MoSh: Motion and Shape Capture from Sparse Markers

Matthew Loper* Naureen Mahmood† Michael J. Black‡
Max Planck Institute for Intelligent Systems, Tübingen, Germany

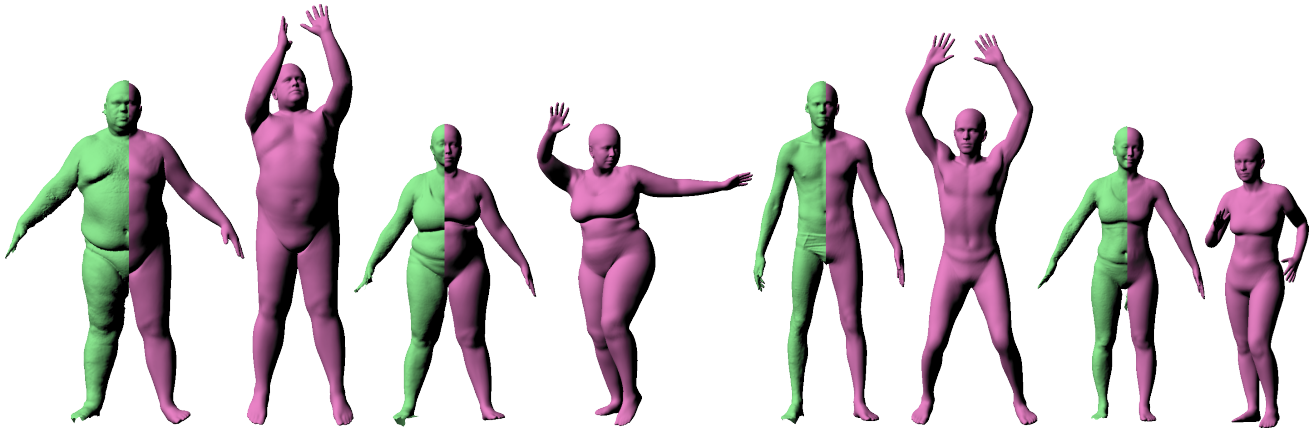

Figure 1: Shape from mocap. *MoSh computes body shape and pose from standard mocap marker sets. Bodies in purple are estimated from 67 mocap markers, while scans in green are captured with a high-resolution 3D body scanner. Split-color bodies compare the shape estimated from sparse markers with scans. MoSh needs only sparse mocap marker data to create animations (purple posed bodies) with a level of realism that is difficult to achieve with standard skeleton-based mocap methods.*

Abstract

Marker-based motion capture (mocap) is widely criticized as producing lifeless animations. We argue that important information about body surface motion is present in standard marker sets but is lost in extracting a skeleton. We demonstrate a new approach called MoSh (**Mo**tion and **Sh**ape capture), that automatically extracts this detail from mocap data. MoSh estimates body *shape* and pose together using sparse marker data by exploiting a parametric model of the human body. In contrast to previous work, MoSh solves for the *marker locations* relative to the body and estimates accurate body shape directly from the markers without the use of 3D scans; this effectively turns a mocap system into an approximate body scanner. MoSh is able to capture *soft tissue motions* directly from markers by allowing body shape to vary over time. We evaluate the effect of different marker sets on pose and shape accuracy and propose a new sparse marker set for capturing soft-tissue motion. We illustrate MoSh by recovering body shape, pose, and soft-tissue motion from archival mocap data and using this to produce animations with subtlety and realism. We also show *soft-tissue motion retargeting* to new characters and show how to magnify the 3D deformations of soft tissue to create animations with appealing exaggerations.

*e-mail:mloper@tue.mpg.de
†e-mail:nmahmood@tue.mpg.de
‡e-mail:black@tue.mpg.de

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

Keywords: Human animation, motion capture, shape capture, soft tissue motion

Links: DL PDF WEB

1 Introduction

While marker-based motion capture (mocap) is widely used to animate human characters in films and games, it is also widely criticized as producing lifeless and unnatural motions. We argue that this is the result of "indirecting" through a skeleton that acts as a proxy for the human movement. In standard mocap, visible 3D markers on the body surface are used to infer the unobserved skeleton. This skeleton is then used to animate a 3D model and what is rendered is the visible body surface. While typical protocols place markers on parts of the body that move as rigidly as possible, soft-tissue motion always affects surface marker motion. Since non-rigid motions of surface markers are treated as noise, subtle information about body motion is lost in the process of going from the non-rigid body surface to the rigid, articulated, skeleton representation. We argue that these non-rigid marker motions are not noise, but rather correspond to subtle surface motions that are important for realistic animation.

We present a new method called MoSh (for **Mo**tion and **Sh**ape capture) that replaces the skeleton with a 3D parametric body model. Given a standard marker set, MoSh simultaneously estimates the marker locations on a proxy 3D body model, estimates the body shape, and recovers the articulated body pose. By allowing body shape to vary over time, MoSh is able to capture the non-rigid motion of soft tissue. Previous work on the mocap of such motions relies on large marker sets [Park and Hodgins 2006; Park and Hodgins 2008]. In contrast, we show that significant soft tissue motion is present in small marker sets and that capturing it results in more

nuanced and lifelike animations. MoSh also recovers qualitatively and metrically accurate body shapes from small numbers of markers; Fig. 1 shows body shapes and poses recovered with 67 markers and compares the body shapes with 3D scans. While fine details are missing, MoSh enables users of standard mocap to obtain reasonable 3D body shapes from markers alone.

The basic version of MoSh has five core components. 1) MoSh uses a parametric 3D body model that realistically represents a wide range of natural body shapes, poses, and pose-dependent deformations. For this we use a learned statistical body model based on SCAPE [Anguelov et al. 2005]. 2) Marker placement on the human body varies across subjects and sessions, consequently we do not assume that the exact marker placement is known. Instead, a key contribution of MoSh is that it solves for the observed marker locations relative to the 3D body model. 3) MoSh also simultaneously solves for the 3D body shape of the person that best explains the observed 3D mocap marker data. 4) Steps 2 and 3 above require that we also simultaneously solve for 3D body pose. Components 2–4 are all embodied in a single objective function and we optimize this for a subset of the mocap sequence. 5) In a second stage, MoSh uses the computed body shape and marker locations on the body, to estimate body pose throughout a mocap session.

This basic method produces appealing animations but the assumption of a single body shape across the session does not account for the *dynamics of soft tissue*; for example, the jiggling of fat during jumping. Currently there are no practical technologies for easily capturing these soft-tissue motions. Previous methods have used large marker sets [Park and Hodgins 2006] but these are time consuming to apply, difficult to label, and suffer from occlusion. These methods also do not apply to archival data. Video-based surface capture methods offer the potential for even greater realism [de Aguiar et al. 2008; Stark and Hilton 2007] but are not yet mature and are not widely adopted. To capture soft-tissue deformation, we allow the body shape to change over time to better fit the marker motions. Our solution uses a low-dimensional shape model to make it practical and penalizes deviations from the fixed body shape estimated without soft-tissue deformation. We make an assumption that these deformations can be approximated *within the space of static human body shape variations*; that is, we model the soft-tissue deformations of an individual effectively by changing their identity. Given a sufficiently rich space of body shape variation, this works surprisingly well.

While we can estimate body shape and pose from standard marker sets and archival mocap sequences, we go further to design *additional marker sets* with greater or fewer markers. Using a principled objective function, and a training set of 3D body meshes, we evaluate the effect of different marker sets on the accuracy of body shape and pose capture. While the standard 47-marker set that is often used for motion capture (e.g. in the CMU dataset) works surprisingly well for recovering both shape and pose, we find that an expanded set, with 20 additional markers, captures more soft tissue motion.

We validate the method with nearly 800 mocap sequences. Since no body scanner or other hardware is required, MoSh can be applied to archival mocap data. To demonstrate this we reconstruct gender, shape, and motion of 39 subjects in the CMU mocap dataset using 47 markers. The resulting animations are nuanced and lifelike and the body shapes qualitatively match reference video. For quantitative evaluation, we scanned twenty subjects with widely different body shapes and performed MoSh with different numbers of markers.

MoSh can be used directly for animation or as a reference for animators. In the accompanying video we show that we can change the body shape to retarget the mocap sequence to new bodies (cf. [Anguelov et al. 2005]). This transfer works for any character with the same topology as our body model. We align several cartoon characters to our mesh and then animate them without the labor-intensive process of developing a rigged model or retargeting the skeletal motions. The animations include the transfer of soft tissue motions and we show further how these motions can be magnified to produce interesting animations with exaggerated soft-tissue dynamics.

In summary, the main contribution of MoSh is that it provides a fully automated method for "mining" lifelike body shape, pose, and soft-tissue motions from sparse marker sets. This makes MoSh appropriate for processing archival mocap. By using the same (or slightly augmented) marker sets, MoSh complements, existing marker-based mocap in that animators can extract standard skeletal models from the markers, MoSh meshes, or both.

2 Prior work

There is an extensive literature on (and commercial solutions for) estimating skeleton proxies from marker sets. Since MoSh does not use a skeleton, we do not review these methods here. Instead, we focus on several key themes in the literature that more directly relate to our work: fitting models to sparse markers, dense marker sets, and surface capture.

