The space of human body shapes:
reconstruction and parameterization from range scans

Brett Allen   Brian Curless   Zoran Popović
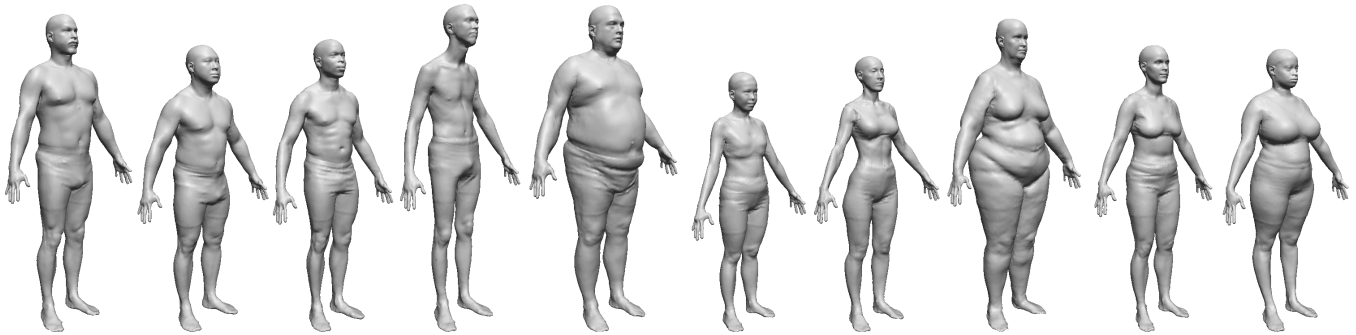
University of Washington

Figure 1: The CAESAR data set is a collection of whole-body range scans of a wide variety of individuals. Shown here are several range scans that have been hole-filled and fit to a common parameterization using our framework. Once this process is complete, we can analyze the variation in body shape in order to synthesize new individuals or edit existing ones.

Abstract

We develop a novel method for fitting high-resolution template meshes to detailed human body range scans with sparse 3D markers. We formulate an optimization problem in which the degrees of freedom are an affine transformation at each template vertex. The objective function is a weighted combination of three measures: proximity of transformed vertices to the range data, similarity between neighboring transformations, and proximity of sparse markers at corresponding locations on the template and target surface. We solve for the transformations with a non-linear optimizer, run at two resolutions to speed convergence. We demonstrate reconstruction and consistent parameterization of 250 human body models. With this parameterized set, we explore a variety of applications for human body modeling, including: morphing, texture transfer, statistical analysis of shape, model fitting from sparse markers, feature analysis to modify multiple correlated parameters (such as the weight and height of an individual), and transfer of surface detail and animation controls from a template to fitted models.

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

Keywords: deformations, morphing, non-rigid registration, synthetic actors

1 Introduction

The human body comes in all shapes and sizes, from ballet dancers to sumo wrestlers. Many attempts have been made to measure and categorize the scope of human body variation. For example, the photographic technique of Sheldon et al. [1940] characterizes physique using three parameters: *endomorphy*, the presence of soft roundness in the body; *mesomorphy*, the predominance of hardness and muscularity; and *ectomorphy*, the presence of linearity and skinniness. The field of anthropometry, the study of human measurement, uses combinations of bodily lengths and perimeters to analyze body shape in a numerical way.

Understanding and characterizing the range of human body shape variation has applications ranging from better ergonomic design of human spaces (e.g., chairs, car compartments, and clothing) to easier modeling of realistic human characters for computer animation. The shortcomings of high level characterizations and sparse anthropometric measurements, particularly for body modeling, is that they do not capture the detailed shape variations needed for realism.

One avenue for creating detailed human models is 3D scanning technology. However, starting from a range scan, substantial effort is needed to process the noisy and incomplete surface into a model suitable for animation. Further, the result of this effort is a model corresponding to a single individual that tells us little about the space of human shapes. Moreover, in the absence of a characterization of this space, editing a body model in a way that yields a plausible, novel individual is not trivial.

In this paper, we propose a method for creating a whole-body morphable model based on 3D scanned examples in the spirit of Blanz and Vetter's morphable face model [1999]. We begin with a set of 250 scans of different body types taken from a larger corpus of data (Section 1.1). By bringing these scans into full correspondence with each other, a difficult task in the context of related work (Section 2), we are able to morph between individuals, and begin to characterize and explore the space of probable body shapes.
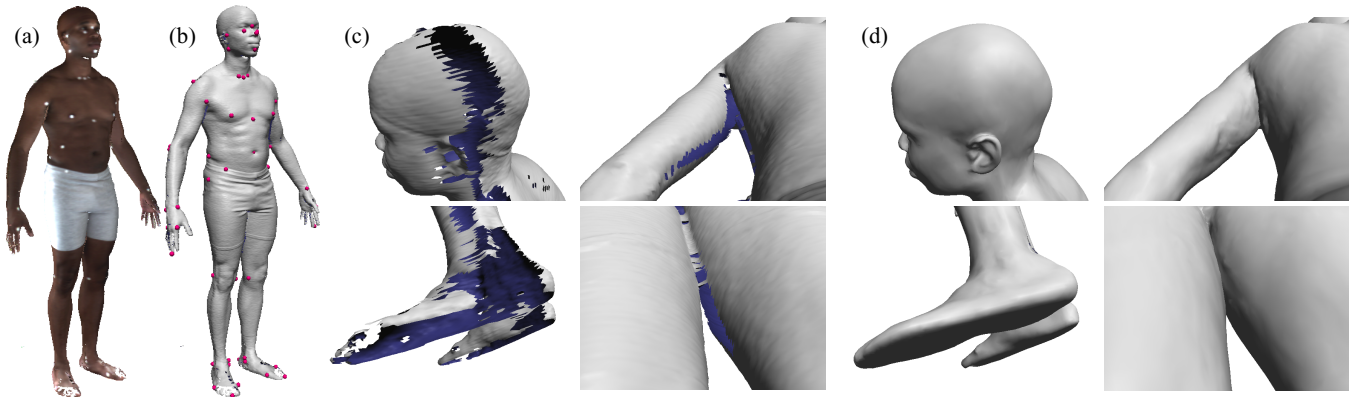
Figure 2: Parameterization of one of the CAESAR subjects. (a) Original scan, rendered with color texture (the white dots are the markers). (b) Scanned surface without texture. The marker positions are shown as red spheres. (c) Detail of holes in the scanned data, caused by occlusions and grazing angle views. Backfacing polygons are tinted blue. In clockwise order: the head, the underarm, between the legs, the feet. Note that erroneous polygons bridging the legs have been introduced by the mesh-stitching process. (d) Detail of difficult areas after template-based parameterization and hole filling (Section 3).

