

Capturing and Animating Skin Deformation in Human Motion

Sang Il Park*

Jessica K. Hodgins†

School of Computer Science
Carnegie Mellon University



Figure 1: Capture and animation of the dynamic motion of the surface of the human body.

Abstract

During dynamic activities, the surface of the human body moves in many subtle but visually significant ways: bending, bulging, jiggling, and stretching. We present a technique for capturing and animating those motions using a commercial motion capture system and approximately 350 markers. Although the number of markers is significantly larger than that used in conventional motion capture, it is only a sparse representation of the true shape of the body. We supplement this sparse sample with a detailed, actor-specific surface model. The motion of the skin can then be computed by segmenting the markers into the motion of a set of rigid parts and a residual deformation (approximated first as a quadratic transformation and then with radial basis functions). We demonstrate the power of this approach by capturing flexing muscles, high frequency motions, and abrupt decelerations on several actors. We compare these results both to conventional motion capture and skinning and to synchronized video of the actors.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation

Keywords: human animation, motion capture, skin deformation

1 Introduction

Optical motion capture has been used very successfully to create compelling human animations for movies and sports video games; however, it provides only a much simplified version of what we would see if we were to view a person actually performing those actions. The data contain an approximation of the motion of the

skeleton but miss such subtle effects as the bulging of muscles and the jiggling of flesh. The current state of the art for whole body capture uses a set of 40-60 markers and reduces it to the rigid body motion of 15-22 segments. To the extent possible, the markers are placed on joint axes and bony landmarks so that they can more easily be used to approximate the motion of the skeleton. Biomechanical invariants are often used to reduce the number of markers to less than the number required to fully specify the orientation of each limb.

In this paper, we take a different approach to motion capture and use a very large set of markers (approximately 350) placed not on bony landmarks but on the muscular and fleshy parts of the body. Our goal is to obtain not only the motion of the skeleton but also the motion of the surface of the skin.

We accurately reconstruct the motion of the surface of the body by applying the three-dimensional trajectories for this dense marker set to a subject-specific polygonal model (Figure 1). The polygonal model is first optimized to fit the three-dimensional locations of the markers from a static pose. During the motion, the rigid body motion of the dense marker set is extracted and the remaining motion of the markers is used to compute local deformations of the polygonal model. The position of occluded markers is estimated from the locations of neighboring markers using a local model of the surface shape. The deformations of the marker set allow the muscle shapes in the polygonal model to grow and shrink and the fleshy areas to move dynamically.

To demonstrate the viability of this technique, we captured the motion of two subjects: a male football player (university-level offensive lineman) and a female professional belly dancer. Both subjects exhibited significant muscle and skin deformation when they performed dynamic activities. To evaluate the results, we compared motion captured with the dense marker set to synchronized video and to similar motions captured with a standard marker set and rendered using the same model and the skinning techniques available in commercial software.

2 Background

Data-driven approaches such as the one described in this paper are only one possible way to create an animation of the deformations

*e-mail: sipark@cs.cmu.edu

†e-mail: jkh@cmu.edu

of the human body as it moves. Other approaches that have been used successfully include skinning from example and anatomically based models such as physical simulations of muscles. We now describe each technique and the classes of the deformations it is capable of modeling. We discuss the face and hands separately because some techniques are applicable to those body parts in particular.

The first techniques to be developed required the modeler to specify the contribution of each bone to the position of the vertices by painting weights on the model (this approach is nicely described by Lewis and colleagues [2000]). These techniques, which have been variously called skeleton subspace deformation, single weight enveloping, and simply skinning, are easy to implement and fast to compute and therefore remain in use today. Unfortunately, with the basic implementation of this algorithm, no set of weights will prevent collapsing joints or the “candy wrapper” effect because the volume of the body is not preserved. A number of recent solutions attempt to fix these problems in an automatic fashion without incurring significant computational cost: interpolating spherical rotations rather than performing linear interpolation [Kavan and Zara 2005], approximating the model by swept ellipsoids [Hyun et al. 2005], adding deformable chunks under the surface of the skin to provide a very simple model of muscle and fat [Guo and Wong 2005], or constructing a simple anatomical model from the outside in [Pratscher et al. 2005]. These techniques model the bending of the skin around joints but cannot show dynamic effects such as jiggling of the flesh or muscle bulging due to exertion.

The next set of techniques also determine how the surface should deform as a function of the pose of the character but use examples to provide hints about the shape of the skin at key poses. Lewis and his colleagues [2000] combined this idea with skeleton subspace deformation. Sloan and his fellow researchers [2001] used radial basis functions to interpolate efficiently among a set of examples. Wang and Phillips [2002] allowed the skinning to have more degrees of freedom (and therefore more weights). Mohr and Gleicher [2003] used a similar approach except that they introduced the degrees of freedom required to incorporate the examples as additional joints rather than as skinning weights. Each of these techniques used pose-specific models for the character. The examples were created by hand and leveraged the insight of the modeler regarding the shape of the body and the poses to be specified. These techniques can model changes in the shape of the body as a function of pose but do not model dynamic effects such as changes in shape as a function of the torque being applied at a joint.

Other data-driven approaches rely on scanned models for the example poses. Allen and his colleagues [2002] used scanned data for the upper body with correlated poses from motion capture to interpolate a subdivision surface model and animate new sequences of motion capture data. The SCAPE system used a set of high resolution scans for one subject to animate the body shapes of new subjects based on a single scan and a set of marker positions [Anguelov et al. 2005]. They decoupled the motion into a rigid body component and a residual deformation as we do. As with the previous data-driven approaches, this system captured joint angle-specific deformations that were present in the data but could not capture dynamic effects in the surface of the skin.

Another alternative is to continuously capture skin deformation as Sand and his colleagues [2003] did and as we do. These data will necessarily be at a lower resolution than a laser scan. Sand and his colleagues used a conventional marker set to capture the motion of the skeleton and then extracted the surface model from silhouette information captured with three video cameras. Because silhouettes do not completely specify the geometry at any moment in time, they generalized observations based on the joint angles of neighboring limbs. This system could not capture dynamic effects that were not

visible in the silhouette.

There is a large body of work on modeling and simulating the underlying musculature of the human body (for example [Scheepers et al. 1997; Wilhelms and Gelder 1997; Nedel and Thalmann 2000; Teran et al. 2005a]). A number of these systems tackle the very difficult problem of not only modeling the complex anatomy but also simulating a part of its functionality. In particular, a several of these systems model the flex of muscles as they apply joint torques. Recently, Teran and his colleagues [2005b] proposed modeling the quasi-static movement of flesh with finite elements. Simple dynamic models have been used to add dynamic effects to the motion of skin [Larboulette et al. 2005]. No simulations have been constructed that combine a complete model of the human body with a dynamic simulation and control system in part because of the difficulty of controlling it to perform a task. However, the control system for some behaviors can be simpler, as Zordan and his colleagues [2004] demonstrated with their anatomical model and control of breathing. The most anatomically accurate models do not show the motion of the skin but show muscle shape and the resulting action (for example [Dong et al. 2002; Lemos et al. 2005]).

