

A Data-Driven Shape Model for Human Body Reconstruction from Photos

Hyewon Seo
CGAL, ChungNam National University
DaeJeon, Korea
hseo@cs.cnu.ac.kr

Young In Yeo
VR Lab., KAIST
DaeJeon, Korea
yyiguy@vr.kaist.ac.kr

Kwangyun Wohn
VR Lab., KAIST
DaeJeon, Korea
wohn@vr.kaist.ac.kr

Abstract

In this paper, we present a data-driven approach to the problem of reconstructing human body models from 2D images. One of the key tasks in reconstructing the 3D model from image data is shape recovery, a task done until now in utterly geometric way, in the domain of human body modeling. In contrast, we use data-driven deformable models that are obtained in a priori, where a collection of range scans of real human body is structured and statically processed. We use a sparse set of feature points and silhouette data both on the input images and on the template model to optimize the deformation parameters, such that the resulting model best matches the given silhouette. In the presence of ambiguity either from the noise or missing views, our technique has a bias towards representing as much as possible the previously acquired shape. We then generate texture coordinates by projecting the modified template model onto the front and back images. Our technique has shown to reconstruct successfully human body models from minimum number of orthogonal images.

Keywords: Image-based reconstruction, silhouette extraction, human body modeling, range scan data, principal component analysis.

1. Introduction

Engaging real people (whether in their static form or in motion) allows to convincingly depict people in digital worlds. Indeed, the problem of digital human body modeling from measurement is being actively and successfully addressed by image-based and hybrid techniques. During its formative years, researchers have focused on developing methods for modeling appearance and movements of real people observed either from 2D photos or video sequences [5][6][8][11]. These efforts traditionally use silhouette

information from multi-view images for determining the shape and the texture.

Today, range scans are becoming more and more available and hence much of the focus of graphics research has been shifted to the acquisition of human body models from 3D range scans [1][2][13]. The measurements acquired from such scanning devices provide rich set of shape information which otherwise requires considerable amount of time and effort by experienced CG software users. Whole body scanners however remain by far more expensive, difficult to use, and provides limited accessibility in comparison to 2D imaging devices.

In this paper, we propose a system for reconstructing human body model from minimum number of multi-view images, exploiting the quality shape captured by range scans. The distinguishing aspect of our modeler in comparison to existing image-based body modeler is that it employs a data-driven deformable body model that has been constructed from range scans of real bodies [12]. We exploit the quality shape as well as statistical information collected from laser-scanned body database, in the presence of ambiguity, noise, or underconstraints caused by missing views.

We aim at an on-line clothing store [4] in which users can try on garment items on their ‘virtual’ replica. Hence we limit our focus to reconstructing lightly clothed subjects.

2. Related work

In this section we give a brief review of some of the most significant works, which relates mainly to image-based human body modeling. While image-based model reconstruction has been at the center of digital human modeling across several research groups, the majority of research progress in this avenue falls into the category of facial modeling [7]. This is perhaps

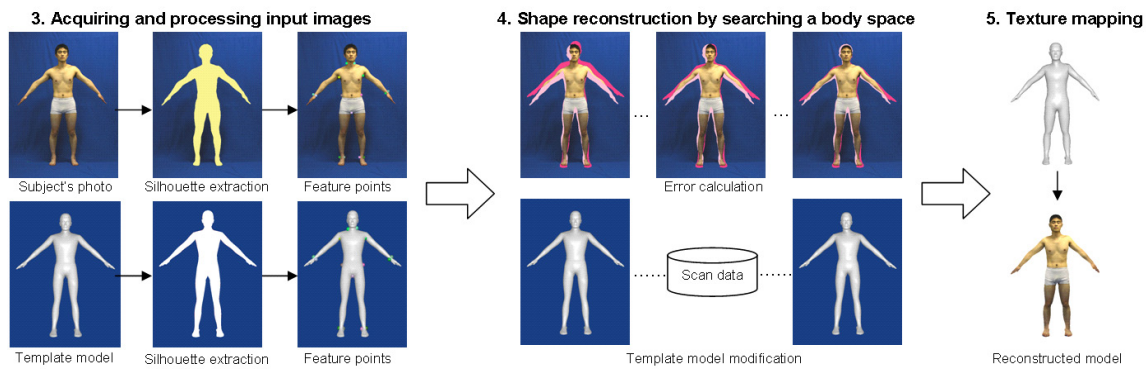


Figure 1: Overview of our modeler

primarily due to the complex articulated structure and high degree of self-occlusion exhibited in our bodies.

