12

Tracking Clothed People

B. Rosenhahn[1], U. G. Kersting[2], K. Powell[2], T. Brox[3] and H.-P. Seidel[1]

[1] Max Planck Institute for Informatics, Stuhlsatzhausenweg 85,
   D-66123 Saarbrücken, Germany
   rosenhahn@mpi-inf.mpg.de

[2] Department of Sport and Exercise Science
   The University of Auckland, New Zealand

[3] CVPR Group, University of Bonn, Germany
   brox@cs.uni-bonn.de

Summary. This chapter presents an approach for motion capturing (MoCap) of dressed people. A cloth draping method is embedded in a silhouette based MoCap system and an error functional is formalized to minimize image errors with respect to silhouettes, pose and kinematic chain parameters, the cloth draping components and external forces. Furthermore, Parzen-Rosenblatt densities on static pose configurations are used to stabilize tracking in highly noisy image sequences. We report on various experiments with two types of clothes, namely a skirt and a pair of shorts. Finally we compare the angles of the MoCap system with results from a commercially available marker based tracking system. The experiments show, that we are less than one dregree above the error range of marker based tracking systems, though body parts are occluded with cloth.

12.1 Introduction

Classical motion capture (MoCap) comprises techniques for recording the movements of real objects such as humans or animals [35]. In biomechanical settings, it is aimed at analyzing captured data to quantify the movement of body segments, e.g. for clinical studies, diagnostics of orthopaedic patients or to help athletes to understand and improve their performances. It has also grown increasingly important as a source of motion data for computer animation. Surveys on existing methods for MoCap can be found in [21, 22, 11]. Well known and commercially available marker based tracking systems exist, e.g. those provided by Motion Analysis, Vicon or Simi [20]. The use of markers comes along with intrinsic problems, e.g. incorrect identification of markers, tracking failures, the need for special laboratory environments and lighting conditions and the fact that people may not feel comfortable with markers attached to the body. This can lead to unnatural motion patterns. As well, marker based systems are designed to track the motion of the markers themselves, and thus it must be assumed that the recorded motion of the markers is identical to the motion of the underlying human segments. Since human segments are not truly

Fig. 12.1. (A) Input: A multi-view image sequence (4 cameras, one cropped image is shown). (B) The algorithm determines the cloth parameters and joint configuration of the underlined leg model. (C) Cloth and leg configuration in a virtual environment. (D) Plain leg configuration.

rigid this assumption may cause problems, especially in highly dynamic movements typically seen in sporting activities. For these reasons, marker-less tracking is an important field of research that requires knowledge in biomechanics, computer vision and computer graphics.

Typically, researchers working in the area of computer vision prefer simplified human body models for MoCap, e.g., stick, ellipsoidal, cylindric or skeleton models [3, 4, 19, 10, 13]. In computer graphics advanced object modeling and texture mapping techniques for human motions are well known [17, 6, 5, 36], but image processing or pose estimation techniques (if available) are often simplified.

In [9] a shape-from-silhouettes approach is applied to track human beings and incorporates surface point clouds with skeleton models. One of the subjects even wears a pair of shorts, but the cloth is not explicitly modeled and simply treated as rigid component. Furthermore, the authors just perform a quantitative error analysis on synthetic data, whereas in the present study a second (commercial) marker based tracking system is used for comparison.

A recent work of us [28] combines silhouette based pose estimation with more realistic human models: These are represented by free-form surface patches and local morphing along the surface patches is applied to gain a realistic human model within silhouette based MoCap. Also a comparison with a marker based system is performed indicating a stable system. In this setup, the subjects have to wear a body suit to ensure an accurate matching between the silhouettes and the surface models of the legs. Unfortunately, body suits may be uncomfortable to wear in contrast to loose clothing (shirts, shorts, skirts etc.). The subjects also move slightly different in body suits compared to being in clothes since all body parts (even unfavorable ones) are clearly visible. The incorporation of cloth models would also simplify the analysis of outdoor scenes and arbitrary sporting activities. It is for these reasons that we are interested in a MoCap system which also incorporates cloth models. A first version of our approach has been presented in [29].

Cloth draping [12, 14, 18, 34] is a well known research topic in computer graphics. Virtual clothing can be moved and rendered so that it blends seamlessly with motion and appearance in movie scenes. The motion of fabrics is determined by bending, stretching and shearing parameters, as well as external forces, aerodynamic effects and collisions. For this reason the estimation of cloth simulation parameters is es-

sential and can be done by video [2, 7, 25, 24] or range data [16] analysis. Existing approaches can be roughly divided into geometrically or physically based ones. Physical approaches model cloth behavior by using potential and kinetic energies. The cloth itself is often represented as a particle grid in a spring-mass scheme or by using finite elements [18]. Geometric approaches [34] model cloths by using other mechanics theories which are often determined empirically. These methods can be very fast computationally but are often criticized as being not very appealing visually.

The Chapter is build upon the foundations and basic tracking system described in Chapter 11. This comprises techniques for image segmentation with level sets, pose estimation of kinematic chains, shape registration based on ICP or optic flow, motion prediction by optic flow, and a prior on joint angle configurations. The focus of this work is now the embedding of a clothing model within the MoCap system. To make the Chapter self-contained, we repeat foundations in the next section. In Section three, we continue with introducing a kinematically motivated cloth draping model and a deformation model, which allow deformation of the particle mesh of a cloth with respect to oncoming external forces. The proposed draping method belongs to the class of geometric approaches [34] for cloth draping. The advantages for choosing this class are two-fold: Firstly, we need a model which supports time efficiency, since cloth draping is needed in one of the innermost loops for minimization of the used error functional. Secondly, it should be easy to implement and based on the same parametric representation as the used free-form surface patches. This allows a direct integration into the MoCap system. In section four we will explain how to minimize the cloth draping and external forces within an error functional for silhouette based MoCap. This allows us to determine joint positions of the legs even if they are partially occluded (e.g. by skirts). We present MoCap results of a subject wearing a skirt and a pair of shorts and perform a quantitative error analysis. Section five concludes with a summary.

12.1.1 Contributions

In this chapter we inform about the following main contributions:

