EUROGRAPHICS 2009 / P. Dutré and M. Stamminger
(Guest Editors)

Volume 28 (2009), Number 2

A Statistical Model of Human Pose and Body Shape

N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H.-P. Seidel

MPI Informatik, Saarbrücken, Germany

Abstract

Generation and animation of realistic humans is an essential part of many projects in today's media industry. Especially, the games and special effects industry heavily depend on realistic human animation. In this work a unified model that describes both, human pose and body shape is introduced which allows us to accurately model muscle deformations not only as a function of pose but also dependent on the physique of the subject. Coupled with the model's ability to generate arbitrary human body shapes, it severely simplifies the generation of highly realistic character animations. A learning based approach is trained on approximately 550 full body 3D laser scans taken of 114 subjects. Scan registration is performed using a non-rigid deformation technique. Then, a rotation invariant encoding of the acquired exemplars permits the computation of a statistical model that simultaneously encodes pose and body shape. Finally, morphing or generating meshes according to several constraints simultaneously can be achieved by training semantically meaningful regressors.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling

1. Introduction

The automatic generation and animation of realistic humans is an increasingly important discipline in the computer graphics community. The applications of a system that simultaneously models pose and body shape include crowd generation for movie or game projects, creation of custom avatars for online shopping or games, or usability testing of virtual prototypes. But also other problems such as human tracking or even biometric applications can benefit from such a model.

Realistic results for human animation can be obtained by simulating the tissue deformation on top of modeled skeletal bones [SPCM97, DCKY02]. This approach has been researched extensively but involves a lot of manual modeling since not only the surface but also the bones, muscles, and other tissues have to be designed. Additionally, these methods tend to be computationally expensive since they involve physically based tissue simulation.

In order to reduce the required amount of manual modeling, several systems have been proposed that attempt to create general human models from 3D scans. One of the major difficulties with most previously suggested methods like

Figure 1: *Muscle bulging is not only a function of the underlying skeleton but is instead closely correlated with the physique of the subject.*

SCAPE [ASK*05] or [ACP03, SMT04, WPP07, WSLG07] is that they rely on different means for encoding shape and pose. Pose (and the effects thereof, e.g., muscle bulging) are stored with the help of an underlying skeleton while body shape is encoded using variational methods or envelope skinning. During animation, the outcomes of the two methods have to be combined in an additional step.

Allen et al., on the other hand, presented a method that learns skinning weights for corrective enveloping from 3D-scan data [ACPH06]. They propose to use a maximum a

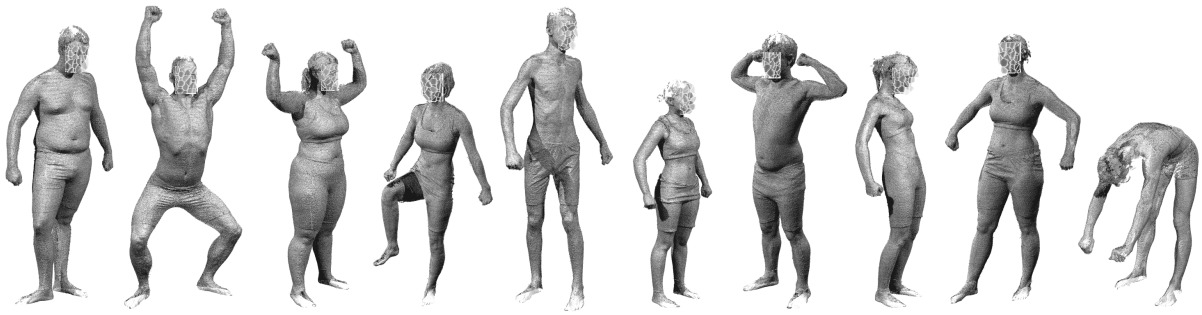

Figure 2: *A few examples of scans included in our database.*

posteriori estimation for solving a highly nonlinear function which simultaneously describes pose, skinning weights, bone parameters, and vertex positions. The authors are able to change weight or height of a character during animation and the muscle deformation looks significantly more realistic than linear blend skinning. However, since this function has a high number of degrees of freedom, the optimization procedure is very expensive. Additionally, the number of support poses that can be computed per subject is limited by the complexity of the approach, which in turn bounds the achievable realism.

A major benefit of the shared encoding is that it is easily possible to encode correlations between pose and body shape, e.g., the body surface deformation generated by the motion performed by an athletic person exhibits different properties than the same motion carried out by a person with less pronounced skeletal muscles (cf. Fig. 1).

The model we propose is based on statistical analysis of over 550 full body 3D scans taken of 114 subjects. Additionally, all subjects are measured with a commercially available impedance spectroscopy body fat scale and a medical grade pulse oximeter.

A sophisticated semi-automatic non-rigid model registration technique is used to bring the scans into semantic correspondence, i.e. to fit a template model to every scan. This approach is similar to [ARV07]. The registration ultimately leads to a semantically unified description of all scans. I.e. the same vertices denote the same semantic position on the surface of a scan.

Models based on statistical analysis of human body shape that operate directly on vertex positions have the distinct disadvantage that body parts that are rotated have a completely differently representation. For this reason most previous techniques [ASK*05, SMT04, WPP07, WSLG07] encode the surface relative to an embedded skeleton. Instead, we opt to use an encoding that is locally translation and rotation invariant in the sense that a body part retains its exact encoding if it is translated or rotated relative to the rest of the body. For example, the encoding of a hand remains unchanged even if the arm is raised and rotated between two scans. A similar encoding has recently been proposed by Kircher and Garland [KG08] in the context of mesh editing.