The central contribution of this paper is a template-based non-rigid registration technique for establishing a point-to-point correspondence among a set of surfaces with the same overall structure, but substantial variation in shape, such as human bodies acquired in similar poses. We formulate an optimization problem to solve for an affine transformation at each vertex of a high-resolution template using an objective function that trades off fit to the range data, fit to scattered fiducials (known markers), and smoothness of the transformations over the surface (Section 3). Our approach is robust in the face of incomplete surface data and fills in missing and poorly captured areas using domain knowledge inherent in the template surface. We require a set of feature markers to initialize the registration, although we show that once enough shapes have been matched, we do not require markers to match additional shapes. We use our fitting algorithm to create a consistent parameterization for our entire set of whole-body scans.

In addition, we demonstrate the utility of our approach by presenting a variety of applications for creating human digital characters (Section 4). These applications include somewhat conventional techniques such as transferring texture from one individual to another, morphing between shapes, and principal component analysis (PCA) of the shape space for automatic synthesis of novel individuals and for markerless matching. In addition, we demonstrate a form of feature analysis that enables modifying individuals by editing multiple correlated attributes (such as height and weight), plausible shape synthesis using only markers, and transfer of animation controls (skeletal and skinning) between the reconstructed models. We conclude the paper with some discussion and ideas for future work (Section 5).

1.1 Data set

Our source of whole-body 3D laser range scans is the Civilian American and European Surface Anthropometry Resource Project (CAESAR). The CAESAR project collected thousands of range scans of volunteers aged 18–65 in the United States and Europe. Each subject wore gray cotton bicycle shorts and a latex cap to cover the hair; the women also wore gray sports bras. Prior to scanning, 74 white markers were placed on the subject at anthropometric landmarks, typically at points where bones can be palpated through the skin (see Figure 2a and b). The 3D location of each landmark was then extracted from the range scan. In addition, anthropometric measurements were taken using traditional methods,

and demographic data such as age, weight, and ethnic group were recorded.

The raw range data for each individual consists of four simultaneous scans from a Cyberware whole body scanner. These data were combined into surface reconstructions using mesh stitching software. Each reconstructed mesh contains 250,000-350,000 triangles, with per-vertex color information. The reconstructed meshes are not complete (see Figure 2c), due to occlusions and grazing angle views. During the mesh-stitching step, each vertex was assigned a "confidence" value, as described by Turk and Levoy [1994], so that less reliable data are marked with lower confidence. For our experiment, we used a subset of the meshes in the CAESAR dataset, consisting of 125 male and 125 female scans with a wide variety of body types and ethnicities.

2 Related work

In this section, we discuss related work in the areas of modeling shape variation from examples, finding mutually consistent surface representations, filling holes in scanned data, and non-rigid surface registration.

The idea of using real-world data to model the variation of human shape has been applied to heads and faces several times. De-Carlo et al. [1998] use a corpus of anthropometric facial measurements to model the variation in face shapes. Blanz and Vetter [1999] also model facial variation, this time using dense surface and color data. They use the term *morphable model* to describe the idea of creating a single surface representation that can be adapted to fit all of the example faces. Using a polygon mesh representation, each vertex's position and color may vary between examples, but its semantic identity must be the same; e.g., if a vertex is located at the tip of the nose in one face, then it should be located at the tip of the nose in all faces. Thus, the main challenge in constructing the morphable model is to reparameterize the example surfaces so that they have a consistent representation. Since their head scans have cylindrical parameterization, Blanz and Vetter align the features using a modified version of 2D optical flow.

In the case of whole body models, finding a consistent representation becomes more difficult, as whole bodies cannot be parameterized cylindrically. Praun et al. [2001] describe a technique to establish an *n*-way correspondence between arbitrary meshes of the same topological type with feature markers. Unfortunately, whole-body range scans contain numerous holes (see Figure 2c) that pre-

vent us from using matching algorithms, such as Praun's, that rely on having complete surfaces.

Filling holes is a challenging problem in its own right, as discussed by Davis et al. [2002]. Their method and other recent, direct hole-free reconstruction methods [Carr et al. 2001; Whitaker 1998] have the nice feature that holes are filled in a smooth manner. However, while smooth hole-filling is reasonable in some areas, such as the top of the head and possibly in the underarm, other areas should not be filled smoothly. For example, the soles of the feet are cleanly cut off in the CAESAR scans, and so fair surface filling would create a smooth bulbous protrusion on the bottoms of the feet. The region between the legs is even more challenging, as many reconstruction techniques will erroneously bridge the right and left legs, as shown in Figure 2c. Here, the problem is not to fill the holes, but to add them.

The parameterization method described in our previous work [Allen et al. 2002] might seem to be a candidate for solving this problem. There, we start from a subdivision template that resembles the range surface, then re-parameterize the surface by sampling it along the template normals to construct a set of displacement maps, and finally perform smooth filling in displacement space. (A related displacement-mapped technique, without hole-filling, was also developed by Hilton et al. [2002].) Here smoothness is defined relative to the template surface, so that, for example, the soles of the feet would be filled in flat. However, to avoid crossing of sample rays, displacement-mapped subdivision requires that the template surface already be a fairly close match to the original surface [Lee et al. 2000], which is not trivial to achieve automatically considering the enormous variation in body shapes.

Kähler et al. [2002] parameterize incomplete head scans by deforming a template mesh to fit the scanned surface. Their technique has the additional benefit that holes in the scanned surface are filled in with geometry from the template surface, creating a more realistic, complete model. Their deformation is initialized using volumetric radial basis functions. The non-rigid registration technique of Szeliski and Lavallée [1994] also defines a deformation over a volume, in their case using spline functions. Although these approaches work well for largely convex objects, such as the human head, we have found that volumetric deformations are not as suitable for entire bodies. The difficulty is that branching parts, such as the legs, have surfaces that are close together spatially, but far apart geodesically. As a result, unless the deformation function is defined to an extremely high level of detail, one cannot formulate a volumetric deformation that affects each branch independently. In our work, we formulate a deformation directly on the body surface, rather than over an entire volume.

