RECONSTRUCTING HUMAN SHAPE AND MOTION FROM MULTI-VIEW VIDEO

Edilson de Aguiar, Christian Theobalt, Marcus Magnor and Hans-Peter Seidel

MPI Informatik, Saarbrücken, Germany.
{edeaguia,theobalt,magnor,hpseidel}@mpi-sb.mpg.de

Keywords: image based modeling, shape reconstruction, mesh deformation, free-viewpoint video

Abstract

In model-based free-viewpoint video, a detailed representation of the time-varying geometry of a real-word scene is used to generate renditions of it from novel viewpoints. In this paper, we present a method for reconstructing such a dynamic geometry model of a human actor from multi-view video. In a two-step procedure, first the spatio-temporally consistent shape and poses of a generic human body model are estimated by means of a silhouette-based analysis-by-synthesis method. In a second step, subtle details in surface geometry that are specific to each particular time step are recovered by enforcing a color-consistency criterion. By this means, we generate a realistic representation of the time-varying geometry of a moving person that also reproduces these dynamic surface variations.

1 Introduction

In a free-viewpoint video, the 3D appearance of a real-world scene is represented in such a way that artificial renditions of the scene from arbitrary novel viewpoints can be generated. To this end, model-based free-viewpoint video methods employ explicit models of dynamic scene geometry and dynamic surface texture. Typically, these dynamic scene models are directly reconstructed from multi-view video footage, or an a priori model is adapted to match the scene appearance in input image data. We have demonstrated that a model-based algorithm, which simultaneously captures the shape, the motion and the texture of a moving person from multi-view video, can generate highly realistic free-viewpoint renditions of human actors [3]. In our original method, an enhanced texturing approach enabled us to create realistic scene renditions although only approximate scene geometry was available that could not model subtle time-varying shape details. To bridge this gap, we present in this paper a method that also reconstructs these dynamic geometry details on the body model from multi-view video.

To serve this purpose, we have developed a two-step algorithm that modifies the pose and the surface geometry of a template human body model such that it matches the appearance of a human actor in all time steps of input video footage. In the first step, a set of anthropomorphic shape and pose parameters are estimated via a silhouette-based analysis-by-synthesis approach to obtain a spatio-temporally silhouette-consistent body representation. In a second step, subtle dynamic changes in surface geometry that are not captured by the silhouette-consistent model are reconstructed via color-consistency-based mesh deformation.

2 Related Work

If convincing novel viewpoint renditions of a dynamic real world scene shall be generated, a representation of the time-varying scene geometry is required. Many approaches have been proposed in the literature to reconstruct such representations of arbitrary scenes from image or video data. Shape-from-silhouette methods reconstruct the geometry from multi-view silhouette images or video streams. Examples are image-based [12, 18] or polyhedral visual hull [11] methods, as well as approaches performing voxel-based [13] or point-based [6] reconstruction. Stereo methods have also been applied to reconstruct and render dynamic scenes [19, 8].

If the moving subject in the scene is a person, structural knowledge about the human body can be exploited to facilitate shape reconstruction. In [1], laser range scans of a human torso and the arms are employed to reconstruct a model of body deformation. The poses of a kinematic skeleton are measured by means of a marker-based motion capture system. In [17], a model of surface deformation for the complete human body is reconstructed by jointly employing a marker-based motion capture method and multi-view silhouette matching. Both approaches produce highly detailed deformable bodies, but their commitment to a marker-based motion capture system makes it hard to use them for 3D video reconstruction.

More similar to our method is the work presented in [15], where a sophisticated body model comprising a kinematic skeleton and a deformable surface geometry is fitted to video data. The work presented in [7] is also closely related to our approach. Here a procedure based on silhouette, stereo and feature correspondence information is described, which is able to reconstruct a human model from multiple camera views.