The resulting shape coefficients and their variances are analyzed using a statistical method. Similar to Blanz et al. [BV99], regression functions are trained to correlate them to semantically significant values like weight, body fat content, or pose. Allen et al. [ACP03] uses a simple linear method to generate models conforming to a set of semantic constraints. Allen et al. do not show quantitative analysis of the accuracy of their morphing functions. In contrast Seo and Magnenat-Thalmann describe a method for generating human bodies given a number of high level semantic constraints [SMT04] and evaluate the accuracy of linear regression based morphing functions. In Section 5 we present a quantitative analysis comparing linear and non-linear regression functions for various body measures. Scherbaum et al. [SSSB07] have concluded in the context of face morphing that although non-linear regression functions are numerically more accurate, the visual difference to the linear counterpart is minimal.

Thus, morphing a scan to conform to a given set of constraints only involves solving a single linear equation system in the minimum norm sense. We present several applications of our model, including animation of skeleton data produced by a motion capture system (Sec. 5.4), fitting of 3D scan data (Sec. 2), and generation of realistic avatars given only a sparse number of semantic constraints (Sec. 5.2).

The primary contributions can be summarized as follows:

- A large statistical model of human bodies that jointly encodes pose and body shape is introduced. The primary advantage of this approach is that correlations between body shape and pose are encoded in addition to the pure information on pose and shape.
- A locally translation and rotation invariant encoding for the meshes is introduced. This fundamentally non-linear transformation subsequently allows us to compute a linear statistical model on the data.
- A quantitative evaluation of linear and non-linear regres-

sors for morphing according to semantically meaningful constraints.

Additionally, an extension to a semi-automatic non-rigid registration scheme is proposed that establishes semantic correspondence between the scans, a handle based paradigm for modeling human body shapes is presented, and our database of approx. 550 full body 3D scans is made available for scientific purposes [XXX08].

The rest of the paper is structured as follows. In the following Section the scan database and the registration scheme that brings scans into semantic correspondence are introduced. The rotation invariant model representation is described in Section 3. The regression framework is presented in Section 4, applications and experimental evaluation results are shown in Section 5, and a brief summary is given in Section 6.

2. Scan Database & Registration

The procedure to register all scans from the database has to be robust and almost fully automatic since the massive number of scans does not allow tedious manual intervention to be performed on every scan. We have consequently opted for a two stage process. First, skeleton-based deformation of the template is used to estimate pose and rough size of the scanned subject. Then a non-rigid deformation technique is employed to fit the template to the scanned point cloud. This is necessary and desirable since the stability of non-rigid registration increases significantly when the initial mesh configuration is close to the point cloud that is to be matched. For complex poses (cf. Fig. 2) direct non-rigid deformation converges very slowly. In contrast, due to the limited number of degrees of freedom, a skeleton based method converges quickly even in extreme cases. Using skeleton fitting for a first estimate also allows us to use the resulting data for pose regression.

The process depends only on a few manually placed correspondences for each scan as the scans do not feature any prescribed landmarks as present for example in the CAESAR database [RDP99]. The template, an example of a labeled scan, the skeleton based fitting, and the final registration result are shown in Figure 4.

2.1. Scan Database

The project is based on a database of dense full body 3D scans, captured using a Vitronic laser scanner. Of the 114 subjects aged 17 to 61, 59 are male and 55 female. In addition to one standard pose that allows the creation of a shape-only model, all subjects are scanned in at least 9 poses selected randomly from a set of 34 poses, see Fig. 3 for the scan distribution. Unlike SCAPE or more recently [ACPH06] we sample the space more densely. This allows our model to capture pose-body-shape correlations more

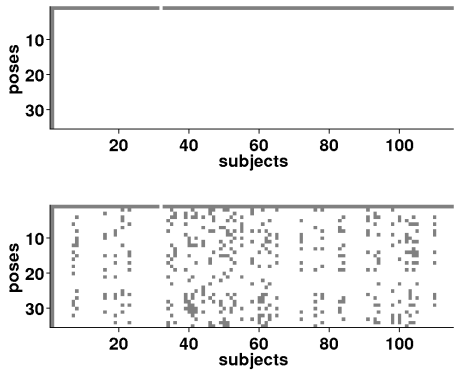

Figure 3: *Every dot marks a scan taken from a subject in a given pose.* ***Top****: SCAPE-like approach - only one subject is scanned in different poses* ***Bottom****: Our model - all subjects are scanned in at least 10 poses. Although grey dots are not taken into account during training a SCAPE-like set of scans is included.*

easily. Furthermore, sex, age, and self-declared fitness level are noted. Likewise, a number of measures are captured with a commercially available impedance spectroscopy body fat scale, namely weight, body fat percentage, percentage of muscle tissue, water content, and bone weight. We also use a medical grade pulse oximeter to capture the oxygenation of the subjects' hemoglobin and their pulse. The scans are arranged in a database $\mathbf{S}_{s,p}$, where s is the subject identifier and p the pose, containing the points and their respective normals as generated by the scanner. This database is available to the scientific public [XXX08].

In order to create a unified model of all the captured data, the scans have to be brought into semantic correspondence. The simplest and most common way to achieve this is to fit a single template model to every scan (cf. [BV99, ACP03, SMT04, ASK*05]).

We create a symmetric template by triangulating a 3D scan, enforcing symmetry by hand and manually fitting a skeleton into the model. Skinning is performed using the technique by Baran and Popović [BP07]. The template t can be parameterized by the parameter vector $\mathbf{Y}$ consisting of n points $\mathbf{P}^t = \mathbf{p}_1^t, \cdots, \mathbf{p}_n^t$ and l triangles $\mathbf{T}^t = \mathbf{t}_1^t, \cdots, \mathbf{t}_l^t$.

