, and even for some degree of clustering. However, they also lack a way to clearly relate different documents and display levels of similarity. Other approaches (such as the one by Carey et al. [CHR03]) combine a number of different strategies to allow various views of the same document set, potentially improving focusing and analysis tasks.

Maps resulting from the techniques mentioned above are meant to present various properties of documents, including content similarity, co-citation, term co-occurrence and attribute-value matchings, such as author or publication date. For a detailed description of available techniques for text document mapping and its applications, systems and challenges, we refer the reader to Borner et al. [BCB03].

Few systems for general multi-dimensional visualization add projections in their tool suite. The *Hybrid Information Visualization Environment* (HIVE) [RC03], however, implements clustering and layout algorithms to enable exploratory visualization of multidimensional data. An intuitive configuration process by the user is its main strength. Nevertheless, the layout algorithms implemented are approximations of the original FDP model, which fail to generate visualizations that group (separate) the data items based on their similarity (dissimilarity).

Projections have nevertheless been intensively explored in the context of visualizing unstructured data, particularly text document collections. Two typical layout approaches in this context are *InfoSky* [AKS*02b, GSAK04] and *Galaxies* [WTP*95, Wis99], the later one incorporated into the IN-SPIRE^{TM} system (see http://in-spire.pnl.gov/). Both systems display documents as points in 2D space following a similarity-based layout. Clustering and projections are employed to position groups of documents and to reduce the number of distance calculations. IN-SPIRE builds that into a dimensional reduction approach, similar to Single Value Decomposition or Latent Semantic Indexing [PRTV98, DFK*04]. InfoSky also creates a hierarchical display by embedding structure into the similarity relationships, so a user can focus in and out analogously to manipulating a telescope.

Within our research group, we have also developed — and made freely available — a tool suite for multidimensional visualization based on projections, called PEx (the Projection Explorer — http://infoserver.lcad.icmc.usp.br/infovis2/PEx). PEx is an experimental platform written in Java that generates visual maps of multi-dimensional data sets based on classical and novel projection and point placement strategies, with particular focus on text processing.

One way to measure the similarity between text documents in our case is by calculating cosine the distance [MPL06] between the vector representations. The other is using the "Kolmogorov distance" [TMP07]. This

measure is based on an approximation of the Kolmogorov complexity [LV97] using compression algorithms to compare one document against another. This can be done using the documents in their original form or, in some cases, in part of their original form. The result is not sensitive to tuning or to text dimensionality.

Visual representations of document maps assume various forms of 2D and 3D graphs and triangulations. Also popular is the landscape-type of display, generating heightfields over triangulations of the projected points. Landscape plots have been the choice of many useful presentations of documents before ([Wei01], [Wis99], [Cha93], [CC92]).

4.3. Visual text mining

The wide availability of information stored in the form of digital files has broadened our information universe to an amazing extent. The problem of extracting useful information from text collections is too often faced both by professionals and ordinary computer users, as they interact with, e.g., file systems, digital libraries, email communication, patent databases, news sites, and the Web in general. In this context, to extract information can mean both to look for a particular piece of knowledge, or to browse the collection in order to gain general knowledge about its contents and to infer relationships amongst documents that meet some user interest. Informed decisions many times depend on being able to explore this type of overwhelming information fast.

Users face difficulties in executing such exploratory search tasks, particularly when manipulating large text collections. Consider, for example, someone who just got a sheer amount of search results displayed as a ranked list. The user will likely inspect only a small subset of the whole list, and in doing so will perhaps miss connections between distinct pieces of information that might be of interest or relate to the query somehow. A ranked results list is generally adequate to meet specific information needs, but little encouraging if the user is in any sense trying to achieve a general view or a network of knowledge within a domain.

Help might come from text mining techniques [WIZD04] embedded into search and browsing engines. Semi-automatic mining algorithms can extract semantic patterns, such as similarity and associations from ordinary text documents or queries very well. Classification and clustering are typical text mining tasks. Their goal is to segment the corpus into groups of documents according to their similarity, where a criterion must be defined to compute similarity. Similar documents, as stated by this criterion, should be placed in the same group, dissimilar documents must be assigned to distinct groups. In classification, assignment of a document to a particular group is a supervised process in which the set of groups, or classes, is known beforehand. Clustering is a type of unsupervised classification, that is, the class distribution is not given and data set segmentation is based solely on features extracted from documents. Partitioning into groups induces an organization that can help explore individual segments of the data set and obtain an overview of the collection from the characterization of the groups.

Classification and clustering have their drawbacks. Classification requires previous knowledge about the collection, which may not be available. For clustering, establishing similarity or dissimilarity amongst documents is not straightforward. Indeed, perception of similarity between text documents is sometimes subjective and depends on semantics and domain characteristics. This means that in some situations, a correct grouping would have assign certain 'border-line' documents to more than one group, and it may be that, after closer inspection, documents in the same group do not seem to be that well related to each other. So, grouping algorithms tend to hide these concepts, frequently very important in some exploration scenarios. Additionally, few techniques are capable of expressing additional relationships, intra-cluster or intra-class, leaving the user with 'pockets' of similar documents but no explanatory view of what is inside each pocket or what types of relationships exist within a group.

Other machine learning techniques, such as association rules extraction, can be used to extend support for user centered exploration of text data sets, since relationships within a document can be inferred based upon their sharing common terms. A term consists of one or multiple meaningful words. Given a set of terms, an association rule has the form $(t_1, t_2, t_3, ...) \Leftarrow (t_i, t_{i+1}, ..., t_n)$, meaning that for a document, the common occurrence of those terms on the right-hand side of the rule implies that all the terms on the left-hand side also occur. Two measures usually applied to express the importance of a particular rule are *support* and *confidence*. The support of a rule informs how frequently the terms on both sides of that rule appear together in the document set, and its confidence informs the percentage of the documents in the set that, having all right-hand side terms, also have the terms on the left-hand side. For instance, a rule expressed by $(beer, chocolate) \Leftarrow (peanuts, candy, aspirin)$ with a 30% support implies that all the five terms in the rule appear together in 30% of the whole collection. A confidence of 80% would mean that, from the collection, 80% of the documents that contain beer and chocolate also contain the other three terms in the rule. Mining of association rules has been applied to extract meaningful relationships between text documents based on their content [CNT04].

An association rule mining algorithm scans the data in search for possible associations that can be established for given support and confidence thresholds. The logical intuition in analyzing the rules so obtained is that documents that share term co-occurrences are likely to be addressing related issues, and therefore can be considered similar or correlated somehow. Nonetheless, the process easily generates a huge number of rules — the actual number being exponential in the number of terms. As a consequence, out of thousands of

association rules characteristically generated by a mining algorithm, maybe only a very small subset is of real interest to the user's task, impairing the exploration process. A filtering process is therefore necessary (and many times troublesome) to identify relevant rules.

In the pursuit of solutions capable of better matching user needs, the field of Visual Text Mining is gaining strength. It exploits the synergy between mining and abstract information visualization to create interactive visual representations of document collections for browsing and querying. The ability to couple visual representation with cues provided by mining algorithms is proving to be valuable to users trying to identify patterns with high semantical content, both globally and locally.

Chen [Che04] suggests that users build an internal cognitive map when navigating through a visual information space, analogous to real world navigation. In fact, many user-driven visual text mining approaches rely on so-called document maps — visual information spaces for user navigation that, similarly to geographical maps, spatially reflect one or more properties of the documents that may be of interest. A document map may be built from extracted information, such as co-citations or common citations, presence and distribution of topics, co-authorship, etc. It can also be based on text characterization such as content similarity. Maps can support a variety of exploratory tasks, connecting users with their own cognitive map while circumventing some of the inherent complexities of the underlying information space.

Displaying these maps in a form visually analog to familiar geographical maps has been pointed out as a major strength of map-based interfaces to document collections [Sku02]. Wise [Wis99] strongly argues in favor of an "ecological approach" to text visualization, in which visualizations are grounded in human perception capabilities — his sample visualizations are analogues to night sky and terrain models, whose interpretation is eased by capabilities wired into our brains as a result of our biological heritage. Several approaches exist for creating and displaying document maps. The underlying concept, however, is that proximity in the visual maps reflects some measure of document similarity.

5. Projection Based Visualization and its Application to Visual Text Mining

5.1. Projection techniques and point placement strategies

Multidimensional projection techniques can be split into two major groups, according to the functions f employed: *linear projection techniques*; and *nonlinear projection techniques*.

Linear projection techniques create linear combinations of the data attributes, defining them in a new orthogonal basis of low dimension. A widely known linear technique is *Principal Component Analysis (PCA)* or *Karhunen-Loève Expansion* [Jol86]. PCA is a second-order technique, that is, it employs information embedded in the covariance matrix of the data. For m attributes, a covariance matrix is a $C_{m\times m}$ matrix whose element c_{ij} denotes the covariance between data attributes i and j. The covariance indicates the degree of linear relationship between the two attributes. Second-order techniques are particularly suitable for data presenting Gaussian (normal) distributions, since in this case it captures almost all data distribution.

The process adopted by PCA is to create the covariance matrix of the data, then decompose it into m eigenvectors with m eigenvalues. The first p eigenvectors with the largest eigenvalues are selected to transform the m-dimensional space into a p-dimensional space that retains the major variance of the data. The variance can be successfully captured even if $p \ll m$. Indeed, PCA is the technique that produces the best results in terms of the information loss. It retains as much as possible the relative distances among the data instances, whilst projecting them onto a low-dimensional space.

Although they perform well on Gaussian data, in handling data with nonlinear structures, such as clusters of arbitrary shapes or curved manifolds, linear techniques typically fail to capture the relevant patterns. In such cases, nonlinear techniques are better candidates. Rather than relying on linear combinations of the attributes, nonlinear techniques attempt to minimize a function of the information loss incurred in the projection. Normally, this function is based on the dissimilarities amongst the m-dimensional instances and on distances among the p-dimensional points. Hence, it does not require representing the original data as vectors, it is sufficient to have a mechanism to measure instance dissimilarity in the high-dimensional space.

Since nonlinear techniques perform an optimization process, their iterative nature is an additional advantage. Thereby, a user can observe the execution of the projection process and interrupt it if convenient. Another interesting feature is that adding new subsets of instances only requires a limited number of additional iterations. Linear techniques, on the other hand, demand the overall process to be entirely redone.

One example of a nonlinear projection technique is *Multidimensional Scaling (MDS)* [CC00]. Sprang from the psychophysics domain, MDS actually comprises a class of techniques aimed at mapping instances belonging to an m-dimensional space into instances on a p-dimensional space ($p \leq m$), striving to keep some distance relations. A well-known example of MDS technique is called *Sammon Mapping* [Sam64]. It starts by defining a function that indicates the amount of information loss incurred in the projection, and then applies an iterative nonlinear optimization method based on the gradient of such function to find a (local) min-

imum. This function is presented in Equation 1. One observes that it will reach a minimum when the dissimilarities $\delta(x_i,x_j)$ amongst the m-dimensional instances are close to the distances $d(f(x_i),f(x_j))$ between the p-dimensional points. Some normalization is also applied to favor the definition of more compact spaces for the projection.

$$S = \frac{1}{\sum_{i<j} \delta(x_i,x_j)} \sum_{i<j} \frac{(d(f(x_i),f(x_j)) - \delta(x_i,x_j))^2}{\delta(x_i,x_j)} \qquad (1)$$

Amongst the various MDS techniques, the simplest ones are those based on *Force-Directed Placement (FDP)* [FR91, Cha96]. Originally proposed as a graph drawing heuristic, the FDP model aims at bringing a system composed of instances connected by imaginary springs into an equilibrium state. Initially, the instances are randomly placed in the system, and the forces generated by the springs are employed to iteratively push and pull the instances until reaching an equilibrium. In order to apply the FDP model as an MDS technique the spring forces must be proportional to the difference between the dissimilarity $\delta(x_i,x_j)$ among the m-dimensional instances, and the distances $d(f(x_i),f(x_j))$ among the p-dimensional points.

One example of the former strategy was presented in [TMN03], called *Force Scheme*. Different from the original idea of Eades, where each instance is moved once per iteration, Force moves each instance $n-1$ times on an iteration. Thus, less iterations are necessary to bring the system to an equilibrium state. Although it reduces the model's complexity, each iteration is still $O(n^2)$. Aiming at reducing this complexity, Paulovich and Minghim [PM06] proposed a new method where the instances are first clustered, and the Force is applied considering the instances of each separated cluster, defining a model whose complexity is $O(n^{\frac{3}{2}})$. The core idea of ProjClus is to calculate the centroids of the initial clusters, then project these centroids onto the plane. Next, it separately projects each set of points defined by the clusters onto the same plane; then the technique assembles the final layout, positioning the clusters' projections according to the projection of their centroids. A related approach was applied by Andrews [KGM*01].

An example of a technique that also reduces the complexity of the iterations of FDP was presented by Chalmers [Cha96]. This approach reduces the complexity of an iteration using data samples in order to determine which instances are connected by the imaginary springs. Although this approach make the iterations linear, the model complexity is still high due to the n iterations necessary to create the layout, being $O(n^2)$. Aiming at reducing this complexity, another approach was presented in [MRC02] (and extended in [MRC03]), which defines a FDP model with complexity $O(n^{\frac{3}{2}})$. In this approach a random sample S of $\sqrt{n}$ instances is first projected using the Chalmers approach. Following, the remaining instances are interpolated from these instances. The process for the interpolation is the one that makes this technique $O(n^{\frac{3}{2}})$, therefore, Morrison et al. [MC04] suggest a modification of this interpolation reducing the final complexity to $O(n^{\frac{5}{4}})$, and Jourdan and Melancon [JM04] suggest a further approach to reduce it to $O(n\ log\ n)$.

Besides ProjClus, we have developed two other distinct nonlinear projection techniques, all illustrated before in Section 2: *Least-Square Projection (LSP)* [PNML06] and a point placement strategy that builds a similarity tree, called N-J (*neighbor-joining*) tree map [CPMT07].

The LSP technique is a generalization of an approach for mesh-recovering and mesh-editing in order to deal with high dimensional spaces. In this technique, a subset of m-dimensional points are projected onto the plane, and the remaining points are projected using an interpolation strategy that considers only the neighborhood of the m-dimensional points.

ProjClus and LSP are high-precision and fast projection techniques, which are suitable to handle points belonging to nonlinear sparse spaces — such as the one generated from document vector representations (see Section 3.1.15) — since they take into account neighborhoods of the original points.

One approach for point placement that differs from the conventional view of grouping by proximity on the 2D plane, is the creation and drawing of similarity trees that reflect the similarity relationship calculated by the measure employed [CPMT07]. To do that, the technique uses a phylogeny reconstruction algorithm, called neighbor-joining. Points that are closer together in the whole set are joined in a tree and replaced in the list by a 'combined' node which then enters in the search process together with the remaining points in the data set. The N-J tree is capable of representing relationships that allows the user to quickly recover information detected by the similarity metric.

5.2. Mapping text collections via projections and point placement

The result is a planar mapping of a point set X representing the documents (we call this 2D placement a *map*), which can be used to explore the document collection represented by X. As is going to be illustrated here, the techniques we have developed are capable of mapping documents in such a way that text data dealing with common subjects form groups that are visually apart from other groups, thus allowing identification of patterns in the text set. An example of such a map is given in Figure11. It shows a document map of a collection of 574 scientific papers belonging to three different subjects, previously manually classified. In the pictures, different classes have different colors (red is Case-Based Reasoning (CBR), blue is Information Retrieval

(IR), and green is Inductive Logic Programming (ILP)). The small greenish group in the rectangle shows the placement of five papers on sonification (the use of sound to display information). That particular group refers to the previously-mentioned five papers reflecting the evolution of our sonification system, which are strongly correlated. Figure11(a) shows that they are adequately placed in the same neighborhood. Figure11(b) demonstrates that, when mapped in conjunction with the remaining papers, this group was placed in the outer boundary of the map and nearby papers in the data set whose main subject was audio retrieval (they are the points with colors other than red in Figure11(c). Another aspect that can be noted in the map is the fact that the CBR and ILP groups are more compact, whereas the IR set is spread out over the entire map, a visual analogy to the characteristics of the subject areas, the first two being more mature and therefore having a set of their own nomenclature and techniques, while IR relies on resources from various sources including those of ILP and CBR.

5.3. Topic extraction and visualization

5.3.1. Topic identification by covariance

Although a projection can help locating groups of related documents, the initial map does not reveal information within these groups. In order to extract information about these groups, a group can be selected and a label can be generated that aims at identifying the main topics discussed within that group.

From a group of selected documents, a label is constructed by an algorithm that first chooses the pair of words with the highest covariance in the vector representation of the group. Then, for each remaining (non-selected) word it computes the mean of the covariances relative to these first two words. If the covariance of two terms approaches zero, it means that the terms are independent. If this is a significant value relative to the highest covariance, e.g., above a defined percentage threshold (in this paper we used 50%) the word is added to the label. This is the way the labeling by term covariance is implemented in PEx, and that was used in the illustrations of Section 2. In those pictures, the first two terms on the left side of a label are the ones that present the greatest covariance, and are indicative of the main topic in the group.

The term covariance is calculated according to Equation 6.

$$cov(t_i,t_j) = \frac{1}{n-1}\sum_{k=1}^{n}(t_{ki}-\bar{t}_i)(t_{kj}-\bar{t}_j) \qquad (2)$$

where $\bar{t}_i$ is the mean of the i^{th} term t_i, and t_{ki} and t_{kj} are the values of the i^{th} and j^{th} terms for the k^{th} document.

In order to illustrate the representativeness of the term covariance further, the examples of Figures 11 and 12 have defined the labeling in a slightly different manner than the one illustrated before, and currently implemented in PEx. Figure 12 shows a document map of the same data set of Figure 11, showing the resulting labels after group selection by the user.

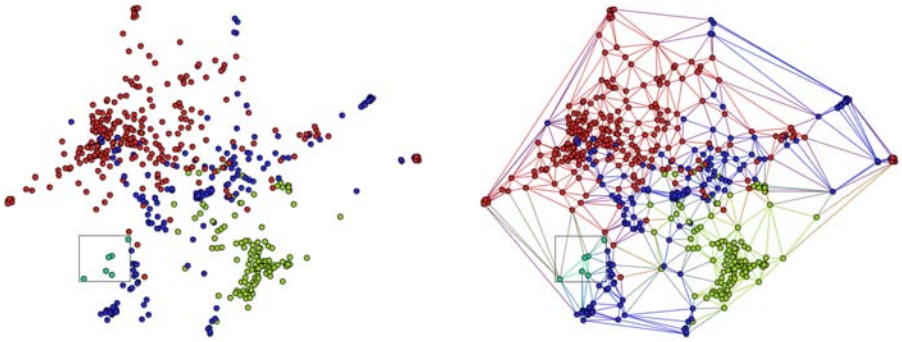
The first two terms on the left side of a label are the ones that present the greatest covariance. The third one on the right side is the term which has the greatest *mean covariance* according to the labels already chosen. The mean covariance is calculated as the mean between the covariances of a term taking two other terms. The numbers between parenthesis are the covariances.

If the covariance approaches zero, it means that the terms are independent. On the label, this number can indicate if the three terms are equally related or if only the first two are. If the number on the left side is much greater than the number on the right side, it means that the third term is not deeply related to the other two. But if these numbers are close, it means that the three terms occur frequently together.

For instance, in Figure 12(b), considering the label *[learning,logic](60.44) >> [programming](44.07)* and dividing the right covariance by the left we get $\frac{44.07}{60.44} \approx 0.73$. Calculating the same rate for the label *[information,retrieval](101.76) >> [text](26.56)*, we get $\frac{26.56}{101.76} \approx 0.26$. Thus, it is possible to infer that the terms "learning" and "logic" are much more related with "programming" than the terms "information" and "retrieval" are with "text".

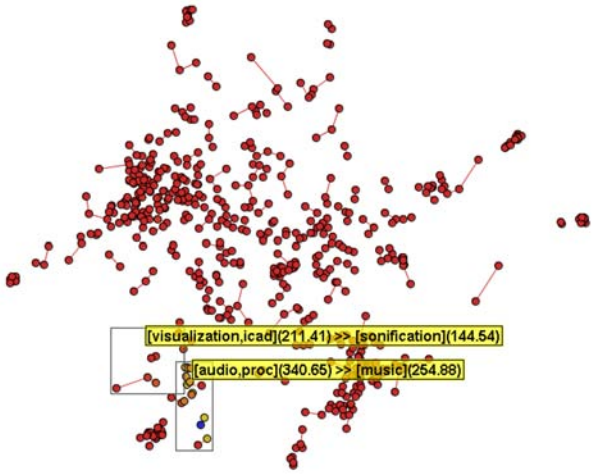
Notice that the covariance is not a causality relation between terms, that is, it does not indicate that one term occurs due to the other one. It is only a measure of the degree to which two terms vary together. For example, on the map of Figure 12(b), which is a view of part of the Figure 12(a), the terms "secret" and "sharing" present a high degree of covariance, but it is not possible to establish that the term "secret" will cause the term "sharing" or the other way round. Another important aspect is that using two terms only, it is not normally possible to direct infer why they are related. This is the point where the third term (as well as the others) of the label can help. The third term for the terms "secret" and "sharing" is "schemes", thus it is possible to infer that the main topic of the document used to create such label is somehow related to cryptography.

Covariance is useful to identify main topics in a group of documents. A more advanced technique is necessary to obtain more detailed automatic topic extraction, exploring further the sub-topics in a group as well as additional topics of documents that may not match the main topics extracted by covariance. The next section presents one such technique, which handles detailed and automatic topic extraction using association rules. For a more detailed view of that subject, we add a previously written technical report in Appendix B. We also refer to [LPMP07] and [PLOM07].



(a) Mapping for the three-class text collection. Detail is the sonification test group.

(b) Delaunay Triangulation of the Map in Figure 11(a) .



(c) Placement of sonification papers midst papers of audio information retrieval papers

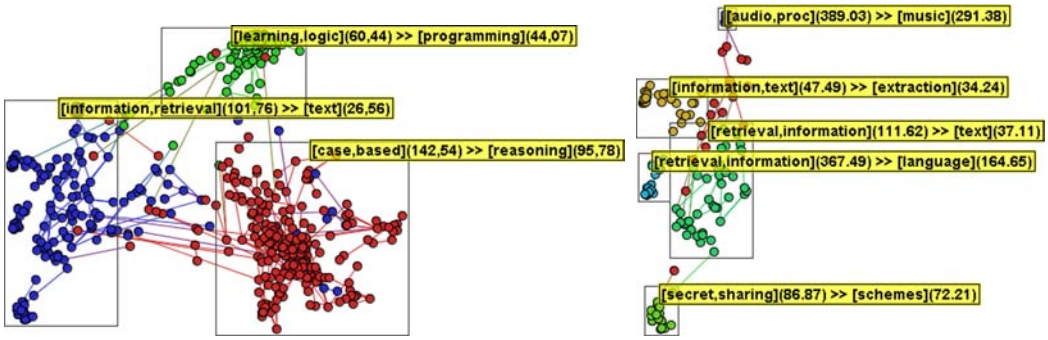
Figure 11: *Document map of a collection of scientific papers belonging to three different main subjects, manually classified, plus a test group (inside the rectangle).*

5.3.2. Topic extraction by sequential covering induction of association rules

An appropriate set of association rules derived from a collection of text documents can be used to describe a context in which a term appears or also the context or topic related to a subset of documents. When mining association rules from text, an association rule (AR) is an implication of the form $X \Rightarrow Y$, where $X \cap Y = \emptyset$ and they are both subsets of L, where $L = l_1, l_2, \ldots, l_m$ is a set of literals or items (each item representing a term from the bag of words). A transaction T is a set of items $T \subseteq L$ that represents a document from the corpus C. The rule $X \Rightarrow Y$ holds in the document set C with confidence c if $c\%$ of the documents in C that contain X also contain Y. The rule $X \Rightarrow Y$ has support $s\%$ in C, if $s\%$ of documents in C contain $X \cup Y$.

A rule R_i: $X \Rightarrow Y$ with significant support in C means that a set of terms $l_i \in X \cup Y$ are found together in a large subset of documents from C. Moreover, if R_i has high confidence, then the occurrence of X (the body) means high probability of the occurrence of Y (the head), at least in C.

In [LPMP07], an algorithm is presented to produce and select good association rules in order to describe the main subject or topic in a selected set of documents S_k. That algorithm deals with the problem of the large number of association rules as follows. Instead of post-pruning the set of rules generated from all documents in S_k by some rule quality measure, it generates rules that contain at least one term (seed) present in a selected set of most relevant terms. The term relevance is given by a weight that favors terms with higher frequency in the selection than in the rest of the corpus. An arbitrary number of rules, usually 1– 3, with the highest term weight summation is then selected for display as label for the group selected. Eventually, some documents



(a) Document map of a scientific papers collection in three different subjects.

(b) A view of part of the map in Figure 12(a) .

Figure 12: *Example of labels for document maps.*

carry topics that are not described by these selected rules, that is, they are not covered by the rule.

Another algorithm uses a sequential covering strategy in order to extract topic-targeted rules from the documents uncovered by the main rule. Additionally, instead of just selecting the n highest weighted rules, it keeps only those rules that provide additional covering over the previous ones. This reduces the amount of redundancy found in the selected rules, considering that if two rules cover the same subset of documents, only the one with the highest weight is kept.

The Algorithm has two nested loops. In the inner loop, the iterative algorithm for generating and ranking rules over a selected area S_k of the projection leads to a set of rules SR. This process is repeated (the outer loop) removing covered documents during the iteration until no document is left or the rule generation process outputs an empty set of rules. The function coverage (AR_i, U) of Algorithm 1 denotes the number of documents from U that support the rule AR_i.

Even considering that each selected rule must cover documents not covered by any other previously selected rule, a considerable number of rules may be generated for a given selection of documents. To reduce cluttering, we apply a rule grouping strategy that relies both on literals in the rule and on overlap of covered documents. Intuitively, if rules share some of their literals and cover roughly the same set of documents, they are likely to represent the same topic, but described in a slightly different way. Thus, instead of presenting a label for each rule, only one label per group of rules is displayed. A rule AR_i joins a group of rules G_j, if (i) AR_i has at least one term $t \in T_{G_j}$, where T_{G_j} is the set of literals found in rules already in the group G_j, and (ii) there is an overlap between the set of documents covered by the rule, and the set of documents that are covered by any of the rules in the group. This document coverage overlap should be equal to or higher than a given constant α. Here, we used $\alpha = 0.5$. To compute the overlap, we use:

$$overlap = \frac{|\cap D_{AR_i}, D_{G_j}|}{min(\{|D_{AR_i}|, |D_{G_j}|\})}$$

where D_{AR_i} is the set of documents covered by the rule AR_i, and

$$D_{G_j} = \bigcup_{AR_i \in G_j} D_{AR_i}$$

If a rule does not join any group, a new group is formed with that rule as an initial element. The label of a group of rules begins with the highest weighting term taken from the rule with the highest support in the group. The label shows also this high support rule followed by all the terms that are found in the other rules of the group and that have not yet appeared. For instance, the label "mosque:[suicide]←[mosque], (27/20%;90%):iraq:baghdad:" is given to a rule group where the rule "[suicide]←[mosque]" has the highest support among the rules in that group (27 documents or 20% of selected points S_k), "mosque" is its term with the highest weight, and 90% is its confidence. In the example, "iraq" and "baghdad" are the terms found in other rules from the group that does not appear in the highest support rule.

Figure 25 demonstrates the potential of this technique in aiding exploration of document collections. To build the map, the flash news data set used in Section 2 was employed. The picture presents the whole set of documents colored according to a clustering algorithm in PEx. Labels have been generated for each of the clusters and a few of them were selected for display.

Labels such as those, which reflect topics in the data set, can be summarized by displaying the main rule and a list of the remaining terms, ordering them according to rule coverage and term relevance within the rules. This prevents too

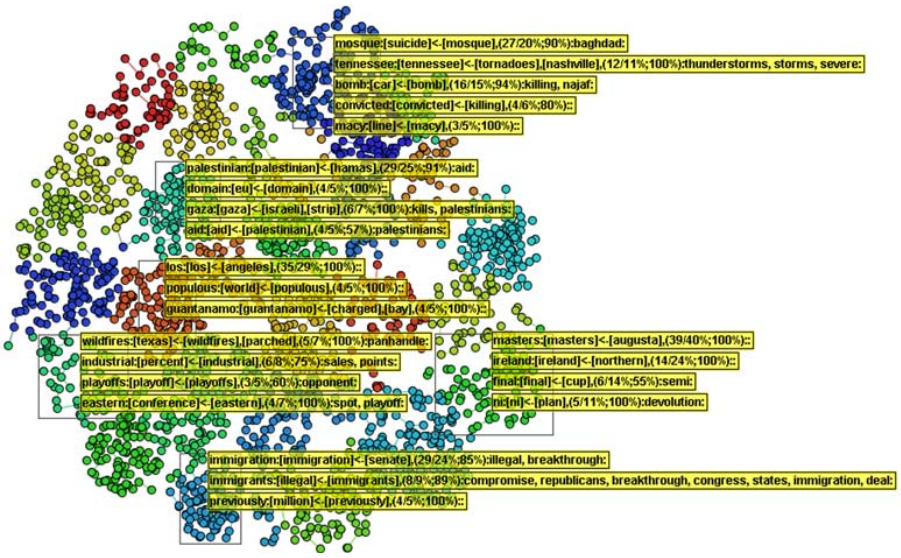


Figure 13: *Document map of the flash news data set. Projection by ProjClus. Individual colors denote clusters of documents. Shown labels were generated from groups of rules inducted using some of the document groups.*

much cluttering of the image, although it hides the various degrees of abstraction revealed by the rules themselves.

5.4. Further Examples

See Slides attached.

6. Conclusions, Current Challenges, Future Trends

The field of mapping documents for content analysis, particularly in the light of new trends in visual and text analytics, is extremely challenging and raises many complex issues.

The first one relates to the extremely different applications that rely on text and, within them, the varying tasks associated with the exploration of such maps. There are applications as diverse as Web search information gathering and analysis of medical information from patient records. As illustrated by some of the examples in these tutorial notes, not only text can be transformed into table data, but also table data can be transformed into textual information. The matter of customization is an important one and there are many parts of the various techniques for text mapping that can be adjusted to fit particular data sets and particular tasks.

The problem of customization also bears relationship to one of the critical challenges of these applications: the preprocessing stage of the mining and visualization processes. Stopwords definition is many times user controlled and is very rarely trivial. Yet, it is a necessary step in most applications of mining, visualization and visual mining. In the case of applications that involve safety of high human involvement (such as medicine, forensics, detection of illegal material), particular depth is due in the definition of vocabulary and in linguistic processing before conclusions can be reached.

The issue of what is important to define as similar in a particular setup of application and task also points to the development of new approaches for calculation of similarities between pairs of documents. Proper strategies can be developed that take into account the homogeneity and semantic levels of a text collection as well as the context in which it is being analyzed. The results of employing Association Rules for topics thus far indicates that there are levels of abstraction that can be extracted employing similar approaches with the target of finding an appropriate combination of concepts or a relevant vocabulary for exploration tasks.

The issue of quality of projections is still open. The classical stress parameter, targeted for optimization strategies for point placement, is of little help to reflect effectiveness of grouping and separation of documents by their content. Additional paradigms and procedures need to be developed.

Closer integration of visual aids with mining is also expected to be one of the hottest issues to be addressed in the coming years. Also are the coordination and multiple views between different visual representations, and even between different mining algorithms. Information gathered from the visual process can be fed into algorithms such as classification. Conversely, mining results can be reflected back onto the visualizations, either by changing its visual (and why not

aural and tactile) attributes, or by feeding into the parameter adjustment of the visualization pipeline [Hei07].

Naturally, a crucial issue regarding textual data sets is scalability. While a few techniques can handle large numbers of documents, there is a compromise in the quality of the resulting map regarding content associations. The current state of this technology seems to point towards two types of analysis tasks: one for massive quantities of text, which relies on efficient algorithms (e.g., visually supported or based on mining and intelligent search strategies) to reduce the data set; and a second type that takes a few thousand individual documents to be visually explored in an integrated mining and visualization environment such as that presented here.

Finally, two issues of great importance, touched upon only superficially in this material due to space constraints, are the interaction paradigms and the evolution in time of a particular data set. New visual approaches imply different strategies for exploration, particularly when coupled with processing by mining algorithms. The issue of time is also extremely pertinent in many applications of textual exploration (e.g., medicine, news, research topics evolution, etc.). They need to be addressed through proper models, and, in the case of the exploration techniques, inspired by user evaluations of current available technology.

7. Acknowledgments

This work has been mostly supported by the Brazilian research support agencies FAPESP (grants 03/02815-0, 04/01756-2, 04/09888-5 and 04/07866-4) and CNPq (grant 304758/2005-1). Haim Levkowitz's research was partially conducted while he was a Fullbright US Scholar to Brazil, August 2004 – January 2005.

The authors wish to acknowledge the following profound contributions to the results behind the research reflected here:

- Fernando Vieira Paulovich, PhD student
- Roberto Pinho, PhD student
- Guilherme Pimentel Telles, research partner, ICMC
- Alneu de Andrade Lopes, research partner, ICMC
- Maria Cristina Ferreira de Oliveira, research partner, ICMC
- Luis Gustavo Nonato, research partner, ICMC
- Lionis Watanabe, Wagner Franchin and Ana Maria Cuadros Valdivia, MsC students
- Pedro Vilela and Renato Rodrigues, undergrad students

Rosane Minghim also acknowledges the enthusiasm and productive research exchange currently in place with Dr. Anton Heijs, Treparel Inc., The Netherlands, and Charl Botha, Technological University of Delft, The Netherlands.

Appendices

A. Technical Report: VISUAL MAPPING OF TEXT COLLECTIONS USING AN APPROXIMATION OF THE KOLMOGOROV COMPLEXITY

Guilheme P. Telles, Rosane Minghim and Fernando V. Paulovich

{gpt,rminghim,paulovic}@icmc.usp.br
ICMC - University of São Paulo,Brazil

The generation of content-based text maps is an important issue to support exploration of information and to help find relevant reading material in increasingly complex document databases. Most techniques that help relate or visualize texts rely on a vector representation that is, at its best, ad-hoc as to its parametrization. This paper presents a novel approach capable of generating a map of documents without the painstaking pre-processing steps, by comparing text against text through an approximation of the Kolmogorov complexity. The similarity measure taken from that analysis is then used to map data in 2D by applying fast multidimensional projection techniques (instead of dimensionality reduction or random initial point placement). The resulting maps show a high degree of content separation and good grouping of similar documents. The approach can be used to map text collections in a variety of applications and the map can be interacted with to further explore text groups. By avoiding vector representation our technique decreases the bias characteristic of that approach and the need for user knowledge of the process. The approach also lends itself to incremental processing for reduction of computational costs.

A.1. Introduction

In a set up of document collections, such as text databases or Internet search results, it is not easy to locate reading material even with an effective ranking system. In the applications targeted here the goal of a user's analysis is to locate relevant documents to examine or study amongst a considerable number of recovered texts. We approach that by studying effective mapping strategies that are capable of coding relationships amongst documents geometrically and visually. An interactive map based on document content can be explored to locate a text relevant to a query or to another target text and to find groups of similar or related documents. This is usually done after a pre-filtering process that has narrowed the collection down to a range of few hundred to a couple of thousand texts. Typical applications are research, education and training.

Most text mapping strategies are based on clustering or dimensional reduction that rely on text vector space representations whereby, after a considerably lengthy pre-processing step, texts in a collection are represented as a vector of many dimensions. Each dimension is a relevant term in the text set. One of the problems with this largely used vector approach is that the dimension of the final data set can reach the thousands easily. It is a well know fact that, as the dimension gets larger, the ability to properly infer vector distances gets impaired, prompting the need for some sort of attribute selection.

As far as processing speed is concerned, the transformation of that representation into a map of some sort may involve either dimensional reduction or clustering techniques, which can be rather slow themselves, or on faster projections, without significant quality loss ([MPL06, PNML06]). The fact remains, though, that the pre-processing step involved in all of them include various procedures, such as stopwords elimination and stemming, that can be affected by various parameter settings. Other adjustments for feature selection and frequency analysis are often necessary. These adjustments have large influence on the outcome, sometimes prompting more than a few iterations before the result is satisfactory, and making the process too sensitive to change in the subject target texts. In most real cases, therefore, it is not possible to tell in advance what the right tuning for the pre-processing is.

This paper proposes a novel approach to text mapping based on direct comparison between texts contents without the need for any preprocessing, by overriding the vector representation altogether. We calculate an approximation of Kolmogorov complexity ([LV97]) as a similarity measure between texts. That is used as a distance value to generate multi-dimensional projections into 2D space by means of a fast approach [PNML06]. The resulting maps can be interacted with in a form that allows further exploration.

The following section offers a review of relevant text mapping literature an the role of projections in its context. Section A.3 describes the multi-dimensional projection procedure employed here step by step. A summary of the theory and implementation of the Kolmogorov 'distance' is given in Section A.4. Results and comparison to other maps are given in Section A.5, which is followed by analysis, conclusion and further work discussion.

A.2. Previous Work

Due to the complexity and variety of the information and scenarios involved in text examination, alternative means of mappings text sets must be sought. Here we review the works in the literature that deal with this problem that, in our view, cover the main issues relating to visual mapping techniques for documents.

A number of different techniques for visualization of textual results from Web and other searches have being deployed ([ABY03], [LA00], [SCL*99], [BY96], [Cha96]). While these techniques are capable of displaying large text bodies, they tend to make location of relevant reading material more troublesome. Our focus in this work is to provide

complementary tools to support mapping of documents in a way that helps locate neighboring similarities between texts and groups of texts. So we assume a pre-filtering task that reduces the universe of targeted documents to a few hundreds (maybe thousands) of texts in a few areas of interest (not necessarily pre-determined).

Many techniques for text visualization exist that search for a representation of the content of an individual text (e.g. [MWBF98], [RES99, RES98]), of text collections (e.g. [LA00], [BCG*99], [Wei01]), or of themes approached in texts (e.g. [HHWN02], [Wis99], [WTP*95]) in order to meet the above mentioned targets.

Usually text processing tasks employ the vector space model [SB88, Sal91] whereby texts are represented as points in a vector space. In this representation each text is a vector with dimensions represented by terms (n-grams). The vector coordinates are the weights of the terms based on their frequency. Typically, dimensions reach the thousands even for small to medium databases. Transformation of a text collection into a vector space is preceded by elimination of non-influential words (such as stopwords), reduction of words to their radicals (stemming), and frequency counting of some sort (various exist). The initial representation is followed by reduction in space dimensions, typically involving cutting off words that are too frequent or too rare in that particular collection, and clustering dimensions to generate new 'combined' attributes, in an attempt to overcome the dimensionality curse.

The most common way to extract structure from a text collection is by applying some sort of dimensional reduction technique over the resulting vector representation. This is the case of systems based on Multi-dimensional Scaling (MDS), Principal Component Analysis (PCA) or Latent Semantic Analysis (LSA), that work with statistical measures for subspace reduction, and Self-Organizing Maps (SOM), that employ neural computation ([Wei01], [Wis99], [BCG*99], [KHLK98], [WTP*95]). Those techniques can be used to plot the original data in bi-dimensional (2D) space, when dimension is reduced to 2.

Although dimensionality reduction is a natural processing trend for texts, these types of techniques have high computational costs and low adaptability to incremental processing. Multidimensional reduction techniques also cause other difficulties, such as [HWR03]: high information loss when applied directly to two dimensions (for display); reduction in input dimensions do not seem to affect greatly the outcome; and there is an inherent discretization problem associated with techniques such as SOM, by which individual documents in groups are not distinguishable. For the target of this work, dimension reduction poses and additional problem: when used to display the results in 2D, the mappings to subspaces may define groups of 'similar documents', but locally it is not possible to relate neighboring texts. In a work by Lopes et. al [LMMP06] LSA has been successfully applied for the generation of document maps with high local content relationships, but the high computational cost remains a problem as well as the handling of vector transformations.

Another recurring strategy for dealing with the organization of information from a text collection is document clustering ([CHR03], [RK04]), many times employed in combination with dimensional reduction and SOM ([IR01], [Wei01], [LA00]). They provide a way of relating documents with varying success rates. When clustering techniques are applied, here too the intra-cluster relations are not given as a result. However, they are very useful to provide general overviews of large collections, although they usually have to be interpreted by users with certain level of expertise.

Point placement strategies and force-based point placement improvement have been used before to generate document displays [Cha96, GSAK04, AKS*02a]. Although still based on vector representation, they can avoid partly or completely the extensive calculations needed in dimensional reduction techniques by starting with a semi-random point placement and re-adjusting their position based on attraction by similarity.

There are approaches that completely avoid the problem of high dimensionality by simply ordering the most used terms in the text and employing the first N terms [RES98]. These strategies work well for single text representation and for association of a limited number of texts, and even for some degree of clustering. However, it also lacks a way of clearly relating different documents and displaying levels of similarity. Other approaches (such as the one by Carey and others [CHR03]) combine a number of different strategies to allow various views of the same document set, potentially improving focusing and analysis tasks.

Final maps resulting from the techniques mentioned above are meant to analyze a number of properties of documents, including similarity, co-citation, term co-occurrence and various others. We refer to the work of Katy Borner and others [BCB03] for a detailed description of the previously available techniques for text mapping and its applications, systems and challenges. A few systems are being developed dedicated to viewing maps from multi-dimensional data and some of them are particularly dedicated to text collections. One recently published system [GSAK04] adds representational power to the conventional ways of plotting text as points in 2D by separating their contents in thematic areas and handling levels of interaction by hierarchical organization.