One approach that has been extensively investigated is model-based techniques. Lee et al [8] and Hilton et al [6][5] have gathered silhouette observations from multi-view images such that they are used to transform a template humanoid model. Affine transformation has been followed by geometric deformation of the prior surface model. They use feature point locations along the silhouette to find the correspondence among different views and to generate consistent texture coordinates. More recently, Sand et al [11] use multi-view recordings to derive the skeleton configuration of a moving subject, which subsequently derives the skin surface shape. These works show how a prior knowledge can be used to avoid difficulties of general reconstruction. However, they do not accumulate observations which can efficiently be used to handle uncertainties.

The strength of gathering information from collective observation has been illustrated in face model acquisition by Blanz and Vetter [3]. They described a face modeler in which a prior knowledge collected from a set of head scan data is exploited to find the optimizing surface and texture parameters that best fit the given image data. While these methods are quite powerful, they have not been applied to image-based reconstruction of an entire human body.

These considerations lead us to look for a more robust approach to image-based human body modeling. Similarly to their work, our technique adopts a model parameterization scheme based on a collection of observations of real bodies. Also adopted is a surface optimization framework, in order to match multiple camera images. Our technique

however handles complex articulated structure of the entire human body, and still runs at an arguably interactive speed.

3. Acquiring and processing input images

3.1. Taking photographs and virtual camera setup

We take three photographs, each from the front, the side, and the back of the subject. Using a single camera, they are temporally distinct views. We assume that the subjects are lightly clothed. To simplify the combinatorial complexity of the human shape and posture, we require the subject to stand in the specific posture; the limbs are straight and apart from the torso as shown in **Figure 2(a)**. In this paper, the images were captured by the Nikon coolpix 5000 camera. To facilitate automatic silhouette detection, we take photos in front of a blue backdrop.

We now describe the camera arrangements and projection matrix we use for projecting the template model onto the image space. The main idea is to simulate virtual camera as closely as possible to the physical setup. This allows us to use input images directly for the silhouette comparison without additional process such as normalization. In our experiments, we assume that the camera lens and the projection plane are parallel and the lens distortion is small enough to be disregarded. Other physical camera parameters are calibrated as the camera is 5m distant from the screen and 1.2m from the ground, with the normalized focal length 49.1mm in a standard 35mm film camera. From the following equation we obtain the fovy of 39.2° , which

determines the perspective projection matrix used for projecting the template model:

$$Fovy = 2 \times \tan^{-1} \left(\frac{35}{2 \times focal_length} \right)$$

3.2. Silhouette extraction

Several different methods can be used to extract silhouettes from photos. The method we use is a standard background subtraction to isolate silhouettes from images using a color key. Among several color models, we use the Hue-Saturation-Value (HSV) color model. With color keys in H and S, shadows have been successfully labeled as background and thus we could obtain fairly good silhouette extraction as shown in *Figure 2(b)*.

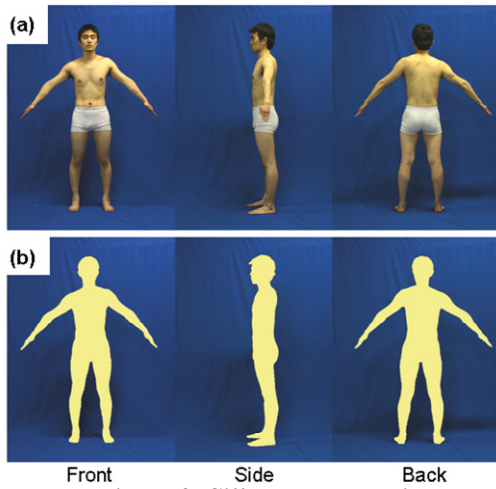


Figure 2: Silhouette extraction

4. Shape reconstruction by searching a body space

In recent years Seo and Magnenat-Thalmann have introduced a parameterized body modeler that makes use of range scans of real human bodies as examples [13]. At the center of that modeler is a representation, which is also adopted here. In that modeler, each of the scanned body shape is represented by combining two shape vectors, which respectively encode the skeleton-driven deformation (SDD) and vertex displacement of a template model that is necessary to reproduce its shape. By collecting these shape vectors from multiple subjects, we obtain what we call the body space.

Building such body space permits us to easily and efficiently explore the coefficients, thus giving us means to change shapes via searching. Given a set of images, we reconstruct the shape by estimating the

coefficients via an optimization procedure. The key point is that instances of the models are deformed in ways that are found in the example set.

