Video-based Reconstruction of Animatable Human Characters

Carsten Stoll, MPI Informatik
Juergen Gall, ETH Zurich
Edilson de Aguiar, Disney Research
Sebastian Thrun, Stanford University
Christian Theobalt, MPI Informatik

Figure 1: *Frames of a real-time animation that was created with a fully-animatable character model that we reconstructed from multi-view video. An animation artist created a new texture for the model and simply applied a motion capture file from the internet to obtain this new animation. The moving geometry and the cloth behavior are as lifelike as on the real subject.*

Abstract

We present a new performance capture approach that incorporates a physically-based cloth model to reconstruct a rigged fully-animatable virtual double of a real person in loose apparel from multi-view video recordings. Our algorithm only requires a minimum of manual interaction. Without the use of optical markers in the scene, our algorithm first reconstructs skeleton motion and detailed time-varying surface geometry of a real person from a reference video sequence. These captured reference performance data are then analyzed to automatically identify non-rigidly deforming pieces of apparel on the animated geometry. For each piece of apparel, parameters of a physically-based real-time cloth simulation model are estimated, and surface geometry of occluded body regions is approximated. The reconstructed character model comprises a skeleton-based representation for the actual body parts and a physically-based simulation model for the apparel. In contrast to previous performance capture methods, we can now also create new real-time animations of actors captured in general apparel.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—; I.4.8 [Image Processing and Computer Vision]: Scene Analysis—;

Keywords: performance capture, markerless motion capture, multi-view reconstruction, game characters

ACM Reference Format
Stoll, C., Gall, J., de Aguiar, E., Thrun, S., Theobalt, C. 2010. Video-based Reconstruction of Animatable Human Characters. *ACM Trans. Graph. 29*, 6, Article 139 (December 2010), 10 pages. DOI = 10.1145/1866158.1866161 http://doi.acm.org/10.1145/1866158.1866161.

1 Introduction

Currently, there is a notable gap in visual quality and detail between virtual humans in movie productions, which are entirely rendered off-line, and virtual characters in real-time applications, like computer games. Ever more powerful computers and specialized hardware and software for real-time physics and cloth simulation (e.g., Nvidia PhysX^{TM} or Havok^{TM}) will soon enable us to narrow this gap. Eventually the complexity and visual fidelity of real-time characters will start to rival that of today's virtual film actors.

The flip-side is that when it comes to character design, developers of games and consumer graphics applications will soon be facing quality requirements that come much closer to those of today's movie productions. More detailed surface geometry will have to be designed or scanned, more detailed textures will have to be painted or photographed, more detailed motion will have to be key-framed or captured, and cloth simulation models will have to be properly tuned to enable real-time rendering. Fortunately, various acquisition technologies exist to support the animator in parts of the character design process: motion capture systems to obtain skeleton motion, 3D scanners to measure static geometry, or specialized measurement devices to analyze cloth material parameters [Kawabata 1980]. However, using such non-integrated tools can be very cumbersome, expensive, and time-consuming. In consequence, even movie productions still rely heavily on manual design and often measure the effort to create virtual actors in man-months of work. This is a level of complexity that most game developers will not be able to afford.

The design process for complex future game characters would become faster and more affordable if an integrated capture technology existed that could measure all aspects of an animation, such as geometry, motion, material values, and appearance, from real subjects. A first step towards this goal has been taken by performance capture approaches which enable the reconstruction of detailed motion and time-varying geometry of humans from multi-view video recordings [Bradley et al. 2008; Vlasic et al. 2008; de Aguiar et al. 2008b]. Unfortunately, performances captured with these approaches can merely be played back, but their motions and deformations cannot be arbitrarily modified by an animator.

Therefore, we propose a new marker-less method to capture skeletal motion, 3D geometry, and the dynamics of the character's apparel from multi-view video. This representation is fully-animatable, i.e., it allows us to freely modify the skeletal motion of the character and create new animations where cloth and surface deformations look as lifelike as on the real subject. Our goal is not the reconstruction of models physically accurate models; rather, we aim for easy and integrated reconstruction of plausible rich performance models that serve as a basis for modification by the animator and that can be realistically animated in real-time. Our main contributions are:

- A new approach to capture plausible fully-animatable virtual humans from sparse video recordings of real world actors in loose apparel that comprise: detailed surface geometry, a fully-rigged skeleton, and a cloth model for loose pieces of apparel.
- An algorithm to automatically identify wavy cloth regions and to approximate geometry that is occluded by the attire.
- An algorithm to estimate plausible cloth simulation parameters from the entire apparel of a subject in motion.
- A method to easily create new real-time animations with the reconstructed human, without needing a new recording.

To achieve our goal, we have to carefully select and design solutions to a large variety of challenging algorithmic sub-problems. Only the proposed combination of these solutions enables reconstruction of animatable models. The integration of our data into real-time applications is simple and our cloth simulation framework is compatible with a widely-used real-time physics library. We show results on a variety of example sequences and also confirm the high fidelity of our reconstructions in a user study.

2 Related Work

In our research, we build upon previous work in many different disciplines in graphics and vision. In the following, we limit our overview to a representative subset.

Our algorithm is an evolution of previous performance capture approaches that reconstruct non-modifiable models of dynamic shape and motion of real-world actors from video. Today's performance capture algorithms are based on marker-less motion capture algorithms from computer vision [Poppe 2007]. Most of these original methods only work for people with skin-tight clothing, merely provide skeletal motion parameters, and thus fail to recover detailed dynamic shape models, let alone cloth simulation parameters.

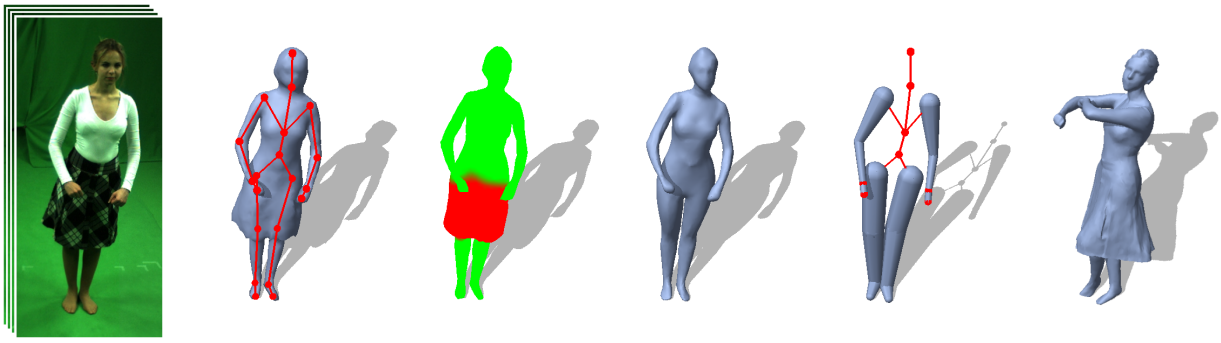
In the fields of image-based rendering and 3D video, several methods for dynamic geometry reconstruction have been proposed [Matusik et al. 2000; Zitnick et al. 2004; Waschbüsch et al. 2005], with the final goal of rendering novel viewpoints of the scene. Unfortunately, the level of detail in the captured geometry is typically not sufficient for reconstruction of an animatable model. Furthermore, most methods fall short in recovering actual motion parameters, and if they do, such as in the case of model-based 3D video [Carranza et al. 2003], people in arbitrary loose clothing cannot be handled. One step towards reconstructing an animatable representation of a human was taken by Allen et al. [2006], who learn a deformation model of the naked shoulder and torso from body scans in different postures. Sand et al. [2003] use a similar idea and learn pose-dependent body deformations by using marker-based motion capture and silhouette matching. Similarly, the SCAPE model learns variations in human body shape and surface deformation from laser scans of people [Anguelov et al. 2005]. [Park and Hodgins 2008] learn a data-driven model of skin and muscle deformation from dense marker-based motion capture data. All these methods are suitable for more-or-less naked people, but none of them can handle people in loose everyday apparel.