The hands and face are in some sense special cases for the problem of animating skin motion: the face because its planarity makes it more amenable to capture and the hands because their bony anatomical structure makes them more amenable to anatomical modeling [Albrecht et al. 2003; Kurihara and Miyata 2004]. Because the motion of the face cannot be reasonably approximated by rigid body motion, facial animation has motivated the study of techniques for cleaning and using dense marker sets [Guenter et al. 1998; Lin and Ouhyoung 2005]. Facial deformations do not include significant occlusions and are appropriate for capture techniques such as structured light [Zhang et al. 2004; Wang et al. 2004].

3 Overview

We present a data-driven approach to capturing and animating the surface of a human character’s skin. The data is captured by placing several hundred small markers on an actor. Although the number of markers is large, the motion of the markers does not fully capture the motion of the skin. To supplement this sparse sample, we use a high-resolution, subject-specific surface model. We address two main issues in this work: processing missing/noisy markers and deforming the reference model to match the marker motion. Because of the large number of small markers and occlusions by other body parts, the three-dimensional (3D) positions of the markers from the motion capture device frequently exhibit missing and disconnected segments. We employ a local model defined on each marker by its neighbors to merge the disconnected segments and fill in the missing markers. To best fit the marker positions to the reference model, we divide the marker motion into a set of rigid motions and approximate the residual deformation of each part. These two contributions are discussed in detail in the next two sections. Section 6 discusses our results and compares the animated motion to alternative techniques for model deformation and to synchronized video and conventional motion capture. The last section of the paper discusses the limitations of our approach and future work.

4 Data Collection and Cleaning

We captured the movement of the markers with a commercial optical motion capture system consisting of 12 near infra-red cameras with 4 megapixel resolution capturing at a rate of 120

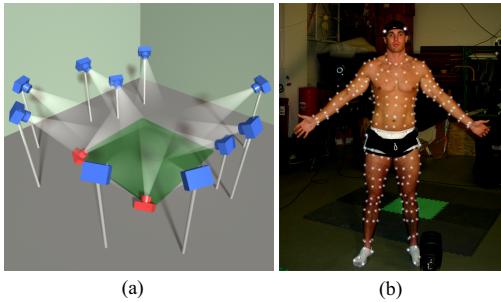


Figure 2: Capture setup: (a) Twelve cameras surrounding a small capture region. Two cameras (shown in red) were aimed up rather than down to capture downward facing markers; (b) 350 small markers attached to the subject’s body.

frames/second [Vicon Motion Systems 2006]. To increase accuracy, we positioned the cameras close to the subject with a small capture region (approximately 2 m by 2 m by 2.5 m high). Two cameras were aimed up rather than down to capture markers facing toward the ground. Figure 2(a) shows the camera configuration.

We placed approximately 350 reflective markers on the subject’s body. To capture the subtle movement of the skin effectively, we chose small markers (diameter 3.0 mm) with a hemispherical shape to minimize the offset of the markers from the body. Although it is not necessary to have an even distribution of the markers on the body, we drew an approximate grid on the subject’s body, and placed the markers on that grid. The average distance between two neighboring markers was 4-5 cm. We supplemented the grid pattern with additional markers in areas where more resolution would likely be needed (the tip of the elbow and the point on the lower part of the shoulder blade, for example). Figure 2(b) shows the marker placement.

We reconstructed the 3D positions of the markers with the VICON IQ 2.0 software [Vicon Motion Systems 2006]. In conventional motion capture, reconstruction is aided by a skeleton model and a rigid link assumption; however, because of the significant deformations in our captures, we could not make this assumption. This assumption is also not valid for motion capture of the face, but our whole body capture was fully 3D and therefore contained many more occlusions than are seen in the more nearly 2D data captured from the face. Occlusions are difficult to handle because they occur simultaneously in regions.

In the next section, we present a novel method for cleaning and recovering damaged marker data based on a local reference frame defined at each marker and the spatial relationship with neighboring markers. We first define the local reference frame and then explain how it can be used to clean and recover corrupted data by identifying trajectories that can be merged and facilitating hole filling. The final step in the cleaning process is smoothing.

4.1 Local Reference Frame

We first select a pose as the *reference pose* and use it to estimate the local shape of the body. We selected the pose shown in Figure 2(b) as the reference pose because few markers were occluded. We assume that subjects begin each motion in the reference pose and that there are no missing markers in the first frame. This last assumption was reasonable for our experiments because only two or three markers near the armpit were occluded and these could be filled in manually by referring to frames where they were visible. We create

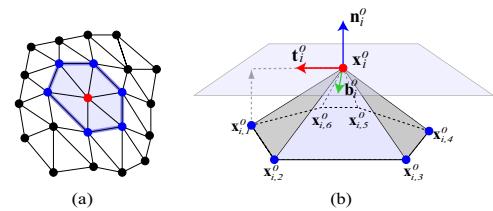


Figure 3: (a) a marker (red dot) and its one-ring neighbors (blue dots); (b) a local reference frame on the marker.

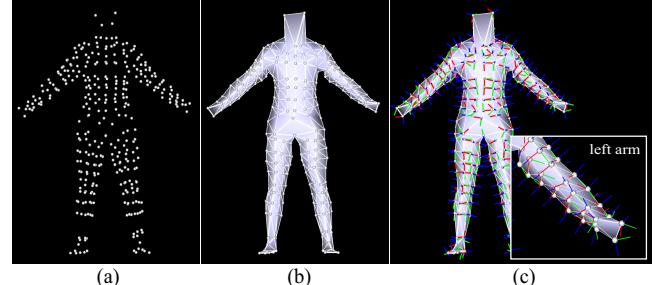


Figure 4: The reference pose: (a) markers; (b) marker surface; (c) local frame at each marker.

a mesh from the marker positions in the reference pose. Neighboring markers are identified based on the geodesic distance along the skin. This process was done manually for our experiments. We call the resulting mesh the *marker surface* (Figure 4(a) and (b)). The indices of the vertices on the surface are assigned to the markers as labels. The construction of the marker surface is done only once for each subject.