In general the methods discussed above lack the ability to determine levels of associations between texts contents. Others are computationally expensive. Faster mapping approaches with the ability of associating texts by similarity have been put forward [MPL06, PNML06, PM06]. Their gain in processing time is attained by using projection techniques,

which are faster compared to dimension reduction and also provide an initial point placement prone to speed up force-based improvement schemes. But those too are based on the vector representations.

Text vector representations, although very useful and largely employed, many times are cause for concern. They tend to impose bias and be difficult to tune. Depending on the choice of pre-processing parametrizations (such as Luhn's cut, vocabulary, types of stemming, types of frequency count, type of feature clustering or selection), the outcome of the analysis and displays of text collections can be highly affected. That also makes it difficult for lay users to employ the representation inside its usual context without knowledge of the processing, pre-processing and visualization techniques.

Another issue that impairs general use of vector-based techniques is its adaptability to incremental processing. Adding new texts to a previously existing collection mostly implies in rebuilding the visualization (including pre-processing) almost from start, propagating possible limitations to every map formation.

The technique presented here is based on projection techniques that have been proven useful to group and separate data when the similarity measure is sufficiently powerful [TMN03, MPL06]. To be able to function, these projections need only the distance between data points (in our case, texts), and not the original data themselves. The contribution of our technique is in producing such a distance between texts through a similarity measure that completely avoids the conventional vector representation. This way it eliminates the need for the ad-hoc pre-processing steps necessary to build that representation and avoids the problem of treating high dimensionality, a real trouble for texts. This measure is based on an approximation of the Kolmogorov complexity [LV97] and is calculated by comparing text against text. This can be done using the texts in their original form or, in some cases, in part of their original form. The results have reproduced separation and grouping advantages of previous methods and is not sensitive to tuning or to text dimensionality. These features also make the method easier to use. Opposed to most vector based mappings, it is naturally adaptable to incremental processing, with affordable storage overhead.

The visual representation adopted here is the landscape-type of display, which is very useful due to its ability to reveal information without resorting to highly attentive perceptual processes. Additionally, surfaces are highly interactive and familiar to most users. Landscape plots have been the choice of many useful presentations of texts before ([Wei01], [Wis99], [Cha93], [CC92]). This feature combined with the absence of pre-processing allows interpretation of mappings even by users with little expertise in the field. The surface representation of our technique is enriched by mapping further significant information to visual attributes (such as lines, colors and height) and aural attributes (such as pitch and timbre). The final map can be explored to the advantage of users interested in having an overview of a set of texts, locating important texts in corpora, or finding useful associations between texts, thus selecting material to read or study.

A.3. Projection techniques for text visualization

Methods of data projection into lower dimensions have the advantage over dimensionality reduction that they are much faster and, depending on the type of projection, good for incremental processing. A previous work [TMN03] has shown the advantages of projection techniques based on distance metrics to obtain useful views of multi-dimensional data sets. When the distance calculation captures significant data set features, they are capable of separating the data collection into groups and result in good association between neighboring individuals. Additionally, those techniques lend themselves to landscape plots. Their application for mapping texts represented in the vector space analogy was tested before with satisfactory results [MPL06]. In this work the same types of projections are used to map points into a plane except the distance between texts is now calculated without generating the vector representation (see Section A.4).

Different from other techniques that can be used to map data into 2D or 3D, such as dimensional reduction, clustering, or point placement strategies that start from a random or semi-random 2D display, the goal of distance-based projection techniques – e.g. Fastmap [FL95] and NNP [TMN03] – is to place a set of points defined in multi-dimensional space in another space such that the relative distances between points are preserved as much as possible. The degree to which that distance cannot be preserved is called the error of the projection. For projections into a bi-dimensional plane, this problem can be stated as:

Let X be a set of points in $\mathbb{R}^n$ and $d : \mathbb{R}^n \rightarrow \mathbb{R}$ be a criterion of proximity between points in $\mathbb{R}^n$. Find a set of points P in $\mathbb{R}^2$ such that if $\alpha : X \rightarrow P$ is a bijective relation and $d_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a proximity criterion in $\mathbb{R}^2$, then $|d(x_i,x_j) - d_2(\alpha(x_i),\alpha(x_j))|$ is minimum for every pair of points $x_i,x_j \in X$.

The set P is called a projection. In this type of projection, it is of great importance the definition of a proper proximity criterion, calculated in our case from the Kolmogorov complexity estimation.

We have used two projection algorithms in our experiments, with similar results: Fastmap and NNP. The first realizes hyperplane projection and the second performs geometric placement in neighborhoods. Each projection is improved by a projection improvement scheme called Force [TMN03], that enhances those projections by recovering part of the information lost during the mapping process, in a similar procedure as that adopted before by other researchers, such as

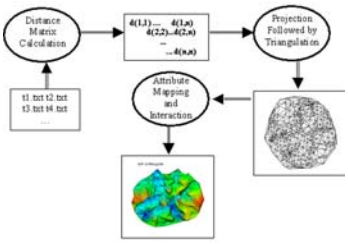


Figure 14: *The whole mapping process. No pre-processing actually needed.*

Chalmers [Cha96]. What the Force scheme does is iteratively approach points projected too far and repel points that were projected closer than they should have. It does that by a fraction of the ideal distance at each iteration. The computational cost of such improvement is minimum due to the initial point positioning by projection. The Force scheme also provides an overall measurement of the projection error (see [TMN03]).

The resulting points, now in 2D space, are connected by triangulation, thus to produce a surface, allowing interaction for exploration of text content. On top of that surface, text properties can be mapped to display attributes to support exploration. In our case, color, height and sound were employed to complete the map. Each attribute (color, height or sound) can map text clustering, or text category (in case it is pre-determined), as well as additional information such as year of publication, number of citations, rank, and so on.

The complete set of steps taken to build a map based on projection from Kolmogorov distance (or any other distance measure for that matter) is the following:

1. calculation of a triangular distance matrix comparing all texts and judging their similarity;
2. projection the points (texts) onto bidimensional space using a fast algorithm, followed by an improvement strategy;
3. triangulation of the data.

Figure 14 illustrates the complete process, and shows one possible resulting map. On top of the map, color (as well as height) were used to show clustering of the projected texts.

The calculation of the distance between texts is presented in the next section.

A.4. Kolmogorov Complexity as a means to define distance between texts

Intuitively, the Kolmogorov complexity is a measure of the amount of information that a message contains. It can also be seen as a measure of randomness of a string or as the length of a string that results after perfect compression. An extensive treatment on the Kolmogorov complexity appears in Li and Vitányi's book [LV97]. A text can be seen as a string, so the discussion that follows applies to texts directly.

Formally, the Kolmogorov complexity of a string y, $K(y)$, is the size of the smallest algorithm that outputs y. Any formal notion of algorithm can be applied, such as Turing machines. The conditional Kolmogorov complexity of a string y given a string x, $K(y|x)$ is the size of the smallest algorithm that outputs y when x is given as input. Intuitively the conditional Kolmogorov complexity is the amount of information in y that is not known by x. The Kolmogorov complexity considered here is the prefix version, where algorithms are considered to be prefix-free, that is, no algorithm is a proper prefix of another.

The Kolmogorov complexity is not computable but it can be approximated using compression. Let xy denote the concatenation of strings x and y. Li and coworkers [LBC*01] defined the normalized distance between x and y

$$d(x,y) = 1 - \frac{K(x) - K(x|y)}{K(xy)} \tag{3}$$

and showed that $d(x,y)$ is a metric up to logarithmic additive terms.

Let $C(x)$ denote the length of the compressed version of a string x. Cilibrasi and Vitányi have shown that the normalized compression distance

$$NCD(x,y) = \frac{C(xy) - \min\{C(x),C(y)\}}{max\{C(x),C(y)\}} \tag{4}$$

is a quasi-universal metric when the compressor in use is normal.

Kolmogorov complexity have been applied theoretical computer science [PSS81], clustering [CV05], plagiarism detection [CLMS02], phylogeny [LBC*01] and others [CV05].

In this work we used the LZW compression algorithm [Wel84] to evaluate an approximation of $d(x,y)$ denoted $d'(x,y)$:

$$d'(x,y) = 1 - \frac{C(y) - C(y|x)}{C(xy)} \tag{5}$$

We have evaluated the terms $C(y)$, $C(y|x)$ and $C(xy)$ as follows. The term $C(y)$ is the number of table positions that LZW would output while compressing y minus the number of table positions that LZW would output when the input is a string of $|y|$ equal symbols plus one. The subtraction is a tentative to overcome the limitations imposed by the very nature of LZW. The term $C(xy)$ is evaluated in the same way. The term $C(y|x)$ is the number of table positions that LZW

would output while compressing y, without counting the positions that belong to the compression table for x. That is, a table position is counted only if it does not belong to the table constructed during the compression of x. Our algorithm was implemented in Perl, using a hash for the table.

In our tests, we have also used the CompLearn [Paga] package together with gzip [Pagb] (chosen for speed) to evaluate NCD. We have noticed slightly better results using $d'(x,y)$ (see Section A.5), although with poorer performance.

Generating a distance table for n texts $t_1, t_2, \ldots, t_n$ with lengths $l_1, l_2, \ldots, l_n$ requires computing $C(t_i)$ for every text, and $C(t_i|t_j)$ and $C(t_i t_j)$ for every pair such that $i \neq j$. Then the cost of the distance table construction is $O(s^2)$ on the average, where $s = \sum_{i=1}^{n} l_i$ and with average cost $O(1)$ for a hash operation.

Adding a new text t_{n+1} to the set does not require recomputing the whole matrix. If we store the values of $C(t_i)$ and the LZW table produced for every t_i, it is enough to calculate $C(t_i|t_{n+1})$ and $C(t_i t_{n+1})$ for $1 \leq i \leq n$, at cost $O(s+l_{i+1})$ on the average. The length of LZW table for a text of length l is $O(l)$, so storing them requires an affordable amount of space.

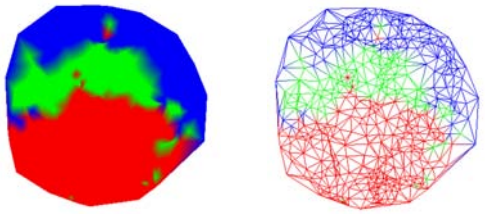
A.5. Results

In the remaining of this text we shall refer to the similarity calculation based on NCD (Equation 4) as NCD and we shall refer to our similarity calculation based on compression by LZW (Equation 5) as k-lzw. In general, we refer to distance calculation based on Kolmogorov complexity approximation Kolmogorov distances or k-distances.

One of the goals of our maps is being able to distribute the files onto a surface automatically, allowing proximity of closely related contents, association between documents and groups of documents, and display of additional information related to them on the same map. By doing projection based on distance metrics it has been possible to obtain results that match those goals, provided the distance measure is capable of coding well content relationships.

To evaluate the ability of a particular distance measure to map corpus content, we have used text collections previously classified as belonging to general areas of knowledge. This classification is based purely on the source of the papers, and therefore some degree of overlapping is expected due to the presence of similar concepts or techniques cross-areas.

Pseudo-classes on the maps were colored to reflect mapping results visually, showing how those documents of the same class were distributed over the surface. Figure 15 shows the result of one final map from k-lzw for a corpus made out of papers from three basic areas: Inductive Logic Programming (ILP), Case-based Reasoning (CBR) and Information Retrieval (IR). It shows that this mapping is capable of keeping most of the documents of the same class in the same region of the map.



(a) Flat surface. (b) Wireframe Model.

Figure 15: *Mapping of scientific documents related to three different primary sources (CBR is red; ILP is green; IR is blue.).*

Processing times for calculation of k-distances for a whole data set are quite high. The set in Figure 15, comprising 574 files took 2h30min to process from scratch in a Pentium IV processor of 3GHz. Opposite to conventional techniques based on frequency count, though, this type of processing is incremental and adding new documents does not require recalculation of previously obtained results. Only the new distances must be processed as databases increase in size, as mentioned in Section A.4.

Table 1 gives details of the documents used to process the remaining tests. The first scientific papers data set (corpus1) included title, authors, abstract and references from a number of texts. CBR and ILP subsets were taken from journals on those subjects. The IR and SON (sonification) subsets were articles obtained as a result of Internet searches pre-filtered to comply to those pseudo-classes. Those were all collected by members of our team. The corpus2 set was recovered from an Internet repository and comprehends files in the ISI format on the subjects of Bibliographic Coupling (BC), Cocitation Analysis (SC), Milgrams (MG) and Information Visualization (IV)(from ella.slis.indiana.edu/~ katy/outgoing/hitcite/{bc,sc,mb,iv}.txt). The remaining sets are messages from news discussion groups recovered from an internet repository (see Hettich, S. and Bay, S. D. (1999). The UCI KDD Archive [http://kdd.ics.uci.edu]. Irvine, CA: University of California, Department of Information and Computer Science).

Previous results have shown that distance based projections of text collections based on vector representation with conventional distance metrics (such as cosine) can provide a point placement with good separation between general subject areas as well as good grouping of similar documents. This process, that entails vector representation possibly followed by k-means attribute combination, followed by projection with improvement scheme, has been called IDMAP (Interactive Document Map) [MPL06]. We call the mapping realized here (projection with improvement scheme based on

Table 1: *Datasets used in the tests*

Set	Areas	General Content	Files
corpus1	CBR+IR+ILP+SON	Scientific Documents	675
corpus2	SC+BC+MG+IV	ISI Files	1624
message1	atheism+graphics	discussion group messages	200
message2	atheism+graphics +baseball	discussion groups messages	300
message3	seven varied subjects	discussion group messages	700
message4	ten varied subjects	discussion group messages	1000

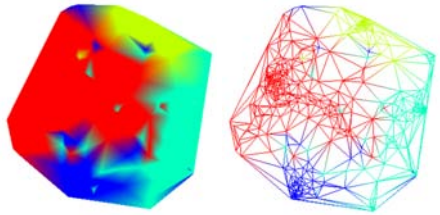
k-distances) Kolmomap. We compare these two approaches for visual result.

Figures 16 and 17 show the visual results of some of the data sets in both IDMAP and Kolmomap. It can be seen that the result is comparable to that obtained with IDMAP in terms of region placement and sub-grouping. Separation is better with IDMAP in some cases, but it tends to jam similar documents into pockets, making it more difficult to interpret neighboring relationships in dense regions.

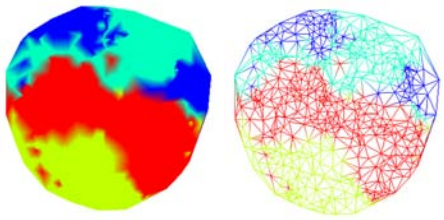
In 'more generic texts' the intrinsic content relationship is a lot less obvious than in academic or scientific papers. Figure 17 shows the generation of Kolmomaps and IDMAPs from messages in discussion groups (data sets message1 and message2 in Table 1). Pseudo-class in this case is the theme of the discussion groups. The maps show that Kolmomaps are capable of distinguishing subjects in that context much better than IDMAPs.

Once the range of themes starts to broaden and specialize, Kolmomaps tend to mix messages coming from distinct subgroups. This can be understood in the light of the fact that discourse tends to be full of common expressions and terms, regardless of the subject under discussion. Figure 18 shows that fact.

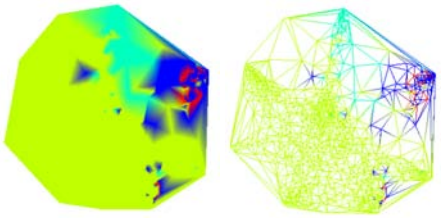
Observing the pictures in Figure 18, it can be noted that Kolmomaps can still determine various 'pockets' of similar documents and also that the mixture is not uniform, that is, at least in the body with 700 messages, there is overall mixture of 'reds' and 'blues' as well as 'greens' and 'yellows'. Blues and reds are sci.space and alt.religion, and greens and yellows are various 'comp' subjects and 'forsale' subjects. Therefore even with lesser distinction between pseudo-classes than the 3-theme case, there is an underlying pattern reflected by the Kolmogorov distances. IDMAP, as might be expected from the previous message tests, was not capable of any significant separation between subjects.



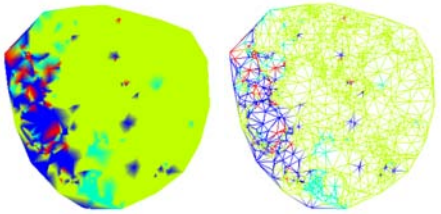
(a) IDMAP results from corpus1 set.



(b) Kolmomap results from corpus1 set.

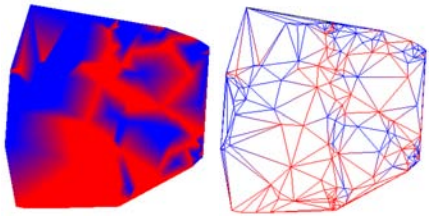


(c) IDMAP results from corpus2 set.

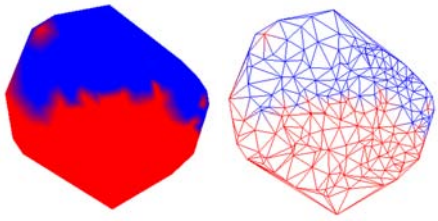


(d) Kolmomap results from corpus2 set. In dataset corpus2, one part of the corpus - in yellow - 1236 files - is a lot larger than any of the other three

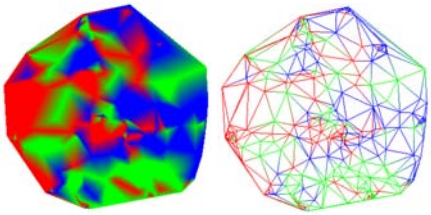
Figure 16: *IDMAP and Kolmomap of the two bibliographic data sets. See Table 1.*



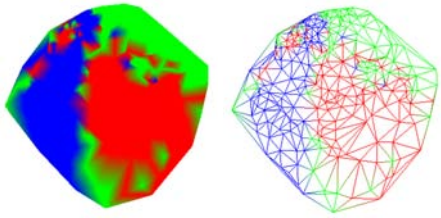
(a) IDMAPs of message1.



(b) Kolmomaps of message1.

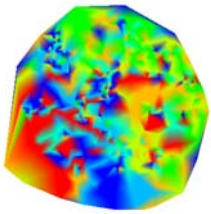


(c) IDMAPs of message2.

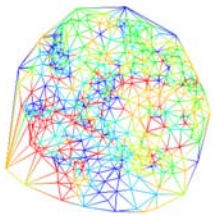


(d) Kolmomaps of message2.

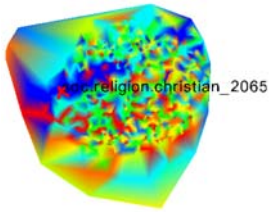
Figure 17: *IDMAP and Kolmomap of the newsgroups data sets.*



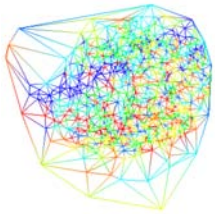
(a)



(b)

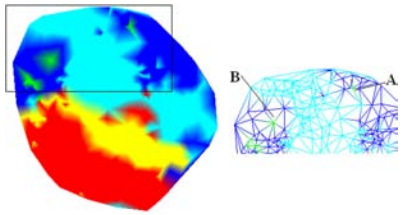


(c)

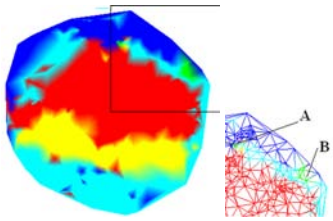


(d)

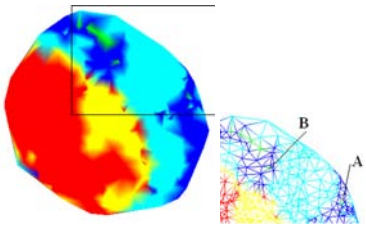
Figure 18: *Kolmomaps of message3 and message4 data sets.*



(a) Kolmomap



(b) maps based on NCD estimation.

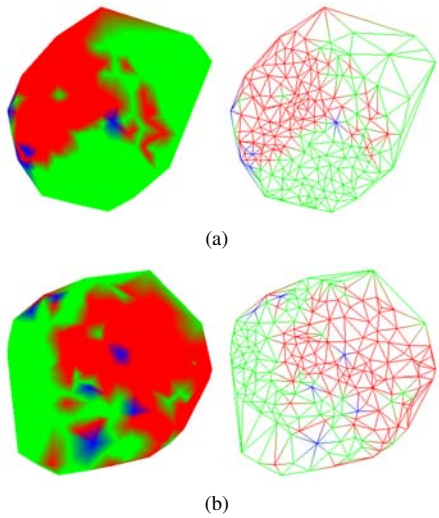


(c) IDMAPs.

Figure 19: *Maps with highly correlated 'intruders' (in green)a).*

Detailed exploration of various maps have shown a close relationship between proximity in the map and the expected proximity of document content with few exceptions.

In order to test the capability of content based point placement, two tests were performed. In one of them 5 'intruders' (documents not previously processed in the set) were added to the map. They belonged to the general class of sonification, but were not in the initial set. All five had at least two common authors and represented and evolution of the same sonification system over the years. Additionally to those, two other papers were added. Again two of the authors at least were repeated, but the main subject of the paper was not sonification. One of them mentioned sonification in-



(a)

(b)

Figure 20: *News-maps with uncorrelated 'intruders' (in blue).*

cidentally (paper A), the other didn't (paper B). Figure 19 shows the map for that test. In that case the Kolmomaps projected those related papers together (Figures 19a and 19b). The papers that were not primarily on sonification were still mapped within the general sonification pseudo-class, indicating the importance of author's names (and most likely some self-reference too) in the context. However, it can be seen that, for the k-lzw implementation, the fact that the content itself was diverse from that of the neighbors, made paper B push them away and form an isle within the group. NCD also provided approximation between most of the related papers, but provided less distinction for that text with content diverse from sonification. IDMAP did not perform as well (Figure 19a), placing the files within the scope of the general sonification groups but making less distinction of those without direct relation with sonification.

The second test meant to compare the two different approximations of Kolmogorov distances (k-lzw and NCD) further. In this test, we added intruders (also 5) in the messages test set message1 that were not related to any of the previous two newsgroups. The two groups were comp.graphics and alt.atheism and the intruders (in blue) belonged to the theme of rec.sports.baseball. Figure 20 presents the results of that placement. It shows that for the k-lzw map the new points are pushed away from the previously existing sets of messages (either to the border of the map, or, in one case, between the two groups). NCD did not perform as well, mixing half of the intruders within either one of the previously formed groups.

Projection errors according to calculations published in a previous work [TMN03] are presented in Table 2. Those values support the evidence that the mixture in maps of larger

Table 2: *Approximate Projection Errors*

Set	k-lzw	NCD	IDMAP
corpus 1	0.25	0.33	0.2
corpus 2	0.25	0.32	0.17
message1	0.17	0.25	0.21
message2	0.17	0.25	0.21
message3	0.23	0.3	0.18
message4	0.17	0.25	0.17

human communication files (news groups with larger number of subjects and messages) is caused by message content instead of the projection itself.

Assigning attributes to the map components (vertices, for instance) can show one or more additional degrees of information on top of the landscape map. Various attributes can be visually mapped to color, height, isocurves, etc., allowing further analysis of relationships between the documents represented in the map. Previous maps in this text have shown pseudo-class by using color. Figure 21a shows the same data mapped redundantly to height for the corpus1 dataset. In that case, different classes are 'plateau layers' in the map and misplaced (or pseudo-misplaced as the case may be) points can be easily located as peaks or depressions in the topography of the maps.

To highlight proximity relationships further, one attribute mapping found useful was to perform a hierarchical clustering of the projected data. A hierarchical clustering can group points closer together in various levels. By doing that, and then showing the depth of the cluster on the maps, it is possible to have visual aid to locate pockets of closely related documents. Figure 21b illustrates the corpus1 data set after submission to single-link hierarchical clustering. Both color and height are mapping the depth of the document in the clustering tree. The dark blue (and highest) areas show where the points are closer together; the following levels or grouping are lighter blue, green, orange, yellow, and red (lowest), the latter showing the points (documents) that are most isolated from their 'neighbors'.

The former mappings can be combined to produce composed visualizations of multiple attributes. Figure 21c shows clustering mapped to height and class mapped to color for the same dataset, and Figure 21d shows the mapping inverse to that (height is class, color is cluster).

The map can be interacted with for exploration. In our case, a tool for exploration of that type of representation, called Spider Cursor, is under development. The current version allows walking over the surface using a cursor that highlights the edges to neighbors (see Figure 21e). It also allows the user to choose how numeric attributes will be mapped to visual attributes such as color, height, level curves, glyphs and sound. Paths can be marked, selected and cut to extract parts of the map (Figure 21f). One attribute, possibly literal,

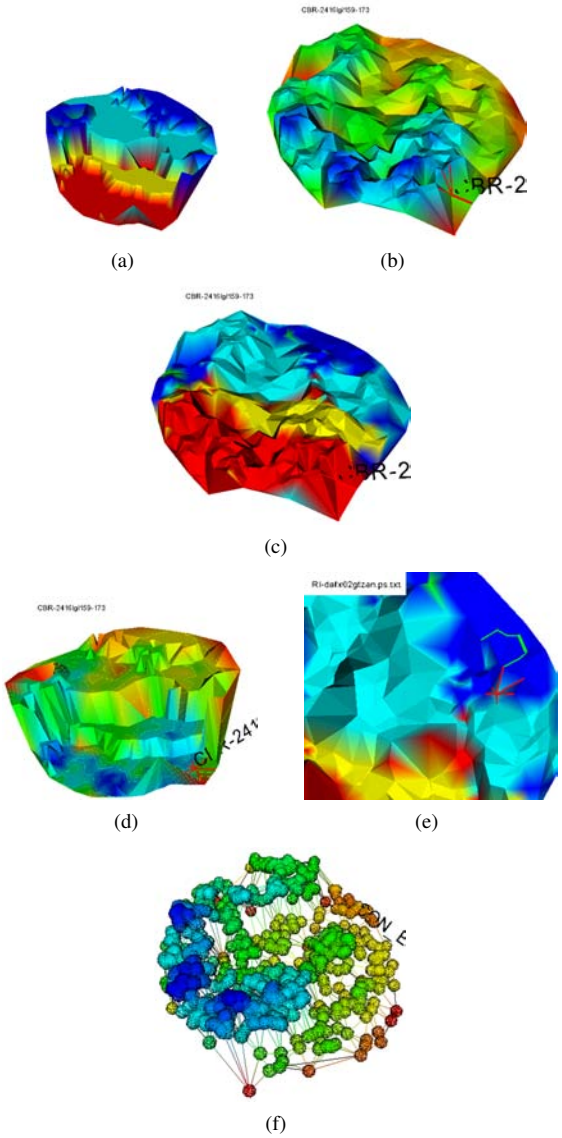


Figure 21: *Various possibilities of attribute mapping and interaction with the corpus1 map.*

is shown on a window on the map. In our examples, they were the labels of the files containing the document.

In visual terms, Kolmomaps were very successful both in distributing documents in larger subject areas and in grouping files with similar contents. The fact that it does not require term analysis makes it a good option to process (and organize) corpora of documents to be analyzed in an interactive session.

The process of forming the distance matrix for a whole corpus is costly, particularly if processed from scratch. Table 3 shows the times to process each data set in full at once, with no intermediate storage (that is, the matrix was not built incrementally). The projection itself, including hierarchical clustering, is done in a matter of few seconds.

Table 3: *Times for distance calculation (h - hours, m - minutes)*

Processor Pentium IV - 3GHz		
Set	k-lzw	NCD
corpus 2	11.3 h	5.5h
message3	2h	17 m
message4	4.2h	39 m

A.6. Conclusions

The technique presented here of similarity calculation via Kolmogorov complexity taken from raw texts has proven useful in conjunction with distance-based projections to build maps of texts. Those distances have shown to separate general content areas and to generate proximity between similar texts, facts reflected by the projection based on the distances.

Kolmomaps have shown to work well for texts in a variety of contexts. It is a good step towards helping text organization by content in groups of documents in a context, where selection of pre-filtered reading material is necessary (such as research, education, training in technology etc.).

Kolmomaps have high computational cost in it distance calculation stage. However, some facts about the time to generate maps of documents based on distances by Kolmogorov complexity estimation should be phrased:

- No pre-processing is needed. The approach allows immediate application to text documents without the need to realize (or understand) vector transformation. No lengthy editions are necessary either. That in itself saves time in the whole analysis process.
- Opposed to vector representation schemes, this process is incremental. For data sets that grow – and they usually do –, it is possible to store intermediate distances and symbol tables, speeding up the calculation for a new text to be added.
- The projections are actually very fast, in the order of a few seconds (including improvement, hierarchical clustering and display).
- Although NCD is faster, k-lzw estimation has resulted in more consistent map behavior. It was also noted that k-lzw time increase showed more regular behavior than NCD as the document bodies grew larger.

The maps using k-distances compared well with IDMAPs. Interpretation of the resulting maps is easy to learn. In a short time, users learn that being close means related content, and

longer edges mean that the distance is larger than to those with smaller edges.

Further work is planned in analyzing a number of text corpora, in data structures for storing and recovering map results, and in adding extra semantic layers on top of this map to reflect other dimensions, such as co-citation and relevance. To our knowledge, the conditional Kolmogorov complexity was not evaluated our way before, so it remains to show that this measure is a (quasi-)metric, as the experimental results suggest.

Scalability to larger data sets was not an issue here. Rather, we have been researching methods to help distinguish important relationships in text data sets of a manageable size but still too large to have the user make sense of it on his/her own efficiently. However, we believe that with that the approach can be scalable a further level, by developing an incremental data organization approach to go with the mappings.

B. Technical Report: CONTENT-BASED DOCUMENT MAPS USING FAST PROJECTIONS AND TOPIC DETECTION

Fernando V. Paulovich, Rosane Minghim, Roberto Pinho, and Alneu A. Lopes

{paulovic,rminghim,rpinho,alneu}@icmc.usp.br
ICMC - University of São Paulo,Brazil

This article presents an approach to build visual maps of text collections based on their content and using multidimensional projections. A tool, called Projection Explorer (PEx) was built to support the steps of this approach. We also present two novel techniques to integrate text mining into the exploration process that help extract topics discussed within the document collection. The first is based on term co-variance and the second is based on seeded generation and weighted filtering of association rules. Topics and subtopics in groups of documents are extracted and this is coupled with the capability of various levels of exploration of the map. This combination of techniques is demonstrated to help the user to gather a general view of the content of a document collection as well as locate and relate subtleties within a textual data set.

B.1. Multidimensional Projections for Mapping Collections of Documents

Examining text is crucial for many different types of applications. Even applications that rely on data other than textual (such as images, signals, simulations) sometimes have alternative text-based output. The troublesome task of interpreting content and extracting useful information from a document collection is the target of efforts in various areas of computer science. The combination of Text Mining and Visualization (or Visual Text Mining — VTM) is concerned with developing tools to support users extracting meaning, drawing conclusions and making decisions without extensive reading, since it is hardly possible to rely on actually going through content in detail due to time restraints found in most applications.

We describe here recent results towards building visual maps of documents based on their content through an integration of mining and visualization algorithms and techniques. The underlying approach taken here defines a placement of the documents on a plane taking into account their content.

Based on the computation of similarity amongst textual documents, they are projected as points onto a two-dimensional space with the target that similar documents should be projected close to each other and that dissimilar documents should end up far apart.

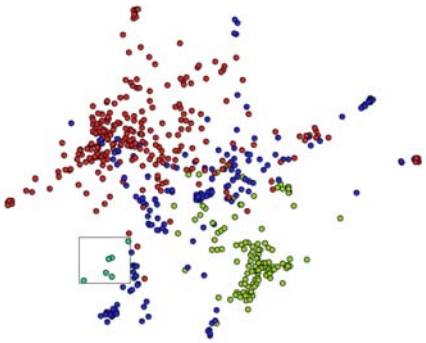
In order to create this graphical representation, a multidimensional projection technique is used. Formally, let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of m-dimensional data, with $\delta(x_i, x_j)$ a dissimilarity (distance) measure between two m-dimensional data instances, and let $Y = \{y_1, y_2, \ldots, y_n\}$ be a set of points into a p-dimensional space, with $p = \{1, 2, 3\}$ and $d(y_i, y_j)$ a (Euclidean) distance between two points of the projected space. A multidimensional projection technique can be described as a bijective function $f : X \rightarrow Y$ that seeks to make $|\delta(x_i, x_j) - d(f(x_i), f(x_j))|$ as close to zero as possible, $\forall x_i, x_j \in X$ [TMN03].

The result is a planar mapping of a point set X representing the documents (we call this 2D placement a *map*), which can be used to explore the document collection represented by X. As it is going to be illustrated in the course of this text, the techniques we have developed are capable of mapping documents in such a way that text data dealing with common subjects form groups that are visually apart from other groups, thus allowing identification of patterns in the text set. An example of such a map is given in Figure22. It shows a document map of a collection of 574 scientific papers belonging to three different subjects, previously manually classified. In the pictures, different classes have different colors (red is Case-Based Reasoning (CBR), blue is Information Retrieval (IR), and green is Inductive Logic Programming (ILP)). The small greenish group in the rectangle shows the placement of five papers on sonification (the use of sound to display information). That particular group refers to five papers reflecting the evolution of a sonification system we have developed, and are strongly correlated. Figure22a) shows that they are placed adequately in the same neighborhood. Figure22c) demonstrates that, when mapped in conjunction with the remaining papers, this group was placed in the outer boundary of the map and nearby papers in the data set whose main subject was audio retrieval (they are the points with colors other than red in Figure22c). Another aspect that can be noted in the map is the fact that the groups of CBR and ILP are more compact, whilst the IR set is spread out on the map, a visual analogy to the characteristics of the subject areas, the fist two being more mature and therefore having a set of own nomenclature and techniques, whereas IR relies on resources from various sources, including those of ILP and CBR.

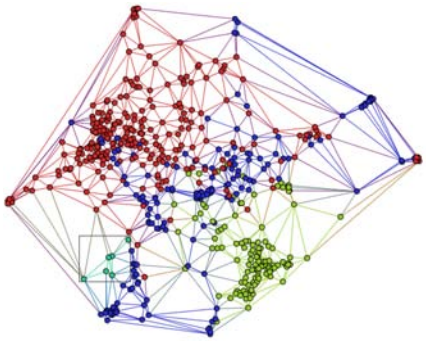
Multidimensional projection techniques can be classified into two major groups, according to the functions f employed: *linear projection techniques*; and *nonlinear projection techniques*.

Linear projection techniques, such as *Principal Component Analysis (PCA)* [Jol86], create linear combinations of the data attributes, defining them in a new orthogonal basis of small dimension. Although this type of technique performs well on Gaussian data, in handling data with nonlinear structures, such as clusters of arbitrary shapes or curved manifolds, it typically fails to capture relevant patterns. In such cases, nonlinear techniques are better candidates.

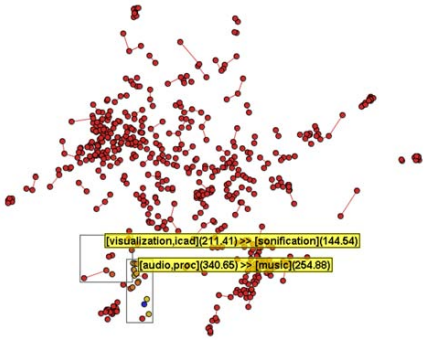
Rather than relying on linear combinations of the attributes, nonlinear techniques attempt to minimize a func-



(a) Point Placement for the text collection. Detail is the sonification test group.



(b) Delaunay Triangulation of Point placement in Figure B.1 .



(c) Placement of sonification papers midst papers of audio information retrieval papers

Figure 22: *Document map of a collection of scientific papers belonging to three different main subjects, manually classified, plus a test group (inside the rectangle).*

tion of the information loss incurred in the projection. Normally, this function is based on the dissimilarities amongst the m-dimensional instances and on distances among the p-dimensional points.

One example of a nonlinear projection technique is *Multidimensional Scaling (MDS)* [CC00]. Sprang from the psychophysics domain, MDS actually comprises a class of techniques aimed at mapping instances belonging to an m-dimensional space into instances on a p-dimensional space ($p \leq m$), striving to keep some distance relations.

Amongst the various MDS techniques, the simplest ones are those based on *Force-Directed Placement (FDP)* [FR91, Cha96]. Originally proposed as a graph drawing heuristic, the FDP model aims at bringing a system composed of instances connected by imaginary springs into an equilibrium state. Instances are initially placed randomly and the spring forces iteratively push and pull them until reaching an equilibrium. To apply the FDP model as an MDS technique the spring forces must be proportional to the difference between the dissimilarity $\delta(x_i, x_j)$ among the m-dimensional instances, and the distances $d(f(x_i), f(x_j))$ among the p-dimensional points.

We have developed two distinct nonlinear projection techniques: *Projection by Clustering (ProjClus)* [PM06] and *Least-Square Projection (LSP)* [PNML06].

The core idea of ProjClus is to create and project clusters of m-dimensional points. After initial clustering, the ProjClus algorithm calculates the centroids of these clusters, and projects these centroids onto the plane. Next, it separately projects each set of points defined by the clusters to the plane; and finally the technique assembles the final layout positioning the clusters' projections according to the projection of their centroids.

The LSP technique is a generalization of an approach for mesh-recovering and mesh-editing in order to deal with high dimensional spaces. In this technique, a subset of m-dimensional points are projected onto the plane, and the remaining points are projected using an interpolation strategy that considers only the neighborhood of the m-dimensional points.

ProjClus and LSP are high-precision and fast projection techniques that are suitable to handle points belonging to nonlinear sparse spaces — such as the one generated from document vector representations (see Section B.2.3) — since they take into account neighborhoods of the original points.

Both LSP and ProjClus are available within a tool called Projection Explorer (PEx), available at: http://www.lcad.icmc.usp.br/∼paulovic/pex. PEx makes available a set of tools to support every step and parametrization of the maps presented here, as well as a number of graph based and text based exploration tools. After high quality projection, exploration based on document content can take place in a more friendly environment. For that to occur, we have developed a number of novel techniques for helping the user to extract the main subjects being discussed and to focus on a particular area to locate details of interest.

B.2. Exploring Content-based Document Maps

This section presents two novel techniques (and its support interactive tools) for visual analysis of content based on topic extraction. Covariance and Generation of Association Rules are the basis for these techniques.

B.2.1. Topic Identification by Covariance

Although a projection can help locating groups of related documents, the initial map does not reveal information within these groups. In order to extract information about these groups, a group can be selected and a label can be generated which aims at identifying the main topics discussed within that group. A covariance-based label is composed by the three distinct terms that have the highest covariance considering only the selected documents. The covariance of two terms is a measure of the degree to which they vary together. The covariance is calculated according to Equation 6.

$$cov(t_i, t_j) = \frac{1}{n-1} \sum_{k=1}^{n} (t_{ki} - \bar{t}_i)(t_{kj} - \bar{t}_j) \qquad (6)$$

where $\bar{t}_i$ is the mean of the i^{th} term t_i, and t_{ki} and t_{kj} are the values of the i^{th} and j^{th} terms for the k^{th} document.

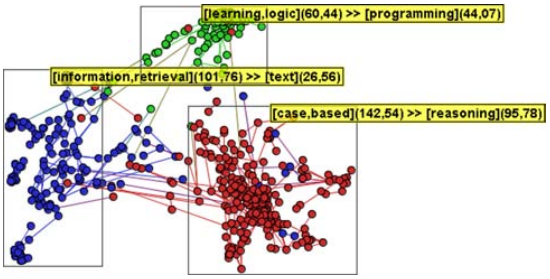
Figure 23 shows another map of the same data set of Figure 22, showing the resulting labels after group selection by the user.

The first two terms on the left side of a label are the ones that present the greatest covariance. The third one on the right side is the term which has the greatest *mean covariance* according to the labels already chosen. The mean covariance is calculated as the mean between the covariances of a term taking two other terms. The numbers between parenthesis are the covariances.

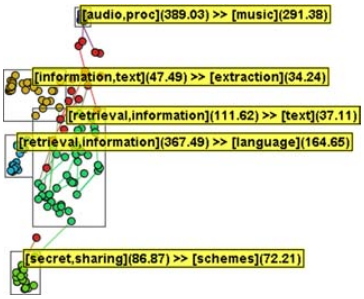
If the covariance approaches zero, it means that the terms are independent. On the label, this number can indicate if the three terms are equally related or if only the first two are. If the number on the left side is much greater than the number on the right side, it means that the third term is not deeply related to the other two. But if these numbers are close, it means that the three terms occur frequently together.

For instance, in Figure 23(b), considering the label *[learning,logic](60.44) >> [programming](44.07)* and dividing the right covariance by the left we get $\frac{44.07}{60.44} \approx 0.73$. Calculating the same rate for the label *[information,retrieval](101.76) >> [text](26.56)*, we get $\frac{26.56}{101.76} \approx 0.26$. Thus, it is possible to infer that the terms "learning" and "logic" are much more related with "programming" than the terms "information" and "retrieval" are with "text".

