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 [Radmacher 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 seemingly impossible camera movements such as the ‘bullet time’ (Matrix) sequence recorded with multiple cameras with staggered exposure times.

- It encompasses 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 volumetric 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 Helmholtz 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 Narasimmhan 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, Sayaj 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 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-Illumination

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 2005] 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 \(L_i(u,v,\theta_i,\phi_i)\) gives complete control over the incident light field on an object, \(L_i(u,v,\theta_i,\phi_i)\). 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 \(I = \sum w_i I_i\). 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 sets 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 et 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 object. 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. Malzbender 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. Winnenmoller et al. [Winnenmoller 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].

© The Eurographics Association 2007
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 refocussing 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 micro-lenses 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


Time


Exposure


Focus


Illumination


Passive Illumination


Polarization


© The Eurographics Association 2007

Wavelength


Location


Matting


Techniques

General


Graph Cuts


© The Eurographics Association 2007

Recognition (CVPR), June 1997, Puerto Rico

Gradient Domain

313–318.


Smoothing, Bilateral and Trilateral Filter


CHOUDHURY, P., TUMBLIN, J., "The Trilateral Filter for High Contrast Images and Meshes". Proc. of the

Video Textures

Conference on Computer Graphics and Interactive Techniques International Conference on Computer Graphics and

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,"

TODD ZICKLER, PETER N. BELHUMEUR, AND DAVID J. KRIEGMAN, "Helmholtz Stereopsis: Exploiting
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

© The Eurographics Association 2007

**Motion**


**Transfer and denoising**

**Flash to no-flash**


**Noise**


Eric P. Bennett and Leonard McMillan "Video Enhancement Using Per-Pixel Virtual Exposures" SIGGRAPH 2005

**Geometric Operations**

**Panorama**


Synthetic Aperture and Multi-views


REN NG, Fourier Slice Photography, SIGGRAPH 2005


A. STERN and B. JAVIDI, "3-D computational synthetic aperture integral imaging (COMPSAI)," Opt. Express 11, 2446-2451 (2003),


Deblurring and Superresolution


Dynamic Range


© The Eurographics Association 2007


**Active Illumination**


© The Eurographics Association 2007
Smart, Unconventional Cameras

MEMS Technology


High Speed Imaging


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


Programmable SIMD


© The Eurographics Association 2007
Advanced, Programmable, Demodulating Cameras and Temporal Correlation

CANESTA Inc, 2004
PIXIM Inc, 2004
FOVEON Inc, 2004
JENOPTIK Inc, 2004
IVP Inc, Ranger Camera, 2004


Others


Tumblin, J., Agrawal, A. and Raskar, R. Why I want a Gradient Camera, IEEE CVPR 2005

Optics


Zand, J., Coded aperture imaging in high energy astronomy, NASA Laboratory for High Energy Astrophysics (LHEA) at NASA’s GSFC, 1996.

Light field optics


© The Eurographics Association 2007


**Fourier optics**


Computational Photography

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

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 Computationelia, Photograp, 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.

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), both 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.
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

Computational Photography:
Optics, Sensors and Computations

Computational Photography

Novel Illumination
Computational Photography

Novel Cameras

Generalized Sensor

Processing

Generalized Optics

Novel Illumination

Light Source

Scene: 3D Ray Modulator

Display

Recreate 4D Lightfield

Scene: 3D Ray Modulator

Computational Photography

Novel Cameras

Generalized Sensor

Processing

Generalized Optics

Novel Illumination

Light Source

Scene: 3D Ray Modulator

Display

Recreate 4D Lightfield

Scene: 3D Ray Modulator

Computational Photography

Novel Cameras

Generalized Sensor

Processing

Generalized Optics

Novel Illumination

Light Source

Scene: 3D Ray Modulator

Display

Recreate 4D Lightfield

Scene: 3D Ray Modulator

Radial Stereoscopic Imaging

(Kathrin Nayar, Columbia University)

Dual photography

from diffuse reflections

Sen et al., Siggraph 2005

Fluttered Shutter Photography

Blurred Taxi, 200 Pixel Blur

Deblur Ffittered Shutter Photography

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:
  – 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?