1. A so-called kinematic cloth draping method is proposed. It belongs to the class of geometric cloth draping methods and is well suited to be embedded in a MoCap-system due to the use of a joint model.
2. The cloth draping is extended by including a deformation model which allows to adapt the cloth draping to external forces, the scene dynamics or speed of movement.
3. The main contribution is to incorporate the cloth draping algorithm in a silhouette based MoCap system. This allows for determining the joint configurations even when parts of the person are covered with fabrics (see Figure 12.1).
4. Finally we perform a quantitative error analysis. This is realized by comparing the MoCap-results with a (commercially available) marker based tracking system. The analysis shows that we get stable results and can compete with the error range of marker based tracking systems.
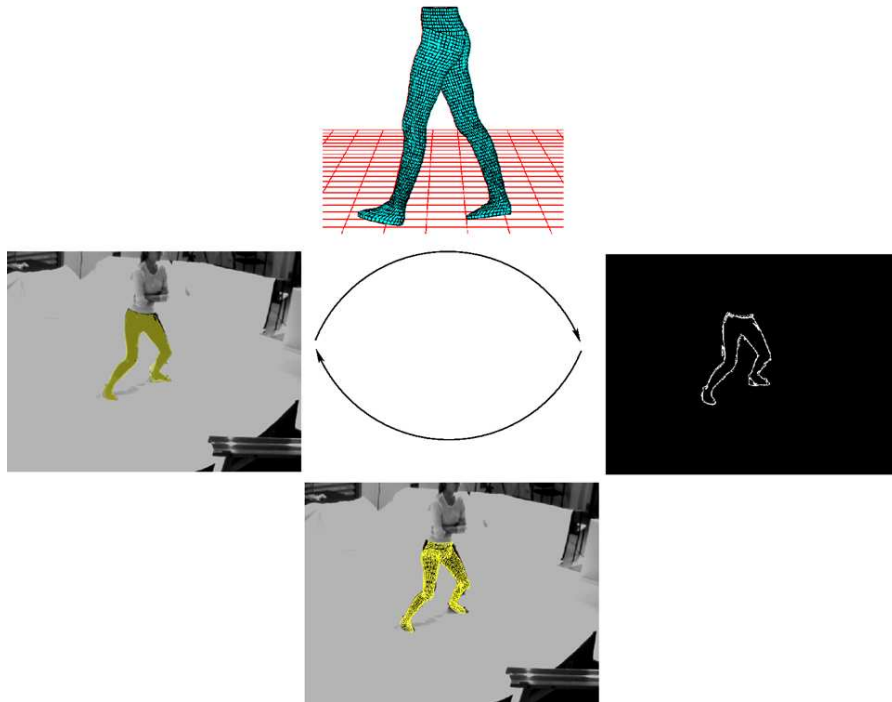
Fig. 12.2. The MoCap system in [28]: **Top:** The model of the person is assumed. **Left:** The object contours are extracted in the input images. **Right:** These are used for correspondence pose estimation. The pose result is applied as shape prior for the segmentation process and the process is iterated.

12.2 Foundations: Silhouette based MoCap

This work is based on a marker-less MoCap system [30, 28]. In this system, the human being is represented in terms of free-form surface patches, joint indices are added to each surface node and the joint positions are assumed. This allows to generate arbitrary body configurations, steered through joint angles. The corresponding counterparts in the images are 2D silhouettes: These are used to reconstruct 3D ray bundles and a spatial distance constraint is minimized to determine the position and orientation of the surface mesh and the joint angles. In this section we will give a brief summary of the MoCap system. These foundations are needed later to explain concisely, where and how the cloth draping approach is incorporated. A more detailed survey can be found in Chapter 11.

12.2.1 Silhouette extraction

In order to estimate the pose from silhouettes, these silhouettes have to be extracted first, which comes down to a classical segmentation problem. In the system described here, the segmentation is based on a level set representation of contours.

Fig. 12.3. Silhouette extraction based on level set functions. Left: Initial segmentation. Right: Segmentation result.

A level set function $\Phi \in \Omega \mapsto \mathbb{R}$ splits the image domain Ω into the foreground region Ω_1 and background region Ω_2 with $\Phi(x) > 0$ if $x \in \Omega_1$ and $\Phi(x) < 0$ if $x \in \Omega_2$. The zero-level line thus marks the boundary between both regions. In order to make the representation unique, the level set functions are supposed to be signed distance functions.

Both regions are analyzed with respect to their feature distribution. The feature space may contain, e.g., gray value, color, or texture features. The key idea is to evolve the contour such that the two regions maximize the a-posteriori probability. Usually, one assumes a priori a smooth contour, but more sophisticated shape priors can be incorporated as well, as shown in Section 12.2.4. Maximization of the posterior can be reformulated as minimization of the following energy functional:

$$E(\Phi, p_1, p_2) = -\int_{\Omega} \big(H(\Phi(x)) \log p_1 + (1 - H(\Phi(x))) \log p_2 + \nu |\nabla H(\Phi(x))| \big) dx \quad (12.1)$$

where $\nu > 0$ is a weighting parameter and $H(s)$ is a regularized version of the Heaviside function, e.g. the error function. The probability densities p_i describe the region model. We use a local Gaussian distribution. These densities are estimated according to the *expectation-maximization principle*. Having the level set function initialized with some contour, the probability densities within the two regions can be estimated. Contour and probability densities are then updated in an iterative manner. An illustration can be seen in Figure 12.3. The left picture depicts the initialization of the contour. The right one shows the estimated (stationary) contour after 50 iterations. As can be seen, the legs and the skirt are well extracted, but there are some problems in the area of the feet region caused by shadows. Incorporation of a shape prior greatly reduces such effects.

12.2.2 Pose estimation

Assuming an extracted image contour and the silhouette of the projected surface mesh, the closest point correspondences between both contours are used to define a set of corresponding 3D lines and 3D points. Then a 3D point-line based pose estimation algorithm for kinematic chains is applied to minimize the spatial distance between both contours: For point based pose estimation each line is modeled as a 3D Plücker line $L_i = (n_i, m_i)$, with a (unit) direction n_i and moment m_i [23]. There exist different ways to represent projection rays. As we have to minimize distances between correspondences, it is advantageous to use an implicit representation for a 3-D line. It allows instantaneously to determine the distance between a point and a line. A Plücker line $L = (n, m)$ is given as a unit vector n and a moment m with $m = x \times n$ for a given point x on the line. An advantage of this representation is its uniqueness (apart from possible sign changes). Moreover, the incidence of a point x on a line $L = (n, m)$ can be expressed as

$$x \in L \Leftrightarrow x \times n - m = 0. \tag{12.2}$$

This equation provides us with an error vector. Let $L = (n, m)$, with $m = v \times n$ as shown in Figure 12.4, and $x = x_1 + x_2$, with $x \notin L$ and $x_2 \perp n$.
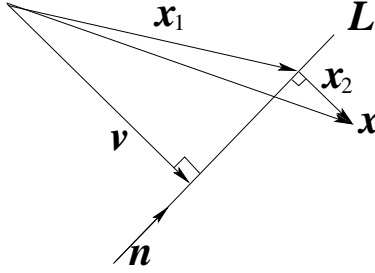

Fig. 12.4. Comparison of a 3-D point x with a 3-D line L.

Since $x_1 \times n = m$, $x_2 \perp n$, and $\|n\| = 1$, we have

$$\|x \times n - m\| = \|x_1 \times n + x_2 \times n - m\| = \|x_2 \times n\| = \|x_2\| \tag{12.3}$$

where $\| \cdot \|$ denotes the Euclidean norm. This means that $x \times n - m$ in (12.2) results in the (rotated) perpendicular error vector to line L.
The 3D rigid motion is expressed as exponential form

$$\boldsymbol{M} = \exp(\theta \hat{\xi}) = \exp \begin{pmatrix} \theta \hat{\omega} & \theta \boldsymbol{v} \\ 0_{3\times 1} & 0 \end{pmatrix} \tag{12.4}$$

where $\theta \hat{\xi}$ is the matrix representation of a twist $\xi \in se(3) = \{(\boldsymbol{v}, \hat{\omega}) | \boldsymbol{v} \in \mathbb{R}^3, \hat{\omega} \in so(3)\}$, with $so(3) = \{\boldsymbol{A} \in \mathbb{R}^{3\times 3} | \boldsymbol{A} = -\boldsymbol{A}^T\}$. The Lie algebra $so(3)$ is the tangential space of the 3D rotations. Its elements are (scaled) rotation axes, which can either be represented as a 3D vector or skew symmetric matrix,

$$\theta\omega = \theta\begin{pmatrix}\omega_1 \\ \omega_2 \\ \omega_3\end{pmatrix}, \text{ with } \|\omega\|_2 = 1 \quad \text{or} \quad \theta\hat{\omega} = \theta\begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix}. \quad (12.5)$$

A twist ξ contains six parameters and can be scaled to $\theta\xi$ for a unit vector ω. The parameter $\theta \in \mathbb{R}$ corresponds to the motion velocity (i.e., the rotation velocity and pitch). For varying θ, the motion can be identified as screw motion around an axis in space. The six twist components can either be represented as a 6D vector or as a 4×4 matrix,

$$\theta\xi = \theta(\omega_1, \omega_2, \omega_3, v_1, v_2, v_3)^T, \|\omega\|_2 = 1, \quad \theta\hat{\xi} = \theta\begin{pmatrix} 0 & -\omega_3 & \omega_2 & v_1 \\ \omega_3 & 0 & -\omega_1 & v_2 \\ -\omega_2 & \omega_1 & 0 & v_3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.(12.6)$$

To reconstruct a group action $\boldsymbol{M} \in SE(3)$ from a given twist, the exponential function $\exp(\theta\hat{\xi}) = \sum_{k=0}^{\infty} \frac{(\theta\hat{\xi})^k}{k!} = \boldsymbol{M} \in SE(3)$ must be computed. This can be done efficiently by using the Rodriguez formula [23].