Our matching technique is based on an energy-minimization framework, similar to the framework of Marschner et al. [2000]. Marschner et al. regularize their fitting process using a surface smoothness term. Instead of using surface smoothness, our optimization minimizes variation of the deformation itself, so that holes in the mesh are filled in with detail from the template surface. Feldmar and Ayache [1994] describe a registration technique based on matching surface points, normals, and curvature while maintaining a similar affine transformation within spherical regions of space. Our smoothness term resembles Feldmar and Ayache's "locally affine deformations," but we do not use surface normals or curvature, as these can vary greatly between bodies. Further, our smoothness term is defined directly over the surface, rather than within a spherical volume.

3 Algorithm

We now describe our technique for fitting a template surface, $\mathcal{T}$, to a scanned example surface, $\mathcal{D}$. Each of these surfaces is represented as a triangle mesh (although any surface representation could be
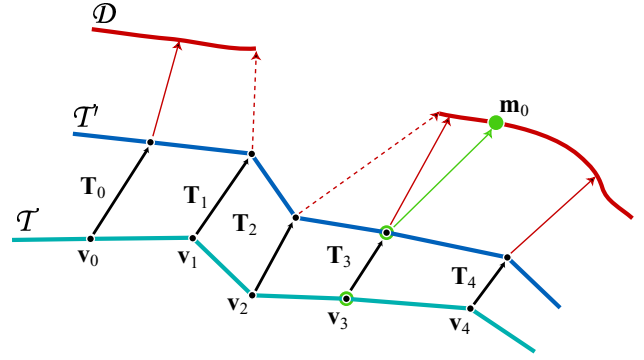

Figure 3: Summary of our matching framework. We want to find a set of affine transformations $\mathbf{T}_i$, that, when applied to the vertices $\mathbf{v}_i$ of the template surface $\mathcal{T}$, result in a new surface $\mathcal{T}'$ that matches the target surface $\mathcal{D}$. This diagram shows the match in progress; $\mathcal{T}'$ is moving towards $\mathcal{D}$, but has not yet reached it. The match proceeds by minimizing three error terms. The *data error*, indicated by the red arrows, is a weighted sum of the squared distances between the transformed template surface and $\mathcal{D}$. Note that the dashed red arrows do not contribute to the data error because the nearest point on $\mathcal{D}$ is a hole boundary. The *smoothness error* penalizes differences between adjacent $\mathbf{T}_i$ transformations. The *marker error* penalizes distance between the marker points on the transformed surface and on $\mathcal{D}$ (here $\mathbf{v}_3$ is associated with $\mathbf{m}_0$).

used for $\mathcal{D}$). To accomplish the match, we employ an optimization framework. Each vertex $\mathbf{v}_i$ in the template surface is influenced by a 4×4 affine transformation matrix $\mathbf{T}_i$. These transformation matrices comprise the degrees of freedom in our optimization, i.e., twelve degrees of freedom per vertex to define an affine transformation. We wish to find a set of transformations that move all of the points in $\mathcal{T}$ to a deformed surface $\mathcal{T}'$, such that $\mathcal{T}'$ matches well with $\mathcal{D}$.

We evaluate the quality of the match using a set of error functions: data error, smoothness error, and marker error. These error terms are summarized in Figure 3 and described in detail in the following three sections. Subsequently, we describe the optimization framework used to find a minimum-error solution. We then show how this approach creates a complete mesh, where missing data in the scan is suitably filled in using the template.

3.1 Data error

The first criterion of a good match is that the template surface should be as close as possible to the target surface. To this end, we define a data objective term E_d as the sum of the squared distances between each vertex in the template surface and the example surface:

$$E_d = \sum_{i=1}^{n} w_i \, \text{dist}^2(\mathbf{T}_i \mathbf{v}_i, \mathcal{D}), \qquad (1)$$

where n is the number of vertices in $\mathcal{T}$, w_i is a weighting term to control the influence of data in different regions (Section 3.5), and the dist() function computes the distance to the closest compatible point on $\mathcal{D}$.

We consider a point on $\mathcal{T}'$ and a point on $\mathcal{D}$ to be compatible if the surface normals at each point are no more than $90°$ apart (so that front-facing surfaces will not be matched to back-facing surfaces), and the distance between them is within a threshold (we use a threshold of 10 cm in our experiments). These criteria are used in the rigid registration technique of Turk and Levoy [1994]. In fact, if we had forced all of the $\mathbf{T}_i$ to be a single rigid body

transformation, then minimizing this data term would be virtually identical to the method of Turk and Levoy.

To accelerate the minimum-distance calculation, we precompute a hierarchical bounding box structure for $\mathcal{D}$, so that the closest triangles are checked first.

3.2 Smoothness error

Of course, simply moving each vertex in $\mathcal{T}$ to its closest point in $\mathcal{D}$ will not result in a very attractive mesh, because neighboring parts of $\mathcal{T}$ could get mapped to disparate parts of $\mathcal{D}$, and vice-versa. Further, there are infinitely many affine transformations that will have the same effect on a single vertex; our problem is clearly underconstrained using only E_d.

To constrain the problem, we introduce a smoothness error, E_s. By smoothness, we are not referring to smoothness of the deformed surface itself, but rather smoothness of the actual deformation applied to the template surface. In particular, we require affine transformations applied within a region of the surface to be as similar as possible. We formulate this constraint to apply between every two points that are adjacent in the mesh $\mathcal{T}$:

$$E_s = \sum_{\{i,j|\{\mathbf{v}_i,\mathbf{v}_j\}\in edges(\mathcal{T})\}} ||\mathbf{T}_i - \mathbf{T}_j||_F^2 \qquad (2)$$

where $||\cdot||_F$ is the Frobenius norm.

