Estimating Body Shape of Dressed Humans

Nils Hasler[a], Carsten Stoll[a], Bodo Rosenhahn[b], Thorsten Thormählen[a], Hans-Peter Seidel[a]

[a] *MPI Informatik*
[b] *Hannover University*

Abstract

The paper presents a method to estimate the detailed 3D body shape of a person even if heavy or loose clothing is worn. The approach is based on a space of human shapes, learned from a large database of registered body scans. Together with this database we use as input a 3D scan or model of the person wearing clothes and apply a fitting method, based on ICP (iterated closest point) registration and Laplacian mesh deformation. The statistical model of human body shapes enforces that the model stays within the space of human shapes. The method therefore allows us to compute the most likely shape and pose of the subject, even if it is heavily occluded or body parts are not visible. Several experiments demonstrate the applicability and accuracy of our approach to recover occluded or missing body parts from 3D laser scans.

Key words: Human Modelling, Shape Estimation, Surface Fitting, Statistical Modelling

1. Introduction

Creating convincing 3D models of humans is a difficult task. For many applications, for example automatic dress size measurement and virtual try-ons, but also for virtual stunt men in movie productions, virtual models of real persons have to be created that are as detailed as possible and faithfully represent the true body skin surface. This can be achieved, for example, by 3D scanning a naked human (using e.g. a laser scanner or structured light based systems). While this generates ground truth data, the procedure is unconvenient or even undesirable in many situations and additionally may result in models containing errors and holes. For example, customers in a shop will often not feel comfortable taking off their clothes for 3D scanner based body measurement. Wearing skin-tight apparel is the next best approximation, but still not feasible in many scenarios. In other situations a 3D scanner is not available and thus it is only possible to create approximate geometry using visual hulls and stereo information from several cameras, resulting in very coarse geometry. So, a method for estimating body shape and biometric measures from coarse, noisy, incomplete 3D data is desirable.

To solve these problems we present a system that is capable of estimating the shape of a human body covered or partially covered by clothes given coarse, noisy, hole riddled or even partial 3D geometry. This is achieved using a statistical model of human body shapes and poses, which is similar to work by Anguelov et al. [1], Weber et al. [2], and Wang et al. [3]. Our system takes a 3D scan or 3D model as input, which can be created, e.g., with full body 3D laser scanners, multi-view stereo methods, or structured light scanners.

The approach works by fitting the statistical model [4] to the recorded data with an iterative approach, while maintaining that the resulting estimation stays in the space of body shapes spanned by the model. This allows us to estimate the body shape of subjects wearing wide and obstructive apparel. While the generated model is a plausible representation of the subject's body, it is, depending on the clothes, not an exact match but rather a best estimate based on what we can perceive. Even for humans it is difficult to guess the body shape of persons wearing, for example, a long coat. Some biometric measures on the other hand, like height, leg length or arm length, can be calculated relatively accurately though.

Please note that our method can not actually "see" through clothing, unlike for example the controversial backscatter X-Ray machines deployed at some airports today. We believe that the privacy of the subject is consequently not invaded by our technique. Nevertheless, care should be taken when employing the technology.

Balan and Black [5] have recently presented a system based on the SCAPE model [1] which allows them to estimate the body shape of dressed persons given a number of multi-view images or video sequences. The subjects are allowed to wear arbitrary clothes but have to be captured in a number of different poses or in a longer animation sequence. Their approach also relies on a color based segmentation of the scans into skin and dressed parts, which is used to apply differently weighted error functions in the segmented regions. In contrast, our method is designed to work without a segmentation and from a single input frame. While our input contains more information than a single multi-view input image, significantly more information about the shape of a person can be extracted from several such multi-view frames when pose and clothing are varied.