2.2. Skeleton Based Pose Estimation

The skeleton based fitting employs an approach commonly used in marker-less motion capture systems [BM98]. Any rigid body motion can be modeled as a single rotation around a chosen axis followed by a suitable translation. Together, the transformation can be stored as a twist ξ with 6 degrees of freedom. See Murray et al. [MSZ94] for mathematical properties and a more in depth description. The deformation

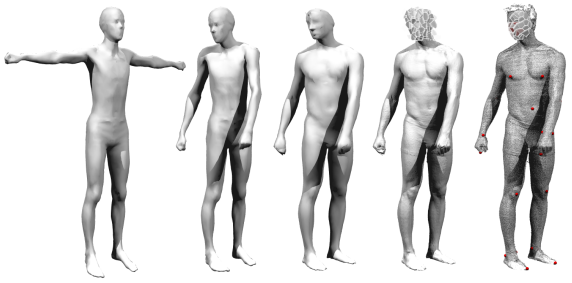

Figure 4: *The registration pipeline.* ***Left to right:*** *The template model, the result after pose fitting, the result after non-rigid registration, transfer of captured surface details, and the original scan annotated with manually selected landmarks are shown.*

of the template model is additionally governed by a kinematic chain with k joints which arises from the embedded skeleton. Only simple revolute joints are considered, which can be parameterized by a single angle γ_i. By parameterizing the complete pose of a person as a vector $\Xi = [\xi, \gamma_1 \dots \gamma_k]$ we can easily generate a linear system of constraint equations to optimize the pose. An ICP style optimisation scheme similar to Bregler et al. [BMP04] is used that generates up to three constraint equations per point of the template surface. Additionally, the manually selected landmark coordinates are used to ensure global stability of the fitting process. The results from this step are used on the one hand as training data to learn regression functions that aim to modify the pose of a subject. On the other hand, the extracted pose can be used to initialize the non-rigid registration technique described in the following.

2.3. Non-Rigid Registration

The posed template from Section 2.2 is used as initialization for a more detailed non-rigid registration step, that captures the remaining details of the current scan. The procedure follows the ideas presented in the work by Allen et al. [ACP03] and Amberg et al. [ARV07]. The registration is expressed as a set of 3×4 affine transformation matrices $\mathbf{T}_i$ associated with each vertex $\mathbf{p}_i^t$ of the posed template, which are organized in a single $4n \times 3$ matrix

$$\mathbf{X} = [\mathbf{T} \dots \mathbf{T}_n]^\top \tag{1}$$

We define the cost function $E(\mathbf{X})$ of our deformation as a combination of three energy terms: $E_d(\mathbf{X})$, which penalizes distance between template and target surface, $E_s(\mathbf{X})$, which acts as a regularization term to generate a smooth deformation, and finally $E_l(\mathbf{X})$, which is a simple landmark distance term,

$$E(\mathbf{X}) = \alpha E_d(\mathbf{X}) + \beta E_s(\mathbf{X}) + \gamma E_l(\mathbf{X}). \tag{2}$$

Unlike Amberg et al., we do not express the distance term using the closest point of the target surface from the template vertices, but as a projection onto a plane fitted to a local patch of the target. Amberg et al. are forced to handle vertices bordering on holes in the target surface specially to avoid artefacts. This is not necessary if the inverse procedure is used and allows us to skip the border labeling step.

To find this projection, we first find the closest mesh vertex $\mathbf{p}_i^t$ for each target surface point $\mathbf{p}_j^s$. We discard matches where the angle between the respective normals are above a threshold $\varepsilon_n = 30^\circ$ or the distance between the points is bigger than $\varepsilon_d = 50\,\text{mm}$. We now go through all mesh vertices $\mathbf{p}_i^t$ and gather the set P_i of all points that were matched to it. We then use a least-squares procedure to fit a local plane to them and project the point $\mathbf{p}_i^t$ onto that plane, resulting in a target point $\tilde{\mathbf{p}}_i^t$ unless P_i contains fewer than 4 points. In the latter case the set is discarded and we assign the closest point of the surface $\mathbf{p}_j^s$ as the target position $\tilde{\mathbf{p}}_i^t$ unless this point also fails the requirements given above. Our distance energy term can now simply be expressed as

$$E_d(\mathbf{X}) = \sum \| \tilde{\mathbf{p}}_i^t - \mathbf{T}_i \mathbf{p}_i^t \|^2 \tag{3}$$

The regularization term $E_s(\mathbf{X})$ in Amberg's original paper is expressed as the Frobenius-Norm of the transformation matrices of neighboring vertices in the mesh. This original term does not take into account irregular sampling of the mesh, and thus may exhibit some artifacts. Additionally, in our case missing data needs to be extrapolated only in localized regions where we have holes in the scan. In our experiments we therefore confine the second order differences of the transformation matrices by applying a Laplacian constraint with cotangent edge weights. This regularization term tries to make changes in the transformation matrices over the mesh as smooth as possible, and not as similar as possible as originally proposed. The regularization term can be written as

$$E_s(\mathbf{X}) = \sum \| \sum_j w_{ij} (\mathbf{T}_i - \mathbf{T}_j) \|^F \tag{4}$$

where w_{ij} are the cotangent Laplacian weights based on the original template mesh configuration and $\|\|^F$ denotes the Frobenius Norm.