4.1. Parameterization of body shape

The parameterization method we use is based on the previous work [13]. Given a set of example body shapes represented as vectors, we apply principal component analysis (PCA) to both joint and displacement vectors. The result is two linear models for the two components:

$$\mathbf{j} = \bar{\mathbf{j}} + \mathbf{P}_j \mathbf{b}_j, \quad \mathbf{d} = \bar{\mathbf{d}} + \mathbf{P}_d \mathbf{b}_d,$$

where $\bar{\mathbf{j}}$ and $\bar{\mathbf{d}}$ are the mean vector, \mathbf{P}_j and \mathbf{P}_d are sets of orthogonal modes of variation, and \mathbf{b}_d and \mathbf{b}_j are the sets of parameters. The appearance of any body models can thus be represented by the set of PC coefficients of joint vector \mathbf{b}_j and that of displacement vector \mathbf{b}_d . A new model can be synthesized for a given pair \mathbf{b}_j , \mathbf{b}_d by deforming the template from vector $\bar{\mathbf{j}}$ and adding the vertex displacement using the map described by \mathbf{d} .

Note that the PCA has the additional benefit that the dimension of the vectors can drastically be reduced without losing the quality of shape. Upon finding the orthogonal basis, the original data vector \mathbf{v} of dimension n can be represented by the projection of itself onto the first M ($\ll n$) eigen vectors that correspond to the M largest eigen values. In this paper, we have used 30 bases both for the \mathbf{b}_d and \mathbf{b}_j . Thus, each body is represented as a set of parameter vector consisting of 30 PC's for the joints and 30 for displacement, giving a total of 60 parameters for the body shape space.

4.2. Silhouette comparison as error metric

One important step in our modeler is to measure the silhouette matching error between the segmented images to projections of the template under deformation. We consider two error terms: The first one is the distance between corresponding feature points. The second one is the silhouette error.

4.2.1. Distance between corresponding feature points.

The first criterion of a good match is the distance between corresponding feature points in the image space (*Figure 3(a)*). We define a distance error term E_d as the sum of the squared distances between each feature point's

location in the data image and its location on the projected image of the template mesh:

$$E_d = \sum_{i=1}^n \text{dist}(P(F_{T,i}), F_{D,i})^2$$

where n is the number of feature points and dist is the Euclidean distance among two pixels in the image, $F_{D,i}$ is the i -th feature pixel in the image, $F_{T,i}$ the corresponding i -th feature point on the template model, and $P:R^3 \rightarrow R^2$ describes the perspective projection of the template mesh to the 2D images.

We consider a sparse set of feature points that are important for calculating joint configurations (scale, and rotation of each joint except for the root that has translation) of the template model. We have found that feature points around the neck, armpits, wrists, crotch and ankles are particularly important, as they undergo relatively high degree of transformation for a matching. Note that they overlap pretty much with anthropometric landmarks as well. 27 points were manually placed on the template mesh prior to projection to the images. On the images, 15 and 12 feature points were defined on the front and side photos, respectively.

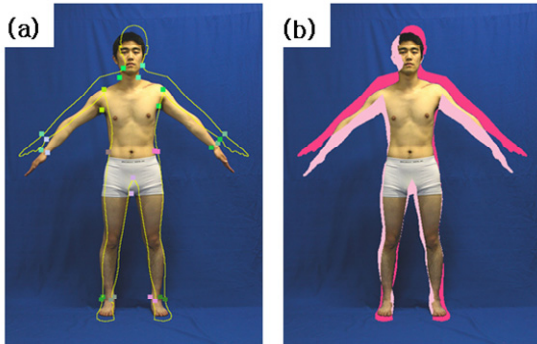


Figure 3: (a) Distance between corresponding feature points (b) Silhouette error

4.2.2. Silhouette error

Using the distances among each feature point location will not result in a successful matching, because even though corresponding feature points are in the same position, actual body shapes can be different from each other. To acquire detailed match of the template model to the image, we define a silhouette error and denote as E_a . By silhouette error we refer to the fraction of pixels for which the projected and observed silhouettes do not match, as shown in **Figure 3(b)**. The number of background pixels that lie inside the projected

template model are summed up with that of foreground pixels that lie outside of it:

$$E_a = \frac{\sum (T(i, j) \cdot \bar{D}(i, j))}{\sum T(i, j)} + \frac{\sum (\bar{T}(i, j) \cdot D(i, j))}{\sum D(i, j)}$$

$T(i, j)$ and $\bar{T}(i, j)$ are the boolean values indicating if the pixel at location (i, j) is inside and outside of the template model, respectively. $D(i, j)$ and $\bar{D}(i, j)$ are 1 if the pixel located at (i, j) is foreground and background, respectively. This notion of non-overlapping area is effectively equivalent to the silhouette error used by Sand et al [11]. Note that the information on arms is taken only from the front/back view.

