Computers & Graphics 33 (2009) 211–216

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Technical Section

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, Germany
[b] Hannover University, Germany

ARTICLE INFO

Article history:
Received 8 December 2008
Received in revised form
2 March 2009
Accepted 3 March 2009

Keywords:
Human modelling
Shape estimation
Surface fitting
Statistical modelling

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.

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.

Note that our method cannot 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

* Corresponding author. Tel.: +49 681 9325 420.
E-mail address: hasler@mpi-inf.mpg.de (N. Hasler).

0097-8493/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cag.2009.03.026

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 nine 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

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

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

3.1. Alignment

For every point of $\mathscr{M}$ the closest point on $\mathscr{S}$ is computed. Matches are dropped if the distance is too big (>10 cm) or normal directions of source and target deviate too strongly ($>30^\circ$). The remaining matches are stored in a list $\mathscr{C}$. Then, the optimal rigid body motion is calculated by minimizing the squared distances of