From Markers to Models. To get body shape from sparse markers, one needs a model of body shape to constrain the problem. There have been several previous approaches. Allen et al. [2003] learn a model of body shape variation in a fixed pose from 3D training scans. Anguelov et al. [2005] go further to learn a model that captures both body shape and non-rigid pose deformation.

Allen et al. show that one can approximately recover an unknown 3D human shape from a sparse set of 74 landmarks. They do this only for a fixed pose since their model does not represent pose variation. Importantly the landmarks are perfect and known; that is, they have the 3D points on the mesh they want to recover and do not need to estimate their location on the mesh. Unlike MoSh this does not address the problem of estimating body shape and pose from mocap markers alone.

Anguelov et al. [2005] show how to animate a SCAPE model from motion capture markers. Their method requires a 3D scan of the subject with the markers on their body. This scan is used for two purposes. First it is used to estimate the 3D shape model of the person; this shape is then held fixed. Second the scanned markers are used to establish correspondence between the scan and the mocap markers. These limitations mean that the approach cannot work on archival mocap data and that a user needs both a 3D body scanner and a mocap system.

It is important to note that Anguelov et al. did not solve the problem addressed by MoSh. They fit a SCAPE model to a 3D body *scan* (what they call shape completion) and with *known marker locations*, animate the model from mocap markers. We go beyond their work to estimate the body shape from *only the sparse mocap markers* without the use of any scan and without knowing their precise location on the body. We do this by simultaneously solving for the marker locations, the shape of the body and the pose using a single objective function and optimization method. Unlike [Anguelov et al. 2005], MoSh is fully automatic and applicable to archival data.

We also go beyond previous work to define new marker sets and evaluate the effect of these on reconstruction accuracy. This provides a guide for practitioners to choose appropriate marker sets.

Dynamics of Soft Tissue. Unlike MoSh, the above work does not address the capture of soft tissue motion. Interestingly, much of the attention paid to soft-tissue motion in the mocap community (particularly within biomechanics) actually focuses on minimizing the effects of soft tissue dynamics [Leardini et al. 2005]. Soft tissue motion means the markers move relative to the bones and this reduces the accuracy of the estimated skeletal models. For animation, we argue that such soft tissue motions are actually critical to making a character look alive.

Dense Marker Sets. To capture soft-tissue motion, previous work has used large, dense, marker sets. Park and Hodgins [2006] use 350 markers to recover skin deformation; in the process, they deform a subject-specific model to the markers and estimate missing marker locations. In later work [Park and Hodgins 2008], they use a large (400-450) marker set for $\approx 10,000$ frames of activity to create a subject-specific model; this model can then be used to recover pose for the same subject in later sessions with a sparse marker set. In these works, the authors visualize soft-tissue deformations on characters resembling the mocap actor. Here we transfer soft-tissue deformations to more stylized characters.

Hong et al. [2010] use 200 markers on the shoulder complex and a data driven approach to infer a model of shoulder articulation. While dense markers can capture rich shape and deformation information, they are not practical for many applications. Placing the markers is time consuming and a large number of markers may limit movement. With these large sets, additional challenges emerge in dealing with inevitable occlusions and marker identification.

Recent work captures skin deformations using a dense set of markers or patterns painted on the body [Bogo et al. 2014; Neumann et al. 2013a]. The work is similar to Park and Hodgins but uses computer vision methods rather than standard mocap markers.

Our work differs in that it conforms to standard mocap practice and is backwards-compatible with existing sparse marker sets. The goal of MoSh is to get more out of sparse markers.

Surface Capture. At the other extreme from sparse markers are methods that capture full 3D meshes at every time instant [de Aguiar et al. 2008; Stark and Hilton 2007]; this can be conceived of as a very dense marker set. Still other methods use a scan of the person and then deform it throughout a sequence [de Aguiar et al. 2007a; Liu et al. 2013]. Existing methods for surface capture rely on multi-camera computer vision algorithms that are computationally expensive compared with commercial marker-based systems. These methods are most applicable to capturing complex surfaces like clothing or breathing [Tsoli et al. 2014] that are difficult to parametrize. In the case of body shape, we find that, together with a parametric body model, a small marker set is already very powerful.

In a related approach, de Aguiar et al. [2007b] use an intermediate template that is animated in a traditional way from mocap markers. They then transfer the template motion to a more complex mesh. Like MoSh this method is motivated by standard practice but it still indirects through a crude proxy, rather than solving directly for shape and pose from markers.

Attribute Capture. The idea that markers contain information about body shape is not new. Livne et al. [2012] use motion capture data to extract socially meaningful attributes, such as gender, age, mental state and personality traits by applying 3D pose tracking to human motion. This work shows that a sparse marker set contains rich information about people and their bodies. MoSh takes a different approach by using the sparse marker data to extract faithful 3D body shape. Like Livne et al., we show that gender can be estimated from markers. Beyond this, we suspect that the full 3D body model can be used to extract additional attributes.

Motion Magnification. There has been recent work on magnifying small motions in video sequences [Wang et al. 2007; Wu et al. 2012; Wadhwa et al. 2013] but less work on magnifying 3D motions. In part this may be because capturing 3D surface motions is difficult. Other work exaggerates mocap skeletal motions using mocap data [Kwon and Lee 2007]. In [Neumann et al. 2013b] they develop methods for spatially localized modeling of deformations and show that these deformations can be edited and exaggerated. In [Jain et al. 2010] they edit body shape to exaggerate it but do not model or amplify non-rigid soft-tissue dynamics. While the exaggeration of facial motion has received some attention, we think ours is the first work to use only sparse marker sets to extract full-body soft-tissue motion for exaggeration.

In summary, MoSh occupies a unique position – it estimates 3D body shape and deformation using existing mocap marker sets. MoSh produces animated bodies directly from mocap markers with a realism that would be time consuming to achieve with standard rigging and skeleton-based methods.

3 Body Model

Extracting body shape from sparse markers is clearly an ill-posed problem; an infinite number of bodies could explain the same marker data. To infer the most likely body we must have a model of human shape that captures the correlations in body shape within the population. For this we use a learned body model that is similar to SCAPE [Anguelov et al. 2005]. It should be noted however that any mesh model could be used, as long as (1) it allows shape and pose variation, and (2) is differentiable with respect to its parameters.

Our body model is a function that returns a triangulated mesh with 10,777 vertices, and is parameterized by a global translation center γ, a vector of pose parameters, θ, a mean shape, μ, and a vector of shape parameters, β. Shape is defined in terms of deformations applied to the triangles of a base template mesh. The surface of the body is described as $S(\beta, \theta, \gamma)$, with the coordinates of vertex k notated $S_k(\beta, \theta, \gamma)$. The body mesh is segmented into parts and each part can undergo a rotation defined by θ. The pose parameters θ consist of 19 angle-axis vectors, whereby length indicates the amount of rotation. Like SCAPE, the function $S(\cdot)$ includes pose-dependent non-rigid deformations that are learned from bodies in a wide range of poses. Body shape is approximated by the mean shape and a linear combination of shape basis vectors; β is a vector of these linear coefficients. This shape basis is learned from deformations of training body shapes using principal component analysis (PCA). In what follows, we represent body shape using 100 principal components.

We train the body shape model from 3803 CAESAR scans of people in an upright pose (approximately 2103 women and 1700 men from the US and EU datasets) [Robinette et al. 2002]. The pose-dependent component of the model is learned from 1832 scans of 78 people (41 women and 37 men) in a wide range of poses. The scans are aligned using the technique in [Hirshberg et al. 2012]. Since the model is trained from an extensive set of scans, it is able to realistically capture a wide range shapes and poses. For details of SCAPE, the reader is referred to [Anguelov et al. 2005].

Note that we train three body shape models: separate models for men and women, plus a gender neutral model. If we know the gender of the subject, we use the appropriate model. If not, we fit the gender-neutral model, infer the gender, and then use a gender-specific model as described below.