On the marker surface, we denote the marker with index i as m_i , $1 \leq i \leq N$, where N is the number of markers, and its one-ring neighbor as $m_{i,k}$, $1 \leq k \leq d_i$, where d_i is the valence of marker i . We denote the position of the i -th marker by \mathbf{x}_i^0 and the position of its one-ring neighbors as $\mathbf{x}_{i,k}^0$ in the reference pose. Using a similar technique to that of Lipman and his colleagues [2005], we define the local frame for marker i in the reference pose with its origin located at \mathbf{x}_i^0 and the triplet $(\mathbf{t}_i^0, \mathbf{b}_i^0, \mathbf{n}_i^0)$, where \mathbf{n}_i^0 is the normal vector of the marker surface at the i -th marker in the reference pose, \mathbf{t}_i^0 is a unit vector with direction equal to the projection of the vector $(\mathbf{x}_{i,1}^0 - \mathbf{x}_i^0)$ onto the tangential plane defined by \mathbf{n}_i^0 , and \mathbf{b}_i^0 is a unit vector orthogonal both to \mathbf{n}_i^0 and \mathbf{t}_i^0 . We call this local frame of the reference pose the *local reference frame*. Note that \mathbf{t}_i^0 can be defined using any of the one-ring neighbors, and our choice of $\mathbf{x}_{i,1}^0$ is arbitrary. Figure 3 illustrates the definition of the local frame.

The position $\hat{\mathbf{x}}_{i,k}^0$ of the k -th 1-ring neighbor measured in the local reference frame of marker i is

$$\hat{\mathbf{x}}_{i,k}^0 = \mathbf{R}_i^0(\mathbf{x}_{i,k}^0 - \mathbf{x}_i^0), \quad (1)$$

where $\mathbf{R}_i^0 \in \mathbb{R}^{3 \times 3}$ is a rotation matrix defined as $[\mathbf{t}_i^0 \ \mathbf{b}_i^0 \ \mathbf{n}_i^0]^T$.

4.2 Merging Disconnected Trajectories

The 3D marker data from the optical motion capture system consists of a set of reconstructed marker trajectories. However, as mentioned above, a marker trajectory may be broken into many partial

trajectories because markers are frequently occluded or confused with a nearby marker. Because the markers at the first frame were already labeled during the construction of the mesh surface, the partial trajectories that include those markers inherit the labels while the other trajectories remain unlabeled. The unlabeled trajectories then need to be merged with labeled trajectories. Estimating which trajectories to merge based on the last known position and velocity of a marker would likely not work well for highly dynamic motions and long occlusions. Instead, we observe that the topology of markers placed on the skin does not change and estimate the position of occluded markers from the position of the visible neighboring markers.

At frame t , we denote the instantaneous set of available markers as Y^t . Suppose that marker m_i is missing while some of its neighbors are available ($m_{i,k} \in Y^t$). We first find the translation $\mathbf{d}_i^t \in \mathbb{R}^{3 \times 1}$ and the rotation $\mathbf{R}_i^t \in \mathbb{R}^{3 \times 3}$ that move available neighbors from their current position $\mathbf{x}_{i,k}^t$ approximately to their reference position $\hat{\mathbf{x}}_{i,k}^0$ in the local reference frame for m_i . Then, we estimate the current position of the marker m_i by bringing the origin of the local reference frame back to the global frame.

The neighbors have not necessarily moved rigidly as a group because of deformations so the calculation of the rigid transformation is a least squares optimization:

$$\operatorname{argmin}_{\mathbf{d}_i^t, \mathbf{R}_i^t} \sum_{k|m_{i,k} \in Y^t}^{d_i} \|\mathbf{R}_i^t \cdot (\mathbf{x}_{i,k}^t + \mathbf{d}_i^t) - \hat{\mathbf{x}}_{i,k}^0\|^2. \quad (2)$$

This is a typical form of the well-known absolute orientation problem in shape matching and several analytical solutions have been published (see [Kanatani 1994] for a comparison of the different schemes). In our implementation, we adapt the method of [Horn 1987].

A minimum of three neighboring markers must be available to compute the rigid transformation. If more than three neighboring markers are available, the estimate should be more robust to noise. Therefore, we perform this computation in a greedy fashion; we first estimate the missing markers that have the largest number of available neighbors and then estimate those with fewer neighbors (including previously estimated marker positions in the calculation).

Now that the current position of all markers has been estimated, we can search for trajectories to merge. For each missing marker m_i , the system searches for an unlabelled trajectory that is within a threshold distance ε_i^1 of the marker's estimated position and which does not overlap in time with the existing trajectory of m_i . The threshold is selected based on the average distance from a marker to its neighbors in the reference pose:

$$\varepsilon_i^1 = \alpha^1 \frac{\sum_{k=1}^{d_i} \|\hat{\mathbf{x}}_{i,k}^0\|}{d_i}, \quad (3)$$

where α^1 is a tolerance.

This matching process is not perfect because a marker from a different part of the body might be close enough to the estimated position of the missing marker to create an incorrect merge. For example, a marker on the upper arm near the armpit might be confused with nearby markers on the side of the torso. To disambiguate marker trajectories from other body parts, we test that the entire candidate trajectory is within the threshold ε_i^2 of the position estimated from the neighbors at each frame. The threshold ε_i^2 is computed in a similar fashion to ε_i^1 with tolerance parameter α^2 . We set α^1 and α^2 to 0.5 and 0.6 in our experiments.

This algorithm may fail to find matching trajectories during motions with extreme local deformations because those deformations do not fit into the α^1 and α^2 tolerances. Thus, we merged the valid unlabeled trajectories manually after the process. Matching trajectories might not exist for some markers in some frames because of occlusion, which results in gaps or holes in the trajectories of the marker. We resolve these in the hole filling process described in the next section.

4.3 Hole Filling

We fill the holes in the merged trajectories by learning a statistical model of the spatial relationship between each marker and its neighbors. We apply Principal Component Analysis (PCA) to the position of each marker and its one-ring neighbors throughout the entire motion. This operation reduces the local deformation around the marker to a low dimensional control space and allows us to reconstruct the marker location by estimating the best fit to the available markers.

We first transform the position of each marker and its neighbors to a local frame. For the marker m_i at frame t , the position of the neighbors in the local frame is

$$\hat{\mathbf{x}}_{i,k}^t = \bar{\mathbf{R}}_i^t (\mathbf{x}_{i,k}^t - \mathbf{x}_i^t), \quad (4)$$

where \mathbf{x}_i^t and $\hat{\mathbf{x}}_{i,k}^t$ are the global position of the marker i and the local position of its k -th neighbor, respectively, and $\bar{\mathbf{R}}_i^t$ is a rotation matrix. Ideally, $\bar{\mathbf{R}}_i^t$ would be the rotation matrix that transforms the marker positions to the local frame as described in Section 4.1, but because we are dealing with damaged data, it may not always be possible to find that transformation (marker $m_{i,1}$ may be missing). Instead, we select the rotation matrix $\bar{\mathbf{R}}_i^t$ that puts $\hat{\mathbf{x}}_{i,k}^t$ as close as possible to the local reference position:

$$\operatorname{argmin}_{\bar{\mathbf{R}}_i^t} \sum_{k|m_{i,k} \in Y^t}^{d_i} \|\hat{\mathbf{x}}_{i,k}^t - \hat{\mathbf{x}}_{i,k}^0\|^2. \quad (5)$$

This definition of the local frame is consistent even with damaged data. The method of [Horn 1987] is used for optimizing Equation 5.

