Computer Vision and Image Understanding 127 (2014) 31–42

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Estimation of human body shape and posture under clothing ☆

Stefanie Wuhrer [a,*], Leonid Pishchulin [b], Alan Brunton [c], Chang Shu [d], Jochen Lang [e]

[a] Cluster of Excellence MMCI, Saarland University, Germany
[b] Max-Planck Institute for Informatics, Campus E14, 66123 Saarbrücken, Germany
[c] Fraunhofer Institute for Computer Graphics Research, Fraunhoferstr. 5, 64283 Darmstadt, Germany
[d] National Research Council of Canada, Canada
[e] School of Electrical Engineering and Computer Science, University of Ottawa, Canada

ARTICLE INFO

Article history:
Received 17 December 2013
Accepted 23 June 2014
Available online 1 July 2014

Keywords:
Digital human shape and posture modeling
Statistical shape space
Geometry processing

ABSTRACT

Estimating the body shape and posture of a dressed human subject in motion represented as a sequence of (possibly incomplete) 3D meshes is important for virtual change rooms and security. To solve this problem, statistical shape spaces encoding human body shape and posture variations are commonly used to constrain the search space for the shape estimate. In this work, we propose a novel method that uses a posture-invariant shape space to model body shape variation combined with a skeleton-based deformation to model posture variation. Our method can estimate the body shape and posture of both static scans and motion sequences of human body scans with clothing that fits relatively closely to the body. In case of motion sequences, our method takes advantage of motion cues to solve for a single body shape estimate along with a sequence of posture estimates. We apply our approach to both static scans and motion sequences and demonstrate that using our method, higher fitting accuracy is achieved than when using a variant of the popular SCAPE model [2,18] as statistical model.

1. Introduction

The problem of estimating the body shape and posture of a dressed human subject is important for various applications, such as virtual change rooms and security. For instance, in virtual change rooms, a dressed user steps in front of a virtual mirror and the system aims to simulate different types of clothing for this user. To this end, such a system requires an accurate estimate of the body shape and posture of the user.

We present an algorithm to estimate the human body shape and posture under clothing from single or multiple 3D input frames that are corrupted by noise and missing data. Our approach assumes that the clothing fits to the body and may fail for loose clothing, such as skirts or wide dresses. When multiple 3D frames of the same human subject are recorded in different postures, these observations provide important cues about the body shape of the subject. The clothing may be more or less loosely draped around a particular body part in different postures, which allows for improved shape estimates based on postures where the clothing is close to the body shape. To utilize these cues, we model body shape independently of body posture, and optimize a single representation of the body shape of the subject along with one pose estimate per frame to fit to a set of input frames. When multiple 3D frames of a subject in motion are recorded with high frame rates, our algorithm takes advantage of the temporal consistency of the acquired data. To reduce the complexity of the problem, our method does not explicitly simulate the clothing, but learns information about likely body shapes using machine learning.

Current solutions to this problem use the SCAPE model [2] to represent the body shape and posture of a human subject. This model represents the body shape in a statistical shape space learned from body scans of multiple subjects acquired in a standard posture and combines this with a representation of body posture learned from body scans of a single subject in multiple postures. A popular variant of SCAPE that performs well in practice is the method by Jain et al. [18] that learns variations in body posture using a skeleton-based deformation. The main disadvantage of these methods is that even when acquiring multiple subjects in standard posture, the postures differ slightly, which leads to a statistical space for body shape that represents a combination of shape and posture changes. Hence, for SCAPE and its variants, shape and posture representations are not properly separated.

To remedy this problem, we propose a method that uses a posture-invariant statistical shape space to model body shape

☆ This paper has been recommended for acceptance by Yaser Sheikh.
* Corresponding author.
E-mail addresses: swuhrer@mmci.uni-saarland.de (S. Wuhrer), leonid@mpi-inf.mpg.de (L. Pishchulin), alan.brunton@igd.fraunhofer.de (A. Brunton), chang.shu@nrc-cnrc.gc.ca (C. Shu), jlang@eecs.uottawa.ca (J. Lang).

http://dx.doi.org/10.1016/j.cviu.2014.06.012
1077-3142/

combined with a skeleton-based deformation to model body posture. Using a posture-invariant statistical shape space for body shape offers the additional advantage that the shape space can be learned based on body scans of multiple subjects acquired in multiple postures, thereby allowing to leverage more of the available training data.

This work makes the following main contributions:

- We present a representation that models human body shape and posture independently. Human body shape is represented by a point in a posture-invariant shape space found using machine learning, and human body posture is represented using skeletal joint angles.
- We present an algorithm to estimate body shape and posture under clothing that fits closely to the body from single or multiple 3D input frames. For multiple input frames, a single representation of body shape is optimized along with a posture estimate per frame to fit to the input frames. This allows to take advantage of important cues about body shape from multiple frames.
- When multiple 3D frames of a subject in motion are recorded with high frame rates, the presented fitting approach is stable as temporal consistency is used for tracking.
- We show experimentally that using our method, higher fitting accuracy is achieved than when using the state of the art variant of SCAPE by Jain et al. [18].