Performance capture algorithms promise to reconstruct detailed scene geometry and surface appearance from unmodified multi-view video streams. Starck and Hilton [2007] use a combination of shape-from-silhouette, multi-view stereo, and spatio-temporal cross-parameterization to capture detailed dynamic scene geometry. Recently, Vlasic et al. showed a multi-view photometric stereo approach to capture dynamic geometry of people recorded under a dome with rapidly switching light conditions [2009]. To facilitate reconstruction of spatio-temporally coherent geometry of people in arbitrary clothing, de Aguiar et al. propose to use deformable meshes created from static full-body laser scans for tracking [2008b]. The methods by Vlasic et al. [2008] and Gall et al. [2009] utilize a skeleton to ease the estimation of the non-rigid mesh deformations.

The biggest disadvantage of all performance capture approaches to date is that the captured result cannot be modified in such a way that physically plausible surface deformations persist (e.g., plausible folds). Mesh sequence editing techniques exist [Kircher and Garland 2006; Xu et al. 2007] that allow for a certain degree of modification. However, they neither allow the same level of control as our fully-rigged character, nor create plausible cloth animations in new poses at a level of quality comparable to our approach. Our method is inspired by motion segmentation methods that identify approximately rigidly moving sections in mesh animations [James and Twigg 2005; de Aguiar et al. 2008a] to identify waving cloth regions and fit a physics-based model to the input performance.

Our final animatable model uses the position-based dynamics approach by Müller et al. [2007] for forward simulation of cloth since it represents a good compromise between speed, simulation accuracy, and complexity of the parameter space. We estimate the parameters of the cloth model from captured performance data. A variety of alternative cloth simulation methods exist in the literature that go beyond plausible simulation and allow for physically accurate simulations (see [Choi and Ko 2005] for an overview of some methods). However, most of them do not satisfy our real-time animation core goals. Many previous approaches for cloth estimation from real world samples reconstruct deforming cloth geometry only, but not a physically-based forward simulation model. Some methods use multi-view feature matching and stereo for vision-based deforming geometry capture of square cloth samples [Pritchard and Heidrich 2003]. Other algorithms rely on special marker patterns printed on the fabric, as well as on an a priori geometry model of a piece of apparel, to measure time-varying cloth geometry from multi-view video [Scholz et al. 2005; White et al. 2007]. The latter approach also learns a simple data-driven deformation model which can approximate the wrinkling of fabric in new poses. Recently, Bradley et al. [2008] proposed a new approach to capture deforming cloth geometry from multi-view video without explicit markers in the scene. An extension of their method allows image-based reinsertion of fine scale geometric folds that could originally not be captured in the geometry [Popa et al. 2009].

So far, only few algorithms estimate parameters of a physics-based cloth model from images. Bhat et al. [2003] learn such parameters from waving square fabric samples that were recorded with a real-time structured light system. Conceptually related is the approach by Shi et al. [2008] who estimate parameters for simulating secondary skin deformation. In contrast, our method neither requires an a priori shape model for individual pieces of apparel, nor any form of visual pattern in the scene, or any off-line physical material measurements from fabric samples. Our method fully-automatically identifies all cloth regions on the entire moving human and recovers plausible simulation parameters.



(a) (b) (c) (d) (e) (f)

Figure 2: *Overview of our processing pipeline: (a) multi-view video sequence of a reference performance; (b) performance capture result: skeleton motion + deforming surface; (c) cloth segmentation (regions of loose apparel in red); (d) Statistical body model fitted to reference geometry; (e) estimated collision proxies. (f) After optimal cloth simulation parameters are found, arbitrary new animations can be created.*

3 Overview

It is our goal to reconstruct an animatable human character in general clothing from unmodified multi-view video streams, with little manual intervention. The first step in our algorithm is recording a so-called *reference sequence* of the person whose model is to be reconstructed (Fig. 2(a)). In this sequence, the person walks around for a few seconds in the same attire that should be part of the final model. From this reference sequence, we reconstruct a so-called *reference performance* which comprises a deforming mesh sequence that represents the detailed dynamic surface geometry of the moving actor, as well as a sequence of joint angle parameters of an underlying skeleton (Fig. 2(b), Sect. 4).

The dynamic surface geometry of the reference performance is automatically decomposed into approximately rigid body parts and non-rigidly deforming pieces of attire (Fig. 2(c)). In the same stage of the algorithm, we fit a statistical body model to the reference performance (Fig. 2(d)) and reconstruct approximate collision geometry (Fig. 2(e)). These collision proxies are needed for believable and fast cloth simulation (Sect. 5). After segmentation, we fit a physics-based cloth simulation model to each region of the deforming fabric (Sect. 6).

The output is an animatable performance model comprising a skeleton with surface skinning for non-cloth regions, collision proxies, and a physics-based simulation model for apparel. Arbitrary new animations of the character with realistic dynamic surface appearance can be created by simply changing skeletal motion parameters or material parameters of fabrics (Fig. 2(f) and Sect. 7).

4 Capturing the Reference Performance

Prior to recording the reference performance, we acquire a model of the subject that comprises two components: a surface mesh and an underlying bone skeleton. To measure the surface mesh $\mathcal{M}$, we take a static full-body laser scan $\mathcal{M}_{high}$ of the actor wearing the attire for the reference sequence, which we then decimate to roughly $n_{tri} = 5000$ triangles. The decimation is necessary to keep both performance capture and, later, performance simulation times in reasonable bounds. The original laser scan may contain holes, some of them due to bad measurements and some of them due to occlusion, such as at the underside of the skirt. We first automatically fill in all holes by means of Poisson surface reconstruction to create a closed surface. Some originally occluded regions may have been filled in a "semantically" wrong way, such as the sheet of triangles connecting the lower rim of skirt and legs. We have developed a semi-automatic tool that suggests regions of "semantically" incorrect triangles. For this we detect regions where the surface reconstruction closed large holes in the original scan. The user can then refine this suggestion manually. Implausible triangles are thereby marked as invalid. They are weighted down in the reference performance capture and excluded from the animatable model estimation.

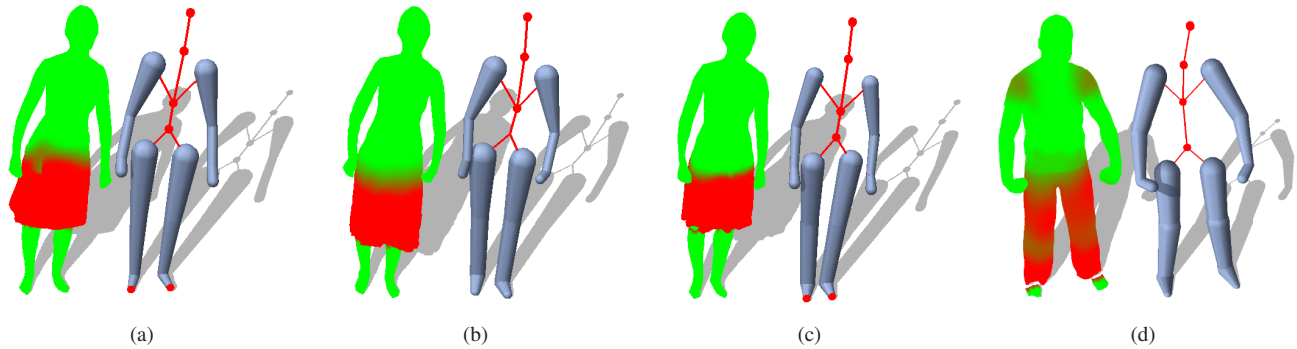
After the surface is defined, a skeleton with 36 degrees of freedom is inserted into the mesh by manually marking the joint positions. Thereafter, the method of [Baran and Popović 2007] is employed to assign each vertex a weight that is used to deform the surface according to the skeleton pose by means of quaternion-blend skinning [Kavan et al. 2007].