The final term $E_l(\mathbf{X})$ is a simple landmark distance term:

$$E_l(\mathbf{X}) = \sum \| \tilde{\mathbf{l}}_i^t - \mathbf{T}_i \mathbf{p}_i^t \|^2 \tag{5}$$

The general registration procedure follows the work of Amberg and coworkers. The energy term (2) can be written as a linear system and solved in the least-squares sense for a given configuration. This process is iterated $t_{max} = 1500$ times, during which we change the energy function weighting terms in the following way:

$$\beta = k_0 \cdot e^{\lambda t} \tag{6}$$

where t is the number of the iteration, $k_0 = 3000$, and

$$\lambda = \ln \left(\frac{k_0}{k_\infty} \right) / t_{max} \tag{7}$$

with $k_\infty = 0.01$. Additionally, α can be kept constant at 1 and $\gamma = k_\gamma \cdot \beta$ with $k_\gamma = 2$.

3. Model Representation

It is useful to encode the registered models in such a way that the relevant differences between a pair of scans can easily be extracted. E.g., if the subject raises an arm between two considered scans, we want the representation of the hand of the person to be identical, although both position and rotation of the hand relative to the main body have changed. A common method to achieve this is to embed a skeleton into the model and encode the surface relative to the skeleton [ASK*05]. This method, however, is forced to resort to interpolation when a given surface point cannot be assigned uniquely to a single bone of the skeleton. Additionally, the correlation between pose and body shape becomes harder to capture. So in the following we describe an encoding that allows us to describe both pose and body shape in a unified way. As we show in Section 5.1, this non-linear transformation allows us to work with linear functions for modifying pose or body shape without significant loss of accuracy.

Translational invariance can easily be achieved by using variational approaches, as for example introduced by Yu et al. [YZX*04] or Sorkine et al. [SCOL*04]. The mesh can be reconstructed given only the original connectivity and the triangle gradients by solving a sparse linear Poisson system. We can also edit and modify the shape by 'exploding' it, applying an arbitrary transformation to each triangle separately. If we then solve the so modified linear system, we effectively stitch the triangles back together in such a way that the prescribed triangles transformations are maintained as good as possible. This fundamental idea has been used for shape editing (as for example in [ZRKS05]), but also forms the basis of SCAPE [ASK*05].

Unfortunately, this variational representation is not rotational-invariant, meaning that the same shape will be encoded differently depending on its orientation. To remedy this, we encode each triangle as a transformation $\mathbf{U}_i$ relative to a rest-pose triangle $\mathbf{t}_i$. This transformation can be split up into a rotation $\mathbf{R}_i$ and a remaining stretching deformation $\mathbf{S}_i$ using polar-decomposition. The stretching deformation is by construction rotation-invariant, which means that we only need to construct a relative encoding for the rotation matrices $\mathbf{R}_i$. This can be achieved by storing relative rotations between pairs of neighboring triangles, i.e.

$$\mathbf{R}_{i,j} = \mathbf{R}_i \cdot \mathbf{R}_j^{-1} \tag{8}$$

where i and j are neighboring triangles. So, for every triangle three relative rotations connecting it with its neighbors can be generated. This encoding may seem wasteful as three rotations are stored instead of just one but this redundancy significantly improves the stability of the reconstruction when a deformation is applied to the encoded model. Recently,

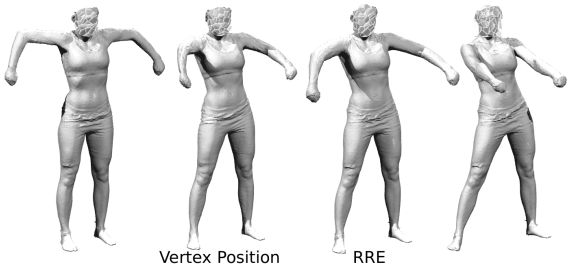

Figure 5: *Due to the relative rotation encoding (RRE) direct linear interpolation of two scans (left and right) results in realistic intermediate poses (middle right) whereas linear interpolation of vertex positions fails as can be seen, e.g., in the subject's degenerated right arm (middle left).*

Kircher and Garland [KG08] have introduced a similar representation for editing mesh animations.

Reconstructing a mesh from this encoding involves solving two sparse linear systems. First we need to reconstruct $\mathbf{U}_i = \mathbf{R}_i\mathbf{S}_i$. Then a Poisson reconstruction yields the complete mesh. Creating the per-triangle rotations $\mathbf{R}_i$ from the relative rotations $\mathbf{R}_{i,j}$ requires solving a sparse linear system, which can be created by reordering Equation (8). For every set of neighbouring triangles equations of the form

$$\mathbf{R}_{i,j} \cdot \mathbf{R}_j - \mathbf{R}_i = \mathbf{0} \tag{9}$$

are added to a sparse linear equation system. As long as the model is encoded and decoded without modification, the rotations $\mathbf{R}_i$ we receive from this system are identical to the input up to floating point accuracy and a global rotation $\mathbf{R}_g$. However, if any modification is applied to the encoded model, the resulting matrices $\mathbf{R}_i$ are not necessarily pure rotation matrices but may contain scale or shear components. To improve the stability of the reconstruction we perform matrix ortho-normalization of the resulting $\mathbf{R}_i$ using singular value decomposition.