In the motion capture community several researchers have developed methods to deal with wide clothing. Rosenhahn et al. [6] described a system that allows them to track loosely dressed persons in multi-view video. However, they do require a priori knowledge of both the body geometry and the clothes. Some good tracking results of loosely dressed persons have recently been presented by de Aguiar et al. [7] and Vlasic et al. [8]. Similarly, Starck and Hilton [9] present a system for capturing the performance of actors using multi-view camera systems in studio environments. However, neither paper addresses the underlying body geometry and track only the surface deformation. Balan et al. [10] on the other hand, use a SCAPE based model to track humans. They are thus able to estimate body shape from multi-view video but are restricted to tight-fitting garments.

Our contributions in this paper are the following:

- We present a method for estimating the body shape of a dressed person from a 3D scan or 3D mesh model. The approach is robust to severe contortion of the surface caused for example by loose clothing, hair, noise corruption, or large holes in the data.
- We demonstrate that the method can be applied to partial scan registration and estimation of biometric parameters.

The rest of the paper is structured as follows: Section 2 describes the model of human body shapes we use, Section 3 details the fitting procedure, experimental results and evaluations are presented in Section 4, and a summary concludes the paper in Section 5.

2. Human Body Space

Our model of human body shapes is based on a database of approx. 550 3D scans of 114 subjects. All subjects are scanned in a base pose, and some subjects are additionally scanned in 9 poses chosen randomly from a set of 34 poses that were selected to span the range of motion of average humans. In order to achieve semantic correspondence between the scans they are registered with a non-rigid registration technique similar to [11].

2.1. Model Representation

It is not desirable to generate a statistical model directly from the scans because non-linear transformations caused by the inherent skeleton cannot be captured easily by a linear statistical model. Previous solutions to this problem embed a skeleton into the model and store vertex positions relative to the associated bones [1]. Instead, we opt to use a surface encoding of the models that is invariant to both, translation and rotation [12, 4]. That way, the local transformations expected to occur when describing the space of human body shapes (scaling) and poses (local rotations) can be described by a linear model. A similar encoding using vertices instead of triangles has been presented by Lipman et al. [13].

Translational invariance can be achieved by employing variational methods (see [14] for an overview). Reconstruction involves solving a sparse linear system to recover vertex positions from their relative encoding. Rotational invariance is more difficult to achieve. We accomplish this by decoupling rotation and stretching of the model triangles. First, all triangles $\mathbf{t}_i$ are represented relative to the corresponding triangles $\mathbf{r}_i$ of a reference model, i.e., $\mathbf{t}_i = \mathbf{T}_i \cdot \mathbf{r}_i$. The matrix $\mathbf{T}_i$ can be factorized into a rotation $\mathbf{R}_i$ and a stretch/shear part $\mathbf{S}_i$ using polar decomposition [15]. Finally, only the relative rotation of a triangle to its neighbors

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

is stored. During reconstruction, given a rotation for one of the triangles, the actual rotation matrices can be recovered by solving a linear system. For increased stability a re-orthonormalization step of the rotation matrices can be added.

Concatenating the rotations represented by rotation vectors and the components of the stretch matrices, yields a high dimensional representation of human bodies that can be approximated linearly with respect to the most common deformations occurring in the combined body shape and pose space. Running principal component analysis (PCA) on the set of 3D scans yields a matrix of eigenvectors $\mathbf{E}$, describing the combined body shape and pose space and a set of low dimensional descriptors $\mathbf{s}$ of a scan $\mathbf{m}$ such that

$$\mathbf{m} = \mathbf{E} \cdot \mathbf{s} + \mathbf{a}, \tag{2}$$

where $\mathbf{m}$ is a model in the relative rotation encoding and $\mathbf{a}$ the average model. Every eigenvector of $\mathbf{E}$ corresponds to properties of the encoded human with different scales of influence on the body shape. However, if an unknown body shape is to be represented in the human body shape space a least squares system needs to be solved

$$\arg\min_{\mathbf{s}} \ (\mathbf{m} - \mathbf{E} \cdot \mathbf{s} + \mathbf{a})^2. \tag{3}$$