In our work, we extend the approach presented in [3] such that it also captures subtle time-varying geometry details in a generic human body model. To this end, we do not only resort

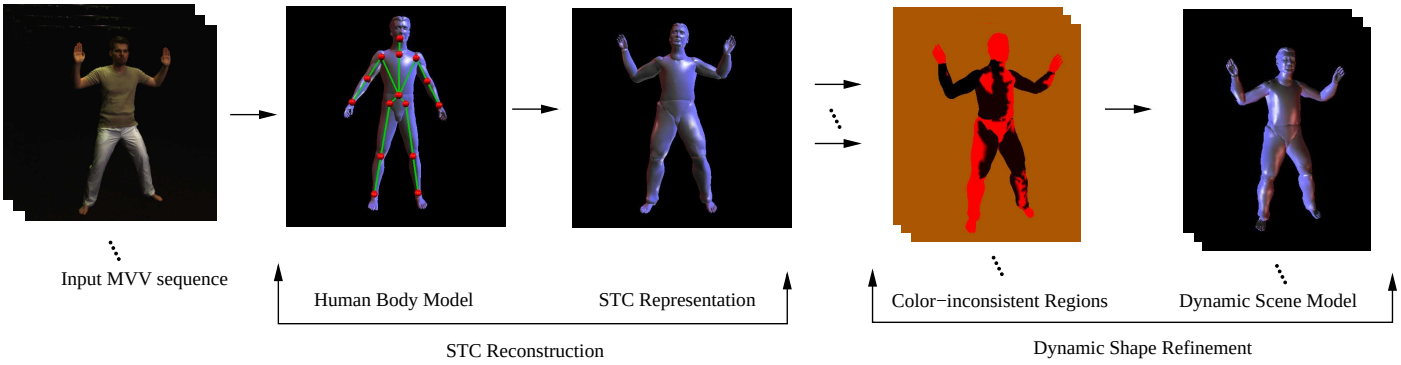


Figure 1: Visualization of the interplay of the individual processing steps in our approach.

to multi-view silhouette matching, as we have done it before, but also make use of a color-consistency criterion. In principle, subtle geometry inaccuracies in a silhouette-fitted body model can be compensated in rendered free-viewpoint videos by applying a clever texture generation scheme. Nonetheless, it can further improve the visual quality of a 3D video if a more detailed geometry representation is available that captures the time-varying body shape more accurately.

3 Overview

Fig. 1 illustrates the algorithmic workflow between the main components in our approach. The system expects multiple synchronized video streams (so-called multi-view video or MVV streams) of an arbitrarily moving person as input (Sect. 4). Our scene representation is based on a generic human body model that comprises a triangle mesh surface representation and an underlying kinematic skeleton (Sect. 5). We employ a two-step procedure to reproduce the shape and the pose of the actor in the real world at each time step of video with our body model.

In the first step (Sect. 6), we apply a silhouette-based analysis-by-synthesis method to estimate the correct body pose of the model at each time step of video. We also estimate body shape parameters that make the model globally consistent with the person's silhouettes at each time step. Unfortunately, this spatio-temporally consistent scene representation (STC representation) does not reproduce subtle time-varying changes in the geometry of the body surface.

We recover these dynamic geometry variations by computing appropriate vertex displacements for each time step of video separately (Sect. 7). To this end, we jointly employ a color- and silhouette-consistency criterion to identify slightly inaccurate surface regions of the body model which are then appropriately deformed by means of a Laplacian interpolation.

The output of our method is a dynamic scene model for a moving human actor that also reproduces time-varying geometry variations (Sect. 8).

4 Multi-view Video Recording

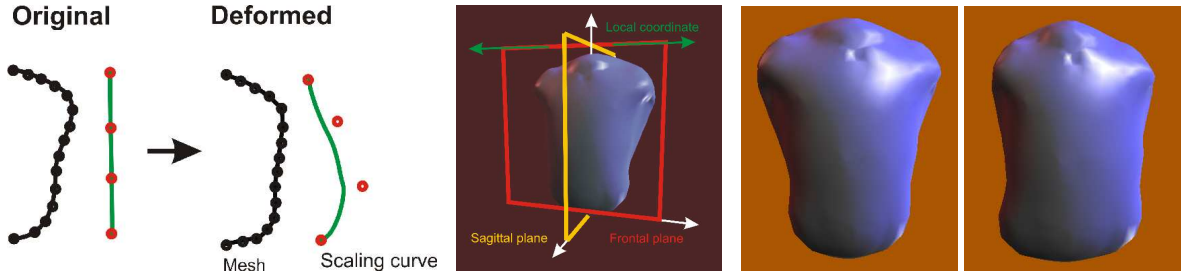
The multi-view video (MVV) sequences used as input to our approach are recorded in our multi-view studio. It enables us to capture an area of approximately 4x4x3m with eight frame-synchronized video cameras, which are placed in a convergent arrangement around the center of the scene. Imperx^{TM} MDC-1004 cameras are employed, featuring a 1004x1004 CCD sensor and delivering 25 fps. The cameras are calibrated into a common coordinate frame. Color-consistency across cameras is ensured by applying a color-space transformation to each camera stream.