PHYSICAL

3D Scene light sources, BRDFs, shapes, positions, movements, …

Exposure, Control, tone map

Light & Optics

Epyepoint position, movement, projection, …

Display RGB(x,y,t)

PERCEIVED

Scene light sources, BRDFs, shapes, positions, movements, …

Eyeepoint position, movement, projection, …

Photo: A Tangible Record

Editable, storable as Film or Pixels
**Ultimate Photographic Goals**

**PHYSICAL**
- 3D Scene: light sources, BRDFs, shapes, positions, movements.
- Eyepoint: position, movement, projection.

**PERCEIVED or UNDERSTOOD**
- Light & Optics
- Visual Stimulus
- Vision

**Photo: A Tangible Record**
- Scene estimates we can capture, edit, store, display.

**Meaning...**

---

**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

---

**Missing: Expressive Time Manipulations**

What other ways better reveal appearance to human viewers? (Without direct shape measurement?)
Can you understand this shape better?

---

**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:

- Light Sources
- Sensors
- Optics
- Optical elements
- Processing
- Display
- Rays
Review: How many Rays in a 3-D Scene?

A 4-D set of infinitesimal members.

Imagine:
- Convex Enclosure of a 3D scene
- Inward-facing ray camera at every surface point
- Pick the rays you need for ANY camera outside.
- 4D surface of cameras,
  2D ray set for each camera,
  $\rightarrow$ 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: $R_{ij} = \frac{\text{outgoing}}{\text{incoming}}$

Future Photography:
Novel Cameras
Novel Displays
Novel Illuminators
Lights
General Optics: 4D Ray Benders
Generalized Sensor
Generalized Processing
4D Ray Sampler
Generalized Display
Looking for light rays
Recreated 4D Light field
Scene: 4D Ray Modulator

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…)
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…)

---

**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.

---

**‘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?
Computational Photography

Understanding Film-Like Photography
or ‘from 2D Pixels to 4D Rays’

Jack Tumblin
Northwestern University

Rays and the ‘Thin Lens Law’

- Focal length $f$: where parallel rays converge
- Focus at infinity: Adjust for $S_2 = f$
- Closer Focus? Larger $S_2$

Not One Ray, but a Bundle of Rays

- BUT Ray model isn’t perfect: ignores diffraction
- Lens, aperture set the point-spread-function (PSF)

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 $S_2$

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

Try it Live! Physlets:

http://webphysics.davidson.edu/Applets/Optics/intro.html

Try it Live! Physlets:

http://webphysics.davidson.edu/Applets/Optics/intro.html
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

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: $f(\lambda)$
  - color fringes near edges of image

Chromatic Aberration

- Lens Design Fix: Multi-element lenses
  - Complex, expensive, many tradeoffs!
- Computed Fix: Geometric warp for R,G,B. 

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.

Film-like Color Sensing

- Visible Light: narrow band of e’mag. spectrum
  $\lambda \approx 400-700 \text{ nm (nm = 10^{-9} \text{ meter wavelength)}}$
- (humans:<1 octave → honey bees: 3-4 octaves)
  do honey bees sense harmonics, see color ‘chords’?