In this naïve representation the influence of eigenvectors corresponding to small eigenvalues is overemphasized. This problem can be alleviated by dividing each eigenvector $\mathbf{e}_i$ by its eigenvalue e_i, yielding a matrix $\mathbf{W}$ of whitened coefficients (see [16]). In this new representation, every scaled eigenvector has the desired influence. Projecting a 3D model into the space of human shapes is equivalent to

$$\mathbf{s} = \mathbf{W}^{+} \cdot (\mathbf{m} - \mathbf{a}), \tag{4}$$

where $\mathbf{W}^{+}$ is the pseudo-inverse of $\mathbf{W}$. As a result of whitening the coefficients, the least-squares solution of Equation (4) results in a model $\mathbf{m}$ that is as close to the average human in a space that evenly describes all human traits as possible.

3. Fitting

Fitting a human model $\mathcal{M}$ to a 3D scan or model $\mathcal{S}$ is done with an iterative approach as illustrated in Figure 1. We start with a sparse set of user specified correspondence points. Marking feet, hands, ellbows, and head is usually sufficient. We then iterate three steps until convergence. In the first step $\mathcal{M}$ is aligned rigidly to $\mathcal{S}$ by finding the set of closest points from $\mathcal{M}$ to $\mathcal{S}$ and minimizing the squared distance. Next, the matches are used to drive a least-squares Laplacian deformation, moving $\mathcal{M}$ closer to $\mathcal{S}$. As this action normally moves the model out of the space spanned by the statistical model of human bodies, we finally project $\mathcal{M}$ back into the human body shape space. In the following we describe the three main steps in more detail.

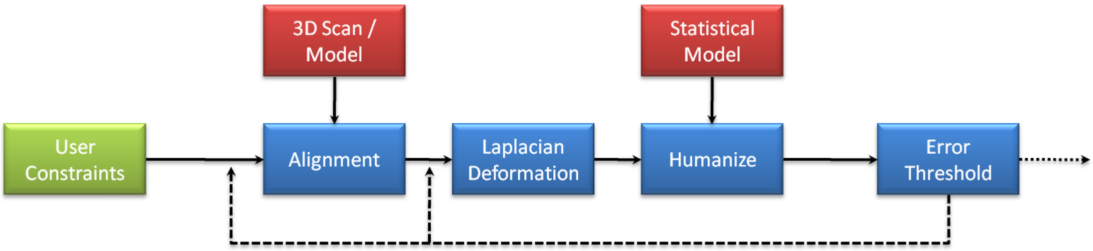

Figure 1: Overview of the fitting algorithm

3.1. Alignment

For every point of $\mathcal{M}$ the closest point on $\mathcal{S}$ is computed. Matches are dropped if the distance is too big (> 10 cm) or normal directions of source and target deviate too strongly (> 30°). The remaining matches are stored in a list C. Then, the optimal rigid body motion is calculated by minimizing the squared distances of the matches in C. Then, $\mathcal{M}$ is transformed accordingly. This procedure is iterated until the mean residual error q of matches converges. The ICP procedure is necessary because the best alignment can only be computed for a given set of matches. It is possible, and in our experience quite likely, that given the configuration after a single alignment step, a better set of matches can be found because the objects are now aligned more closely.

3.2. Laplacian Deformation

Next, $\mathcal{M}$ is deformed using a simple linear least-squares Laplacian mesh deformation [17]. Specifically, the following energy is minimized:

$$\arg\min_{\mathbf{x}} \ (\mathbf{Lx} - \mathbf{d})^2 + (\mathbf{Cx} - \mathbf{c})^2, \tag{5}$$