Next, we build a PCA model for each marker using the frames in which the marker and all of its neighbors are present. PCA allows us to represent the neighborhood of a marker (in the local frame) in a reduced orthogonal basis while retaining the desired accuracy (99% in our implementation). The system uses this model to estimate the positions of missing markers by selecting the coefficient of each basis vector to minimize the squared distance between the reconstructed positions of the available neighbors and their measured positions. If a marker is missing in a frame, the PCA models for each of its neighbors are used to estimate a value for the marker. Those values are averaged to improve the robustness of the estimation.

4.4 Smoothing

We apply a time-domain filter to each marker trajectory to reduce noise. Because the rigid body motions of the human body are large in most of our experiments, filtering the global trajectory would adversely affect both the rigid body motion and the shape of the body. Instead, we filter in the local frame defined in Equation (2). The system applies a smoothing filter several times to the local trajectory based on the update rule $\hat{\mathbf{x}}_i^t \leftarrow \hat{\mathbf{x}}_i^t - \lambda \Delta^4 \hat{\mathbf{x}}_i^t$, where λ is a damping factor controlling the rate of convergence and Δ is a derivative

operator in time. The filtered position in the local frame is then transformed back to the global frame.

5 Skin Animation

Given the locations of the markers at each frame, the skin animation problem can be formulated as a scattered data interpolation problem of deforming the vertices of the detailed model to match the marker locations. However, the space of deformations for humans is likely too complex to be approximated by a standard linear interpolation technique such as radial basis interpolation.

Instead, we reduce the complexity of the problem by factoring the deformation space into two pieces. We first segment the full body into a set of rigid parts, and then focus on the local deformation that remains after the rigid body motion is removed. The residual deformation still consists of higher-order deformations such as twisting and bulging, and we represent it as a linear combination of a set of predefined primitive deformation modes. Finally, the remaining residuals are resolved with radial basis interpolation.

5.1 Registration

We created a detailed model of each subject by photographing the subject and asking a professional artist to make a model for the subject using commercial modeling software (Maya 7). We provided size information for the subject by importing the marker surface for the reference pose into Maya. In addition to serving as a reference, the photographs were used to create textures for the model.

Although the artist used the marker surface to size the model, the final model is not a perfect fit to the reference marker surface. We adjust the detailed surface to fit the marker surface using optimization. We adopt the method of Allen and his colleagues [2003] for registering two surfaces and find a global transformation for each vertex that minimizes an objective function with three terms: minimize the sum of the distance from each marker to the surface, preserve the original shape of the detailed surface model by minimizing the sum of the difference between the global transformations of pairs of neighboring vertices, and maintain a correspondence between the markers and the vertices at 40 important landmarks of the body (elbows, knees, etc.). Initially, we assign weights only to the second and third terms so that the model is registered globally first. After the convergence, we perform the optimization again with equal weight for all the terms.

5.2 Segmentation and Weighting

For near-rigid segmentation, we used the method of James and Twigg [2005], in which mean shift clustering is used to group triangles having similar rotational movement. We apply this method to the marker surface captured while the subject was demonstrating the range of motion of each of his or her joints. Figure 5(a) shows the result after mean shift clustering. The large deformation around the upper arms and shoulders is not handled well and mean shift generates many groups containing just one triangle. We manually adjust the result so that the body is divided into 17 parts (Figure 5(b)). This segmentation is done only once for each subject.

We assign a set of weight values $\alpha_{i,p}$, $1 \leq p \leq N_p$, to the markers, where $\alpha_{i,p}$ is the weight value of the marker m_i for part p , N_p is the total number of parts and $\sum_{p=1}^{N_p} \alpha_{i,p} = 1$. The weight for a part is 1

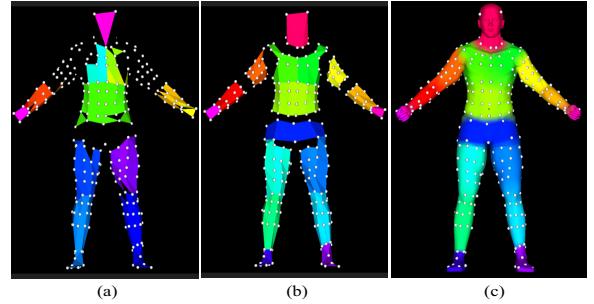


Figure 5: Near-rigid segmentation: (a) after the mean shift segmentation (triangles of the same color belong to the same group while the black colored triangles do not belong to any group); (b) after the manual adjustment of the segmentation; (c) weight distribution on the detailed surface.

if the marker is inside the part and 0 for all other parts. For markers on the boundary between parts, we determine the weight values by computing the likelihood that the marker belongs to each neighboring part. The system compares the marker's motion to the rigid body motion of the neighboring parts (as defined by the markers that are labeled as fully inside the part). Based on the error value, we employ a gaussian model for computing the likelihood that a marker belongs to a neighboring part [Anguelov et al. 2004], and we assign the marker to the part if its weight is more than a pre-defined threshold (0.2 in our experiments). Markers may belong to multiple parts.

Finally, the weight for each vertex of the detailed model is determined by finding the nearest triangle of the marker surface and assigning the weight obtained by interpolating the weights of the three markers based on the barycentric coordinate of the projected vertex position on the triangle. We denote the weight value for part p of the vertex j as $\beta_{j,p}$. Figure 5(c) shows the weight distribution of the detailed model.

5.3 Deformation

Based on the segmentation, we factor the full body motion into a rigid body transformation of the segmented parts and their local deformation. The local deformation is further divided into two terms: a quadratic deformation and its residual.

We first define the local frame for each rigid part in the reference pose. We call this frame the *local part reference frame*. The origin of the local frame is the average of the reference pose position of all markers assigned to that part. The three axes of the local frame are the three eigenvectors of the covariance matrix of the markers' positions. We define the rigid translation and rotation of the part p which transform a global position into its local position as $\tilde{\mathbf{d}}_p^0 \in \mathbb{R}^{3 \times 1}$ and $\tilde{\mathbf{R}}_p^0 \in \mathbb{R}^{3 \times 3}$. We denote the position of the i -th member marker assigned to part p as $\mathbf{p}_{p,i}$, $1 \leq i \leq Y_p$, where Y_p is the number of markers assigned to part p . For a position $\mathbf{p}_{p,i}^0$ at the reference pose, the local position $\hat{\mathbf{p}}_{p,i}^0$ is computed as $\hat{\mathbf{p}}_{p,i}^0 = \tilde{\mathbf{R}}_p^0(\mathbf{p}_{p,i}^0 + \tilde{\mathbf{d}}_p^0)$. We call this the *local part reference position* of the marker.