For pose estimation the reconstructed Plücker lines are combined with the screw representation for rigid motions:

Incidence of the transformed 3D point X_i with the 3D ray $L_i = (n_i, m_i)$ can be expressed as

$$(\exp(\theta\hat{\xi})X_i)_{3\times 1} \times n_i - m_i = 0. \quad (12.7)$$

Since $\exp(\theta\hat{\xi})X_i$ is a 4D vector, the homogeneous component (which is 1) is neglected to evaluate the cross product with n_i. Then the equation is linearized and iterated, see [28].

Joints are expressed as special screws with no pitch of the form $\theta_j\hat{\xi}_j$ with known $\hat{\xi}_j$ (the location of the rotation axes is part of the model) and unknown joint angle θ_j. The constraint equation of an ith point on a jth joint has the form

$$(\exp(\theta\hat{\xi})\exp(\theta_1\hat{\xi}_1)\ldots\exp(\theta_j\hat{\xi}_j)X_i)_{3\times 1} \times n_i - m_i = 0 \quad (12.8)$$

which is linearized in the same way as the rigid body motion itself. It leads to three linear equations with the six unknown pose parameters and j unknown joint angles.

12.2.3 Shape Registration

The goal of shape registration can be formulated as follows: Given a certain distance measure; the task is to determine one transformations that leads to the minimum distance between shapes. A very popular shape matching method working on such representations is the iterated closest point (ICP) algorithm [1]. Given two finite sets P and Q of points. The (original) ICP algorithm calculates a rigid transformation T and attempts to ensure $TP \subseteq Q$.

1. **Nearest point search**: for each point $p \in P$ find the closest point $q \in Q$.
2. **Compute registration**: determine the transformation T that minimizes the sum of squared distances between pairs of closest points (p, q).
3. **Transform**: apply the transformation T to all points in set P.
4. **Iterate**: repeat step 1 to 3 until the algorithm converges.

This algorithm converges to the next local minimum of the sum of squared distances between closest points. A good initial estimate is required to ensure convergence to the sought solution. Unwanted solutions may be found if the sought transformation is too large, e.g. many shapes have a convergence radius in the area of $20°$ [8], or if the point sets do not provide sufficient information for a unique solution.

The original ICP algorithm has been modified in order to improve the rate of convergence and to register partially overlapping sets of points. Zhang [37] uses a modified cost function based on robust statistics to limit the influence of outliers. Other approaches aim at the avoidance of local minima during registration subsuming the use of Fourier descriptors [31], color information [15], or curvature features [33].

The advantages of ICP algorithms are obvious: they are easy to implement and will provide good results, if the sought transformation is not too large [8]. For our tracking system we compute correspondences between points on image silhouettes to the surface mesh with the ICP algorithm presented in [31].

In Chapter 11 and [27] it is further explained how an alternative matching procedure, by using the optic flow can be used to improve the convergence rate and convergence radius. In this work we make use of both matchers to register a surface model to an image silhouette.

12.2.4 Combined pose estimation and segmentation

Since segmentation and pose estimation can both benefit from each other, it is sensible to couple both problems in a joint optimization problem. To this end, the energy functional for image segmentation in (12.1) is extended by an additional term that integrates the surface model:

$$E(\Phi, \theta\xi) = -\int_{\Omega} \big(H(\Phi)\log p_1 + (1 - H(\Phi))\log p_2\big)\, dx + \nu \int_{\Omega} |\nabla H(\Phi)|\, dx$$
$$+ \lambda \underbrace{\int_{\Omega} (\Phi - \Phi_0(\theta\xi))^2\, dx}_{\text{Shape}}. \tag{12.9}$$

The quadratic error measure in the shape term has been proposed in the context of 2D shape priors, e.g. in [32]. The prior $\Phi_0 \in \Omega \to \mathbb{R}$ is assumed to be represented by the signed distance function. This means in our case, $\Phi_0(x)$ yields the distance of x to the silhouette of the projected object surface.

Given the contour Φ, the pose estimation method from Section 12.2.2 minimizes the shape term in (12.9). Minimizing (12.9) with respect to the contour Φ, on the other hand, leads to the gradient descent equation

$$\partial_t \Phi = H'(\Phi)\left(\log\frac{p_1}{p_2} + \nu\,\nabla^\top\left(\frac{\nabla\Phi}{|\nabla\Phi|}\right)\right) + 2\lambda\,(\Phi_0(\theta\xi) - \Phi). \tag{12.10}$$

The total energy is minimized by iterating both minimization procedures. Both iteration steps minimize the distance between Φ and Φ_0. While the pose estimation method draws Φ_0 towards Φ, thereby respecting the constraint of a rigid motion, (12.10) in return draws the curve Φ towards Φ_0, thereby respecting the data in the image.

12.2.5 Quantitative error analysis

A lack of many studies (e.g. [19]) is that only a visual feedback about the pose result is given, by overlaying the pose result with the image data.
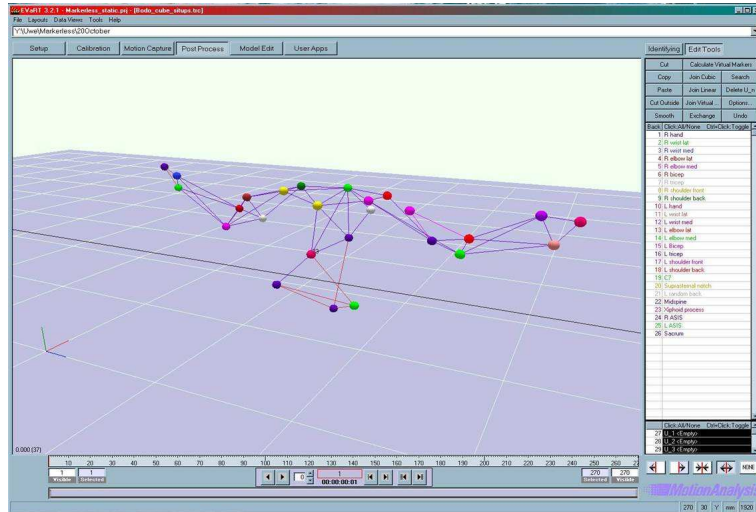