where $\mathbf{L}$ and $\mathbf{d}$ are a Laplacian system with cotangent weights, and $\mathbf{C}$ and $\mathbf{c}$ represent the constraints C computed in the previous step weighted by the importance function $W(i)$. If the person in the target scan is wearing tight fitting clothes, generating uniform weights is sufficient to produce convincing results. In case of wider and more obstructive clothes though this scheme fails. One main observation leading to an improved weighting function is that the human body always lies either exactly *on* the target surface or *beneath* it. Thus it is important to weight matches that constrain vertices, which lie on the outside (as determined by normal direction) of the target surface stronger than those which lie on the inside of the target. In case of a given segmentation of $\mathcal{S}$ computed with prior knowledge, for example the skin color detection [5] or garment detection employing a model of the clothes as in [18], we could further modify the importance function to reflect this information.

3.3. Humanization

Once the mesh has been deformed with the given constraints, it needs to be projected back into the space of human body shapes defined by the statistical model described in Section 2 because we are not interested in just fitting the surface of the scan but to find the human body shape that best fits the scan. In principle this step projects the unconstrained solution back onto the solution manifold.

This is achieved by transforming the current model $\mathcal{M}$ into the relative rotation encoding $\mathbf{m}$. The model is then projected into the space of human body shapes using Equation (4). Since $\mathbf{W}^{+}$ can be precomputed, this step reduces to a matrix vector multiplication. The result of which is a closest fit of $\mathcal{M}$ in the space of human body shapes. Due to the limited dimension of the shape descriptor $\mathbf{s}$, shapes that are not human body shapes cannot be represented easily. Reconstructing $\mathcal{M}'$ from $\mathbf{s}$ using Equation (1) with subsequent Poisson reconstruction [19] yields the humanized model in Euclidean space.

3.4. Error Evaluation

Given the humanized mesh, we can calculate the mean length q' of the matches generated in the alignment step. If q' is lower than q we continue with the alignment step, otherwise the Laplacian Deformation is repeated with reduced weights for the matches. After a fixed number of iterations (10) with $q' > q$ the algorithm terminates.

As for all ICP methods, the initial configuration has to be fairly close to the solution for the approach to converge. This issue can be avoided if a few markers are placed manually on the target surface. Then, to generate the initial configuration the same algorithm is run with these fixed matches.

4. Experiments

In this section several experiments are evaluated. First, the hidden body geometry of several persons is estimated and biometric measures are extracted and compared to the true values. Then, registration bootstrapping, a technique for improving the quality of scan registration and increasing the size of the scan database, is demonstrated. Last, the same technique is applied to shape estimation from partially missing and severely noise corrupted data. A full statistical analysis, however, is not provided here and remains as future work.

The experimental setup is conceivably simple. A full body 3D scanner is used to scan the subjects, which takes about 10 seconds. Then, using a custom tool markers are selected (2 min). Finally, the proposed approach is run. The runtime of our Matlab implementation of the algorithm is in the order of minutes per scan. E.g. for the scan shown in figure 3 b, marker

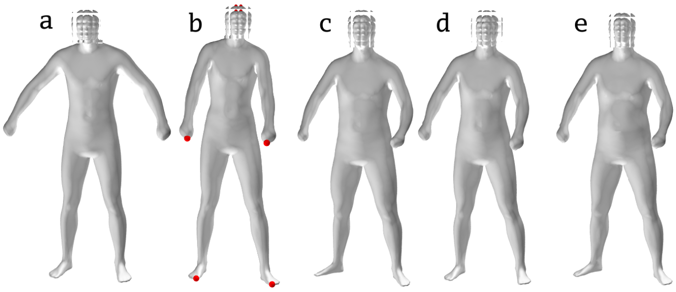

Figure 2: The optimization procedure is based on iterative deformation of the model using Laplacian editing followed by a projection of the result onto the solution manifold spanned by the model of human body shapes. This figure shows some of the steps performed during the optimization. In **a** the initial mesh (the average man) and in **b** the result of the marker based fitting is shown. **c** and **d** display a Laplace deformation step followed by a humanization at different points in the optimization. Lastly, the final humanized result is displayed in **e**.

based registration takes 3.5 min (20 iterations) and the surface registration 8 min (22 iterations).