The reference performance of an actor is filmed by 8 synchronized and calibrated cameras (40 fps, 1004×1004 pixels) placed in a roughly circular arrangement around the scene. The image silhouettes are extracted by chroma-keying. The multi-view video streams are further processed by a fully-automatic marker-less performance capture method [Gall et al. 2009]. This method jointly captures the pose of the skeleton as well as the non-rigid deformation of the surface mesh (such as folds in clothing) over time, even for people wearing loose apparel that occludes large parts of the body (Fig. 2(b)).

5 Cloth Segmentation and Approximation of Collision Geometry

Skin deformation, as well as the deformation of rather tight clothing, are locally nearly rigid deformations and can be approximated by surface skinning. This does, however, not hold true for surface regions which represent loose non-rigidly deforming pieces of apparel. We therefore analyze the deforming surface geometry of the reference performance and automatically segment it in order to separate loose apparel from largely rigid surface regions whose motions are well described by skinning (Fig. 3).

If the deformation of a body part is approximately rigid, mutual distances of vertices on that part hardly change over time. Based on this insight we derive the following simplified-yet-efficient rule to decide if a vertex $\mathbf{v_i}$ lies on a cloth segment: In the first mesh pose $\mathcal{M}(1)$ of the reference sequence, a ray is shot from vertex $\mathbf{v_i}$ in the negative local normal direction and intersected with the mesh surface on the opposite side. Vertex $\mathbf{v_i}$ is paired with the inter-



(a) (b) (c) (d)

Figure 3: *Cloth segmentation and collision proxies for different clothing styles (red=cloth, green=approximately rigid): (a) skirt s1, dancing sequence; (b) skirt s2, walking seq.; (c) dress s4, dancing seq.; (d) capoeira seq. s5 - cloth regions and transition areas at boundaries are reliably identified. Note that in (a) a mostly rigid woolen cord was correctly excluded from the cloth region (dent in the red area).*

section point $\mathbf{p_i}$, which is described on the intersected face at sub-triangulation accuracy using barycentric coordinates. We now compute the standard deviation $\sigma(\mathbf{v_i}) = \sigma(d(\mathbf{v_i}, \mathbf{p_i}))$ of the Euclidean distance $d(\mathbf{v_i}, \mathbf{p_i})$ over the entire reference sequence. If this standard deviation exceeds a threshold t_{rigid} we conclude that $\mathbf{v_i}$ lies on a non-rigidly deforming surface area likely to be cloth. In other words, we find non-rigidly deforming surface regions by searching for sections with non-persistent cross-sectional areas. This is a weaker criterion than testing preservation of mutual distances between all points on a segment, but is much more efficient to compute and performs well in practice.

Since our final animatable performance model combines physically-based and skeleton-based animation, we need to make sure that the transition between the two modalities is seamless. We transform the $\sigma(\mathbf{v_i})$ values into "*clothness weights*" $\delta(\mathbf{v_i}) \in [0, 1]$:

$$\delta(\mathbf{v_i}) = \text{cl}\left(\frac{\sigma(\mathbf{v_i}) - t_{rigid}}{t_{cloth} - t_{rigid}}\right). \quad (1)$$

Here, the function cl clamps the weights to $[0, 1]$. t_{cloth} represents a threshold on $\sigma(\mathbf{v_i})$ above which vertex motion is fully determined by the physics simulation. The motion of a vertex with $\delta(\mathbf{v_i}) = 0$ is fully-controlled by skeleton motion and skinning, a vertex with $\delta(\mathbf{v_i}) = 1$ is fully cloth model controlled. Vertices with $\delta(\mathbf{v_i}) \in]0, 1[$ lie in the transition zone between cloth and skeleton simulation (e.g., the rim of the skirt and parts of the pants in Fig. 3). In Sect. 6.2 we explain how to blend the two animation types in those areas. To create a smooth clothness distribution and fill in potential holes in the estimation, we also perform a diffusion of the δ's on the surface.

For both the estimation of cloth simulation parameters, as well as for creating new animations, it is essential that collisions of the fabric with the body are plausibly simulated. Unfortunately, the true shape of the body geometry under wide attire is not directly visible and is thus neither part of the captured mesh sequence nor the skeleton model itself. We therefore resort to a statistical mesh model $\mathcal{M}_s$ (Fig. 2(d)) that jointly encodes the space of naked human body shapes and body poses [Hasler et al. 2009a] in a lower-dimensional parameter space. This model has been learned from 550 scans of 114 individuals (59 male, 55 female subjects - aged 17 to 61) in different body poses. In this model, a body mesh of particular shape and pose is parameterized with a vector of principal component coefficients $\mathbf{s}$ such that $\mathcal{M}_s = \mathbf{E} \cdot \mathbf{s} + \mathcal{M}_\mathbf{a}$, with $\mathbf{E}$ being the matrix of eigenvectors (principal components), and $\mathcal{M}_a$ being the average human body.

Unlike [Balan and Black 2008], who fit a statistical model directly to image data, we fit the statistical model to our surface scan such that its shape comes as close as possible to the scan in all areas where the clothness weight is 0. Thereby, we exploit the fact that the model globally encodes the space of body shapes, and fitting it to regions where the underling body is well exposed allows for inference of geometry in occluded regions. Fitting is performed in a semi-automatic way similar to [Hasler et al. 2009b]. First, a few corresponding landmark points on the scan and the statistical model (hands, feet, and head) are marked. Thereafter, a combination of rigid and non-rigid alignment to closest points on the scan is applied, followed by a projection onto the statistical model's PCA space. Unlike [Hasler et al. 2009b] we exploit the cloth segmentation and only use points on our scan that have low clothness weight as fitting constraint. Since the statistical model also encodes pose-dependent deformations of the naked body, it is sufficient to fit it to only one pose of our model. As efficient collision checks are of utmost importance for real-time animation, we approximate the geometry of the statistical surface model in regions occluded by cloth. These conical collision proxies approximate the occluded geometry well and allow for efficient collision checks. Although we do not use the full statistical model for collisions, we use it in new animations for rendering the shape of occluded body parts, such as the legs under the skirt (see Sect. 7).

6 Estimating a Cloth Model

For cloth simulation we use the position-based dynamics approach presented in [Müller et al. 2007]. Many previous cloth simulation methods do not use vertex positions explicitly as part of the simulation state, but compute them by integrating velocities. In contrast, in position-based dynamics vertex positions are explicit members of the simulation state which allows for direct position manipulation during simulation. Internally, simulation results are computed by a combination of Euler integration and a Gauss-Seidel-like solver for constraint projection. The position-based dynamics framework bears a couple of advantages in our problem setting: it allows for plausible real-time simulation, it is very stable, it can handle non-linear constraints, and it allows us to formulate all our simulation constraints using the same mathematical framework. Here, we only briefly review the main concepts of our approach and refer the interested reader to the original paper by Müller et al. for details on the physics-based simulation.

Each piece of cloth is geometrically represented by the respective surface mesh region found by our segmentation approach (Sect. 5). The simulated geometry comprises vertices with positions $\mathbf{v_i}$ and velocities $\mathbf{u_i}$. The masses m_i for each vertex are approximated by

the respective areas on the surface. In addition, the solver expects a set of n_c constraints C_j acting on the positions of the mesh vertices, as well as a set of external forces acting on the fabric $\mathbf{f}$. In analogy to the original paper by Müller et al., we represent each constraint C_j by means of a function defined over a subset of vertex positions. The strength of each constraint is further controlled by a stiffness parameter k_j.