Equiluminant Curve defines ‘luminance’ vs. wavelength
Film-like Color Sensing

- Visible Light: narrow band of emag spectrum
- $\lambda = 400-700$ nm (nm = 10-9 meter wavelength)
- At least 3 spectral bands required (e.g., R, G, B)

Color Sensing

- 3-chip: vs. 1-chip: quality vs. cost

1-Chip Color Sensing: Bayer Grid

- Estimate RGB at ‘G’ cells from neighboring values

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

\[ I(x) = I(x-1) + G(x) \]

Cumulative sum

Intensity Gradient in 2D

Gradient at $x,y$ as Forward Differences

\[ G_x(x,y) = I(x+1,y) - I(x,y) \]
\[ G_y(x,y) = I(x,y+1) - I(x,y) \]
\[ G(x,y) = \langle G_x, G_y \rangle \]
Intensity Gradient Vectors in Images

Gradient Vector

Image Intensity Gradients in 2D

Sanity Check:
Recovering Original Image

Solve Poisson Equation, 2D linear system

Intensity Gradient Manipulation

A Common Pipeline

Modify Gradients

Graph and Images

Credits: Jianbo Shi

Agrawala et al, Digital Photomontage, Siggraph 2004
**Segmentation = Graph partition**

\[ G = (V, E) \]

- **V**: graph node
- **E**: edges connecting nodes
- **wij**: Edge weight
- **pixel**: Image pixel
- **E**: Link to neighboring pixels
- **Pixel similarity**

---

**Minimum Cost Cuts in a graph**

Cut: Set of edges whose removal makes a graph disconnected

\[ S_{ij} : \text{Similarity between pixel } i \text{ and pixel } j \]

Cost of a cut,

\[ \text{cut}(A, \overline{A}) = \sum_{i \in A, j \in \overline{A}} S_{i,j} \]

---

**Graph Cuts for Segmentation and Mosaicing**

Cut \sim String on a height field

---

**Bilateral Filtering**

- Start with Gaussian filtering
  - Here, input is a step function + noise
Start with Gaussian filtering
• Spatial Gaussian f
\[
J(x) = f \ast I(x)
\]
Output is blurred
\[
J(x) = f \ast I(x)
\]
Gaussian filter as weighted average
• Weight of \( \xi \) 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

$$J(x) = \frac{1}{k(x)} \sum_{\zeta} f(x, \zeta) g(\|I(\zeta) - I(x)\|) I(\zeta)$$

Bilateral filtering is non-linear

[Tomasi and Manduchi 1998]

* The weights are different for each output pixel

$$J(x) = \frac{1}{k(x)} \sum_{\zeta} f(x, \zeta) g(\|I(\zeta) - I(x)\|) I(\zeta)$$

Bilateral Filtering

* Unilateral filtering
  - Smoothing using filtering
* Bilateral filtering
  - Edge-preserving smoothing

Input

Gaussian Smoothing

Lightness-Adaptive 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

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

Agrawala et al., Digital Photomontage SIGGRAPH 2004

NEAR

FAR

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)

`Debevec'97 (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/m²
- Range of Typical Displays: from ~1 to ~100 cd/m²

`??`
**FUSION: Multispectral Wavelengths**

Vegetation Mapping of the Forest

SAR | Optical Landsat

**Color2Gray: Salience-Preserving Color Removal**

SIGGRAPH 2006
Gooch, Olsen, Tumblin, Gooch

New Method

Color Original | Grayscale

**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**

\[ \text{crunch}(x) = \alpha \cdot \tanh(x/\alpha) \]

\[ \alpha = 5 \quad \alpha = 10 \quad \alpha = 25 \]
Which hue changes should be darker in grayscale?

Set a polarity for the color wheel:

\[ \text{sign}(\| \Delta C_{ij} \|) = \text{sign}(\Delta C_{ij} \cdot v) \]

Original Color2Grey Color2Grey+Color

**Color Difference Space**

\[ \mathbf{v} = (\cos \theta, \sin \theta) \]

**RECONSTRUCTION: Clear Day from Foggy Days**

Two Different Foggy Conditions

Clear Day Image

Deweathering

Time: 3:00 PM

Time: 5:30 PM

**Varying Polarization**

Varying Polarization

Yoav Y. Schechner, Nir Karpel 2005

Best polarization state

Worst polarization state

Best polarization state

Recovered image

Left: The raw image is 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

**Non-photorealistic Camera:**

Depth Edge Detection and Stylized Rendering using Multi-Flash Imaging

Ramesh Raskar, Karhan Tan, Rogerio Feris, Jingyi Yu, Matthew Turk

Mitsubishi Electric Research Labs (MERL), Cambridge, MA

U of California at Santa Barbara
Our Method

Canny

RECONSTRUCT: Depth Discontinuities

Internal and external
Shape boundaries, Occluding contour, Silhouettes
Both...