4.1. Hidden Body Geometry

We evaluate our technique for estimating hidden body geometry, on the one hand, by showing overlays of the scan with the estimated geometry and on the other, by extracting biometric measures from the estimations. Overlays of resulting estimations are shown in Figure 3. As can be seen, the estimated body shapes are highly plausible and fit well into the overlaid 3D scans. For a scan as shown in Figure 4 it is extremely difficult to estimate measures such as dress size or body weight of the Santa impersonator because the thick coat generates an ambiguous situation that is even difficult for humans to resolve. However, some measures, such as the length of arms and legs and his total height, can be recovered quite well (cf. Table 1). Note that the input data for the algorithm does not have to be generated by a 3D scanner. Structured light scanners or multi-view stereo techniques provide sufficient 3D geometry.

The progression of the optimization is displayed in Figure 2. Starting from the average man, markers are used to get an initial pose estimate. Please note that the body shape of this initial estimate (Fig. 2 b) differs significantly from the final result (Fig. 2 e). This indicates that the markers merely stabilize the optimization but do not contribute considerably to the shape estimate. The initial estimate is taller, thinner and uses outstretched instead of slightly bent arms to reach the hand markers.

Estimating biometric measures given a 3D model of a human is a difficult problem. Two of the simplest and most prominent solutions include computing the measures directly on the estimated 3D model or employing the statistical model of human body shapes to learn functions that compute the desired measures. Some measures, such as weight, cannot be computed directly from a given mesh of a human. However, even for length measurements that can easily be computed on a mesh surface, we found that fitting a filtered linear function to the statistical model achieves better results [4]. The measures summarized in Table 1 are consequently computed by training a linear function on the statistical human body shape model. For dressed humans, a weight factor of 10 between matches of vertices lying outside vs. inside the target surface was found to be optimal.

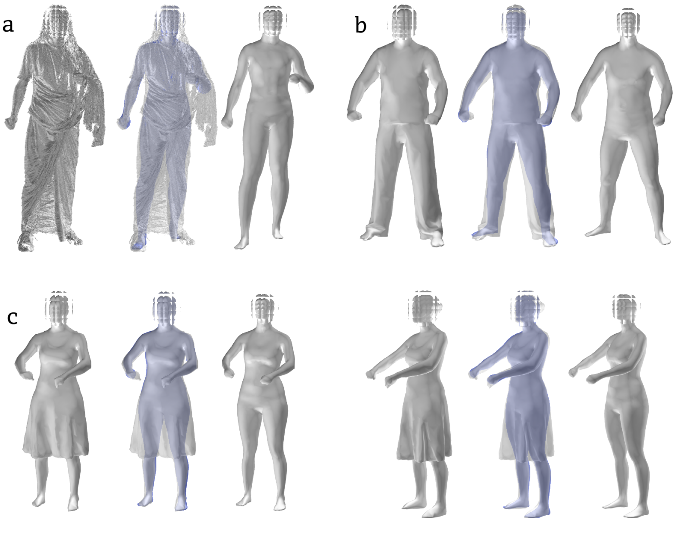

Figure 3: Body shape estimation can be performed given 3D data generated by a laser scanner (a) or by silhouette based multi camera systems (b and c).

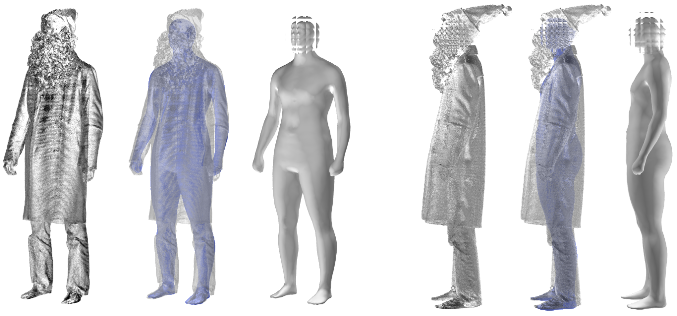

Figure 4: Even for humans, it is impossible to accurately estimate the body shape of a person wearing heavy clothing. So, the estimation of some biometric measures can only be achieved with limited accuracy. Generally, lengths, e.g. arm length or body height can still be estimated but circumference based measures (waist girth, weight, etc.) are obscured.

