

Research Article

Human Model Adaptation for Multiview Markerless Motion Capture

Dianyong Zhang,^{1,2} Zhenjiang Miao,^{1,2} and Shengyong Chen³

¹ Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China

² Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China

³ College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

Correspondence should be addressed to Dianyong Zhang; 07112064@bjtu.edu.cn

Received 28 November 2012; Accepted 17 January 2013

Academic Editor: Carlo Cattani

Copyright © 2013 Dianyong Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An approach to automatic modeling of individual human bodies using complex shape and pose information. The aim is to address the need for human shape and pose model generation for markerless motion capture. With multi-view markerless motion capture, three-dimensional morphable models are learned from an existing database of registered body scans in different shapes and poses. We estimate the body skeleton and pose parameters from the visual hull mesh reconstructed from multiple human silhouettes. Pose variation of body shapes is implemented by the defined underlying skeleton. The shape parameters are estimated by fitting the morphable model to the silhouettes. It is done relying on extracted silhouettes only. An error function is defined to measure how well the human model fits the input data, and minimize it to get the good estimate result. Further, experiments on some data show the robustness of the method, where the body shape and the initial pose can be obtained automatically.

1. Introduction

In the area of computer vision, researchers have been working for many years to obtain and analyze real 3D information of an object. Lots of new methods have been explored by researchers in the recent years [1–5]. As a special subject, the detection and recovery of human shapes and their 3D poses from images or videos are important problems in computer vision area. There are many potential applications in diverse fields such as motion capture, interactive computer games, industry design, sports or medical purpose, interfaces for human computer interaction (HCI), surveillance, and robotics. Accurate detailed human model can be used for markerless motion capture tracking a subject individual model, which includes information on both body shape and pose. Model-based motion capture is especially suited to markerless motion capture because it can constrain the search space by defining the degrees of freedom of the human skeleton. Initialization of human motion capture always requires the definition of humanoid model approximating the

shape, appearance, kinematic structure, and initial pose of the subject to be tracked.

A lot of model-based pose estimation algorithms are based on minimizing an error function that measures how well the 3D model fits the images. So, the good initialization result is very important. The majority of algorithms for human pose estimation use a manually initialized generic model with limb lengths and shapes which approximate the individual. Automate initialization will improve the quality of tracking result. Accurate initialization is base of the right pose estimation. A limited number of researchers do the automate initialization to improve the tracking result by reconstructing 3D model of the subject from a single or multiple view images. Detailed 3D human shape and pose estimations from multi-view images are still difficult problems that does not exist a satisfactory solution.

In this paper, as illustrated in Figure 1, we use multi-views to recover a detailed 3D representation of the person. Our parametric representation of the body is based on a 3D morphable human model with an underlying predefined

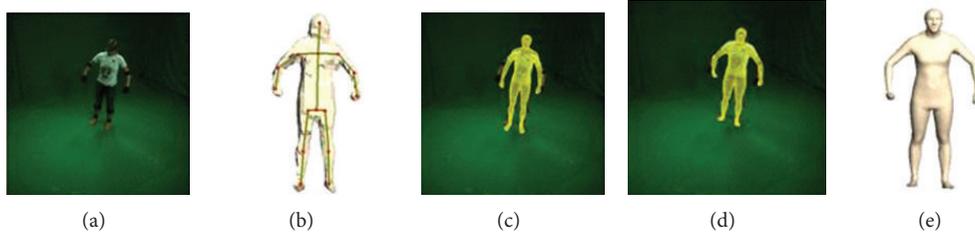


FIGURE 1: Overview of our system workflow: (a) input multi-view images, (b) visual hull mesh from multi images and get the skeleton, (c) original 3D morphable model projected to the image, (d) shape and pose parameters to fit the images, and (e) output estimated model.

skeleton. The models are generated from a template mesh given by a deformable human model that is learned for a scan database [6] over 550 full body 3D scans taken of 114 undressed subjects. Previous contribution [7] estimates human shape and pose from multiple camera views using the popular SCAPE model, but it does not provide an underlying skeleton. We propose a method to automatically adjust a morphable human body model to fit the first frame of the markerless motion capture data. The goal of our job is to obtain a human body model as much adjusted as possible to the individual subject we are going to track. The estimated refined shape and skeleton pose serve as initialization for the next frame to be tracked.