4.2.3. Combining the error

In this work, we use weighted sum of the two error terms defined above:

$$E = \alpha E_d + (1 - \alpha) E_a$$

At the first iterations we need to quickly search for joint parameters, hence we set $\alpha = 1$. Feature points from both the frontal and the side images are measured. Next, we further improve the fitting accuracy by setting $\alpha = 0.3$. The deformable model is first fit to the frontal image and then side image error is added. Finally, the displacement map is explored with the same setting. At each iteration, we combine the errors from the frontal image and the side image, so that the fitting of the template to frontal and side images can simultaneously be handled.

4.3. Deformation by searching

The next step is the geometry acquisition process, which adjusts the PC coefficients \mathbf{b}_j and \mathbf{b}_d of the template model so that the resulting surface produces a silhouette as close as possible to that of the input image. The template model is first registered with the images using a sparse set of manually defined feature points.

We then find the solution in a coarse-to-fine manner. Since the deformation is parameterized with PCA space for each of the vector components, we first find the optimizing joint parameter \mathbf{b}_j , followed by the subsequent search for the \mathbf{b}_d in the displacement vector space. Our optimization technique is based on a direction set method [10], and repeats search-deform-compare loop until the resulting shape best fits the extracted silhouette – The algorithm generates a body shape from the current coefficients, projects the body model

onto 2D space, and updates the coefficients according to the silhouette difference. The first set of iterations is performed by optimizing only the first coefficients controlling the first PC. In subsequent iterations, more and more PCs are added.

4.4. Shape deformation by skinning update

In recent years the most common technique for the character skin deformation is the skeleton driven deformation (SDD) technique [9]. As the character undergoes animation, the vertex positions comprising the skin mesh are calculated based on the weighted combination of the joint matrices. At the dress pose, each vertex is registered with its local offset, which is defined by its relevant position to each of the local frame of its influencing joint. When the pose changes, the deformed position of a vertex v is determined by interpolating its rigidly transformed local offsets.

Once the surface has been modified through the displacement, the local offsets need to be adapted accordingly, in order to maintain appropriate skin deformation. This is achieved by readjusting the local offsets with the displacement.

In *Figure 4*, a vertex v has two local offsets, p_1 defined on joint J_1 and p_2 on J_2 , and weight values w_1 and w_2 . This can be denoted as follows:

$$v = \sum_{i=0}^k w_i J_i(p_i)$$

where k is the number of influencing joints for a vertex v , $J_i(p_i)$ represents the transformation from the local offset p_i to the world coordinate system. The set of vertices S on the body shape is modified to S' and we newly compute the position v' from the following equation.

$$v' = \sum_{i=0}^k w_i J_i(p_i) + \Delta v = \sum_{i=0}^k w_i J_i(p_i + \Delta p_i)$$

Skinning update is made by adding Δp_i to each local offset p_i . The weight values remain identical, because they determine how much each joint influences the vertex and are usually defined by designers.

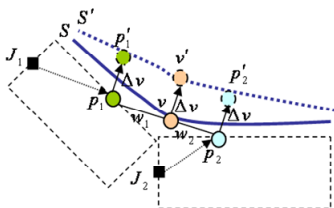


Figure 4: The skinning update on a vertex v

5. Texture mapping

We generate texture coordinate according to the angle between vertex normal and the view direction. If this angle is between $-\pi/2$ and $\pi/2$, we project the vertex on the deformed template surface onto the front image. The other vertices are projected on to the back image. Although we require consistent poses through each view, they may be slightly different from one view to another. Therefore, prior to the second projection to the back image, we adjust the posture of the model so that it matches the silhouette data on the back image.

6. Experimental results

Whole body scans recorded with laser scanners (TecmathTM and CyberwearTM) were used in our experiments. The dataset was originally captured for made-to-measure garment retails. Analogous to those subjects for the image-based reconstruction, these subjects for the three dimensional scan were lightly clothed without accessories, and were in a standing posture with arms and legs slightly apart.



Figure 5: Reconstructed models from photos

We have applied our modeler to a variety of example images. Some of the models generated from our modeler are shown in *Figure 5*. In all examples, we matched the template model built from the first 30 joint and 30 displacement principal components that were derived from the whole body scan dataset as described in the previous section. Once the silhouettes have been extracted, the whole matching procedure was performed in less than 1000 iterations for each principal component. On a Pentium 4 processor, computation time was 5~6 minutes.

