Image-based Animation of Clothes

A. Hilsmann and P. Eisert[†]

Humboldt University of Berlin and Fraunhofer Heinrich-Hertz Institute, Berlin, Germany

Abstract

We propose a pose-dependent image-based rendering approach for the visualization of clothes with very high rendering quality. Our representation combines body-pose-dependent geometry and appearance. A geometric model accounts for low-resolution shape adaptation, e.g. animation and view interpolation, while small details as well as complex shading/reflection properties are accounted for through numerous images. Information on shading, texture distortion and silhouette at fine wrinkles are extracted from the images to allow later texture replacement. The image-based representations are estimated in advance from real samples of clothes captured in an offline process, thus shifting computational complexity into the training phase. For rendering, pose dependent geometry and appearance are interpolated and merged from the stored representations.

Categories and Subject Descriptors (according to ACM CCS): I.3.8 [Computer Graphics]: Applications—

1. Introduction

We propose a new image-based representation for the visualization of articulated objects, such as clothes worn by a person. The application we are targeting at is the realistic visualization of garments in augmented reality environments, where virtual clothes are rendered into real video material. The most crucial point in this scenario is the photo-realistic visualization of the virtual clothes rendered onto a moving human body in real-time. Existing approaches to cloth visualization focus on high-quality cloth simulation and rendering but are far from working in real-time, e.g. [BMF03]; others provide a real-time visualization, but are quite limited in visual quality or to retexturing of small regions [PLF05, HE09].

We propose a new representation for clothes that combines rough 3D geometry with body-pose dependent appearance. The use of real images leads to very natural looking results while the geometric information allows animation of the model. As the pose-dependent acquisition of the image-based representation is performed a-priori, the computational complexity during visualization is reduced and realistic visual rendering in real-time will be possible. Our approach is mainly driven by the following assumptions:

- We are rather aiming at a plausible photo-realistic and perceptually correct visualization of clothes than at accurate and correct reconstruction.
- Texture and shading represent strong cues for the perception of shape such that fine details can be modeled by images, carrying information on texture distortion, shading and silhouette (*appearance*). Hence, rough shape is modeled in a geometric model to allow animation while small details are represented by appearance.
- Wrinkles and creases of clothes are mainly influenced by a person's pose, such that our approach models wrinkling behavior as a function of the underlying pose (concentrating on types of clothing that roughly follow a person's shape). We assume that wrinkling behavior is preliminary affected by the nearest joints. Hence, we can establish a database of image-based representations for different views and poses and interpolate and merge any new pose from the captured representations in the database.

Most image-based approaches, e.g. [DCJM96, BBM*01, CTMS03], focus on view-dependency and do not allow for geometric modification or animation of the object. Especially, very little research has been done in body-pose dependent image-based rendering techniques. Recently, new methods for space-time interpolation of complex captured scenes were proposed in [LLB*10]. Instead of interpolation between images taken at different times, our aim is to allow animation and interpolation in pose-space.

[†] The work presented in this paper is funded by the German Science Foundation, DFG EI524/2-1.