By minimizing the change in deformation over the template surface, we prevent adjacent parts of the template surface from being mapped to disparate parts of the example surface. The E_s term also encourages similarly-shaped features to be mapped to each other. For example, flattening out the template's nose into a cheek and then raising another nose from the other cheek will be penalized more than just translating or rotating the nose into place.

3.3 Marker error

Using the E_d and E_s terms would be sufficient if the template and example mesh were initially very close to each other. In the more common situation, where $\mathcal{T}$ and $\mathcal{D}$ are not close, the optimization can become stuck in local minima. For example, if the left arm begins to align with the right arm, it is unlikely that a gradient descent algorithm would ever back up and get the correct alignment. Indeed, a trivial global minimum exists where all of the affine transformations are set to a zero scale and the (now zero-dimensional) mesh is translated onto the example surface.

To avoid these undesirable minima, we identify a set of points on the example surface that correspond to known points on the template surface. These points are simply the anthropometric markers that were placed on the subjects prior to scanning (see Figure 2a and b). We call the 3D location of the markers on the example surface $\mathbf{m}_{1...m}$, and the corresponding vertex index of each marker on the template surface $\kappa_{1...m}$. The marker error term E_m minimizes the distance between each marker's location on the template surface and its location on the example surface:

$$E_m = \sum_{i=1}^{m} ||\mathbf{T}_{\kappa_i}\mathbf{v}_{\kappa_i} - \mathbf{m}_i||^2 \qquad (3)$$

In addition to preventing undesirable minima, this term also encourages the correspondence to be correct at the marker locations. The markers represent points whose correspondence to the template is known *a priori*, and so we can make use of this fact in our optimization. However, we do not require that *all* salient features have markers. (If we did, then we would need many more markers than are present in the CAESAR data!) The smoothness and data error terms alone are capable of aligning areas of similar shape, as long as local minima can be avoided.

3.4 Combining the error

Our complete objective function E is the weighted sum of the three error functions:

$$E = \alpha E_d + \beta E_s + \gamma E_m, \qquad (4)$$

where the weights α, β, and γ are tuned to guide the optimization as described below. We run the optimization using L-BFGS-B, a quasi-Newtonian solver [Zhu et al. 1997].

One drawback of the formulation of E_s is that it is very localized; changes to the affine transformation need to diffuse through the mesh neighbor-by-neighbor with each iteration of the solver. This locality leads to slow convergence and makes it easy to get trapped in local minima. We avoid this problem by taking a multiresolution approach. Using the adaptive parameterization framework of Lee et al. [1998], we generate a high and a low resolution version of our template mesh, and the relationship between the vertices of each. We first run our optimization using the low resolution version of $\mathcal{T}$ and a smoothed version of $\mathcal{D}$. This optimization runs quickly, after which the transformation matrices are upsampled to the high-resolution version of $\mathcal{T}$, and we complete the optimization at full resolution.

We also vary the weights, α, β, and γ, so that features move freely and match up in the early stages, and then finally the data term is allowed to dominate. Although the marker data is useful for global optimization, we found that the placement of the markers was somewhat unreliable. To reduce the effect of variable marker placement, we reduce the weight of the marker term in the final stages of the optimization. The overall optimization schedule is as follows:

At low resolution:

1. Fit the markers first: $\alpha = 0$, $\beta = 1$, $\gamma = 10$
2. Allow the data term to contribute: $\alpha = 1$, $\beta = 1$, $\gamma = 10$

At high resolution:

3. Continue the optimization: $\alpha = 1$, $\beta = 1$, $\gamma = 10$
4. Allow the data term to dominate: $\alpha = 10$, $\beta = 1$, $\gamma = 1$

3.5 Hole-filling

We now explain how our algorithm fills in missing data using domain information. Suppose that the closest point on $\mathcal{D}$ to a transformed template point $\mathbf{T}_i\mathbf{v}_i$ is located on a boundary edge of $\mathcal{D}$ (as shown by the dashed red lines in Figure 3). In this situation we set the weight w_i in E_d to zero, so that the transformations $\mathbf{T}_i$ will only be affected by the smoothness term, E_s. As a result, holes in the example mesh will be filled in by seamlessly transformed parts of the template surface.

In addition to setting w_i to zero where there is no data, we also wish to downweight the importance of poor data, i.e., surface data near the holes and samples acquired at grazing angles. Since each vertex in the CAESAR mesh has a confidence value based on these criteria, we simply set w_i to the barycentrically interpolated confidence value of the closest point on $\mathcal{D}$. (In practice, we scale and clamp the confidence values so that the range $0 \ldots 0.2$ maps to a w_i in the range $0 \ldots 1$.) Because the weights taper gradually to zero near holes, we obtain a smooth blend between regions with good data and regions with no data.

In some areas, such as the ears and the fingers, the scanned data is particularly poor, containing only scattered fragments of the true surface. Matching these fragments automatically to the detailed template surface is quite difficult. Instead, we provide a mechanism for manually identifying areas on the template that are known to scan poorly, and then favor the template surface over the scanned surface when fitting these areas. In the marked areas, we modify the data term's w_i coefficient using a multiplicative factor of
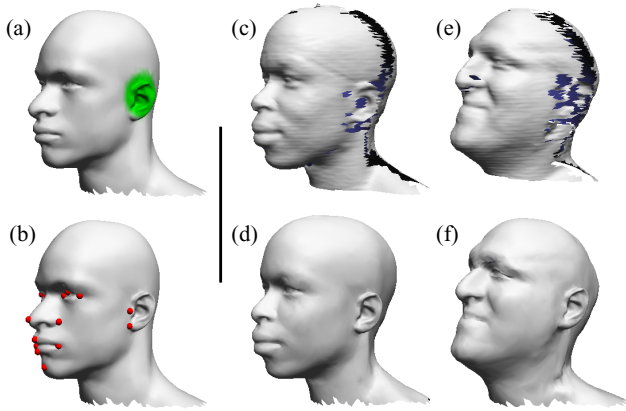
Figure 4: Using a template mesh to synthesize detail lost in the scan. (a) The template mesh. Since we know the ear does not scan well, we weight the ear vertices to have a zero data-fitting term (shown in green). (b) Since the template mesh does not have the CAESAR markers, we use a different set of markers based on visually-identifiable features to ensure good correspondence. (c) A head of one of the subjects. Interior surfaces are tinted blue. (d) The template head has been deformed to match the scanned head. Note that the ear has been filled in. (e) Another scanned head, with a substantially different pose and appearance from the template. (f) The template mapped to (e). The holes have been filled in, and the template ear has been plausibly rotated and scaled.
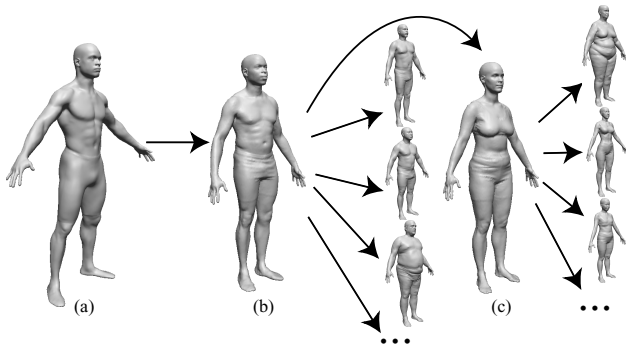