Notice that the covariance is not a causality relation between terms, that is, it does not indicate that one term occurs due to the other one. It is only a measure of the degree to which two terms vary together. For example, on the map of Figure 23(b), which is a view of part of the Figure 23(a), the terms "secret" and "sharing" present a high degree of covariance, but it is not possible to establish that the term "secret" will cause the term "sharing" or the other way round. Another important aspect is that using two terms only, it is not normally possible to direct infer why they are related. This is the point where the third term of the label can help. The third term for the terms "secret" and "sharing" is "schemes", thus it is possible to conclude that the main topic of the document used to create such label is somehow related to cryptography.



(a) Document map of a scientific papers' collection in three different subjects.



(b) A view of part of the map in Figure 23(a) .

Figure 23: *Example of labels for document maps.*

Inside PEx there are two ways to define the groups of documents used to generate labels. Either the user selects a rectangular area on the projection or a clustering of the documents after projection is performed to suggest a particular grouping and generate labels all at once. Two types of clustering are available in PEx, k-means and density based clustering. The labels presented on Figure 23(b) were automatically created using the density based clustering approach.

Figure 24 presents a map of a data set composed by text files including titles, authors and abstracts of all papers published by IEEE InfoVis Conference from year 2000 to 2005 and by the International Conference on Information Visualisation (IV), also published by IEEE CS Press, from year 2000 to 2006. 1020 textual documents are mapped. Some selected groups and their labels by co-variance are presented

in that picture. Darker nodes are InfoVis papers and lighter ones are IV papers. It could be observed that doing a search within the map, papers about graph based visualization, image based techniques, network, web and various others subjects concentrated on small neighborhoods in the map. Also, a certain distribution of subjects between the conferences can be observed, revealing the types of paper presented in each.

The process employed to explore a document map using covariance-based labels is a step-wise refinement process. For example, the view presented in Figure 23(b) is labeled as *[information,retrieval](101,76) >> [text](26,56)* in Figure 23(a). Thus, when such group is explored in detail, the words "information" and "retrieval" can be ignored for calculating the covariance since we already know that these documents deal mainly with "information" and "retrieval". The third term "text" is not ignored in this analysis since the right covariance is much lower than the left one. In the case where the two covariances present similar values, the three terms can be discarded. Lists of terms to be ignored when extracting topics can be edited by users of PEx.

Although co-variance based topic extraction is very useful for a first view of the possible subjects around the map, further information can be automatically extracted to support revealing subtleties and laying out more subjects under discussion in a particular group. The next technique contributes to achieving precisely that goal.

B.2.2. Topic Extraction by Sequential Covering Induction of Association Rules

An appropriate set of association rules derived from a collection of text documents can be used to describe a context in which a term appears or also the context or topic related to a subset of documents. When mining association rules from text, an association rule (AR) is an implication of the form $X \Rightarrow Y$, where $X \cap Y = \emptyset$ and they are both subsets of L, where $L = l_1, l_2, \ldots, l_m$ is a set of literals or items (each item representing a term from the bag of words). A transaction T is a set of items $T \subseteq L$ that represents a document from the corpus C. The rule $X \Rightarrow Y$ holds in the document set C with confidence c if $c\%$ of the documents in C that contain X also contain Y. The rule $X \Rightarrow Y$ has support $s\%$ in C, if $s\%$ of documents in C contain $X \cup Y$.

A rule R_i: $X \Rightarrow Y$ with significant support in C means that a set of terms $l_i \in X \cup Y$ are found together in a large subset of documents from C. Moreover, if R_i has high confidence, then the occurrence of X (the body) means high probability for the occurrence of Y (the head), at least in C.

In a previous work [LPMP07], we presented an algorithm to produce and select good association rules in order to describe the main subject or topic in a selected set of documents S_k. That algorithm deals with the problem of the large number of association rules as follows. Instead of post-pruning the set of rules generated from all documents in S_k by some rule quality measure, it generates rules which contain at least one term (seed) present in a selected set of most relevant terms. The term relevance is given by a weight that favors terms with higher frequency in the selection than in the rest of the corpus. An arbitrary number of rules, usually 1–3, with the highest term weight summation is then selected for display as label for the group selected. Eventually, some documents carry topics that are not described by these selected rules, that is, they are not covered by the rule.

Here, we present a new algorithm that uses a sequential covering strategy in order to extract topic targeted rules from the documents uncovered by the main rule. Additionally, instead of just selecting the n most weighted rules, we keep only those rules which provide additional covering over the previous ones. This reduces the amount of redundancy found in the selected rules, considering that if two rules cover the same subset of documents, only the one with the highest weight is kept.

The Algorithm (detailed below) has two nested loops. In the inner loop, the iterative algorithm for generating and ranking rules over a selected area S_k of the projection leads to a set of rules SR. This process is repeated (outer loop) removing covered documents during the iteration until no remaining document is left or the rule generation process outputs an empty set of rules. The function coverage(AR_i, U) on Algorithm 1 denotes the number of documents from U that support the rule AR_i.

Even considering that each selected rule must cover documents not covered by any other previously selected rule, a considerable number of rules may be generated for a given selection of documents. To reduce cluttering, we apply a rule grouping strategy that relies both on literals in the rule and on overlap of covered documents. Intuitively, if rules share some of their literals and cover roughly the same set of documents, they are likely to represent the same topic, but described in a slightly different way. Thus, instead of presenting a label for each rule, only one label per group of rules is displayed. A rule AR_i joins a group of rules G_j, if (i) AR_i has at least one term $t \in T_{G_j}$, where T_{G_j} is the set of literals found in rules already in the group G_j, and (ii) there is an overlap between the set of documents covered by the rule, and the set of documents that are covered by any of the rules in the group. This document coverage overlap should be equal or higher than a given constant α. Here, we used $\alpha = 0.5$. For computing the overlap, we use:

$$overlap = \frac{|\cap D_{AR_i}, D_{G_j}|}{min(\{|D_{AR_i}|, |D_{G_j}|\})}$$

where D_{AR_i} is the set of documents covered by the rule AR_i, and

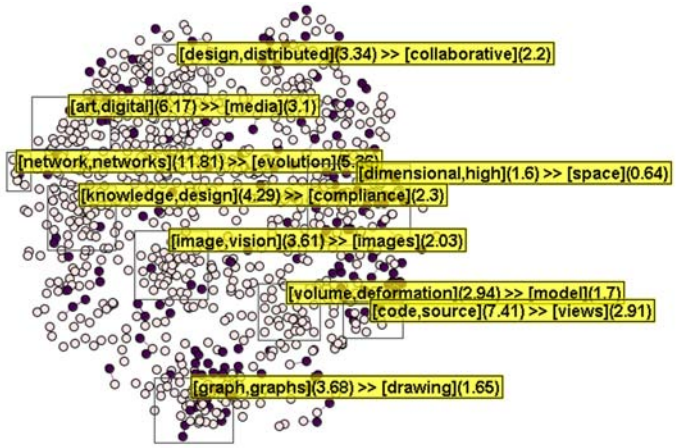


Figure 24: *A map of IV and InfoVis Papers from 2000 to 2006.*

Algorithm 1 - Sequential Covering of document selection

Input: S_k % k selected points
% (document × term matrix ($M_{k,l}$))
Corpus C % n points, (document × term matrix ($M_{n,l}$))
Bag_of_words % l literals of matrix document × term
Output: RuleSet % Set of rules that covers selection S_k

Algorithm
$RuleSet = \emptyset$
$U \leftarrow S_k$ % uncovered documents
do
    $SR \leftarrow$ call iterative generation and rank of rules over U [LPMP07]
    $NR \leftarrow |SR|$ % cardinality of SR
    //rule selection
    while($SR \neq \emptyset$)
        remove most ranked rule AR_i from SR
        if coverage$(AR_i, U) > 0$
            $RuleSet = RuleSet \cup \{AR_i\}$
            $U = U -$ Set of documents covered by AR_i
        end if
    end while
while($U \neq \emptyset$ and $NR > 0$)
return $RuleSet$

$$D_{G_j} = \bigcup_{AR_i \in G_j} D_{AR_i}$$

If a rule does not join any group, a new group is formed with that rule as an initial element. The label for a group of rules begins with the highest weighting term taken from the rule with the highest support in the group. The label shows also this high support rule followed by all the terms that are found in the other rules of the group and that have not yet appeared. For instance, the label "mosque:[suicide]←[mosque], (27/20%;90%):iraq:baghdad:" is given to a rule group where the rule "[suicide]←[mosque]" has the highest support among the rules in that group (27 documents or 20% of selected points S_k), "mosque" is its term with the highest weight, and 90% is its confidence. In the example, "iraq" and "baghdad" are the terms found in other rules from the group that does not appear in the highest support rule.

To demonstrate the potential of this technique in aiding document collection exploration, we used a corpus that is composed by 2684 RSS news feed articles, collected from

BBC, Reuters, CNN, and Associated Press sites, during two days in April 2006. Figure 25 presents the whole set of documents colored according to a clustering algorithm in PEx. Labels have been generated for each of the clusters and a few of them were selected for display.

Figure B.2.2 zooms in on a region that has almost all results for the search "iraq". Over that display, a user selected two regions that were richer in documents with "iraq". The labels shown (Figure B.2.2) reveal that the two regions indeed refer to the topic "iraq" from different perspectives. The region on top clearly deals with terror attacks, while the one on the bottom is more concerned about political and legal aspects of the issue. Further inspecting the top region, the user then selected two small clusters of documents inside it, which allowed him to discover that each of them dealt with a particular terror attack (Figure B.2.2). In the latter figure documents were colored using terms from the two main rules from the display labels in a boolean search query ("(mosque and suicide) or (car and bomb)"). This was done by a text search operation of PEx, in order to highlight documents covered on both selections. The same is true for individual rules, if the user chooses not to use the rule grouping strategy. Using PEx, the user can inspected one or several documents from the selected regions and confirm what has been hinted by the labels.

In the following text, the complete process from text to interaction supported by these tools is summarized.

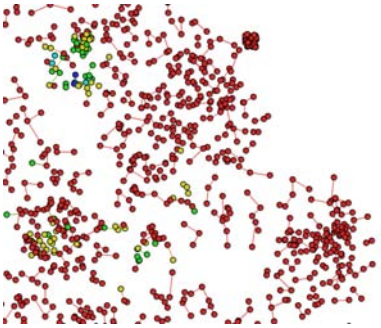
B.2.3. Overview of the Mapping Generation Process

This section provides some details on the steps to build a map from a document collection by PEx. From the data set (given as separate files for each document in ASCII format), the process followed to generate and explore content-based document maps is summarized in Figure 27. This process is composed by two main tasks, *creating the map* and *exploring the map*

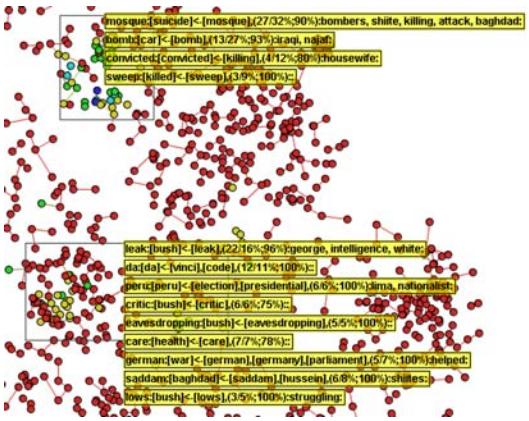
After text data collection, the documents are pre-processed (1) in order to establish their vector representation. This process converts each document into a vector in a multidimensional space (each term or combination of term representing a dimension), and uses a vector-based distance metric to compute the dissimilarities between documents. The steps to accomplish text pre-processing are [WIZD04]:

1. Non-representative words (stopwords) are eliminated from the documents, such as prepositions, articles, and other words not relevant to context;
2. The remaining words are converted into their radicals (stemming) using *Porter's* stemming algorithm;
3. The terms are counted in each document in order to determine their frequencies;
4. Luhn´s upper and lower cut-off are applied to eliminate terms that are too frequent and too rare
5. The terms frequencies on each document are weighted according to *term-frequency inverse document-frequency (tfidf)* measure. In tfidf, if a term appears in almost every document, its "representation power" is decreased since such term is not capable of distinguish individual documents on the collection.
6. The vectors formed are normalized in order to have unity size. Thus, documents that have different sizes but similar content can be correctly compared.

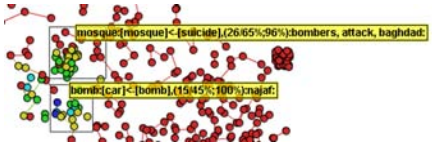
PEx offers various visual tools in order to help the user to choose the parameters of this documents process.



(a) Color by search "iraq". Red documents have no occurrence of the word.



(b) Two user-selected regions with corresponding labels for groups of rules.



(c) Detailed Zoom and selection of two regions. Documents colored by search to highlight documents that support either the rule "mosque ← suicide" or "car ← bomb"

Figure 26: *Zoom and selections over document map of 2684 RSS news feed articles.*

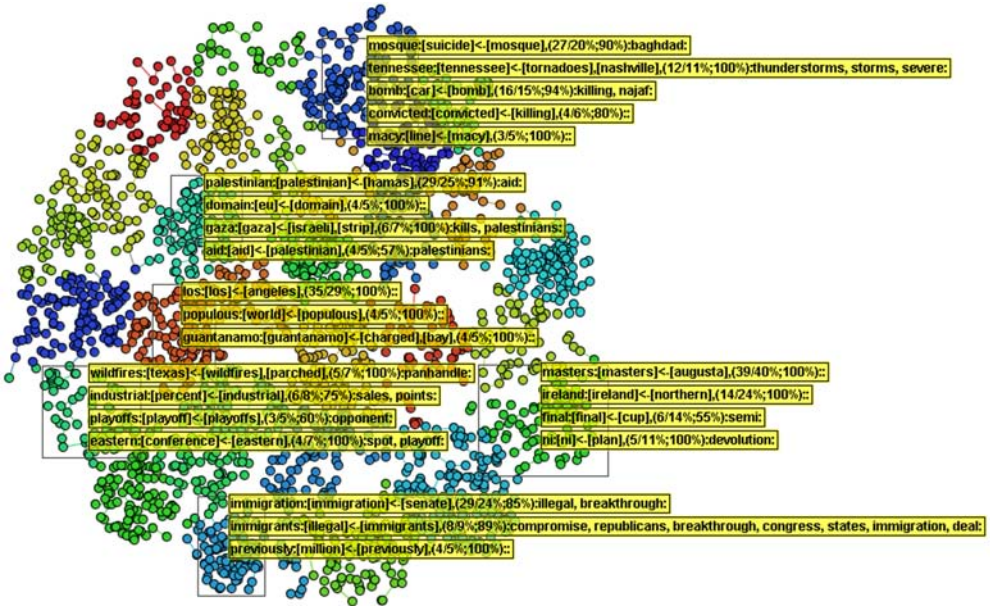


Figure 25: *Document map of 2684 RSS news feed articles, collected from BBC, Reuters, CNN, and Associated Press sites, during two days in April 2006. Projection by ProjClus. Individual colors denotes clusters of documents. Shown labels were generated from groups of rules inducted using some of the clusters of documents.*

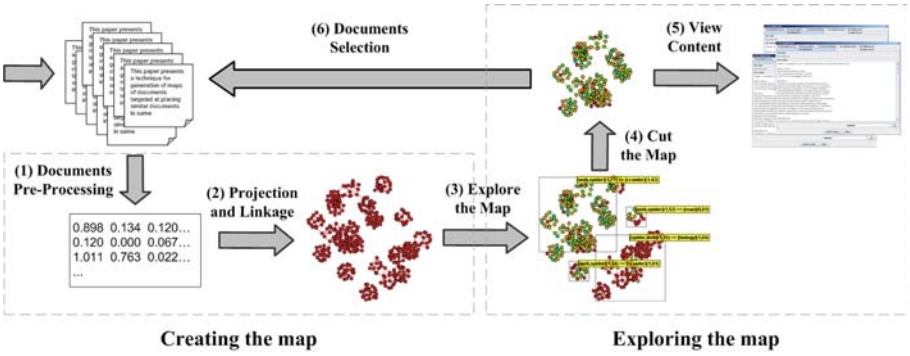


Figure 27: *The suggested process which can be used with PEx to help the user to identify useful reading material.*

The final result of the vector representation is a document x terms matrix coding documents' coordinates. From that matrix, a measure of distance between two documents is determined using a cosine-based metric applied on the vectors that represent them [FL95]. This kind of metric is preferred instead of other most common metrics, such as Euclidean, since this is less sensitive to the sparsity of the space, and the document x terms matrix normally is very sparse.

The multi-dimensional projection is then generated from such calculated distances.

Projected points can be connected to in various forms to help explore the map. In PEx, a triangulation or a connection by next neighbors of any order can be accomplished. Attributes can be mapped to color and size of the nodes, and node content can be loaded as illustrated above. Various graph manipulation tools are available and point placement after projection can also be performed based on any of the available edge determination alternatives.

The map can be updated by eliminating nodes and groups of nodes, moving them and coloring by certain algorithms.

For instance, the same metric that allowed the calculation of document distances can be used to color code each node on the resulting graph based on its distance to a particular document. Figure 28 presents the same map shown

in Figure 24. In this picture, the nodes with colors different from red are the most similar to the paper entitled "Semi-Automatic Image Annotation Using Frequent Keyword Mining".

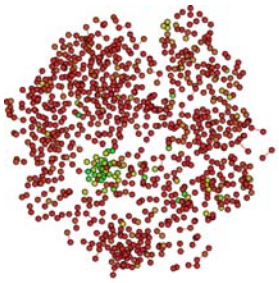


Figure 28: *Color coding by distance from a particular element of the map.*

The process presented here can be understood as a visual process based on content similarity amongst documents which aims at supporting the user to search, find and focus on sets of relevant documents in a large variety of applications and to find particularities of text collections.

As an alternative to the documents vector representation, an approximation of the Kolmogorov complexity has been used, called *Normal Compress Distance (NCD)* [CV05]. NCD was originally proposed as a metric between DNA sequences, but has been successfully employed to documents [TMP07]. As the main advantage, NCD does not need to pre-process the document collection in order to define the similarity between documents nor it needs information about the entire collection such as needed to calculate tfidf.

B.3. Further Remarks

The knowledge discovery process is, by its very nature, highly interactive and iterative. However, when humans are dealing with large sources of explorative information, it is very difficult even to drive the mining process. Visual text mining techniques in general and the novel ones presented here presented here are able to leverage the user interaction, speeding up the process that gives rise to useful and novel knowledge. Scanning, filtering, brushing, zooming, topic extraction and automatic labeling enable users to quickly get feedback and reformulate or adjust data preprocessing criteria in order to achieve a better display on which he/she can concentrate on.

As further steps to increase the power of the presented tools we intend to use terms drawn from the process of rule generation to build a representational space that at the same time present relevant concepts within a text data set and can offer various levels of generalization amongst texts. Various treatments of the distance measures and of the subtleties of the projections are also being studied further and formalized within a large project on visual analytics.

References

[ABY03] ALONSO O., BAEZA-YATES R.: Alternative implementation techniques for web text visualization. In *Proc. First Latin American Web Congress (LA-WEB 2003)* (Santiago, Chile, November 2003), IEEE Computer Society, IEEE Press, pp. 202–203.

[AF03] ASLAM J. A., FROST M.: An information-theoretic measure for document similarity. In *SIGIR '03: Proc. 26th annual international ACM SIGIR conference on Research and development in informaion retrieval* (New York, NY, USA, 2003), ACM Press, pp. 449–450.

[AKS*02a] ANDREWS K., KIENREICH W., SABOL V., BECKER J., DROSCHL G., KAPPE F., GRANITZER M., AUER P., TOCHTERMANN K.: The infosky visual explorer: exploiting hierarchical structure and document similarities. *Information Visualization Journal 1* (2002), pp. 166–181.

[AKS*02b] ANDREWS K., KIENREICH W., SABOL V., BECKER J., DROSCHL G., KAPPE F., GRANITZER M., AUER P., TOCHTERMANN K.: The infosky visual explorer: exploiting hierarchical structure and document similarities. *Information Visualization 1*, 3/4 (2002), pp. 166–181.

[AOL93] ALLEN R. B., OBRY P., LITTMAN M.: An interface for navigating clustered document sets returned by queries. In *Proc. ACM Conf. Organizational Computing Systems* (1993), pp. 166–171.

[BCB03] BORNER K., CHEN C., BOYACK K.: Visualizing knowledge domains. *Annual Review of Informtion Science & Technology 37* (2003), pp. 1–51.

[BCG*99] BOOKER A., CONDLIFF M., GREAVES M., HOLT F., A.KAO, PIERCE D., POTEET S., WU Y.-J.: Visualizing text data sets. *Computing in Science and Eng. 1*, 4 (1999), pp. 26–35.

[BEW95] BECKER R. A., EICK S. G., WILKS A. R.: Visualizing network data. *IEEE Tran. Visualization and Computer Graphics 1*, 1 (March 1995), pp. 16–28.

[BY96] BAEZA-YATES M. R.: Visualizing large answers in text databases. In *Proc. Int. Workshop on Advanced User Interfaces (AVI'96)* (1996), ACM Press, pp. 101–107.

[BYRN99] BAEZA-YATES R., RIBEIRO-NETO B.: *Modern information retrieval*. Addison-Wesley Harlow, England, 1999.

[CC92] CHALMERS M., CHITSON P.: Bead: explorations in information visualization. In *15th Int. ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR'92)* (1992), IEEE CS Press, pp. 330–337.

[CC00] COX T. F., COX M. A. A.: *Multidimensional Scaling*, second ed. Chapman & Hall/CRC, 2000.

[Cha93] CHALMERS M.: Using a landscape methaphor to represent a corpus of documents. In *Spatial Information Theory: A Theoretical Basis for GIS* (1993), Frank A. U., Campari I., (Eds.), vol. 716 of *Lecture Notes in Computer Science*, Springer, pp. 377–390.

[Cha96] CHALMERS M.: A linear iteration time layout algorithm for visualising high-dimensional data. In *Proc. 7th IEEE Visualization, (VIS)'96* (Los Alamitos, CA, USA, 1996), IEEE Computer Society Press, pp. 127–132.

[Che73] CHERNOFF H.:. *Journ. Amer. Stat. Assoc. 68*, 361 (1973).

[Che04] CHEN C.: *Information Visualization: Beyond the Horizon*. Springer, 2004.

[CHR03] CAREY M., HEESCH D. C., RUGER S. M.: A visualization tool for document searching and browsing. In *Proc. Intl Conf. Distri. Multimedia Sys.* (2003).

[CLMS02] CHEN X., LI M., MCKINNON B., SEKER A.: *A theory of uncheatable program plagiarism detection and its practical implementation*. Tech. rep., UCSB, 2002.

[CNT04] CHERFI H., NAPOLI A., TOUSSAINT Y.: Towards a Text Mining Methodology Using Frequent Itemsets and Association Rules. *Soft Computing Journal 11* (2004), pp. 431–441.

[CPMT07] CUADROS A. M., PAULOVICH F. V., MINGHIM R., TELLES G. P.: Point placement by phylogenetic trees and its application to visual analysis of document collections (conditionally accepted). IEEE CS Press.

[CV05] CILIBRASI R., VITÁNYI P.: Clustering by compression. *IEEE Trans. Information Theory 51*, 4 (2005), pp. 1546–1555.

[DFK*04] DRINEAS P., FRIEZE A., KANNAN R., VEMPALA S., VINAY V.: Clustering large graphs via the singular value decomposition. In *Machine Learning* (2004), vol. 56, pp. 9–33.

[Eic96] EICK S. G.: Aspects of network visualization. *IEEE Computer Graphics and Applications 16*, 2 (March 1996), pp. 69–72.

[FFW91] FOWLER R., FOWLER W., WILSON B.: Integrating query, thesaurus, and documents through a common visual representation. In *14th Int. ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR'91)* (1991), pp. 142–151.

[FL95] FALOUTSOS C., LIN K.: Fastmap: A fast algorithm for indexing, datamining and visualization of traditional and multimedia databases. In *ACM SIGMOD Intl. Conf. Management of Data* (San Jose-CA, USA, 1995), ACM Press: New York, pp. 163–174.

[FR91] FRUCHTERMAN T. M. J., REINGOLD E. M.:

Graph drawing by force-directed placement. *Software — Practice and Experience 21*, 11 (1991), pp. 1129–1164.

[GSAK04] GRANITZER M., SABOL W. K. V., ANDREWS K., KLIEBER W.: Evaluating a system for interactive exploration of large, hierarchically structured document repositories. In *Information Visualization 2004* (Austing- TX, USA, 2004), IEEE CS Press, pp. 127–132.

[H.91] H. L.: Color icons: Merging color and texture perception for integrated visualization of multiple parameters. In *Proc. IEEE Visualization'91* (October 1991).

[HCC*97] HAN J., CHIANG J. Y., CHEE S., CHEN J., CHEN Q., CHENG S., GONG W., KAMBER M., KOPERSKI K., LIU G., LU Y., STEFANOVIC N., WINSTONE L., XIA B. B., ZAIANE O. R., ZHANG S., ZHU H.: Dbminer: a system for data mining in relational databases and data warehouses. In *CASCON'97: Proc. 1997 Conf. Centre for Advanced Studies on Collaborative research* (1997), IBM Press, p. 8.

[Hei07] HEIJS A.: Panel paper: Requirements for coordinated multiple view visualization systems for industrial applications (to appear). In *CMV '07: Proc. 5th Int. Conf. Coordinated & Multiple Views in Exploratory Visualization* (2007), IEEE CS Press.

[HHWN02] HAVRE S., HETZLER E., WHITNEY P., NOWELL L.: Themeriver: Visualizing thematic changes in large document collections. *IEEE Trans. Visual. and Comp. Graphics 8*, 1 (Jan-Mar 2002), pp. 9–20.

[Hof99] HOFFMAN P. E.: *Table Visualizations: a formal model and its applications*. PhD thesis, University of Massachusetts Lowell, Computer Science Department, Lowell, MA, USA, 1999.

[HP96] HEARST M. A., PEDERSEN J. O.: Reexamining the cluster hypothesis: Scatter/gather on retrieval results. In *Proc. 19th Int. ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR'96)* (1996), pp. 76–84.

[HWR03] HUANG S., WARD M., RUNDENSTEINER E.: *Exploration of dimensionality reduction for text visualization*. Tech. Rep. Technical Report TR-03-14, Worcester Polytechnic Institute, 2003.

[IR01] IRITANO S., RUFFOLO M.: Managing the knowledge contained in electronic documents: a clustering method for text mining. In *12th DEXA Workshop* (2001), IEEE CS Press, pp. 454–458.

[JM04] JOURDAN F., MELANCON G.: Multiscale hybrid mds. In *IV '04: Proc. Eighth Int. Conf. Information Visualisation* (Washington, DC, USA, 2004), IEEE Computer Society, pp. 388–393.

[Jol86] JOLLIFFE I. T.: *Principal Component Analysis*. Springer-Verlag, 1986.

[KGM*01] K.ANDREWS, GÜTL C., MOSER J., SABOL V., LACKNER W.: Search result visualisation with xFIND. In *Proc. Int. Workshop on User Interfaces to Data Intensive Systems (UIDIS 2001)* (Zurich, Switzerland, May 2001), IEEE CS Press, pp. 50–58.

[KHLK98] KASKI S., HONKELA T., LAGUS K., KOHONEN T.: Websom - self-organizing maps of document collections. *Neurocomputing 1*, 1-3 (1998), pp. 110–117.

[Koh97] KOHONEN T.: Exploration of very large databases by self-organizing maps. In *Proc. IEEE Int. Conf. Neural Networks* (1997), pp. 1–6.

[LA00] LEUSKI A., ALLAN J.: Lighthouse: Showing the way to relevant information. In *InfoVIS* (2000), IEEE Computer Society Press, pp. 125–130.

[LBC*01] LI M., BADGER J. H., CHEN X., KWONG S., KEARNEY P., ZHANG H.: An information-based sequence distance and its application to whole mitochondrial genome phylogeny. *Bioinformatics 17*, 2 (2001), pp. 149–154.

[LC96] LEOUSKI A. V., CROFT W. B.: *An evaluation of techniques for clustering search results*. Tech. Rep. IR-76, Department of Computer Science, University of Massachusetts, Amherst, 1996.

[Lin91] LIN X.: A self-organizing semantic map for information retrieval. In *14th Int. ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR'91)* (1991), pp. 262–269.

[LMMP06] LOPES A., MINGHIM R., MELO V., PAULOVICH F.: Mapping texts through dimensionality reduction and visualization techniques for interactive exploration of document collections. In *Visualization and Data Analysis 2006 — Proc. SPIE-IS&T Electronic Imaging* (San Jose, California, 2006), Erbacher R. F., Roberts J. C., Gröhn M. T., Borner K., (Eds.), vol. 6060, SPIE, p. 60600T.

[LPMP07] LOPES A., PINHO R., MINGHIM R., PAULOVICH F.: Visual text mining using association rules. *Computers & Graphics Journal, Special Issue on Visual Analytics 31*, 3 (June 2007), pp. 316–326.

[LV97] LI M., VITÁNYI P.: *An Introduction to Kolmogorov Complexity and Its Applications*, 2nd ed. Springer Verlag, 1997.

[MB94] MACEDONIA M. R., BRUTZMAN D. P.: Mbone provides audio and video across the internet. *IEEE Computer 27*, 4 (April 1994), pp. 30–36.

[MC04] MORRISON A., CHALMERS M.: A pivot-based routine for improved parent-finding in hybrid mds. *Information Visualization 3*, 2 (2004), pp. 109–122.

[MDE87] MCCORMICK B. H., DEFANTI T. A., BROWN M. D. (EDS.): Visualization in scientific computing. *Computer Graphics 21*, 6 (1987).

[Mer97] MERKL D.: Exploration of text collections with hierarchical feature map. In *Proc. 20th Int. ACM SIGIR*

Conference on Research and Development in Information Retrieval (SIGIR'97) (1997), pp. 186–195.

[MGTS90] MIHALISIN T., GAWLINSKI E., TIMLIN J., SCHWEGLER J.: Visualizing scalar field on a dimensional lattice. In *Proc. IEEE Visualization '90, IEEE CS Press* (1990), pp. 255–262.

[MLN*05] MINGHIM R., LEVKOWITZ H., NONATO L. G., WATANABE L., SALVADOR V., LOPES H., PESCO S., TAVARES G.: Spider cursor: A simple verstile interaction tool for data visualization and exploration. In *Proc. GRAPHITE 2005 3rd Int. Conf., Comp. Graph. and Interac. Tech. in Australasia and SE Asia* (Dunedin, New Zeland, 2005), ACM Press, pp. 307–314.

[MPL06] MINGHIM R., PAULOVICH F., LOPES A.: Content-based text mapping using multi-dimensional projections for exploration of document collections. In *Visualization and Data Analysis 2006, Proc. SPIE-IS&T Electronic Imaging* (San Jose, California, 2006), Erbacher R. F., Roberts J. C., Gröhn M. T., Borner K., (Eds.), vol. 6060, SPIE, p. 60600S.

[MRC02] MORRISON A., ROSS G., CHALMERS M.: A hybrid layout algorithm for sub-quadratic multidimensional scaling. In *INFOVIS '02: Proc. IEEE Symposium on Information Visualization* (Washington, DC, USA, 2002), IEEE Computer Society, p. 152.

[MRC03] MORRISON A., ROSS G., CHALMERS M.: Fast multidimensional scaling through sampling, springs and interpolation. *Information Visualization 2*, 1 (2003), pp. 68–77.

[MTS91] MIHALISIN T., TIMLIN J., SCHWEGLER J.: Visualizing multivariate functions, data and distributions. *IEEE Computer Graphics & Applications 11*, 3 (1991), pp. 28–34.

[MWBF98] MILLER N. E., WONG P. C., BREWSTER M., FOOTE H.: Topic islands — a wavelet-based text visualization system. In *Proc. Visualization '98* (Research Triangle Park, North Carolina, United States, 1998), IEEE CS Press, pp. 189–196.

[OL03] OLIVEIRA M., LEVKOWITZ H.: From visual data exploration to visual data mining: a survey. *IEEE Trans. on Vis. Comp. Graph. 9*, 3 (2003), pp. 378–394.

[Paga] PAGE C. H.:. http://complearn.sourceforge.net/.

[Pagb] PAGE G. H.:. www.gzip.org.

[PG88] PICKETT R. M., GRINSTEIN G. G.: Iconographic displays for visualizing multidimensional data. In *Proc. IEEE Conf. Systems, Man and Cybernetics, Vol. I* (1988), pp. pages 514–519.

[PLOM07] PINHO R., LOPES A. A., OLIVEIRA M. C. F., MIGHIM R.: Topic extraction from document collections using visualization and locally weighted association rules. In *submitted to PKDD* (2007).

[PM06] PAULOVICH F. V., MINGHIM R.: Text map explorer: a tool to create and explore document maps, 2006.

[PNML06] PAULOVICH F. V., NONATO L. G., MINGHIM R., LEVKOVITZ H.: Visual mapping of text collections through a fast high precision projection technique. In *Proc. 10th Int. Conf. on Information Visualisation, London, UK* (Washington, DC, USA, 2006), IEEE Computer Society, pp. 282–290.

[PNML07] PAULOVICH F. V., NONATO L. G., MINGHIM R., LEVKOWITZ H.: Least square projection: a fast high precision multidimensional projection technique and its application to document mapping (provisionaly accepted). *IEEE Trans. Vis. Comput. Graph.* (2007).

[PRTV98] PAPADIMITRIOU C. H., RAGHAVAN P., TARNAKI H., VEMPALA S.: Latent semantic indexing: A probabilistic analysis. In *Proc. 17th ACM Symposium on the Principles of Database Systems* (1998), pp. 159–168.

[PSS81] PAUL W., SEIFERAS J., SIMON J.: An information-theoretic approach to time bounds for on-line computation. *J. Comput. Syst. Sci. 23*, 2 (1981), pp. 108–126.

[RC03] ROSS G., CHALMERS M.: A visual workspace for constructing hybrid multidimensional scaling algorithms and coordinating multiple views. *Information Visualization 2*, 4 (2003), pp. 247–257.

[RES98] ROHRER R. M., EBERT D. S., SIBERT J. L.: The shape of shakespeare: Visualizing text using implicit surfaces. In *IEEE InfoVis'98* (1998), IEEE Press, pp. 121–129.

[RES99] ROHRER R. M., EBERT D. S., SIBERT J. L.: A shape-based visual interface for text retrieval. *IEEE Computer Graphics and Applications* (September/October 1999).

[RK04] RASMUSSEN M., KARYPIS G.: *gCLUTO — An Interactive Clustering, Visualization, and Analysis System.* Tech. Rep. CSE/UMN TR 04-021, Univ. of Minnesota, Dep. of Computer Science and Engineering, 2004.

[SA98] SWAN R., ALLAN J.: Aspect windows, 3-D visualizations, and indirect comparisons of information retrieval system. In *Proc. 21st Int. ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR'98)* (1998).

[Sal91] SALTON G.: Developments in automatic text retrieval. *Science 253* (1991), pp. 974–980.

[Sam64] SAMMON J. W.: A nonlinear mapping for data structure analysis. In *IEEE Transactions on Computers* (May 1964), vol. C-18, pp. 401–409.

[SB88] SALTON G., BUCKLEY C.: Term-weighting approaches in automatic text retrieval. *Inf. Process. Manage. 24*, 5 (1988), pp. 513–523.

[SCL*99] SEBRECHTS M., CUGINI J., LASKOWSKI S., VASILAKIS J., MILLER M.: Visualization of search results: A comparative evaluation of text, 2d, and 3d interfaces. In *22nd ACM-SIGIR Conf. Research and Development in Information Retrieval* (1999), ACM Press, pp. 3–10.

[Sku02] SKUPIN A.: A cartographic approach to visualizing conference aborner2003bstracts. *IEEE Computer Graphics and Applications Volume 22, Issue 1* (2002), pp. 50–58.

[SN87] SAITOU N., NEI M.: The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol. 4*, 4 (1987), pp. 406–425.

[TC89] THOMPSON R. H., CROFT W. B.: Support for browsing in an intelligent text retrieval system. *Int. Journal of Man-Machine Studies 30*, 6 (1989), pp. 639–668.

[TMN03] TEJADA E., MINGHIM R., NONATO L.: On improved projection techniques to support visual exploration of multidimensional data sets. *Information Visualization Journal 2*, 4 (2003), pp. 218–231.

[TMP07] TELLES G., MINGHIM R., PAULOVICH F.: Normalized compression distances for visual analysis of document collections. *Computer & Graphics, Special Issue on Visual Analytics 31*, 3 (June 2007), pp. 327–337.

[vWvL93] VAN WIJK J. J., VAN LIERE R.: Hyperslice: visualization of scalar functions of many variables. In *VIS '93: Proc. 4th Conf. on Visualization '93* (1993), pp. 119–125.

[Wei01] WEIPPL E.: Visualizing content based relations in texts. In *Proc. 2nd Australian Conf. User interface* (Queensland, Australia, 2001), IEEE Computer Society, IEEE CS Press, pp. 34–41.

[Wel84] WELCH T.: A technique for high-performance data compression. *IEEE Computer 17*, 6 (1984), pp. 8–19.

[Wis99] WISE J. A.: The ecological approach to text visualization. *J. of the American Soc. for Inf. Sci. 50*, 13 (November 1999), pp. 1224–1233.

[WIZD04] WEISS S., INDURKHYA N., ZHANG T., DAMERAU F.: *Text Mining: Predictive Methods for Analyzing Unstructured Information*. Springer, 2004.

[WTP*95] WISE J. A., THOMAS J. J., PENNOCK K., LANTRIP D., POTTIER M., SCHUR A., CROW V.: Visualizing the non-visual: spatial analysis and interaction with information for text documents. In *Readings in information visualization: using vision to think* (San Francisco, CA, USA, 1995), Morgan Kaufmann Publishers Inc., pp. 442–450.

[Zam98] ZAMIR O.: *Visualization of search results in document retrieval systems*. General examination report, University of Washington, 1998.

[ZE98] ZAMIR O., ETZIONI O.: Web document clustering: a feasibility demonstration. In *Proc. 21st Int. ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR'98)* (1998), pp. 46–54.

[ZE00] ZAMIR O., ETZIONI O.: *Grouper: A Dynamic Clustering Interface to Web Search Results*. Tech. rep., University of Washington, 2000.

[ZK05] ZHAO Y., KARYPIS G.: Topic-driven clustering for document datasets. In *SIAM 2005 Data Mining Conf.* (2005).

Visual Mining of Text Collections

Rosane Minghim
ICMC – University of São Paulo
São Carlos
Brazil

Haim Levkowitz
IVPR – University of Massachusetts Lowell
USA

Visual Mining of Text Collections

- Overview, Motivation, Goals
- Test Cases
- Basic Concepts
- From Visualization to Visual Text Mining
- Visual Text Mining Using Projectinons and Point Placement Estrategies
- Conclusions

Overview, Motivation, Goals

- Collections of Text
  - Metadata
  - Content
- Multi-disciplinar
- Visual Data Mining
- Visual Text Mining
- Visual Analytics

- Applications
  - Information gathering
  - Searching
  - Detection
  - Discovering the Unexpected
  - Relating to predetermined documents

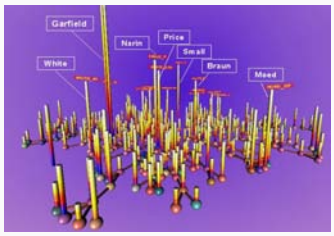
- Applications
  - Survey
  - Study and Education
  - Patent Search
  - News finding
  - Forensics
  - ...