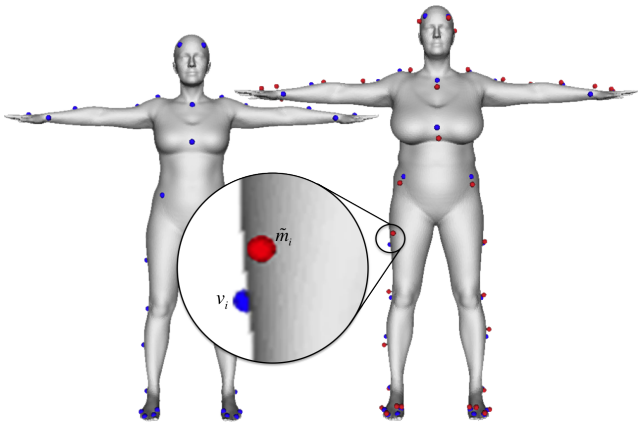

Figure 2: Optimizing shape and markers. *Left: initial guess of markers, v_i, on the template shape in the canonical pose (blue). Right: Shape and marker locations after optimization. Optimized marker locations, $\tilde{m}_i$, are shown in red. Note that they have moved (see inset).*

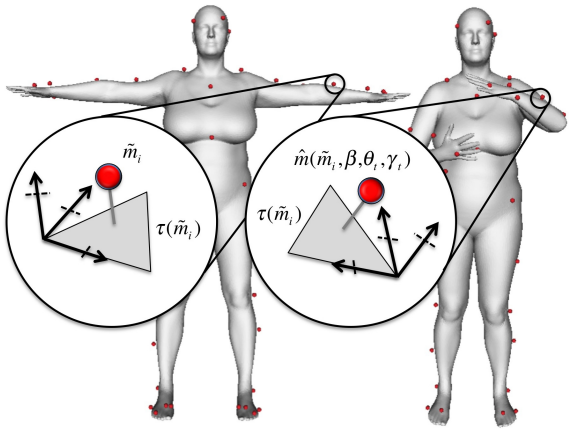

Figure 3: Marker transformations. *In the latent coordinate space (left) we project a marker, $\tilde{m}_i$ into a basis defined by the nearest vertex: specifically by its normal, an arbitrary normalized edge, and the cross product between them. This provides a pose invariant representation for the marker. When the body pose changes (right), we then compute the location of the marker, $\hat{m}(\tilde{m}_i, \beta, \theta_t, \gamma_t)$, in the observed frame.*

4 Markers on the Body and in the World

Mocap markers extend from the human body to varying degrees and are placed on the body manually. Precise placement can be difficult, particularly on heavy subjects where fat makes it difficult to palpate boney locations. The result is that we cannot expect to know the exact marker locations in advance. The first step of MoSh solves for the marker locations, relative to a template body mesh, for a given mocap sequence (or collection of sequences for one subject).

4.1 Defining a Marker Set

We assume that we know the number of markers and their approximate location relative to a reference template mesh. The only manual part of MoSh occurs if a user wants to use a new marker set. In this case they need to identify a template vertex for each marker. Notationally, we say a user creates a mapping $h(i)$ from marker indices, i, to vertex indices on the template. Each marker requires the user-specification of an expected distance d_i from the marker center to the skin surface. Both the location and the distance can be approximate since we optimize these for each subject.

To parameterize marker locations with respect to the body, we introduce a latent coordinate system that contains markers and our body model in a neutral pose, γ_0, θ_0, as in Fig. 2 (left). The purpose of this latent coordinate system is to model the relationship between the body surface and the markers in a pose-independent, translation-independent, fashion. This relationship is then transferred to meshes in observed mocap frames.

We then denote the default position of the markers, v_i, as

$$v_i(\beta) \equiv S_{h(i)}(\beta, \theta_0, \gamma_0) + d_i N_{h(i)}(\beta, \theta_0, \gamma_0), \qquad (1)$$

where $N_k(\beta, \theta, \gamma)$ indicates the vertex normal for index k given body model parameters. Thus $v_i(\beta)$ is the position of the model vertex, offset by a user-prescribed distance, d_i, from the surface, in the latent coordinate system, corresponding to marker i. These are illustrated as blue balls in Fig. 2.

Defining the marker set needs to be done once and then it is used for any subject captured with that marker set. For example, we did this once for the 47-marker Vicon set and used this for all mocap sequences in the CMU database.

4.2 Parameterizing Markers

The default markers, v_i, are approximate and below we optimize to solve for the body shape, β, and the actual location of the latent markers, $\tilde{m}_i$, for a given subject and mocap sequence. Let $\tilde{M}$ denote the collection of latent markers. Notationally, we use i to indicate marker number and t to indicate the mocap sequence frame number. Observed markers are denoted $m_{i,t}$ individually and M_t together. From a collection of M_t we estimate the latent markers $\tilde{M}$, shown as red balls in Fig. 2.

To that end, we define a function $\hat{m}(\tilde{m}_i, \beta, \theta_t, \gamma_t)$ that maps latent markers to the world given a particular shape, pose, and location of the body. We call these "simulated markers". Intuitively, we want to solve for the shape, pose, body location, and latent marker locations $\tilde{m}_i$ such that, when projected into the mocap sequence, the simulated markers match the observed markers M_t.

This requires a mapping from local surface geometry to a 3D marker position that can be transferred from the latent coordinate system to the observed markers resulting from different poses. We represent a marker position in an orthonormal basis defined by its nearest triangle in the latent coordinate system. We define that basis by three vectors: the triangle normal, one of the triangle's normalized edges, and the cross product between those two. This is geometrically depicted in Fig. 3 (left).

We denote the rigid transformation matrix that projects $\tilde{m}$ into the basis for closest triangle $\tau(\tilde{m})$ in the mesh, as $B_{\tau(\tilde{m})}(\beta, \theta, \gamma)$. We then define a simulated marker position $\hat{m}(\cdot)$ as

$$\hat{m}^*(\tilde{m}, \beta, \theta_t, \gamma_t) = B_{\tau(\tilde{m})}(\beta, \theta_t, \gamma_t) B^{-1}_{\tau(\tilde{m})}(\beta, \theta_0, \gamma_0) \tilde{m}^* \qquad (2)$$

where $\tilde{m}^* = [\tilde{m}^T, 1]^T$ and $\hat{m}^*(\cdot) = [\hat{m}(\cdot)^T, 1]^T$ denote the marker locations in homogeneous coordinates. Equation 2 can be seen as having two steps. First, the matrix $B^{-1}_{\tau(\tilde{m})}(\beta, \theta_0, \gamma_0)$ transforms $\tilde{m}^*$ from a 3D *latent*-space position into a coordinate vector in the space of its local basis. In the second step, $B_{\tau(\tilde{m})}(\beta, \theta_t, \gamma_t)$ maps this coordinate vector into a 3D *observed*-space position, $\hat{m}^*(\cdot)$, defined by the specific position and pose, γ_t, θ_t. This is illustrated in Fig. 3 (right).

With the marker parameterization defined, we next define the objective functions we use to estimate marker positions, shape, pose, and nonrigid motion.

5 Objectives

Let sequences of body pose $\theta_{1..n}$, and position $\gamma_{1..n}$, with n time instants be denoted as Θ and Γ respectively. We wish to estimate the latent markers $\tilde{M}$, poses Θ, body locations Γ, and body shape β, such that the simulated markers $\hat{m}(\cdot)$, match the observed markers $m_{i,t}$. To do so we define an objective function with several terms.

The data term, E_D, is the sum of squared distances between simulated and observed landmarks:

$$E_D(\tilde{M}, \beta, \Theta, \Gamma) = \sum_{i,t} ||\hat{m}(\tilde{m}_i, \beta, \theta_t, \gamma_t) - m_{i,t}||^2. \quad (3)$$

Note that distances are measured in *cm*.

A surface distance energy term, E_S, encourages markers to keep a prescribed distance from the body surface in the latent coordinate system. Let $r(x, S)$ denote the signed distance of a 3D location x to the surface S. Then

$$E_S(\beta, \tilde{M}) = \sum_i ||r(\tilde{m}_i, S(\beta, \theta_0, \gamma_0)) - d_i||^2. \quad (4)$$

Since the marker locations are roughly known to begin with, we penalize estimated latent markers if they deviate from this. The energy term E_I regularizes the adjusted marker towards its original position

$$E_I(\beta, \tilde{M}) = \sum_i ||\tilde{m}_i - v_i(\beta)||^2. \quad (5)$$

We also define pose and shape priors to regularize the estimation of body shape and pose. These are modeled as Gaussian, with their statistics $\mu_\beta, \mu_\theta, \Sigma_\beta, \Sigma_\theta$ computed from the pose and shape training data used to train our body model. We regularize β and θ_t by penalizing the squared Mahalanobis distance from the mean shape and pose:

$$E_\beta(\beta) = (\beta - \mu_\beta)^T \Sigma_\beta^{-1} (\beta - \mu_\beta) \quad (6)$$

$$E_\theta(\Theta) = \sum_t (\theta_t - \mu_\theta)^T \Sigma_\theta^{-1} (\theta_t - \mu_\theta). \quad (7)$$

We also add a velocity constancy term E_u that helps to smooth marker noise by a small amount:

$$E_u(\Theta) = \sum_{t=2}^{n} ||\theta_t - 2\theta_{t-1} + \theta_{t-2}||^2. \quad (8)$$

Our objective in total is the sum of these terms, each weighted by its own weight, λ:

$$E(\tilde{M}, \beta, \Theta, \Gamma) = \sum_{\omega \in \{D,S,\theta,\beta,I,u\}} \lambda_\omega E_\omega(\cdot). \quad (9)$$

6 Optimization

The objective function above is quite general and it enables us to solve a variety of problems depending on what we minimize and what we hold constant. In all cases, optimization uses Powell's dogleg method [Nocedal and Wright 2006], with Gauss-Newton Hessian approximation. The gradients of the objective function are computed with algorithmic differentiation [Griewank and Walther 2008], which applies the chain rule to the objective function; for this we use an auto-differentiation package called Chumpy [Loper 2014]. Only the differentiation of the body model $S_k(\beta, \theta, \gamma)$ and the signed mesh distance $r(x, S)$ were done by hand, to improve runtime performance.