The rigid transformation $(\tilde{\mathbf{d}}_p^t$ and $\tilde{\mathbf{R}}_p^t)$ for a given part p at frame t is computed so that it brings the position of all member markers as close as possible to their local part reference positions using the absolute orientation method [Horn 1987]. The local position $\hat{\mathbf{p}}_{p,i}^t$ is computed as $\hat{\mathbf{p}}_{p,i}^t = \tilde{\mathbf{R}}_p^t(\mathbf{p}_{p,i}^t + \tilde{\mathbf{d}}_p^t)$. The remaining error between

$\hat{\mathbf{p}}_{p,i}^t$ and the reference position for that marker is the local deformation.

We approximate the local deformation of each part with a continuous deformation field. Human body movements include non-linear deformations such as twisting, bulging and bending. Inspired by the work of Müller and his colleagues [2005], we choose quadratic deformations to represent the local deformation. With this approach, a complex non-linear deformation is modeled as a linear combination of 3×9 basic deformation modes. The quadratic transformation is defined as a matrix $\tilde{\mathbf{A}} = [\mathbf{A}_1 \mathbf{A}_2 \mathbf{A}_3] \in \mathbb{R}^{3 \times 9}$, where $\mathbf{A}_1 \in \mathbb{R}^{3 \times 3}$ corresponds to a linear transformation, $\mathbf{A}_2 \in \mathbb{R}^{3 \times 3}$ and $\mathbf{A}_3 \in \mathbb{R}^{3 \times 3}$ are a pure quadratic and a mixed quadratic transformation, respectively. Given a 3D position $\mathbf{p} = [p_x \ p_y \ p_z]^T$, the quadratic transformation provides a new transformed position $\tilde{\mathbf{p}}$: $\tilde{\mathbf{p}} = \tilde{\mathbf{A}}\mathbf{q}$, where $\mathbf{q} = [p_x, p_y, p_z, p_x^2, p_y^2, p_z^2, p_xp_y, p_yp_z, p_zp_x]^T$ is a 9 dimensional quadratic vector corresponding to \mathbf{p} .

At each frame, we compute the components of the quadratic transformation $\tilde{\mathbf{A}}_p^t$ of part p such that the transformation brings the local part reference position of all member markers as close as possible to their local positions at frame t . We use the pseudo inverse to solve for this transformation:

$$\tilde{\mathbf{A}}_p^t = \hat{\mathbf{P}}^t (\mathbf{Q}^0)^T \left[\mathbf{Q}^0 (\mathbf{Q}^0)^T \right]^{-1}, \quad (6)$$

where $\hat{\mathbf{P}}^t = [\hat{\mathbf{p}}_{p,1}^t, \dots, \hat{\mathbf{p}}_{p,Y_p}^t] \in \mathbb{R}^{3 \times Y_p}$, $\mathbf{Q}^0 = [\mathbf{q}_{p,1}^0, \dots, \mathbf{q}_{p,Y_p}^0] \in \mathbb{R}^{9 \times Y_p}$ and $\mathbf{q}_{p,i}^0$ is a quadratic vector corresponding to the local part reference position $\mathbf{p}_{p,i}^0$.

Finally, given the transformed position $\tilde{\mathbf{p}}_{p,i}^t = \tilde{\mathbf{A}}_p^t \mathbf{q}_{p,i}^0$, we use radial basis interpolation to resolve the remaining residual $\mathbf{r}_{p,i}^t = \mathbf{p}_{p,i}^t - \tilde{\mathbf{p}}_{p,i}^t$ by determining the weight vector $\mathbf{w}_{p,i}^t$ such that

$$\mathbf{r}_{p,j}^t = \sum_{i=1}^{Y_p} \mathbf{w}_{p,i}^t \phi \left(\frac{\|\tilde{\mathbf{p}}_{p,i}^t - \tilde{\mathbf{p}}_{p,j}^t\|}{\sigma_p} \right), \text{ for } 1 \leq j \leq Y_p, \quad (7)$$

where $\phi(\cdot)$ is a radial basis function and σ_p is a dilation factor for part p . In our experiments, we use a cubic B-spline as the radial basis function and set the dilation factor to be twice the maximum distance between two nearest member markers for each part.

Provided $\tilde{\mathbf{d}}_p^t$, $\tilde{\mathbf{R}}_p^t$, $\tilde{\mathbf{A}}_p^t$ and $\mathbf{w}_p^t = \{w_{p,1}^t, \dots, w_{p,Y_p}^t\}$ at frame t , any given position $\hat{\mathbf{p}}_p^0$ represented in the local part reference frame of part p can be transformed back to its global position \mathbf{p}^t as

$$\mathbf{p}^t = (\tilde{\mathbf{R}}_p^t)^{-1} \cdot \left(\tilde{\mathbf{A}}_p^t \hat{\mathbf{q}}_p^0 + \sum_i \mathbf{w}_{p,i}^t \phi \left(\frac{\|\tilde{\mathbf{A}}_p^t \hat{\mathbf{q}}_p^0 - \tilde{\mathbf{p}}_{p,i}^t\|}{\sigma_p} \right) \right) - \tilde{\mathbf{d}}_p^t, \quad (8)$$

where $\hat{\mathbf{q}}_p^0$ is the quadratic vector of $\hat{\mathbf{p}}_p^0$. Consequently, for a given position \mathbf{v}_j^0 of the vertex j of the detailed model in the reference pose, its deformed position is computed as

$$\mathbf{v}_j^t = \sum_{p=1}^{N_p} \beta_{j,p} \mathbf{v}_{j,p}^t, \quad (9)$$

where $\mathbf{v}_{j,p}^t$ is the deformed positions related to part p obtained by Equation (8). We animate the model by transferring the reference position using Equation (9).