We employ several types of simulation constraints: collision constraints, stretching and bending constraints, as well as additional so-called blend constraints that enable us to seamlessly interpolate between cloth and skeleton simulation, as explained in Sect. 6.1. Stretching constraints take the form

$$C_{stretch}(\mathbf{v_a}, \mathbf{v_b}) = \|\mathbf{v_a} - \mathbf{v_b}\| - l_{a,b}\,, \tag{2}$$

where $l_{a,b}$ is the rest-length of the edge between $\mathbf{v_a}$ and $\mathbf{v_b}$, which is the length of the respective edge in the decimated body scan $\mathcal{M}$. The bending constraint evaluates to

$$C_{bend}(\mathbf{v_a}, \mathbf{v_b}, \mathbf{v_c}, \mathbf{v_d}) = acos(\mathbf{n_l} \cdot \mathbf{n_r}) - \phi\,, \tag{3}$$

where $\mathbf{v_a}$, $\mathbf{v_b}$, $\mathbf{v_c}$, and $\mathbf{v_d}$ are the positions of vertices of two adjacent triangles that share a common edge, and $\mathbf{n_l}$ and $\mathbf{n_r}$ are the respective triangle normals. ϕ is the rest-angle between the two triangle normals.

We use spatial hashing to quickly find collisions and self-intersections, yielding additional collision constraints with fixed stiffness $k_{coll} = 1.0$. Finally, we include so-called blend constraints, which are defined for each vertex with a $\delta(\mathbf{v_a}) \in]0, 1[$ and read

$$C_{blend}(\mathbf{v_a}) = \mathbf{p_a}\,, \tag{4}$$

where $\mathbf{p_a}$ is the position the vertex should reach. The stiffness of this constraint, $k_{blend}(\mathbf{v_a})$, is vertex-specific and explained in more detail in Sect. 6.1.

External forces comprise gravity $\mathbf{f_g}$ and drag $\mathbf{f_d}$. Unlike the original paper by Müller et al., we use a more sophisticated model for air-resistance similar to Bhat et al. [2003]. The air drag force depends quadratically on the velocity in direction of the surface normal, and linearly in tangential direction:

$$\mathbf{f_d}(\mathbf{v_i}) = -0.5A(\mathbf{v_i})(d_n\|\mathbf{u_{i,n}}\|^2\mathbf{n} + d_t\mathbf{u_{i,t}})\,. \tag{5}$$

Here, $A(\mathbf{v_i})$ is the area of the surface patch around the vertex $\mathbf{v_i}$ (in our case equal to m_i). $\mathbf{u_{i,n}}$ and $\mathbf{u_{i,t}}$ are the components of $\mathbf{u_i}$ in direction of normal $\mathbf{n}$ and tangential to it, respectively. The constants d_n and d_t are the drag coefficients.

Friction and restitution are handled by directly manipulating velocities of colliding vertices. To this end, vertex velocities are damped in the direction perpendicular to the collision normal by a factor $1 - k_{friction}$, and reflected in the direction of the collision normal.

For each cloth region on the model, the six parameters

$$\rho = (d_n, d_t, k_{bend}, k_{stretch}, k_{friction})$$

have therefore to be determined, i.e., the two drag coefficients, stretching and bending stiffness, and friction.

In theory, the number of simulation steps n_{iter} per frame of video could be considered a free parameter of our optimization problem as well. However, we aim for a character that can be simulated in real-time, and in real-time applications the number of iterations is usually fixed in order to balance resources. Therefore, we employ the same number of iterations ($n_{iter} = 24$) during both parameter optimization and the creation of novel animations.

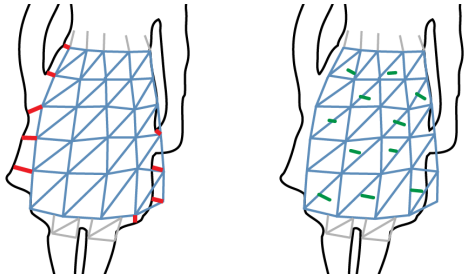


Figure 4: *Two components of the error function on the cloth section (light blue) shown in 2D. (a) silhouette distance error: red lines between reprojected and measured silhouette points; (b) SIFT distance: green lines between predicted SIFT feature locations and measured SIFT feature locations.*

6.1 Combining Cloth and Skeleton Animation

Before explaining how to estimate cloth parameters, we describe how new poses of our final fully-animatable performance model are created. Positions of vertices with $\delta(\mathbf{v_i}) = 0$ are solely determined by the current joint parameters of the skeleton and dual quaternion skinning. Similarly, the positions of all pure cloth vertices with $\delta(\mathbf{v_i}) = 1$ are determined by the physical simulation described in Sect. 6. For blend vertices, $\delta(\mathbf{v_i}) \in]0, 1[$, the new pose is jointly determined by cloth and skeleton simulation. To this end, we apply the blend constraints from Eq.(4), which push the points towards the positions according to the skeleton-based animation with a stiffness of $k_{blend}(v_i) = 1 - \delta(\mathbf{v_i})$.

6.2 Cloth Parameter Estimation

The cloth parameters ρ are found by running a numerical optimization that strives to minimize cloth deformation in the reference sequence. We measure two properties of the simulated animation: **1)** the alignment of the cloth region's silhouette edges with silhouette boundaries in all input images over time, $E_{sil}(\rho, t)$, and **2)** the alignment of the reprojected cloth with robust SIFT features [Lowe 1999] in the interior of the fabric in all camera views, $E_{sift}(\rho, t)$.

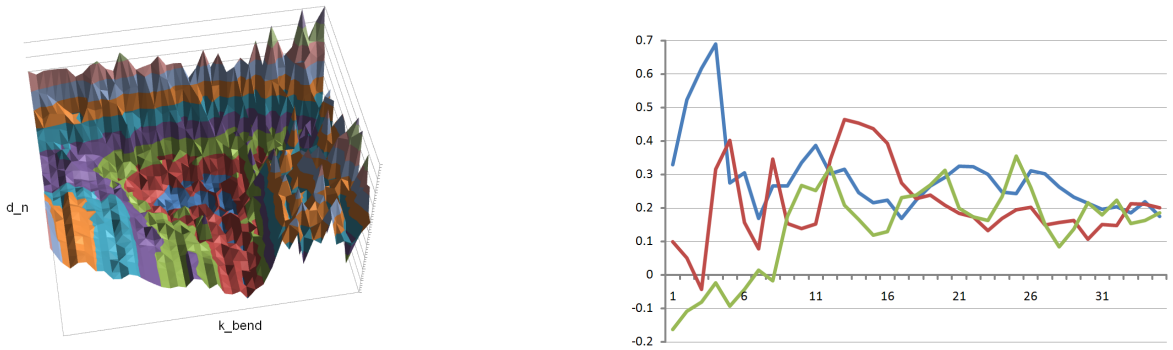
For each time step of the video t and each input camera view c, a certain set of vertices on the cloth segment under consideration should project onto the silhouette boundary of the respective input frame. This set can be easily identified and we call it the set of silhouette rim vertices $\mathcal{V}_{c,t}$ (Fig. 4(a)). The silhouette error then evaluates to

$$E_{sil}(\rho, t) = \frac{1}{n_{sil}} \sum_c \sum_{\mathbf{v} \in \mathcal{V}_{\mathbf{c,t}}} d^2_{im}(\mathbf{q}_{S_{c,t}}(\mathbf{v}), \mathbf{q}'_{S_{c,t}}(\mathbf{v}))\,, \tag{6}$$

where $d_{im}(\mathbf{q}_{S_{c,t}}(\mathbf{v}), \mathbf{q}'_{S_{c,t}}(\mathbf{v}))$ is the image space distance between the reprojection $\mathbf{q}'(\mathbf{v})$ of silhouette rim vertex $\mathbf{v}$ into silhouette image $S_{c,t}$ and the closest boundary point of the measured silhouette in the same image, $\mathbf{q}_{S_{c,t}}(\mathbf{v})$. n_{sil} is the number of all rim vertices from all $\mathcal{V}_{c,t}$ in this frame.