The remaining sections of this paper are organized as follows. In the following section, we will present the relevant previous work on detailed human model estimation. In Section 3, we describe the morphable model with PCA method and the predefined skeleton information. In Section 4, we present the visual hull model and get the initial pose parameters. In Section 5, we present the fitting algorithm in details. In Section 6, we demonstrate the estimation result of our solution. We will conclude this paper in Section 7.

2. Related Works

The papers [9, 10] present comprehensive survey of existing related techniques in motion capture research area of computer vision. A model-based markerless motion capture system can be broken down into four processes: initialization, tracking, pose estimation, and recognition. The initialization step is concerned with two things: the initial pose of a subject and the model representing the subject. Shape and pose initialization can be obtained by manual adaptation or using automatic methods, the latter methods still have some limitations, such as the requirement of a specific pose or predefined motion. The priori model can be of several kinds: kinematic skeleton, shape, and color priors. Many approaches employ kinematic body models, it is hard for them to capture motion, let alone detailed body shape. The majority of algorithms continue to use a manually initialized generic model with limb lengths and shapes which approximate the individual. Because of the difference of vision, the manual initialization often cannot get the well initialization result. Over the past 10 years, there have been substantial research works [11–14] in the automatic initialization of human model

shape from multiple view images. For improved accuracy in tracking, these approaches reconstruct an articulated model which approximates the shape of a specific subject. Because of few images, people cannot get the accurate body shape information, and furthermore, the shape of the subject can differ from person to person.

Our goal is to get the detailed shape and pose body model fast and accurately. Body shape is estimated from a single-view or a multi-view, which is performed by projecting the 3D model onto the images and building a cost function that minimizes the distance between the projected model and the 2D silhouettes. A popular parametric model SCAPE (Shape Completion and Animation for People) [15] is a data-driven method for building body shapes with different poses and individual body shapes. This model has recently been adopted as morphable model to estimate human body shape from monocular or multi-view images [7, 16–19]. Balan et al. [7] have adopted this model closer to observed silhouettes to capture more detailed body deformations.

Most recently, the approach has been used to infer pose and shape from a single image. Guan et al. [17] have considered more visual cues, shading cues, internal edges, and silhouettes to fit the SCAPE model to an uncalibrated single image with the body height constrained. Sigal et al. [16] describe a discriminative model based on the mixture of experts to estimate SCAPE model parameters from monocular and multicameras image silhouettes. Chen et al. [20] proposed a probabilistic generative method that models 3D deformable shape variations and infers 3D shapes from a single silhouette. They use nonlinear optimization to map the 3D shape data into a low-dimensional manifold, expressing shape variations by a few latent variables. Pons-Moll et al. [21] proposed a hybrid tracker approach that combined correspondence based local optimization with five inertial sensors placed at a human body, although they can obtain a much accurate and detailed human tracker, they need additional sensors, and they use the 3D scan model. Lee et al. [22] present a near-automatic method to get a 3D face model by PCA according to the set of input silhouettes. However, for the initialization phase, the pose is known approximately as the subject adopts a specified position in most of papers, such as T position or others. Many approaches still require manual or semiautomatic model positioning for bootstrapping the algorithm.

Benjamin et al. [14] presented a method for generic pose initialization, and it is for the skeleton pose and conclude the

coarse body shape information. Gall et al. [12] introduce an approach for global optimization that is for human motion capturing called as interacting simulated annealing (ISA), they use prior knowledge learned for training motion data as the soft constraint. Many algorithms are based on minimizing an error function that measures how well 3D model fits the image. Our approach is most similar to the work of Jain et al. [23] and Balan et al. [7], but they both estimate pose parameters simultaneously and this will cost more time than estimating pose and shape parameters separately.