For each MVV sequence that serves as input to our method, the person first strikes an initialization pose for a short moment, and thereafter is free to move arbitrarily. In a post-processing step, the silhouette of the person in each frame is extracted via color-based background subtraction.

5 An Adaptable Human Body Model

We employ a template human body model whose shape and proportions can be customized in order to optimally reproduce the appearance of a person in the real world (Fig. 3a). The kinematics of the model are represented by means of a skeleton comprising 16 segments and 17 joints that provide 35 pose parameters in total. The surface geometry of each segment is represented via a closed triangle mesh.

Two sets of anthropomorphic shape parameters are provided which are modified during reconstruction of a spatio-temporally consistent scene representation (Sect. 6). The first set consists of a uniform scaling parameter for each bone. The second set of parameters consists of four sets of control values



(a) (b) (c)

Figure 2: (a) Scaling curve before and after deformation; (b) Local sagittal and frontal planes for a body segment; in each of these planes two scaling curves are defined that scale the segment's geometry in opposite directions; (c) Torso segment before (left) and after (right) free-form deformation.

for each segment, which define one-dimensional B-spline scaling curves that enable free-form geometry deformation.

In addition to the anthropomorphic parameters, the model also provides additional shape refinement parameters that enable a more detailed shape control. For each vertex, a displacement in the direction of the local normal can be specified. The per-vertex displacement parameters are modified during our dynamic shape refinement procedure (Sect. 7).

6 Spatio-temporally Consistent Shape Reconstruction

In the first major step of our dynamic scene model reconstruction method, we estimate a spatio-temporally consistent geometry description, a so-called STC representation. The STC representation has two characteristic features: Firstly, the pose of the generic human body model matches the pose of the real-world actor at each time step of video. Secondly, a constant set of anthropomorphic shape parameters for the body has been identified such that the model is multi-view silhouette-consistent in several body poses.

To reconstruct the STC representation, we extend the silhouette-based analysis-by-synthesis approach originally proposed in [3]. This method performs an optimization search in the anthropomorphic shape and the pose parameters in order to maximize the overlap between the silhouette of the reprojected model and the image silhouette in all camera views. The energy function EF that numerically assesses this overlap sums up the number of set pixels in binary XOR images between image and model silhouettes from all camera perspectives. By this means, we first customize the skeleton dimensions of the template human body model (Sect. 6.1). Thereafter, in an iterative procedure we find an optimal set of pose parameters for each time step of video and a globally optimal set of B-spline scaling parameters (Sect. 6.2).

6.1 Adapting the Skeleton

To customize the bone lengths of the default skeleton we only employ the eight input frames depicting the person in the initialization pose. Our skeleton rescaling method is an iterative procedure that alternates between an optimization of pose parameters and an estimation of uniform scaling parameters.

In the first step of each iteration the uniform scaling parameters of all body segments are adjusted. The second step of each iteration uses the rescaled body model and computes an estimate of the body pose parameters. These two steps are repeated several times. Fig. 3b shows the model after the kinematic skeleton has been shape-adapted to the proportions of our test subject.

6.2 Joint Pose Estimation and Spatio-temporal Free-form Deformation

We have developed a novel spatio-temporal free-form deformation scheme that deforms the individual segment geometries until they are in accordance with the actor's body in multiple body poses. The deformation of each individual segment is controlled by the anthropomorphic B-spline parameters. For each of the 16 triangle meshes, four local B-spline curves are defined. Each of these curves scales the geometry in one specific local coordinate direction (Fig. 2a). Two of the scaling curves deform the mesh in two opposite scaling directions in the sagittal plane, the two others do the same in the local frontal plane (Fig. 2b).