2. Related work

The problem of estimating the body shape and posture of humans occurs in many applications and has been researched extensively in computer vision and computer graphics. Many methods focus on estimating the posture of a subject in an image or a 3D scan *without* aiming to predict the body shape (e.g. [16,4,27]). Other methods aim to track a *given* human shape that may include detailed clothing across a sequence of images or 3D scans in order to capture the acquired motion without using markers (e.g. [8,10,29,9,12]).

In this work, we are interested in estimating both the body shape and posture of *any* human subject represented as a 3D mesh that was acquired while wearing clothing. To achieve this goal, we need a model that can represent different body shapes in different postures. Statistical shape models have been shown to be a suitable representation in this case.

Statistical shape models learn a probability distribution from a database of 3D shapes. To perform statistics on the shapes, the shapes need to be in full correspondence. Allen et al. [1] proposed a method to compute correspondences between human bodies in a standard posture and to learn a shape model using principal component analysis (PCA). This technique has the drawback that small variations in posture are not separated from shape variations. To remedy this, multiple follow-up methods have been proposed. Hasler et al. [14] analyze body shape and posture jointly by performing PCA on a rotation-invariant encoding of the model's triangles. While this method models different postures, it cannot directly be constrained to have a constant body shape and different poses for the same subject captured in multiple postures. With the goal of analyzing body shape independently of posture, Wuhrer et al. [32] propose to perform PCA on a shape representation based on localized Laplace coordinates of the mesh. In this work, we combine this shape space with a skeleton-based deformation model that allows to vary the body posture.

Several methods have been proposed to decorrelate the variations due to body shape and posture changes, which allow to vary body shape and posture independently. The most popular of these models is the SCAPE model [2], which combines a body shape model computed by performing PCA on a population of 3D models captured in a standard posture with a posture model computed by analyzing near-rigid body parts (corresponding to bones) of a single body shape in multiple postures. Chen et al. [7] recently proposed to improve this model by adding multi-linear shape models for each part of the SCAPE model, thereby enabling more realistic deformation behavior near joints of the body. Neophytou and Hilton [23] proposed an alternative statistical model that consists of a shape space learned as PCA space on normalized postures and a pose space that is learned from different subjects in different postures.

Several authors have proposed to use statistical shape models to estimate human body shape and posture under clothing. Most of these methods use the SCAPE model as statistical model. Muendermann et al. [22] proposed a method to track human motion captured using a set of synchronized video streams. The approach samples the human body shape space learned using SCAPE and initializes the body shape of the subject in the video to its closest sample in terms of height and volume. The approach then tracks the pose of the subject using an iterative closest point method, where joints are modeled as soft constraints. Balan and Black [5] used the SCAPE model to estimate the body shape and posture of a dressed subject from a set of input images. The method proceeds by optimizing the shape and posture parameters of the SCAPE model to find a human body that optimally projects to the observed silhouettes. If the same subject is given in multiple poses, the shape of the subject is assumed to be constant across all poses, and the model optimizes one set of shape parameters and several sets of posture parameters to fit the model to the observed input images. Weiss et al. [30] used a similar technique to fit a SCAPE model to a Kinect scan. Zhou et al. [33] used a SCAPE model to modify an input image. They learned a correlation between the SCAPE model parameters and semantic parameters, such as the body weight, which allows them to modify an instance of the SCAPE model to appear to have higher or lower body weight. The approach first optimizes a learned SCAPE model to fit to the input image, changes the shape of the 3D reconstruction of the subject, and modifies the input image, such that the silhouette of the modified subject is close to the projection of the changed 3D shape. Jain et al. [18] extended this approach to allow for the modification of video sequences. They used a slightly modified version of the SCAPE model that does not learn a subject-specific pose deformation of the triangles. Helten et al. [15] proposed a real-time full body tracker based on the Kinect. They first acquire the shape of a subject in a fixed posture using a Kinect, and then track the posture of the subject over time using the modified SCAPE model by Jain et al. [18] while fixing the shape parameters.

A notable exception to using the SCAPE model is the approach by Hasler et al. [13], which uses a rotation-invariant shape space [14] to estimate body shapes under clothing. Recently, Perbet et al. [25] proposed an approach based on localized manifold learning that was shown to lead to accurate body shape estimates. While these methods have been shown to perform well on static scans, they are less suitable to predict body shape and postures from motion sequences as the body shape cannot be controlled independently of posture in these shape spaces.

In this work, we are interested in fitting a single body shape estimate and multiple body posture estimates to a given sequence of scans, which requires a shape space that models variations of body shape and posture independently. The variant of the SCAPE model proposed by Jain et al. [18] is a commonly used state-of-the-art method that has been shown to lead to accurate body shape and posture estimates and that models shape and posture variations independently. We propose a new shape space that combines a posture-invariant statistical shape model with a skeleton-based deformation, and show that this model can fit more