- Maps of text Collections
  - Based on Relationships (Borner & Chen)
    - Co-authorship, co-citation
  - Based on Content
    - Similarity and Grouping
    - Common underlying subject
    - ➔Topics

Relationships : Topic Busts and co-word

(Mane and Borner) 2004

Relationships : Citation and Co-citation

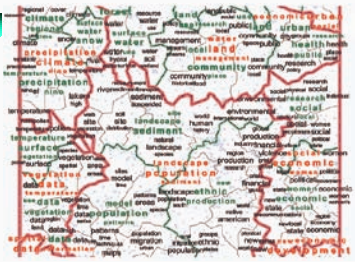


(Borner) (2003)

Content-based Text Mapping

- Approach 1: Dimension reduction
  ex. MSD, SVD, PCA
- Approach 2: Point Placement (PP)
- Approach 3: Clustering
- Approach 4: Projections
  ex. FASPMAP, NNP, LSP

Content - based

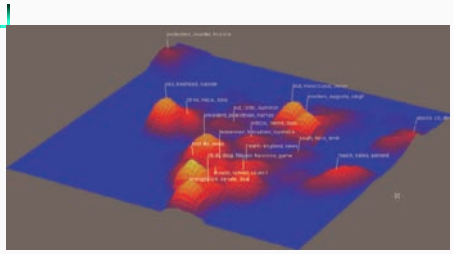


(Skupin)
(2002)
(abstracts)
SOM

Content - based

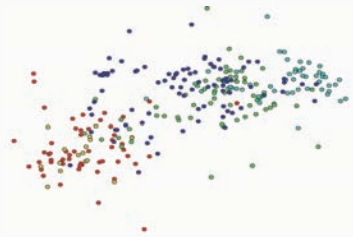
(Dimensional
Reduction)
News flash
IN-SPIRE
(PNL)

Content - based



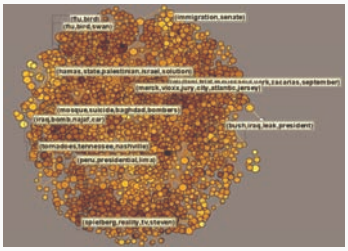
(Surface
View)
IN-SPIRE

Content – based by Projections



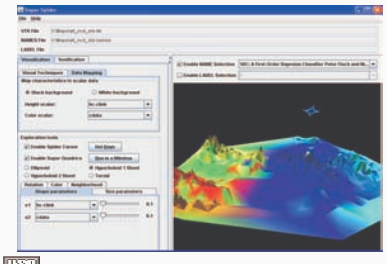
(Projections: LSP)

Content – based by Projections



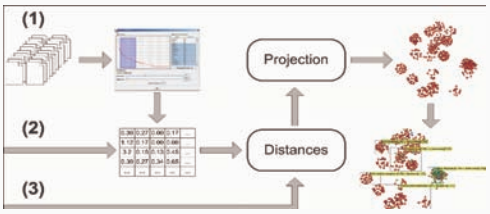
(PEx Projection Explorer (Graph View)

Content – based by Projections



(Super Spider) (Surface View)

Content – based by Projections



2 – Test Cases

See main Text

3 - Basic Concepts

3.1 Text Preprocessing and IR

3.2 Data and text mining

3.3 Projection techniques

3.4 Visual representations: graphs, surfaces, volumes, triangulations

3.1 Text Preprocessing

1. Stopwords elimination
2. Extraction of words radicals (stemming)
3. Creation of n-grams
4. Frequency count and Luhn's lower cut (n-grams appearing less then x times are ignored)
5. Weighting process (*term-frequency inverse document-frequency - (tfidf)*)

Result is a Vector Model

- Attributes: terms (n-grams)
- Value: term weight
- Table Data

Vector Representation – term weighting

- tf – term frequency
- tfidf – tf x idf = tf x inverse document frequency

$$w_{ik} = tf_{ik} \times log\left(\frac{N}{n_k}\right)$$

Vector Representation

	$term_1$	$term_2$	$term_3$	$term_4$	...	$term_m$
Doc_1	0.92	0.62	0.92	0.10	...	0.67
Doc_2	0.13	0.11	1.00	0.34	...	0.33
Doc_3	0.52	0.00	0.00	0.44	...	0.77
...	...	...	...	...	...	...
Doc_n	0.02	0.12	0.22	0.92	...	0.00

Vector Representation – Similarity calculatin

EUCLIDEAN

$$sim_{i,j} = \sqrt{(w_{i,1} - w_{j,1})^2 + \ldots + (w_{i,k} - w_{j,k})^2}$$

MANHATAN

$$sim_{i,j} = |w_{i,1} - w_{j,1}| + \ldots + |w_{i,1} - w_{j,1}|$$

COSINE

$$sim_{i,j} = \frac{(w_{i,1} \times w_{j,1}) + \ldots + (w_{i,k} \times w_{j,k})}{(w_{i,1}^2 + \ldots + w_{i,k}^2) \times (w_{j,1}^2 + \ldots + w_{j,k}^2)}$$

Vector Representation – distance calculation

$$dis(doc_i, doc_j) = \sqrt{2 * (1 - sim(doc_i, doc_j))}$$

$$sim(doc_i, doc_j) = \frac{doc_i \times doc_j}{\|doc_i\| * \|doc_j\|}$$

Visualization (e.g.)

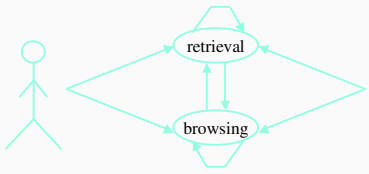
- Attribute Reduction
  - Co-clustering
  - PCA
  - SVD

  ↓ followed by
- Projection by Dimension Reduction

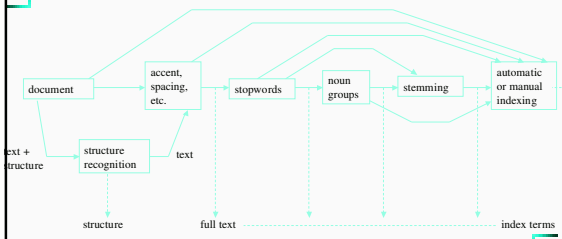
Alternatively

- Similarity Calculation
  - Ex: NCD Normalized Compression Distance
    - Approximation of Kolmogorov Complexity
- Point Placement (e.g. Force-directed)

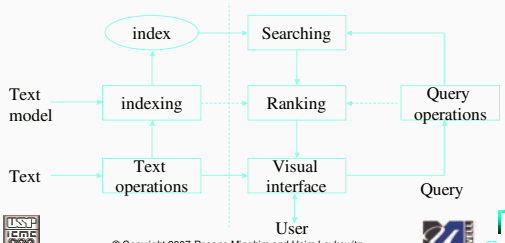
The retrieval process (high level)



Logical view of doc



The retrieval process (details)



==> 2 view of IR problem

- Computer-centered …
- Human-centered …

Computer-centered

- IR problem
  - build efficient indexes
  - process user queries with high performance
  - develop ranking algorithms
  - ==> improve `quality' of answer set

Human-centered view

- IR problem
  - study behavior of user
  - understand main needs
  - determine how such understanding affects organization and operation of retrieval system
- ==> keyword based query processing
  - might be seen as unlikely to yield good solution in long run

first generation

- Basically
  - automation of previous technologies
    - E.g., card catalogs)
  - allowed searches based on
    - author name
    - title

Second generation

- Increased search functionality added
- ==> search by
  - subject headings
  - keywords
  - some more complex query facilities

Third generation

- Currently being deployed
- Focus on
  - improved graphical interfaces
  - electronic forms
  - hypertext features
  - open system architectures

Three dramatic and fundamental changes

- First
  - access to various info sources a lot cheaper
  - ==> reaching wider audience than ever possible before

Second

- Advances in digital communication
  - ==> greater access to networks
  - ==> access
    - Distantly
    - Quickly
      - few seconds

Third

- Freedom to post whatever
- ==> popularity of Web
- For the first time in history
  - Most people have free access to large publishing medium

- Web (and modern digital libraries) as highly interactive medium
- ==> exchange messages, photos, documents, software, videos
- `Chat' :convenient + low cost
- Can do it at time of preference
  - ==> more convenience
- ==> high interactivity
  - fundamental and current shift in communication paradigm.
    - (Searching the Web + digital libraries)

Future: 3 main questions to be addressed

- First,
  - despite high interactivity, people still find it difficult (impossible?) to retrieve information relevant to their information needs
  - ==> which techniques ==> retrieval of higher quality?

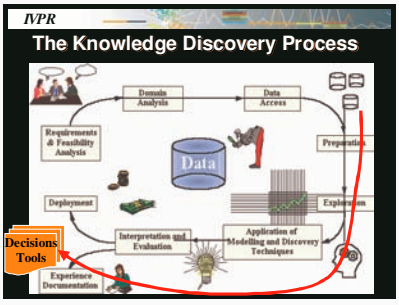
Second

- Ever increasing demand for access
- ==> quick response more and more a pressing factor
- ==> which techniques ==>
  - faster indexes
  - smaller query response times?

Third

- Quality of retrieval task greatly affected by user interaction with system
- ==> how better understanding of user behavior affect design and deployment of new information retrieval strategies?

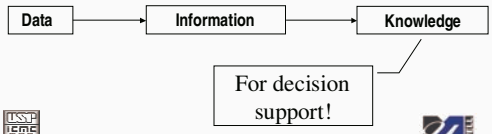
3.2 Data and Text Mining



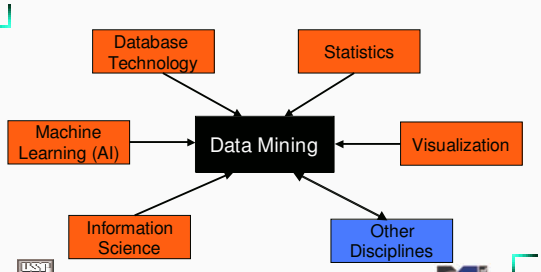
- Data Exploration is the process of searching and analyzing databases to discover implicit but potentially useful information

Goals of Data Exploration

- Convey information
- Discover new knowledge
- Identify structure, patterns, anomalies, trends, relationships



A Confluence of Multiple Disciplines



Data Mining Tasks & Techniques

Major Data Mining Tasks
- Summarization
- Association
- Classification
- Prediction
- Clustering
- Time-Series Analysis

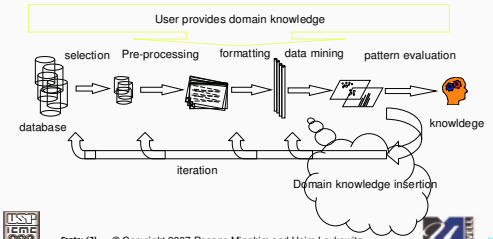
using

Statistical Tools
- **Missing Value Imputation**
- **Normalizations**
- **Error & Variational Analysis**
- **Confidence Estimates**

based on

Major Techniques
- Linear Regression Trees
- Non-Linear Regression
- MARS
- Naïve Bayes
- K-Means and K-Median
- Neural Networks
- Association Rules
- Decision Trees
- Principal Curve Analysis
- Support Vector Machines
- Genetic Algorithms

User-centered discovery process



Fonte: (2)

Definition [Ankerst 2000]

- VDM – visual data mining:
  - step in process of knowledge discovery / extraction (KDD)
  - utilizes visualization as communication channel between computer and user
  - to support identification of new and interpretable patterns

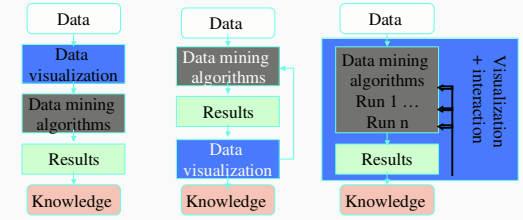
- Last two phases of process:
  - Data mining
  - Evaluation
- Three VDM categories

Categories [Ankerst 2000]

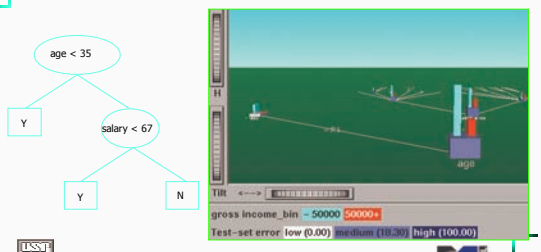
- Data visualization
  - User in full control of search in search space
  - Limit / focus search space
- Visualize mining results
  - Support interpretation of extracted models

- Visualize mining intermediate results
  - Guide search
  - Provide domain knowledge
    - E.g., adapt generic kernel (for different application) with user's intervention

Categories [Ankerst 2000]

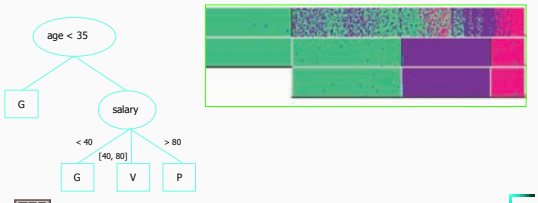


Classification tree visualization [MineSet] -- pre



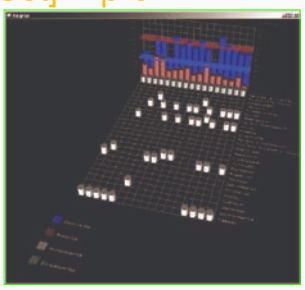
Fonte: (2)

Classification tree visualization [MineSet] -- integrated



Fonte: (2)

Classification tree visualization [MineSet] -- pre

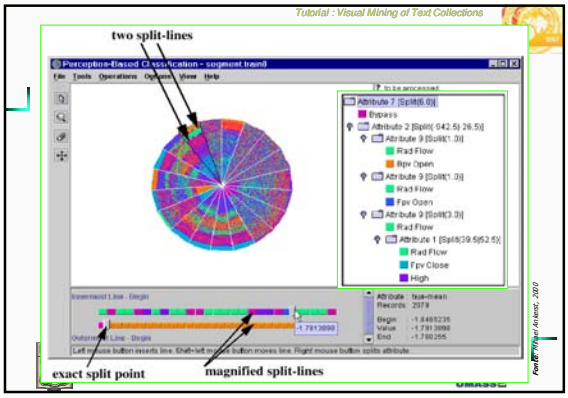


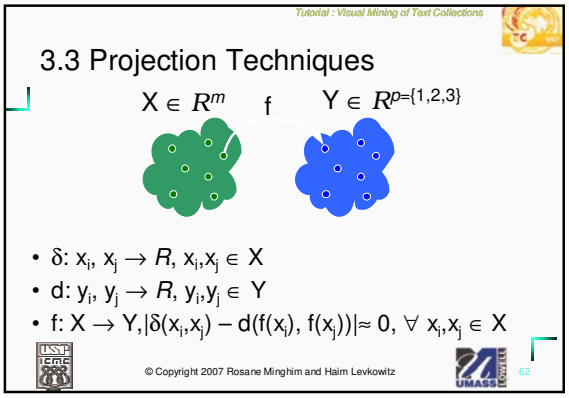
Categories [Wong 1999]

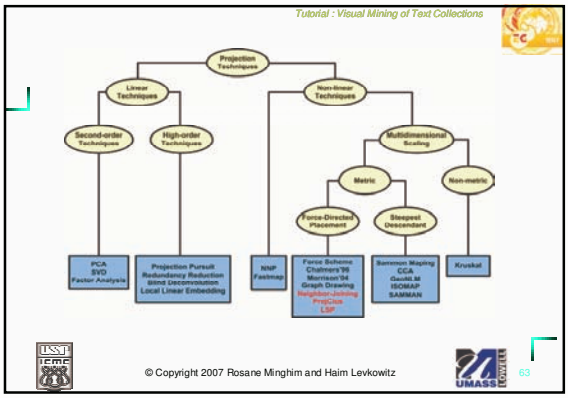
- Loosely coupled
  - Visualization detached from analytic mining strategies
  - Support pro-processing, interpret results ...
  - Limited approach: limits of both...

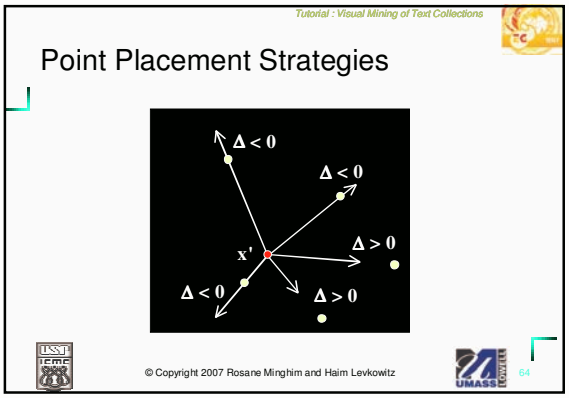
- Tightly coupled
  - Visualization integrated with mining analytic strategy
  - More user control and understanding of analytic process
    - Support decision making
  - Create visual representations of search space

- Mining with visualization
  - E.g., visual environment to generate classification trees (*Ankerst et al. 1999, 2000*)
- Visualization with mining
  - E.g., Hierarchical Parallel Coordinates (*Fua et al. 1999*)
- Examples with tight coupling ...



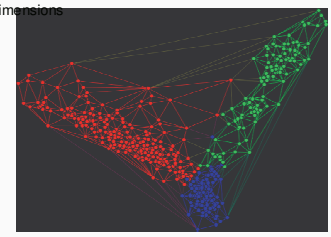




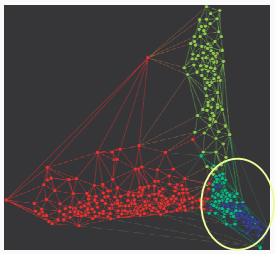


Problems PCA

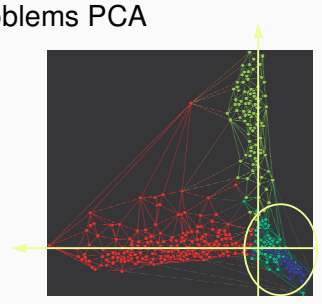
390 dimensions



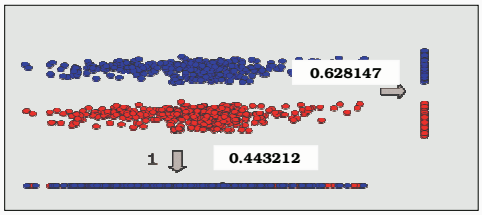
Problems PCA



Problems PCA

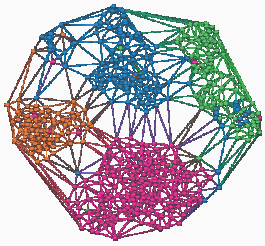


Problems with Stress measurements

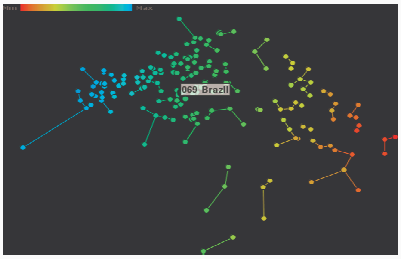


$$Stress = \sqrt{\frac{\sum_{i<j}(d(f(x_i),f(x_j)) - \delta(x_i,x_j))^2}{\sum_{i<j} d(f(x_i),f(x_j))^2}}$$

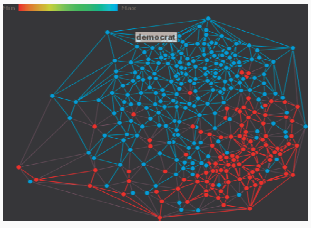
Example: Text Collections



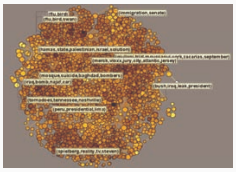
Example: ONU HDI



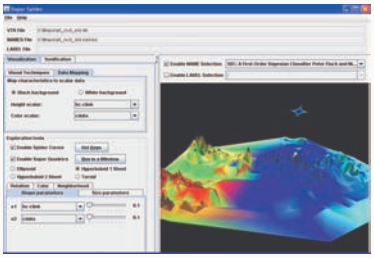
Example: Voting US Parlament



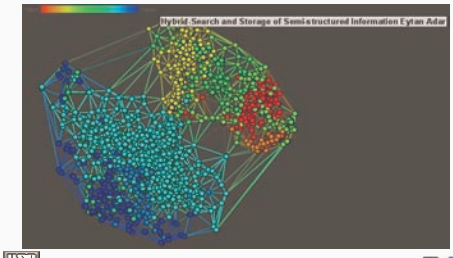
3.4 Visual representations: graphs, surfaces, volumes, triangulations



3.4 Visual representations: graphs, surfaces, volumes, triangulations



3.4 Visual representations: graphs, surfaces, volumes, triangulations



4. From Visualization To Visual Text Mining

4.1 Visualization Techniques for Multidimensional Data

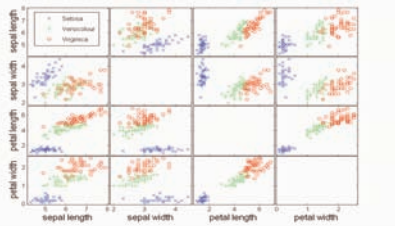
4.2 Visualization techniques and systems for handling document collections

4.3 Visual Text Mining

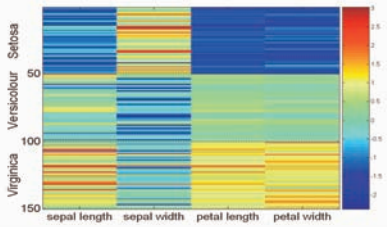
4. From Visualization To Visual Text Mining

4.1 Visualization Techniques for Multidimensional Data

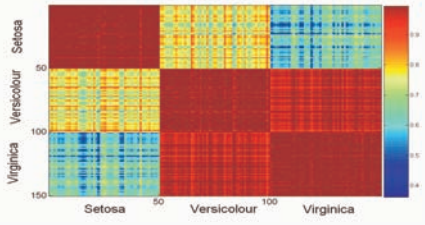
Scatter Plot Matrix



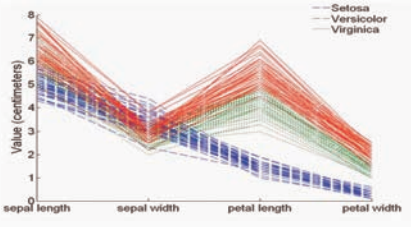
Data Matrix



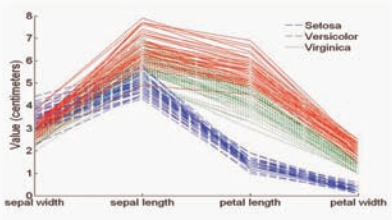
Correlation Matrix



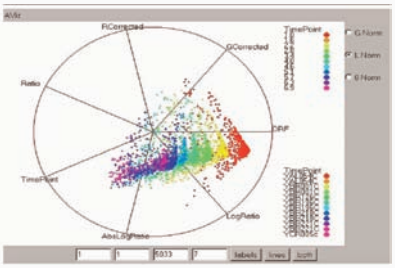
Parallel Coordinates



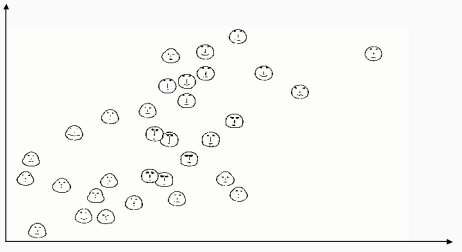
Parallel Coordinates



RadViz

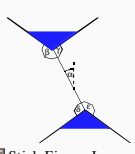


Chernoff Faces



Stick Figures

- properties of a line icon
- two attributes of the data are mapped to the display axes
- the remaining attributes are mapped to the angle/length of the limbs
- texture patterns in the visualization show certain data characteristics

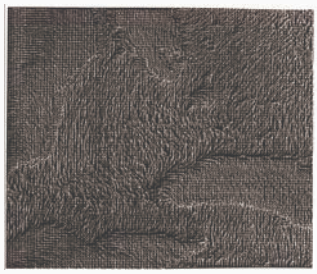


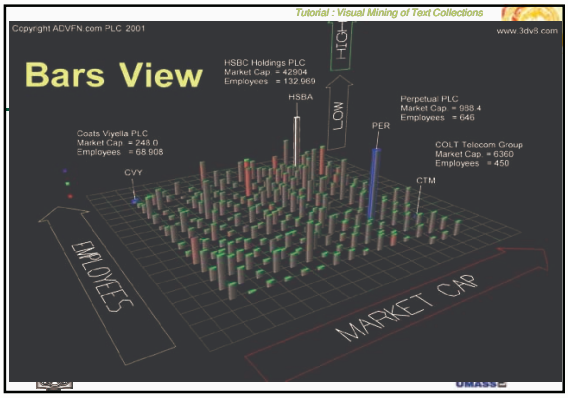
Stick Figure Icon

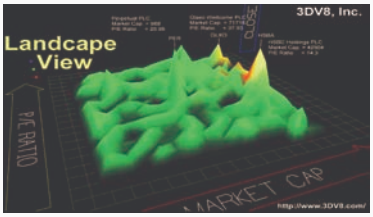
A Family of Stick Figures

Stick Figure Icon

5-dimensional NOAA image data from the great lake region





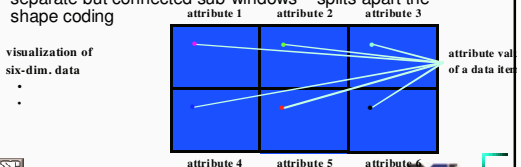


Pixel-oriented Techniques

- use each pixel to visualize one data value
  - about 1.3 million data values can thus be displayed at one time
- Need to map each data value to a color
- Thus need a color map and an interpretation for the map
- Also called dense pixel displays

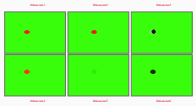
Dense Pixel Displays

- each attribute value is represented by one colored pixel (the value ranges of the attributes are mapped to a fixed color map)
- the attribute values for each attribute are presented in separate but connected sub-windows – splits apart the shape coding



Three Questions

How should the *pixels be arranged* within the subwindows?



Are alternative *shapes* of the sub-windows possible?

What is an appropriate ordering of the dimensions?

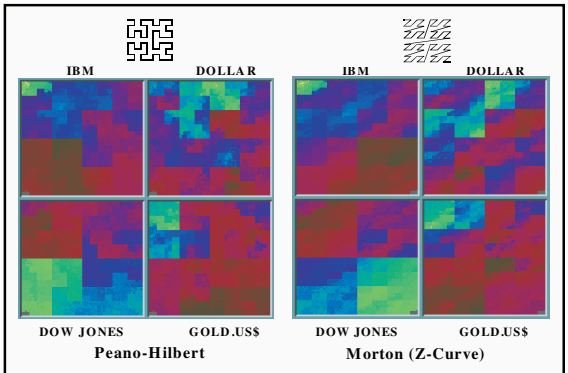
Arrangement of Pixels

Given: Ordered Set of *n* data items $\{a_1, ..., a_n\}$
consisting of *k* data values each $\left(a_i^1, \cdots, a_i^k\right)$

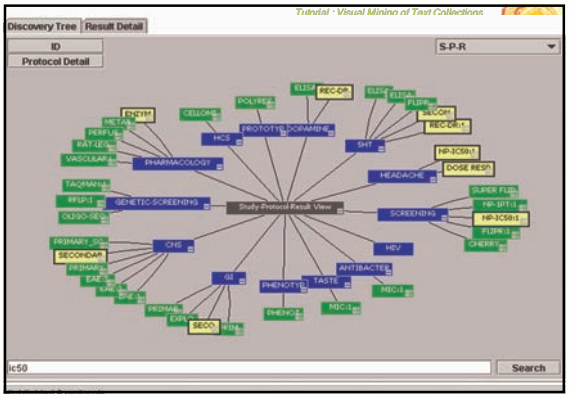
Goal: Two-dimensional arrangement of the data values, i.e. bijective mapping $f: \{1...n\} \rightarrow \{1...b\}'\{1...h\}$ $(n \leq b * h)$, such that the function

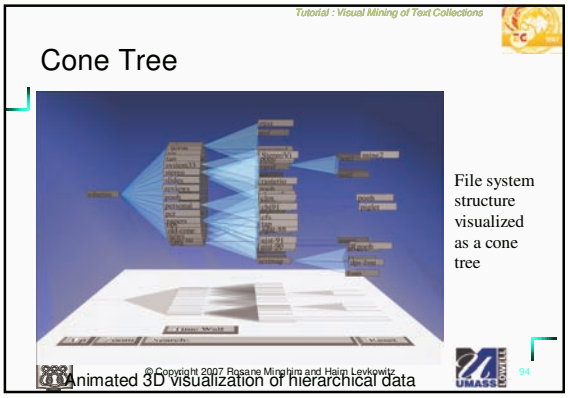
$$\sum_{i=1}^{n} \sum_{j=1}^{n} \left| d\left(f(i), f(j)\right) - d\left((0,0), \left(b \cdot \sqrt{\frac{|i-j|}{n}}, h \cdot \sqrt{\frac{|i-j|}{n}}\right)\right)\right|$$

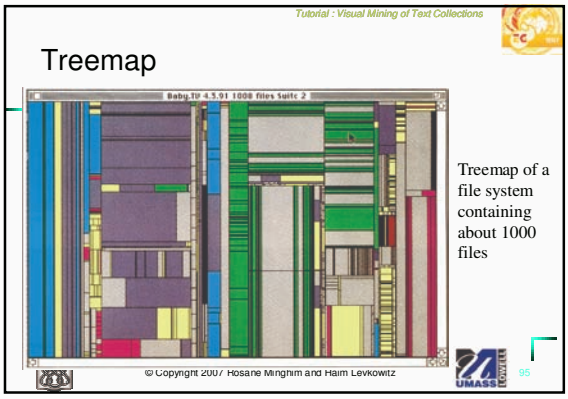
is minimal, where $d(f(i), f(j))$ is the L^p-distance ($p = 1, 2$) of the pixels belonging to a_i and a_j

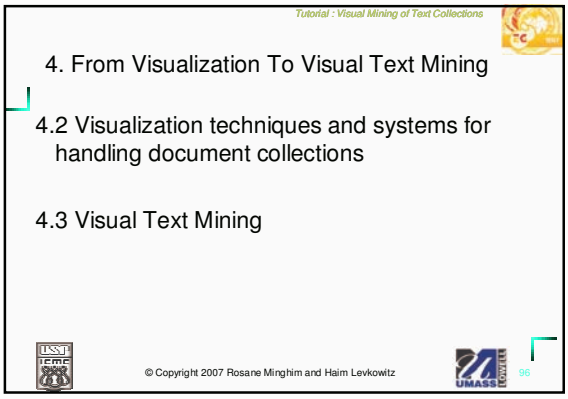


Space-Filling Curves







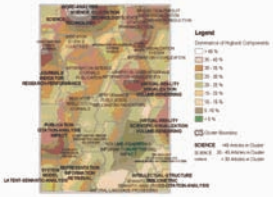


IN-SPIRE

- Spatial Paradigm for Information Retrieval - Pacific Nortwest National Laboratories
- Two Visualization Metaphors:
  - Galaxies – dimensional reduction
  - Themescape

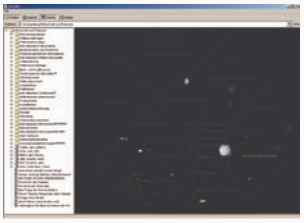
SOM based

- **Self-Organization Maps (SOMs) cartográficos (ex. Skurpin 2002)**



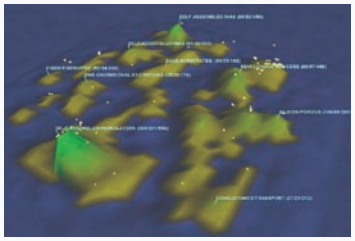
InfoSky

Granitzer (Granitzer et al., 2004) also employs galaxy metaphor



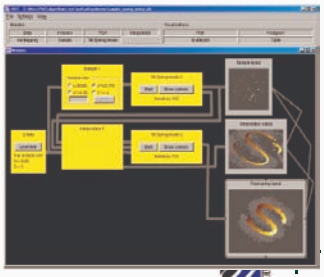
VxInsight

- Sandia National Laboratories, mountain metaphor (Boyack et al., 2002).



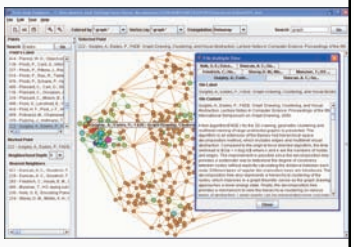
HIVE (Ross and Chalmers 2003)

- Interconnected components:
  - Import
  - Transform
  - Render multi-dim data



Projection Explorer (PEx)

- Projection and Point placement
- Precision
- Graphs and surfaces (Super Spider)



5. Projection Based Visualization and its application to Visual Text Mining

5.1 Projection Techniques and Point Placement Strategies

5.2 Mapping Text Collections via Projections and Point Placement

5.3 Topic Extraction and Visualization

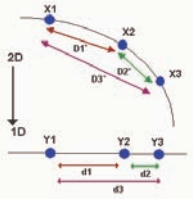
5.4 Further Examples

5. Projection Based Visualization and its application to Visual Text Mining

5.1 Projection Techniques and Point Placement Strategies

Ex: Sammon Mapping

- Let **X** be the points in the original space R^n, we apply a distance measure d_{ij}^* between Xi an Xj., and find **Y, the projected point**, ex. R^2 and d_{ij} the Euclidean distance between them.
- Sammon's method applies an error function to measure the target.



Least-Square Projection (LSP)

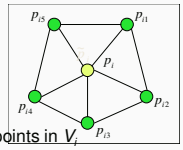
- $S = \{p_1, p_2, \ldots, p_n\} \in R^m$
- Preserving as much as possible the neighborhood relationship amongst points
- The neighborhood is defined by choosing the k-nearest neighbors of a point
- Least square Meshes

Least-Square Projection (LSP)

- Least square Meshes
- Three main steps
  - Select a subset of *S* (*control points)* and project it onto R^d
  - Determine the neighborhood of the points
  - Build a linear system whose solutions are the Cartesian coordinates of the points p_i in R^d

Building the Linear System: Laplacian Matrix

- Let $V_i = \{p_{i1},\ldots,p_{iki}\}$ be a set of k_i points in a neighborhood of a point p_i and be the coordinates of p_i in R^d



- each p_i is the centroid of the points in V_i

Building the Linear System: Laplacian Matrix

$$L\mathbf{x}_1=0,\ L\mathbf{x}_2=0,\ \ldots,\ L\mathbf{x}_d=0$$

where $x_1, x_2, \ldots, x_d$ are the vectors containing the Cartesian coordinates of the points and L is the matrix given by

$$Lij = \begin{cases} 1 & i = j \\ -\dfrac{1}{ki} & p_j \in V_i \\ 0 & otherwise \end{cases}$$

$$\mathbf{L} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ 0 \end{pmatrix}$$

Building the Linear System: Adding Control Points

$$Cij = \begin{cases} 1 & p_j \text{ is a control point} \\ 0 & \text{otherwise} \end{cases}$$

$$b_i = \begin{cases} 0 & i \le n \\ x_{p_{c_i}} & n < i \le n + nc \end{cases}$$

$$\mathbf{L} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ c_1 \\ c_2 \end{pmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

Solving the System

- The system is solved in a least-square sense
- The unique analytical solution is

$$\|Ax-b\|^2$$

$$A^T A\,\mathbf{x} = A^T\mathbf{b} \quad \Rightarrow \quad \mathbf{x} = (A^T A)^{-1} A^T\mathbf{b}$$

- is symmetric and sparse and can be solve using *Cholesky* factorization

Choosing the Control Points

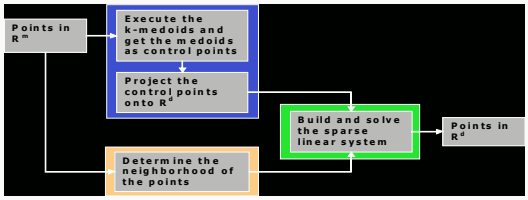
- In order to select the control points
  - the space R^m is split into *nc* clusters using k-medoids.
  - the control points are the medoids of each cluster

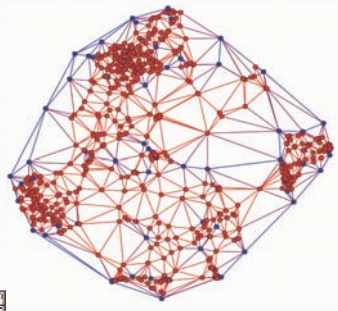
Choosing the Control Points

- Once the control points are chosen, these points are projected onto R^d through a fast dimensionality reduction method
  - Fastmap or NNP [2]
  - Further improvement of the projection using FORCE [2]

[2] E. Tejada, R. Minghim, and L. Nonato. On improved projection techniques to support visual exploration of multidimensional data sets. *Information Visualization Journal*, 2(4):218–231, 2003.

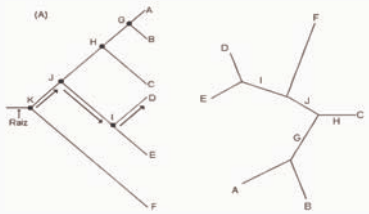
LSP: Overview





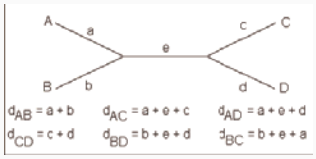
Control points in blue

Point Placement by Phylogenetic Tree Construction Algorithms (N-J Trees)



Point Placement by Phylogenetic Tree Construction Algorithms (N-J Trees)

$$d_{AB} + d_{CD} \leq \max\left(d_{AC} + d_{BD}, d_{AD} + d_{BC}\right)$$



Algorithm Neighbor-joining

Input: distance matrix

1. Criate a star tree for n objects.
2. Iteration
   1. Select a node pair (i,j) with smaller Sij (branch size)

$$S_{ij} = \frac{1}{2(n-2)} \sum_{k=3}^{N} (D_{ik} + D_{jk}) + \frac{1}{2} D_{ij} + \frac{1}{n-2} \sum_{\forall x \neq i,x \neq j} D_{ij}$$

   2. Combine nodes i and j in a new node and calculate the branch size of the new node.

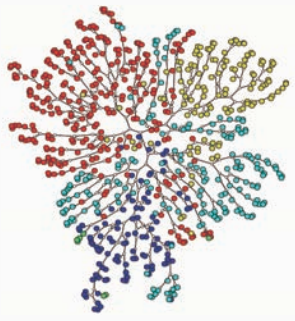
$$L_{ix} = \frac{D_{ij} + D_{iz} - D_{jz}}{2} \qquad L_{jx} = \frac{D_{ij} + D_{jz} - D_{iz}}{2}$$

Algorithm Neighbor-joining

3. Calculate new distance matrix, computing the new distances from the new node to the remaining nodes.

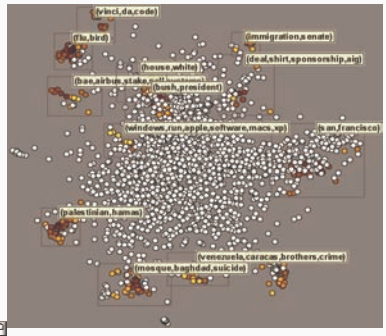
$$D_{(i-j),k} = \frac{(D_{ik} + D_{jk})}{2} \qquad (3 \leq k \leq N)$$

4. Eliminate previous nodes i and j
5. If n>2 then iterate again.



5.2 Mapping Text Collections via Projections and Point Placement

- Positioning and labeling

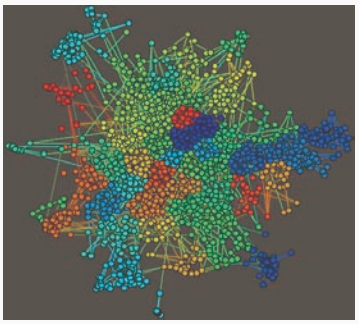


5.2 Mapping Text Collections via Projections and Point Placement

- Detailing topics

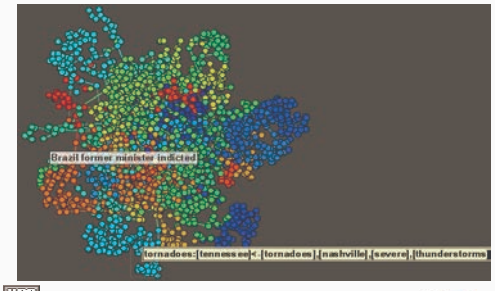
5.2 Mapping Text Collections via Projections and Point Placement

- Finding Relationships



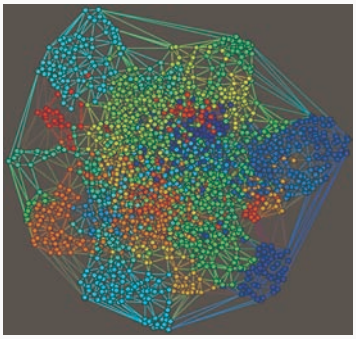
5.2 Mapping Text Collections via Projections and Point Placement

- Untangling



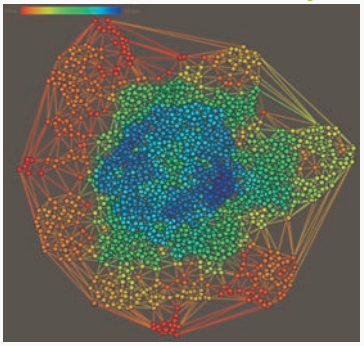
5.2 Mapping Text Collections via Projections and Point Placement

- Building a Surface - meshing



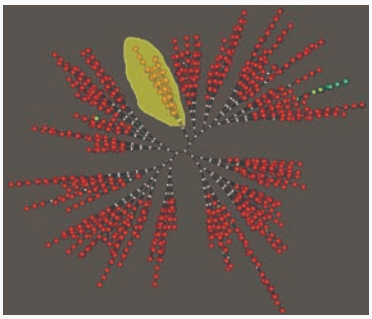
5.2 Mapping Text Collections via Projections and Point Placement

- Coloring by degree of proximity



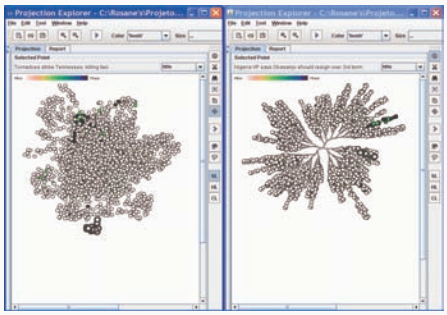
5.2 Mapping Text Collections via Projections and Point Placement

- Alternate view (N-J Tree)



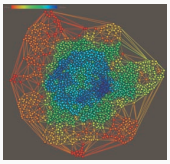
5.2 Mapping Text Collections via Projections and Point Placement

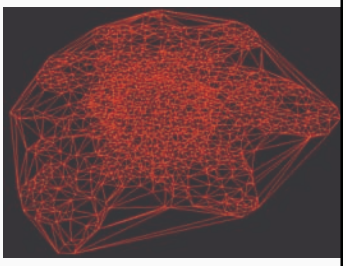
- Coordinating

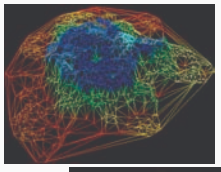


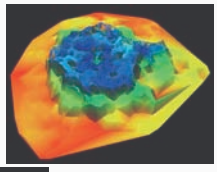
5.2 Mapping Text Collections via Projections and Point Placement

- Building a Surface









5.3 Topic Extraction and Visualization

- Topic Definition by Covariance
- Topic Extraction by Seeded Generation of Association Rules (pruning by relevant terms)
- Labeling and Viewing

Topic Extraction and Visualization

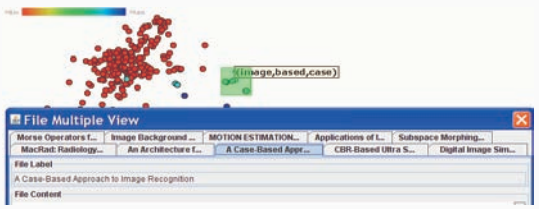
Topic Definition by Covariance

- Pair of words with highest covariance

$$cov(t_i, t_j) = \frac{1}{n-1} \sum_{k=1}^{n} (t_{ki} - \bar{t}_i)(t_{kj} - \bar{t}_j)$$

- For all the other words, highest mean covariance compared to first two.
- Add to label if above threshold.