Figure 5: We begin with a hole-free, artist-generated mesh (a), and map it to one of the CAESAR meshes using a set of 58 manually selected, visually identifiable landmarks. We then use the resulting mesh (b), and 72 of the CAESAR markers (plus two we added), as a template for all of the male scans. For the female scans, we first map our male template to one of the female subjects, and then use the resulting mesh as a template (c).

zero, tapering towards 1 at the boundary of the marked area. As a result, the transformation smoothness dominates in the marked regions, and the template geometry is carried into place. As shown in Figure 4, this technique can have a kind of super-resolution effect, where detail that was not available in the range data can be drawn from the template.

4 Applications

We used our matching algorithm to create a hole-free and mutually consistent surface parameterization of 250 range scans, using the workflow illustrated in Figure 5. To bootstrap the process, we
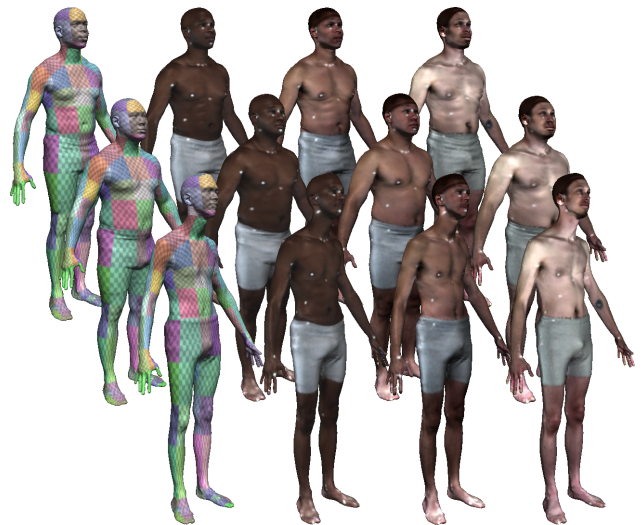

Figure 6: To test the quality of our matching algorithm, we apply the same texture (each column) to three different meshes. The mesh in each row is identical. On the left, we use a checkerboard pattern to verify that features match up. The right-hand 3×3 matrix of renderings use the textures extracted from the range scans. (The people along the diagonal have their original textures.)

matched a high quality, artist-generated mesh to one of the CAESAR scans using 58 manually selected landmarks. This fitted mesh served as a template for fitting to the remaining models with the help of the CAESAR markers. Of the 74 CAESAR original markers, the two located on the lower ribs varied in placement to such an extent that we omitted them. To compensate, we manually introduced a new marker at the navel in each scan, as well as a new marker at the tip of each nose to improve the matching on the face.

In the remainder of this section, we demonstrate how the representation provided by our matching algorithm can be used to analyze, create, and edit detailed human body shapes.

4.1 Transfer of textures and morphing

As in Praun et al. [2001], once we have a consistent parameterization, we can transfer texture maps between any pair of meshes. Although this is a simple application, its success hinges on the quality of our matching algorithm. Figure 6 demonstrates transferring texture between three subjects.

Similarly, we can morph between any two subjects by taking linear combinations of the vertices. Figure 7 demonstrates this application. In order to create a good morph between individuals, it is critical that all features are well-aligned; otherwise, features will cross-fade instead of moving. Notice that even features that were not given markers, such as the bottom of the breasts and the waistline, morph smoothly.

4.2 Principal component analysis

Principal component analysis (PCA) has been used to analyze facial features [Praun et al. 2001; Blanz and Vetter 1999; Turk and Pentland 1991]. The main advantage is data compression, since the vectors with low variance can be discarded, and thus the full data set does not need to be retained in order to closely approximate the original examples.

Suppose we match k scanned examples, and our template surface has n vertices. We stack the vertices of the parameterized scans into k column vectors $\mathbf{s}_i$ of height $3n$. Let the average of $\{\mathbf{s}_i\}$ be $\bar{\mathbf{s}}$, and
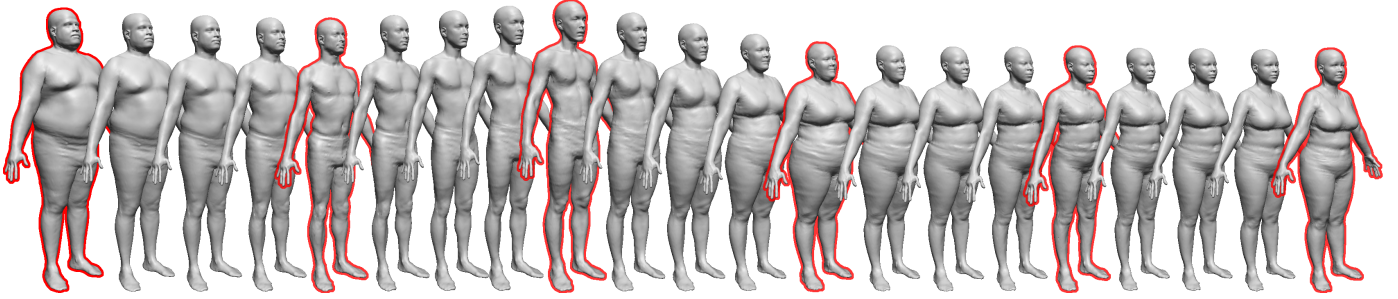
Figure 7: Morphing between individuals. Each of the keyframe models (outlined) are generated from a Gaussian distribution in PCA space. These synthesized individuals have their own character, distinct from those of the original scanned individuals. The in-between models are created by linearly interpolating the vertices of the keyframes.