Table 1: The number of trajectories before and after the merging process, the number of incorrectly labeled partial trajectories, the number of manual corrections for the missing unlabeled trajectories, and the number of holes before the hole filling process. The number of markers was 350.

example motions (# of frames)	# of trajectories		incorrectly labeled (error rate)	manual merging	total # of holes
	before merging	after merging			
flexing (917)	859	379	0 (0.0%)	4	25674
golfing (868)	2368	439	10 (0.4%)	32	36232
punching (574)	2247	409	25 (1.1%)	44	22948
jump rope (974)	5006	538	32 (0.6%)	64	42050

Table 2: The accuracy of the hole-filling algorithm for the punching motion. The error values are normalized by the average distance to the neighbors.

region	abdomen	elbow	thigh	knee
Avg. Error	0.017	0.022	0.020	0.023
Max. Error	0.052	0.062	0.045	0.051

6 Experimental Results

We first show the effectiveness of the clean-up algorithm. Table 1 shows the number of trajectories before and after the merging process and the number of holes before the hole filling process. More dynamic motions tend to have more disconnected trajectories. This table also shows the number of partial trajectories assigned wrong labels and the number of manual corrections required to merge missing unlabeled trajectories. Those wrong and missing trajectories occurred mainly around the elbows during large deformations. Because of noise, the number of the trajectories after merging is still more than the number of markers; the extra trajectories usually contained severe (one or two frames long) outliers and are ignored in the later stages of processing. We tested the accuracy of our hole-filling algorithm by deleting a marker that was seen in the data and reconstructing it (Table 2). The average error was about 2%, which is acceptable for visual quality. Figure 6 shows that our hole filling algorithm works well even though a large number of the markers are missing.

Next, we compare our deformation method with three different methods: rigid deformation for each part without resolving residuals (Figure 7(b)), quadratic deformation without resolving residuals (Figure 7(c)) and rigid deformation with resolving residuals (Figure 7(d)). Our method, quadratic deformation with resolving residuals, is shown in Figure 7(e). Quadratic deformation gives a smoother deformation and follows the marker positions better due to its higher order deformation terms. Because the residuals are significantly smaller in quadratic deformation than in the rigid one, the results after resolving the residuals are also better when the quadratic deformations are used.

We applied our method to various behaviors including slow motions such as breathing (Figure 8(a)), and golfing (Figure 1) as well as highly dynamic motions such as punching (Figure 1), jumping rope (Figure 8(b)) and belly dancing (Figure 9). We compare our synthesized results with video taken during the capture (Figure 8(c) and (d)). We also compared our results with those from conventional motion capture (Figure 8(e) and (f)). Subjects wearing conventional marker sets were asked to perform similar motions to those captured with the 350 marker set. As shown in the figures, our method successfully captured the expanding and contracting of breathing and the dynamic deformations of the surface of the torso during jumping. Conventional methods failed to capture these effects.

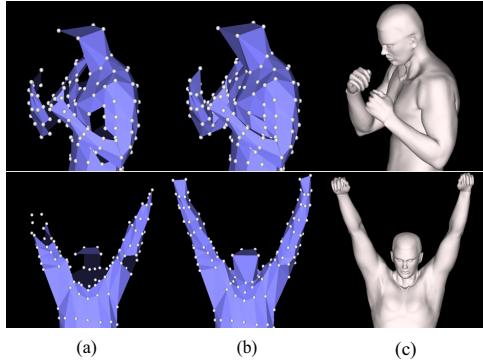


Figure 6: Hole filling: (a) markers with holes; (b) after hole filling; (c) a detailed surface model

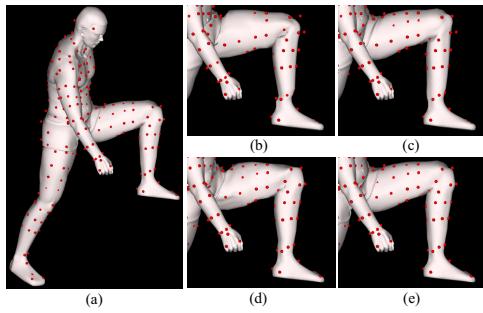


Figure 7: Comparison of different deformation methods; (a) an example pose with the markers; (b) rigid deformation without resolving residuals; (c) quadratic deformation without resolving residuals; (d) rigid deformation with residuals resolved; (e) quadratic deformations with residuals resolved (our method).

7 Discussion

In this paper we have demonstrated that data captured from a large set of markers can be used to animate the natural bending, bulging, and jiggling of the human form. Our contribution is twofold. First, we provide algorithms for estimating the location of missing markers. Because the human body is articulated and our marker set is dense, markers are frequently occluded by other parts of the body. Therefore, missing markers are far more common in our captures than in facial animation, the other domain where a dense marker set has been used. Our second contribution is an algorithm for deforming a subject-specific model to match the trajectories of the markers. We do this in two stages, first capturing the rigid body motion of the markers and then resolving the residual deformations. Our approach to this problem allowed us to animate the motion of a muscular male punching and jumping rope and the performance of a professional belly dancer. The animated deformations of their bodies are dramatic and perhaps unexpected, but we demonstrate that they are a good match to video captured simultaneously. We also demonstrate that the subtleties of the skin motion would not have been captured by a standard marker set working in concert with the skinning algorithms available in commercial animation software.

Although we chose to use hand-designed models, we believe that this approach would also work on a scanned model if such a model were available for our subjects. The models used in the SCAPE system had approximately 50,000 polygons [Anguelov et al. 2005]; our polygonal models had 54,000 polygons so the resolution is similar.

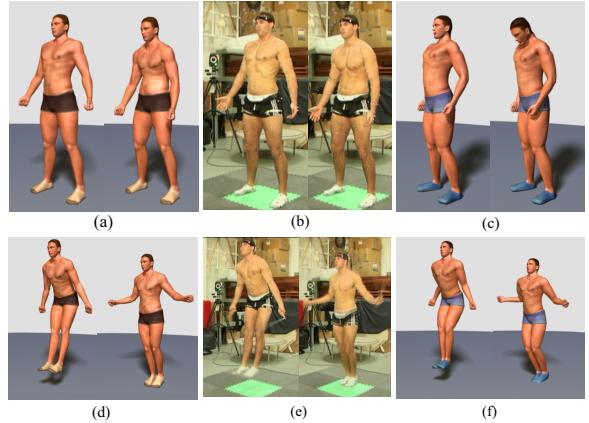


Figure 8: Comparison between our method and conventional motion capture: (a) and (d) breathing and jumping rope from our motion; (b) and (e) stills of video taken during capture (c) and (f) similar motions from conventional motion capture.

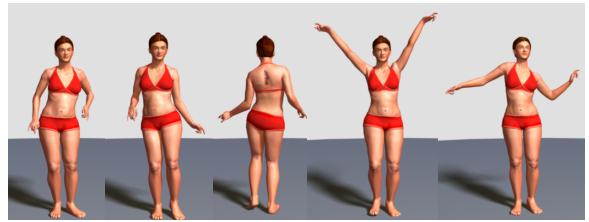


Figure 9: The belly dance

One limitation of our approach is its smaller capture area because the cameras are placed close to the subject for accuracy. Thus, motions requiring a large space such as running cannot be captured in our current system.