Topic Definition by Covariance



Topic Extraction and Visualization

Topic Extraction using Association Rules

- Use relevant words as seeds
- Prune the case by rule weighting

Topic Extraction using Association Rules

Transactions	Items
1	Trousers, t-shirt, snickers
2	T-shirt, snickers
3	shorts, snickers
4	Trousers, sandals

Frequent Itemsets	Support
{snickers}	75%
{Trousers}	50%
{T-shirt}	50%
{T-shirt, snickers}	50%

Min. support = 50% (2 transactions).
Min. confidence = 50%.

Topic Extraction using Association Rules

tenis ⟶ t-shirt

$$support = support(\{snickers, t\text{-}shirt\}) = 50\%$$

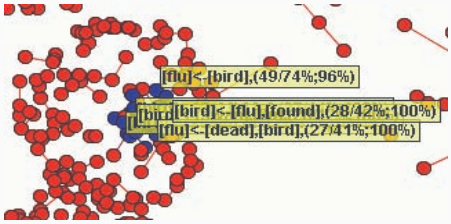
$$confidence = \frac{support(\{snickers, t-shirt\})}{support(\{t-shirt\})} = \frac{50}{50} = 100\%$$

T-shirt ⟶ snickers

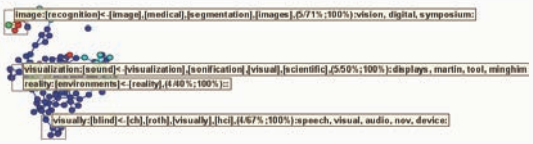
$$support = support(\{snickers, t-shirt\}) = 50\%$$

$$confidence = \frac{support(\{snickers, t-shirt\})}{support(\{snicker s\})} = \frac{50}{75} = 66{,}6\%$$

Topic Extraction using Association Rules (example)



Topic Extraction using Association Rules (example)



Topic Extraction using Association Rules

- Topics using AR
  - Term co-occurance in documents <=> subject
  - Transaction => Document
  - Item => term

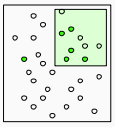
Topic Extraction using Association Rules

- Issues
  - Discovered rules amount
  - Term relevance (items)
  - Rule relevance measure (filtering)

  - High Sup. & conf. => few interesting rules
  - Low Sup. & conf. => huge amount of rules

Locally weighted and seeded AR

- Weighting Terms and Rules



$$w_{i,j} = \frac{\sum_{j=1}^{k} Tf_{i,j}}{\sum_{j=1}^{k} Tf_i}$$

$wi,s = 5/6 = \mathbf{0.83333}$

Steps

1. S: set of user selected documents
2. Picked 10 most relevant terms

$$W_{t_j S_k} = \frac{\sum T f_{t_j S_k}}{\sum T f_{t_j C}}$$

Steps

1. Initial item sets: Tr x T
   - Relevant Terms x All Terms
2. Items Sets discovered by Apriori altorighm
3. Sorted by weight:

$$\sum W_{t_j S_k}$$

Steps

6. Highest weight item set selected
7. Covered documents removed from S
8. Further item sets are selected if there is support over residual S (repeats 6 e 7)

9. If all items sets are considered and |S residual| >0, repeats whole process with residual S.

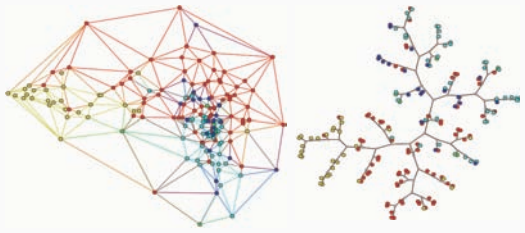
Sequential covering with Multiple restart

- Variance and Coverage
- Partitioning Strategies
- Grid
  - Resize
  - Slide
- Cluster
  - Cluster number

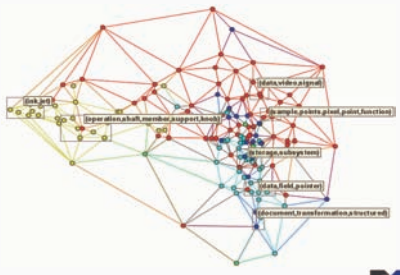
5.4 Further Examples

- RSS Patent Data, recovered from the Web http://www.freepatentsonline.com/
- Case 1:
  - 170 files
  - Graphics processing, printer, database, document, ai

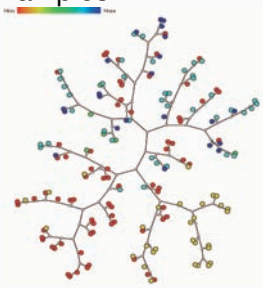
Further Examples

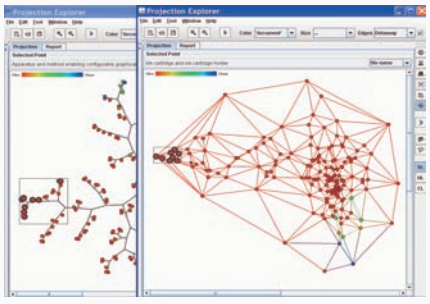


Further Examples

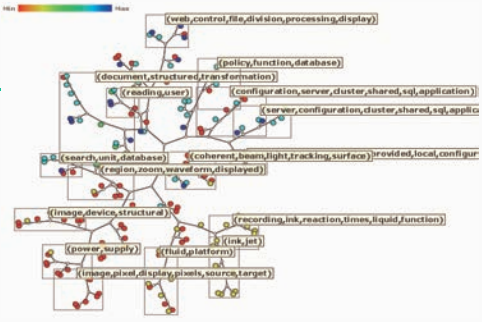


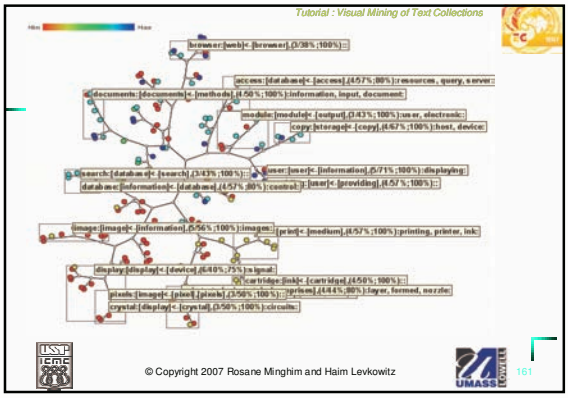
Further Examples

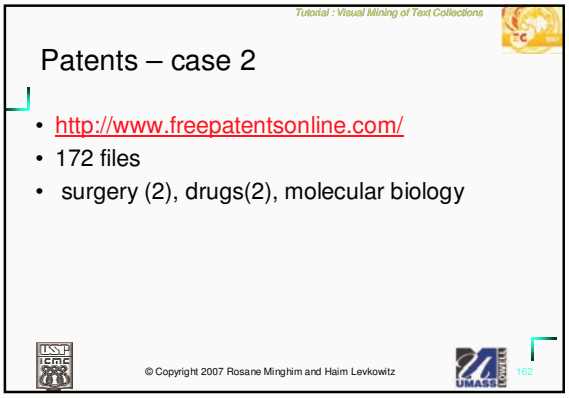


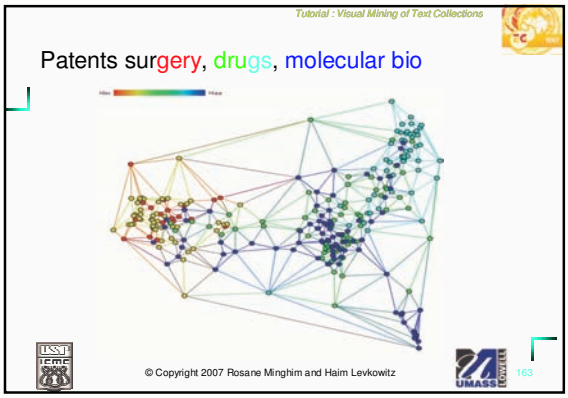


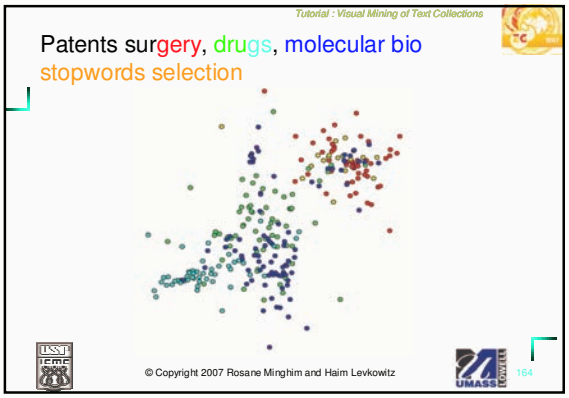
(ink jet, document)



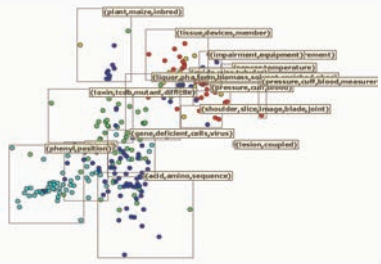




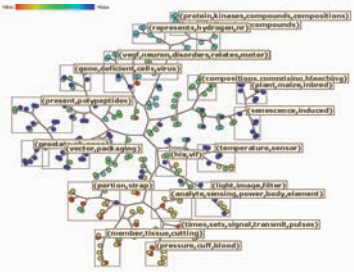




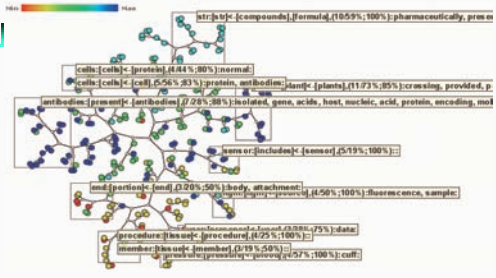
Patents surgery, drugs, molecular bio topics



Patents surgery, drugs, molecular bio



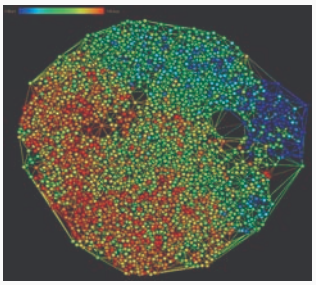
Patents surgery, drugs, molecular bio



Further Examples

- Cattle performance data
  - Translated to text from categorical information, e.g.,
    - Ranges of weight to words such as:
      {weight_below_fifty_percent;
      weight_between_fifty_seventy_five; etc..}
  - 9135 individuals

Cattle performance data



Colored by word 'top'

6. Conclusions, Challenges, Trends

- Other M-D data
- Measures of similarity
- Metrics for quality of projections and point placement
- Interaction and Exploration Paradigms
- Time representation

Conclusions, Challenges, Trends

- Vector Representation
  - Stopword lists/vocabulary
  - Alternative representations
  - Heterogeneous vs. Homogeneous
- Scalability: two levels?
  - Massive
  - Close examination

Conclusions, Challenges, Trends

- Evaluation
- Customization
- Further coupling with mining
  - Supervised/non-supervised
- Coordination, Multiple Views

EG:1022

EG:1023

3D Shape Description and Matching Based on Properties of Real Functions

Eurographics 2007 Tutorial T12

Organizers

Bianca Falcidieno (CNR - IMATI - GE, Italy)
Michela Spagnuolo (CNR - IMATI - GE, Italy)

Speakers

Bianca Falcidieno (CNR - IMATI - GE, Italy)
Daniela Giorgi (CNR - IMATI - GE, Italy)
Simone Marini (CNR - IMATI - GE, Italy)
Giuseppe Patan (CNR - IMATI - GE, Italy)
Michela Spagnuolo (CNR - IMATI - GE, Italy)

EG:1024

3D shape description and matching based on properties of real functions

S. Biasotti[1], B. Falcidieno[1], P. Frosini[2], D. Giorgi[1], C. Landi[3], S. Marini [1], G. Patané[1] and M. Spagnuolo[1]

[1] CNR - IMATI - GE, Italy
[2] University of Modena e Reggio Emilia, Italy
[3] University of Bologna, Italy

Abstract

This tutorial covers a variety of methods for 3D shape matching and retrieval that are characterized by the use of a real-valued function defined on the shape (mapping function) to derive its signature. The methods are discussed following an abstract conceptual framework that distinguishes among the three main components of these class of shape matching methods: shape analysis*, via the application of the mapping function, shape* description*, via the construction of a signature, and* comparison*, via the definition of a distance measure.*
Goal of the tutorial is to facilitate the understanding of the performance of the various methods by a methodical analysis of the properties of various methods at the three different stages.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Line and Curve Generation

1. Introduction

3D shape matching and retrieval are key aspects in the current panorama of search engines. Shape models carry a high value with them, and search engines able to retrieve this type of visual media would be surely useful to speed-up content design, re-use and processing. Keyword-based searching is simply not sufficient to achieve the necessary capability of resource exploration for 3D. Therefore, a variety of methods have been proposed in the literature to tackle the problem with different approaches that span from coarse filters suited to browse very large 3D repositories on the web, to domain-specific approaches.

Generally speaking, shape matching methods rely on the computation of a shape *description*, also called signature, that effectively captures some essential features of the object. The shape descriptions are then compared using an appropriate computational technique able to translate the similarity between objects into some distance between descriptors. The majority of the methods proposed in the literature mainly focus on geometric aspects, that is, the description characterizes the spatial distribution or extent of the object in the 3D space [NK01, OFCD02, KFR03]. From a practical point of view, the main advantage of these methods is that they do not make specific assumption on the topology of the digital models and the computational efficiency. Conversely, these methods generally fail in supporting more elaborate shape comparisons, such as partial matching or sub-part correspondence where the similarity has to be evaluated in terms of presence and similarity of features in the shapes. In this case, more sophisticated descriptions should be used, in order to properly characterize the essential features and store them in an efficient and salient structure. Several approaches to shape characterization have been adopted in the literature (e.g. curvature, level-sets, enclosed spheres), yielding to different structuring methods (e.g. patch segmentation, Reeb graph, skeletons, medial axis).

Given the complexity of the problem, understanding and evaluating the performance of methods for 3D matching is not an easy task: first of all, there is neither a single *best* shape characterization nor a single *best* similarity measure, and the solution largely depends on the type of shapes to be analyzed and on the application domains. Recently, a 3D shape retrieval contest has been proposed – SHREC – whose general objective is to eval-

uate the performances of 3D-shape retrieval algorithms http://www.aimatshape.net/event/SHREC/ . The initial results of the contest provided the first opportunity to analyze the various algorithms, their strengths, as well as their weaknesses, using a common test collection which allows a direct comparison of algorithms. A single test collection necessarily delivers only a partial view of the whole picture, and for this reason the contest quickly moved towards a multi-track organization, for partial and whole matching, polygon soup and watertight model matching, as well as a number of context-specific benchmarks, for example for mechanical part matching, molecule matching, or 3D face matching.

2. Tutorial focus and contribution

While the performance of retrieval can be evaluated in quantitative terms using appropriate benchmarks and ground truth, it is not easy to understand the contribution to the results of the various components of the retrieval system. The results, indeed, depend both on the shape descriptions and the comparison tools, which are very often quite intertwined. Moreover, existing surveys [BKS*05, TV04, BP06] mainly focus on a classification and discussion of geometry-oriented methods, which target the conversion of statistical and geometric shape analysis into feature vectors or histograms. The comparison among methods usually addresses properties of admissible input representations and formats, invariance of the description with respect to a transformation class, and retrieval performance.

The goal of the tutorial is to facilitate the understanding of the performance of the various methods by a methodical analysis of the properties of various methods at the three different stages of an abstract conceptual framework which distinguishes among the three main components of these class of shape matching methods: shape *analysis*, via the application of some mathematical technique, shape *description*, via the construction of a signature, and *comparison*, via the definition of a distance measure. More precisely, we will analyze in depth methods that approach the analysis phase by making use of the properties provided by some real function f, called the mapping function, defined on the surface $\mathcal{M}$ representing the 3D object. Therefore, the underlying conceptual framework is structured in three-steps:

1. choice and evaluation of the real functions f_i on $3D$ shapes $\mathcal{M}_i$;
2. construction of *high-level descriptors* $\mathcal{G}_i$ of $\mathcal{M}_i$, using f_i;
3. choice of the comparison techniques to be used for the set of shapes and descriptors $\{(\mathcal{M}_i, \mathcal{G}_i)\}_i$.

We believe that the discussion of the properties at the three levels will facilitate the evaluation of theoretical and practical performances of the methods, will indicate more precisely the strength and weaknesses of the methods, and will also suggest a way for adopting different shape descriptors according to the properties and invariants that one wishes to investigate. The choice of the real function and the nature of the descriptor play indeed the role of the "lens" through which we look at the properties of the shape. The generality and flexibility of the framework is of interest for a wide research community with applications to visualization and topological modeling. In this tutorial, we will overview and analyze a large set of solutions, evaluate their effectiveness, and discuss perspectives, open issues, and future developments.

3. Outline

The proposed tutorial relies on recent survey work of the authors in related fields, see [BFF*06, Mar05, BAB*07].

The updated version of the slides presented at Eurographics 2007 will be made available at the following URL: http://www.ge.imati.cnr.it/ima/smg/training.html

In the following, we outline the main items that we plan to discuss in the tutorial, by giving for each group a synthetic description of the methods and a summary of the most relevant references, which will be discussed in detail and with examples and emphasis on shape matching applications.

3.1. Shape matching: motivations and challenges

The first part of the tutorial will provide an introduction to the tutorial, explain the rationale of the presentation, and introduce some of the main challenges of the topic area and its perspective impact in a number of crucial applications.

3.2. Properties of the real functions

A variety of different functions have been used in the shape matching literature for characterizing relevant features of objects. In general, the availability of *a-priori* information on the classes of the input database can be used to select the mapping functions which are best suited to identify specific shape features (e.g., protrusions), thus constraining the retrieval to match them with a higher degree of importance with respect to other features. This par of the tutorial will provide some introductory definitions on the basic concepts that will be discussed, concerning critical points, Morse function, level sets and briefly introduce their discretization [Ban70, Ban67, GP74, Mil63]. Following, a variety of real-valued functions will be presented and discussed, grouped into four main categories according to their definition, domain and properties:

- the *height* [SKK91, FK97] function is among the most intuitive and simple choices for analysing the shape of an object; since it depends on the direction considered, its usage is preferred for applications in which objects have a natural predefined direction (Figure 1(a)). A more elaborate characterization of the shape according to differences in the elevation value is provided by the *elevation* [AEHW06] function, which derives from the traditional height function but aims at a rotation invariant

analysis. The notion of elevation captured by this function measures how much a point is relevant in its normal direction with respect to its neighbourhood. The elevation function is defined by pairing the critical points of the height function in all directions.

- Shape properties can be effectively characterized by measuring distances between feature points or by evaluating the elongation of the shape. In this broad class, the analysis approaches based on the *geodesic distance* generally provide and isometry invariant characterization of a shape [BBK06a]. Geodesic distance has been applied in several settings, in particular for the evaluation the geodesic distance of mesh vertices from selected feature points [MP02, EK03], and for averaging all geodesic distances among the vertices [HSKK01, KT03, GSCO07] .
  The *Euclidean distance from a point* $\mathbf{p} \in \mathbb{R}^3$ [FK97, SV01] (e.g., the barycentre of $\mathcal{M}$, Figure 1(b)) has also been used, as it is invariant to the shape embedding and detects protrusions (resp. hollows) of $\mathcal{M}$ with respect to $\mathbf{p}$ as regions of influence of maxima (resp. minima) f.
- *curvature-based analysis* have been frequently used to characterize the shape of 3D objects; generally, curvature-based analysis are rather sensible to noise or small features and to the quality of the shape discretization in terms of sampling density and tiny triangles. More robust computation is achieved either using variations of the curvature evaluation function (e.g. [GCO06]), polynomial surface fitting [ZP01], or with a multi-scale curvature evaluation where details are discarded [MPS*04].
- The *local diameters* function [GSCO07] aims at measuring the shape by computing the *diameter* of the volume enclosed by the surface. Therefore, it provides a volumetric rather than a boundary characterization, similarly to the *distance tranforms* [DS06] which is more focused on the medial axis radius.
- If the shapes to be compared do not exhibit a uniform structure, *harmonic* [NGH04, Flo97, PP93] and *Laplacian-based* functions [RWP06, DBG*06] may provide a new and powerful set of descriptors for shape analysis as they are intrinsically defined by the Laplacian matrix of the shape (see Figure 1(c-d)).
  We will discuss the numerical (in)stability of extraction of this type of functions from the Laplacian matrix of the shape $\mathcal{M}$, a very relevant aspect that has to be considered to understand at which extent this instability affects the descriptor of $\mathcal{M}$, and eventually the matching algorithm [GV89].

The presentation and discussion of the above-cited functions will be carried out considering:

- the *saliency* of f, as its ability to identify relevant shape features of $\mathcal{M}$;
- the *smoothness* degree of f, meant as its behaviour with respect to the number, nature and properties of its critical points;

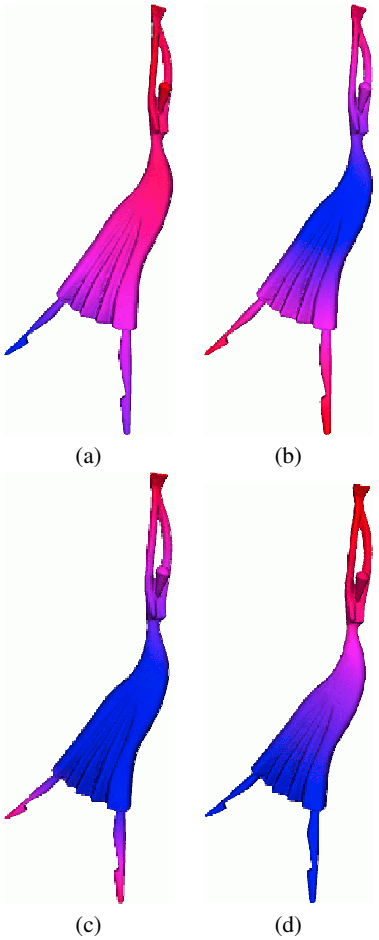


Figure 1: *(a) Height function, (b) Euclidean distance from the center of mass, (c) harmonic function, (d) first eigenfunction of the Laplacian matrix of the model.*

- the *stability* of f with respect to its discretization and computation on $\mathcal{M}$;
- the *robustness* of f, that is, the variation of f with respect to small geometric changes of the shape $\mathcal{M}$;
- the *degree of freedom* (DoF) and the number of *heuristics* used in the definition and evaluation of f.
- the *efficiency* of f in terms of the computational cost required by its evaluation on $\mathcal{M}$;
- the *invariance* of f to transformation groups;
- the hypotheses or restrictions on the *input*.

The analysis of the properties and the potentialities of the fs will provide an insight into the formalization of function suites, beyond a generic best-practice or rule-of-thumbs.

3.3. Properties of the shape descriptors

In the literature, it is quite common that functions used to analyse the shape are directly associated to a corresponding signature, or shape descriptor. For some of the methods this association is exclusive, meaning that no other function can be used to produce the same descriptor, while for other methods the descriptor is *parametric* with respect to the choice of the function.

Among shape descriptors that are parametric with respect to the choice of f, we will present:

- Reeb graphs [Ree46, CMEH*03, HSKK01, ABS03, Bia04, TS05, BFS00], size theory [Fro90, FL01, FL99, dFL06, FM99, BCF*07] and persistent homology tools [ELZ02, CZCG04, CZCG05, WAB*05, ZC05, CSEH05, CSEH07] are topological descriptors that root in Morse theory. When the function f varies, a collection of descriptors may be obtained. For any f, these descriptors code the shape by the configuration of elements or properties that characterize the topological evolution of level sets or lower level sets of f, see Figure 2;
- descriptors that decompose a function f given over simpler basis functions; examples are the spherical harmonic shape decompositions [KFR03, Vra04, VSR01] and wavelets-based methods [LTN06].

Among shape descriptors that exclusively linked to a specific choice of f, we will present:

- descriptors based on quantities extracted by intrinsic shape functions, such as the spectrum of the Laplace-Beltrami operator [RWP06, RWP07, NRW*07];
- descriptors built on isometry invariant quantities, as for example the geodesic function [JZ06, JZ05, EK03, BBK05, BBK06b, BBK06a] or the curvature [ZP01, GCO06];
- the pose-oblivious shape signature [GSCO07], that associate to $\mathcal{M}$ histograms of the distribution over the shape of two real functions, the first related to surface and the second to volume information;
- the centerline skeleton that connects feature points through the geodesic distance [MP02])

The shape descriptors will be presented from a theoretical and computational point of view, providing examples and results to assess different aspects, in particular:

- the *saliency* of the descriptor, that is its ability to capture the structure of the shape in terms of its features;
- the *concisness* of the descriptor, that is its ability to minimize the memory needed to store the descriptor while maximizing the amount of information represented; this property is related also to the type of output produced;
- the *robustness* with respect to small changes of the shape;
- the *unicity* of the descriptor: once the theoretical methodology for extracting the descriptor, the algorithm, and possible parameters have been chosen, the descriptor is unique;

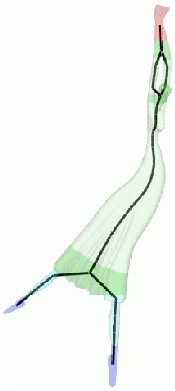


Figure 2: *(a) Reeb graph of the first eigenfunction of the Laplacian matrix of the model.*

- the *completness* in the sense that the same descriptor cannot be associated to different shapes;
- the *invariance* of the descriptor to transformation groups;
- the *degree of freedom* (DoF) and the number of *heuristics* used in the construction of the descriptor;
- the hypotheses or restrictions on the *input*;
- the *efficiency* of the descriptor in terms of the computational cost required by its construction.

3.4. Comparison methodologies

Although the surveyed descriptors are inspired by the same idea of quantifying geometric properties conveyed by f, there are substantial differences in the shape interpretation they provide and in the structures used to encode the shape information. In particular, the type of structure produced strongly influences the choice of the methods adopted for the final shape comparison step. The methodologies will be presented following a logical grouping according to the type of coding of the shape descriptor:

- the similarity between descriptors encoded as *histograms, feature vectors, or matrix structures* is evaluated by linear algebraic or statistical techniques [KFR03, Vra04, LTN06];
- the similarity among descriptors stored as *graphs* is generally evaluated by graph-matching techniques [HSKK01, SSGD03, LK03, CDS*05, BSRS04, ZSm*05, BRS06, BMSF06] (see Figure 3).
- the similarity between combinatorial descriptors is measured by friendly and computationally efficient tools, such as persistence diagrams and formal series [dFL06, BCF*07, CSEH07].

The methodologies will be presented and discussed highlighting their properties in terms of the following characteristics:

- the *properties* of the similarity measure that characterize

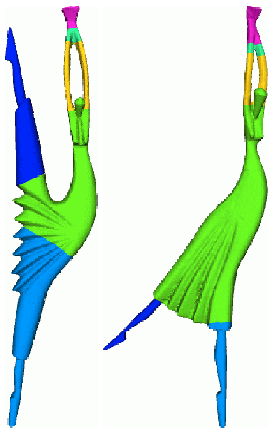


Figure 3: *Sub-part correspondence obtained using the graph comparison method defined in [BMSF06].*

it as a metric, semi-metric, or pseudo-distance [VH01, Tve77, SJ99];

- the *robustness* of the measure with respect to small changes of the shape;
- the *type of comparison* provide by the measure, in terms of supporting global, partial or sub.part correspondence;
- the *type of information*: according to the type of information stored and the way it is coded in the descriptor, the measure of similarity may take into account geometric, topological or structural information;
- the *efficiency* in terms of computational complexity required to evaluate the measure;
- the *application scenario* in which the comparison is performed.

3.5. Conclusions and future perspectives

In the conclusive part of the tutorial, we will try to provide a coherent comparison of the various techniques at the three levels of the framework, based on the analysis provided for all the aspects discussed. Obviously, the tutorial does not claim either to be an exhaustive survey of the wealth of existing methods for 3D matching or to examine all technical details of each single method. Rather, the objective of the comparison is to give a structured presentation of the methods in terms of the several properties of the descriptors and comparison tools, that are often not discussed in details in existing surveys. We believe that the presentation and discussions organized in this manner should serve as a basis for extending the performance analysis beyond standard precision-recall diagrams and help the user to understand if the reasons of good or bad retrieval results depend, for instance, on an insufficient efficacy of the descriptor, on an intrinsic instability of the function, or also on an inappropriate comparison tool.

Finally, we will list a series of topics deserving further research, such as the role of invariance with respect to transformation groups, the concurrent use of more than a single characterizing function, and the need to balance the use of geometrical and topological information for accurate shape descriptions. Last but not least, we will also address issues related to the emerging use of semantic indicators to perfom matching and retrieval, based either on (semi)-automatic annotation of shapes or in supervised classification and prototype extraction.

4. Authors' CVs

Two research groups are involved:

The *Shape Modeling Group at CNR-IMATI-GE* works since years on topics related to geometric modelling with the main aim to describe the shape of objects through geometric and topological reasoning techniques. Lately, the research themes focus on broadening the role of traditional modelling with the definition of new representations, encapsulating also knowledge technologies methodologies, able to express also the semantic level at which the perception of shape is encoded. In this field, CNR-IMATI-GE is leading the FP6 European Project NoE AIM@SHAPE.

The team *Vision Mathematics of the Univ. of Bologna*, Dept. of Mathematics, works at the use of topology and geometry in robotic applications since 1988. Mainly, the team deals with computer vision by means of a shape descriptor (the Size Functions) conceived and developed by P. Frosini. But the group interests cover a fairly wide area reaching from the abstractions of manifold topology to robot navigation and to concrete application projects.

Bianca Falcidieno is Research Director at CNR and head of the Shape Modelling Group, working in the field of Applied Mathematics and Computer Science, with applications in Computer Graphics, Geographic Information Systems, and Industrial Design. She is Editor in chief of the International Journal Shape Modelling, member of the Steering Committee of Shape Modeling International (SMI), and author of more than 200 scientific refereed papers and books. Bianca Falcidieno is the coordinator of the FP6 NoE AIM@SHAPE.

Patrizio Frosini is assistant professor in the Faculty of Engineering at the Univ. of Bologna. He is a member of the ARCES group at the Univ. of Bologna. He received the PhD degree in Mathematics from the Univ. of Florence (1991). His research interests include the study of geometrical-topological methods for shape comparison and related applications in Computer Vision.

Claudia Landi is assistant professor at the Univ. of Modena and Reggio Emilia in Reggio Emilia (Italy). She obtained a PhD in Mathematics in 2000, at the University of Pisa. Since 1994 she is member of the Vision Mathematics Group of the University of Bologna. Her main research interest is shape description via geometry and topology.

Michela Spagnuolo is senior researcher at CNR-IMATI-GE and received the Ph.D. in Computer Science Engineering, at the INSA, Lyon, France (1997). Her research interests are related to shape-based approaches to modeling digital shapes, computational topology techniques for shape analysis, geometric reasoning for the extraction of shape features from discrete surface models, and geometric models for coding uncertainty in data samples (fuzzy-based modelling). She is a member of the Steering Committee of Shape Modeling International (SMI).

Silvia Biasotti is researcher at CNR-IMATI-GE and received a Ph.D. in Mathematics and Applications at the Univ. of Genoa (2004). Her research interests include computational topology, shape abstraction and skeleton representation of polyhedral surfaces.

Daniela Giorgi is research fellow at CNR-IMATI-GE and received a Ph.D. in Applied Mathematics at the Univ. of Padua (2006). Her research interests are in Pattern Recognition and topological methods for shape analysis.

Simone Marini is researcher at CNR-IMATI-GE and received a Ph.D. in Electronic and Computer Engineering at the Univ. of Genova (2005). His main interests concern evaluation of 3D shape similarity, graph comparison, and ontological representation of scientific concepts.

Giuseppe Patané is researcher at CNR-IMATI-GE and received a Ph.D. in Mathematics and Applications at the Univ. of Genova (2005). His research interests include numerical analysis (implicit surfaces), shape analysis, computational geometry (topological graphs, local and global parameterization).

Acknowledgements

This work has been partially supported by the EC-IST FP6 Network of Excellence AIM@SHAPE. Models are courtesy of the AIM@SHAPE repository http://www.aimatshape.net.

References

[ABS03] ATTENE M., BIASOTTI S., SPAGNUOLO M.: Shape understanding by contour-driven retiling. *The Visual Computer 19*, 2-3 (2003), 127–138.

[AEHW06] AGARWAL P. K., EDELSBRUNNER H., HARER J., WANG Y.: Extreme elevation on a 2-manifold. *Discrete Comput. Geom. 36*, 4 (2006), 553–572.

[BAB*07] BIASOTTI S., ATTALI D., BOISSONNAT J.-D., EDELSBRUNNER H., ELBER G., MORTARA M., DI BAJA G. S., SPAGNUOLO M., TANASE M., VELTKAMP R.: Skeletal structures. In *Shape Analysis and Structuring*, Floriani L. D., Spagnuolo M., (Eds.). Springer-Verlag, 2007.

[Ban67] BANCHOFF T.: Critical points and curvature for embedded polyhedra. *Journal of Differential Geometry 1* (1967), 245–256.

[Ban70] BANCHOFF T. F.: Critical points and curvature for embedded polyhedral surfaces. *Am. math. Monthly 77* (1970), 475–485.

[BBK05] BRONSTEIN A. M., BRONSTEIN M. M., KIMMEL R.: On isometric embedding of facial surfaces into $\mathbb{S}^3$. In *Proceedings International Conference on Scale Space and PDE Methods in Computer Vision* (2005), vol. 3459 of *Lecture Notes in Computer Science*, pp. 622–631.

[BBK06a] BRONSTEIN A. M., BRONSTEIN M. M., KIMMEL R.: Efficient computation of isometry-invariant distances between surfaces. *SIAM Journal on Scientific Computing 28*, 5 (2006), 1812–1836.

[BBK06b] BRONSTEIN A. M., BRONSTEIN M. M., KIMMEL R.: Generalized multidimensional scaling: A framework for isometry-invariant partial surface matching. *Proceedings National Academy of Science 103*, 5 (2006), 1168–1172.

[BCF*07] BIASOTTI S., CERRI A., FROSINI P., GIORGI D., LANDI C.: *Multidimensional size functions for shape comparison.* Tech. Rep. 4, IMATI, 2007.

[BFF*06] BIASOTTI S., FLORIANI L. D., FALCIDIENO B., FROSINI P., GIORGI D., LANDI C., PAPALEO L., SPAGNUOLO M.: *Geometrical-topological properties of real functions for describing shapes.* Tech. Rep. 5, IMATI, 2006.

[BFS00] BIASOTTI S., FALCIDIENO B., SPAGNUOLO M.: Extended Reeb Graphs for surface understanding and description. In *Discrete Geometry for Computer Imagery Conference* (2000), Borgefors G., di Baja G. S., (Eds.), vol. 1953 of *Lecture Notes in Computer Science*, Springer, pp. 185–197.

[Bia04] BIASOTTI S.: *Computational Topology Methods for Shape Modelling Applications*. PhD thesis, Universitá degli Studi di Genova, May 2004.

[BKS*05] BUSTOS B., KEIM D. A., SAUPE D., SCHRECK T., VRANIĆ D. V.: Feature-based similarity search in 3D object databases. *ACM Computing Surveys 37*, 4 (December 2005), 345–387.

[BMSF06] BIASOTTI S., MARINI S., SPAGNUOLO M., FALCIDIENO B.: Sub-part correspondence by structural descriptors of 3D shapes. *Computer-Aided Design 38*, 9 (2006), 1002–1019.

[BP06] BIMBO A. D., PALA P.: Content-based retrieval of 3D models. *ACM Trans. on Multimedia Computing, Communications and Applications 2*, 1 (2006), 20–43.

[BRS06] BESPALOV D., REGLI W. C., SHOKOUFANDEH A.: Local feature extraction and matching partial objects. *Computer-Aided Design 38*, 9 (2006), 1020–1037.

[BSRS04] BESPALOV D., SHOKOUFANDEH A., REGLI W. C., SUN W.: Local feature extraction using scale-space decomposition. In *ASME Design Engineering Tech-*

nical Conferences, Computers and Information in Engineering Conference Conference (DETC 2004-57702) (Sep 2004), ASME Pres.

[CDS*05] CORNEA N. D., DEMIRCI M. F., SILVER D., SHOKOUFANDEH A., DICKINSON S. J., KANTOR P. B.: 3D object retrieval using many-to-many matching of curve skeletons. In *Proc. of Shape Modeling and Applications 2005 (SMI' 05)* (2005), IEEE Computer Society, pp. 368–373.

[CMEH*03] COLE-MCLAUGHLIN K., EDELSBRUNNER H., HARER J., NATARAJAN V., PASCUCCI V.: Loops in Reeb graphs of 2-manifolds. In *Proc. of the ACM Symposium on Computational Geometry* (2003), ACM Press, pp. 344–350.

[CSEH05] COHEN-STEINER D., EDELSBRUNNER H., HARER J.: Stability of persistence diagrams. In *Proceedings of SoCG 2005: ACM Symposium on Computational Geometry* (2005), pp. 263–271.

[CSEH07] COHEN-STEINER D., EDELSBRUNNER H., HARER J.: Stability of persistence diagrams. *Discrete Computational Geometry 37* (2007), 103–120.

[CZCG04] COLLINS G., ZOMORODIAN A., CARLSSON A., GUIBAS L. J.: A barcodes shape descriptor for curve point cloud data. *Computers and Graphics 28* (2004), 881–894.

[CZCG05] CARLSSON G., ZOMORODIAN A., COLLINS A., GUIBAS L. J.: Persistence barcodes for shapes. *Intern. Journal of Shape Modeling 11*, 2 (2005), 149–187.

[DBG*06] DONG S., BREMER P.-T., GARLAND M., PASCUCCI V., HART J.: Spectral surface quadrangulation. *ACM SIGGRAPH 25*, 3 (August 2006). Proceedings SIGGRAPH 2006.

[dFL06] D'AMICO M., FROSINI P., LANDI C.: Using matching distance in size theory: a survey. *International Journal of Imaging Systems and Technology 16*, 5 (2006), 154–161.

[DS06] DEY T. K., SUN J.: Defining and computing curve-skeletons with medial geodesic function. In *Proc. of Symposium on Geometry Processing* (2006), pp. 143–152.

[EK03] ELAD A., KIMMEL R.: On bending invariant signatures for surfaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence 25*, 10 (2003), 1285–1295.

[ELZ02] EDELSBRUNNER H., LETSCHER D., ZOMORODIAN A.: Topological persistence and simplification. *Discrete Computational Geometry 28* (2002), 511–533.

[FK97] FOMENKO A., KUNII T. L.: *Topological Modelling for Visualization*. Springer Verlag, 1997.

[FL99] FROSINI P., LANDI C.: Size theory as a topological tool for computer vision. *Pattern Recognition and Image Analysis 9* (1999), 596–603.

[FL01] FROSINI P., LANDI C.: Size functions and formal series. *Appl. Algebra Engrg. Comm. Comput. 12* (2001), 327–349.

[Flo97] FLOATER M. S.: Parametrization and smooth approximation of surface triangulations. *Computer Aided Geometric Design 14*, 3 (1997), 231–250.

[FM99] FROSINI P., MULAZZANI M.: Size homotopy groups for computation of natural size distances. *Bull. Belg. Math. Soc. 6* (1999), 455–464.

[Fro90] FROSINI P.: A distance for similarity classes of submanifolds of a Euclidean space. *Bull. Austral. Math. Soc. 42* (1990), 407–416.

[GCO06] GAL R., COHEN-OR D.: Salient geometric features for partial shape matching and similarity. *ACM Transactions on Graphics 25*, 1 (2006), 130–150.

[GP74] GUILLEMIN V., POLLACK A.: *Differential Topology.* Englewood Cliffs, NJ: Prentice-Hall, 1974.

[GSCO07] GAL R., SHAMIR A., COHEN-OR D.: Pose-oblivious shape signature. *IEEE Transactions on Visualization and Computer Graphics 13*, 2 (2007), 261–271.

[GV89] GOLUB G., VANLOAN G.: *Matrix Computations*. John Hopkins University Press, 2nd. edition, 1989.

[HSKK01] HILAGA M., SHINAGAWA Y., KOHMURA T., KUNII T. L.: Topology matching for fully automatic similarity estimation of 3D shapes. In *ACM SIGGRAPH* (2001), pp. 203–212.

[JZ05] JAIN V., ZHANG H.: A spectral approach to shape-based retrieval of articulated 3d models. *CAD 37*, 5 (2005), 509–530.