For each time step of the video and each camera view, we also compute a set of SIFT features, $\mathcal{F}_{c,t}$. In addition, we establish correspondences between features in two subsequent time steps for each camera view (Fig. 4(b)). The feature-based energy term is defined as

$$E_{sift}(\rho, t) = \frac{1}{n_{sift}} \sum_c \sum_{\mathbf{e} \in \mathcal{F}_{c,t}} d^2_{im}(\mathbf{o}_{I_{c,t}}(\mathbf{e}), \mathbf{o}'_{I_{c,t}}(\mathbf{e}))\,, \tag{7}$$



(a) (b)

Figure 5: *(a) Plot of the energy function E_{fit} over d_n and k_{bend}. A clear minimal energy area can be observed near the center. (b) Robustness of CMA-ES optimization for a single material: Several optimizations initialized with different particle distributions converge to the correct parameter region after only a few number of particle generations (X-axis: generation; Y-axis: distribution mean shown for normalized $k_{friction}$ component).*

where $\mathbf{o}_{I_{c,t}}(\mathbf{e})$ is the 2D image position of feature $\mathbf{e}$ in image $I_{c,t}$. $\mathbf{o}'_{I_{c,t}}(\mathbf{e})$ is the predicted image position of feature $\mathbf{e}$ at time t and n_{sift} is the number of all features from all $\mathcal{F}_{c,t}$ in this frame. For each camera c, these predicted image positions are obtained from feature positions at the previous time step as follows: Each feature $\mathbf{e}$ at time step $t-1$ is projected back onto the final cloth surface at $t-1$ using the camera matrix of camera c. For each feature $\mathbf{e}$, this yields a 3D position on the surface of the mesh $\mathbf{p_{c,t-1}}(\mathbf{e})$, expressed in barycentric coordinates relative to the enclosing mesh triangle. The position at the current time step t, $\mathbf{p'_{c,t}}(\mathbf{e})$, is predicted from $\mathbf{p_{c,t-1}}(\mathbf{e})$ by the cloth simulation. The predicted image positions $\mathbf{o}'_{I_{c,t}}(\mathbf{e})$ are then obtained by reprojecting $\mathbf{p'_{c,t}}(\mathbf{e})$ back into each respective camera view.

Both error terms are evaluated over the entire reference sequence and their contributions are combined, yielding the overall fitting error

$$E_{fit}(\rho) = \frac{1}{N}\sum_{t=1}^{N}(\alpha E_{sil}(\rho,t) + \beta E_{sift}(\rho,t))\,. \qquad (8)$$

Here, the values $\alpha = 1$ and $\beta = 10$ are empirically determined weights that are kept constant for all our estimations. The combination of silhouette and SIFT features is essential for our goodness-of-fit measure. While silhouette data is important to assess the overall appearance of the cloth boundaries, the SIFT features provide a lot of information on cloth behavior from the inner regions of the cloth in all images.

When evaluating E_{fit}, both E_{sil} and E_{sift} are evaluated for each frame t after the new model pose is determined according to the method in Sect. 6.1. We intentionally formulate our energy function in terms of image features of the original reference sequence, and not in terms of a 3D comparison to the tracked meshes from the reference performance. In this way we stay as close as possible to the measured data and prevent unintended fitting to potential inaccuracies in the tracked 3D reference performance (see Sect. 7 for an 3D RMS error comparison to a synthetic ground truth sequence). In addition, our formulations of E_{sil} and E_{sift} are memory efficient as they only require storage of 2D silhouette rim points and 2D feature locations for each input frame.

The error function E_{fit} is non-convex and exhibits many local minima, as shown in Fig. 5 for two of the six simulation parameters (d_n and k_{bend}). Despite the local extrema, a clear region of minimal error is apparent. This multi-modalness comes as no surprise since the cloth simulation behavior is highly non-linear and causes potentially large changes in geometric cloth appearance for only small changes in ρ. Quasi-Newton or conjugate gradient based optimizers do not work very well on such energy functions. In addition, analytic gradients are not easily available which would make costly numerical approximations necessary.

Evolution Strategy with Covariance Matrix Adaptation To tackle the challenging energy functional, we apply the evolution strategy with covariance matrix adaption (CMA-ES) [Hansen et al. 2009] to find a minimum of the energy function Eq.(8). In the following we briefly describe its core concepts and refer the reader to the original paper for a more in-depth discussion.

CMA-ES is an iterative sampling-based algorithm. In each iteration (called a generation) the energies of a set of random samples from the parameter space ρ are evaluated. The distribution of these samples is determined by a multivariate Gaussian. The samples are then used to re-estimate the covariance matrix of the Gaussian distributions, after which a new generation of samples is generated. By iterating this process, the optimizer explores the search space and finds a local minimum in the energy function.

Intuitively, the CMA-ES algorithm strives to estimate a covariance matrix that provides a local second-order approximation of the underlying energy function (similar to the inverse Hessian in a Quasi-Newton approach) without requiring analytic gradients. Since the covariance matrix update analyzes the best samples from several previous generations and not only the most recent one, the optimization can be performed with a small population. This not only makes optimization faster, it also creates distributions that provide: a) high variance in the best sampling direction, b) an increase of the likelihood of good samples to be redrawn in new generations. In combination this enables sampling in the correct direction, i.e., a placement of samples along the actual local gradient direction rather than a uniform placement of samples around the mean of the distribution. At the same time, the evolution strategy prevents premature convergence to local minima.

In our case, we stop the optimization after 512 function evaluations, or when the variance along all dimensions is smaller than 0.025. We return the mean of the final distribution as the best simulation parameters. The best particle population size per generation is automatically determined and typically equals 9 in our experiments. We initialize all optimizations with the same mean and covariance.

In practice, we rescale all elements of the parameter vector ρ (from the original parameter range shown in Tab. 1) such that the feasible solutions lie within $[0, 1]$. In order to limit the search space without disturbing the sampling procedure at the boundary, we augment the energy function in Eq.(8) with a quadratic penalty term as in [Hansen et al. 2009], yielding our final energy function:

$$E_{CMA}(\rho) = \begin{cases} E_{fit}(\rho) & \rho \in [0,1]^6 \\ E_{fit}(\rho_r) + \omega \cdot \|\rho - \rho_r\|^2 & \text{otherwise} \end{cases},$$

where ρ_r is the nearest boundary value to ρ. Parameter ω is fixed to 16 for all experiments.

When deciding for the best optimizer, we experimented with a simulated annealing (SA) approach as well as CMA-ES. Although we were able to achieve plausible material parameters with both approaches, convergence of SA depends strongly on the chosen annealing strategy parameters, whereas CMA-ES automatically adjusts its strategy. In our case, simulated annealing usually requires $2 - 4$ times more sample evaluations than the evolutionary approach. Further, CMA-ES proves to be very robust and stably converges to a plausible solution after relatively few sample generations, even if the initial particle distribution varies a lot (Fig. 5(b)).

7 Results and Discussion

We captured 14 performances using the method described in Sect. 4, each comprising between 400 and 1100 frames. In those sequences, a male and a female actor wear five different types of apparel. The female subject wears: a skirt with medium thickness (s1), a long skirt made from a lighter material (s2), a medium length skirt of a comparably stiff material (s3), and a dress (s4). For the male subject we reconstructed a capoeira sequence in loose pants (clothing style s5). For each attire, we captured several different sequences with motion styles ranging from walking and dancing to kicking and turning in the capoeira scene.

All renderings of new animations in the paper and the accompanying video were created by real-time transfer of the pose of the decimated mesh $\mathcal{M}$ used for simulation to the high resolution body scan $\mathcal{M}_{high}$ using a method similar to [de Aguiar et al. 2008b]. In the following, we discuss several aspects of our approach.