The geometry of one individual segment is deformed by simultaneously finding four optimal sets of N local control values. Each set of control values specifies one of the local scaling curves. The criterion that guides the

optimization search is the previously mentioned silhouette-XOR energy function, EF. We have tested two numerical optimization schemes to minimize this error function in the deformation parameters, the LBFGS-B method [2] and Powell's method [16]. Both methods tend to converge to similar solutions but the former one is preferable since it converges much faster.

We also performed experiments to determine the best number of control values to be used for each scaling curve. Many control values can capture much of the details of a body segment. On the other hand, the optimization takes longer and often does not converge. Therefore, we have decided to use only 4 control values for each scaling curve ($N = 4$), totalling 16 parameters for each segment, which is a good compromise between optimization speed and shape modeling precision. Fig. 2c shows the torso segment before and after free-form deformation.

We have developed a spatio-temporal optimization procedure that employs the previously described principle to shape-adapt the geometry of all body segments. It allows us to robustly infer deformation values that correctly reproduce the geometry of an actor not only in one but in several body poses. Since the stance of the skeleton changes over time, we apply a two-step iterative procedure that alternates between pose determination and segment deformation.

In the first step of each iteration, the pose parameters of the model at each time step of video are estimated using the silhouette-based analysis-by-synthesis approach.

In the second step, the B-spline control values for each of the 16 segments are computed by means of the previously described optimization scheme. To this end, K time steps of video are automatically selected out of the M time steps that the input video sequence contains. We find scaling parameters that optimally reproduce the shape of the segments in all of these K body poses simultaneously. A modified energy function EF_R sums over the silhouette-XOR contributions EF_I at each of these K time steps, $EF_R = \sum_{I=1}^{K} EF_I$. We do not optimize the shape of all 16 segments at the same time but in a sequence that complies with the skeleton hierarchy of the body model. First the deformation of the torso is computed, thereafter the scaling of the upper arms and upper legs, and so on.

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

Figure 3: (a) Adaptable generic human body model; (b) initial model after skeleton rescaling; (c) model after one and (d) several iteration of the spatio-temporal free-form deformation scheme.

Optionally, the two-step optimization procedure can be repeated several times. At the end, the model possesses a spatio-temporally silhouette-consistent shape (Fig. 3d).

7 Dynamic Shape Refinement

The STC scene representation that we have now at our disposition is globally silhouette-consistent with a number of time steps of the input video sequence. However, although the match is globally optimal, it may not exactly match the actor's silhouettes at each individual time step. In particular, subtle changes in body shape that are due to muscle bulging or deformation of the apparel are not modeled in the geometry. Furthermore, certain types of geometry features, such as concavities on the body surface, can not be captured from silhouette images alone.

In order to capture these dynamic details in the surface geometry, we compute per-vertex displacements for each time step of video individually. To this end, we also exploit the color information in the input video frames.

Assuming a purely Lambertian surface reflectance, we estimate appropriate per-vertex displacements by jointly optimizing a multi-view color-consistency and a multi-view silhouette-consistency measure. Regularization terms that assess mesh distortions and visibility changes are also employed.

The following subsequent steps are performed for each body segment and each time step of video:

- Regions on the body model are identified in which the STC representation does not match the actor's shape according to a color-consistency measure (Sect. 7.1).
- A number of random seed vertices in each region are displaced along their local normal directions (Sect. 7.2).
- Using the seeds' displacements, all vertices belonging to the color-inconsistent region are displaced by means of a Laplace interpolation (Sect. 7.3).

7.1 Identification of Color-inconsistent Regions

We use color information to identify, for each time step of video individually, those regions of the body geometry which do not

fully comply with the appearance of the actor in the input video images. To numerically assess the geometry misalignment, we compute for each vertex a color-consistency measure similar to the one described in [5]. To this end, for vertex j with 3D coordinate v_j we first average the colors of those image pixels from all camera views to which vertex j projects:

$$\overline{I}(v_j) = \frac{\sum_{i=1}^{K} \gamma_i(v_j) I_i(v_j)}{\sum_{i=1}^{N} \gamma_i(v_j)}. \quad (1)$$

In (1), K is the number of input images, $\gamma_i(v_j)$ is the visibility function for image i. It is 1 if the vertex j is visible in image i, and 0 otherwise. $I_i(v_j)$ denotes the color of j in image i. The color-consistency value for vertex j is then computed as:

$$E_I(v_j) = \frac{\sum_{i=1}^{N} \gamma_i(v_j)[I_i(v_j) - \overline{I}(v_j)]^2}{\sum_{i=1}^{N} \gamma_i(v_j)}. \quad (2)$$