Now that a reversible procedure for encoding a model in a locally rotation invariant way has been described, we can think about how exactly the different components of the description are represented. The main requirement of the encoding is that linear interpolation leads to intact representations. Shear matrices are already in a suitable format if only one half of the symmetric matrices are stored. Rotation matrices, however, are badly suited for direct interpolation. Evaluation of a number of different encodings leads directly to rotation vectors because this representation allows easy interpolation and unlike quaternions all possible combinations of values are valid and, in contrast to Euler Angles, the encoding does not suffer from gimbal lock [Ple89]. Additionally, in order to further linearise the encoding space, all parameters are stored relative to the corresponding triangle of the mean model, which is constructed by averaging all components of all models in the relative encoding. The re-

sulting final encoding has 15 degrees of freedom per triangle (nine for rotation and six for in-plane deformation).

The advantage of this complex representation is that it is hard to generate inconsistent meshes and that many common deformations, namely scaling during shape morphing and rotation during pose modification, are linear operations. This improves the quality of trained regression functions significantly and allows us to use linear regressors without visible loss of quality. As shown in Figure 5 it is even possible to linearly interpolate between two poses of a subject and obtain realistic results.

Unfortunately, high frequency information, as present for example in the wrinkles of the pants the subjects are wearing, is very hard to represent with our model. Subjects sometimes adjust the fit of the pants between scans and the intra subject variance of wrinkles is even higher. So we opted to use a simple detail transfer procedure to re-add high frequency information after morphing, similar to displacement subdivision surfaces by Lee et al. [LMH00]. This step improves firstly the accuracy of the estimated regression functions as noise is removed, secondly the efficiency since the computationally intensive steps operate on lower quality meshes, and thirdly the visual quality as high frequency information is retained during morphing instead of getting smoothed out. It works as follows: After fitting, the base mesh is subdivided using the simple mid-edge scheme and projected onto the scan. The offsets of the subdivision vertices in normal direction of the base mesh are stored with each triangle. During recall the mesh is subdivided again and the stored offsets are added.

4. Regression

In Section 3 a joint model encoding human pose and shape has been presented. In this section regression is introduced as a powerful tool to incorporate further information such as gender, height, or joint angle into the model to obtain a fully statistical and morphable model. Starting from a set of encoded scans $\mathbf{A} = [\mathbf{a_1} \dots \mathbf{a_i}]$ and a set of labels, attaching a given semantic value $\mathbf{l} = [l_1 \dots l_i]$ to every scan, we compute a gradient direction $\mathbf{s}$ and a corresponding offset o such that

$$\arg\min_{[o\mathbf{s}]} [\mathbf{1}|\mathbf{A}] \begin{bmatrix} o \\ \mathbf{s} \end{bmatrix} - \mathbf{l} \tag{10}$$

This function can simply be computed in a least-squares sense. However, the achievable generalisation is limited. Much better results can be achieved if cross validation is used to reject components of the gradient that primarily contribute noise. The easiest way to achieve this is to compute the PCA of the database and project $\mathbf{A}$ into this space. Then, cross validation is performed to choose the number of components that yield the best generalisation. The classical approach discards components in the order the PCA suggests, starting with the elements corresponding to the smallest eigenvalues. However, this was found to be suboptimal since we are not interested in rejecting noise in the global sense of the spanned space. However, we can observe that components that contribute primarily noise are not correlated strongly with a given semantic gradient. So by first solving Equation (10) and then rejecting components in ascending order in which they contribute, we are able to significantly improve generalisation of the function. A comparison of cross validation accuracy for both methods plus a non-linear Support Vector Regression based technique is shown in Table 1. The accuracy of our filtered approach is very close to the precision of the non-linear approach.

Obviously, morphing a subject $\mathbf{d}$ to conform to a semantic constraint k just involves walking along the function's gradient $\mathbf{d}' = \mathbf{d} + \mathbf{v}$ where $\mathbf{v} = \lambda\mathbf{s}$ for a suitable λ. For several semantic constraints, the solution is slightly more involved since the gradients are not necessarily orthogonal. Basically we want to find a new gradient $\mathbf{z}$, where

$$\frac{\langle \mathbf{z}, \mathbf{v_i} \rangle}{\|\mathbf{v_i}\|} = \|\mathbf{v_i}\|. \tag{11}$$

Equation (11) is used to generate one row for every constraint $\mathbf{v_i}$. Solving the frequently underdeterminded system for $\mathbf{z}$ in the minimum norm sense yields a solution that changes the subject as little as possible.

4.1. Semantic Model Basis

It is interesting to rotate the original PCA basis such that the first vectors correspond to semantically meaningful directions since this allows morphing a subject while keeping some constraints constant. E.g. increasing body height normally results in additional weight. However, by keeping weight constant while increasing height results in slimmer subjects.

The Gram-Schmidt algorithm [GVL96] is used to span the subspace of the original PCA space such that it is orthogonal to all given semantic morphing vectors. Then PCA is used to generate a basis for the remaining subspace. The new basis and all morphing vectors now span the original PCA space. Thereafter, all scans are transformed to the new base. The reconstruction of these models using only the subspace not spanned by semantic variables leads to a representation in which all models are invariant to the semantic constraint variables. The PCA of these reconstructions in combination with the semantic constraint vectors represents the final basis. The desirable properties of this basis are that the first vectors directly represent semantically meaningful gradient directions and the remaining human body shape space is spanned by a PCA basis.

5. Applications and Results

In this section applications of our model are presented. First the foundation of our model is validated by quantitatively

Regressor	Height	Weight	Body Fat	Muscles	Waist
simple linear	3.68	3.03	3.38	5.63	2.78
filtered linear	1.43	1.33	2.02	2.41	0.945
non-linear	1.15	1.17	1.66	2.38	0.858

Table 1: *Root mean squared errors estimated by 10-fold cross validation for different semantic variables and regressors on a model that contains every subject exactly once. Weight is measured in kg, Body Fat, and Muscles in %, and Height and Waist Girth in cm.*

evaluating the accuracy of different semantics based morphing strategies. Then character modeling schemes are described and last our approach for animating characters based on skeletal angles is introduced.