Fig. 12.5. Screen shot of the used EVA solver from Motion Analysis
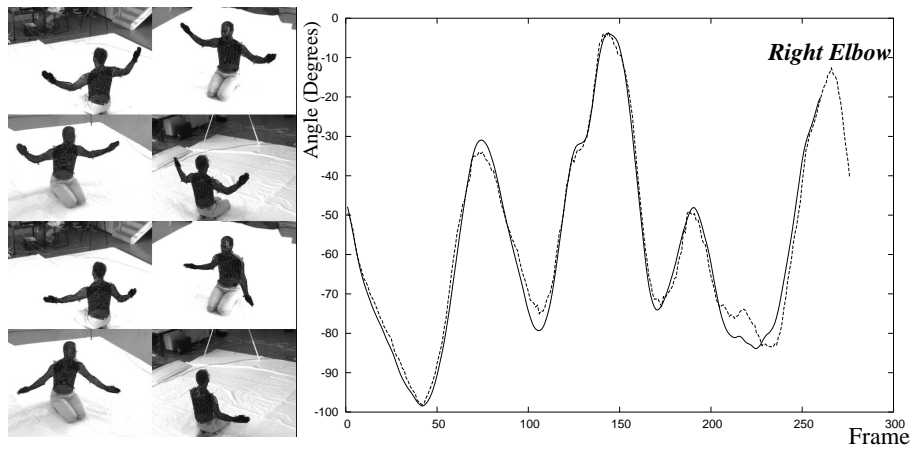

Fig. 12.6. Tracked arms: The angle diagrams show the elbow values of the Motion analysis system (dotted) and the silhouette system (solid).
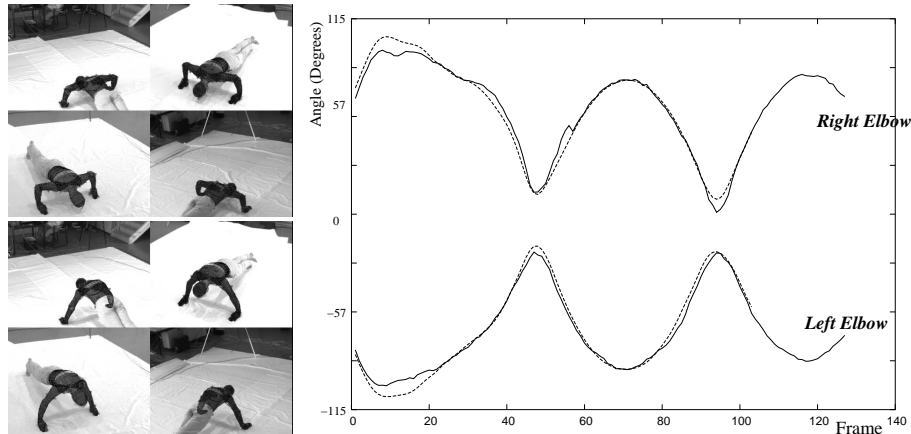
Fig. 12.7. Tracked Push-ups: The angle diagrams show the elbow values of the Motion analysis system (dotted) and the silhouette system (solid).

To enable a quantitative error analysis, we use a commercial marker based tracking system for a comparison. Here, we use the Motion Analysis software [20], with an 8-Falcon-camera system. For data capture we use the Eva 3.2.1 software and the Motion Analysis Solver Interface 2.0 for inverse kinematics computing. In this system a human has to wear a body suit and retroflective markers are attached to it. Around each camera is a strobe light led ring and a red-filter is in front of each lens. This gives very strong image signals of the markers in each camera. These are treated as point markers which are reconstructed in the eight-camera system. Figure 12.5 shows a screen shot of the Motion Analysis system. The system is calibrated by using a wand-calibration method. Due to the filter in front of the images we had to use a second camera set-up which provides *real* image data. This camera system is calibrated by using a calibration cube. After calibration, both camera systems are calibrated with respect to each other. Then we generate a stick-model from the point markers including joint centers and orientations. This results in a complete calibrated set-up we use for a system comparison.

Figure 12.6 shows the first test sequence, where the subject is just moving the arms forwards and backwards. The diagram on the right side shows the estimated angles of the right elbow. The marker results are given as dotted lines and the silhouette results in solid lines. The overall error between both angles diagrams is 2.3 degrees, including the tracking failure between frames 200 till 250.

Figure 12.7 shows the second test sequence, where the subject is performing a series of push-ups. Here the elbow angles are much more characteristic and also well comparable. The overall error is 1.7 degrees. Both sequences contain partial occlusions in certain frames. But this can be handled from the algorithm.

In [26] eight biomechanical measurement systems are compared (including the Motion Analysis system). A rotation experiment is performed, which shows that the RMS[4] errors are typically within three degrees. Our error measures fit in this range quite well.

---

[4] root mean square

12.3 Kinematic cloth draping

To integrate a clothing model in the MoCap system, we decided to use a geometric approach. The main reason is that cloth draping is needed in one of the innermost loops for pose estimation and segmentation. Therefore it must be very fast. In our case we need around 400 iterations for each frame to converge to a solution. A cloth draping algorithm in the area of seconds would require hours to calculate the pose of one frame and weeks for a whole sequence.
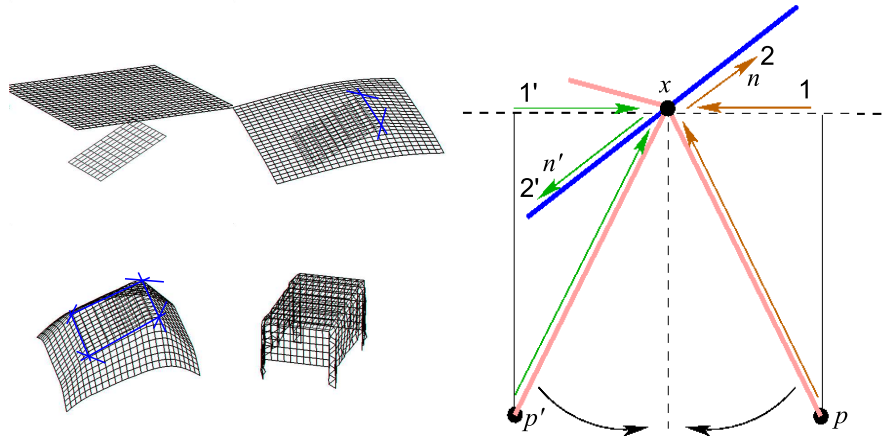

Fig. 12.8. The cloth draping principle. Joints are used to deform the cloth while draping on the surface mesh.

Fig. 12.9. Draping of the skirt model

We decided to model the skirt as a string-system with underlined kinematic chains: The main principle is visualized on the left in Figure 12.8 for a piece of cloth falling on a plane. The piece of cloth is represented as a particle grid, a set of points with known topology. While lowering the cloth, the distance of each cloth point to the ground plane is determined. If the distance between one point on the cloth to the surface is below a threshold, the point is set as a fixed-point, see the top right image on the left of Figure 12.8. Now the remaining points are not allowed to *fall*
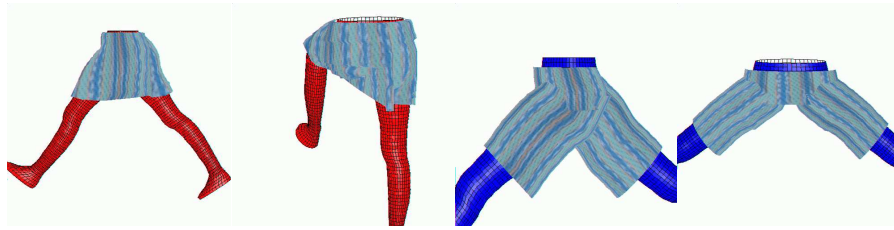
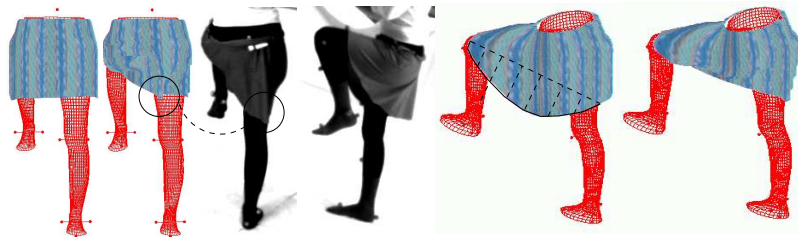**Fig. 12.10.** Cloth draping of a skirt and shorts in a simulation environment.