Segmentation and Cloth Parameter Estimation For each style of apparel, we choose one reference sequence from which to reconstruct an animatable performance, Table 1. For skirt s1, we also reconstructed animatable models from two different reference motions (walking and dancing) to verify stability. Our segmentation approach was able to reliably identify the loose cloth areas in all reference scenes using a fixed set of segmentation thresholds (Fig. 3). The segmentation of s1 in two different reference sequences is almost identical, which demonstrates the stability of our method. The segmentation of the pants in the capoeira sequence shows that our algorithm also faithfully handles tighter clothing which is less wavy than the skirts. Our algorithm correctly identifies the legs of the trousers and the sleeves of the t-shirt as slightly wider apparel (Fig. 3(d)).

Table 1 lists the set of cloth simulation parameters which we estimated for the different clothing styles. As best shown in the video, the visual draping characteristics of all reconstructed fabrics closely match the draping behaviors of their real world counterparts. For instance, the stiffest skirt s3 (Fig. 6(top)) produces far less folds than the light dress s4 (Fig. 6(bottom)) which moves rather vividly.

Creating New Animations Arbitrary new motions of reconstructed models can be created in only a few minutes by applying new skeletal motion data. One source of new motion data used in this paper are additional sequences measured with the method from Sect. 4, i.e., we captured additional motions not used as reference performances. For some animations we used one of 15 motion capture files that were downloaded from an online motion capture database and retargeted. Note that arbitrary motion capture files or key-frame animations by an animator would be equally feasible. In all new animations, both the overall appearance of the body and the deformation behavior of the cloth are very lifelike since even subtle deformation details are realistically reproduced (Fig. 6 and accompanying video). It is also easily possible for an artist to modify material parameters, or to apply a new texture to the newly created animations, as shown in Fig. 1.

The mesh pre-processing step described in Sect. 4 may lead to holes in the geometry (e.g., underneath the skirt), that may become visible during new animations. Therefore, for visualization we render the respective parts of the statistical body model to approximate the shape of occluded body geometry (such as the legs) and thereby close the holes.

As mentioned before, real-time rendering performance is maintained by rendering new poses based on the decimated surface model $\mathcal{M}$ and real-time pose transfer to the high-resolution scan $\mathcal{M}_{high}$. Using $\mathcal{M}_{high}$ directly for both reconstruction and rendering leads to slightly more detailed results but precludes real-time rendering.

Validation and User Study Our goal is to conveniently capture models that enable realistic and physically plausible simulations in real-time, but not models with material parameters of strict physical accuracy. Therefore, comparison of our estimated cloth parameters to material properties measured with material science test methods [Kawabata 1980] is not a reasonable way to validate our reconstructions. Instead, we apply our reconstruction method to a synthetic video data set that was created by rendering one of our performance models fitted with a synthetic texture back into the input camera views. Our estimated material parameters $\rho_{est} = (5.4, 5.6, 0.75, 1.0, 0.47)$ are similar to the ground truth set used for creating the sequence $\rho_{art} = (4.0, 4.9, 0.7, 1.0, 0.5)$. As shown in the video, the visual appearance of the reconstruction is almost indistinguishable from ground truth. This is also numerically confirmed by the low average 3D RMS vertex distance error of 0.6 cm between the cloth parts of the data sets over time.

We also performed a user study with 52 participants to validate that our system is able to plausibly reproduce the behavior of a specific set of clothes. Each of the participants was shown a web page containing three videos. Each video contained one input camera view of the dancing sequence of skirt s3 and a simulated 3D model generated with our algorithm from the same viewpoint. One video (video C) showed the simulation parameters found by our estimation algorithm (Table 1), while the other two videos used different material parameters taken from the other reconstructed skirts (see additional material). In a first experiment, participants were asked to evaluate how similar the skirt's fabric in our animated model behaves in comparison to the input video. Five options were given to assess the similarity to the input video in percent: totally different - 0%, different - 25%, looks similar - 50%, almost the same - 75%, the same - 100%. The results were: 56.73% (A), 54.32% (B), and 67.30% (C). The result confirms that the majority of participants correctly identified video C (generated with our estimated parameters) as most similar to the real skirt (ANOVA $p < 0.01$). A similar result was achieved by our second experiment, where we asked the subjects to rank the simulations according to the closeness between simulated cloth behavior and reference input (1=most similar, 3=most dissimilar). Here, 27 subjects correctly ranked video

sequence	set name	d_n $[0, 32]$	d_t $(\times 10^3)$ $[0, 32]$	k_{bend} $[0, 1]$	$k_{stretch}$ $]0, 1]$	$k_{friction}$ $[0, 1]$	$\#frames$
s1a: dancing	ρ_{1a}	7.054	0.119	0.195	1.0	0.129	1000
s1b: walking	ρ_{1b}	5.198	0.553	0.134	1.0	0.401	800
s2: walking	ρ_2	0.876	1.603	0.575	1.0	0.790	1100
s3: dancing	ρ_3	0.774	0.452	0.888	1.0	0.070	600
s4: dancing	ρ_4	1.184	0.621	0.100	1.0	0.228	800
s5: capoeira	ρ_5	1.727	3.587	0.382	1.0	0.558	400

Table 1: *Estimated cloth simulation parameters for the different skirts, the dress, and the pants, each made of a different fabric.* ρ_{1a} *and* ρ_{1b} *were estimated for the same skirt from different reference sequences.*

C as the closest match (Wilcoxon test, $p < 0.05$). In both experiments, the tests (for the differences in mean) show that both results are clearly statistically significant. Therefore, we conclude that we can capture and simulate cloth behavior in a plausible way.

Timings and Practical Considerations It takes around 2.5 hours to reconstruct a reference performance of 1000 frames on an Intel Core 2 Duo 3.0 GHz with the approach from Sect. 4. All sequences we captured, both reference sequences and skeleton motions only used for new animations, were tracked fully-automatically. Segmentation of cloth takes approximately 5 seconds. Using a single-threaded implementation for computations, estimation of cloth simulation parameters takes approximately $3-5$ hours for a reference sequence consisting of roughly 1000 frames. There is room for notable speedups through parallel parameter sample evaluation (Sect. 6.2) which we leave for future work. Overall, the entire reconstruction requires very little manual interaction, e.g., fitting of the statistical body model or the kinematic skeleton only.

The actual computation of new poses and the transfer to the higher resolution mesh of about $25k$ triangles run in real-time at 60 fps on the same machine. The cloth simulation currently runs with the same step size as the capture rate of our camera system (i.e., 40 fps). However, it would be a straightforward process to decouple simulation and animation frame rates by expressing fractional steps using linear blending between two updates, as is usually done in real-time simulation scenarios.

Applications Character animators for real-time graphics applications like games and networked virtual environments will benefit from our integrated acquisition approach. For instance, our algorithm could be used to capture animatable avatar models, such as personalized game characters. In most cases, our method will not immediately deliver the final result to be used in the application, but provides animators with detailed models that can be conveniently adjusted to their needs in little time. Our system is thus a powerful add-on to the animator's toolbox.

Design Choices Our approach requires careful selection and adaptation of existing solution strategies, as well as the design of new algorithmic solutions to many challenging sub-problems.

We tested several alternative strategies for cloth segmentation, such as: 1) Analyzing the variation of surface to skeleton bone distance over time, or 2) thresholding the distance of the surface scan (with apparel) to the fitted statistical body model which is used for collision proxy creation. Our proposed method from Sect. 5 is much simpler to implement and produces better results than any of the above. Alternative 1) frequently leads to implausible segmentations were the skeleton model does not reflect the anatomical reality faithfully, such as in the shoulder area. Alternative 2) fails since motion information is not used and actually loose apparel may happen to lie flat on the body in the scanned pose. As an additional benefit our segmentation approach would also work for reference performances captured with purely surface-based capture approaches, such as [de Aguiar et al. 2008b].