$\mathbf{u}_i$ be $\mathbf{s}_i - \bar{\mathbf{s}}$. We assemble the $\mathbf{u}_i$ into a $3n \times (k-1)$ matrix $\mathbf{U}$. Principal component analysis of $\mathbf{U}$ yields a set of principal vectors $\mathbf{c}_{1...k-1}$, each of size $3n$. Associated with each principal vector $\mathbf{c}_i$ is a variance σ_i^2, and the vectors are sorted so that $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_{k-1}^2$.

We can use these variance terms to synthesize new random individuals. By sampling from the Gaussian distribution that the PCA represents, we can create an unlimited number of new individuals who, for the most part, have a realistic appearance, but do not look like any particular individual from the example set. A few randomly-generated models are outlined in red in Figure 7. (Note that we run PCA separately on the male and female data.)

4.3 Feature analysis

Principal component analysis helps to characterize the space of human body variation, but it does not provide a direct way to explore the range of bodies with intuitive controls, such as height, weight, age, and sex. Blanz and Vetter [1999] devise such controls for single variables using linear regression. Here we show how to relate several variables simultaneously by learning a linear mapping between the controls and the PCA weights. If we have l such controls, the mapping can be represented as a $(k-1) \times (l+1)$ matrix, $\mathbf{M}$:

$$\mathbf{M} \begin{bmatrix} f_1 & \cdots & f_l & 1 \end{bmatrix}^T = \mathbf{p}, \qquad (5)$$

where f_i are the feature values of an individual, and $\mathbf{p}$ are the corresponding PCA weights.

We can draw feature information from the demographic data associated with each CAESAR scan. After assembling the feature vectors into an $(l+1) \times k$ feature matrix $\mathbf{F}$, we solve for $\mathbf{M}$ as

$$\mathbf{M} = \mathbf{P}\mathbf{F}^+, \qquad (6)$$

where $\mathbf{F}^+$ is the pseudoinverse of $\mathbf{F}$. We can then create a new feature vector, e.g., a desired height and weight, and create an average-looking individual with those characteristics, as shown in the left part of Figure 10 on the last page of this paper. (Since this method is a linear approximation, and since weight is roughly proportional to volume, we actually use the cube root of the weight, to make it comparable with the height measurements.)

In addition, we can create delta-feature vectors of the form:

$$\Delta\mathbf{f} = \begin{bmatrix} \Delta f_1 & \cdots & \Delta f_l & 0 \end{bmatrix}^T \qquad (7)$$

where each Δf_i is the difference between a target feature value and the actual feature value for an individual. By adding $\Delta\mathbf{p} = \mathbf{M}\Delta\mathbf{f}$ to the PCA weights of that individual, we can edit their features, e.g., making them gain or lose weight, and/or become taller or shorter, as shown in the right part of Figure 10.
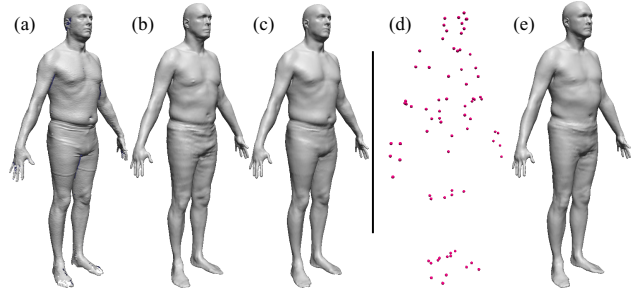

Figure 8: PCA-based fitting. (a) A scanned mesh that was not included in the data set previously, and does not resemble any of the other scans. (b) A surface match using PCA weights and no marker data. (c) Using (b) as a template surface, we get a good match to the surface using our original method without markers. (d) Next, we demonstrate using very sparse data; in this case, only the 74 marker points. (e) A surface match using PCA weights and no surface data.

4.4 Markerless matching

Principal component analysis also gives us a way to search the space of possible bodies given partial data. Instead of finding a smooth set of transformations applied to each vertex (as described in section 3.2), we can search for a set of principle component weights that match the data. This is similar to the bootstrapping technique of Blanz and Vetter [1999].

Suppose we have a body scan without any marker data. If the template surface is close enough to the new scan, then we can use the same optimization as before, but if the new scan is substantially different then the match will fail. In this case, we search in PCA space instead of transformation space, and replace E_s with the following term indicating the likelihood of a particular set of PCA weights:

$$E_p = \sum_{i=1}^{k'} (p_i/\sigma_i)^2, \qquad (8)$$

where the p_i are the PCA weights, σ_i^2 are the corresponding variances, and k' is the number of components used.

The new data term is similar to the one in Section 3.1, except we are matching against the PCA-reconstructed surface, $\mathbf{r}$:

$$\mathbf{r} = \bar{\mathbf{s}} + \sum_{j=1}^{k'} p_j \mathbf{c}_j \qquad (9)$$

$$E'_d = \sum_{i=1}^{n} w_i \, \mathrm{dist}^2([r_{3i} \ r_{3i+1} \ r_{3i+2}]^T, \mathcal{D}) \qquad (10)$$

6

The overall error that we optimize is a weighted sum of E_p and E'_d.

As in Blanz and Vetter [1999], we set k' to be small initially, and increase it in stages. Once a closest fit is found using this optimization, we use the reconstructed shape as the template surface for our original algorithm (minus the marker term) and complete the fit. Figure 8a–c demonstrates this approach.

4.5 Marker-only matching

We now consider the converse situation, where no surface data is available, and we have *only* the marker data, as shown in Figure 8d. One could get just marker data using less expensive equipment than a laser range scanner (e.g., using a handful of calibrated photographs of a stationary subject). Using the E_p term from the previous section, and a similarly modified E_m term, we can estimate the approximate shape of the subject (Figure 8e).