DOI: 10.2312/conf/EG2012/short/069-072

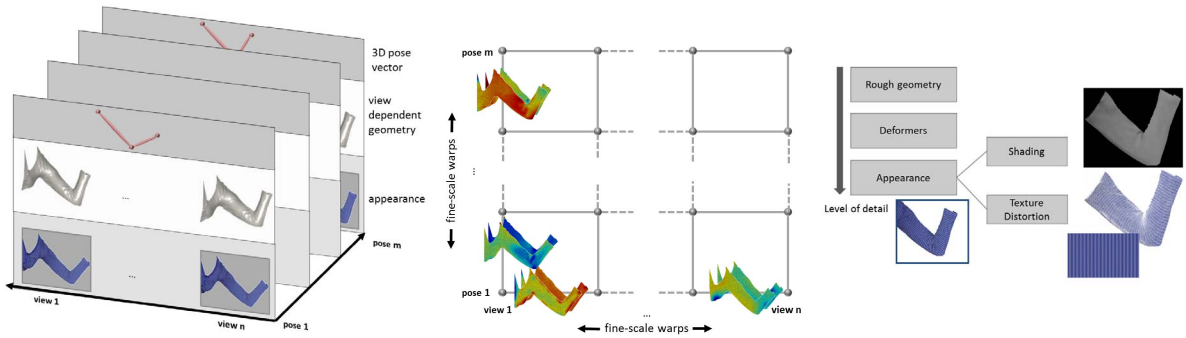

Figure 1: *Left: Base model representation consisting of a multi-view/multi-pose image set, silhouette information, view-dependent geometry and 3D pose information. Center: Additional fine-scale mesh warps are estimated between neighboring views as well as neighboring poses on top of the geometry-guided warps to achieve view-/pose consistency. Right: Appearance is separated into texture albedo, shading and texture distortion to allow retexturing using [HSE11].*

The assumption, that fine-scale wrinkling is pose-dependent also appears in [WHRO10], a recent example-based cloth simulation approach. Here, fine-scale wrinkling geometry is learned for different poses in advance for each joint separately and added onto a large-scale cloth simulation during rendering. Our approach makes the same assumption but models wrinkling behavior in appearance instead of geometry. In [XLS*11] the texture of a human body model is synthesized from a database of images to synthesize video sequences of human performance.

This paper presents work in progress. We describe the concept of the proposed representation in section 2. In section 3 we describe how we use our a representation for view-/pose-interpolation using the example of a reduced scenario. By *reduced* we mean, that in the example we do not use a full body-model but show view- and pose-interpolation of an arm-bending sequence with realistic wrinkling behavior, thus leaving merging representations of different body parts of a full model for future work (section 4).

2. Pose-dependent Geometry and Appearance Model

Our proposed image-based representation combines view- and body-pose-dependent geometry and appearance. The main idea is to capture a database of images from different (calibrated) view points showing different body poses to be able to interpolate and merge such a representation for any new incoming pose from the database. The database will consist of different levels of detail (see figure 1). A coarse geometric model together with a skeleton model (representing body poses as a vector of joint angles) accounts for dominant motions, e.g. view interpolation and coarse animation. Additional fine-scale warps are needed to achieve photo-consistency during rendering. Very fine wrinkles will be captured by appearance and represented by a large number of images, accounting for texture distortion and shading. At the silhouette, fine details are accounted for through an alpha-mask associated to the captured images. To allow not only manipulation of shape, but also of appearance, we separate the appearance information into albedo, shading and spatial texture distortion such that retexturing of the piece of cloth is possible while preserving shading and texture distortion properties of the original image [HSE11]. The representation is illustrated in figure 1 and described in the following.

Our representation consists of a set of multi-view images $\mathcal{I}_{v,p}$ taken from different view points v and showing different body poses p. Each image is associated with a silhouette image $\mathcal{S}_{v,p}$ that can be generated by e.g. semi-automatic or automatic segmentation.
The generation of new views and poses from the database is based on mesh-based warps representing correspondences between *neighboring* images (similar views or poses). These warps are separated into a *geometry-guided warp* and an additional *fine-scale warp*. Let $\mathcal{M}^2_{v,p}$ denote a 2D triangular mesh on image $\mathcal{I}_{v,p}$, generated by triangulating the silhouette $\mathbf{S}_{v,p}$ using [She96]. Further let $\mathbf{V}_{v,p} = [\mathbf{v}_1^T \dots \mathbf{v}_k^T]^T$ denote the concatenated vertices of this mesh. For each vertex, we estimate a displacement vector to its corresponding position in a second image, such that a warp $\mathcal{W} = \{\mathbf{V}, \mathbf{V}^*\}$ can be represented by vertex positions in the original and the warped image. To estimate the warps between image pairs we use an image-based optimization approach [HE09].