4.2. Registration Bootstrapping

Given, for example, a scan, as shown in Figure 5 a, we apply the surface registration procedure described in [4]. First, a skeleton based pose estimation is performed (Fig. 5 b) to generate a starting point for the subsequent non-rigid surface registration shown in Figure 5 c. Unfortunately, due to the large difference in body shape of template and scan a registration error occurs in the left armpit. This is a result of the skeleton based initial pose estimation which does not model different body shapes. So the starting configuration for the non-rigid surface registration step may be quite far away from the target surface, if the body shape is significantly different. This is not much of a problem in smoothly varying areas, such as the chest, but in unfortunate circumstances, the mesh can self-intersect or creases may develop. These problems can be alleviated, if a better initial guess can be generated (cf. Fig. 5 d). Our approach is

	Height	Arm Length	Leg Length	Weight	Waist Girth
Santa	178	61	77	76	75
full	178	60	77	71	75
partial	181	57	79	64	75
ground truth	182	57	79	63	74
Toga	175	61	80	69	73
ground truth	179	59	84	67	74

Table 1: Biometric measures of one person: First dressed as a **Santa** impersonator (Fig. 4), and second wearing every day clothing (Fig. 7) using the **full** and a **partial** scan, as well as manually acquired ground truth values for comparison. Additionally, biometric measures of the subject wearing a **toga** shown in Figure 3 a are compared to ground truth. Lengths are measured in cm and weight in kg.

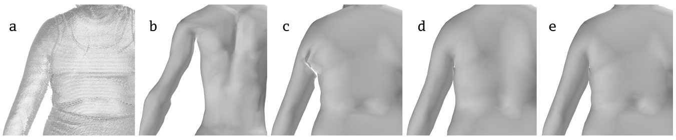

Figure 5: The quality of registration can be improved by using the statistical model to generate an improved starting point for the surface fitting step. The figure shows an input scan (a), a skeleton based initial guess (b), the surface fitting result using that initial guess (c), the initial guess generated by our system (d), and the final surface fitting result.

similar to the bootstrapping for facial scans described by Blanz et al. [20]. We perform the same fitting algorithm as described above for naked/tightly dressed scans with uniform constraint weights. Figure 5 e shows the improvements when bootstrapping is applied.

This procedure can not only be applied to increase the database size but to improve the quality of scans already part of the database. In fact, the model shown in Figure 5 was already part of the database.

It may seem surprising, that for a model which is already in the database, the gradient descent based fitting procedure does not arrive at the exact representation of the scan. Since fitting starts from the average model and registration errors, as shown in Figure 5 c, can be considered outliers, the gradient descent based registration technique is unable to find that specific minimum of the cost function.

An example of employing the bootstrapping procedure to a scan that is not in the database is shown in Figure 6. The initial guess models the scan very well already. So the surface fitting step is only required to fill in minor details instead of being responsible for performing major alignment of template and target surface.

4.3. Scan Completion

In a similar vein, it is possible to estimate body shape from incomplete scans as present, for example, when structured light or range scanners are used. In Figure 7 a a laser scan acquired from a single direction is shown in Figure 7 a. In comparison the full scan and the resulting body shapes are shown in Figure 7 b. The two reconstructions are very similar as also evidenced in Table 1.