Our animations and models were specific to the particular actors who performed for us. Although it might be possible to parameterize the deformations seen in our capture sessions with joint angles (as others have done) and perhaps also with velocity in a world coordinate system for dynamic effects or joint torque computed from inverse dynamics for muscle bulging, we believe that it would be very difficult and probably impossible to generalize the dynamic captured motion to an animated character with a significantly different body type. The skin motion of a heavy person, for example, will not look realistic when applied to a skinny character.

Because we use the actual three-dimensional locations of the markers, we should be able to capture fine details that systems that use an approximation to the marker locations or a much sparser marker set will likely miss. For example, the pointy bone, the *olecranon*, on the outside of a bent elbow is sometimes smoothed out with other approaches, but an additional marker placed on the elbow captured that information in our system. We added markers to supplement our nominal grid that captured some of these details but a more anatomically based procedure for placing the markers might capture more of these subtle details.

During hole-filling, we determined the position of the missing markers by averaging the estimated positions from neighbors. However, the contributions of neighbors might be weighted with less weight given to the neighbors that were estimated rather than measured. When learning the PCA model, we only use the frames where a marker and all its neighbors are available, which decreases

the number of samples. Although this approach gave a reasonable result in our experiments, an enhanced PCA such as [Shum et al. 1995] could be applied to create a more comprehensive statistical model.

Our marker set was quite large and suiting up a subject for capture took about an hour in our experiments. Therefore, it would might be worth exploring whether these dynamic effects could be captured with a more parsimonious marker set by combining a conventional marker set with additional markers placed only on the areas of the body expected to deform such as the belly or biceps. A smaller marker set might be effective if the captured information was supplemented by a detailed anatomical model as was done for the face by Sifakis and his colleagues [2005].

Acknowledgements

The authors would like to thank Moshe Mahler for his help in modeling and rendering and Justin Macey for his assistance in the motion capture. This work was supported in part by NSF grant IIS-0326322 and CNS-0196217, and the partial support for the first author was provided by IT Scholarship Program of IITA (Institute for Information Technology Advancement) & MIC (Ministry of Information and Communication), Korea. Autodesk donated their MAYA software.

References

- ALBRECHT, I., HABER, J., AND SEIDEL, H.-P. 2003. Construction and animation of anatomically based human hand models. In *2003 ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 98–109.
- ALLEN, B., CURLESS, B., AND POPOVIĆ, Z. 2002. Articulated body deformation from range scan data. *ACM Transactions on Graphics* 21, 3, 612–619.
- ALLEN, B., CURLESS, B., AND POPOVIĆ, Z. 2003. The space of human body shapes: Reconstruction and parameterization from range scans. *ACM Transactions on Graphics* 22, 3, 587–594.
- ANGUELOV, D., KOLLER, D., PANG, H., SRINIVASAN, P., AND THRUN, S. 2004. Recovering articulated object models from 3d range data. In *the 20th Conference on Uncertainty in Artificial Intelligence*, 18–26.
- ANGUELOV, D., SRINIVASAN, P., KOLLER, D., THRUN, S., RODGERS, J., AND DAVIS, J. 2005. Scape: shape completion and animation of people. *ACM Transactions on Graphics* 24, 3, 408–416.
- CHADWICK, J. E., HAUMANN, D. R., AND PARENT, R. E. 1989. Layered construction for deformable animated characters. *Computer Graphics (Proceedings of SIGGRAPH 89)* 23, 3, 243–252.
- CHAI, J., XIAO, J., AND HODGINS, J. 2003. Vision-based control of 3d facial animation. In *2003 ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 193–206.
- CHOE, B., LEE, H., AND KO, H.-S. 2001. Performance-driven muscle-based facial animation. *The Journal of Visualization and Computer Animation* 12, 2, 67–79.
- CHUANG, E., AND BREGLER, C. 2005. Mood swings: expressive speech animation. *ACM Transactions on Graphics* 24, 2, 331–347.
- COSKER, D., PADDOCK, S., MARSHALL, D., ROSIN, P. L., AND RUSHTON, S. 2004. Towards perceptually realistic talking heads: models, methods and mcgurk. In *APGV 2004*, 151–157.
- DONG, F., CLAPWORTHY, G. J., KROKOS, M. A., AND YAO, J. 2002. An anatomy-based approach to human muscle modeling and deformation. *IEEE Transactions on Visualization and Computer Graphics* 8, 2, 154–170.
- FIDALEO, D., AND NEUMANN, U. 2004. Analysis of co-articulation regions for performance-driven facial animation. *Computer Animation and Virtual Worlds* 15, 1, 15–26.
- GUENTER, B., GRIMM, C., WOOD, D., MALVAR, H., AND PIGHIN, F. 1998. Making faces. In *Proceedings of SIGGRAPH 98*, Computer Graphics Proceedings, Annual Conference Series, 55–66.
- GUO, Z., AND WONG, K. C. 2005. Skinning with deformable chunks. *Computer Graphics Forum* 24, 3, 373–382.
- HORN, B. K. P. 1987. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A* 4, 4, 629–642.
- HUANG, K.-S., CHANG, C.-F., HSU, Y.-Y., AND YANG, S.-N. 2005. Key probe: a technique for animation keyframe extraction. *The Visual Computer* 21, 8-10, 532–541.
- HWANG, B.-W., AND LEE, S.-W. 2003. Reconstruction of partially damaged face images based on a morphable face model. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 3, 365–372.
- HYUN, D.-E., YOON, S.-H., CHANG, J.-W., SEONG, J.-K., KIM, M.-S., AND JÜTTLER, B. 2005. Sweep-based human deformation. *The Visual Computer* 21, 8-10, 542–550.
- IGARASHI, T., MOSCOVICH, T., AND HUGHES, J. F. 2005. As-rigid-as-possible shape manipulation. *ACM Transactions on Graphics* 24, 3, 1134–1141.
- JAMES, D. L., AND TWIGG, C. D. 2005. Skinning mesh animations. *ACM Transactions on Graphics* 24, 3, 399–407.
- KANATANI, K. 1994. Analysis of 3-d rotation fitting. *IEEE Trans. Pattern Anal. Mach. Intell.* 16, 5, 543–549.
- KAVAN, L., AND ZARA, J. 2005. Spherical blend skinning: A real-time deformation of articulated models. In *2005 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, ACM Press, 9–16.
- KSHIRSAGAR, S., MOLET, T., AND MAGNENAT-THALMANN, N. 2001. Principal components of expressive speech animation. In *Computer Graphics International 2001*, 38–44.
- KURIHARA, T., AND MIYATA, N. 2004. Modeling deformable human hands from medical images. In *2004 ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 355–363.
- LA TORRE, F. D., AND BLACK, M. J. 2001. Dynamic coupled component analysis. In *CVPR*, 643–650.
- LARBOULETTE, C., CANI, M.-P., AND ARNALDI, B. 2005. Dynamic skinning: adding real-time dynamic effects to an existing character animation. In *Spring Conference on Computer Graphics 2005*, 87–93.
- LEMOS, R. R., ROKNE, J., BARANOSKI, G. V. G., KAWAKAMI, Y., AND KURIHARA, T. 2005. Modeling and simulating the deformation of human skeletal muscle based on anatomy and physiology. *Computer Animation and Virtual Worlds* 16, 3-4, 319–330.