From vertex correspondences between neighboring views at a fixed pose and camera calibration information, we can assign a depth to each vertex and such generate a mesh-based depth map $\mathcal{M}^3_{v,p}$ (view-dependent geometry) for each image $\mathcal{I}_{v,p}$. This geometry will be used for coarse view interpolation and animation. Similar to many image-based rendering techniques, e.g. [BBM*01], to render a new view of a virtual camera, we search for the nearest k camera views, project the corresponding depth meshes into the new view and use the such established 2D vertex correspondences to warp the corresponding images to the new virtual view (*geometry-guided view warping*). The warped images are blended according to the angular distance between the new virtual camera view

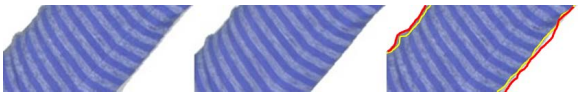

Figure 2: *Left two images: blending without and with fine-scale warps. Instead of blending in the new silhouette information, the vertex displacements gradually move the silhouette to that of the new view, e.g. the yellow silhouette onto the red one in the right image.*

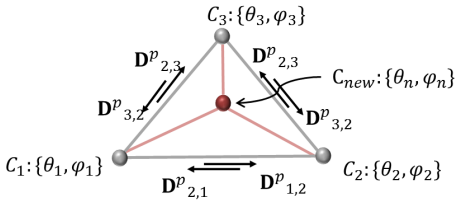

Figure 3: *Camera views are weighted according to the barycentric coordinates of the new camera view in azimuth/elevation angle space of the nearest 3 camera views. Fine-scale warps are estimated between the image pairs.*

and the k camera views. To compensate for inaccurate depth maps and also to avoid the typical *blending-in-effect* at the silhouette, we estimate additional fine-scale warps between neighboring views on top of the geometry-guided warp as follows. After warping an image onto a neighboring view using geometry-guided view warping, we optimize an additional warp that fine-scale aligns the two images using [HE09]. We denote the vertex displacements of a fine scale-warp at a fixed pose p from view v_i to view v_j by $\mathbf{D}^p_{v_i,v_j}$. The idea is similar to floating textures [EDM*08], but the *float field* is estimated a-priori and interpolated during rendering. This has the effect that the silhouette (as well as the texture) rather *moves* to its new position during interpolation instead of *blending in* (figure 2) yielding a time-consistent view interpolation result.

For animation, each view-dependent geometry is associated with a 3D skeleton model whose joint angles represent the 3D pose. The skeleton can either be fitted manually or automatically, e.g. using a morphable human body model. Skinning weights between vertices of each view dependent depth-mesh $\mathcal{M}^3_{v,p}$ and each skeleton joint are calculated using [BP07]. To render a new pose, we animate the nearest (in view- and pose-space) view-dependent depth meshes to the new pose using linear blend skinning, project the animated depth meshes into the new view and use the such established 2D vertex positions to warp the corresponding images (*geometry-guided warping*). The skinning weights produce smooth and intuitive deformations for vertices attached to bones but of course do not capture the *real* deformation of clothes when e.g. an arm is bent. To assure photo-consistency between poses, we estimate additional fine-scale warps (represented by vertex displacements $\mathbf{D}^v_{p_i,p_j}$) on top of the geometry-guided pose warp by first warping two images of the same view but different poses onto each other using

Figure 4: *Retextured sample pose.*

the geometry-guided pose warp and then optimizing for an additional fine-scale warp that registers the two images accurately.

Finally, from the image correspondences we can determine affine color correction models between the image pairs which are directly applied to all images. To allow modification of the appearance by *retexuring* (figure 4), we separate the images into a shading map, texture albedo and texture distortion using [HSE11]. Texture distortion and shading information are represented as a (geometric and photometric) warp from an estimated undeformed reference texture to the images in the database. For a view- and pose- consistent retexturing result, the complete procedure of [HSE11] needs to be done only once, and texture distortion as well as shading map information can be optimized for each image in the database using the estimated correspondences between views and poses. The full model representation (figure1) is determined a-priori for pre-defined poses and views in a training phase.