4.6 Instrumentation transfer

Beyond providing tools for realistic human body analysis and modeling, we hope to create figures that can be readily animated. To animate an articulated figure, we first need to define a skeleton for controlling the pose, and then associate each vertex's position with the skeleton in some way. This association process is called skinning, and a variety of techniques are used in popular animation packages. In this paper, we assume that one of the meshes has been properly instrumented with a skeleton for animation. This instrumentation can be done manually, or using a semi-automatic process such as the one proposed by Hilton et al. [2002].

Once we have instrumented one model, we would like to transfer its skeleton and skinning information to other parameterized scans, or to synthesized or edited characters. To transfer a skeleton, we begin by choosing 2–3 points on the surface to act as markers for each joint in the skeleton. These points can be the original anthropometric markers or other points; the main criterion is that their position is approximately rigid with respect to their associated joint. We then calculate the local position of these markers in the joint's coordinate frame. Having chosen a set of vertices as markers on one mesh, we know the location of those markers on any other mesh because of our consistent parameterization. Using inverse kinematics, we can then solve for the skeleton pose and bone lengths that give the best match between each marker's position in the joint coordinate frame and its global position derived from the mesh. This approach is not precise, since the marker's local position is assumed to be fixed, whereas in reality the local position depends on body thickness. However, with enough markers a reasonable skeleton can be determined for animation purposes, as shown in Figure 9.

Once the skeleton transfer is complete, the skinning information must be transferred as well. We employ a skinning scheme based on per-vertex weights. In this case, the transfer is trivial: since the vertices in each mesh are in correspondence, the weights can be directly copied.

5 Discussion and future work

In this section, we summarize some of the insights gained from this research and suggest a few future directions.

First of all, we found that, as a general reconstruction strategy, our template-based method works fairly well in practice. We were able to match all of our scanned examples to a reasonable degree. In less than 5% of the examples, the lips were misaligned, due largely to the paucity and variable placement of the CAESAR markers on the face.

One assumption made during this work is that the pose of the template is similar (though not necessarily identical) to the target
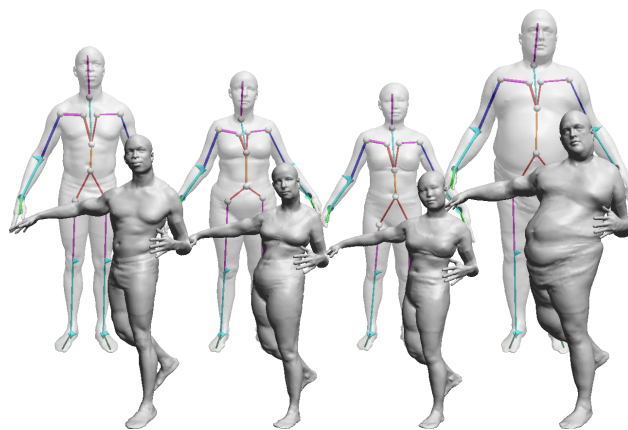

Figure 9: Skeleton transfer. We manually created a skeleton and skinning method for the scanned individual in the top left. The skeletons for the other three scanned individuals in the top row were generated automatically. In the bottom row, we show each of the parameterized scans put into a new pose using the skeleton and transferred skinning weights.

surface. If the poses are quite different, then the optimized template has to contain locally dissimilar transformations at bending joints, something that we currently penalize. An area for future work is to employ a posable template that tries to match the pose of the character in addition to the other fitting criteria. Interestingly, we also found that the small variations in pose that were present in our dataset, while not problematic for our fitting procedure, did impact the PCA analysis. Some of the components corresponded roughly to features one might expect, such as height variation and approximate body types (or both), but a number of them also clearly included pose variations. By factoring out pose, we would expect to achieve a more compact PCA representation. Indeed, such a model could also be used to accomplish objectives such as body shape estimation from photographs of bodies in arbitrary poses, in the spirit of Blanz and Vetter's [1999] work on human faces.

Our PCA analysis is really only suggestive of the kind of information we might learn from human body datasets. Our development of the space of body shapes is based on a relatively small dataset, and indeed we hope to incorporate more of the CAESAR scans in the future. Still, PCA is just one tool in the statistician's toolbox – a tool that sees the data as samples drawn from a single, multi-dimensional Gaussian distribution. Applying more sophisticated analyses (e.g., mixtures of Gaussians) to determine the "true" landscape of human shape variations remains an area for future work.

Finally, although we demonstrate transfer of animation parameters such as a skeleton and skinning weights, the quality of the results is only as good as the skinning algorithm used on the template. Transferring more sophisticated surface motions, e.g. employing example-based methods developed by a number of researchers [Lewis et al. 2000; Sloan et al. 2001; Allen et al. 2002], could lead to more sophisticated and compelling animation transfer.

6 Acknowledgments

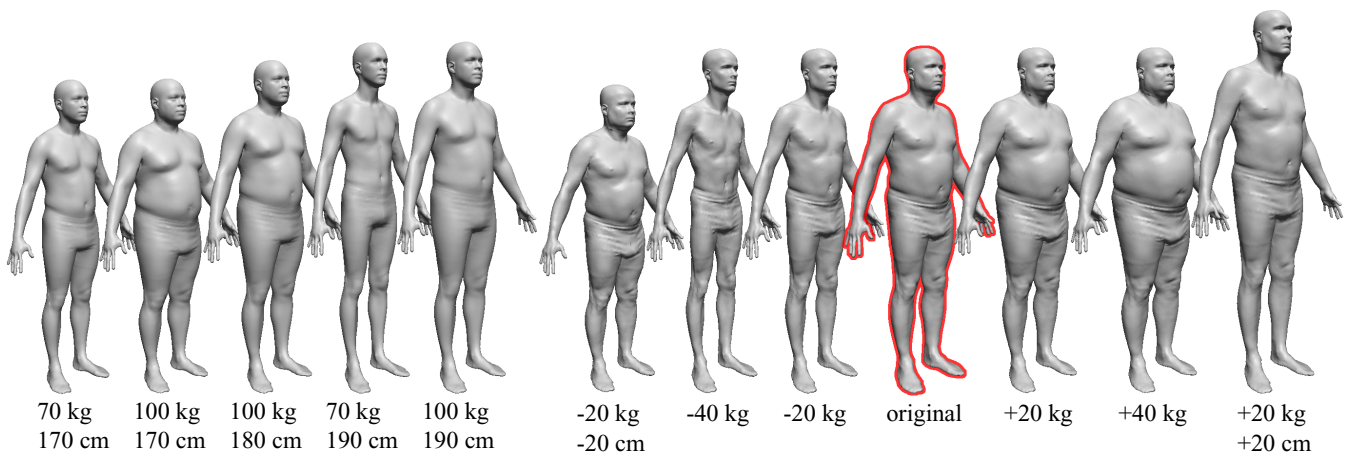| 70 kg | 100 kg | 100 kg | 70 kg | 100 kg | -20 kg | -40 kg | -20 kg | original | +20 kg | +40 kg | +20 kg |
| 170 cm | 170 cm | 180 cm | 190 cm | 190 cm | -20 cm | | | | | | +20 cm |