5.1. Semantic Morphing

In line with research conducted by Allen et al. [ACP03] and Seo and Magnenat-Thalmann [SMT04] we present body shape morphing driven by high level semantic variables. We also conduct quantitative analysis of the accuracy of the trained functions. In Table 1 mean squared errors generated by 10-fold cross validation of different semantic functions are presented. It is doubtful that changes within the range of these error bounds would be perceptible. The data also shows that, as a result of the non-linear relative rotation encoding, the achievable accuracy is only slightly better when using non-linear rather than linear regression. This assumption is confirmed by the observations summarised in Table 2. Mean and standard deviation of the angles between morphing directions for selected functions computed for all subjects in the shape only model are shown. All of these values are small indicating that morphing directions are highly collinear independent of the position in shape space. Apparently, the relative rotation encoding linearises the space sufficiently so that it is now admissible to use a linear function to represent the changes.

Unfortunately, many semantically meaningful functions cannot easily be evaluated quantitatively. For example, in order to define a muscledness function the subjects in the database have to be labelled somehow. Yet, it is hard for a human judge to assign a number to the muscledness of a person. It is much easier to compare two given scans and decide on the more muscled subject. Each random pairing of scans defines a gradient direction. The judge only chooses the sign of the gradient towards greater muscularity. By first normalizing and then averaging the gradients, a general muscularity function can be generated. Results of applying the function are shown in Figure 6.

Since morphing functions can operate directly on the relative encoding, it is trivial to constrain morphing to selected body parts. A multiplicative mask allows deformation to occur in selected areas. The reconstruction process spreads out the error arising at the edge of the selected region evenly, preventing the development of steps in the surface. Figure 6 shows selective morphing of the upper body using the muscledness function.

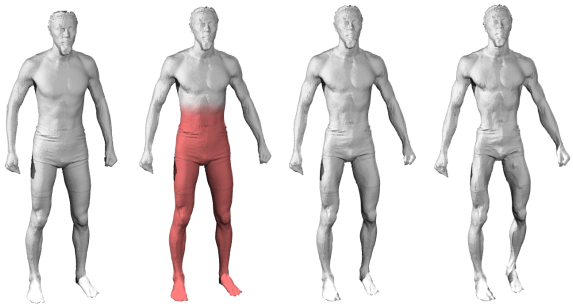

Figure 6: *The effects of applying the muscledness function.* ***Left to right****: Original, a selective mask was applied to increase only upper body muscles, full body muscle augmentation, and an extremely muscular caricature.*

	Height	Weight	Body Fat	Waist Girth
μ	2.04	1.47	2.78	2.00
σ	0.688	0.520	0.853	0.715

Table 2: *Mean and standard deviation of angles (in degrees) between morphing directions computed by the non-linear model for all subjects.*

5.2. Character Generation

As shown recently, it is essential for the perception of the diversity of a crowd that the body shapes of the characters differ significantly [MLD*08]. It is consequently important to have a simple method that allows the generation of diverse body shapes. Yet, it may also be important to be able to tightly control a generated character's body shape. Employ-

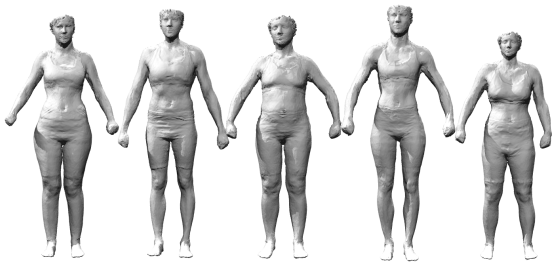

Figure 7: *Several women were randomly generated using the semantic basis. We applied the constraints sex = female and weight = 65 kg. As expected the taller the woman the slimmer she is.*

ing our model both objectives can be achieved by combining two different approaches.

The PCA projects the largest variances of a dataset in the first components while noise like features are displaced to the last components. This is a most welcome feature for many applications. For the purpose of generating a random character that exhibits a unique physique while creating a natural, human look, a PCA based shape-only model is the tool of choice. We use a model consisting only of scans of subjects in the resting pose. The model that contains all scans cannot be used for this purpose since it contains pose dependent components in the first PCA vectors. So, a randomly generated character with the limited model would exhibit not just changes in shape but also in pose. Yet, for animation the generated characters can be plugged directly into the full model as described in Section 5.4.

The advantage of the PCA based technique is that the diversity of the generated characters is very high. On the downside, control over the type of generated character is fairly low. Neither gender nor body weight or height can easily be controlled as the PCA vectors do not, unlike often assumed, directly pertain to a single semantically relevant measure. We can, however, take a given starting point, generated e.g. with the above technique and apply the morphing described in Section 4 to enforce a given set of constraints. Unfortunately, this approach, although workable, may produce suboptimal results since the morphed distance may be large, introducing artifacts on the way.

It is much more efficient to use the semantic basis introduced in Section 4.1. This technique allows us to specify semantic constraints such as height between 1.70 m and 1.90 m and sex between 0.9 and 1.1 male and allow the system to fill in the details. Ultimately, this approach is used to generate the models displayed in Figure 7 whereas Figure 8 shows a crowd that was generated without constraints.