- LEWIS, J. P., CORDNER, M., AND FONG, N. 2000. Pose space deformations: A unified approach to shape interpolation and skeleton-driven deformation. In *Proceedings of ACM SIGGRAPH 2000*, Computer Graphics Proceedings, Annual Conference Series, 165–172.
- LIN, I.-C., AND OUHYOUNG, M. 2005. Mirror mocap: Automatic and efficient capture of dense 3d facial motion parameters from video. *The Visual Computer* 21, 6, 355–372.
- LIN, I.-C., YENG, J.-S., AND OUHYOUNG, M. 2002. Extracting 3d facial animation parameters from multiview video clips. *IEEE Computer Graphics & Applications* 22, 6, 72–80.
- LIPMAN, Y., SORKINE, O., LEVIN, D., AND COHEN-OR, D. 2005. Linear rotation-invariant coordinates for meshes. *ACM Transactions on Graphics* 24, 3, 479–487.
- MAGNENAT-THALMANN, N., AND THALMANN, D. 2005. Virtual humans: thirty years of research, what next? *The Visual Computer* 21, 12, 997–1015.
- MOHR, A., AND GLEICHER, M. 2003. Building efficient, accurate character skins from examples. *ACM Transactions on Graphics* 22, 3, 562–568.
- MÜLLER, M., HEIDELBERGER, B., TESCHNER, M., AND GROSS, M. 2005. Meshless deformations based on shape matching. *ACM Transactions on Graphics* 24, 3, 471–478.
- NEDEL, L. P., AND THALMANN, D. 2000. Anatomic modeling of deformable human bodies. *The Visual Computer* 16, 6, 306–321.
- PRATSCHER, M., COLEMAN, P., LASZLO, J., AND SINGH, K. 2005. Outside-in anatomy based character rigging. In *2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, 329–338.
- PRONOST, N., DUMONT, G., BERILLON, G., AND NICOLAS, G. 2006. Morphological and stance interpolations in database for simulating bipedalism of virtual humans. *The Visual Computer* 22, 1, 4–13.
- SAND, P., McMILLAN, L., AND POPOVIĆ, J. 2003. Continuous capture of skin deformation. *ACM Transactions on Graphics* 22, 3, 578–586.
- SCHEEPERS, F., PARENT, R. E., CARLSON, W. E., AND MAY, S. F. 1997. Anatomy-based modeling of the human musculature. In *Proceedings of SIGGRAPH 97*, Computer Graphics Proceedings, Annual Conference Series, 163–172.
- SEO, H., AND MAGNENAT-THALMANN, N. 2003. An automatic modeling of human bodies from sizing parameters. In *2003 ACM Symposium on Interactive 3D Graphics*, 19–26.
- SEO, H., CORDIER, F., AND MAGNENAT-THALMANN, N. 2003. Synthesizing animatable body models with parameterized shape modifications. In *2003 ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 120–125.
- SHUM, H.-Y., IKEUCHI, K., AND REDDY, R. 1995. Principal component analysis with missing data and its application to polyhedral object modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17, 9, 854–867.
- SIFAKIS, E., NEVEROV, I., AND FEDKIW, R. 2005. Automatic determination of facial muscle activations from sparse motion capture marker data. *ACM Transactions on Graphics* 24, 3, 417–425.
- SINGH, K., AND KOKKEVIS, E. 2000. Skinning characters using surface oriented free-form deformations. In *Graphics Interface*, 35–42.
- SLOAN, P.-P. J., III, C. F. R., AND COHEN, M. F. 2001. Shape by example. In *2001 ACM Symposium on Interactive 3D Graphics*, 135–144.
- SUMNER, R. W., ZWICKER, M., GOTSMAN, C., AND POPOVIĆ, J. 2005. Mesh-based inverse kinematics. *ACM Transactions on Graphics* 24, 3, 488–495.
- SUN, W., HILTON, A., SMITH, R., AND ILLINGWORTH, J. 2001. Layered animation of captured data. *The Visual Computer* 17, 8, 457–474.
- TERAN, J., SIFAKIS, E., BLEMKER, S. S., NG-THOW-HING, V., LAU, C., AND FEDKIW, R. 2005. Creating and simulating skeletal muscle from the visible human data set. *IEEE Transactions on Visualization and Computer Graphics* 11, 3, 317–328.
- TERAN, J., SIFAKIS, E., IRVING, G., AND FEDKIW, R. 2005. Robust quasistatic finite elements and flesh simulation. In *2005 ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 181–190.
- VICON MOTION SYSTEMS, 2006. <http://www.vicon.com/>.
- VLASIC, D., BRAND, M., PFISTER, H., AND POPOVIĆ, J. 2005. Face transfer with multilinear models. *ACM Transactions on Graphics* 24, 3, 426–433.
- WALLRAVEN, C., BREIDT, M., CUNNINGHAM, D. W., AND BÜLTHOFF, H. H. 2005. Psychophysical evaluation of animated facial expressions. In *APGV 2005*, 17–24.
- WANG, X. C., AND PHILLIPS, C. 2002. Multi-weight enveloping: Least-squares approximation techniques for skin animation. In *ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 129–138.
- WANG, Y., HUANG, X., LEE, C.-S., ZHANG, S., LI, Z., SAMARAS, D., METAXAS, D., ELGAMMAL, A., AND HUANG, P. 2004. High resolution acquisition, learning and transfer of dynamic 3-d facial expressions. *Computer Graphics Forum* 23, 3, 677–686.
- WILHELM, J., AND GELDER, A. V. 1997. Anatomically based modeling. In *Proceedings of SIGGRAPH 97*, Computer Graphics Proceedings, Annual Conference Series, 173–180.
- ZALEWSKI, L., AND GONG, S. 2005. 2d statistical models of facial expressions for realistic 3d avatar animation. In *CVPR*, 217–222.
- ZHANG, L., SNAVELY, N., CURLESS, B., AND SEITZ, S. M. 2004. Spacetime faces: high resolution capture for modeling and animation. *ACM Transactions on Graphics* 23, 3, 548–558.
- ZORDAN, V. B., AND HORST, N. C. V. D. 2003. Mapping optical motion capture data to skeletal motion using a physical model. In *2003 ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 245–250.
- ZORDAN, V. B., CELLY, B., CHIU, B., AND DILORENZO, P. C. 2004. Breathe easy: model and control of simulated respiration for animation. In *2004 ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 29–37.