In contrast to related approaches from the literature, e.g., [Bhat et al. 2003], our cloth parameter fitting uses rich 3D feature information captured from around a subject wearing an actual piece of apparel, not only data from a sheet of cloth. Our input data therefore reflect the real cloth behavior in interaction with the body more realistically. Moreover, our energy function spatio-temporally evaluates the quality of the simulation dynamics (by evaluating predicted feature locations). Both of these aspects are important for the quality of our results.

Discussion Our approach is subject to some limitations. We cannot handle apparel that shows arbitrary topology changes while moving, such as the opening up of a coat. Additional manual post-processing is needed in this case. Further on, the quality of cloth segmentation and the plausibility of estimated cloth parameters is dependent on the type of reference sequence captured. If there is insufficient motion in the cloth, we will not be able to identify wavy pieces of apparel automatically. However, this is not a principal limitation since in a general recording scenario it is trivial to instruct the person to move in such a way that the motion of all pieces of apparel are sufficiently prominent.

As seen in the error plot (Fig. 5(a)), the energy function has several local minima near the global minimum. Each of them has an error value very close to the minimal error. Any simulation run with parameters from this restricted region will be perceptually almost indistinguishable from the optimum according to our energy function. While we cannot guarantee to find the global optimum our optimization will typically find one of the near local minima that reproduces the cloth behavior in a perceptually plausible way. The restricted regions of plausible parameters are clearly distinct for different materials. This enables our algorithm to faithfully disambiguate between materials with different physical properties. This has been confirmed in our user study.

We estimated the parameters for the same skirt s1 from two different reference sequences (a walking sequence and a dancing sequence, see parameter sets ρ_{1a} and ρ_{1b} in Table 1). While the parameter sets are not identical, cross-validation against parameter sets of s1 to s5 confirms that ρ_{1a} and ρ_{1b} represent the same material. If the reference sequence does not demonstrate the behavior of the material in a reasonable way (for example when recording a scene without significant motion), the optimization will produce a different parameter set than for a highly dynamic scene of the same apparel. This can be fixed by properly instructing the performer.

Currently, tracking errors in the skeleton estimation phase of our pipeline can cause jittering artifacts in the final results. In the supplemental video this can sometimes be seen around the hands and

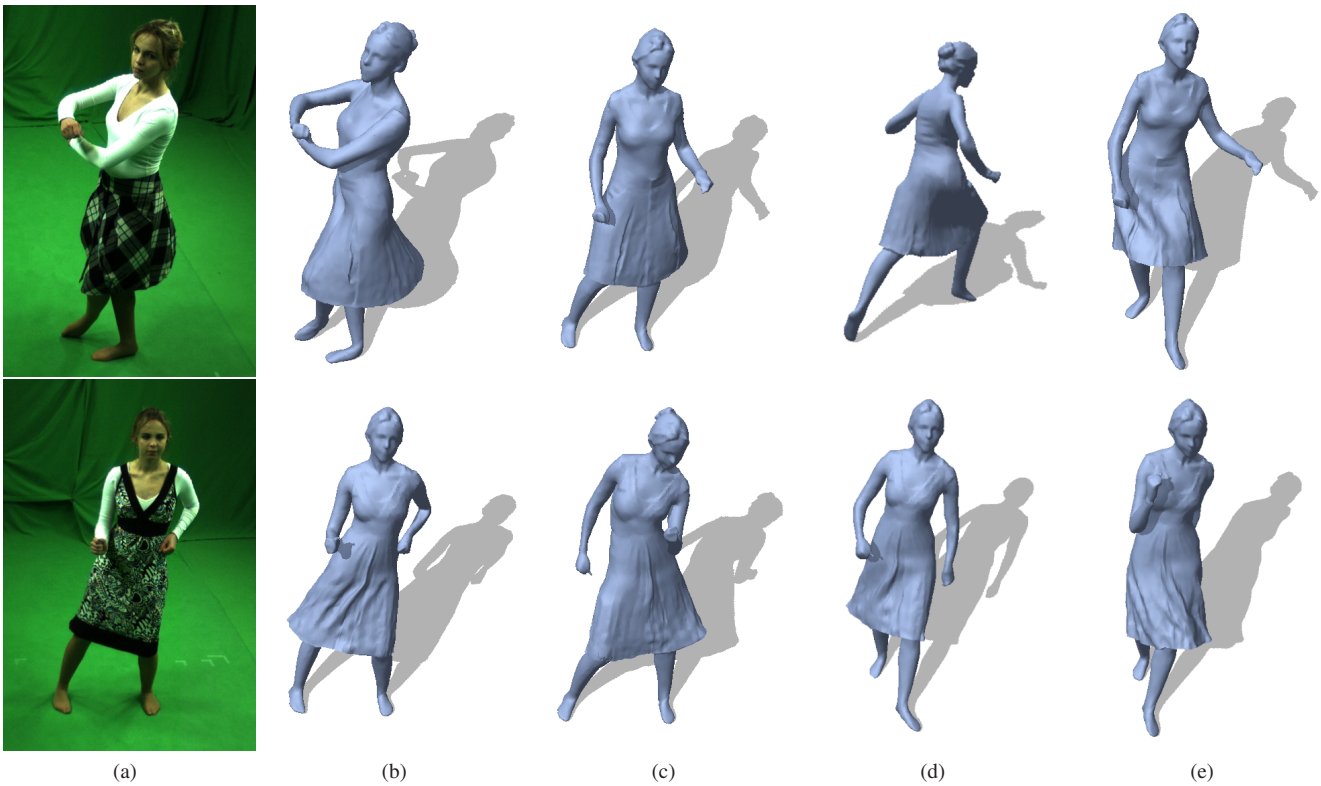


Figure 6: *Results for skirts s3, and the dress model s4. Column (a) input frame from reference sequence; (b) the reconstructed animatable model corresponding to the reference frame; (c)-(e) subsequent frames of a newly created animation for which only motion parameters of the underlying skeleton were given (see also accompanying video).*

hips. These errors mostly occur because of silhouette ambiguities (including actual errors in the segmentation and differences in pose between 3D scan and video like open hand versus closed hand), or missing image features in some parts of the videos. Adding additional constraints like joint limits or smoothing trajectories would alleviate some of these issues. The tracking errors may have an impact on the estimated cloth parameters and it would be interesting to evaluate this influence in more detail as part of future work. Nevertheless, even in the presence of these artifacts we are able to estimate good simulation parameters that produce visually convincing results.

Our method estimates only a single set of material parameters for all cloth regions of a character. In future work we plan to estimate more fine-grained spatially-varying properties, so that structural features in cloth, such as seams, can be represented more faithfully. Our assumption that non-rigidly moving regions are cloth is reasonable for most types of scenes. Other simulation models would be required to properly model hair or strongly deforming tissue.

The cloth simulation parameters are not independent of the step size and the number of iterations. We always have to estimate parameters for exactly the same simulation settings that we also use for creating new animations. We have experimentally explored that dependency by animating skirt s1 with lower numbers of simulation iterations than used for reconstruction. Even if we reduce the number of iterations by 25%, the 3D RMS vertex distance error to the animation with the correct number of iterations is only 1.8 cm (for the cloth region). This corresponds to only minor visual differences. Therefore, it is fair to assume that within this range of iterations the same parameters can be used.

It was our goal to develop a convenient method for capturing human character models of much higher quality and detail than the models usually found in today's real-time graphics applications. Even though our reconstructed materials are not physically accurate, our results show that highly plausible new animations can be rendered in real-time. Admittedly, the reconstructions shown in this paper would probably fall short of the quality requirements of movie productions. However, we are confident that the basic estimation principle scales to such scenarios and movie animators may still find our reconstructions helpful as a starting point. Some quality gains can already be obtained by giving up real-time performance and using the high-resolution meshes throughout the pipeline, as well as by using the complete statistical body model as collision geometry.