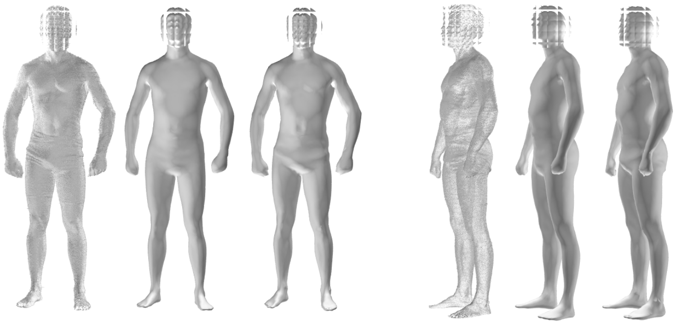

Figure 6: Using a bootstrapping technique, it is easily possible to increase the size of the scan database. Here, the input scan, the initial, model based estimate, and the final surface fit are shown. Since the initial estimate is already very close to the scan surface, a high quality semantic registration can be reached.

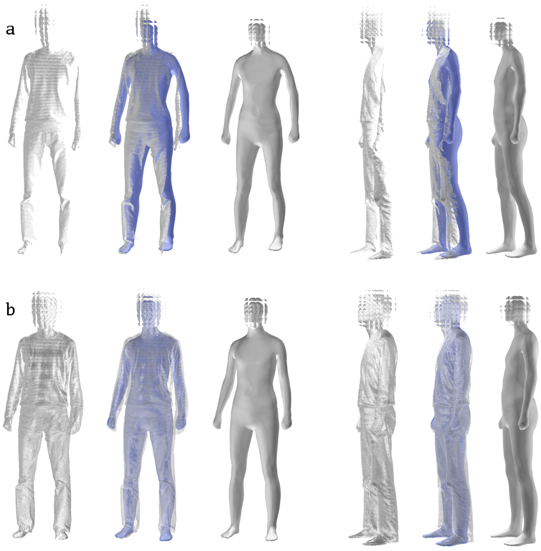

Figure 7: The model is fitted to a single view 3D scan of a subject wearing every day garments (**top**) as well as to the corresponding full multi-view 3D scan (**bottom**).

4.4. Noise

Robustness to noise is an important property for any algorithm working on real world input data. In Figure 8 Gaussian noise is added to a 3D scan. Then, the unmodified algorithm described in Section 3 is run on the data. The result looks plausible and the pose is only slightly misestimated.

5. Summary

The paper deals with the estimation of a detailed 3D body shape given a 3D scan or 3D model of a person wearing heavy or loose clothing. An ICP based Laplacian mesh deformation approach is driven towards a given point cloud or 3D model. The key is that by leveraging the statistical model of human shapes, it is possible to enforce the humanness of the resulting body shape. This strong prior allows us to predict the shape of humans from partial or noisy 3D scans. We have presented several experiments to demonstrate the applicability and accuracy of the approach to recover occluded or missing body parts

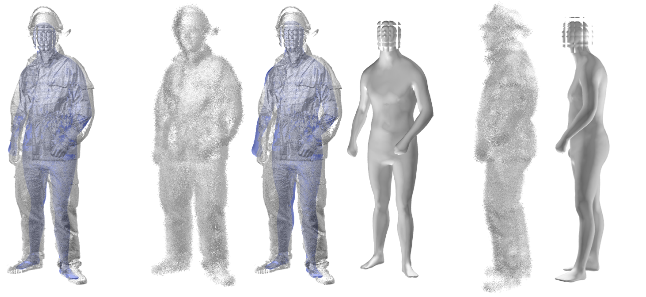

Figure 8: The approach is robust even to severe noise corruption. Here we analyze the body of a fully dressed emergency medical technician (EMT). On the left the uncorrupted scan overlaid with a model fit to that scan is shown. The rest of the figure shows the noise corrupted scan, the result overlaid with the original scan, and the result on its own.

from 3D scans or models. In addition, we have shown that the estimation of biometric measures is possible even under unfavorable conditions.