3. Image-based Animation of Clothes

In this section, we describe how we generate images of new views (accounting for a rigid pose change of the body to the cameras) and body-poses from the image-based representation for a reduced scenario (arm bending sequence, 1 joint, 3 cameras, 10 poses). We parameterize camera positions based on azimuth θ and elevation ϕ angles $C_v : \{\theta_v, \phi_v\}$ and poses based on joint angles α_p. Given a new virtual camera position C_{new} and a new body pose α_{new}, we determine the set of the two nearest poses $\mathsf{p} = \{p_1, p_2\}$ (based on differences between the joint angles) and the set of the three nearest cameras $\mathsf{v} = \{v_1, v_2, v_3\}$.

The geometry-guided warps $\mathcal{W}_{v,p} = \{\mathbf{V}_{v,p}, \mathbf{V}^*_{v,p}\}, v \in \mathsf{v}, p \in \mathsf{p}$ for the corresponding images are determined by animating the depth meshes $\mathcal{M}^3_{v,p}$ to the new pose using linear blend skinning and projecting them into the new view yielding 2D vertex positions $\mathbf{V}^*_{v,p}$. To these vertex positions we now add the weighted fine-scale displacements between views and poses. In view space we interpolate based on barycentric coordinates in $\{\theta, \phi\}$-space (similar to [LLB*10]) of the three cameras (figure 3). This assures that if the virtual view falls exactly onto one of the views in the database, the weight of that particular view is set to one while all others fall to zero. In pose-space we interpolate linearly. The displacements are added according to these weights to each warp

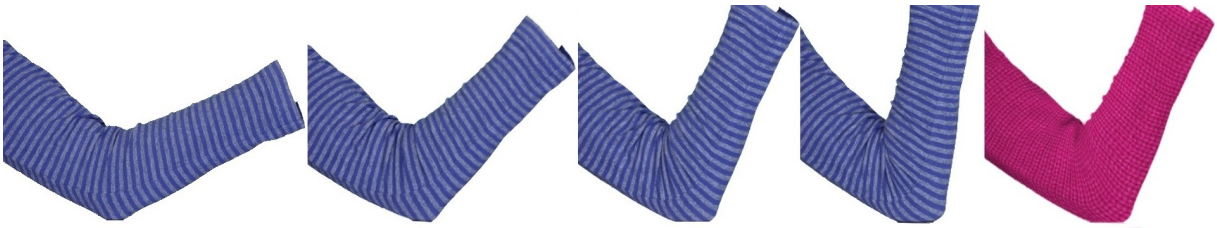

Figure 5: *Details of an arm bending sequence interpolating between the first and fourth images. The second and third images are synthetically generated in-between poses. Note how the wrinkling behavior is perceptually correct. This can be better seen in the accompanying video material. The rightmost image shows a retextured in-between pose.*

$\mathcal{W}_{v,p}, v \in \mathsf{v}, p \in \mathsf{p}$:

$$\mathbf{V}^*_{v,p} \leftarrow \mathbf{V}^*_{v,p} + \sum_{v_i \in \mathsf{v} \setminus v} \beta_{v_i} \cdot \mathbf{D}^p_{v,v_i} + w_p \cdot \mathbf{D}^v_{p,p_j}$$

with $w_p = \frac{|\alpha_p - \alpha_{new}|}{|\alpha_{p_1} - \alpha_{p_2}|}$ and $p_j = \mathsf{p} \setminus p$.

The images $\mathcal{I}_{v,p}$ are now warped using the such established vertex correspondences $\mathcal{W}_{v,p} = \{\mathbf{V}_{v,p}, \mathbf{V}^*_{v,p}\}$. Blending the images is performed using barycentric coordinates in view space and linear interpolation in pose space:

$$\mathcal{I}_{new} = \sum_{p \in \mathsf{p}} w_p \cdot \left(\sum_{v \in \mathsf{v}} \beta_v \cdot \mathcal{I}^*_{v,p} \right)$$

where $\mathcal{I}^*_{v,p}$ denotes the warped images. The image-based generation of new views and poses can either be performed with the original images or with retextured versions using texture analysis and shading information from the original images.