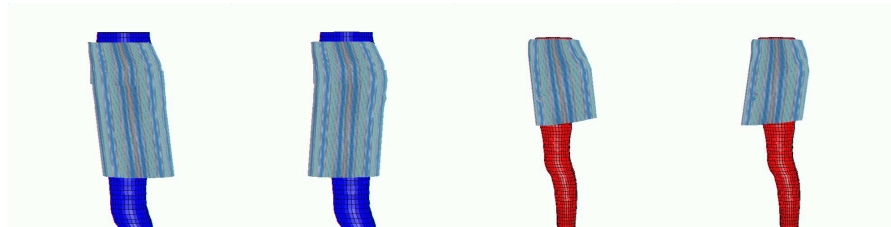**Fig. 12.11.** Reconstraining the skirts' length.

Fig. 12.12. External forces, e.g. virtual wind on the shorts (left) and the skirt (right). Visualized is frontal and backward wind.

downwards anymore. Instead, for each point, the nearest fixed-point is determined and a joint (perpendicular to the particle point) is used to rotate the free point along the joint axis through the fixed point. The used joint axes are marked as blue lines in Figure 12.8. The image on the right in Figure 12.8 shows the geometric principle to determine the twist for rotation around a fixed point: The blue line represents a mesh of the rigid body, x is the fixed point and the (right) pink line segment connects x to a particle p of the cloth. The direction between both points is projected onto the y-plane of the fixed point (1). The direction is then rotated around 90 degrees (2), leading to the rotation axis n. The point pairs $(n, x \times n)$ are the components of the twist, see equation (12.6). While lowering the cloth, free particles not touching a second rigid point, will swing below the fixed point (e.g. p'). This leads to an opposite rotation (indicated with (1'), (2') and n') and the particle swings back again, resulting in a natural swinging draping pattern. The draping velocity is steered through a rotation velocity θ, which is set to 2 degrees during iteration. Since all points either become fixed points, or result in a stationary configuration
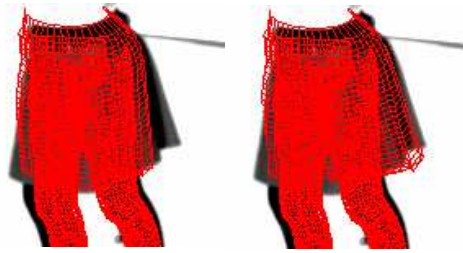
Fig. 12.13. Left: Overlaid pose result without external forces. Right: Overlaid pose result including external forces.

while swinging backwards and forwards, we constantly use 50 iterations to drape the cloth. The remaining images on the left in Figure 12.8 show the ongoing draping and the final result. Figure 12.9 shows the cloth draping steps of a skirt model.

Figure 12.10 shows example images of a skirt and a pair of shorts falling on the leg model. The skirt is modeled as a 2-parametric mesh model. Due to the use of general rotations, the internal distances in the particle mesh cannot change with respect to one of these dimensions, since a rotation maintains the distance between the involved points. However, this is not the case for the second sampling dimension. For this reason, the skirt needs to be re-constrained after draping. This is visualized in Figure 12.11: If a stretching parameter is exceeded, the particles are reconstrained to minimal distance to each other. This is only done for the non-fixed points (i.e. for those which are not touching the skin). It results in a better appearance especially for certain leg configurations.

Figure 12.11 shows that even the creases are maintained. In this case, shorts are simpler since they are modeled as cylinders, transformed together with the legs and then draped.

To improve the dynamic behavior of clothing during movements, we also add external forces to the cloth draping. We continue with the cloth-draping in the following way: dependent on the direction of a force we determine a joint on the nearest fixed point for each free point on the surface mesh with the joint direction being perpendicular to the force direction. Now we rotate the free point around this axis dependent on the force amount (expressed as an angle) or until the cloth is touching the underlying surface. Figure 12.12 shows examples of the shorts and skirt with frontal or backward virtual wind acting as external force. The external forces and their directions are later part of the minimization function during pose tracking. Figure 12.13 visualizes the effect of the used deformation model. Since the motion dynamics of the cloth are determined dynamically, we need no information about the cloth type or weight since they are implicitly determined from the minimized cloth dynamics in the image data; we only need the measurements of the cloth.

12.4 Combined cloth draping and MoCap

The assumptions are as follows: We assume the representation of a subject's lower torso (i.e. for the hip and legs) in terms of free-form surface patches. We also assume known joint positions along the legs. Furthermore we assume the wearing of
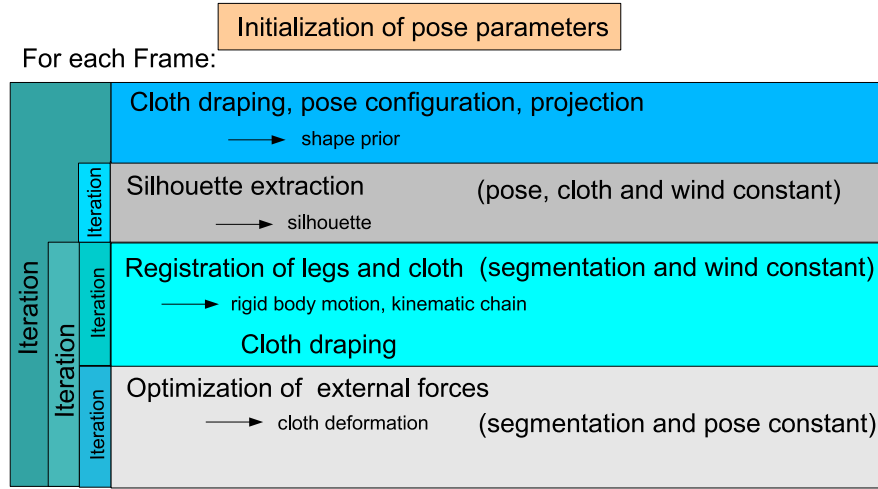
Fig. 12.14. The basic algorithm for combined cloth draping and motion capturing

a skirt or shorts with known measures. The person is walking or stepping in a four-camera setup. These cameras are triggered and calibrated with respect to one world coordinate system. The task is to determine the pose of the model and the joint configuration. For this we minimize the image error between the projected surface meshes to the extracted image silhouettes. The unknowns are the pose, kinematic chain and the cloth parameters (external forces, cloth thickness, etc.). The task can be represented as an error functional as follows:

$$E(\Phi, p_1, p_2, \theta\xi, \theta_1, \ldots, \theta_n, c, w) =$$

$$\underbrace{-\int_\Omega \big(H(\Phi)\log p_1 + (1 - H(\Phi))\log p_2 + \nu|\nabla H(\Phi)|\big)\,dx}_{\text{segmentation}}$$

$$\underbrace{+\lambda\int_\Omega(\Phi - \Phi_0(\underbrace{\theta\xi, \theta_1, \ldots, \theta_n}_{\text{pose and kinematic chain,}}, \underbrace{c, w}_{\text{external forces}}))dx}_{\text{shape error}}$$