[JZ06] JAIN V., ZHANG H.: Robust 3d shape correspondence in the spectral domain. In *Proc. of Shape Modeling International* (2006), pp. 118–129.

[KFR03] KAZHDAN M., FUNKHOUSER T., RUSINKIEWICZ S.: Rotation invariant spherical harmonic representation of 3D shape descriptors. In *Proc. of Symposium on Geometry Processing* (2003), pp. 156–165.

[KT03] KATZ S., TAL A.: Hierarchical mesh decomposition using fuzzy clustering and cuts. *ACM SIGGRAPH* (July 2003), 954–961.

[LK03] LEYMARIE F. F., KIMIA B. B.: Computation of the shock scaffold for unorganized point clouds in 3d. In *Computer Vision and Pattern Recognition* (June 2003), vol. 1, pp. 821–827.

[LTN06] LAGA H., TAKAHASHI H., NAKAJIMA M.: Spherical wavelet descriptors for content-based 3d model retrieval. In *Proc. of Shape Modeling and Applications* (2006), pp. 15–25.

[Mar05] MARINI S.: *3D Shape Similarity Through Structural Descriptors*. PhD thesis, University of Genova, April 2005.

[Mil63] MILNOR J.: *Morse Theory.* Princeton University Press, New Jersey, 1963.

[MP02] MORTARA M., PATANÉ G.: Shape-covering for skeleton extraction. *Int. Journal of Shape Modelling 8*, 2 (2002), 245–252.

[MPS*04] MORTARA M., PATANE G., SPAGNUOLO M., FALCIDIENO B., ROSSIGNAC J.: Blowing bubbles for multi-scale analysis and decomposition of triangle meshes. *Algorithmica 38*, 1 (2004), 227–248.

[NGH04] NI X., GARLAND M., HART J. C.: Fair Morse functions for extracting the topological structure of a surface mesh. *ACM SIGGRAPH 23*, 3 (2004), 613–622.

[NK01] NOVOTNI M., KLEIN R.: A geometric approach to 3D object comparison. In *Proc. of Shape Modelling and Applications* (2001), pp. 167–175.

[NRW*07] NIETHAMMER M., REUTER M., WOLTER F.-E., BOUIX S., PEINECKE N., KOO M.-S., SHENTON M.: Global medical shape analysis using the laplace-beltrami spectrum. In *MICCAI07, 10th International Converence on Medical Image Computing and Computer Assisted Intervention* (2007).

[OFCD02] OSADA R., FUNKHOUSER T., CHAZELLE B., DOBKIN D.: Shape distributions. *ACM Trans. Graph. 21*, 4 (2002), 807–832.

[PP93] PINKALL U., POLTHIER K.: Computing discrete minimal surfaces and their conjugates. *Experimental Mathematics* (1993), 15–36.

[Ree46] REEB G.: Sur les points singuli'ers d'une forme de Pfaff complètement intégrable ou d'une fonction numèrique. *Comptes Rendu Hebdomadaires des Séances de l'Académie des Sciences 222* (1946), 847–849.

[RWP06] REUTER M., WOLTER F.-E., PEINECKE N.: Laplace-Beltrami spectra as Shape-DNA of surfaces and solids. *Computer-Aided Design 38*, 4 (2006), 342–366.

[RWP07] REUTER M., WOLTER F.-E., PEINECKE N.: Laplace spectra as fingerprints for image recognition. *Computer-Aided Design 39* (2007), 460–476.

[SJ99] S S., JAIN R.: Similarity measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 9 (September 1999), 871–883.

[SKK91] SHINAGAWA Y., KUNII T. L., KERGOSIAN Y. L.: Surface coding based on Morse theory. *IEEE Computer Graphics and Applications 11* (1991), 66–78.

[SSGD03] SUNDAR H., SILVER D., GAGVANI N., DICKINSON S.: Skeleton based shape matching and retrieval. In *Proc. of Shape Modelling and Applications* (2003), pp. 130–139.

[SV01] SAUPE D., VRANIC D. V.: 3d model retrieval with spherical harmonics and moments. In *Proceedings of the 23rd DAGM-Symposium on Pattern Recognition* (London, UK, 2001), Springer-Verlag, pp. 392–397.

[TS05] TUNG T., SCHMITT F.: The Augmented Multiresolution Reeb Graph approach for content-based retrieval of 3D shapes. *Int. Journal of Shape Modelling 11*, 1 (June 2005), 91–120.

[TV04] TANGELDER J., VELTKAMP R.: A survey of content based 3d shape retrieval methods. In *Proceedings of Shape Modeling Applications, 2004* (2004), pp. 145–156.

[Tve77] TVERSKY A.: Features of similarity. *Psychological Review 84* (1977), 327–352.

[VH01] VELTKAMP R. C., HAGENDOORN M.: State-of-the-Art in Shape Matching. In *Principles of Visual Information Retrieval*, Lew M., (Ed.). Springer-Verlag, 2001, pp. 87–119.

[Vra04] VRANIC D.: *3D model retrieval*. PhD thesis, University Leipzig, June 2004.

[VSR01] VRANIC D., SAUPE D., RICHTER J.: Tool for 3D-object retrieval: Karhunen-Loeve transform and spherical harmonics. In *Proc. of IEEE 2001 Workshop on Multimedia Signal Processing* (2001), Dugelay J.-L., Rose K., (Eds.), pp. 293–298.

[WAB*05] WANG Y., AGARWAL P. K., BROWN P., EDELSBRUNNER H., RUDOLPH J.: Coarse and reliable geometric alignment for protein docking. In *Proceedings of Pacific Symposium on Biocomputing* (2005), pp. 64–75.

[ZC05] ZOMORODIAN A., CARLSSON G.: Computing persistent homology. *Discrete Computational Geometry 33* (2005), 249–274.

[ZP01] ZAHARIA T., PRETEUX F. J.: 3D-shape-based retrieval within the MPEG-7 framework. In *Nonlinear Image Processing and Pattern Analysis XII, Edward R. Dougherty; Jaakko T. Astola; Eds.* (May 2001), Dougherty E. R., Astola J. T., (Eds.), vol. 4304 of *Proc. of the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pp. 133–145.

[ZSm*05] ZHANG J., SIDDIQI K., MACRINI D., SHOKOUFANDEH A., DICKINSON S.: Retrieving articulated 3-D models using medial surfaces and their graph spectra. In *Proc. of Int. Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition* (2005).

3D shape description and matching based on properties of real functions

Eurographics 2007 Tutorial T12

S. Biasotti[1], B. Falcidieno[1], P. Frosini[2], D. Giorgi[1], C. Landi[3], S. Marini[1], G. Patané[1],M. Spagnuolo[1]

[1]CNR-IMATI-GE , [2]University of Bologna, [3]University of Modena e Reggio Emilia, Italy

AIM@SHAPE

CNR-IMATI-GE

- The network of CNR research institutes, which are distributed all over the national territory, is multidisciplinary: it has competences in the field of health and biology, of computer science, of environment and climate, of chemistry and physics, of behavioural, economic and social sciences
- 107 institutes in 11 departments

Istituto di Matematica Applicata e Tecnologie Informatiche, Genova

CNR-IMATI-GE

- Applied mathematics and information technology
  - Geometric Modelling
  - Computational Geometry and Topology
  - High Performance Computing
  - Computational Electromagnetics
  - Cognitive models based on ICT
- Applications
  - Product Design, Spatial Data Handling, Ergonomics, Bioinformatics
- Staff
  - 21 staff members
  - 10 contract researchers (PostDoc)
  - 1 PhD

3D shape description and matching based on properties of real functions

Shape matching: motivations and challenges

Eurographics 2007 Tutorial T12

Speaker:

Michela Spagnuolo

CNR-IMATI-GE - Italy

3D shapes

- 3D shapes are digital representations of either physically existing objects or virtual objects that can be processed by computer applications

why is 3D retrieval important ?

3D users

- Professionals
  - Product Modeling, CAD/CAM
  - Design
  - Cultural Heritage
  - Gaming
  - Virtual Environments
  - Medicine
  - Bioinformatics
  - Architects
  - Archeology
- **Non professionals**

why is 3D retrieval important ?

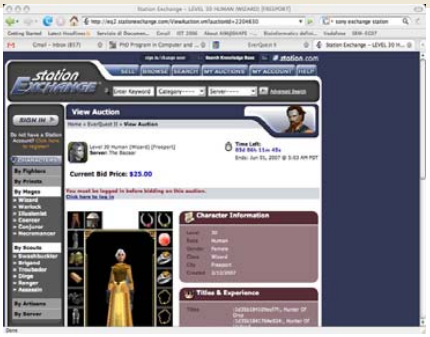
3D retrieval: why ?

- ✓ Due to great technological advances, 3D content is poised to become the 4th wave of multi-media
  - 3D shapes can be easily digitized
  - modelling and processing of 3D shapes are mature research fields (geometric modelling & computer graphics)
  - 3D content can be delivered easily as most of the PCs connected to the Internet are now equipped with high-performance 3D graphics hardware

3D retrieval : why ?

- ✓ plenty of online stores selling 3D models
- ✓ target customers
  - gaming industry, entertainment, simulation
- ✓ industrial impact foreseen at short/medium-term
  - Product Design, 3D-TV, medical sector, gaming, ..
  - SecondLife, EverQuest II, Sony Exchange Station
- ✓ everyday life impact foreseen at long-term
  - *shapes*.google.com
  - 3D acquired and *streamed* on mobile devices

3D media: why ?



3D media: why ?

3D retrieval

- ✓ 3D models carry a high value with them
- ✓ Search & retrieval is necessary for supporting content creation and re-use :
  - Designers
  - Product models
  - Drug design
  - Gaming assets
- ✓ Search & retrieval is necessary for supporting analysis and classification (*understanding*):
  - Medicine
  - Bioinformatics

3D media: problems and peculiarities

- ✓ Data deluge for 3D content is even larger than for 2D content
  - Storage problems
  - Delivery problems
- ✓ Complete shape description
  - Recognition/classification of 2D content has to handle occlusion problems, while 3D content is based on the full representation of the shape
- ✓ Consistency
  - The representation space has the same dimension of the object space

Shape Matching

- Given a query shape ***S***, does ***S*** belong to the repository ?

- Aims the search at a *specific* shape – *precise match* –
- Examples: bioinformatics, drug design, copyright protection

Shape Retrieval

- Given a query shape ***S*** does the repository contain an object **similar** to ***S*** ?

- Aims the search at a *category* level – *similarity assessment* –
- Examples: design, 3D content authoring

Shape Classification

- Given a classified repository and a query shape ***S***, find the class ***S*** belongs to

- Aims the match at a *category* level – *similarity evaluation* –
- Examples: bioinformatics, medical applications

Shape Matching and Retrieval

- Global vs partial match

- Correspondence between parts in similar objects

- Similarity among parts in *dissimilar* objects

Shape similarity

- Driven by the concept of similarity that one wants to implement: **many different *flavours* of similarity !**

geometric congruence

structural equivalence

functional equivalence

"semantic" equivalence

Shape similarity

"semantic" equivalence

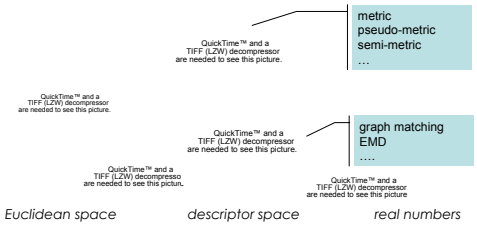
geometric congruence

structural equivalence

functional equivalence

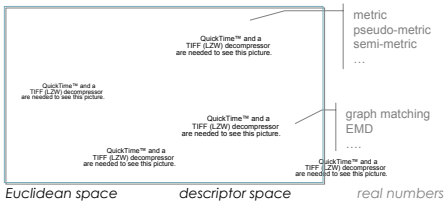
Shape M&R pipeline

- Shape matching is done by associating a description, or signature, to the shapes and by defining a distance, or dissimilarity measure, between descriptors



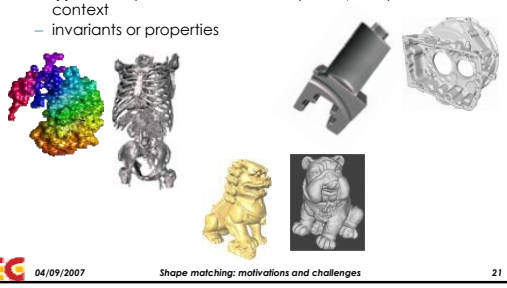
Shape descriptors

- Capture properties of the shape
  - shape matching **context** is very relevant here
  - usually driven by invariants
- Reduce the complexity of the process
- Plenty of different approaches and methods



Shape descriptors

- Descriptors have to take into account the **context** in terms of
  - ***type*** of shapes and their variability/complexity in the context
  - invariants or properties



What tools to *describe* a shape ?

- Geometry
  - *3D shape descriptors based on geometric descriptors (e.g., shape distributions, spherical harmonics, PCA, ..)*
- Structure
  - *3D shape descriptors based on the configuration of features (e.g., skeletons)*
- Semantics
  - *3D shape ontologies and domain conceptualization (e.g., metadata, ontology, reasoners and inference)*

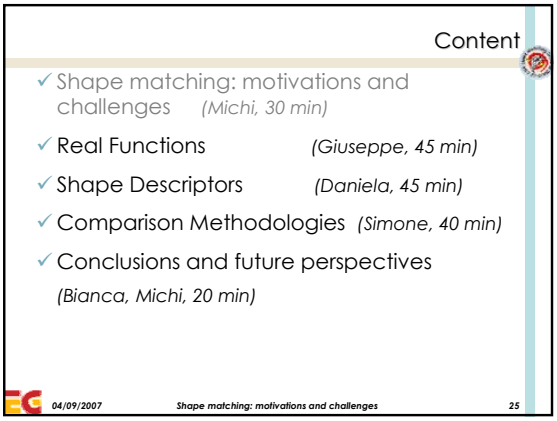
Tutorial focus

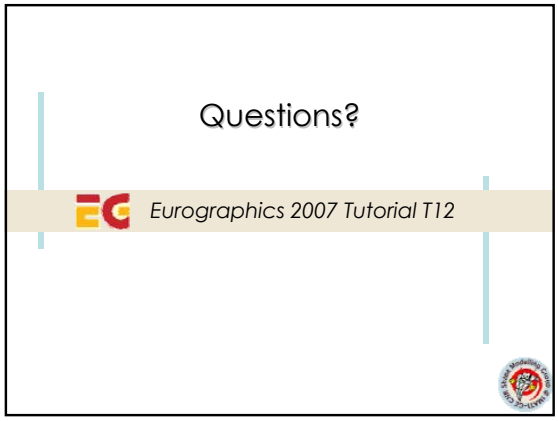
- Among all the shape descriptors, we will focus on those that are defined by the application of a real-valued mapping function *f* that is applied to the shape
  - different functions *f* provide different insights
  - different signatures can be built on tje results of shape analysis with real-valued functions
  - Different comparison methodologies can be applied to signatures

Shape M&R pipeline

- How to measure the performance of shape retrieval systems ?

SHREC'07 - Shape Retrieval Contest 2007





3D Shape Description and Matching Based on Properties of Real Functions

Real functions

Eurographics 2007 Tutorial T12

Speaker

Giuseppe Patanè

CNR-IMATI-GE - Italy

Outline

- ✓ Differential properties of real functions defined on 3D shapes.
- ✓ Discrete case:
  - critical points and level sets: definition and properties
  - functions:
    - definition (ie., height and elevation, Euclidean-, curvature- and geodesic-based functions, local diameters, harmonic function, Laplacian eigenfunctions)
    - properties (ie., saliency, smoothness, stability, robusteness, degrees of freedom and heuristics, efficiency, invariance).

Critical points [GP76,Mil63]

- ✓ Given a smooth function *f* defined on a manifold:
  - a point x is called ***regular*** if the differential df_x is surjective
  - a point x is called ***critical*** is the differential df_x is the zero map
  - a critical point x is called ***non-degenerate*** if the Hessian matrix *H* of *f* is non-singular at x
  - if x is a non-degenerate critical point of *f*, then the number λ of negative eigenvalues of *H* is called the ***index*** of x.

Morse functions and critical points

- ✓ In other words:
  - a point *p* is ***critical*** for *f* if:

$$\frac{\partial f}{\partial x_1}(p)=0,\frac{\partial f}{\partial x_2}(p)=0,\ldots,\frac{\partial f}{\partial x_k}(p)=0$$

  - *f* is ***Morse*** at *p* if:

$$\left|H_f(p)\right|=\left|\frac{\partial^2 f}{\partial x_i \partial x_j}(p)\right|\neq 0.$$

- ✓ The definition of critical points is *local* and *sensitive* to small perturbations of the surface.
- ✓ The function *f* is called ***simple*** if any pair x, *y* of distinct critical points verifies $f(x) \neq f(y)$.

Morse functions

- ✓ On any smooth compact manifold there exist Morse functions.
- ✓ Morse functions are everywhere dense in the space of all smooth functions on the manifold.
- ✓ On a compact manifold, any Morse function has only a finite number of critical points .
- ✓ The set *S* of all simple Morse functions is everywhere dense in the set of all Morse functions.
- ✓ Examples are: *height function, distance functions, geodesic distance, etc.*

Morse theory & critical points

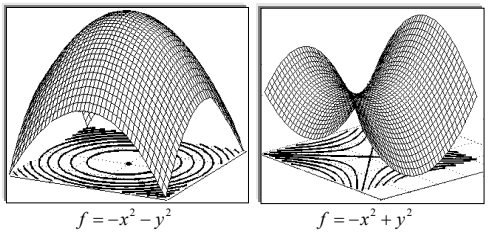
- ✓ **Morse Lemma.**
  In a neighbourhood of each non-degenerate critical point *p*, the function *f* can be expressed as:

$$f = f(p)-(y_1)^2-\ldots-(y_\lambda)^2+(y_{\lambda+1})^2+\ldots(y_n)^2$$

  where λ is the *index* of the critical point.
- ✓ **Euler formula.**

  #maxima - #saddles + #minima= $\chi(S)$.

Morse theory & critical points

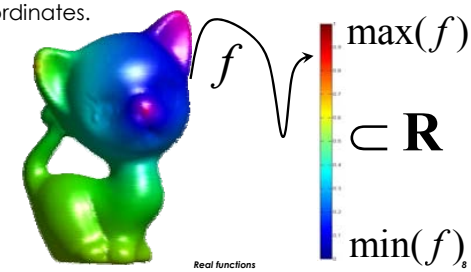


$f = -x^2 - y^2$ — **Maximum λ=2**

$f = -x^2 + y^2$ — **Saddle λ=1**

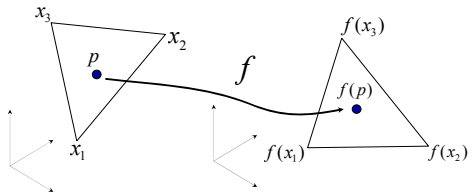
Functions on 3D shapes: discrete case

Define *f* on the mesh vertices and extend *f* to the edges and faces by using barycentric coordinates.



Linear approximation

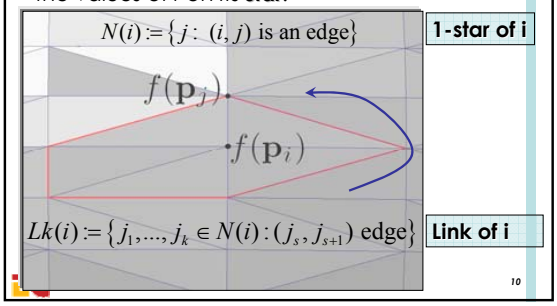
f is uniquely determined by its values on the surface vertices of M.



$$p = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 \Leftrightarrow f(p) = \lambda_1 f(x_1) + \lambda_2 f(x_2) + \lambda_3 f(x_3)$$
$$\lambda_i \geq 0, i = 1,2,3, \lambda_1 + \lambda_2 + \lambda_3 = 1$$

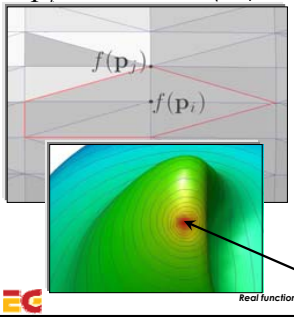
Critical points classification [Ban67]

✓ Each vertex p_i of M is classified according to the values of *f* on its **star**.



Critical points: minimum/maximum

✓ p_i is a **maximum** (resp., minimum) if



$$f(p_i) > f(p_j)$$
$$j \in N(i)$$
$$(\text{resp.}, f(p_i) < f(p_j))$$

$$\gamma_\alpha := \{p \in M : f(p) = \alpha\}$$

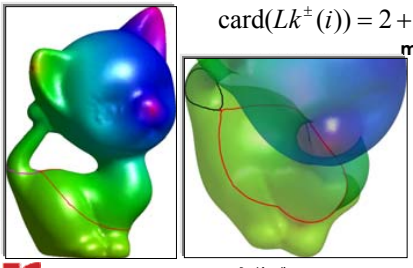
Level sets of f

Critical points: saddle

✓ Let $Lk^{\pm}(i) := \{(j_s, j_{s+1}) \in Lk(i) : f(j_{s+1}) > f(p_i) > f(j_s)\}$ be the mixed link of *i*. Then, p_i is a saddle if

$$\text{card}(Lk^{\pm}(i)) = 2 + 2m, \; m \geq 1$$

multiplicity



Critical point properties: discrete case [Ban67,Ban70]

- If *f* is **general** (ie., *f(x) ≠f(y)*, whenever x and y are distinct vertices of *M*), then the critical points of *(M,f)*
  - satisfy the Euler formula

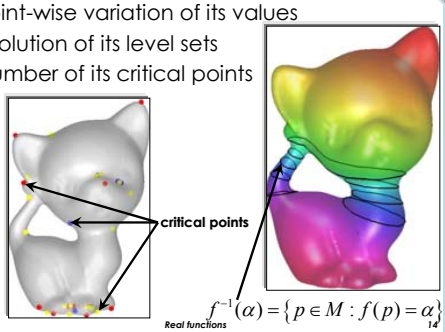
$$\chi(M) = \#\text{minima} - \#\text{saddles} + \#\text{maxima},$$

where saddles are counted with their multiplicity *m*.

  - are located where the topological changes of *(M,f)* happen.

How can we study the behavior of functions on M?

- Point-wise variation of its values
- Evolution of its level sets
- Number of its critical points
- …



Evaluating the properties of *f*

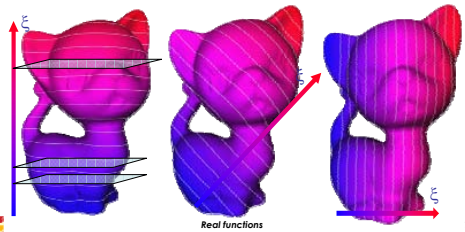
- **Saliency:** ability to measure the shape features we are focusing on.
- **Smoothness:** behavior of *f* wrt the *nature* of its critical points.
- **Stability** wrt discretization and computation.
- **Robustness:** low variation of the *f* values wrt small changes of the shape.
- **DoF and heuristics:** number and type of parameters involved in the definition and/or computation of *f*.
- **Efficiency:** computation cost.
- **Invariance** wrt a group of transformations.

Examples

- Common choices of *f* are:
  - Height and elevation
  - Euclidean-based
  - curvature-based
  - geodesic-based
  - local diameters
  - harmonic functions
  - Laplacian eigenfunctions
  - …

Height function [Ban70,FK97]

- Given a direction ξ, the height function value at $p \in M$ with respect to ξ is defined as $f_\xi(p) := \langle p, \xi \rangle$.
- The level sets correspond to the intersection of the surface with planes orthogonal to the direction ξ.



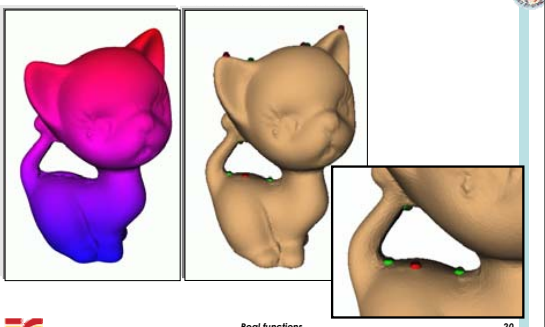
Height function

- **Saliency:** *f* is able to identify the shape features of *M* along the direction ξ.
- **Smoothness:**
  - Critical points are points whose normal is parallel to the direction ξ.
  - In the continuous case, almost all height functions are Morse (ie, the critical points are non degenerate).
- **Stability:** exact evaluation/computation of *f* and interpolation on the faces/edges of *M*.

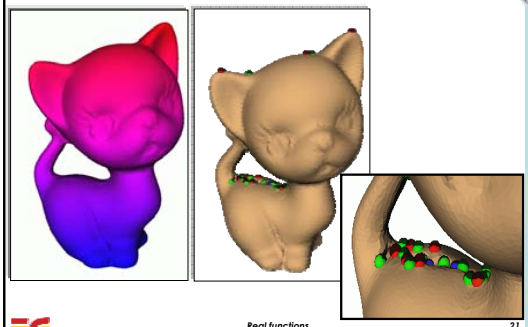
Height function

- ✓ **Robustness:** the computation of *f* is robust, while its critical points aren't.
- ✓ **DoF and heuristic:** the choice of ξ.
- ✓ **Efficiency:** the computational cost is *O(1)*.
- ✓ **Invariance**: the function *f* is
  - invariant to translations
  - dependent on rotations: the recognized properties depend on the chosen direction.

Height function: robustness example

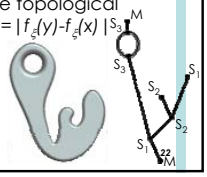


Height function: robustness example



Elevation [AEHW04]

- ✓ For any point x of *M*, there exists at least one direction ξ such that x is a critical point of the height functions f_{ξ} *and* $f_{-\xi}$.
- ✓ Then, for every $\xi \in S^2$
  - let *x, y* be two critical points of the height function wrt the direction ξ,
  - if x, *y* are paired according to the topological persistence, then *pers(x)=pers(y)=* $|f_{\xi}(y)-f_{\xi}(x)|$
  - the elevation is defined as *Elevation(x)=pers(x).*

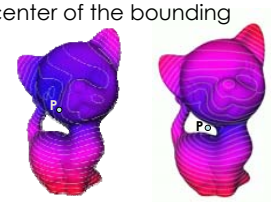


Elevation

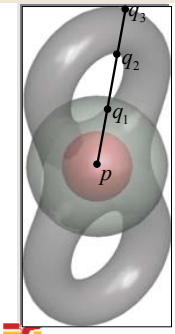
- ✓ **Saliency:** *f* identifies the depression and protrusions of *M* wrt any normal direction.
- ✓ **Smoothness:** *f* is continuous and smooth almost everywhere.
- ✓ **Efficiency:** the overall computational cost for
  - finding the persistence pairs is $O(n \log^2 n)$
  - classifying critical points is $O(n^5 \log^2 n)$.
- ✓ **Invariance:** *f* is invariant to translations and rotations.

Euclidean distance from a point [FK97]

- ✓ $f(x) := \| x - p \|_2$
- ✓ The level sets correspond to the intersection of the surface with a set of spheres centered at the point *p*.
- ✓ Common choices of the point are the barycenter, the center of the bounding sphere, etc.

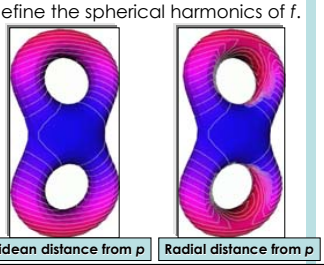


Radial distance from a point [SV01]



$$f(q) := \max\{r \geq 0 : r(q-p) \in M\}$$

In an analogous way, *f* can be defined on the unit sphere and used to define the spherical harmonics of *f*.

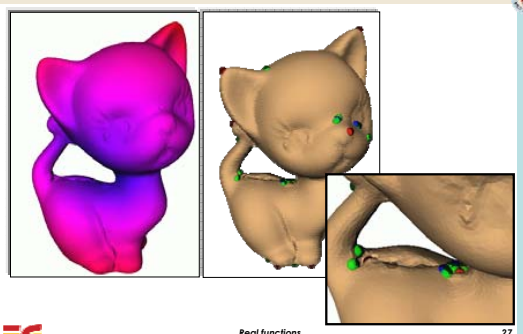


Euclidean distance from *p* | Radial distance from *p*

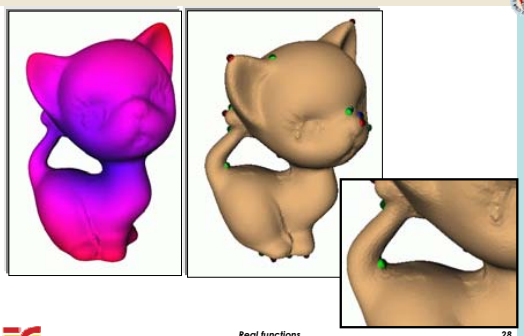
Euclidean distance from a point

- **Saliency:** maxima and minima are located on protrusions and concavities wrt *p*.
- **Smoothness:** in the continuous case, almost all distance functions from a point are Morse.
- **Stability:** in the discrete case, exact computation at the mesh vertices.
- **Robustness:** the computation of *f* is robust, while its critical points aren't.
  - For instance, the distance from the barycenter: due to its dependence on all the vertices, the barycenter is not affected by small perturbations of *M*.

Euclidean distance from a point: robustness example

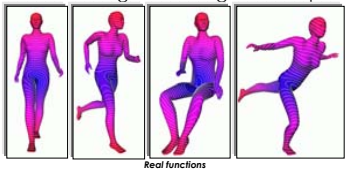


Euclidean distance from a point: robustness example



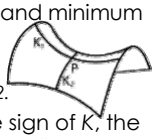
Euclidean distance from a point

- **DoF and heuristics:** the point *p*.
- **Efficiency:** *f* is computed in *O(n)* time.
- **Invariance:**
  - *f* is invariant to translations and rotations of the coordinate system
  - *f* is suitable to distinguish among different poses.

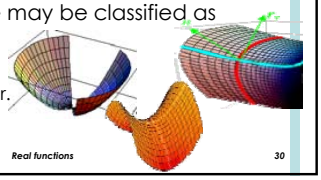


Curvature-based function [GCO06,MPS*04, ZP01]

- The principal curvatures k_1 and k_2 at a point $p \in M$ measure the maximum and minimum bending of a surface at *p*:
  - the Gaussian curvature $K=k_1 k_2$
  - the Mean curvature $H=(k_1+k_2)/2$.



- According to behavior of the sign of *K*, the points of a surface may be classified as
  - elliptic
  - hyperbolic
  - parabolic or planar.

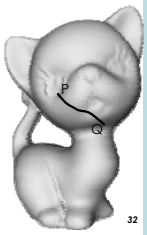


Curvature-based function

- **Saliency:** is provided by the characterization of the local shape as elliptic, hyperbolic, parabolic/planar.
- **Smoothness:** related to the differentiability degree of *M*.
- **Stability:** a coarse surface sampling and an irregular connectivity affect the discretization of the curvature.
- **Robustness:** low degree.
- **DoF and heuristics:** the size of the neighborhood used to compute *K* and *H*.
- **Efficiency:** depends on the size of the neighborhood; at least O(*n*) wrt the 1-star.
- **Invariance:**
  - K is intrinsic, ie it is invariant wrt isometries
  - H is extrinsic and depends on the surface embedding.

Geodesic distance: definition and properties

- Given two points $p,q \in M$, the geodesic distance $g(p,q)$ is the length of the shortest path between *p* and *q*.
- The geodesic distance is invariant to isometric transformations.
- The shortest path is not unique.
- Exact computation in $O(n^2 \log n)$.
- Approximations:
  - Dijkstra [VL99]: $O(n \log n)$
  - [SSK*05]: $O(n \log n)$
  - Fast marching [KS98]: $O(n)$.



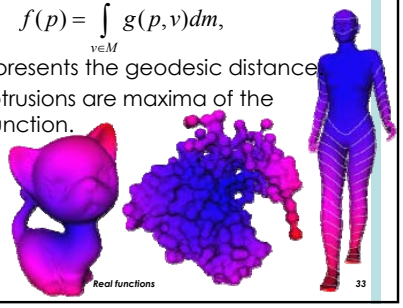
Average geodesic distance [HSKK01]

- The mapping function is defined as

$$f(p) = \int_{v \in M} g(p,v)dm,$$

where *g* represents the geodesic distance

- Surface protrusions are maxima of the mapping function.



Average geodesic function

- Discretized using a set of base points $\{b_1, \ldots, b_n\}$ instead of all mesh vertices:

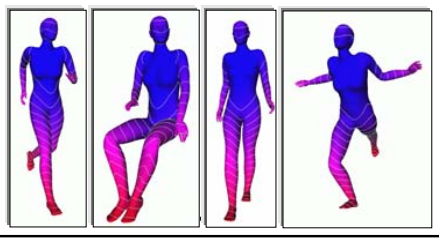
$$f(p) = \sum_i g(p,b_i) area(b_i)$$

where $area(b_i)$ is the influence region of b_i.

- It has been extended to consider also the angle variation along a path [ST03].

Average geodesic distance

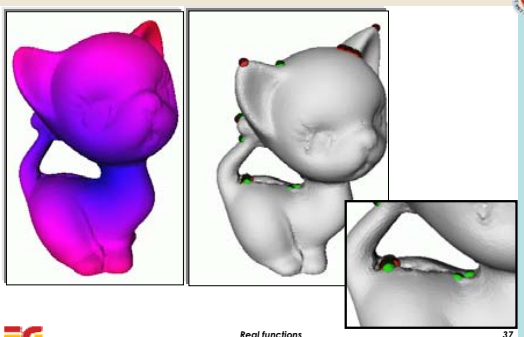
- **Saliency:** *f* discriminates protrusions of *M*.
- **Invariance:** *f* is invariant to isometries, that is, it does not distinguish among different poses of the same surface.



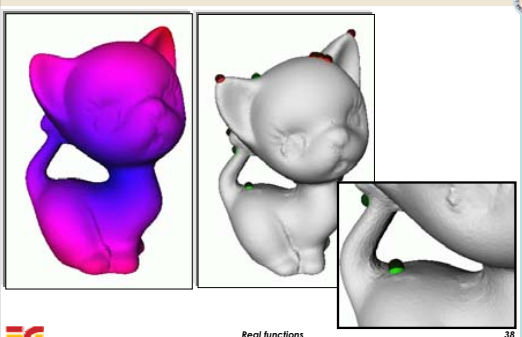
Average geodesic distance

- **Smoothness:** *f* is smooth.
- **Stability:**
  - the discretization and computation depend on the chosen algorithm, eg., Dijkstra [VL99], [SSK*05], fast marching [KS98]
  - generally, a coarse surface sampling and an irregular connectivity affect the discretization of the geodesic distance
  - the instabilities are averaged by the integral in the definition of *f*.
- **Robustness:** *f* is robust to local shape changes.
- **DoF and heuristics:** choice of the base points used to discretize the integral.
- **Efficiency:** depends on the discretization and number of base points. It is computationally expensive using the Dijkstra's algorithm with al vertices as base points: O($n^2 \log n$).

Average geodesic distance: robustness example

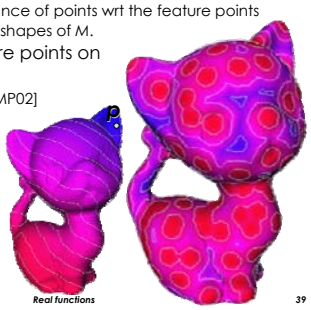


Average geodesic distance: robustness example



Geodesic distance from feature points [MP02,VL99]

- The geodesic distance can be used to
  - measure the importance of points wrt the feature points
  - characterize tubular shapes of *M*.
- Choice of the feature points on the surface:
  - curvature extrema [MP02]
  - user-defined [VL99]
  - or uniform sampling.



Topological distance from curvature extrema [MP02]

- Let *p* be the centroid of a high-curvature region, we define

$$g_p(q) := \min\{k: q \in k\text{-neighborhood}\}.$$

- Given $\{p_1,\ldots, p_k\}$ k feature points, we define g as:

$$g(q) := \min\left\{g_{p_1}(q),\ldots,g_{p_k}(q)\right\}.$$

and

$$f(q) := g_{\max} - g(q).$$

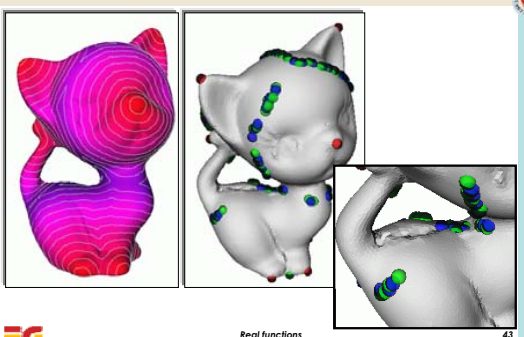
Topological & geodesic distance from curvature extrema

- **Saliency:** *f* discriminates protrusions, especially those that include the curvature extrema as feature points.
- **Smoothness:** low degree.
- **Stability:**
  - topological distance: since *f* is discretized using the connectivity of *M*, the neighborhood expansion is computationally stable
  - geodesic distance: the stability of *f* is affected by the mesh connectivity.

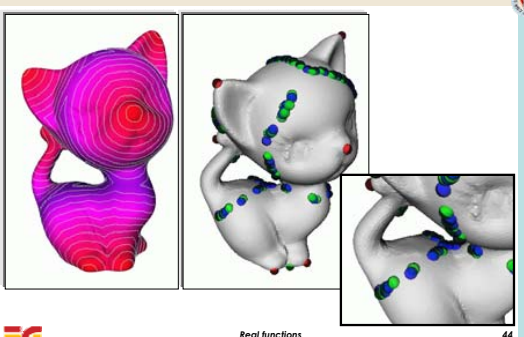
Topological & geodesic distance from curvature extrema

- **Robustness:** the geodesic (resp, topological) distance from feature points is robust wrt small geometric and connectivity (resp, geometric) changes.
- **DoF and heuristics:** choice of the feature points.
- **Efficiency:** the computational cost of the topological expansion is O(*n*) and O(*nlogn*) for the geodesic distance.
- **Invariance:**
  - topological distance: *f* is invariant wrt any transformation that preserves the mesh connectivity
  - geodesic distance: *f* is invariant to isometric transformations.

Topological distance from curvature extrema: robustness example

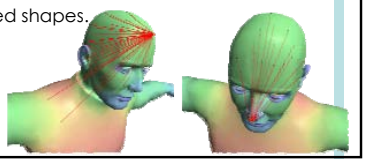


Topological distance from curvature extrema: robustness example



Local diameter shape function [GSC07]

- On a smooth surface, the exact diameter of a shape at a point *p* is the distance to the antipodal point of *p* wrt the direction opposite to the normal at *p*.
- The local diameter function at *p*
  - is a statistical measure of the diameters in a cone around the direction opposite to the normal at *p*.
  - requires closed shapes.



Local diameter shape function

- **Saliency:** morphological characterization of the shape in terms of relative size of its parts.
- **Smoothness:** no guarantees of smoothness for the local shape diameter: it may fail at sites of branching or in particular visibility cones.
- **Stability:** yes.
- **Robustness:** robust to deformations that do not locally alter the shape volume.
- **DoF and heuristics:** no DoF and several heuristics to drive the statistical sampling of the diameters.
- **Efficiency:** $O(n^2)$.
- **Invariance:**
  - invariant to translations and rotations
  - Invariance to pose changes is forced by averaging the values of *f* at the vertices of *M* wrt the values of neighbors.

Harmonic functions [DBG*06]

Smooth functions with a (generally) low number of critical points are achieved by solving the Laplace equation with Dirichlet boundary conditions.

$$\begin{cases} \Delta f(p_i) = 0, i \in I^C & \text{Laplace equation} \\ f(p_i) = \alpha_i, i \in I & \text{Dirichlet boundary conditions} \end{cases}$$

Laplacian matrix of a triangle mesh

$$\Delta f(p_i) = \sum_{j \in N(i)} [f(p_j) - f(p_i)] w_{ij} = 0$$

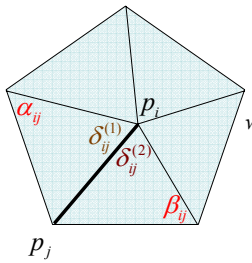
$$L_{ij} = \begin{cases} \sum_{j \in N(i)} w_{ij} & i = j \\ -w_{ij} & (i,j) \text{ edge} \\ 0 & \text{else} \end{cases}$$

$$\bar{L}\bar{f} = b$$

$L \in \mathbf{R}^{n \times n}$, $rank(L) = n-1$

Sub-matrix of *L* | Unknown values of *f* | Know right-hand vector

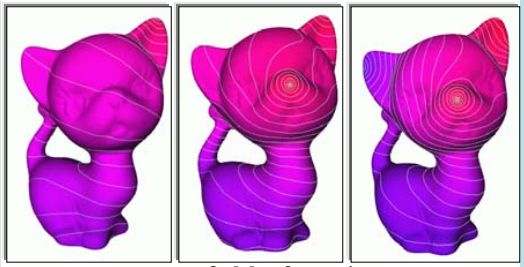
Discretization: weights [Flo97,PP93,…]



$$w_{ij:=}\begin{cases}\dfrac{\cot\alpha_{ij}+\cot\beta_{ij}}{2}\\ \dfrac{\tan(\delta_{ij}^{(1)}/2)+\tan(\delta_{ij}^{(2)}/2)}{\|p_j-p_i\|_2}\\ \dots\end{cases}$$

Harmonic functions

$m=1, M=1, s=2$ $\quad$ $m=3, M=3, s=6$



$m=2, M=2, s=4$

Harmonic functions

- **Saliency:**
  - the choice of the maxima and minima of *f* (→Dirichlet conditions) on feature regions guarantees their characterization through (*M,f*)
  - topological saliency: saddle points are located on the topological handles of *M*, if f has only *1* min and *1* max.
- **Smoothness**: the number of critical points depends on:
  - the number of critical points depends on the Dirichlet boundary conditions and the genus of the input surface
  - using as Dirichlet boundary conditions *1* max & *1* min guarantees to build a harmonic function *f* with a minimal number of critical points (ie., *2g*).
  - $f \in_{\approx} C^2(M,\mathbf{R})$.