By applying a threshold T_{LOD} to the error measure $E_I(v_j)$ we can decide if the vertex j is in a photo-consistent or photo-inconsistent position. If $E_I(v_j) > T_{LOD}$, j is classified as inconsistent, otherwise as consistent. The value T_{LOD} controls how much of the shape inaccuracy shall be corrected via vertex displacements. If T_{LOD} is very small, even slightly shape misalignments are identified. If T_{LOD} is large, only the most significant geometry discrepancies are recovered. Fig. 5a illustrates the influence of varying T_{LOD}.

All photo-inconsistent vertices are clustered into contingent photo-inconsistent surface patches by means of a region growing method.

7.2 Computing Vertex Displacements

We randomly select M vertices out of each color-inconsistent region that we have identified in the previous step. For each vertex $j \in M$ with position v_j we compute a displacement $\vec{r_j}$ in the direction of the local surface normal that minimizes the following energy functional:

$$E(v_j, \vec{r_j}) = \begin{array}{l} w_I E_I(v_j + \vec{r_j}) + w_S E_S(v_j + \vec{r_j}) + \\ w_D E_D(v_j + \vec{r_j}) + w_P E_P(v_j, v_j + \vec{r_j}) \end{array} \quad (3)$$

$E_I(v_j + \vec{r_j})$ is the color-consistency measure from Sect. 7.1.

The term $E_S(v_j + \vec{r_j})$ penalizes vertex positions that project into image plane locations that are very distant from the boundary of the person's silhouette. It evaluates to:

$$E_S(v_j + \vec{r_j}) = \begin{array}{l} w_{in} \sum_{i=1}^{K} E_{Sin,i}(v_j + \vec{r_j}) + \\ w_{out} \sum_{i=1}^{K} E_{Sout,i}(v_j + \vec{r_j}) \end{array} \quad (4)$$

$E_{Sin,i}(v)$ is the value of the inner distance field to the silhouette boundary in the image plane of camera i, evaluated at the projected position of the vertex with 3D position v. $E_{Sout,i}(v)$ is the respective value of the outer distance field. The inner and outer distance fields for each silhouette image (Fig. 4) can be pre-computed by means of the method described in [9].

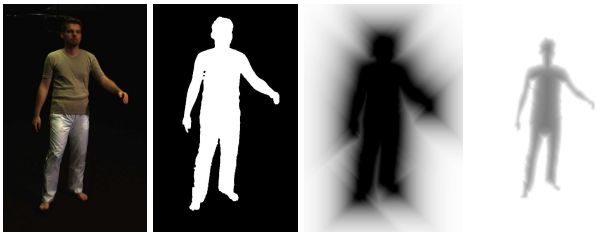


Figure 4: From left to right: original input image, silhouette image, outer distance field and inner distance field.

$E_D(v)$ regularizes the segment's mesh by measuring the distortion of triangles. We employ a distortion measure which is based on the Frobenius norm [14]:

$$\kappa = \frac{a^2 + b^2 + c^2}{4\sqrt{3}A} - 1, \quad (5)$$

where a, b and c are the lengths of a triangle's edges and A is the area of the triangle. For an equilateral triangle the value is 0. For degenerate triangles it approaches infinity. To compute $E_D(v_j + \vec{r_j})$ for a displaced vertex j at position $v_j + \vec{r_j}$, we average the κ values for the triangles adjacent to j.

The term $E_P(v_j, v_j + \vec{r_j})$ penalizes visibility changes that are due to moving a vertex j from position v_j to position $v_j + \vec{r_j}$. It has a large value if in position $v_j + \vec{r_j}$ the number of cameras that sees that vertex is significantly different from the number of cameras that sees it at v_j. If the number of cameras that sees it does not change, $E_P(v_j, v_j + \vec{r_j}) = 0$.

The weights w_I, w_S, w_D, w_P are straightforwardly found through experiments, and are chosen in a way that $E_I(v)$ and $E_S(v)$ dominate. We use the LBFGS-B method [2], a quasi-Newton algorithm, to minimize the energy function $E(v_j, \vec{r_j})$. After calculating the optimal displacement for all M random vertices individually, these displacements are used to smoothly deform the whole region by means of a Laplace interpolation method.