Due to the large number of parameters and unknowns we decided for an iterative minimization scheme, see Figure 12.14: Firstly, the pose, kinematic chain and external forces are kept constant, while the error functional for the segmentation (based on Φ, p_1, p_2) is minimized (section 12.2.1). Then the segmentation and external forces are kept constant while the pose and kinematic chain are determined to fit the surface mesh and the cloth to the silhouettes (section 12.2.2). Finally, different directions of external forces are sampled to refine the pose result (section 12.3). Since all parameters influence each other, the process is iterated until a steady state is reached. In our experiments, we always converged to a local minimum.
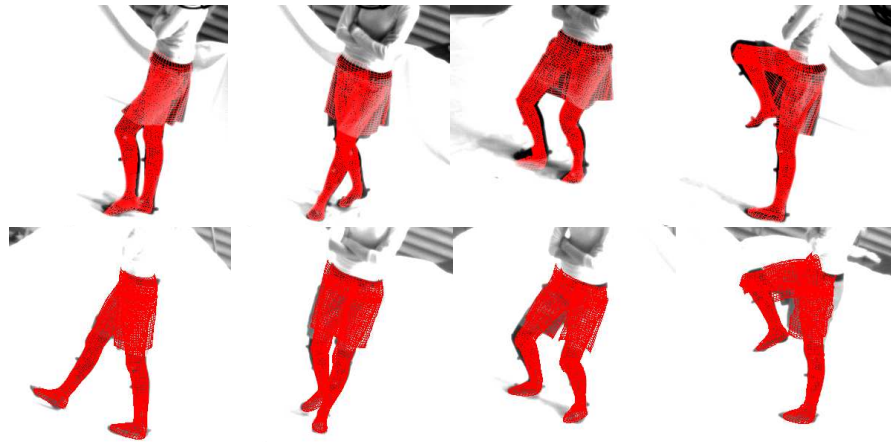
12.5 Experiments

Fig. 12.15. Example sequences for tracking clothed people. **Top row**: walking, leg crossing, knee bending and knee pulling with a skirt. **Bottom row**: walking, leg crossing, knee bending and knee pulling with shorts. The pose is determined from 4 views (just one of the views is shown, images are cropped).

Fig. 12.16. Error during grabbing the images

For the experiments we used a four-camera set up and grabbed image sequences of the lower torso with different motion patterns: The subject was asked to wear the skirt and the shorts while performing walking, leg crossing and turning, knee bending and walking with knees pulled up. We decided on these different patterns, since they are not only of importance for medical studies (e.g. walking), but they are also challenging for the cloth simulator, since the cloth is partially stretched (knee pulling sequence) or hanging down loosely (knee bending). The turning and
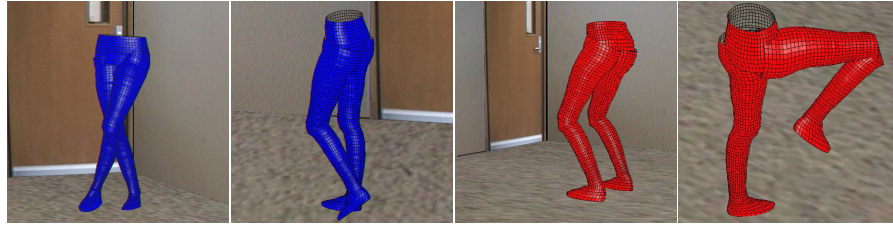
Fig. 12.17. Example leg configurations of the sequences. The examples are taken from the subject wearing the shorts (blue) and the skirt (red) (leg crossing, walking, knee bending, knee pulling).

Fig. 12.18. The set-up for quantitative error analysis: Left/middle, the subject with attached markers. The cloth does not interfere with tracking the markers. Right: A strobe light camera of the used Motion Analysis system.

leg crossing sequence is interesting due to the higher occlusions. Figure 12.15 shows some pose examples for the subject wearing the skirt (top) and shorts (bottom). The pose is visualized by overlaying the projected surface mesh onto the images. Just one of the four camera views is shown. Each sequence consists of 150-240 frames. Figure 12.16 visualizes the stability of our approach: While grabbing the images, a couple of frames were stored completely wrong. These sporadic outliers can be compensated from our algorithm, and a few frames later (see the image on the right) the pose is correct. Figure 12.17 shows leg configurations in a virtual environment. The position of the body and the joints reveal a natural configuration.

Finally, the question about the stability arises. To answer this question, we attached markers to the subject and tracked the sequences simultaneously with the commercially available Motion Analysis system [20]. The markers are attached to the visible parts of the leg and are not disturbed by the cloth, see Figure 12.18. We then compare joint angles for different sequences with the results of the marker based system, similar to Section 12.2.5. The overall errors for both types of cloth varies between 1.5 and 4.5 degrees, which indicates a stable result, see [26]. Table 12.1 summarizes the deviations
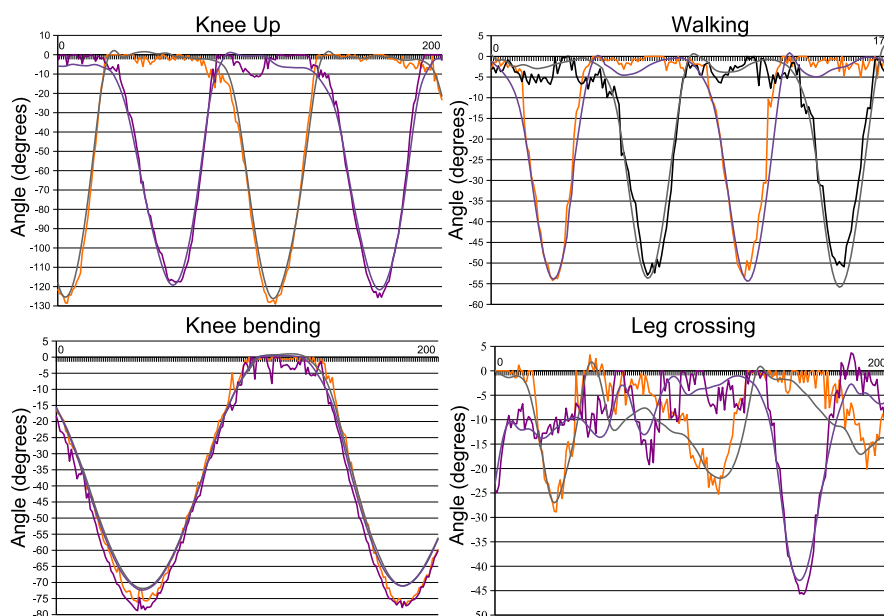
Fig. 12.19. Left: Knee angles from sequences wearing the shorts. **Right**: Knee angles from sequences wearing the skirt. **Top left**: Angles of the knee up sequence. **Bottom left**: Angles of the knee bending sequence. **Top right**: Angles of the walking sequence. **Bottom right**: Angles of the leg crossing sequence.

Table 12.1. Deviations of the left and right knee for different motion sequences.

Skirt			Shorts		
Sequence	Left knee	Right knee	Sequence	Left knee	Right knee
Dancing	3.42	2.95	Dancing	4.0	4.0
Knee-up	3.22	3.43	Knee-up	3.14	4.42
Knee bending	3.33	3.49	Knee bending	2.19	3.54
Walking	2.72	3.1	Walking	1.52	3.38

The diagrams in Figure 12.19 shows the overlay of the knee angles for two skirt and two shorts sequences. The two systems can be identified by the smooth curves from the Motion Analysis system and unsmoothed curves (our system).