Reconstructed Edges

Merged Photos

Merged Result

Reconstruct: Light Source Angles from camera, without calibration

Both...

Raking Camera for Pentimenti

Jack Tumblin, Ankit Mohan, Chi-Yin Cheung, Eric Russell

Dec. 2005: Kirk Vuillemot, Art Institute of Chicago, positions, prepares

"Man with Moustache..." 1915, Pablo Picasso


Jack Tumblin, Jack Tumblin, Ankit Mohan, Chi-Yin Cheung, Eric Russell

New form of optical-- only `Pentimenti' (regrets):

Fresnel Reflectance reveals artist's overpainting, scrapings, & revisions

Raking Camera

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
‘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
  \(120^\circ\) spacing

Reconstruction:
Polynomial Texture Maps

Store just 6 coefficients at each pixel, get Interactive re-lighting...

A Mostly 2-D Method

X-ray Study: we cancel the backing?

Pentimenti:
Can we see, follow the revisions?

Shape-Time Photography

Visually Expressive ‘Time Fragments’

- Duchamp
  - Nude Descending a Staircase

Freeman et al 2003
Concluding Questions

- How should we modify ‘film-like’ photography to better gather the ‘visually essential’ (not just the ‘optically essential’) contents of a scene?
Computational Illumination

Ramesh Raskar
Mitsubishi Electric Research Labs

Course WebPage: http://www.merl.com/people/raskar/photo/

Ramesh Raskar, Computational Illumination

Computational Photography

Novel Cameras

Generalized Optics

4D Light Field

Scene: 8D Ray Modulator

Display

Remote: 4D Lightfield

Processing

Ray Reconstruction

Upto 4D Ray Sampler

4D Incident Lighting

Computational Illumination

Novel Cameras

Generalized Optics

4D Light Field

Scene: 8D Ray Modulator

Display

Remote: 4D Lightfield

Processing

Ray Reconstruction

Upto 4D Ray Sampler

4D Incident Lighting

Programmable 4D Illumination field + time + wavelength

Computational Illumination

‘Smarter’ Lighting Equipment

What Parameters Can We Change?

Edgerton 1930’s

Not Special Cameras but Special Lighting
**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

---

**Denoising Challenging Images**

Available light:
- nice lighting
- noise/blurriness
- color

---

**Flash:**
- details
- color
  - flat/artificial
Transfer detail from flash image to no-flash image

+ original lighting
+ details/sharpness
+ color

Cross-Bilateral Filter based Approach

Build Exposure HDR image

- Multiple images with different exposure
  - Debevec & Malik, Siggraph 97
  - Nayar & Mitsunaga, CVPR 00

Increasing Exposure

Build Flash HDR image
Capturing HDR Image
- Varying Exposure time
- Varying Flash brightness
- Varying both

Flash and Ambient Images
- Agrawal, Raskar, Nayar, Li (Siggraph05)

Intensity Gradient Vectors in Flash and Ambient Images
- Same gradient vector direction
- Flash Gradient Vector
- Ambient Gradient Vector
- No reflections

Different gradient vector direction
- Reflection Ambient Gradient Vector
- Flash Gradient Vector
- With reflections

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

**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

---

**Move shadow back**

**Soften this shadow**

**Make this brighter**

**Remove this specular highlight**

**Make this darker**

---

**Debevec et al. 2002: ‘Light Stage 3’**

**Image-Based Actual Re-lighting**

Debevec et al., SIGG2001

Film the background in Milan,
Measure incoming light,
Matched LA and Milan lighting.