Our models take advantage of the template model and reuse the initial skinning with a simple update (Section 3.4). They have been successfully integrated and reanimated in an animation system, without any additional process. *Figure 6* shows one of our models undergoing animation.



Figure 6: Reconstructed models reanimated

7. Conclusion and future work

In this paper we presented a technique for reconstructing human body model from a limited number of 2D images. Using the three-dimensional body space that has been generated from processing range scans, we propose reconstructing the 3D surface in a different manner than existing modelers. For the shape recovery we start with a deformable template model whose deformation is parameterized with PCA of the scanned body shapes. Given a set of images, the optimizing shape is found by searching the shape space, such that it minimizes the matching error measured between silhouettes. The idea is to start from a space consisting of a few PCs and to increase its size by progressively adding new PCs. This provides us powerful means of matching the template model to the image in a coarse-to-fine manner. In addition, a high level of detail and accuracy is acquired, since our modeller essentially blends multiple shapes of the human body acquired from 3D laser scanners. This constitutes a good complement to geometric methods, which cannot capture detailed shape solely from image input.

Our system is intended for off-line reconstruction of geometry, but it is reasonably efficient and can also be adopted for on-line applications. Additionally, we have

successfully integrated our models in animation systems. We are currently exploring an extension of this technique so that arbitrary view and posture can also be handled. Reconstruction of casually dressed subjects is certainly one way of extending our modeler.

Acknowledgements

This research was supported in part by BK21 project of EECS Department of KAIST, MMRC project funded by SK Telecom, Wearable Ubiquitous Computer (WUC) project funded by Institute of Information Technology Assessment and in part by ChungNam National University.

References

- [1] B. Allen, B. Curless, Z. Popovic, "Articulated body deformation from range scan data", *Proc. ACM SIGGRAPH*, Addison-Wesley, San Antonio, USA, 2002, pp.612–619.
- [2] B. Allen, B. Curless, Z. Popovic. "The space of human body shapes: reconstruction and parameterization from range scans", *Proc. SIGGRAPH '03*, pp.587–594, Addison-Wesley, 2003.
- [3] B. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces", *Proc. SIGGRAPH '99*, Addison-Wesley, pp. 187–194, 1999.
- [4] F. Cordier, H. Seo., N. Magnenat-Thalmann, "Made-to-Measure Technologies for Online Clothing Store", pp. 38–48, *IEEE CG&A special issue on Web Graphics*, January 2003.
- [5] A. Hilton, Beresford,D., Gentils,T., Smith,R.J., Sun,W. and Illingworth,J., "Whole-body modelling of people from multi-view images to populate virtual worlds", *Visual Computer: International Journal of Computer Graphics*, 16(7), pp. 411–436, 2000.
- [6] A. Hilton, J. Starck and G. Collins, "3D Shape Capture to Animated Models", *Proc. First International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT)*, IEEE Press, Padova, Italy, 2002, pp. 246–255.
- [7] Y. Lee, D. Terzopoulos, and K. Waters, "Realistic Modeling for Facial Animation", *Proc. ACM SIGGRAPH'95*, pp. 55–62, 1995.
- [8] W. Lee, J. Gu, and N. Magnenat-Thalmann, "Generating Animatable 3D Virtual Humans from Photographs", *Computer Graphics Forum*, vol. 19, no. 3, *Proc. Eurographics 2000 Interlaken, Switzerland*, August, pp. 1–10, 2000.
- [9] J.P. Lewis, M. Corder, N. Fong, "Pose Space Deformations: A Unified Approach to Shape Interpolation and Skeleton-Driven Deformation", *Proc. SIGGRAPH '00*, Addison-Wesley, pp. 165–172, 2000.
- [10] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C, The art of scientific computing*. Cambridge University Press, 1988.
- [11] P. Sand, L. McMillan, J. Popovic, "Continuous Capture of Skin Deformation", *Proc. ACM SIGGRAPH 2003*, pp.578–586.
- [12] H. Seo, "Parameterized Human Body Modeling", PhD thesis, Departement d'informatique, University of Geneva, 2004.
- [13] H. Seo, N. Magnenat-Thalmann, "An Automatic Modeling of Human Bodies from Sizing Parameters", *ACM SIGGRAPH Symposium on Interactive 3D Graphics*, ACM Press, pp. 19–26, 2003.