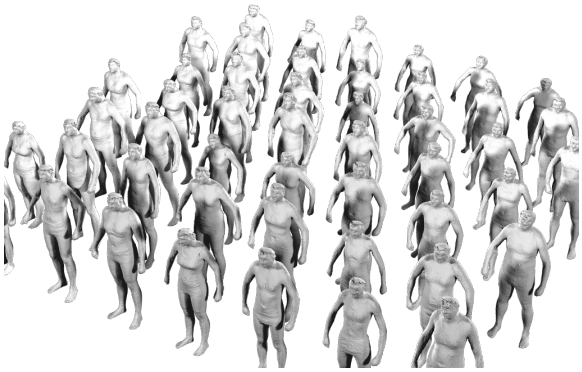

Figure 8: *An example of a randomly generated set of characters.*

Figure 9: *A handle based interface for editing body shapes. The red markers are held in place while the yellow arrows show where attached markers are moved. First the height is increased, then hip width is decreased, and last the crotch is raised.*

5.3. Handle Based Body Shape Modeling

Additional adjustments may be deemed necessary by the responsible artist. Obviously morphing along semantic trajectories is a simple option. In this section a different approach is introduced. By adding moveable handles to the body model a very intuitive way of changing body shape is established. This avenue is opened by our use of Poisson reconstruction in the last step which allows the inclusion of additional positional constraints. The deformation that is required to conform to the constraints is distributed evenly. In fact, mesh editing has been performed using this technique [YZX*04]. However, we want the constraints to influence the body shape realistically instead of just deforming the initial shape in the least squares sense because this inevitably leads to unrealistic distortions. This, however, can easily be mended by projecting the candidate body model into the PCA space spanned by the body shape database since primarily valid body models can be represented in this space. It may thus not be possible to represent the proposed model in shape-space and constraints are not met exactly. So, we iterate between deforming the model and projecting it back into the space of body shapes. After about 10 iterations the system converges. Still, the method acts as a very strong regularizer, so that constraints are frequently not met exactly and it may be necessary to exaggerate them to achieve a desired effect. A simple editing session is shown in Figure 9.

5.4. Animation

Every scan's pose is estimated during registration. Thus we can easily train functions that each change a specific degree of freedom of the pose. That way, we can morph any scan

Figure 10: *By adding positional constraints during reconstruction of deformed models it is possible to ameliorate accumulated pose errors. Here the effect of using positional constraints is demonstrated.* ***Left****: No constraints,* ***Middle****: Constraints,* ***Right****: Reference Pose.*

Figure 11: *Animation result. The subject on the right is crossing his legs. This is significant as no subject in the database has been scanned in a similar pose.*

into a specified pose. Since animations are frequently parameterised by a set of joint angles for every frame, we can simply morph a given model to conform to these constraints to animate it. This works well for most joints and poses.

However, improvements are possible if two minor issues are addressed. Some functions' areas of influence are not localised to the expected area but also include areas on the mirrored side of the body. This is a result of the choice of scanned poses. In most poses arm movements are symmetric. Fortunately, the effect can be compensated for by computing some functions, namely all functions concerning arm movement, only on one side of the body. Similarly, it proved beneficial to split PCA vectors into left and right halfs during reprojection of candidate models to assist independent motion of left and right arms.

Furthermore, absolute positioning accuracy of end effectors can be improved by following Wang et al. [WPP07] who propose to add positional constraints to the tips of limbs during the poisson reconstruction step. This step also serves to correct correlations that were wrongly learned. E.g., in one of the poses the subjects stand on one leg (cf. Fig. 2). In order to keep balance and not to move for the 10 s it takes to perform the scan a very unrelaxed upper body posture was commonly adopted by the subjects. This has led to undesired correlations. Also note that these artifacts cannot be prevented unless fast 3D-scanning of moving subjects is performed, which is not available today at a comparable accuracy. A side-by-side comparison of using end effector constraints vs. not using them is shown in Figure 10. Figure 11 shows several frames from an animation. In one of the frames the subject crosses his legs. This is significant as no pose of a scan in the database is even close to the displayed pose. An animated crowd is displayed in Figure 12.

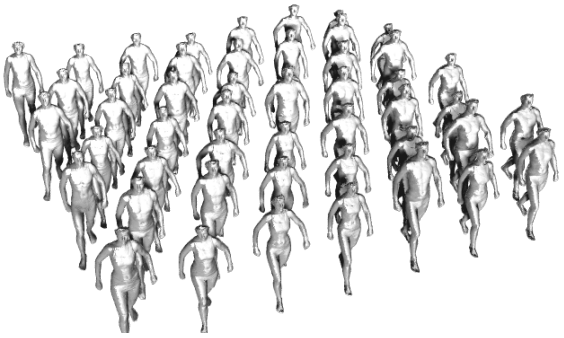

Figure 12: *A walking animation is transfered to 49 character.*

6. Conclusions and Future Work

We describe a model of human pose and body shape. A rotation invariant encoding of the scan database allows us to train semantic regression functions. These in turn can be used to generate and animate human avatars. Since pose and body shape are stored in a single model, correlations between them are automatically exploited to induce realistic muscle bulging and fat deformation during animation.

Comparison of linear and non-linear regression shows that the numerical cross-validation errors are slightly lower for non-linear regression but the visual quality is equivalent. So for display oriented applications linear regression is sufficient.

A designer can perform semantic morphing either on the whole body or by applying a simple mask to selected body parts. This allows the fine grained generation of realistic hu-

man models. Additionally, a handle based body shape editing paradigm provides the artist with an intuitive tool that seamlessly integrates with semantic constraint based editing.

As most of our system is implemented in Matlab the performance is not optimal. Reconstructing one model takes approx. 20 s on a recent machine. However, a more efficient implementation and possibly a hierarchical multi-resolution approach could improve the performance significantly.