12.5.1 Prior knowledge on angle configurations

Chapter 11 also introduced the use of nonparametric density estimates to build additional constraint equations to enforce the algorithm to converge to familiar configurations. It further can be seen as the embedding of soft-constraints to penalize configurations which are uncommon (e.g. to move the arm through the body) and they regularize the equations which results in guaranteed non-singular system of

equations. These advantages can also be seen from the experiments in Figure 12.20, 12.21 and 12.22:
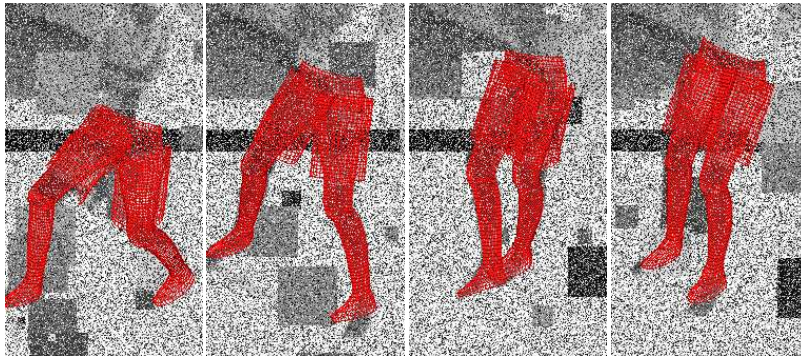

Fig. 12.20. Scissors sequence (cropped images, one view is shown): The images have been distorted with 60% uncorrelated noise, rectangles of random size and gray values and a black stripe across the images.
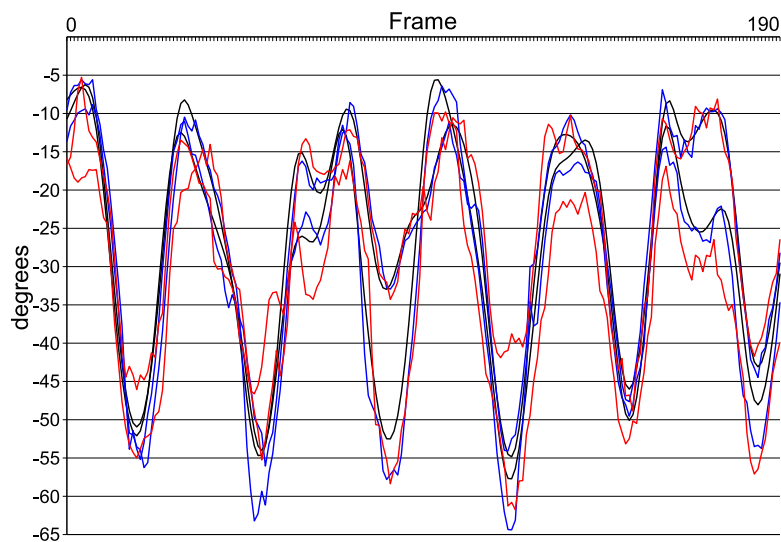

Fig. 12.21. Quantitative error analysis of the scissors sequence (the knee angles are shown). Black: the results of the Motion Analysis system. Blue: the outcome from our marker-less system (deviation: 3.45 and 2.18 degrees). Red: The results from our system for the highly disturbed image data (deviation: 6.49 and 3.0 degrees).
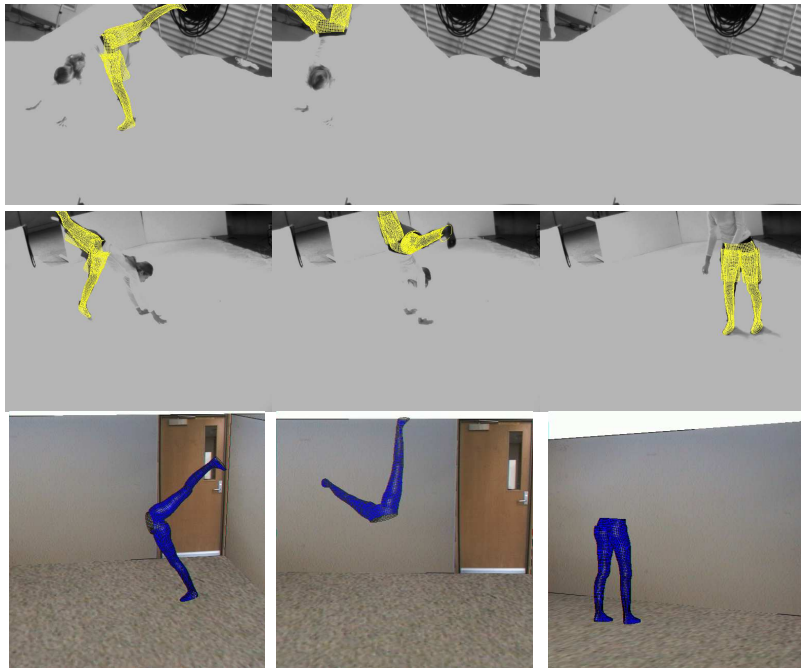
Fig. 12.22. Cartwheel sequence in a Lab environment: The legs are not visible in some key frames, but the prior knowledge allows to give the legs a natural (most likely) configuration. The top images shows two views and three example frames of the sequence. The bottom images show the leg configuration in a virtual environment.

Figure 12.20 shows results of a scissors sequence. The images have been distorted with 60% uncorrelated noise, rectangles of random size and gray values and a black stripe across the images. Due to prior knowledge of scissor jumps, the algorithm is able to track the sequence successfully. The diagram in Figure 12.21 quantifies the result by overlaying the knee angles for the marker results (black) with the undisturbed result (blue) and highly noised (red) image data. The deviation for the plain data is 3.45 and 2.18 degrees, respectively and for the highly disturbed sequence we get 6.49 and 3.0 degrees. Due to the high amount of noise added to the image data, we consider the outcome as a good result.

Figure 12.22 shows results from a cartwheel sequence. During tracking the sequence the legs are not visible in some key frames, but the prior knowledge allows to give the legs a natural (most likely) configuration. The top images shows two views and three example frames of the sequence. The bottom images show the leg configuration in a virtual environment.

12.6 Summary

The contribution presents an approach for motion capture of clothed people. To achieve this we extend a silhouette-based motion capture system, which relies on image silhouettes and free-form surface patches of the body with a cloth draping procedure. We employ a geometric approach based on kinematic chains. We call this cloth draping procedure kinematic cloth draping. This model is very well suited to be embedded in a motion capture system since it allows us to minimize the cloth draping parameters (and external forces) within the same error functional such as the segmentation and pose estimation algorithm. Due to the number of unknowns for the segmentation, pose estimation, joints and cloth parameters, we decided on an iterative solution. The experiments with a skirt and shorts show that the formulated problem can be solved. We are able to determine joint configurations and pose parameters of the kinematic chains, though they are considerably covered with clothes. Indeed, we use the cloth draping appearance in images to recover the joint configuration and simultaneously determine dynamics of the cloth. Furthermore, Parzen-Rosenblatt densities of static joint configurations are used to generate constraint equations which enables a tracking of persons in highly noisy or corrupted image sequences.

To quantify the results, we performed an error analysis by comparing our method with a commercially available marker based tracking system. The experiments show that we are close to the error range of marker based tracking systems [26].

Applications are straightforward: The motion capture results can be used to animate avatars in computer animations, and the angle diagrams can be used for the analysis of sports movements or clinical studies. The possibility of wearing loose clothes is much more comfortable for many people and enables a more natural motion behavior. The presented extension also allows us to analyze outdoor activities, e.g. soccer or other team sports.