8 Conclusion

We have presented a new integrated method for reconstruction of a fully-animatable model of a person in general apparel from multi-view video that incorporates a physically-based cloth model. Our human model is a rigged character with skeleton, skinning weights and cloth parameters, and is designed for real-time rendering. All elements of our animation model have been estimated from data with only minimal manual interaction. In contrast to previous performance capture methods, we can now also create new real-time animations of performers captured in general apparel. Each aspect of the model can be conveniently modified. New animations can be rendered in real-time and look as realistic and physically plausible as if they were captured from the real subject. The high quality of our reconstructions and new animations has been confirmed on a variety of test data sets, as well as through a user study.

Acknowledgements

We would like thank Nils Hasler for help with the statistical body model, Tatjana Feldmann for her performance, and the anonymous reviewers for their helpful comments. We would also like to thank CMU for their support with the audio for the supplemental video. This work has been developed within the Max-Planck-Center for Visual Computing and Communication collaboration, and has been co-financed by the Intel Visual Computing Institute.

References

ALLEN, B., CURLESS, B., POPOVIĆ, Z., AND HERTZMANN, A. 2006. Learning a correlated model of identity and pose-dependent body shape variation for real-time synthesis. In *Proc. of SCA*, 147–156.

ANGUELOV, D., SRINIVASAN, P., KOLLER, D., THRUN, S., RODGERS, J., AND DAVIS, J. 2005. SCAPE: Shape completion and animation of people. In *ACM TOG (Proc. SIGGRAPH '05)*.

BALAN, A. O., AND BLACK, M. J. 2008. The naked truth: Estimating body shape under clothing. In *ECCV (2)*, 15–29.

BARAN, I., AND POPOVIĆ, J. 2007. Automatic rigging and animation of 3d characters. *ACM TOG (Proc. SIGGRAPH '07)*.

BHAT, K. S., TWIGG, C. D., HODGINS, J. K., KHOSLA, P. K., POPOVIĆ, Z. Z., AND SEITZ, S. M. 2003. Estimating cloth simulation parameters from video. In *Proc. of SCA*.

BRADLEY, D., POPA, T., SHEFFER, A., HEIDRICH, W., AND BOUBEKEUR, T. 2008. Markerless garment capture. *ACM TOG (Proc. SIGGRAPH '08)*.

CARRANZA, J., THEOBALT, C., MAGNOR, M., AND SEIDEL, H.-P. 2003. Free-viewpoint video of human actors. In *ACM TOG (Proc. SIGGRAPH '03)*.

CHOI, K.-J., AND KO, H.-S. 2005. Research problems in clothing simulation. *Computer-Aided Design 37*, 6, 585–592.

DE AGUIAR, E., THEOBALT, C., THRUN, S., AND SEIDEL, H.-P. 2008. Automatic conversion of mesh animations into skeleton-based animations. *Proc. Eurographics EG'08*.

DE AGUIAR, E., STOLL, C., THEOBALT, C., AHMED, N., SEIDEL, H.-P., AND THRUN, S. 2008. Performance capture from sparse multi-view video. In *ACM TOG (Proc. SIGGRAPH'08)*.

GALL, J., STOLL, C., DE AGUIAR, E., THEOBALT, C., ROSENHAHN, B., AND SEIDEL, H.-P. 2009. Motion capture using simultaneous skeleton tracking and surface estimation. In *Proc. IEEE CVPR*.

HANSEN, N., NIEDERBERGER, A. S. P., GUZZELLA, L., AND KOUMOUTSAKOS, P. 2009. A method for handling uncertainty in evolutionary optimization with an application to feedback control of combustion. *IEEE TEC 13*, 1.

HASLER, N., STOLL, C., SUNKEL, M., ROSENHAHN, B., AND SEIDEL, H.-P. 2009. A statistical model of human pose and body shape. *CGF (Proc. Eurographics 2009) 28*, 2.

HASLER, N., STOLL, C., ROSENHAHN, B., THORMÄHLEN, T., AND SEIDEL, H.-P. 2009. Estimating body shape of dressed humans. *Comput. Graph. 33*, 3, 211–216.

JAMES, D. L., AND TWIGG, C. D. 2005. Skinning mesh animations. *ACM TOG (Proc. SIGGRAPH'05)*.

KAVAN, L., COLLINS, S., ŽÁRA, J., AND O'SULLIVAN, C. 2007. Skinning with dual quaternions. In *Symposium on Interactive 3D graphics and games*, 39–46.

KAWABATA, S. 1980. *The standardization and analysis of hand evaluation*. Textile Machinery Society of Japan.

KIRCHER, S., AND GARLAND, M. 2006. Editing arbitrarily deforming surface animations. In *ACM TOG (Proc. SIGGRAPH'06)*.

LOWE, D. G. 1999. Object recognition from local scale-invariant features. In *Proc. ICCV*, vol. 2, 1150ff.

MATUSIK, W., BUEHLER, C., RASKAR, R., GORTLER, S., AND MCMILLAN, L. 2000. Image-based visual hulls. In *ACM TOG (Proc. SIGGRAPH'00)*.

MÜLLER, M., HEIDELBERGER, B., HENNIX, M., AND RATCLIFF, J. 2007. Position based dynamics. *Journal Vis. Com.*.

PARK, S. I., AND HODGINS, J. K. 2008. Data-driven modeling of skin and muscle deformation. *ACM TOG (Proc. SIGGRAPH'08)*.

POPA, T., ZHOU, Q., BRADLEY, D., KRAEVOY, V., FU, H., SHEFFER, A., AND HEIDRICH, W. 2009. Wrinkling captured garments using space-time data-driven deformation. *Computer Graphics Forum (Proc. Eurographics) 28*, 2.

POPPE, R. 2007. Vision-based human motion analysis: An overview. *CVIU 108*, 1-2, 4–18.

PRITCHARD, D., AND HEIDRICH, W. 2003. Cloth motion capture. In *Proc. Eurographics EG'03*.

SAND, P., MCMILLAN, L., AND POPOVIĆ, J. 2003. Continuous capture of skin deformation. *ACM TOG (Proc. SIGGRAPH'03)*.

SCHOLZ, V., STICH, T., KECKEISEN, M., WACKER, M., AND MAGNOR, M. 2005. Garment motion capture using color-coded patterns. In *Proc. Eurographics EG'05*.

SHI, X., ZHOU, K., TONG, Y., DESBRUN, M., BAO, H., AND GUO, B. 2008. Example-based dynamic skinning in real time. *ACM TOG (Proc. SIGGRAPH '08)*.

STARCK, J., AND HILTON, A. 2007. Surface capture for performance based animation. *IEEE CGAA 27(3)*, 21–31.

VLASIC, D., BARAN, I., MATUSIK, W., AND POPOVIĆ, J. 2008. Articulated mesh animation from multi-view silhouettes. *ACM TOG (Proc. SIGGRAPH '08)*.

VLASIC, D., PEERS, P., BARAN, I., DEBEVEC, P., POPOVIĆ, J., RUSINKIEWICZ, S., AND MATUSIK, W. 2009. Dynamic shape capture using multi-view photometric stereo. In *ACM TOG (Proc. SIGGRAPH Asia '09)*.

WASCHBÜSCH, M., WÜRMLIN, S., COTTING, D., SADLO, F., AND GROSS, M. 2005. Scalable 3D video of dynamic scenes. In *Proc. Pacific Graphics*, 629–638.

WHITE, R., CRANE, K., AND FORSYTH, D. 2007. Capturing and animating occluded cloth. In *ACM TOG (Proc. SIGGRAPH'07)*.

XU, W., ZHOU, K., YU, Y., TAN, Q., PENG, Q., AND GUO, B. 2007. Gradient domain editing of deforming mesh sequences. In *ACM TOG (Proc. SIGGRAPH '07)*.

ZITNICK, C. L., KANG, S. B., UYTTENDAELE, M., WINDER, S., AND SZELISKI, R. 2004. High-quality video view interpolation using a layered representation. *ACM TOG (Proc. SIGGRAPH '04)*.