It may also be interesting to estimate body shapes from images as this would open a number of applications for example in the human motion capture field or in biometric identification.

References

[ACP03] ALLEN B., CURLESS B., POPOVIĆ Z.: The space of human body shapes: reconstruction and parameterization from range scans. *ACM Transactions on Graphics 22*, 3 (2003), 587–594.

[ACPH06] ALLEN B., CURLESS B., POPOVIĆ Z., HERTZMANN A.: Learning a correlated model of identity and pose-dependent body shape variation for real-time synthesis. In *SCA '06: Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation* (Aire-la-Ville, Switzerland, Switzerland, 2006), Eurographics Association, pp. 147–156.

[ARV07] AMBERG B., ROMDHANI S., VETTER T.: Optimal step nonrigid icp algorithms for surface registration. *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on* (June 2007), 1–8.

[ASK*05] ANGUELOV D., SRINIVASAN P., KOLLER D., THRUN S., RODGERS J., DAVIS J.: Scape: shape completion and animation of people. *ACM Transactions on Graphics 24*, 3 (2005), 408–416.

[BM98] BREGLER C., MALIK J.: Tracking people with twists and exponential maps. In *CVPR '98: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Washington, DC, USA, 1998), IEEE Computer Society, p. 8.

[BMP04] BREGLER C., MALIK J., PULLEN K.: Twist based acquisition and tracking of animal and human kinematics. *International Journal of Computer Vision 56*, 3 (2004), 179–194.

[BP07] BARAN I., POPOVIĆ J.: Automatic rigging and animation of 3d characters. *ACM Transactions on Graphics 26*, 3 (2007), 72.

[BV99] BLANZ V., VETTER T.: A morphable model for the synthesis of 3d faces. In *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1999), ACM Press/Addison-Wesley Publishing Co., pp. 187–194.

[DCKY02] DONG F., CLAPWORTHY G. J., KROKOS M. A., YAO J.: An anatomy-based approach to human muscle modeling and deformation. *IEEE Transactions on Visualization and Computer Graphics 8*, 2 (2002), 154–170.

[GVL96] GOLUB G. H., VAN LOAN C. F.: *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)*. The Johns Hopkins University Press, Oct. 1996.

[KG08] KIRCHER S., GARLAND M.: Free-form motion processing. *ACM Transactions on Graphics 27*, 2 (2008), 1–13.

[LMH00] LEE A., MORETON H., HOPPE H.: Displaced subdivision surfaces. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 2000), ACM Press/Addison-Wesley Publishing Co., pp. 85–94.

[MLD*08] MCDONNELL R., LARKIN M., DOBBYN S., COLLINS S., O'SULLIVAN C.: Clone attack! perception of crowd variety. *ACM Trans. Graph. 27*, 3 (2008), 1–8.

[MSZ94] MURRAY R. M., SASTRY S. S., ZEXIANG L.: *A Mathematical Introduction to Robotic Manipulation*. CRC Press, Inc., Boca Raton, FL, USA, 1994.

[Ple89] PLETINCKX D.: Quaternion calculus as a basic tool in computer graphics. *The Visual Computer 5*, 1 (Jan. 1989), 2–13.

[RDP99] ROBINETTE K., DAANEN H., PAQUET E.: The caesar project: a 3-d surface anthropometry survey. *3-D Digital Imaging and Modeling, 1999. Proceedings. Second International Conference on* (1999), 380–386.

[SCOL*04] SORKINE O., COHEN-OR D., LIPMAN Y., ALEXA M., RÖSSL C., SEIDEL H.-P.: Laplacian surface editing. In *SGP '04: Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing* (New York, NY, USA, 2004), ACM Press, pp. 175–184.

[SMT04] SEO H., MAGNENAT-THALMANN N.: An example-based approach to human body manipulation. *Graphical Models 66*, 1 (January 2004), 1–23.

[SPCM97] SCHEEPERS F., PARENT R. E., CARLSON W. E., MAY S. F.: Anatomy-based modeling of the human musculature. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1997), ACM Press/Addison-Wesley Publishing Co., pp. 163–172.

[SSSB07] SCHERBAUM K., SUNKEL M., SEIDEL H.-P., BLANZ V.: Prediction of individual non-linear aging trajectories of faces. In *The European Association for Computer Graphics, 28th Annual Conference, EUROGRAPHICS 2007* (Prague, Czech Republic, 2007), vol. 26 of *Computer Graphics Forum*, The European Association for Computer Graphics, Blackwell, pp. 285–294.

[WPP07] WANG R. Y., PULLI K., POPOVIĆ J.: Real-time enveloping with rotational regression. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers* (New York, NY, USA, 2007), ACM, p. 73.

[WSLG07] WEBER O., SORKINE O., LIPMAN Y., GOTSMAN C.: Context-aware skeletal shape deformation. *Computer Graphics Forum 26*, 3 (Sept. 2007), 265–274.

[XXX08] web address omitted, Sept. 2008.

[YZX*04] YU Y., ZHOU K., XU D., SHI X., BAO H., GUO B., SHUM H.-Y.: Mesh editing with poisson-based gradient field manipulation. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers* (New York, NY, USA, 2004), ACM, pp. 644–651.

[ZRKS05] ZAYER R., RÖSSL C., KARNI Z., SEIDEL H.-P.: Harmonic guidance for surface deformation. In *Proc. of 26th Annual Eurographics Conference* (Dublin, Ireland, 2005), Alexa M., Marks J., (Eds.), vol. 24 of *Computer Graphics Forum*, Eurographics, Blackwell, pp. 601–609.