7.3 Deformation of Color-inconsistent Regions

Using a Laplace interpolation (see e.g. [4, 10]), each color-inconsistent region is deformed such that it globally complies with the per-vertex displacements. The new positions of the vertices in a region form an approximation to the displacement constraints. Formally, the deformed vertex positions are found via a solution to the Laplace equation

$$\mathbf{Lv} = 0, \quad (6)$$

where **v** is the vector of vertex positions and the matrix **L** is the discrete Laplace operator with

$$\mathbf{L}_{ij} = \begin{cases} valence(i) & \text{if } i \text{ inner vertex and } i = j, \\ -1 & \text{if } i \text{ inner vertex and } j \text{ in 1-neighborhood,} \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

The matrix **L** is singular, and we hence need to add suitable boundary conditions to Eq. 6 in order to solve it. We reformulate the problem as

$$min \left(\begin{pmatrix} \mathbf{L} \\ \mathbf{K} \end{pmatrix} \mathbf{v} - \begin{pmatrix} 0 \\ \mathbf{d} \end{pmatrix} \right)^2 \tag{8}$$

This equation is solved in each of the three Cartesian coordinate directions (x, y and z) separately. The matrix **K** and the vector **d** impose the individual per-vertex constraints (Sect. 7.2) which will be satisfied in least-squares sense:

$$\mathbf{K}_{ij} = \begin{cases} w_i & \text{if a displacement is specified for } i, \\ w_i & \text{if } i \text{ is a boundary vertex,} \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

The elements of **d** are:

$$\mathbf{d}_i = \begin{cases} w_i \cdot (v_i + \vec{r_i}) & \text{if a displacement is specified for } i, \\ w_i \cdot v_i & \text{if } i \text{ is a boundary vertex,} \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

The values w_i are constraint weights, v_i is the position coordinate of vertex i before deformation, and $\vec{r_i}$ is the displacement for i. The least-squares solution to (8) is found by solving the linear system:

$$\begin{pmatrix} \mathbf{L} \\ \mathbf{K} \end{pmatrix}^T \begin{pmatrix} \mathbf{L} \\ \mathbf{K} \end{pmatrix} \mathbf{x} = (\mathbf{L}^2 + \mathbf{K}^2)\mathbf{x} = \begin{pmatrix} \mathbf{L} \\ \mathbf{K} \end{pmatrix}^T \mathbf{d}. \tag{11}$$

Appropriate weights for the displacement constraints are easily found through experiments. After solving the 3 linear systems individually (for x, y and z-coordinate directions), the new deformed body shape that is both color- and silhouette-consistent with all input views is obtained.

To remove temporal noise from the dynamically refined body geometry the vertex displacements are gauss-filtered in a post-processing step.

8 Results

We have tested our method by reconstructing dynamic scene representations from two multi-view video sequences of a male actor. In one sequence (160 time steps) the person walks around in the scene, in the second sequence (210 time steps) the person performs a Tai Chi move. In Fig. 5b the STC representation that has only been reconstructed from silhouette-images, and our color-consistency-refined scene representation are shown in direct comparison. One can clearly see that in the refined model the shape of the shoulder region has been reconstructed with much higher precision. Only by means of our novel shape refinement procedure we could capture these pose-dependent geometry variations even with a segmented body model. The improvements in body geometry also lead to a better visual quality if the model is used for rendering a free-viewpoint video, where the body model is projectively textured with the input video frames. Although clever texture blending can cloak geometry inaccuracies in a purely silhouette-fitted body model [3], the dynamically refined model leads to an even further augmented rendering quality

Fig. 5d illustrates that our method is capable of reconstructing dynamic shape variations in the torso while our test subject performs a Tai Chi move. These shape variations could not have been modeled with a static set of anthropomorphic shape parameters. Fig. 5e shows the time-varying geometry of the leg in a sequence where the test subject is walking around. Fig. 5c visually illustrates the principle of our dynamic shape refinement method. First, only a few color-inconsistent vertices on the body are identified. However, if only these vertices were displaced, little bumps would appear on the body since the rest of the geometry remained unaffected. By means of our Laplace deformation the shape is smoothly deformed into a multi-view color-consistent configuration.