References

[1] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, J. Davis, Scape: shape completion and animation of people, ACM Trans. on Graphics 24 (3) (2005) 408–416.

[2] O. Weber, O. Sorkine, Y. Lipman, C. Gotsman, Context-aware skeletal shape deformation, Computer Graphics Forum 26 (3) (2007) 265–274.

[3] R. Y. Wang, K. Pulli, J. Popović, Real-time enveloping with rotational regression, in: ACM SIGGRAPH Papers, ACM Press, New York, NY, USA, 2007, p. 73. doi:http://doi.acm.org/10.1145/1275808.1276468.

[4] anonymous, provided as supplementary material (2009).

[5] A. O. Balan, M. J. Black, The naked truth: Estimating body shape under clothing, in: D. A. Forsyth, P. H. S. Torr, A. Zisserman (Eds.), Proc. ECCV, Vol. 5303 of Lecture Notes in Computer Science, Springer-Verlag, Marseille, France, 2008, pp. 15–29.

[6] B. Rosenhahn, U. Kersting, K. Powell, R. Klette, G. Klette, H.-P. Seidel, A system for articulated tracking incorporating a clothing model, Machine Vision and Applicationsdoi:10.1007/s00138-006-0046-y.

[7] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, S. Thrun, Performance capture from sparse multi-view video, Vol. 27, ACM Press, 2008.

[8] D. Vlasic, I. Baran, W. Matusik, J. Popović, Articulated mesh animation from multi-view silhouettes, ACM Trans. on Graphics 27 (3) (2008) 1–9.

[9] J. Starck, A. Hilton, Surface capture for performance-based animation, IEEE Comput. Graph. Appl. 27 (3) (2007) 21–31. doi:http://dx.doi.org/10.1109/MCG.2007.68.

[10] A. Balan, L. Sigal, M. Black, J. Davis, H. Haussecker, Detailed human shape and pose from images, Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on (2007) 1–8doi:10.1109/CVPR.2007.383340.

[11] B. Amberg, S. Romdhani, T. Vetter, Optimal step nonrigid icp algorithms for surface registration, Proc. CVPR (2007) 1–8.

[12] S. Kircher, M. Garland, Free-form motion processing, ACM Trans. on Graphics 27 (2) (2008) 1–13.

[13] Y. Lipman, O. Sorkine, D. Levin, D. Cohen-Or, Linear rotation-invariant coordinates for meshes, in: ACM SIGGRAPH Papers, ACM Press, 2005, pp. 479–487.

[14] M. Botsch, O. Sorkine, On linear variational surface deformation methods, IEEE Transactions on Visualization and Computer Graphics 14 (1) (2008) 213–230.

[15] G. J. Murphy, C-*-Algebras and Operator Theory, Academic Press, New York, 1990.

[16] R. O. Duda, P. E. Hart, D. G. Stork, Pattern Classification (2nd Edition), Wiley-Interscience, 2000.

[17] M. Alexa, Differential coordinates for local mesh morphing and deformation, The Visual Computer 19 (2–3) (2003) 105–114.

[18] N. Hasler, B. Rosenhahn, H.-P. Seidel, Reverse engineering garments, in: A. Gagalowicz, W. Philips (Eds.), Mirage, Springer-Verlag, Rocquencourt, France, 2007, pp. 200–211.

[19] Y. Yu, K. Zhou, D. Xu, X. Shi, H. Bao, B. Guo, H.-Y. Shum, Mesh editing with poisson-based gradient field manipulation, in: ACM SIGGRAPH Papers, ACM Press, New York, NY, USA, 2004, pp. 644–651. doi:http://doi.acm.org/10.1145/1186562.1015774.

[20] V. Blanz, T. Vetter, A morphable model for the synthesis of 3d faces, in: ACM SIGGRAPH Papers, ACM Press, New York, NY, USA, 1999, pp. 187–194.