Our main contribution is proposing a new approach to automatic generation 3D shape and pose model fitting the multicameras images using a 3D morphable model. In our system, we get the individual human model with the same pose as the subject we are going to track automatically. Human shape and pose are captured by multiple synchronized and calibrated cameras. And we use background subtraction algorithm to extract body silhouettes [24]. The overview of our system is showed in Figure 1.

3. 3D Morphable Model

Principal component analysis (PCA) is a popular statistical method to extract the most salient directions of data variation from large multidimensional datasets. Our morphable model is based on scan database of [6] over 550 full body 3D scans taken of 114 undressed subjects. All subjects are scanned in a based pose, some subjects are scanned in 9 poses chosen randomly from a set of 34 poses. We apply the PCA model to this database [6], a new human shape model can be generated by learning a linear mapping between shape parameters and PCA weights. Therefore, a human model is given by

$$M(\alpha) = m_0 + \sum_{i=1}^n \alpha_i m_i, \quad (1)$$

where the human shape parameter is $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$, m_i is the i th eigen human model, and m_0 is the mean or average human model. Similar to [23, 25], the morphable model is a combination of a bone skeleton with joints. Like Jain et al. [23] and Gall et al. [25], we drive the body pose by a defined underlying skeleton manually, shown in Figure 2. And the shape parameters can be described by PCA parameters. In our paper, we define 20 human PCA components like [23]. We define a kinematic chain, so the motion of body model can be parameterized by the joint angles. For many years, kinematic chains are widely used in human tracking and motion capture systems. The mesh deformation can be controlled by linear blend skinning (LBS) technique. If V_i is the position of vertex i , T_j is the transformation of the bone j , and $\omega_{i,j}$ is the weight of the j bone for vertex i , LBS gives the position of the transformed vertex i as

$$v'_i = \sum_j \omega_{i,j} (T_j v_i). \quad (2)$$

The bone weights for the vertices mean how much each bone transform affects each vertex. These weights are normalized such that $\sum \omega_{i,j} = 1$.

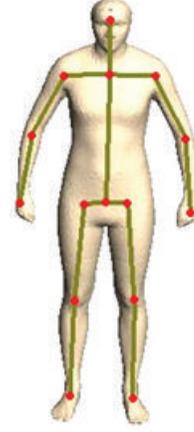


FIGURE 2: The template morphable human model with the underlying skeleton.

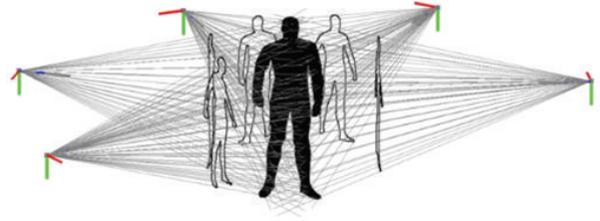


FIGURE 3: The concept of visual hull reconstruction, picture taken from [8].

4. The Visual Hull and Pose Initialization

The pose of a human body model is modelled by a set of joint angles. The pose is computed using the skeleton curve from the visual hull based image silhouettes.

The visual hull is the volume obtained from the interaction of the silhouette cones. The visual hull for the image sequences analyzed is shown in Figure 3. Camera numbers and placements highly influenced the reconstruction quality of the visual hull. 3D volumetric reconstruction from images has long been an important and active research topic. Volumetric approach has been widely used to reconstruct the visual hull from silhouettes for its simplicity and robustness. We build our body model to fit multi-view images, and it just can be used for the markerless motion capture. Not like most of researches, people usually build the body visual hull model using voxel data. We want to get the visual hull mesh model. The visual hull construction process can be shown in Figure 4.