Figure 5 shows close-ups of the elbow in synthetically generated pose images of an arm bending sequence. The left and right most images show images from the database while the second and third images are synthetically generated in-between pose images. The perceptually correct wrinkling behavior is best evaluated in the accompanying video material.

4. Conclusions and Future Work

We presented current work on image-based visualization of clothes from a view- and pose-dependent image database. By separating the appearance information into albedo, shading and spatial texture distortion also changing the appearance (*retexturing*) is possible. Future work includes an extension of our representation to more complex body models. It will be necessary to treat different body poses independently to be able to cover the complete space of possible body poses such that more complex interpolation strategies will be needed.

References

[BBM*01] BUEHLER C., BOSSE M., MCMILLAN L., GORTLER S., COHEN M.: Unstructured Lumigraph Rendering. In *Proc. Conf. on Computer Graphics and Interactive Techniques* (2001), SIGGRAPH 2001, pp. 425–432. 1, 2

[BMF03] BRIDSON R., MARINO S., FEDKIW R.: Simulation of Clothing with Folds and Wrinkles. In *Proc. of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2003), pp. 28–36. 1

[BP07] BARAN I., POPOVIĆ J.: Automatic Rigging and Animation of 3D Characters. *ACM Trans. Graph. 26* (July 2007). 3

[CTMS03] CARRANZA J., THEOBALT C., MAGNOR M. A., SEIDEL H.-P.: Free-Viewpoint Video of Human Actors. *ACM Trans. Graph. 22* (July 2003), 569–577. 1

[DCJM96] DEBEVEC P. E., C. J T., MALIK J.: Modeling and Rendering Architecture from Photographs: A Hybrid Geometry- and Image-based Approach. In *Proc. Conf. on Computer Graphics and Interactive Techniques* (1996), SIGGRAPH 1996. 1

[EDM*08] EISEMANN M., DECKER B. D., MAGNOR M., BEKAERT P., DE AGUIAR E., AHMED N., THEOBALT C., SELLENT A.: Floating Textures. *Computer Graphics Forum (Proc. of Eurographics) 27*, 2 (2008), 409–418. 3

[HE09] HILSMANN A., EISERT P.: Tracking and Retexturing Cloth for Real-Time Virtual Clothing Applications. In *Mirage 2009 - Computer Vision/Computer Graphics Collaboration Techniques and Applications* (Rocquencourt, France, 2009). 1, 2, 3

[HSE11] HILSMANN A., SCHNEIDER D. C., EISERT P.: Warp-based Near-Regular Texture Analysis for Image-based Texture Overlay. In *Vision, Modeling, and Visualization Workshop 2011* (Berlin, Germany, 2011). 2, 3

[LLB*10] LIPSKI C., LINZ C., BERGER K., SELLENT A., MAGNOR M.: Virtual Video Camera: Image-Based Viewpoint Navigation Through Space and Time. *Computer Graphics Forum 29*, 8 (Dec. 2010), 2555–2568. 1, 3

[PLF05] PILET J., LEPETIT V., FUA P.: Augmenting Deformable Objects in Real-Time. In *Proc. Int. Symposium on Mixed and Augmented Reality (ISMAR 2005)* (October 2005). 1

[She96] SHEWCHUK J. R.: Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. In *Applied Computational Geometry: Towards Geometric Engineering*, vol. 1148 of *Lecture Notes in Computer Science*. Springer-Verlag, 1996, pp. 203–222. 2

[WHRO10] WANG H., HECHT F., RAMAMOORTHI R., O'BRIEN J.: Example-based Wrinkle Synthesis for Clothing Animation. *ACM Transactions on Graphics (SIGGRAPH 2010) 29*, 4 (July 2010). 2

[XLS*11] XU F., LIU Y., STOLL C., TOMPKIN J., BHARAJ G., DAI Q., SEIDEL H.-P., KAUTZ J., THEOBALT C.: Video-based Characters - Creating New Human Performances from a Multi-View Video Database. In *Proc. of SIGGRAPH 2011* (2011). 2