There are two main optimization processes. The first estimates time-independent parameters (body shape β and marker placements $\tilde{M}$), while the second estimates time-dependent parameters $\Theta = \{\theta_1 \dots \theta_n\}, \Gamma = \{\gamma_1 \dots \gamma_n\}$.

Body Shape and Latent Markers. For a given mocap sequence (or set of sequences for the same subject), optimization always starts by estimating the latent marker locations $\tilde{M}$, body shape β, poses Θ, and body positions Γ for a subset of the frames. The latent marker locations and the body shape are assumed to be time independent and can be estimated once for the entire sequence (or set of sequences).

Notably, the transformation from latent to observed coordinate systems is continuously re-estimated during the optimization of marker placement. The assignment of nearest neighbors, the local basis itself, and the coefficients relating a marker to that basis undergo continual adjustment to allow refinement of the relationship between markers and the body surface.

The λ values in Eq. 9 are: $\lambda_D = 0.75$, $\lambda_S = 100.0$, $\lambda_I = 0.25$, $\lambda_\beta = 1.0$, $\lambda_\theta = 0.25$, $\lambda_u = 0$.

The λ values were initialized to normalize each term by an estimate of its expected value at the end of the optimization; in particular, the distance-based λ values (λ_D, λ_S, λ_I) have interpretations as inverse variances with units of $\frac{1}{cm^2}$. These λ values were then empirically refined.

The velocity term is not used in this stage ($\lambda_u = 0$) because we are optimizing over random disconnected frames.

To help avoid local optima, the optimization is run in six stages, starting with strong regularization and then gradually decreasing this. Specifically, the regularization weights $\{\lambda_\theta, \lambda_\beta, \lambda_I\}$ are lowered from being multiplied by 40, then by 20, 10, 4, 2, and finally 1. Note that these regularization terms are linear and quadratic in contrast to the data term, which is non-linear. Similar to graduated non-convexity schemes, by increasing the regularization weights we make the objective function more convex, potentially helping the optimization avoid local optima during early stages of the process. In practice we found this to work well.

Computational cost increases with the number of frames used to estimate the parameters since each frame requires its own pose θ_t. For efficiency we perform this optimization using a randomly selected subset of mocap time instants. We ran experiments with different numbers of randomly chosen frames and saw little improvement with more than 12 frames. Consequently we use 12 random frames for all experiments here.

Pose. Motion capture now becomes the problem of estimating the pose of the body, θ_t, and body position, γ_t, at each time instant given the known body shape and latent markers. We initialize the optimization at frame t with the solution at $t-1$ if it is available and then a short optimization is run for each time step.

For pose estimation, the λ values are now: $\lambda_D = 0.75$, $\lambda_S = 0$, $\lambda_I = 0$, $\lambda_\beta = 0$, $\lambda_\theta = 1.0$, $\lambda_u = 6.25$. Note that we now employ the velocity smoothness term, λ_u. A weight of zero means

that this term is not used and the corresponding parameters are not optimized. Specifically, we do not optimize the marker locations or body shape. We do however use a pose prior, $\lambda_\theta = 1.0$, to penalize unlikely poses. Here we do not used the staged regularization because the optimization begins close to the minimum and converges quickly.

Pose and Soft Tissue Motion. In the optimization above we assume body shape and latent marker locations do not change. To capture soft tissue motions we now allow the body shape to vary across the sequence while keeping the marker transformation fixed. We still denote β as a shape estimated in the first stage, but now denote the time-varying deviations in shape from β as $B = \{\beta_1 \dots \beta_n\}$, such that a person's shape at time t is now $\beta + \beta_t$.

To regularize the β_t, we add one additional energy term to Eq. 9:

$$E_\Delta(B) = \sum_t ||\beta_t||^2 \tag{10}$$

and set λ_Δ to 0.25, adding $\lambda_\Delta E_\Delta(\cdot)$ in Eq. 9. This term allows body shape to change over time while regularizing it to not deviate too much from the person's "intrinsic shape", β.

While our body shape training set does not contain examples of soft tissue dynamics, it does capture many shape variations across the population. These are exploited to capture soft tissue deformations during motion. Someone inhaling, for example, might look like a different person with a higher chest or a bigger stomach. When someone jumps up and down, their chest changes in ways that resemble the chests of other people. It is interesting, and perhaps surprising, that the shape variations between people can be used to approximate the shape variation of an individual due to dynamics. Presumably there are soft-tissue deformations that cannot be explained this way but, given sufficiently many training body shapes, and sufficiently many principal components, we posit that a wide range of such deformations are representable. We suspect, however, that training shapes specific to soft-tissue deformations could be used to learn a more concise model. Note further that we do not *model* dynamics of soft tissue, we only approximate what is present in the mocap marker data.

Since standard marker sets are designed for estimating a skeleton, the markers are mostly placed on rigid body structures to minimize soft tissue motion. This is another reason why existing mocap methods lack nuance. Consequently *to capture soft tissue dynamics, we want just the opposite;* we must have markers on the soft tissue. We consider this below.

Run Time. Shape and marker estimation requires about 7 minutes. Pose estimation without soft tissue estimation takes about 1 second per frame; pose estimation with soft tissue estimation requires about 2 seconds per frame.

7 Marker Selection

Body shape estimation from motion capture depends on the number and placement of markers; here we propose a method for constructing a new marker set to improve body surface reconstruction. To be practical a marker set must be simple, make sense to the technician applying it, be repeatable across subjects, and take into account self occlusion, self contact, and the impact on subject movement. Consequently we start with a standard marker set and propose additional symmetrical marker locations for a total of 114 candidate markers (Fig. 4).

We then evaluate these putative markers to determine how important the different markers are for shape recovery. For this we use

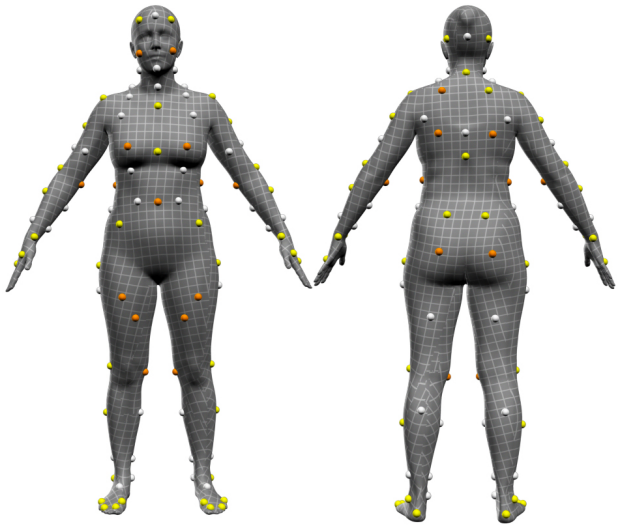

Figure 4: Marker sets. *The union of all markers illustrates the 114 possible markers we considered. Yellow markers correspond to a standard 47-marker Vicon set. The 20 orange markers were found to improve shape estimation the most. The union of yellow and orange markers corresponds to our 67-marker set used for capturing shape and soft-tissue motion. White markers were deemed redundant and were not used.*

a set of 165 meshes of 5 females of different shapes in a variety of poses selected from the FAUST dataset [Bogo et al. 2014]. A template mesh is aligned to each of the 3D scans resulting in a set of registered meshes, R^z, $z = 1 \dots 165$, in which all vertices are in correspondence across the 165 instances. We associate our 114 markers with vertices of the template and then estimate body shape from different subsets of the markers. We evaluate the accuracy of the result in terms of the Euclidean distance between the vertices of the estimated and true mesh. Specifically we compute the root mean squared error (RMSE) over all the vertices (including the subset used for fitting) for all meshes.

More formally, given a maximum number of markers, c, we seek a subset, T, of the mesh vertices, A, that enables the most accurate estimation of body shape. This subset T is the one that minimizes a cost $E_M(T)$; that is

$$T^* = \underset{T \subseteq A, |T|=c}{\arg\min} \; E_M(T). \tag{11}$$

Notationally, we will now abbreviate body model parameters $\{\beta, \theta, \gamma\}$ as P. We will also denote vertex k of registered mesh z as R^z_k. The best parameters $P^*(\{R^z_j | j \in T\})$, given access only to subset T of the vertices for registered mesh z, are defined as

$$P^*(\{R^z_j | j \in T\}) = \arg\min_P \sum_{i \in T} ||S_i(P) - R^z_i||^2. \tag{12}$$

The cost of choosing subset T takes into account the distance between all vertices $i \in A$ across all the registered meshes $z \in Z = \{1 \dots 165\}$

$$E_M(T) = \sum_{i \in A, z \in Z} ||S_i(P^*(\{R^z_j | j \in T\})) - R^z_i||^2. \tag{13}$$

Note that the RMSE is $(E_M(T)/(|A||Z|))^{1/2}$.