Fig. 5a visually illustrates that the color-consistency threshold T_{LOD} (Sect. 7.1) allows the user to control how much dynamic geometry detail is reconstructed in the shape refinement process. The images show a color-coded visualization of photo-inconsistent parts of the body's geometry for one particular time instant of video. Color-consistent parts are shown in red, color-inconsistent parts are shown in black. To generate the images (from left to right), consistency thresholds of 100, 150, 175, 200 and 225 have been used.

Applying our method to the test data, we measured the following runtimes for the individual processing steps: Through experimental evaluation, we have found out that it is sufficient to employ only 5 frames out of the input footage for spatio-temporally consistent shape reconstruction. On a PC featuring a Pentium™ 4 CPU and Nvidia GeForce 6800 GPU one iteration of the skeleton adaptation method takes on average around 1 minute. The runtime of the spatio-temporal free-form deformation strongly depends on how many time steps of video are considered. If 5 time steps are employed it takes around 15 minutes to find optimal scaling parameters. The dynamic shape refinement procedure takes around 4 minutes per time step when $M = 0.1N$ seed vertices are employed, N being the number of vertices belonging to a color-inconsistent region (Sect. 7.2).

Despite the high quality of the reconstructed scene representations, our approach is subject to a few limitations. If vertices belonging to a color-inconsistent region are never seen by any of the cameras at a particular time instant, surface details can not be recovered for this region. Furthermore, the very coarse parts of the triangle meshes do not enable us to capture small surface details, such as small folds in the apparel, which are below the mesh's resolution. In consequence, even on the highest detail level the refined geometry remains a smooth approximation to the true shape. Finally, the color-based shape refinement may fail if the person wears clothes that exhibit a strong view-dependent reflectance variation.

Despite these limitations, we have demonstrated that we can robustly reconstruct a realistic dynamic scene representation of a moving human actor, even though we only employ a rather generic segmented body model.

9 Conclusions

In this paper, we presented an automatic approach to reconstruct a dynamic scene representation of a moving human subject from multi-view video data. Starting from a generic body model, our method first creates a spatio-temporally consistent shape representation that matches the actor's silhouettes at multiple time instants. Subtle surface details at each particular time step, which could not be captured from silhouette images alone, are identified by employing color-consistency and qualitative criteria. The final dynamic scene representation realistically recovers even time-varying geometry details of a moving person by means of a generic segmented body model. In the future, we plan to reconstruct detailed surface reflectance models from multi-view video, and to use them for shape refinement of body surfaces with arbitrary reflectance properties.

References

[1] B. Allen, B. Curless, and Z. Popović. Articulated body deformation from range scan data. 21(3):612–619, 2002.

[2] R. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comp.*, 16(5):1190–1208, 1995.

[3] J. Carranza, C. Theobalt, M. Magnor, and H.-P. Seidel. Free-viewpoint video of human actors. *ACM Trans. Graph.*, 22(3):569–577, 2003.

[4] G. Farin. *Curves and Surfaces for CAGD: A Practical Guide*. Morgan Kaufmann, 1999.

[5] P. Fua and Y. G. Leclerc. Object-centered surface reconstruction: combining multi-image stereo and shading. *Int. J. Comput. Vision*, 16(1):35–55, 1995.

[6] M. H. Gross, S. Würmlin, M. Näf, E. Lamboray, C. P. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. J. Van Gool, S. Lang, K. Strehlke, A. V. Moere, and O. G. Staadt. Blue-c: a spatially immersive display and 3d video portal for telepresence. *ACM Trans. Graph.*, 22(3):819–827, 2003.

[7] A. Hilton and J. Starck. Multiple view reconstruction of people. In *3DPVT*, pages 357–364, 2004.

[8] T. Kanade, P. Rander, and P. J. Narayanan. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE MultiMedia*, 4(1):34–47, 1997.

[9] M. N. Kolountzakis and K. N. Kutulakos. Fast computation of the euclidian distance maps for binary images. *Inf. Process. Lett.*, 43(4):181–184, 1992.

[10] Y. Lipman, O. Sorkine, D. Cohen-Or, D. Levin, C. Rössl, and H.-P. Seidel. Differential coordinates for interactive mesh editing. In Franca Giannini and Alexander Pasko, editors, *Shape Modeling International 2004 (SMI 2004)*, pages 181–190, Genova, Italy, 2004. IEEE.