Light the actress in Los Angeles

Matte the background
Computational Illumination

Novel Cameras

Generalized Sensor

Processing

Ray Reconstruction

Generalized Optics

4D Ray Bender

Programmable 4D Illumination field = time x wavelength

4D Light Field

Display

Scene 4D Ray Modulator

Ramesh Raskar, Computational Illumination

Debevec et al. 2000

Masselus et al. 2002

Malzbender et al. 2002

A 4-D Light Source

Non-photorealistic Camera: Depth Edge Detection and Stylized Rendering using Multi-Flash Imaging

Ramesh Raskar, Karhan Tan, Rogerio Feris, Jingyi Yu, Matthew Turk

Mitsubishi Electric Research Labs, Cambridge, MA

U of California at Santa Barbara

U of North Carolina at Chapel Hill

MultiFlash NPR Camera
Depth Discontinuities

Internal and external Shape boundaries, Occluding contour, Silhouettes

Our Method

Canny Intensity Edge Detection

A New Problem

Shadows Clutter
Highlight Shape Edges Mark moving parts
Mark many colors Basic colors
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

What does synthetic aperture illumination look like?

Underwater confocal imaging with and without SAP

Dual Photography

Dual Photography
**Dual Photography**

- Projector
- Photocell
- 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**

- Book
- Aperture
- Camera
- Projector

- The camera’s view

Sen et al, Siggraph 2005
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

A Night Time Scene:
Objects are Difficult to Understand due to Lack of Context

Dark Bldgs
Reflections on bldgs
Unknown shapes

Enhanced Context:
All features from night scene are preserved, but background is clear

'Well-lit' Bldgs
Reflections in bldgs windows
Tree, Street shapes

Background is captured from day-time scene using the same fixed camera

Result: Enhanced Image

'Smarter' Lighting Equipment

Programmable Parameters
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

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 dependent on distance from plane of focus
- Cubic Phase Plate
  - Defocus nearly independent of distance
  - All points ‘blurred’
  - Deconvolve to get sharper image

Wavefront Coding:
10X Depth of Field

http://www.cdm-optics.com/site/extended_dof.php

Integral Photography

Todor Georgiev et al 2006
Light field photography using a handheld plenoptic camera

Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz and Pat Hanrahan

Prototype camera

Contax medium format camera
Kodak 16-megapixel sensor

x 300 Optics microlens array
125ȝ 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/

Fujis 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

Natural Scene Properties

High Dynamic Range Images

Intensity camera fail to capture range
Gradients saturate at very few isolated pixels

Log Camera Image

Locally Adaptive Gain

Problem: Visible quantization effects at high intensities
Thus the local neighborhood is lost and only isolated values are captured.
Gradient Camera

- Two main features:
  1. Sense difference between neighboring pixel intensity
     - At each pixel, measure $(x', y')$:
     \[ x' = I_{x+1, y} - I_{x, y} \quad \text{and} \quad y' = I_{x, y+1} - I_{x, y} \]
  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
  - Problems: 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 for 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

- Intensity Histogram
- Gradient Histogram

- 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

Camera Pipeline

On-board Hardware

Software

4 Pixel Clusters

Difference between pixels

Local gain adaptive to difference

2D Integration to reconstruct the image

Log

Uniform

GradCam

Detail Preserving

Intensity Histogram

Gradient Histogram

Original image

Uniform quantization 3 bits

Log Uniform quantization 3 bits

Log Uniform gradients quantization 3 bits

Original Image

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

619
**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:
- 1 cm for 2.7 meters

---

**Canesta: Modulated Emitter**

- Time of flight measurement
  - Phase ~ distance
  - Amplitude ~ reflectance

---

**Motion**
Fluttered Shutter Camera
Raskar, Agrawal, Tumblin Siggraph2006

- 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.
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?

Multiperspective Camera?

Agrawal et al, Long-Soome Panoramas, Siggraph 2006

Rademacher et al, McOP, Siggraph 1998

[ 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:
  - All HD 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