Evaluating all possible subsets of 114 markers is infeasible so we take a greedy approach. If we currently have N markers, we remove one, evaluate the cost for the $N-1$ possible sets, and select

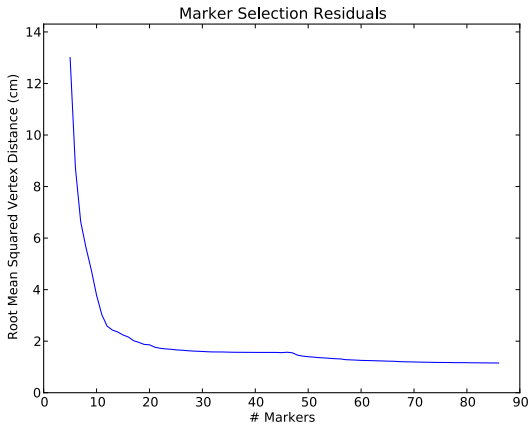

Figure 5: Marker selection residuals. *The plot shows the mesh shape reconstruction error as a function of marker count.*

the deleted marker that produces the lowest error. We remove this marker and repeat.

Figure 4 shows all 114 putative markers. The standard 47-marker set is in yellow. White and orange markers correspond to the set of additional markers that we considered. Using the greedy method, we found that the white markers were not as useful for estimating shape as the orange ones. Figure 5 shows a plot of the RMSE for different numbers of markers. Note that here we start with the 47-marker set and subtract markers from it and add markers to it. Surprisingly one can remove markers from the standard set and still obtain reasonable shape estimates down to about 25 markers. We decided to keep the original set and add the 20 additional (orange) markers. The addition of markers to the 47 results in a noticeable decrease in RMSE. Note that we could obtain similar error to our set of 67 with fewer markers by dropping some of the original 47. To enable comparison with CMU results, however, we decided to preserve the 47 and add to this set.

8 Results

8.1 Quantitative Shape Analysis

We evaluate the first stage of optimization, which computes the body shape and marker locations. To compare estimated body shapes to real ones, we scanned 20 subjects using a high-resolution 3D body scanner (3dMD LLC, Atlanta, GA). Before scanning, all subjects gave informed written consent. Additionally, 10 of the subjects were professional models who signed modeling contracts that allow us to release their full scan data.

We also used a Vicon mocap system (Vicon Motion Systems Ltd, Oxford, UK) to capture subjects with 89 markers. The 89 markers were selected using the marker optimization analysis from the full set of 114 evaluated in Sec. 7. We use at most 67 markers for shape and pose estimation; unused markers prove valuable to evaluate held-out marker error. In all cases we used the optimization with soft-tissue deformation. We processed, and evaluate error using, a total of 73 mocap sequences.

Our goal is to estimate a body shape that minimizes 3D body shape reconstruction error. We measure this error in two different ways: as held-out marker error and as mesh registration error. Held-out marker error reveals how well we can predict marker locations that were not used by the optimization: for example, if we use 47 of the markers to estimate the body shape then we use the remaining markers to estimate held-out error. As shown in Fig. 6 (right), the mean distance for held-out markers drops to approximately 3.4cm when we use 67 markers. Note that these errors include deviations in placing markers on a subject, which can easily exceed a centimeter. Specifically, when we estimate shape from a subset of markers, we do not optimize the placement of the held-out markers. So this error combines human placement error with errors in soft-tissue motion of the held-out markers that are not predicted by the subset used for fitting.

After about 25 markers the improvement is very gradual. This is interesting because it suggests that small marker sets can give good estimates of body shape. Note that this evaluation uses all 73 mocap sequences and hence evaluates how well MoSh explains marker motions due to changes in both shape and pose.

Example 3D scans of several subjects are shown in Fig. 7 (row 1). For each subject we align a template mesh to the scan and this template mesh has the same topology as the MoSh body model (Fig. 7 row two); this produces a registered mesh that we use for evaluation. Note that the registered meshes faithfully represent the scans and conform to the mesh topology of our model but do not have holes. Registration error is a measure of how well we can explain a subject's registered mesh in terms of average vertex-to-vertex mesh distance. Recovered body shapes using 67 markers are shown in Fig. 7 row three. Here we pose the MoSh result in the same pose as the scan. Given that MoSh results in a shape vector β, we adjust $\{\theta, \gamma\}$ for a body model to minimize model-to-registration distance. The heat map in the bottom row of Fig. 7 shows the distance from the MoSh shape to the registered mesh, illustrating how well MoSh approximates the shape from 67 markers.

This registration error is shown in Fig. 6 (left). Registration error behaves much like held-out marker error, except it is uniformly smaller. Unlike the held-out experiment, here we only need to explain shape and not both pose and shape. Shape estimates are obtained from 12 mocap frames and are well constrained.

While large marker sets like those used in [Park and Hodgins 2006] certainly contain more information, we see in Fig. 6 (left) diminishing returns with larger marker sets. The ideal number of markers is likely related to the resolution of the mesh.

To give some insight into what these numbers mean, Fig. 8 shows body shape for one subject reconstructed using different numbers of markers. Here we selected markers based on our greedy evaluation strategy. What is surprising is that with only 10 markers, we get a shape that roughly captures the person's size. Note that the registration error decreases as we add more markers; the numerical results show the registration error in m.

For the 10 models, scans, aligned meshes, mocap sequences, and MoSh fits are provided for research purposes here:

`http://ps.is.tuebingen.mpg.de/project/MoSh`

This data allows others to estimate shape from the same sequences and compare with both the ground truth shape and our results.

8.2 Archival Mocap (CMU)

While we do not have ground truth shape for the CMU dataset, we can evaluate results qualitatively. A visual inspection of shape recovery from CMU can be seen in Fig. 9, where video frames are shown above the bodies and poses estimated from 47 standard markers. To be clear, MoSh does not use this video frame; we show it here only for a visual evaluation of rough shape. Since the CMU

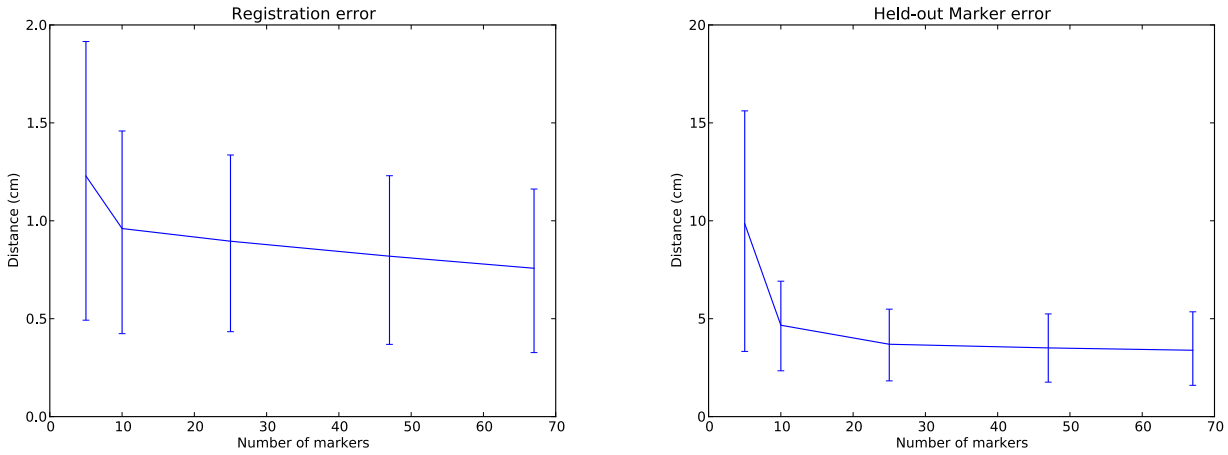

Figure 6: Effects of marker number on reconstruction error. *The mean and standard deviations of distance residuals indicate how the marker number affects reconstruction. Left: Shape reconstruction error. This is computed as the mean absolute distance between the true body shape (as represented by the alignment of the template to a scan) and the body shape estimated by MoSh reposed to match the registered mesh. Right: Held-out marker error across all sequences. This measures errors in both shape and pose but is inflated by marker placement error and marker movement. In both plots,* 68.2% $(\pm 1\sigma)$ *of the residuals are contained between the error bars.*

dataset has no anthropometric data, a quantitative evaluation is not possible.

8.3 Gender Estimation

For the above CMU results we used sequences for which the gender of the subject could be determined using accompanying video footage. Next we ask whether we can estimate gender from the markers automatically (cf. [Livne et al. 2012]). We use a linear support vector machine to predict gender from body model parameters. First we fit a gender-neutral body model to all subjects in the CAESAR dataset to obtain linear shape coefficients. We then train the SVM to predict known gender given the shape parameters. We then evaluate gender classification on body shape parameters estimated by MoSh from the CMU dataset with the gender-neutral body model. For the 39 subjects with known gender we correctly predicted it 89.7% of the time; this is comparable to [Livne et al. 2012], which is not surprising since both methods rely on essentially the same kind of marker data.

8.4 Pose Estimation Results

Given our estimate of intrinsic shape, β, and the marker locations, $\tilde{M}$, we now optimize the pose across a mocap sequence. We compute the pose for 39 subjects across 722 different mocap sequences in the CMU dataset. Figure 10 shows some representative frames from some representative sequences in the CMU dataset. Even with 47 markers we can capture some soft tissue deformation and the results shown here allow body shape deformation over time. The visual nuance of pose reconstruction is difficult to illustrate in a static image but is apparent in the **accompanying video**. Note that this is fully automatic.