[11] T. Matsuyama and T. Takai. Generation, visualization, and editing of 3D video. In *Proc. of 1st International Symposium on 3D Data Processing Visualization and Transmission (3DPVT'02)*, page 234ff, 2002.

[12] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan. Image-based visual hulls. In *Proceedings of ACM SIGGRAPH 00*, pages 369–374, 2000.

[13] S. Moezzi, L.-C. Tai, and P. Gerard. Virtual view generation for 3D digital video. *IEEE MultiMedia*, 4(1):18–26, 1997.

[14] P. P. Pebay and T. J. Baker. A comparison of triangle quality measures. *In Proceedings to the 10th International Meshing Roundtable*, pages 327–340, 2001.

[15] R. Plänkers and P. Fua. Articulated soft objects for video-based body modeling. In *8th IEEE Conference on Computer Vision(ICCV)*, pages 394–401, 2001.

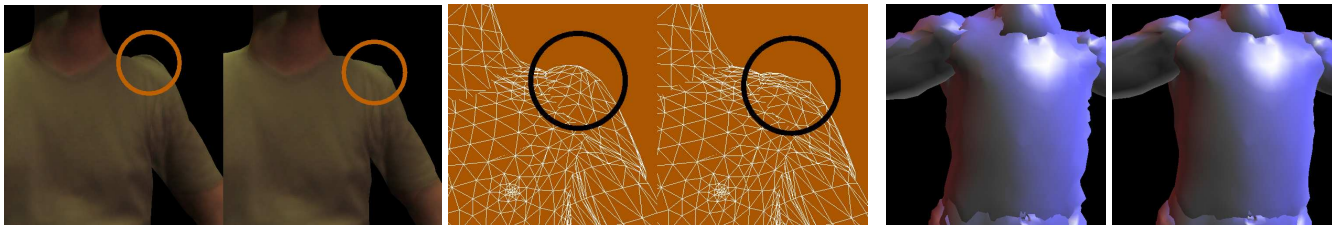
[16] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes*. Cambridge University Press, 1992.

[17] P. Sand, L. McMillan, and J. Popović. Continuous capture of skin deformation. *ACM Trans. Graph.*, 22(3):578–586, 2003.

[18] S. Wuermlin, E. Lamboray, O. G. Staadt, and M. H. Gross. 3d video recorder. In *Proc. of IEEE Pacific Graphics*, pages 325–334, 2002.

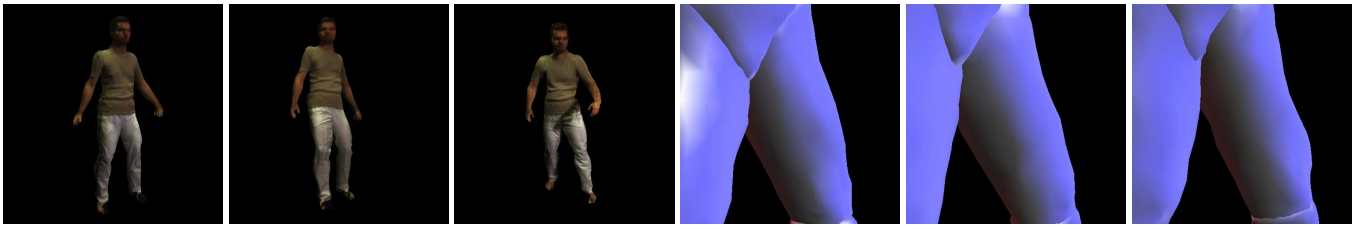
[19] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-quality video view interpolation using a layered representation. *ACM Trans. Graph.*, 23(3):600–608, 2004.

(a)



(b) (c)

(d)



(e)

Figure 5: (a) Different amounts of details can be reconstructed at each time instant by varying T_{LOD}; (b) Improvements in the shoulder geometry by applying the dynamic shape refinement method; (c) Illustration of the dynamic shape refinement process: if only random seed vertices would be displaced, all the remaining geometry would be unaffected; the Laplace interpolation smoothly deforms the whole geometry; (d) Different time instants of a Tai Chi motion; the time-varying shape of the torso has been recovered; (e) Dynamic changes in the leg's geometry while the person is walking.