Figure 10: The left part of this figure demonstrates *feature-based synthesis*, where an individual is created with the required height and weight. On the right, we demonstrate *feature-based editing*. The outlined figure is one of the original subjects, after being parameterized into our system. The gray figures demonstrate a change in height and/or weight. Notice the double-chin in the heaviest example, and the boniness of the thinnest example.

References

ALLEN, B., CURLESS, B., AND POPOVIĆ, Z. 2002. Articulated body deformation from range scan data. 612–619.

BLANZ, V., AND VETTER, T. 1999. A morphable model for the synthesis of 3D faces. In *Proceedings of ACM SIGGRAPH 99*, ACM Press/Addison-Wesley Publishing Co., New York, A. Rockwood, Ed., Computer Graphics Proceedings, Annual Conference Series, 187–194.

CARR, J. C., BEATSON, R. K., CHERRIE, J. B., MITCHELL, T. J., FRIGHT, W. R., MCCALLUM, B. C., AND EVANS, T. R. 2001. Reconstruction and representation of 3D objects with radial basis functions. In *Proceedings of ACM SIGGRAPH 2001*, ACM Press / ACM SIGGRAPH, New York, E. Fiume, Ed., Computer Graphics Proceedings, Annual Conference Series, 67–76.

DAVIS, J., MARSCHNER, S. R., GARR, M., AND LEVOY, M. 2002. Filling holes in complex surfaces using volumetric diffusion. In *Proc. First International Symposium on 3D Data Processing, Visualization, and Transmission*.

DECARLO, D., METAXAS, D., AND STONE, M. 1998. An anthropometric face model using variational techniques. In *Proceedings of ACM SIGGRAPH 98*, ACM Press, Computer Graphics Proceedings, Annual Conference Series, 67–74.

FELDMAR, J., AND AYACHE, N. 1994. Rigid and affine registration of smooth surfaces using differential properties. In *ECCV (2)*, 397–406.

HILTON, A., STARCK, J., AND COLLINS, G. 2002. From 3D shape capture to animated models. In *Proc. First International Symposion on 3D Data Processing, Visualization, and Transmission (3DPVT 2002)*.

KÄHLER, K., HABER, J., YAMAUCHI, H., AND SEIDEL, H.-P. 2002. Head shop: Generating animated head models with anatomical structure. In *Proceedings of the 2002 ACM SIGGRAPH Symposium on Computer Animation*, ACM SIGGRAPH, San Antonio, USA, S. N. Spencer, Ed., Association of Computing Machinery (ACM), 55–64.

LEE, A. W. F., SWELDENS, W., SCHRÖDER, P., COWSAR, L., AND DOBKIN, D. 1998. MAPS: Multiresolution adaptive parameterization of surfaces. In *Proceedings of ACM SIGGRAPH 98*, ACM Press, Computer Graphics Proceedings, Annual Conference Series, 95–104.

LEE, A., MORETON, H., AND HOPPE, H. 2000. Displaced subdivision surfaces. In *Proceedings of ACM SIGGRAPH 2000*, ACM Press / ACM SIGGRAPH / Addison Wesley Longman, K. Akeley, Ed., Computer Graphics Proceedings, Annual Conference Series, 85–94.

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*, ACM Press / ACM SIGGRAPH / Addison Wesley Longman, K. Akeley, Ed., Computer Graphics Proceedings, Annual Conference Series, 165–172.

MARSCHNER, S. R., GUENTER, B., AND RAGHUPATHY, S. 2000. Modeling and rendering for realistic facial animation. In *Proceedings of 11th Eurographics Workshop on Rendering*, 231–242.

PRAUN, E., SWELDENS, W., AND SCHRÖDER, P. 2001. Consistent mesh parameterizations. In *Proceedings of ACM SIGGRAPH 2001*, ACM Press / ACM SIGGRAPH, New York, E. Fiume, Ed., Computer Graphics Proceedings, Annual Conference Series, 179–184.

SHELDON, W. H., STEVENS, S. S., AND TUCKER, W. B. 1940. *The Varieties of Human Physique*. Harper & Brothers Publishers, New York.

SLOAN, P.-P., ROSE, C., AND COHEN, M. F. 2001. Shape by example. In *Proceedings of 2001 Symposium on Interactive 3D Graphics*, 135–143.

SZELISKI, R., AND LAVALLÉE, S. 1994. Matching 3-D anatomical surfaces with non-rigid deformations using octree-splines. In *IEEE Workshop on Biomedical Image Analysis*, IEEE Computer Society, 144–153.

TURK, G., AND LEVOY, M. 1994. Zippered polygon meshes from range images. In *Proceedings of ACM SIGGRAPH 94*, ACM Press, Computer Graphics Proceedings, Annual Conference Series, 311–318.

TURK, M., AND PENTLAND, A. 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience 3*, 1, 71–86.

WHITAKER, R. 1998. A level-set approach to 3-D reconstruction from range data. *International Journal of Computer Vision 29*, 3, 203–231.

ZHU, C., BYRD, R. H., LU, P., AND NOCEDAL, J. 1997. Algorithm 778. L-BFGS-B: Fortran subroutines for Large-Scale bound constrained optimization. *ACM Transactions on Mathematical Software 23*, 4 (Dec.), 550–560.