The best way to evaluate accuracy of pose and shape together is in terms of held out marker error. For this we used 20 subjects and 73 mocap sequences acquired with our extended marker set. We use 67 markers for estimation and 22 to compute held-out error. This error is $3.4cm$ and corresponds to the rightmost point on the right plot in Fig. 6 (right).

With a small marker set, noise in any one marker can have an impact. In the shape estimation stage, the shape and marker placement are estimated from many poses, so variation in any individual marker should not unduly harm shape or marker placement estimation. During pose estimation, velocity constancy helps reduce the effect of single marker noise. Future work should address methods to automatically detect and downweight missing markers or markers that have moved.

9 Soft Tissue Deformation Results

Our body model was learned to represent both shape and pose-dependent deformations from registered meshes of static subjects. Many other subtle body shape deformations were not explicitly learned by our model, including static muscle contraction, breathing, gravity, external forces, and dynamics. What we show is that the space of body shapes learned from different people captures variations in shape that can approximate soft tissue motions. Note that we do not *model* the dynamics of soft tissue. We only fit the effects of such motions that are apparent in the marker data.

Figure 11 shows examples from several sequences. We show the estimated body shape with a single body shape, β, per subject (left image in each pair) and the results allowing deviations, β_t, from this shape (right image in each pair). Note the markers on the chest and belly. Red are the simulated markers predicted by our model and green are the observed markers. With changing body shape, we more accurately fit the markers undergoing soft-tissue deformation. This is not surprising, but what is important is that the shape remains "natural" and continues to look like the person.

Numerically we see the mean observed marker error go down from 0.79cm to 0.62cm with dynamics. Again this is not surprising since we are allowing the shape to deform to fit these markers. We also tested held out marker error; these are markers that were not used to estimate shape. Here too we see the mean error go from 3.41cm to 3.39cm. This is not a significant improvement, but rather a validation that fitting the soft-tissue motion does not hurt held-out marker error. This confirms our subjective impression that the body shape does not deform unnaturally and the non-rigid motions, away from

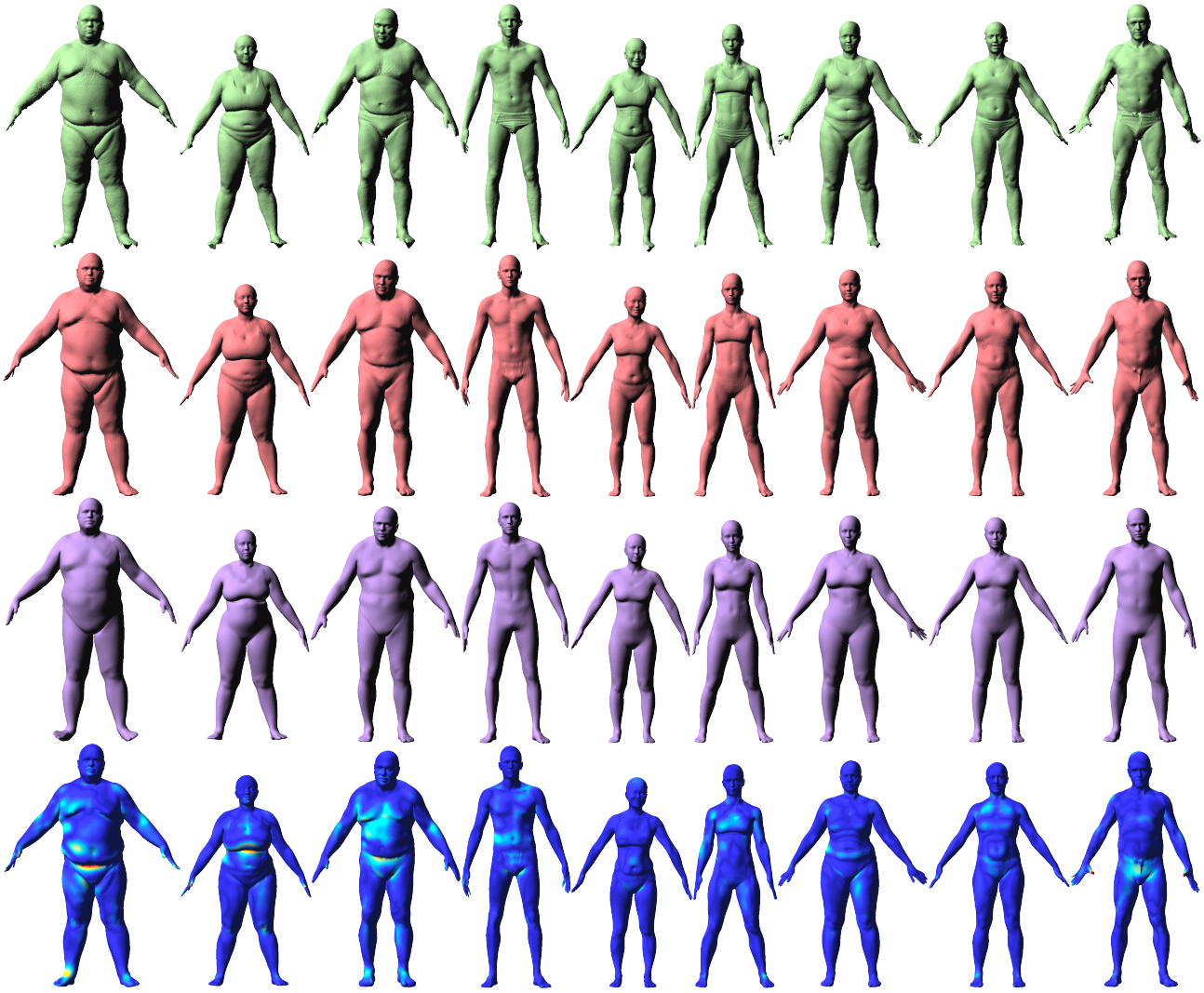

Figure 7: Shape reconstruction. *First row: raw 3D scans from a high-resolution scanner. Second row: registered meshes obtained by precisely aligning a template mesh, with the same topology as our model, to the scans. These registered meshes faithfully capture the body shape and are used for our quantitative analysis. Third row: our model with shape, β, estimated from only 67 markers. Here we estimate the pose, θ, of our model to match the registered meshes to facilitate comparison. Bottom row: Distance between second and third rows. The heat map shows Euclidean distance from the registered mesh to the nearest point on the surface of the body estimated by MoSh; blue means zero and red means $\geq$ 4 cm.*

the tracked markers, reflect realistic body deformations. While, of course, we cannot capture fine ripples with a sparse set of markers, it is surprising how much realistic deformation MoSh can estimate.

See the **accompanying video** for better visualizations and more results. In the video one sees the observed markers "swimming" around relative to the estimated shape when we do not model dynamics. There we also compare 47 markers with our 67-marker set and find that the extra markers placed on the soft tissue are important.

9.1 Exaggerated Soft-Tissue Deformation

Our soft tissue deformations correspond to directions in the space of human body shapes. We can vary the amount of deformation along these directions to either attenuate or amplify the effect. Specifically we magnify the 3D motion by multiplying β_t by a user-specified constant to exaggerate the soft tissue deformations.

This is difficult to show in print but the **video** shows examples of the same sequence with different levels of exaggeration. We found that we could magnify the deformations by a factor of 1.5 or 2 while retaining something like natural motion. Pushing the exaggeration by a factor of 4 sometimes produce interesting effects and, other times, unnatural body shapes.

This tool could be useful to animators to produce reference material since it highlights how soft tissue deforms. It could also be used to create new effects that exaggerate human actions but in a way that is based on physically realistic deformations.

9.2 Soft-Tissue Retargeting

An important use of skeletal mocap data is the retargeting of motion to a new character; the same can be done with MoSh. Consider the stylized characters in Fig. 12 that were downloaded from the Internet. For each character, we deform our template towards the

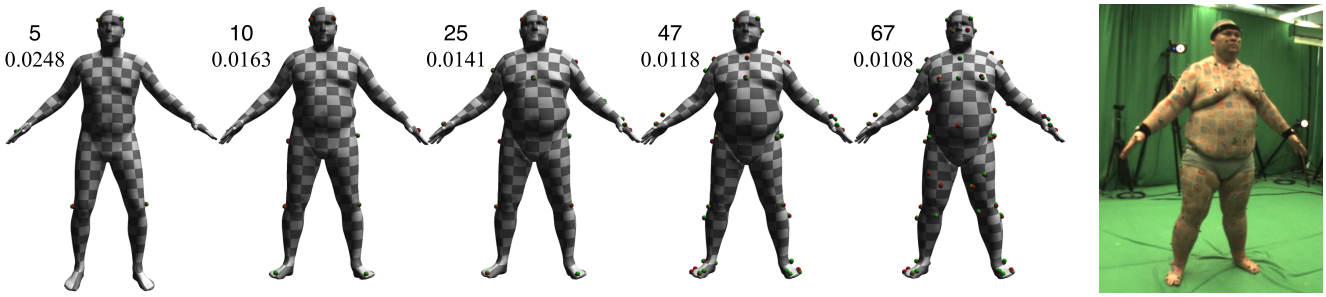

Figure 8: Shape from markers. *We show the effect of the number of markers (5, 10, 25, 47, 67) on the registration error (in m) of the estimated shape. Far right: reference image of the subject.*

Figure 9: CMU bodies. *Extracted shapes (bottom) and reference images (top) for several CMU subjects. Shape and pose is computed with MoSh using 47 Vicon markers only.*

character using regularized registration, initialized by hand-clicked correspondences. To model shape deformations from this character mesh, we simply recenter our PCA model of body shape by replacing our original mean shape, μ, with the character's template deformations. The soft tissue deformation coefficients, β_t, are then simply applied to this new mean shape. We also directly apply the estimated translation, γ_t, and MoSh part rotations, θ_t, to the parts of the new character along with the learned non-rigid pose-dependent shape deformations. This produces plausible animations. Note that, to get realistic soft-tissue transfer, we use human actors with body shapes that resemble the stylized character; see Fig. 12. Of course, these deformations can also be exaggerated.