Harmonic functions

- **Stability:** the Laplace operator is local and uses only the *1*-star of each vertex. Numerical instabilities might be introduced by its discretization:
  - the cotangent weights might be negative if $\alpha_{ij}+\beta_{ij}>\pi$
  - the mean-value weights are always positive and more stable than the cotangent weights.
- **Robustness:** the computation and the properties of *f* are robust wrt changes of the surface and connectivity that do not make unstable the discretization of the Laplace operator.

Harmonic functions: robustness

- Harmonic scalar fields with the same Dirichlet boundary conditions: different postures of the same shape.

Harmonic functions

- **DoF and heuristics:** the choice of the Dirichlet boundary conditions.
- **Efficiency:**
  - solution of a sparse linear system *O(nlogn)*
  - changing the Dirichlet boundary conditions does not require to re-build the coefficient matrix (→re-use its factorization to solve the same problem with different Dirichlet boundary conditions).
- **Invariance:**
  - *f* is invariant wrt isometries
  - with constant weights, *f* is affine invariant.

Eigensystem of the Laplacian matrix [NGH04,RWP06]

- The spectrum of the Laplacian matrix associated to *M* enables to define a set of functions "intrinsically" defined by the input shape.
- Since *L* is symmetric, it has a real eigensystem

$$Lx_i = \lambda_i x_i, \quad i = 1,\ldots,n$$

and

$$\forall y \in \mathbf{R}^n, \quad y = \sum_{i=1}^{n} \alpha_i x_i.$$

Spectrum of the Laplacian Matrix

- Eigenvalues:

$$0 = \lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$$

- Eigenvectors:

$$(x_i, \lambda_i), \quad Lx_i = \lambda_i x_i$$

- i-th function

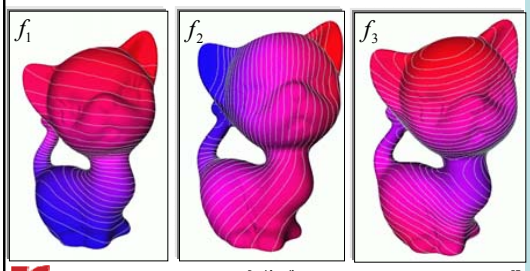
$$f_i : M \to \mathbf{R}$$

$$f_i(p_k) = \sqrt{\lambda_i}[x_i]_k$$

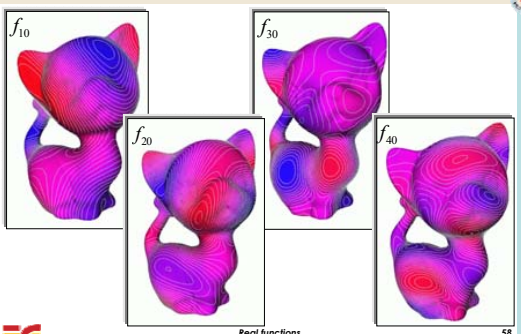
$$i = 2,\ldots,n$$

Laplacian-based functions: examples

- Large set of smooth eigenfunctions with a "generally" low number of critical points.

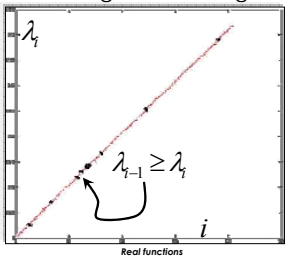


Laplacian-based functions: examples

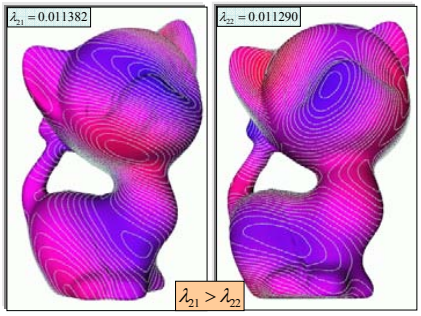


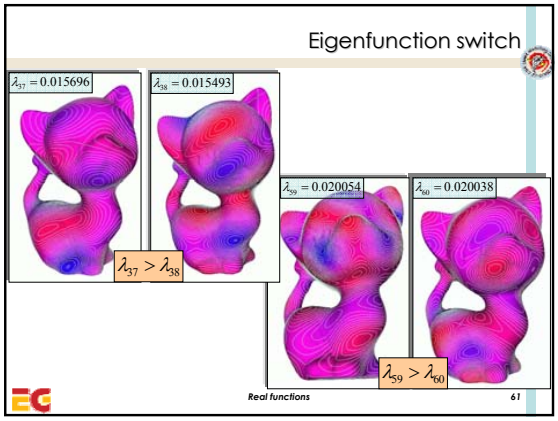
Eigenfunction switch

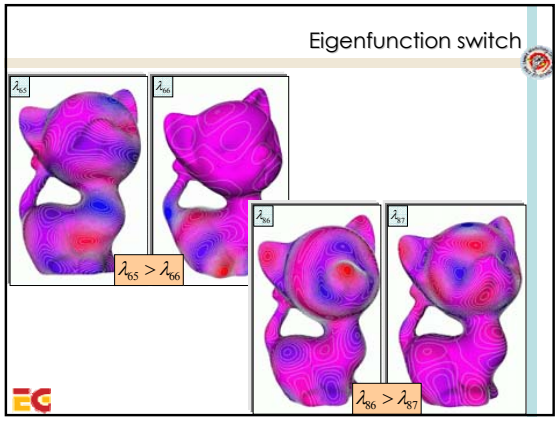
- Generally, the numerical computation of the Laplacian spectrum may switch the order of some eigenvalues/eigenvectors.

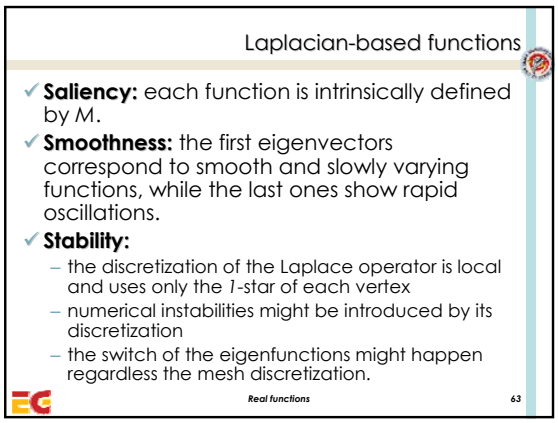


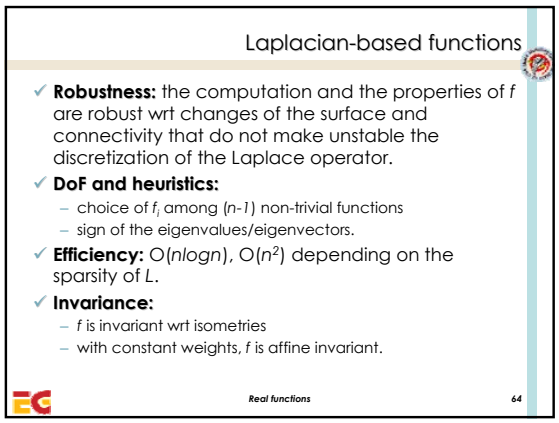
Eigenfunction switch

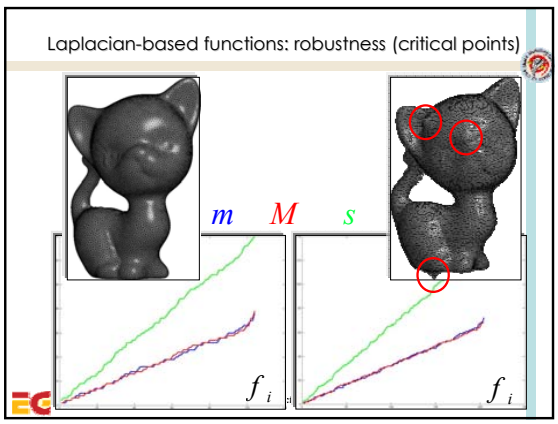


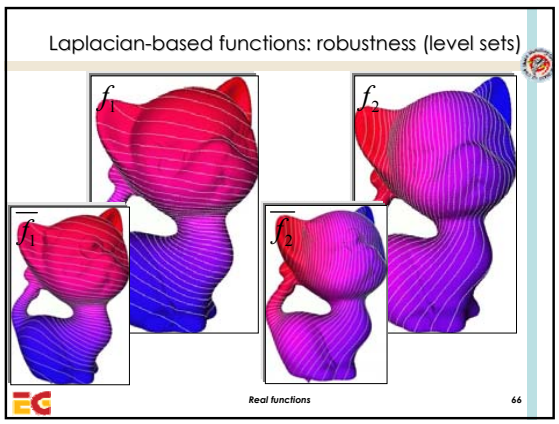

Eigenfunction switch
$\lambda_{37} = 0.015696$
$\lambda_{38} = 0.015493$
$\lambda_{59} = 0.020054$
$\lambda_{60} = 0.020038$
$\lambda_{37} > \lambda_{38}$
$\lambda_{59} > \lambda_{60}$
Real functions
61


Eigenfunction switch
λ_{65}
λ_{66}
λ_{86}
λ_{87}
$\lambda_{65} > \lambda_{66}$
$\lambda_{86} > \lambda_{87}$

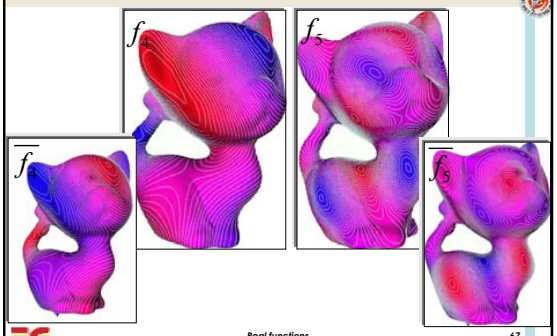

Laplacian-based functions
Saliency: each function is intrinsically defined by M.
Smoothness: the first eigenvectors correspond to smooth and slowly varying functions, while the last ones show rapid oscillations.
Stability:
the discretization of the Laplace operator is local and uses only the 1-star of each vertex
numerical instabilities might be introduced by its discretization
the switch of the eigenfunctions might happen regardless the mesh discretization.
Real functions
63


Laplacian-based functions
Robustness: the computation and the properties of f are robust wrt changes of the surface and connectivity that do not make unstable the discretization of the Laplace operator.
DoF and heuristics:
choice of f_i among (n-1) non-trivial functions
sign of the eigenvalues/eigenvectors.
Efficiency: O(nlogn), O(n^2) depending on the sparsity of L.
Invariance:
f is invariant wrt isometries
with constant weights, f is affine invariant.
Real functions
64


Laplacian-based functions: robustness (critical points)
m
M
s
f_i
f_i


Laplacian-based functions: robustness (level sets)
f_1
f_2
$\overline{f_1}$
$\overline{f_2}$
Real functions
66

Laplacian-based functions: robustness (level sets)



Questions?

Eurographics 2007 Tutorial T12

Appendix: Perturbation Theory for Eigenvalues and Eigenvectors [GV89]

Eurographics 2007 Tutorial T12

Speaker
Giuseppe Patanè
CNR-IMATI-GE - Italy

Perturbation theory: general case

- Right eigenvector $Ax = \lambda x$
- Left eigenvector $y^* A = \lambda y^*$
- If *A* is diagonalizable, $y_i^* x_j = 0, i \neq j$
- Consider the matrix

$$A_\varepsilon := A + \varepsilon B, \qquad |b_{ij}| \leq 1$$

with right eigenvector $x_i(\varepsilon)$ and eigenvalue $\lambda_i(\varepsilon)$.
Pb: Which relation exists between the eigensystem of A, A_ε ?

Eigenvalue perturbation: general case

- For each eigenvalue, the following relation holds

$$\left|\frac{\lambda_i(\varepsilon) - \lambda_i}{\varepsilon}\right| \to \frac{\|B\|_2}{s(\lambda_i)}, \qquad s(\lambda_i) := |y_i^* x_i|.$$

- Then, the above estimation is:
  - proportional to the l_2-conditioning number of the perturbation matrix *B*
  - inversely proportional to the angle between the left and right eigenvectors.

Eigenvalue perturbation: general case

- We note that

$$s(\lambda_i) = \underbrace{|y_i^* x_i|}_{\geq 0} \leq \|y_i\|_2 \, \|x_i\|_2 = 1.$$

- The term $s(\lambda_i)^{-1}$ is called **conditioning number** of the eigenvalue λ_i.
- Then, an eigenvalue is **well-conditioned** iff its conditioning number is not close to zero.

Eigenvalue perturbation: Laplacian matrix

- If the input surface is closed (or with boundary + virtual edges), the Laplacian matrix is symmetric and

$$y_i \equiv x_i, \quad s(\lambda_i) = 1, \quad i = 1, \ldots, n.$$

- Each eigenvalue is well-conditioned and

$$\left| \frac{\lambda_i(\varepsilon) - \lambda_i}{\varepsilon} \right| \to \| B \|_2, \quad \varepsilon \to 0.$$

- The variation of the eigenvalues depends only on the l_2-norm of the perturbation matrix *B*.

Eigenvector perturbation: general case

- For the *i*-th eigenvector, we have

$$\| x_i(\varepsilon) - x_i \|_2 \le \varepsilon \sum_{j \ne i}^{n} \left| \frac{y_j^* B x_i}{(\lambda_i - \lambda_j) s(\lambda_j)} \right| + \mathrm{O}(\varepsilon^2).$$

- Then, the bound depends on:
  - the conditioning number of **each** eigenvalue $s(\lambda_i)$
  - the differences $\lambda_i - \lambda_j$
  - the factors $\beta_{ij} := y_j^* B x_i.$

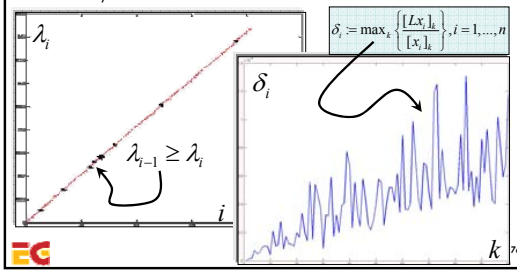
Eigenvector perturbation: general case

- The perturbation in the eigenvector is proportional to the conditioning number of the whole set of eigenvalues.
- If the eigenvalues are close to one another, we may have difficulties in computing the eigenvectors.
- Let *A* have distinct eigenvalues. If for some eigenvalue $s(\lambda) < 1$, then there exists a matrix *E* such that λ is a repeated eigenvalue of (*A*+*E*) and

$$\frac{\|E\|_2}{\|A\|_2} \le \frac{s(\lambda)}{\sqrt{1 - s(\lambda)^2}}.$$

The perversity theorem does not hold

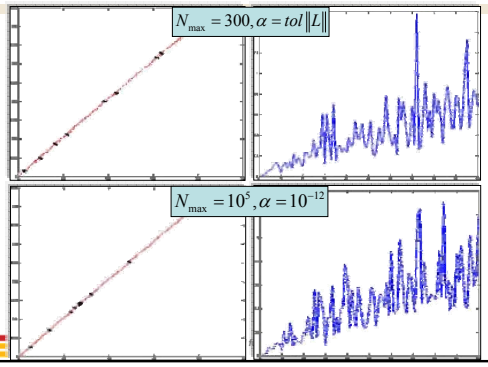
- Then, even if the eigenvalues are distinct, if one eigenvalue is ill-conditioned, the computation of the eigenvalues, and especially the eigenvectors, may be very difficult.



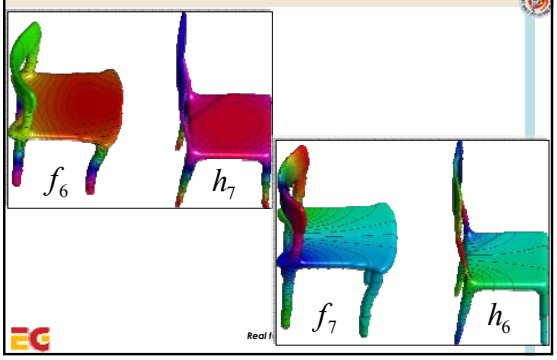
Jacobi iterations and stop criteria

- The (first or last) elements of the eigensystem of the input matrix are evaluated by using the Jacobi method with 2 stop criteria:
  - max. number of iteration $N_{\max}$
  - approximation threshold α.
- Increasing $N_{\max}$ and reducing α do not avoid the switching of eigenvalues and eigenvectors.

Jacobi iterations and stop criteria



Eigenfunction "switch" on different shapes



Eigenfunction switch: discussion

- The switch of the eigenfunctions
  - can happen among the eigenfunctions of the same surface;
  - is strictly correlated to the computation of the eigensystem;
  - a "good" geometry and connectivity (wrt the computation of the entries of *L*) do not guarantee to avoid the switch of the eigenfunctions.

Questions?

Eurographics 2007 Tutorial T12

3D Shape Description and Matching Based on Properties of Real Functions

Shape Descriptors

Eurographics 2007 Tutorial T12

Speaker
Daniela Giorgi
CNR-IMATI-GE - Italy

Outline

- ✓ Descriptors parametric with respect to *f:*
  - Reeb graph
  - Size theory tools
  - Persistent homology tools
  - Descriptors based on spherical decompositions
    - Spherical harmonics
    - Spherical wavelets
- ✓ Descriptors linked to a specific *f:*
  - Shape-DNA
  - Bending invariant signature
  - Spectral embedding
  - Pose-oblivious shape signature
  - Salient geometric features

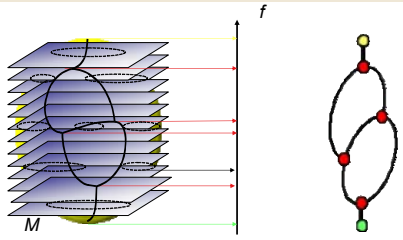
Properties to be discussed

- ✓ **Saliency**: ability to capture the essential features of the shape
- ✓ **Conciseness**: ability to minimize the memory needed to store the descriptors while maximizing the amount of information
- ✓ **Robustness** wrt small changes of the shape
- ✓ **Uniqueness** and **completeness**
- ✓ **Invariance** to transformation groups
- ✓ **DoF and Heuristics** used in the construction of the descriptor
- ✓ **Input:** hypotesis and restrictions
- ✓ **Efficiency**: computational cost

Outline

- ✓ Descriptors parametric with respect to *f:*
  - Reeb graph
  - Size theory tools
  - Persistent homology tools
  - Descriptors based on spherical decompositions
    - Spherical harmonics
    - Spherical wavelets
- ✓ Descriptors linked to a specific *f:*
  - Shape-DNA
  - Bending invariant signature
  - Spectral embedding
  - Pose-oblivious shape signature
  - Pose-oblivious shape signature
  - Salient geometric features

Reeb graph [Ree46]



Reeb graphs are used to store the evolution of the level sets of the mapping function ***f***

Reeb graph [Ree46]

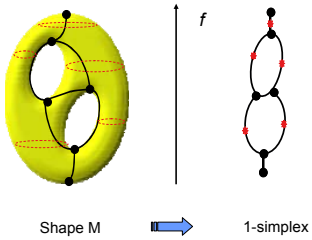
Let M be a compact 2-manifold and *f*: M→R a simple Morse function;

Let "~" be the equivalence relation:

(P, f(P)) ~ (Q, f(Q)) ⇔ *f(P) = f(Q)* and *P* and Q are in the same connected component of $f^{-1}(f(P))$

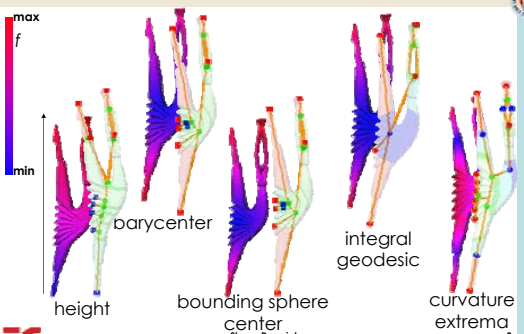
The quotient space on MxR is a finite and connected simplicial complex K of dimension 1, such that the counter-image of each vertex of dim 0 of K is a singular connected component of the level sets of f, and the counter-image of the interior of each simplex of dim 1 is homeomorphic to the topological product of one connected component of the level sets by R

Reeb graph



The 1-simplex is often associated to a geometric embedding (*centerline skeleton*), or used to store additional geometric data

RGs when the function *f* varies



Reeb graph based representations

- Different proposals for descriptors induced by the Reeb graph:
  - Multiresolution Reeb graph (MRG) [HSKK01, BSR06]
  - Augmented Multiresolution Reeb graph (aMRG) [TS05]
  - Extended Reeb graph (ERG) [BFS00,Bia04,BMSF06]

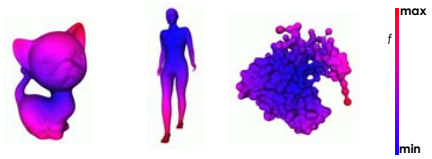
Multiresolution Reeb graph [HSKK01]

- It is defined on the basis of the function:

$$f(p)=\int_{v \text{ in } S} g(p,v)dS$$

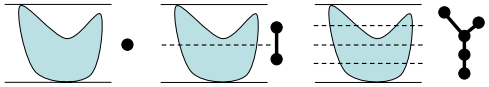
where *g* represents the geodesic distance

- Surface protrusions are maxima of the function *f*

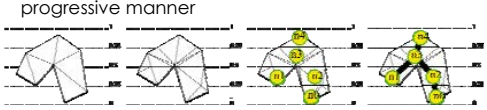


Multiresolution Reeb graph [HSKK01]

- Provides a hierarchical graph encoding



- The graph is extracted inserting contours in a progressive manner



- The area *A* of a region and the relative size *L* of the interval of *f* are associated as attributes to nodes

Augmented Multiresolution Reeb graph [TS05]

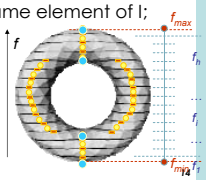
- The descriptor is the same but the nodes are enriched with attributes storing also other geometric measures related to the spatial extent of the regions associated to the nodes
  - Relative volume
  - Statistic measure of the chords
  - Koenderink shape index
  - Statistic orientation of the triangle normals
  - Statistic on the texture (when available)

Multiresolution Reeb graphs [HSKK01,TS05]

- **Saliency**: captures protrusions of the shape
- **Conciseness**: very good conciseness properties due to the synthesis of the information (geometry and topology) in an attributed graph
- **Robustness**: no theoretical results
- **Uniqueness**: fixed the resolution, the MRG is unique
- **Completeness**: no
- **Invariance**: inherited by *f*
- **DoF and heuristics**: the resolution has to chosen
- **Input**: manifold, closed triangle meshes
- **Efficiency**: the cost of the graph extraction is O(*n*+*k*), where *k* is the number of vertices added during the construction

Extended Reeb Graph [Bia04]

- Founds on an extended Reeb equivalence
  - let f:M→R be a real valued function;
  - let I={(f_{min}, f_1),(f_h,f_{max}),(f_i,f_{i+1}),i=1…h-1} ∪ {f_{min},f_1,…f_h, f_{max}} be a partition of [f_{min}, f_{max}];
  - the an extended Reeb equivalence between *P*, *Q* ∈*M* is given by:
    - *f(P)*, *f(Q)* belong to the same element of I;
    - *P*, *Q* belong to the same connected component of $f^{-1}(f(t))$, $t \in I$.



Geometric embedding of the ERG [BMSF06]

- Each arc can be oriented using the growing direction of the mapping function: the ERG is a direct acyclic graph
- Store with each ERG node *n* attributes measuring properties of regions or subparts associated to *n* (eg, using spherical harmonics)
- Store for each ERG arc *e* the number of slices traversed by the arc (arc length)

Extended Reeb Graph [Bia04]

- **Saliency**:
  - captures the structure among the features characterized by the critical points of the mapping function
  - the geometric embedding influences the saliency
  - it preserves the topology of the manifold
- **Conciseness**: very good conciseness properties due to the synthesis of the information (geometry and topology) in an attributed graph
- **Robustness**: no theoretical results
- **Uniqueness:** fixed the partition, the ERG is unique
- **Completeness:** no

Extended Reeb Graph [Bia04]

- **Invariance**: inherited by *f*
- **DoF and heuristics**: the partion has to be chosen
- **Input:** orientable 2-manifold represented by triangle meshes
- **Efficiency**: the cost of the graph extraction is *O(n+k)* (where *k* is the number of vertices added during the construction)

Outline

- Descriptors parametric with respect to *f*:
  - Reeb graph
  - Size theory tools
  - Persistent homology tools
  - Descriptors based on spherical decompositions
    - Spherical harmonics
    - Spherical wavelets
- Descriptors linked to a specific *f*:
  - Shape-DNA
  - Bending invariant signature
  - Spectral embedding
  - Pose-oblivious shape signature
  - Salient geometric features

Size Theory and Size Functions [Fro90]

- Size Theory proposes an approach where Shapes are toplogical spaces endowed with real functions, and comparing shapes means comparing the properties expressed by the real functions
- If two shapes are similar, a homeomorphism between the shapes almost preserving the function values must exist
- How can we measure how well a homeomorphism can preserve the values taken by the considered function?
- In Size Theory two shapes are similar if their the natural pseudo-distance is small

Size Theory and Size Functions [Fro90]

- M,N topological spaces

H a (subset of) the set of all homeomorphisms

$\gamma : M \to N$

Consider two continuous measuring functions

$f : M \to \Re^k, \quad g : N \to \Re^k$

Define the natural pseudo-distance

$$\Theta(\gamma) = \max_{P \in M} \| f(P) - g(\gamma(P)) \|_\infty$$

$$d((M,f),(N,g)) = \begin{cases} \inf_{\gamma \in H} \Theta(\gamma) & \text{if } H \text{ is not empty} \\ +\infty & \text{if } H \text{ is empty} \end{cases}$$

Size Theory and Size Functions [Fro90]

- The natural pseudo-distance is a powerful tool to compare shapes, but difficult to compute (we have to study all the homeomorphisms between two spaces)
- We need a tool to study the natural pseudo-distance
- We can get information from size functions, a mathematical tool providing a lower bound for the natural pseudo-distance

(Multidimensional) Size Functions [FM99,FL99,BCF*07]

$(\mathcal{M}, \vec{\varphi})$ size pair, with $\vec{\varphi} : \mathcal{M} \to \mathbb{R}^k$;

for every $\vec{x} = (x_1, \ldots, x_k)$, $\vec{y} = (y_1, \ldots, y_k) \in \mathbb{R}^k$

- $\vec{x} \preceq \vec{y}$ $(\vec{x} \prec \vec{y}) \overset{def}{\Longleftrightarrow} x_l \leq y_l$ $(x_l < y_l)$, $i = 1, \ldots, k$;
- $\mathcal{M}_{\vec{x}} = \{P \in \mathcal{M} : \vec{\varphi}(P) \preceq \vec{x}\}$;

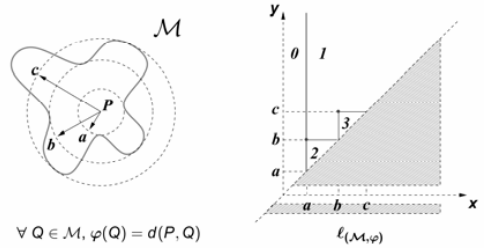
$\Delta^+ = \{(\vec{x}, \vec{y}) \in \mathbb{R}^k \times \mathbb{R}^k : \vec{x} \prec \vec{y}\}$;

Definition

The (multidimensional) size function of the size pair $(\mathcal{M}, \vec{\varphi})$ is the function $\ell_{(\mathcal{M},\vec{\varphi})} : \Delta^+ \to \mathbb{N}$ that takes each $(\vec{x}, \vec{y})$ to the number of connected components of $\mathcal{M}_{\vec{y}}$ that contain at least a point of $\mathcal{M}_{\vec{x}}$.

Example with k=1

$\ell_{(M,\varphi)}(x, y) = \#\{\text{connected components under } y \text{ with a point under } x\}$



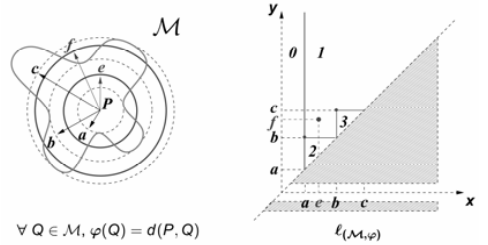
$\forall\, Q \in \mathcal{M},\ \varphi(Q) = d(P, Q)$

$\ell_{(\mathcal{M},\varphi)}$

(Symbols $\vec{x}, \vec{y}, \vec{\varphi}$ are replaced by x, y, φ).

Example with k=1

$\ell_{(M,\varphi)}(x, y) = \#\{\text{connected components under } y \text{ with a point under } x\}$



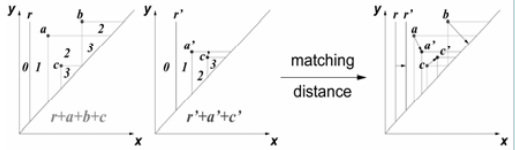
$\forall\, Q \in \mathcal{M},\ \varphi(Q) = d(P, Q)$

$\ell_{(\mathcal{M},\varphi)}$

(Symbols $\vec{x}, \vec{y}, \vec{\varphi}$ are replaced by x, y, φ).

K=1: Representation and matching [FL01,dAFL06]

Each 1-dimensional size function can be represented by a formal series of points representing vertices of triangular region in Δ^+.



K=1: Stability of matching distance [dAFL06]

- Matching Stability Theorem:
  The matching distance satisfy the following stability condition:

$$\max_{P\in\mathcal{M}} |\varphi(P) - \psi(P)| \le \epsilon \Rightarrow d_{match}(\ell_{(\mathcal{M},\varphi)}, \ell_{(\mathcal{M},\psi)}) \le \epsilon.$$

- Lower bound for the natural pseudo-distance:
  Let δ be the matching distance between the two size functions $\ell_{(M,\varphi)}$ e $\ell_{(N,\psi)}$. Then

$$d((M,\varphi),(N,\psi)) \ge \delta.$$

Multidimensional size functions [BCF*07]

- PROBLEMS WITH k>1:
  - How to extend the representation as formal series of points and lines?
  - How to compare size functions with k>1? A direct approach involves working in a domain of $\mathfrak{R}^{2k}$
  - How to obtain stability in computation?
- SOLUTION: there exists a foliation in half-planes of the domain of multidimensional size functions s.t. on each leaf of the foliation the multidimensional size function coincides with a particular 1-dimensional size function; this allow to difine a stable multidimensional matching distance

Multidimensional size functions [BCF*07]

- Define a foliation of $\Delta^+ = \{(\vec{x},\vec{y}) \in \mathbb{R}^k \times \mathbb{R}^k : \vec{x} \prec \vec{y}\}$ by a parameterized family of half-planes:

$$\pi_{(\vec{l},\vec{b})} : \vec{x} = s\vec{l} + \vec{b},\ \vec{y} = t\vec{l} + \vec{b},\ s,t \in \mathbb{R}, s < t$$

with
$\vec{l} = (l_1,\dots,l_k),\ \vec{b} = (b_1,\dots,b_k) \in \mathbb{R}^k,\ \|\vec{l}\| = 1,\ l_i > 0,\ \sum b_i = 0.$

- Given a size pair $(\mathcal{M},\vec{\varphi}),\ \vec{\varphi} : \mathcal{M} \to \mathbb{R}^k$, define a 1D-measuring function for each $(\vec{l},\vec{b})$:

$$F^{\vec{\varphi}}_{(\vec{l},\vec{b})}(P) = \max_{i=1,\dots,k} \left\{ \frac{\varphi_i(P) - b_i}{l_i} \right\}.$$

Reduction Theorem

For every $(\vec{x},\vec{y}) = (s\vec{l} + \vec{b}, t\vec{l} + \vec{b}) \in \pi_{(\vec{l},\vec{b})}$, it holds that

$$\ell_{(\mathcal{M},\vec{\varphi})}(\vec{x},\vec{y}) = \ell_{(\mathcal{M},F^{\vec{\varphi}}_{(\vec{l},\vec{b})})}(s,t).$$

Multidimensional size functions [BCF*07]

- on each leaf of the foliation, size functions can be represented by formal series of points and lines;
- the induced 1D matching distance on each leaf of the foliation is stable wrt small changes of the leaves;
- a multidimensional matching distance can be defined

$$D_{match}\left(\ell_{(\mathcal{M},\vec{\varphi})}, \ell_{(\mathcal{N},\vec{\psi})}\right) = \sup_{(\vec{l},\vec{b})} \min_{i=1,\dots,k} l_i \cdot d_{match}(\ell_{(\mathcal{M},F^{\vec{\varphi}}_{(\vec{l},\vec{b})})}, \ell_{(\mathcal{N},F^{\vec{\psi}}_{(\vec{l},\vec{b})})})$$

- theorems about the stability of the matching distance and the lower bound for the natural pseudo distance can be stated also in the case k>1

Size Functions

- **Saliency**:
  - captures the structure among the features characterized by the critical points of the mapping function
  - connection with the comparison of topological spaces in terms of natural pseudo-distance
- **Conciseness**: very concise combinatorial description
- **Robustness**: theoretically proven robustness wrt small shape changes
- **Unique** but not **complete**

Size Functions

- ✓ **Invariance**: inherits invariance properties from the underlying measuring functions
- ✓ **DoF** and **heuristics**: none
- ✓ **Input**: representation of shapes as size graphs (from discrete or discretized objects)
- ✓ **Efficiency**: the computational complexity is $O(n \log n + m\alpha(2m+n,n))$

Outline

- ✓ Descriptors parametric with respect to *f:*
  - Reeb graph
  - Size theory tools
  - Persistent homology tools
  - Descriptors based on spherical decompositions
    - Spherical harmonics
    - Spherical wavelets
- ✓ Descriptors linked to a specific *f:*
  - Shape-DNA
  - Bending invariant signature
  - Spectral embedding
  - Pose-oblivious shape signature
  - Salient geometric features

Persistent Homology [ELZ02]

✓ The idea of Persistent Homology is to control the placement of topological events in a growing space and assess their relevance according to their life-time

✓ Given a growing complex **K**, represented by a filtration

$$\{K^i\}_{i=\dots,n},\ K^n = K,\ K^i \text{subcomplex of } K^{i+1}$$

the ***j***-persistent ***k***-th homology group of K^i is a group isomorphic to the image of the homomorphism

$$\eta_{ij}^{\ k} : H_k(K^i) \to H_k(K^{i+j})$$

induced by the inclusion of K^i into K^{i+j}

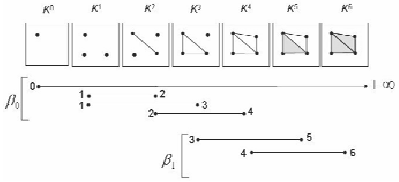
✓ Persistence represents the life-time of cycles in the growing filtration

Persistent Homology [ELZ02]

✓ The persistent homology of a growing complex can be represented by a set of intervals, called persistence intervals: a $\mathscr{P}$- interval is a pair

$$(i,j),\quad i,j \in \mathbb{Z} \cup +\infty,\ 0 \le i < j$$

such that there exists a cycle that is completed at the level ***i*** of the filtration and remains non-bounding until the level ***j***



Persistence Homology and Barcodes [CGCZ05]

✓ The shape of a complex *K* can be described by filtering the complex by the increasing values of a real function

✓ Idea: construct a new complex strictly related to *K*, namely the tangent complex *T(K)* (closure of the space of all tangents to all points in K), and filter it with the function computing the curvature at a point along a tangent direction

✓ The barcode of the shape is the set of $\mathscr{P}$- intervals for the filtered tangent complex

Barcodes and Persistence Diagrams [CSEH07]

✓ More recently $\mathscr{P}$- intervals have been described as point sets in the extended plane, and named persistence diagrams: barcodes are essentially a different representation of persistence diagrams for the tangent complex with the curvature function

✓ The Bottleneck Stability Theorem has been proved:
Let X be a triangulable space with continuous tame functions

$$f, g : X \to \Re$$

Then the persistence diagrams $D(f)$, $D(g)$ satisfy

$$d_B(D(f), D(g)) \le \|f - g\|_\infty$$

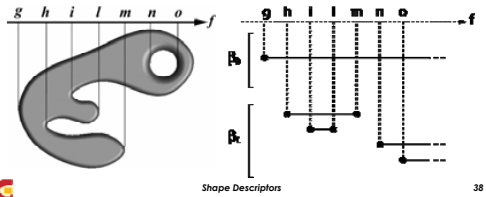
with d_B the Bottleneck distance (also true for the Hausdorff distance)

Persistent homology [ELZ02]

- Recent research directions:
  - Vines and Vineyards [CSEM06]
  - Multidimensional Persistence [CZ07a]
  - Localized Homology [CZ07b]
  - Persistence Intervals [DW07]
  - Extended Persistence[CSEH07]

Persistence barcodes and diagrams [ELZ02,CGCZ05]

- **Saliency**:
  - captures the structure among the features characterized by the critical points of the mapping function
  - the authors have confined themselves to compute barcodes using a specific space and a specific function, but the underlying theory is valid for more general classes of shapes and functions



Persistence barcodes and diagrams [ELZ02,CGCZ05]

- **Conciseness**: very concise combinatorial description
- **Robustness**: theoretically proven robustness wrt small shape changes
- **Unique** but not **complete**
- **Invariance**: inherits invariance properties from the underlying measuring functions
- **DoF** and **heuristics**: none
- **Input**: Barcodes computed on curve PCD and mathematical surfaces, but triangulations and a more general input are admissible
- **Efficiency**: Computing persistent homology requires at most $O(m^3)$, with m the number of simplices

Outline

- Descriptors parametric with respect to *f*:
  - Reeb graph
  - Size theory tools
  - Persistent homology tools
  - Descriptors based on spherical decompositions
    - Spherical harmonics
    - Spherical wavelets
- Descriptors linked to a specific *f*:
  - Shape-DNA
  - Bending invariant signature
  - Spectral embedding
  - Pose-oblivious shape signature
  - Salient geometric features

Spherical Harmonics [VSR01]

- Idea: build multi-resolution feature vectors using the Fourier expansion of a function defined on the sphere
- Represent the spherical function $f: S^2 \rightarrow \mathbf{R}$ (eg. the spherical extent function, measuring the extent of the object in given directions) as

$$f(\theta,\varphi) = \sum_{l\geq 0} \sum_{|m|\leq l} a_{l,m} Y_l^m(\theta,\varphi)$$

- Feature vectors can be extracted from the first rows of coefficients, thereby providing a multiresolution approach

Spherical Harmonics [KFR03]

- Represent a function *f* defined on the sphere through its spherical harmonics and consider the vector of energies (i.e. frequency norms)

$$SH(f) = \left\{ \left\| f_0(\theta,\varphi) \right\|, \left\| f_1(\theta,\varphi) \right\|, \ldots \right\}$$

with f_l the frequency components

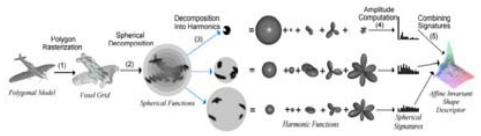
$$f_l(\theta,\varphi) = \sum_{m=-l}^{l} a_{lm} Y_l^m(\theta,\varphi)$$

- Let R be a rotation; then it holds:

$$SH(R(f)) = SH(f)$$

Spherical Harmonics [KFR03]

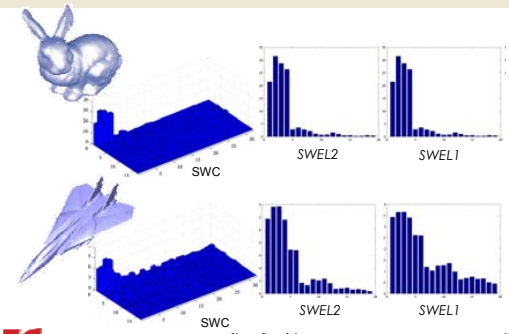
- Extension to voxel description:
  - Restrict the voxel grid to a collection of concentric spheres
  - Represent each spherical restriction in terms of the energy of its frequency decomposition, thus obtaining a 1D descriptor
  - The final descriptor resulting from the analysis of spheres with different radii is a 2D grid indexed by radius and frequency



Spherical Wavelets [LTN06]

- The problem of the sensitivity of the sampling of the spherical function to latitude-longitude parametrization of the sphere is addressed
- A rotation invariant sampling is proposed, relying on the flat octahedron parametrization of the sphere
- A Spherical Wavelet Transform is applied to the spherical shape function
- Resulting descriptors:
  - Matrix of wavelet coefficients (SWC)
  - L1 energy-based feature vector (SWEL1)
  - L2 energy-based feature vector (SWEL2)

Spherical wavelet descriptors [LTN06]