For future works we plan to extend the cloth draping model with more advanced ones [18] and we will compare different draping approaches and parameter optimization schemes in the motion capturing setup.

Acknowledgments

References

1. Besl P. and McKay N. A method for registration of 3D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:239–256, 1992.
2. Bhat K.S., Twigg C.D., Hodgins J.K., Khosla P.K., Popovic Z. and Seitz S.M. Estimating cloth simulation parameters from video. In D.Breen and M.Lin, editors, *Proc. ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 37–51. 2003.
3. Bregler C. and Malik J. Tracking people with twists and exponential maps. In *Proc. Computer Vision and Pattern Recognition*, pages 8–15, Santa Barbara, California, 1998.

4. Bregler C., Malik J. and Pullen K. Twist based acquisition and tracking of animal and human kinetics. *International Journal of Computer Vision*, 56(3):179–194, 2004.

5. Carranza J., Theobalt C., Magnor M.A. and Seidel H.-P. Free-viewpoint video of human actors. In *Proc. SIGGRAPH 2003*, pages 569–577, 2003.

6. Chadwick J.E., Haumann D.R. and Parent R.E. Layered construction for deformable animated characters. *Computer Graphics*, 23(3):243–252, 1989.

7. Chafri H., Gagalowicz A. and Brun R. Determination of fabric viscosity parameters using iterative minimization. In A.Gagalowicz and W.Philips, editors, *Proc. Computer Analysis of Images and Patterns*, volume 3691 of *Lecture Notes in Computer Science*, pages 789–798. Springer, Berlin, 2005.

8. Chetverikov D., Stepanov D. and Krsek P. Robust Euclidean alignment of 3D point sets: The trimmed iterative closest point algorithm. *Image and Vision Computing*, 23(3):299–309, 2005.

9. Cheung K.M., Baker S. and Kanade T. Shape-from-silhouette across time: Part ii: Applications to human modeling and markerless motion tracking. *International Journal of Computer Vision*, 63(3):225–245, 2005.

10. Fua P., Plänkers R. and Thalmann D. Tracking and modeling people in video sequences. *Computer Vision and Image Understanding*, 81(3):285–302, March 2001.

11. Gavrilla D.M. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1):82–92, 1999.

12. Haddon J., Forsyth D. and Parks D. The appearance of clothing. http://http.cs.berkeley.edu/haddon/clothingshade.ps, June 2005.

13. Herda L., Urtasun R. and Fua P. Implicit surface joint limits to constrain video-based motion capture. In T.Pajdla and J.Matas, editors, *Proc. 8th European Conference on Computer Vision*, volume 3022 of *Lecture Notes in Computer Science*, pages 405–418, Prague, May 2004. Springer.

14. House D.H., DeVaul R.W. and Breen D.E. Towards simulating cloth dynamics using interacting particles. *Clothing Science and Technology*, 8(3):75–94, 1996.

15. Johnson A.E. and Kang S.B. Registration and integration of textured 3-D data. In *Proc.International Conference on Recent Advances in 3-D Digital Imaging and Modeling*, pages 234–241. IEEE Computer Society, May 1997.

16. Jojic N. and Huang T.S. Estimating cloth draping parameters from range data. In *Proc. Int. Workshop Synthetic-Natural Hybrid Coding and 3-D Imaging*, pages 73–76. Greece, 1997.

17. Magnenat-Thalmann N., Seo H. and Cordier F. Automatic modeling of virtual humans and body clothing. *Computer Science and Technology*, 19(5):575–584, 2004.

18. Magnenat-Thalmann N. and Volino P. From early draping to haute cotoure models: 20 years of research. *Visual Computing*, 21:506–519, 2005.

19. Mikic I., Trivedi M., Hunter E. and Cosman P. Human body model acquisition and tracking using voxel data. *International Journal of Computer Vision*, 53(3):199–223, 2003.

20. MoCap-System. Motion analysis: A marker based tracking system. www.motionanalysis.com, June 2005.

21. Moeslund T.B. and Granum E. A survey of computer vision based human motion capture. *Computer Vision and Image Understanding*, 81(3):231–268, 2001.

22. Moeslund T.B., Granum E. and Krüger V. A survey of of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2):90–126, 2006.

23. Murray R.M., Li Z. and Sastry S.S. *Mathematical Introduction to Robotic Manipulation.* CRC Press, Baton Rouge, 1994.

24. Povot X. *Deformation Constraints in a Mass-Spring Model to Describe Rigid Cloth Behavior* Graphics Interface '95, Canadian Human-Computer Communications Society, W. A. Davis and P. Prusinkiewicz (editors), pages 147–154,1995.

25. Pritchard D. and Heidrich W. Cloth motion capture. *Eurographics*, 22(3):37–51, 2003.

26. Richards J. The measurement of human motion: A comparison of commercially available systems. *Human Movement Science*, 18:589–602, 1999.

27. Rosenhahn B., Brox T., Cremers D. and Seidel H.-P. A comparison of shape matching methods for contour based pose estimation. In R.Reulke, U.Eckhardt, B.Flach and U.Knauer, editors, *Proc. 11th Int. Workshop Combinatorial Image Analysis*, volume 4040 of *Lecture Notes in Computer Science*, pages 263–276, Berlin, 2006. Springer.

28. Rosenhahn B., Brox T., Kersting U., Smith A., Gurney J. and Klette R. A system for marker-less motion capture. *Künstliche Intelligenz*, (1):45–51, 2006.

29. Rosenhahn B., Kersting U., Powell K. and Seidel H.-P. Cloth x-ray: Mocap of people wearing textiles. In *Accepted: Pattern Recognition, 28th DAGM-symposium*, Lecture Notes in Computer Science, Berlin, Germany, September 2006. Springer.

30. Rosenhahn B., Kersting U., Smith A., Gurney J., Brox T. and Klette R. A system for marker-less human motion estimation. In W.Kropatsch, R.Sablatnig and A.Hanbury, editors, *Pattern Recognition, 27th DAGM-symposium*, volume 3663 of *Lecture Notes in Computer Science*, pages 230–237, Vienna, Austria, September 2005. Springer.

31. Rosenhahn B. and Sommer G.. Pose estimation of free-form objects. In T.Pajdla and J.Matas, editors, *Computer Vision - Proc.8th European Conference on Computer Vision*, volume 3021 of *Lecture Notes in Computer Science*, pages 414–427. Springer, May 2004.

32. Rousson M. and Paragios N. Shape priors for level set representations. In A.Heyden, G.Sparr, M.Nielsen and P.Johansen, editors, *Computer Vision – ECCV 2002*, volume 2351 of *Lecture Notes in Computer Science*, pages 78–92. Springer, Berlin, 2002.

33. Rusinkiewicz S. and Levoy M. Efficient variants of the ICP algorithm. In *Proc.3rdIntl. Conf. on 3-D Digital Imaging and Modeling*, pages 224–231, 2001.

34. Weil J. The synthesis of cloth objects. *Computer Graphics (Proc. SigGraph)*, 20(4):49–54, 1986.

35. Wikipedia. Motion capture. http://en.wikipedia.org/wiki/Motion_capture, September 2005.

36. You L. and Zhang J.J. Fast generation of 3d deformable moving surfaces. *IEEE Trans. Systems, Man and Cybernetics, Part B: Cybernetics*, 33(4):616–615, 2003.

37. Zhang Z. Iterative points matching for registration of free form curves and surfaces. *International Journal of Computer Vision*, 13(2):119–152, 1994.