10 Conclusion and Discussion

MoSh addresses a key criticism of existing motion capture methods. By estimating a changing body shape over time from sparse markers, MoSh captures detailed non-rigid motions of the body that produce lifelike animations. MoSh is completely compatible with existing industry-standard mocap systems. It can be used alone or in conjunction with traditional skeletal mocap since no information is lost and MoSh can use exactly the same markers as current systems. Our hope is that MoSh breathes new life into old mocap datasets and provides an easily adopted tool that extends the value of existing investments in marker-based mocap.

There are several current limitations that present interesting directions for future work. For example, we need to roughly know the marker set and we also assume the markers are in correspondence. We can correct for some mislabeled markers but we still assume a largely labeled dataset. Establishing correspondence and cleaning markers sets is a time consuming part of current mocap practices. It would be interesting to leverage the body model to try to solve these problems automatically. For example, we could also use our simulated markers to detect when a marker is missing or has moved. If a marker moves between sessions we could then update its location on the fly. We could also estimate the noise in each marker independently and take this into account during pose and shape estimation. The estimated body pose could also be used to create a virtual marker sequence that could replace the original. This would provide a principled way of fixing occlusions. Simulating a different set might be useful for methods that extract skeletal data from markers.

The quality of MoSh output is very dependent on the quality of the body model that is used. If our model cannot represent a pose realistically, then the output of MoSh will have artifacts. We observed this for a few poses, for example, both arms pointed forward, elbows straight and palms together. This suggests our pose training

Figure 10: CMU mocap. *Example meshes extracted from the CMU mocap dataset and representative frames from the animation. All shapes and poses are estimated automatically using only 47 markers. See* **accompanying video** *to see these and other results for CMU.*

set should be augmented with new poses.

An interesting direction for future work would be to use other types of body models. For example, it should be possible to replace our model with one that uses linear blend skinning and corrective blend shapes.

Our method for evaluating new marker sets could be used to construct sets to capture specific types of non-rigid deformations such as breathing. If we had 3D mesh *sequences* we could extend our analysis to select marker sets directly relevant for capturing soft-tissue motion. We did not evaluate which poses are most effective for estimating body shape; we simply chose 12 at random. Jointly optimizing the marker set and the poses could make a mocap system a more effective "body scanner;" the body scanning protocol would involve attaching the markers and having the subject assume the prescribed poses.

Our soft-tissue motions are approximations based on sparse markers but result in dense deformations. Since it is easy to acquire the data, it would be interesting to use these to train a more physical model of how soft tissue moves. That is, possibly we could leverage MoSh to learn a more sophisticated body shape model with dynamics. This could allow generalization of soft-tissue motions to new body shapes and movements.

We plan to extend our body model and MoSh methods to include the motion of feet, hands and faces. We think this is relatively straightforward but likely requires a more sophisticated pose prior model than the Gaussian one used here. It may be possible to extend these ideas further for capturing clothing or to couple our marker-based analysis with video or range data. Finally, we are also working on speeding up processing using a multi-resolution model to enable the use of MoSh in virtual production.

Acknowledgements

The CMU mocap data used in this project was obtained from mocap.cs.cmu.edu. That database was created with funding from NSF EIA-0196217. We thank B. Corner for useful feedback, E. Holderness and M. Martin for help with motion capture and scanning, and J. Romero for advice and technical guidance.

References

ALLEN, B., CURLESS, B., AND POPOVIĆ, Z. 2003. The space of human body shapes: Reconstruction and parameterization from range scans. *ACM Trans. Graph. (Proc. SIGGRAPH) 22*, 3, 587–594.

ANGUELOV, D., SRINIVASAN, P., KOLLER, D., THRUN, S., RODGERS, J., AND DAVIS, J. 2005. SCAPE: Shape Completion and Animation of PEople. *ACM Trans. Graph. (Proc. SIGGRAPH 24*, 3, 408–416.

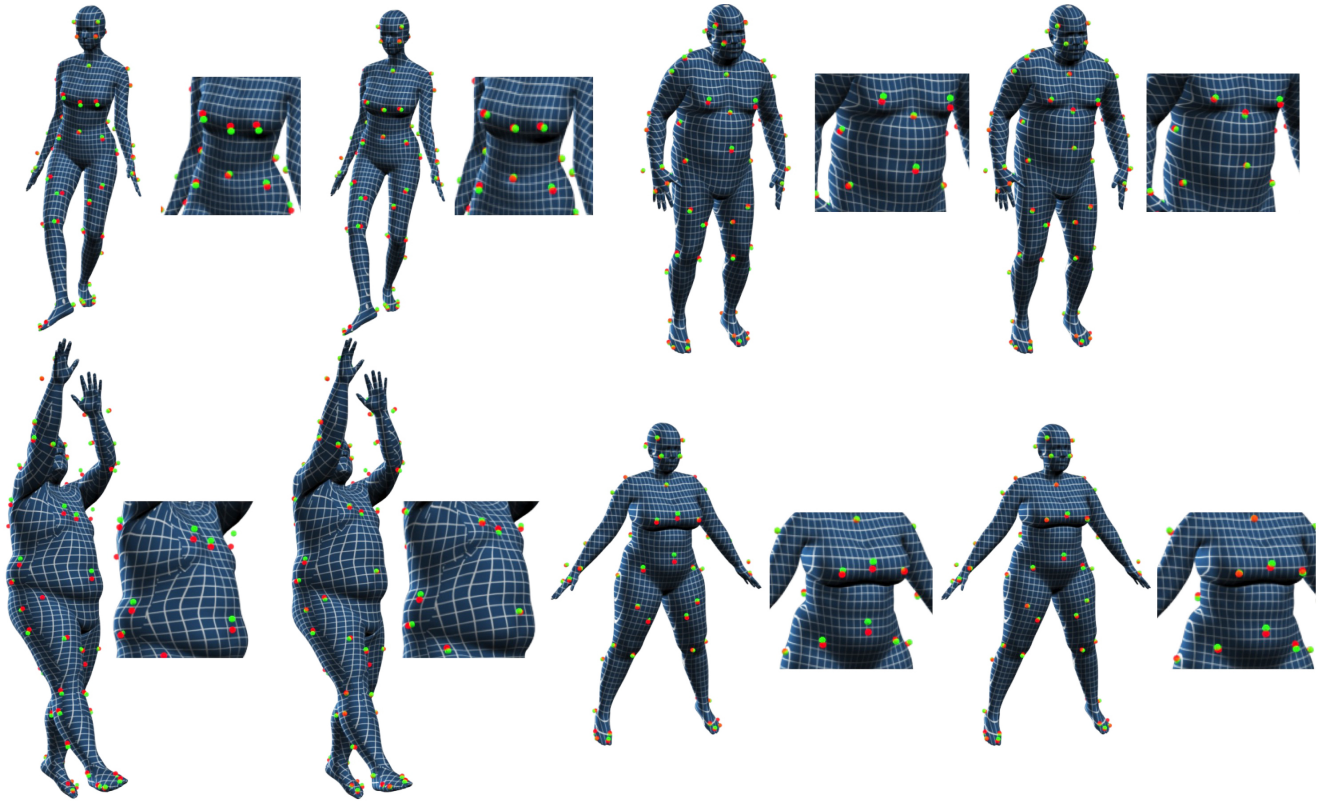

Figure 11: Motion of soft tissue. *Some representative samples are shown. In each pair, the left image is without modeling dynamics (body shape fixed) and the right with with dynamics (body shape varying). Each image shows the full body and a detail region. Green balls correspond to the mocap markers. Red balls correspond to the simulated marker locations. Allowing body shape to change over time better captures soft tissue deformations. Note that, with dynamics, the predicted markers much more closely match the observed markers.*

BOGO, F., ROMERO, J., LOPER, M., AND BLACK, M. J. 2014. FAUST: Dataset and evaluation for 3D mesh registration. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

DE AGUIAR, E., THEOBALT, C., STOLL, C., AND SEIDEL, H.-P. 2007. Marker-less deformable mesh tracking for human shape and motion capture. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 1–8.