Spherical Harmonics and Wavelets [VSR01,KTR03,LTN06]

- **Saliency**: captures the geometrical properties expressed by the spherical function
- **Conciseness**: very concise descriptors (feature vectors or matrices)
- **Robustness**: robustness wrt small changes of the spherical function derived from the decomposition properties
- **Unique**, but not **complete**, since a finite number of coefficients or energies is taken into account
- **Invariance**:
  - Wrt translation
  - Wrt rotation: [VSR01] requires alignment, [KTR03] is invariant only to rotations applied to the sampled input, in [LTN06] SWC requires alignment, while SWEL1 and SWEL2 are rotation invariant

Spherical Harmonics and Wavelets [VSR01,KTR03,LTN06]

- **DoF** and **heuristics**:
  - Sampling and voxelization
  - number of frequency components
  - [LTN06] requires the choice of the wavelet transform
- **Input**: meshes (also poligon soups) and grids
- **Efficiency**: the computational complexity in [KTR03] is $O(n^3)$ with n size of the voxel grid

Outline

- Descriptors parametric with respect to *f*:
  - Reeb graph
  - Size theory tools
  - Persistent homology tools
  - Descriptors based on spherical decompositions
    - Spherical harmonics
    - Spherical wavelets
- Descriptors linked to a specific *f*:
  - Shape-DNA
  - Bending invariant signature
  - Spectral embedding
  - Pose-oblivious shape signature
  - Salient geometric features

Shape DNA [RWP06]

- The shape DNA is the beginning of the spectrum of the Laplace – Beltrami operator, defined for real valued functions on Riemannian manifolds:

Given a Riemannian *n*- manifold M and $f : M \to \Re$ the Laplace – Beltrami operator is

$$\Delta f := div(grad\, f)$$

(different from the discrete Laplacian on graphs)

Shape DNA = $\{\lambda_0 \le \lambda_1 \le ... \le \lambda_m\} \in \Re^m_{\ge 0}$

with λ_i eigenvalues of the Helmholtz equation $\Delta f = -\lambda f$

Shape DNA [RWP06]

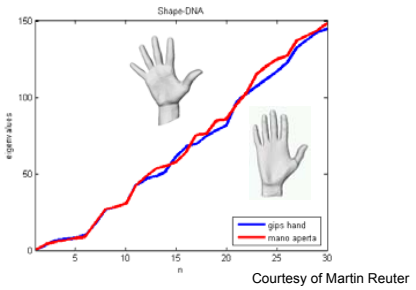
- Numerical computation:

Translate the Helmotz equation into a variational problem and employ the Finite Elements Method (FEM) with m form functions, leading to the generalized eigenvalue problem:

$$AU = \lambda BU$$

with A, B sparse positive (semi-)definite symmetric matrices and $U = (U_1, U_2, ..., U_m)$ eigenfunctions with corresponding eigenvalues λ

Shape DNA [RWP06]



Courtesy of Martin Reuter

Shape DNA [RWP06]

- **Saliency**:
  - description of geometrical and topological properties intrinsic to the object
  - scalable amount of captured shape information, related to the dimension of the cropped spectrum
- **Robustness**: continuously dependent on shape deformations
- **Unique** but not **complete**, since there exist isospectral but not isometrical shapes
- **Invariance** wrt isometry

Shape DNA [RWP06]

- **DoF and heuristics**:
  - choice of the number of eigenvalues in the cropped spectrum
  - choice of the form functions for FEM
- **Input**: parametric surfaces, polygonal meshes, solid polyhedra, but conversion to a dataform supported by the FEM engine is required; independent w.r.t. parametrization
- **Efficiency**: the eigenvalue computation is the most time consuming step

Outline

- Descriptors parametric with respect to *f*:
  - Reeb graph
  - Size theory tools
  - Persistent homology tools
  - Descriptors based on spherical decompositions
    - Spherical harmonics
    - Spherical wavelets
- Descriptors linked to a specific *f*:
  - Shape-DNA
  - Bending invariant signature
  - Spectral embedding
  - Pose-oblivious shape signature
  - Salient geometric features

Bending Invariant Surface Signatures [EK03]

- Geodesic distances between surface points are invariant to surface bending
- Idea: use geodesic distances to define an isometrical embedding of a surface in a small dimensional Euclidean space, in which geodesic distances are approximated by Euclidean ones
- Method: apply a MultiDimensional Scaling (MDS) procedure on a geodesic distance matrix, with geodesics computed via the Fast Marching on Triangulated Domains (FMTD) algorithm

Bending Invariant Surface Signatures [EK03]

- Sample with *n* vertices a given triangulated surface, via iterative Voronoi sampling, and build an *n* x *n* dissimilarity matrix *D*

$$D_{ij} = \left(\delta_{ij}\right)^2$$

with δ_{ij} the geodesic distance between vertices i,j computed following the FMTD algorithm

- Define a dimension *m* for the Euclidean embedding space and apply MDS on the matrix *D*, yelding an *n* x *m* matrix whose rows define the coordinates in $\Re^m$ of the points of the signature surface

Bending Invariant Surface Signatures [EK03]

- These two steps define a bending invariant descriptor, that allows to translate the problem of matching non-rigid objects in various posture into a simpler problem of matching rigid objects

Bending Invariant Surface Signatures [EK03]

- Drawback: embedding in the Euclidean space may introduce metric distortions

- Extension to non-Euclidean embeddings (such as embedding on the sphere [BBK05]) and introduction of Generalized MDS [BBK07]
- In [BBK06] partial surface matching is also addressed, introducing the Partial Embedding distance

Bending Invariant Surface Signatures [EK03]

- **Saliency**:
  - metric properties are captured by the geodesic distance
  - scalable amount of captured shape information, related to the dimension of the embedding
- **Invariance** wrt isometry
- Not **unique**, due to the randomly chosen starting point in the sampling stage, and not **complete**

Bending Invariant Surface Signatures [EK03]

- **Input**: triangulated surfaces
- **DoF and heuristics**:
  - choice of the dimension of the sampling and the embedding
  - choice of the specific MDS algorithm (classical, least squares, fast)
- **Efficiency**:
  - Computing the matrix requires $O(n^2)$, with n the number of sampled vertices
  - the MDS algorithm is at most $O(nN)$, with N number of iterations

Outline

- ✓ Descriptors parametric with respect to *f*:
  - Reeb graph
  - Size theory tools
  - Persistent homology tools
  - Descriptors based on spherical decompositions
    - Spherical harmonics
    - Spherical wavelets
- ✓ Descriptors linked to a specific *f*:
  - Shape-DNA
  - Bending invariant signature
  - Spectral embedding
  - Pose-oblivious shape signature
  - Salient geometric features

Spectral Embedding [JZ07]

- ✓ Ideas similar to [Elad and Kimmel 2003] are developed, introducing a descriptor suitable to compare articulated objects
- ✓ The matrix D is an affinity matrix involving a Gaussian of width σ

$$D_{i,j} = e^{-\frac{\delta_{ij}^2}{\sigma}}$$

with geodesic distances approximated through an heuristic

- ✓ The embedding in $\Re^m$ is given by the first m eigenvectors of the matrix, computed via Nyström approximation

Spectral Embedding [JZ07]

- ✓ The descriptor is given by the embedded surface or by the matrix first eigenvalues

Spectral Embedding [JZ07]

- ✓ **Saliency**:
  - metric properties captured by the geodesic distances
  - the possibility to use affinity matrices based on different functions (e.g. Euclidean or combined distances) is suggested
  - Scalable amount of shape information, related to the embedding dimension
- ✓ **Stability**: problems related to eigenmode switching and eigenmode sign assignment have to be faced
- ✓ **Robustness**: sensitiveness to outliers in the data
- ✓ **Invariance** wrt isometries

Spectral Embedding [JZ07]

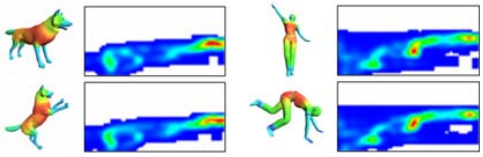
- ✓ **Unique** but not **complete**
- ✓ **Input**: triangulated surfaces (possibly to be repaired)
- ✓ **DoF and heuristics**:
  - sampling rate
  - embedding dimension
  - Gaussian width
  - heuristic to compute the geodesic distance
- ✓ **Efficiency**: $O(Nn\log n+N^3)$ operations required to compute and eigen-decompose the affinity matrix, with n the number of vertices of the mesh and N the number of sampled points

Outline

- ✓ Descriptors parametric with respect to *f*:
  - Reeb graph
  - Size theory tools
  - Persistent homology tools
  - Descriptors based on spherical decompositions
    - Spherical harmonics
    - Spherical wavelets
- ✓ Descriptors linked to a specific *f*:
  - Shape-DNA
  - Bending invariant signature
  - Spectral embedding
  - Pose-oblivious shape signature
  - Salient geometric features

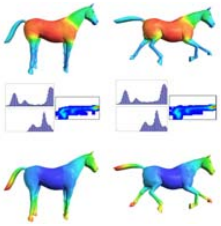
Pose-oblivious shape signature [GCO06]

- The pose-oblivious shape signature is a 2D histogram that combines two scalar functions defined on the boundary surface of the 3D shape.
  - the local diameter function: this function measures the diameter of the 3D shape in the neighbourhood of each vertex
  - the centricity function: this function measures the average geodesic distance from a vertex to all other vertices on the mesh [HSKK01]



Pose-oblivious shape signature [GSCO06]

- The shape signature is an histogram that combines both CF and DF
- The signature is represented as 2D array of scalar values between [0,0] and [1,1]
- Each array bin with values (*x*,*y*) contains the approximated probability of a point on the boundary of the mesh to have a DF value of *x* and a CF value of *y*

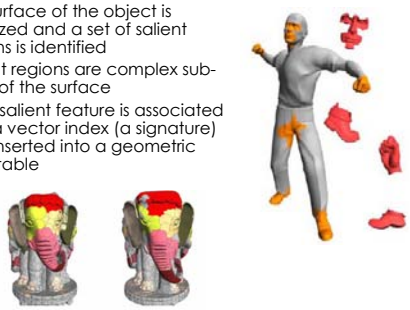


Outline

- Descriptors parametric with respect to *f*:
  - Reeb graph
  - Size theory tools
  - Persistent homology tools
  - Descriptors based on spherical decompositions
    - Spherical harmonics
    - Spherical wavelets
- Descriptors linked to a specific *f*:
  - Shape-DNA
  - Bending invariant signature
  - Spectral embedding
  - Pose-oblivious shape signature
  - Salient geometric features

Salient geometric features [GCO06]

- The surface of the object is analyzed and a set of salient regions is identified
- Salient regions are complex sub-parts of the surface
- Each salient feature is associated with a vector index (a signature) and inserted into a geometric hash table



Salient geometric features [GCO06]

- The local shape descriptor is a point *p* on a surface and its associated quadric patch that approximate the surface in a local neighbourhood of *p*.
- Salient geometric features are obtained by clustering together a set of descriptors such that they have a high curvature relative to their surroundings, and a high variance of curvature values

Questions?

Eurographics 2007 Tutorial T12

AIM@SHAPE

3D Shape Description and Matching Based on Properties of Real Functions

Comparison Methodologies

Eurographics 2007 Tutorial T12

Speaker

Simone Marini

CNR-IMATI-GE - Italy

Evaluating the matching characteristics

- **Properties** of the similarity measure
- **Robustness** of the similarity measure
  - *Low* variation of the measure wrt *small* variations of the shape descriptor
- **Type of comparison**
  - global and/or partial matching
- **Type of information** taken into account
  - geometrical, topological, structural
- **Computational complexity** of the matching algorithm
- **Application context**

Properties of similarity measures

- Let **S** be the set of shape descriptors, the distance measure **d** between two shapes descriptors is defined as:

$$d : S \times S \rightarrow \Re$$

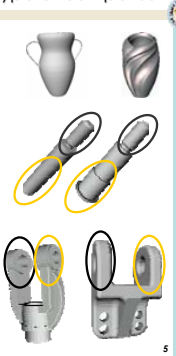
- Properties:
  - $d(x,x) = 0$ (self identity)
  - $d(x,y) > 0,\ x \neq y$ (positivity)
  - $d(x,y) = d(y,x)$ (symmetry)
  - $d(x,z) \leq d(x,y) + d(y,z)$ (triangular inequality)
  - $d(x,z) \leq \max\{d(x,y), d(y,z)\}$ (strong t. i.)

Properties of similarity measures

- The measure properties are grouped as in the following:
  - **semi-metric**: self-identity, positivity, symmetry
  - **pseudo-metric**: self-identity, symmetry, triangular inequality
  - **metric**: a pseudo-metric that satisfies the positivity
  - **ultra-metric**: a metric satisfying strong triangular inequality
- The perceptual space can be approximated by the metric properties ? [Tve77,SJ99]
  - symmetry and triangular inequality should not holds for partial matching
- Metrics can be used for indexing purposes

Type of comparison

- Global Matching:
  - Overall shape comparison
  - Real number representing the similarity estimation between the two objects
- Sub-Part Correspondence:
  - Real number as similarity estimation
  - Mapping among similar sub-parts
- Partial Matching:
  - Real number as similarity estimation
  - Similar sub-parts between objects having different overall shape



Type of information taken into account

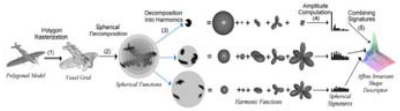
- according to the type of information stored and the way it is coded in the descriptor, the measure of similarity may take into account:
  - geometric information
  - topologic information
  - structural information

Comparison methodologies

- for descriptors represented by matrices and vectors
  - Spherical Harmonic representation [KFR03]
  - Shape DNA [RWP06]
  - Bending Invariant Surface Signatures[EK03,BBK06]
  - Spectral Embedding [JZ07]
  - Pose-oblivious shape signature [GSCO07]
  - Salient geometric features [GCO06]
- for descriptors represented by formal series
  - Size functions [dFL], [dAFL05]
  - Barcodes and persistence diagrams [CZCG05]
- for descriptors represented by graphs
  - Multiresolution Reeb Graphs [HSKK01]
  - Extended Reeb Graphs [BMSF06]

Spherical Harmonic representation [KFR03]

- Represent a function *f* defined on the sphere through its spherical harmonics and consider the vector of energies (i.e. frequency norms)
- Extension to voxel description:
  - Restrict the voxel grid to a collection of concentric spheres
  - Represent each spherical restriction in terms of the energy of its frequency decomposition, thus obtaining a 1D descriptor
  - The final descriptor resulting from the analysis of spheres with different radii is a 2D grid indexed by radius and frequency
- 2D descriptors are compared by using the L_2 norm



Spherical Harmonic representation: matching characteristics [KFR03]

- **Properties** of the similarity measure
  - metric
- **Robusteness** of the similarity measure
  - induced by the properties of metrics
- **Type of comparison**
  - global matching
- **Type of information** taken into account:
  - geometric information
- **Computational complexity** of the matching algorithm
  - linear in the number of entries stored in the 2D array
- **Application context**
  - retrieval of 3D objects, not suitable for articulated objects

Shape DNA [RWP06]

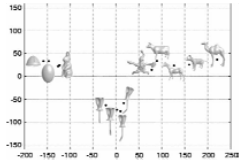
- Shape DNA signatures are *m*-dimensional feature vectors that can be compared using any metric between vectors, e.g. the Euclidean *p*-norm

$$d_p(u,v)=\left(\sum_{i=1}^{m}|u_i-v_i|^p\right)^{\frac{1}{p}}$$

  the Hausdorff distance, the Pearson correlation distance
- according to empirical evidence, d_2 yields good results while being easy to compute

Shape DNA [RWP06]

- Matching results on a small database of meshes, including different classes of deformed models, show a nice clustering of objects



- Other experiments on collections of grey-scale and colour images [RWP07]
- Medical applications on brain surfaces [NRW07], using statistical methods to distinguish populations; extention to 3D brain data

Shape DNA: matching characteristics [RWP06]

- **Properties** of the induced similarity measure
  - metric (using the Euclidean *p*-norm)
- **Robustness** of the similarity measure:
  - induced by the robustness of metrics
- **Type of comparison**: global matching
- **Type of compared information**
  - the descriptor stores geometric and topological information, but it is difficult to control them in a differentiated manner in the definition of the measure
- **Computational complexity** of the matching algorithm
  - p-norms are linear in the number of vertices of the model
- **Application context**:
  - medical applications, suitable for articulated objects

Bending Invariant Surface Signatures [EK03]

- Given the surface signatures, any algorithm to evaluate the similarity of rigid objects can be involved in the comparison step
- Example: Compute the vectors of the first few moments of the surfaces and compute their Euclidean distance

Bending Invariant Surface Signatures [EK03]

- **Properties** of the induced similarity measure
  - Depends on the matching method used
- **Robustness** of the similarity measure
  - Depends on the matching method used
- **Type of comparison**
  - Global or partial matching depending on the matching method used
- **Type of compared information**
  - Depends on the matching method used
- **Computational complexity** of the matching algorithm
  - Depends on the matching method used
- **Application context**
  - face recognition

Spectral Embedding [JZ07]

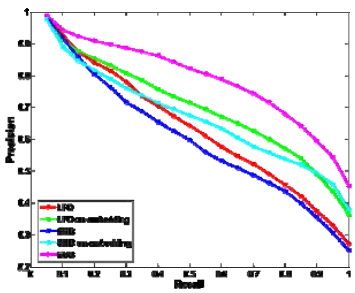
- Compare shapes by computing existing shape descriptors (Light Field, Spherical Harmonics) on spectral embeddings
- Use the vectors of normalized eigenvalues and define:

$$D_{EVD}(Q,S)=\frac{1}{2}\sum_{i=1}^{m}\frac{\left[\left|\lambda_i^Q\right|^{\frac{1}{2}}-\left|\lambda_i^S\right|^{\frac{1}{2}}\right]^2}{\left|\lambda_i^Q\right|^{\frac{1}{2}}+\left|\lambda_i^S\right|^{\frac{1}{2}}}$$

- Compute a correspondence cost derived from the correspondence between the vertices of the two shapes (possibly after a first filter using EVD)

$$D_{CCD}(Q,S)=\sum_{p\in Q}\left\|V_Q(p)-V_S(match(p))\right\|$$

Spectral Embedding [JZ07]



Spectral Embedding: matching characteristics [JZ07]

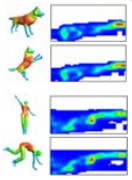
- **Properties** of the induced similarity measure
  - $D_{EVD}(Q,S)=\frac{1}{2}D(f,g)$;
  - $D(f,g)=\chi^2=\int\frac{(f-g)^2}{f+g}$, $f=\left|\lambda_i^Q\right|^{\frac{1}{2}}, g=\left|\lambda_i^S\right|^{\frac{1}{2}}$
  - X^2 is a semi-metrics if f and g are positives
- **Robustness** of the similarity measure
  - induced by the robustness of metrics

Spectral Embedding: matching characteristics [JZ07]

- **Type of comparison**
  - Global matching
- **Type of compared information**
  - geometric and topological information
- **Computational complexity** of the matching algorithm
  - D_{EVD} is linear in the number of vertices of the embedded model
- **Application context**
  - suitable for articulated objects

Pose-oblivious shape signature[GCO06]

- The pose oblivious is a 2D histogram that combines local diameter function and centricity function both defined on the boundary surface of the 3D shape.



- Matching:
  - correlation coefficient:

$$R(P,Q) = \frac{N\sum p_i q_i - \sum p_i \sum q_i}{\sqrt{(N\sum p_i^2 - (\sum p_i)^2)(N\sum q_i^2 - (\sum q_i)^2)}}$$

  - $\chi^2 = D(f,g) = \int \frac{(f-g)^2}{f+g}$

Pose-oblivious shape signature : matching characteristics [GSCO07]

- **Properties** of the similarity measure
  - X^2 is a semi-metrics if *f* and *g* are positives
  - correlation coefficient is a semi-metric
- **Robustness** of the similarity measure
  - induced by the properties of measures
- **Type of comparison**
  - global matching
- **Type of information** taken into account
  - the descriptor stores geometric information
- **Computational complexity** of the matching algorithm
  - linear in the number of entries stored in the 2D array
- **Application context**
  - retrieval of 3D objects, suitable for rticulated objects

Salient geometric features [GCO06]

- Each salient feature is associated with a vector index (a signature) and inserted into a geometric hash table
- Given a query object, its salient feature are extracted and used to query the database for a list of matching features
- The returned features identify the models having larger number of matches.

Salient geometric features [GCO06]

- The vector index used in the hash table encode the following information:
  - area of the salient feature
  - curvature of the salient feature
  - number of local minimum(s) or maximum(s) curvatures in the salient feature
  - the curvature variance in the salient feature
- The similarity between objects is given by the number of correspondence among the salient features

Salient geometric features : matching characteristics [GCO06]

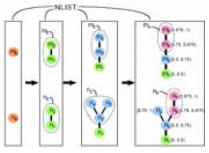
- **Properties** of the similarity measure
  - similarity measure is not proposed by authors
- **Type of comparison**
  - Sub-part correspondence and partial matching.
- **Type of information** taken into account:
  - geometric information
- **Computational complexity** of the matching algorithm
  - depends on the geometric hashing used
- **Application context**
  - retrieval of 3D objects, object alignement

Multiresolution Reeb Graph [HSKK01]

- Similarity between two nodes *P*,*Q* is weighted on their attributes:

$$sim(P,Q) = \alpha \,|\, A(P) - A(Q) \,| + (1-\alpha) \,|\, L(P) - L(Q) \,|\,, \; 0 < \alpha < 1$$

- Nodes with maximal similarity are paired if:
  - Share the same range of *f*
  - Parent nodes are matched
  - Belong to graph paths already matched
- The distance between two MRGs is the sum of all node similarities

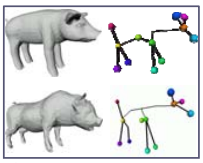


Multiresolution Reeb Graph: matching characteristics [HSKK01]

- **Properties** of the induced similarity measure
  - metric
- **Robustness** of the similarity measure
  - stability properties of metric
- **Type of comparison**
  - global matching (suitable also for sub-part correspondence and partial matching)
- **Type of compared information**
  - structural and geometric information
- **Computational complexity** of the matching algorithm
  - $O(M \cdot (M+N))$ where M and N is the number of nodes of the two multiresolution graphs
- **Application context**
  - Retrieval of free form objects

Extended Reeb Graphs [BMSF06]

- Two ERGs are compared using a graph-matching approach based on the "best common subgraph" detection
- Also sub-part correspondences are recognized
- Heuristics are used to improve
  - Quality of the results
  - Computational time



Extended Reeb Graphs [BMSF06]

- Given G_1 and G_2, two direct, acyclic and attributed graphs:
  - the distance d between two nodes $v_1 \in G_1$ and $v_2 \in G_2$ is

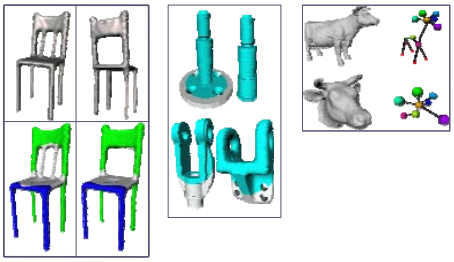
$$d(v_1, v_2) = \frac{w_1 G_S + w_2 St_S + w_3 Sz_S}{3} \qquad w_i \in [0,1], \ \sum w_i = 1$$

  - the distance $D(G_1, G_2)$ depends both on the geometry and the structure of the objects:

$$D(\mathcal{G}1, \mathcal{G}2) = 1 - \frac{\sum_{v \in \varphi}(1 - d(\psi_1(v), \psi_2(v)))}{max(|\mathcal{G}1|, |\mathcal{G}2|)}$$

Extended Reeb Graphs [BMSF06]

- Some examples of sub-part correspondence and partial matching



Extended Reeb Graphs: matching characteristics [BMSF06]

- **Properties** of the Induced similarity measure
  - semi-metric
- **Robustness** of the similarity measure
  - Stability properties of semi-metrics
- **Type of comparison**
  - global matching, sub-part correspondence and partial matching
- **Type of compared information**
  - structural and geometric information
- **Computational complexity** of the matching algorithm
  - $O(n^3)$ where n is $\max\{|G_1|,|G_2|\}$
- **Application context**
  - free form objects and CAD models

Matching distance between 1-dimensional size functions [dAFL06]

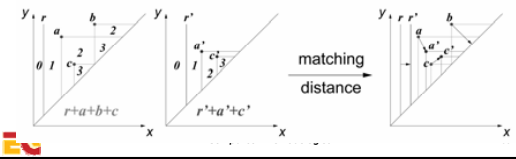
- Two size functions ℓ_1, ℓ_2, with associated formal series C_1 and C_2, can be compared by measuring the reciprocal distances of cornerpoints and cornerlines

$$\delta((x,y),(x',y')) = \min\left\{\max\{|x-x'|,|y-y'|\}, \max\left\{\frac{y-x}{2}, \frac{y'-x'}{2}\right\}\right\}$$

and choosing the matching which minimizes the maximum of these distances

$$d_{match}(\ell_1, \ell_2) = \min_{\sigma} \max_{p \in C_1} \delta(p, \sigma(p))$$

when σ varies among the bijections between C_1 and C_2



Matching distance between 1-dimensional size functions [dAFL06]

- Matching Stability Theorem:
  The matching distance satisfy the following stability condition:

$$\max_{P \in \mathcal{M}} |\varphi(P) - \psi(P)| \leq \epsilon \Rightarrow d_{match}(\ell_{(\mathcal{M},\varphi)}, \ell_{(\mathcal{M},\psi)}) \leq \epsilon.$$

- Lower bound for the natural pseudo-distance:
  Let δ be the matching distance between the two size functions $\ell_{(M,\varphi)}$ e $\ell_{(N,\psi)}$. Then

$$d((M,\varphi),(N,\psi)) \geq \delta.$$

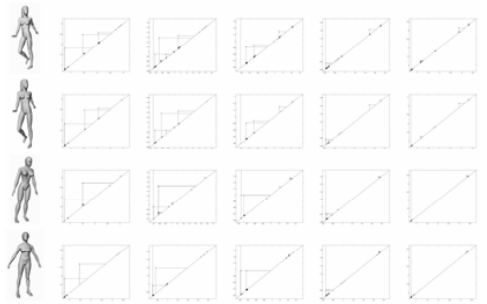
Matching distance between multidimensional size functions [BCF*07]

- On each leaf of a particular foliation of their domain, multidimensional size functions coincide with a particular 1-dimensional size function
- the induced 1D matching distance on each leaf of the foliation is stable wrt small changes of the leaves;
- a multidimensional matching distance can be defined

$$D_{match}\left(\ell_{(\mathcal{M},\vec{\varphi})}, \ell_{(\mathcal{N},\vec{\psi})}\right) = \sup_{(\vec{l},\vec{b})} \min_{i=1,\ldots,k} l_i \cdot d_{match}(\ell_{(\mathcal{M},F^{\vec{\varphi}}_{(\vec{l},\vec{b})})}, \ell_{(\mathcal{N},F^{\vec{\psi}}_{(\vec{l},\vec{b})})})$$

- theorems about the stability of the matching distance and the lower bound for the natural pseudo distance can be stated also in the case k>1

Multidimensional Size Functions [BCF*07]



Size Functions: matching characteristics [dAFL06,BCF*07]

- **Properties** of the induced similarity measure
  - the matching distance is a metric
  - it provides a lower bound for the natural pseudo-distance
- **Robustness** of the similarity measure
  - stability theorem for the matching distance
- **Type of comparison**
  - global matching
- **Type of compared information**
  - geometric-topological
- **Computational complexity** of the matching algorithm
  - $O(n^{2.5})$, where n is the number of cornerpoints taken into account
- **Application context**
  - Medical images, trademarks recognition, 3D retrieval

Barcodes [CZCG05]

- I, J intervals in a barcode, $\delta(I,J) = |I \cup J - I \cap J|$
- A matching between barcodes S_1, S_2 is the set
  $M(S_1,S_2) \subseteq S_1 \times S_2 = \{(I,J)\, s.t.\, I \in S_1, J \in S_2\}$
  s.t. any interval in S_1, S_2 occurs in at most one pair (I,J)
- Distance between S_1, S_2 relative to M

$$D_M(S_1,S_2) = \sum_{(I,J)\in M} \delta(I,J) + \sum_{L \in N} |L|$$

with N the set of non matched intervals

Barcodes [CZCG05]

- Barcode pseudo-metric:

$$D(S_1,S_2) = \min_M D_M(S_1,S_2)$$

- Minimizing D_M is equivalent to maximizing the similarity

$$S_M(S_1,S_2) = \frac{1}{2}\left(\sum_{S_1} |I| + \sum_{S_2} |J| - D_M(S_1,S_2)\right)$$

- Recasting the problem as a graph problem, such minimization is equivalent to the well known maximum weight bipartite matching problem

Barcodes [CZCG05]

- Examples on mathematical surfaces
- Classification results on a set of 80 hand-drawn copies of letters

Persistence Diagrams [CSEH07]

- Describing $\mathscr{F}$-intervals as point sets in the extended plane, i.e. by persistence diagrams, the Bottleneck Stability Theorem has been proved
- Under conditions on the space and the functions *f,g*, it holds that the Bottleneck distance between persistence intervals $D(f)$, $D(g)$ satisfies

$$d_B\big(D(f), D(g)\big) \le \|f-g\|_\infty$$

where d_B is defined as

$$d_B(X,Y) = \inf_\gamma \sup_x \|x-\gamma(x)\|_\infty$$

with X, Y multisets of points, $x \in X$, $y \in Y$ range over all points and γ ranges over all bijections from *X* to *Y*

Barcodes and Persistence Diagrams [CSEH07]

- In terms of persistence diagrams, the distance defined for barcodes can be written

$$d(D_1, D_2) = \inf_\gamma \sum_p \|p-\gamma(p)\|_1$$

with γ ranging in the set of bijections between D_1 and D_2, but this distance does not guarantee the stability property proven for persistence diagrams under the Bottleneck distance

- Under certain assumptions, the Barcode Theorem holds, guaranteeing the stability property under the Bottleneck distance

Barcodes and persistence diagrams: matching characteristics [CZCG05,CSEH07]

- **Properties** of the induced similarity measure
  - pseudo-metric between barcodes
  - metric between persistence diagrams
- **Robustness** of the similarity measure
  - stability theorems for barcodes and persistence diagrams under the Bottleneck distance
- **Type of comparison**
  - global matching
- **Type of compared information**
  - geometric-topological
- **Computational complexity** of the matching algorithm
  - for the pseudo-metric between barcodes, it depends on the algorithm used to minimize $D(s_1, s_2)$
- **Application context:** 3D and curve PCD comparison

Questions?

Eurographics 2007 Tutorial T12

3D shape description and matching based on properties of real functions

Conclusions and future perspectives

Eurographics 2007 Tutorial T12

Speakers

Bianca Falcidieno Michela Spagnuolo

CNR-IMATI-GE - Italy

Shape M&R: remarks

- Shape matching and retrieval is a complex process that involves reasoning at many different levels
  - Perception
  - Query formulation
  - Context understanding and formalization
  - Shape complexity (retrieval in broad or narrow domains)
- Crucial aspects that determine the performance of shape and retrieval systems
  - Saliency of the descriptors
  - Efficiency of the matching
  - *Indexing*

2

Shape M&R: remarks on performances

- Quantitative measures: relatively easy ..
  - Precision/recall, first and second tier, CG, ..
  - *More attention to the ground truth definition*
  - *Flexible classification tools*
- Qualitative measures :
  - more difficult, need users or scenarios and specific models
- Reproducibility of results
  - Executables should be provided
  - Benchmarks for specific shape M&R tasks

3

Evaluation and benchmarking

- SHREC'07 - Shape Retrieval Contest 2007 promoted by AIM@SHAPE, and coordinated by Remco Veltkamp (UU)
  - Organized **every year**, in conjunction with Shape Modeling International – SMI – (next year, Stony Brook, June 2008)
  - **Multi-track**: this year 7 tracks for watertight models, partial matching, protein models, CAD models, relevance feedback, similarity measures, 3D faces

http://www.aimatshape.net/event/SHREC

4

why SHREC

- PSB has limited benchmarking capabilities
- different representation models
  - NOT format, but MODELS: triangle meshes, NURBS, Breps, that may have different characteristics

- different "types of" similarity
  - Form (geometry or structure)
  - Function & Semantics
  - ..

5

Shape M&R: remarks

- The results did not show any method really outperforming others
- Results per single query can give insights on what method is best suited for specific shapes

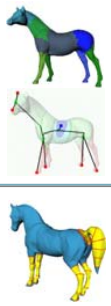
No single *best* method exists

Benchmarking can help the definition of *best practices* to help users selecting the most appropriate retrieval method for the application context, shape category and complexity, type of similarity implied

6

How to *describe* a shape ?

- Geometry
  - *Detect relevant local features*
- Structure
  - *Organize them in a structure*
- Semantics
  - *Use the structure to detect high-level features (semantics)*



7

How to embed semantics in Shape M&R?

- All methods use semantics/knowledge in the shape description process
- Reasoning at semantic level (eg, logic based reasoning) on shape similarity requires the annotation of shapes and/or shape parts
- How can we associate semantic "tags" to shapes or shape parts and use them in a Shape M&R sessions ?

… starting from the shape description step …

8

- is there a "best" method to segment a shape ?

How to use more effectively segmentation tools to **annotate** shapes ?

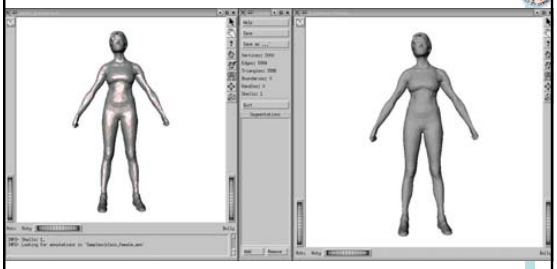
9

Shape Annotator

- Shape features are identified through multiple *segmentation algorithms*
  - A single segmentation algorithm is usually not enough to capture all the feature classes necessary for a satisfactory annotation, even in a single domain
  - **Solution: *Pick*** interesting features from different segmentations
- Compose the best segmentation for a specific domain and **annotate** shape features with concepts formalized by an ontology

10

The ShapeAnnotator

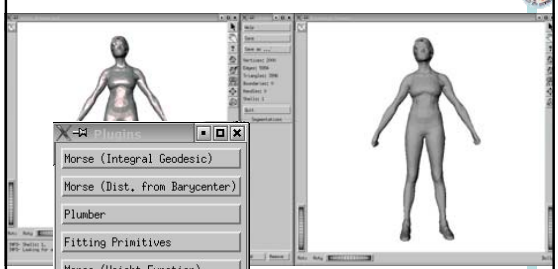


window showing the results of each segmentation plug-in

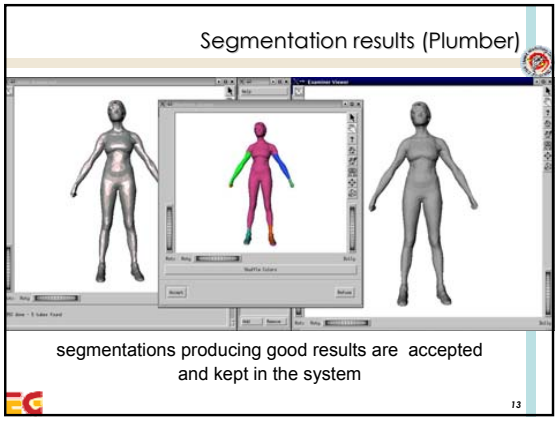
window showing the composition of the final segmentation

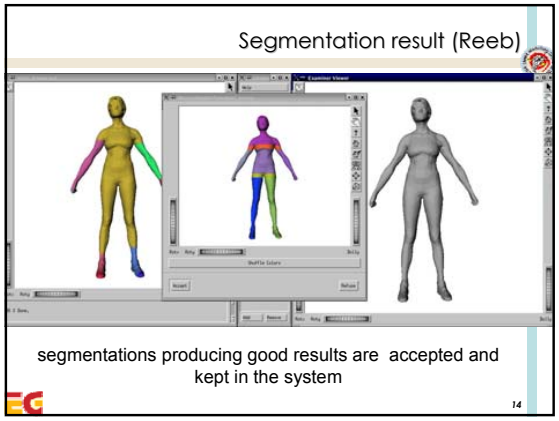
11

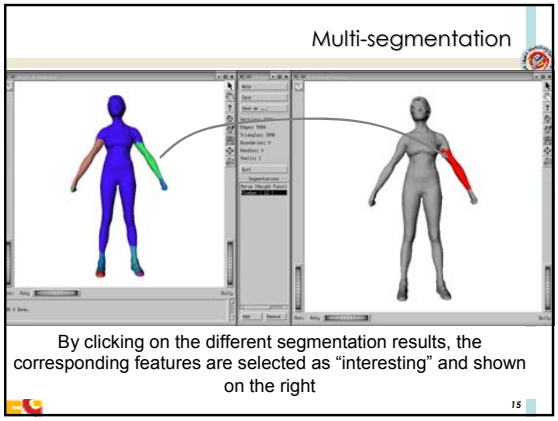
Segmentation plugins

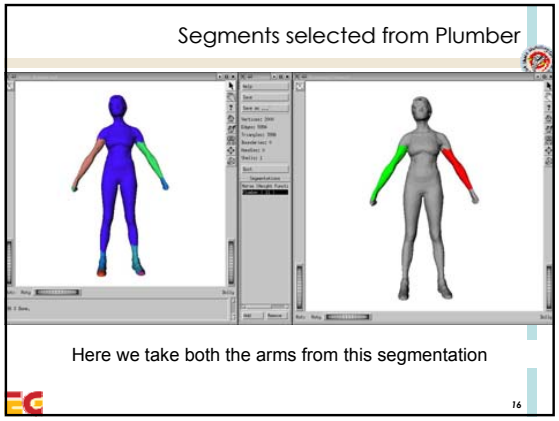


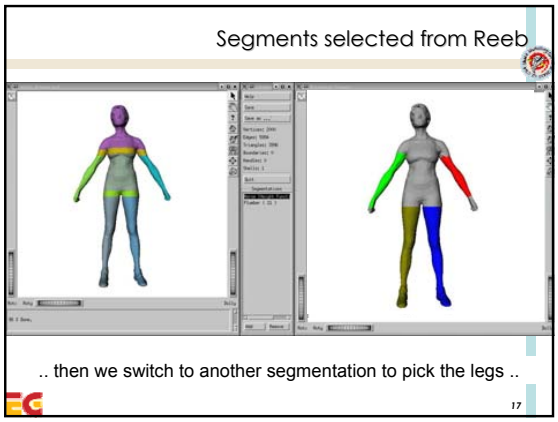
12

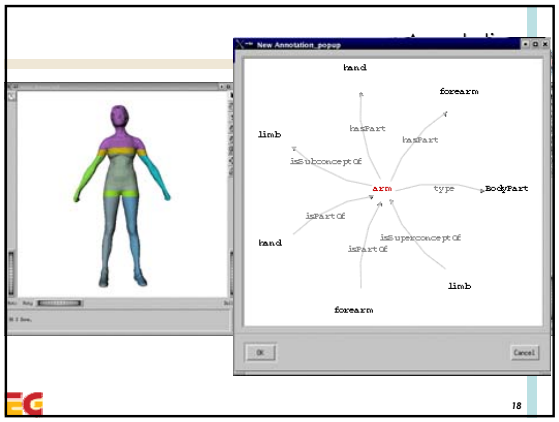




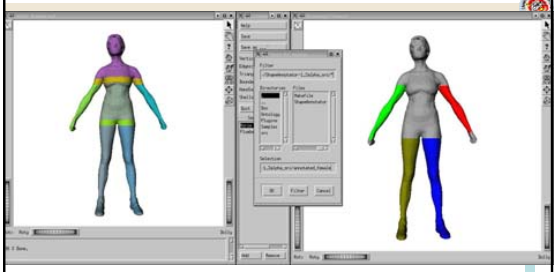








Saving the annotated mesh



The multi-segmented and annotated mesh can be saved and shared in a knowledge base

19

Shape Annotator

- ✓ Submitted papers
- ✓ Technical report available (contacts: Francesco Robbiano, Marco Attene)

20

Shape Annotator and 3D Search

- ✓ From geometry-based to semantics-based search engines for 3D content
  - See the presentation of Francesco, IMATI, this afternoon
  - How to formalize shape descriptions (ShaDe)
- ✓ Mixing geometric and semantic reasoning will allow for text-based query formulation to be used and mixed with geometric similarity evaluation and retrieval

21

Trends/challenges

- ✓ To make the **creation and modification** of 3D shapes, with the associated knowledge, as easy as **cut & paste** for texts
  - the development of appropriate knowledge representations and data structures for **intelligent** 3D shapes management
  - the ability to identify regions in 3D digital shapes that correspond to what humans perceive as regions of interest and tag them appropriately (**semantic annotation**)

22

Trends/challenges

- ✓ To make the **creation and modification** of 3D shapes, with the associated knowledge, as easy as **cut & paste** for texts
  - To develop **3D search engines** based not only on geometry but also on semantics
  - To deliver the **appropriate 3D content in the appropriate modality**
    - development of appropriate visualization techniques including semantic rendering and semantic LODs
    - visual rendering of the relevance of retrieval results

23

... the end ...

Questions?

Eurographics 2007 Tutorial T12

contacts:

{bianca,daniela,simone,patane,michi}@ge.imati.cnr.it