DE AGUIAR, E., ZAYER, R., THEOBALT, C., SEIDEL, H. P., AND MAGNOR, M. 2007. A simple framework for natural animation of digitized models. In *Computer Graphics and Image Processing, 2007. SIBGRAPI 2007. XX Brazilian Symposium on*, 3–10.

DE AGUIAR, E., STOLL, C., THEOBALT, C., AHMED, N., SEIDEL, H.-P., AND THRUN, S. 2008. Performance capture from sparse multi-view video. *ACM Trans. Graph. (Proc. SIGGRAPH) 27*, 3 (Aug.), 98:1–98:10.

GRIEWANK, A., AND WALTHER, A. 2008. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, second ed. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

HIRSHBERG, D. A., LOPER, M., RACHLIN, E., AND BLACK, M. J. 2012. Coregistration: Simultaneous alignment and modeling of articulated 3d shape. In *Computer Vision ECCV 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds., vol. 7577 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 242–255.

HONG, Q. Y., PARK, S. I., AND HODGINS, J. K. 2010. A data-driven segmentation for the shoulder complex. *Computer Graphics Forum 29*, 2, 537–544.

JAIN, A., THORMÄHLEN, T., SEIDEL, H.-P., AND THEOBALT, C. 2010. MovieReshape: Tracking and reshaping of humans in videos. *ACM Transactiosn on Graphics (Proc. SIGGRAPH) 29*, 6 (Dec.), 148:1–148:10.

KWON, J.-Y., AND LEE, I.-K. 2007. Rubber-like exaggeration for character animation. In *Proceedings of the 15th Pacific Conference on Computer Graphics and Applications*, IEEE Computer Society, Washington, DC, USA, PG '07, 18–26.

LEARDINI, A., CHIARI, L., CROCE, U. D., AND CAPPOZZO, A. 2005. Human movement analysis using stereophotogrammetry: Part 3. soft tissue artifact assessment and compensation. *Gait & Posture 21*, 2, 212 – 225.

LIU, Y., GALL, J., STOLL, C., DAI, Q., SEIDEL, H.-P., AND THEOBALT, C. 2013. Markerless motion capture of multiple characters using multiview image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 35*, 11, 2720–2735.

LIVNE, M., SIGAL, L., TROJE, N., AND FLEET, D. 2012. Human attributes from 3D pose tracking. *Computer Vision and Image Understanding 116*, 5, 648–660.

LOPER, M., 2014. Chumpy autodifferentiation library. http://chumpy.org/.

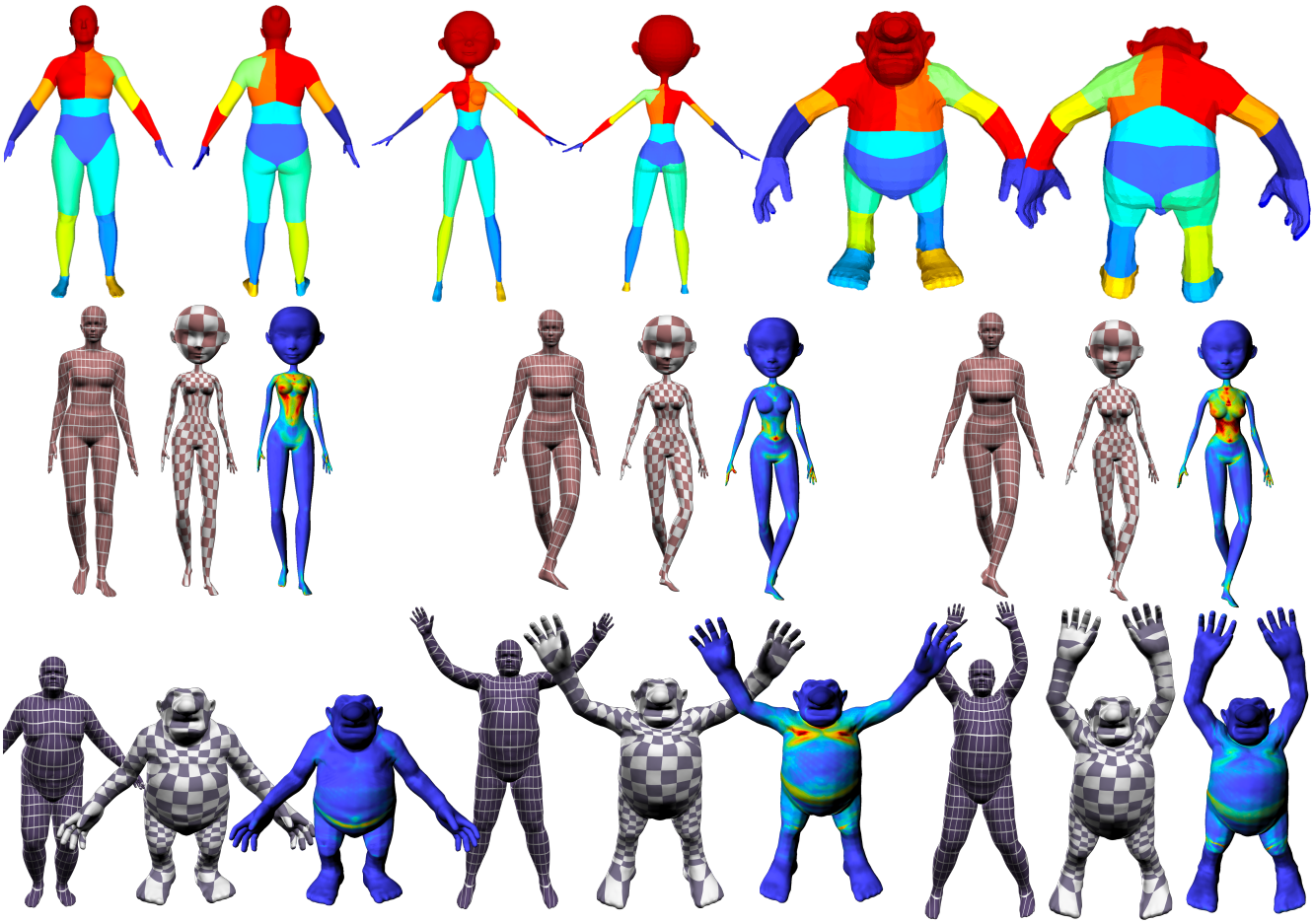

Figure 12: Retargeting soft-tissue motions. *Top row: Body part segmentation for human and stylized characters. Middle row: retargeting pose and soft-tissue motion of an actor (left) to a stylized female character (middle), with heat maps (right) illustrating the percentage of soft-tissue deformation; blue means zero and red means $\geq$ 20 percent deformation. Bottom row: retargeting to another stylized character. See* **accompanying video** *to visualize the soft-tissue motions.*

NEUMANN, T., VARANASI, K., HASLER, N., WACKER, M., MAGNOR, M., AND THEOBALT, C. 2013. Capture and statistical modeling of arm-muscle deformations. *Computer Graphics Forum 32*, 2 (May), 285–294.

NEUMANN, T., VARANASI, K., WENGER, S., WACKER, M., MAGNOR, M., AND THEOBALT, C. 2013. Sparse localized deformation components. *ACM Trans. Graph. 32*, 6 (Nov.), 179:1–179:10.

NOCEDAL, J., AND WRIGHT, S. J. 2006. *Numerical Optimization*, 2nd ed. Springer, New York.

PARK, S. I., AND HODGINS, J. K. 2006. Capturing and animating skin deformation in human motion. *ACM Trans. Graph. (Proc. SIGGRAPH) 25*, 3 (July), 881–889.

PARK, S. I., AND HODGINS, J. K. 2008. Data-driven modeling of skin and muscle deformation. *ACM Trans. Graph. (Proc. SIGGRAPH) 27*, 3 (Aug.), 96:1–96:6.

ROBINETTE, K., BLACKWELL, S., DAANEN, H., BOEHMER, M., FLEMING, S., BRILL, T., HOEFERLIN, D., AND BURNSIDES, D. 2002. Civilian American and European Surface Anthropometry Resource (CAESAR) final report. Tech. Rep. AFRL-HE-WP-TR-2002-0169, US Air Force Research Laboratory.

STARK, J., AND HILTON, A. 2007. Surface capture for performance-based animation. *IEEE Computer Graphics and Applications 27*, 3, 21–31.

TSOLI, A., MAHMOOD, N., AND BLACK, M. J. 2014. Breathing life into shape: Capturing, modeling and animating 3D human breathing. *ACM Trans. Graph., (Proc. SIGGRAPH) 33*, 4 (July), 52:1–52:11.

WADHWA, N., RUBINSTEIN, M., DURAND, F., AND FREEMAN, W. T. 2013. Phase-based video motion processing. *ACM Trans. Graph., (Proc. SIGGRAPH) 32*, 4 (July), 80:1–80:10.

WANG, H., XU, N., RASKAR, R., AND AHUJA, N. 2007. Videoshop: A new framework for spatio-temporal video editing in gradient domain. *Graph. Models 69*, 1, 57–70.

WU, H.-Y., RUBINSTEIN, M., SHIH, E., GUTTAG, J., DURAND, F., AND FREEMAN, W. T. 2012. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph. (Proc. SIGGRAPH) 31*, 4 (July), 65:1–65:8.