Though the visual hull model from the silhouettes tends to overestimate the real body volume of the subject, we can still compute the skeleton pose using it. We use the Baran's algorithm [26] to rig the visual hull mesh model. We define a template skeleton and adjust the skeleton to fit inside the mesh by resizing and positioning its bones and joints. This adjustment can be treated as an optimization problem, which is to compute the skeleton adjustment that can fit better

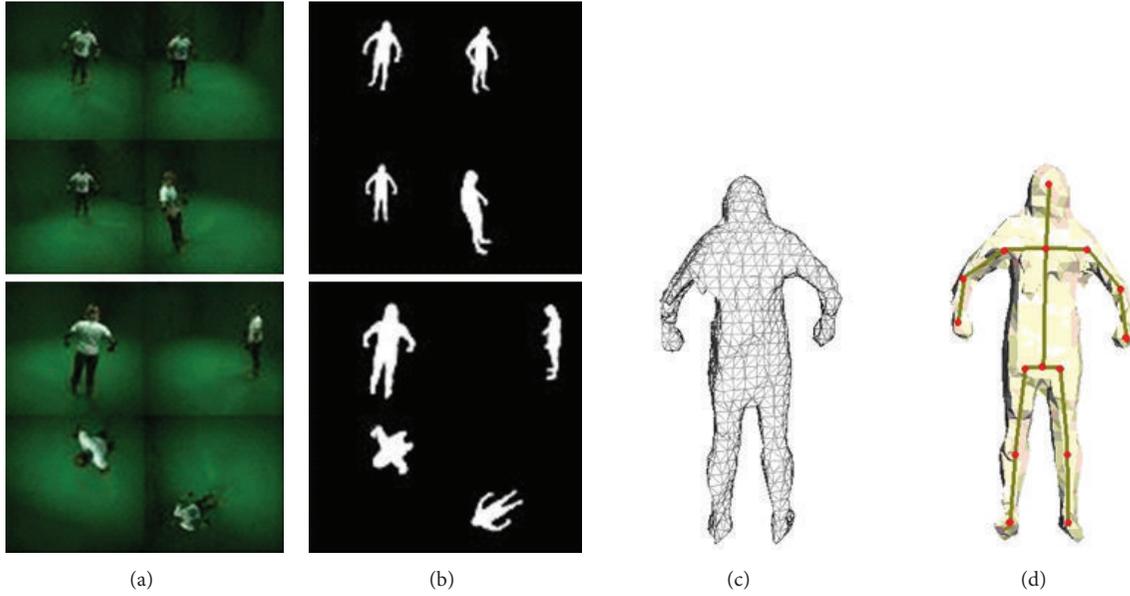


FIGURE 4: Processing images to compute silhouettes and visual hull mesh: (a) input multi-view images, (b) segmentation result, (c) reconstruction visual hull, and (d) extract the skeleton and joints.

inside the mesh while maintaining as much as possible its resemblance with the defined template skeleton.

The recovered position and joint angles of the visual hull can be used to initialize the optimization of the 3D morphable model pose parameters. The number of joints extract from visual hull mesh is 18, we just consider 15 joints not include the foot and waist joints. Like Pishchulin et al. [27], we retarget the skeleton of 3D morphable model to the extracted 3D pose by computing inverse kinematics through minimizing the Euclidean distance between a set of corresponding 3D joint positions: namely, head, neck, thorax, pelvis, left/right knees, ankles, hips, elbows, and wrists. In Section 5, we describe the method of getting the shape parameters in details.

5. Fitting Human Model to Images

In this section, we describe our method for fitting model shape parameters to a set of input silhouette images. This step aims to deform the 3D morphable model to fit the given silhouettes. We assume the human initial pose of the articulated skeleton is known from Section 4. Then, the shape PCA parameters are the only variables that need to be estimated.

5.1. Correspondence between 2D Images and 3D Model. The fitting is fundamental to 2D-3D correspondences. Let $M(\alpha)$ be an individual human model given a vector parameter $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$, the fitting procedure determines the parameter vector α so that the proposed 3D morphable model fits the given 2D silhouettes as well as possible. In the multiple cameras way, each camera has its own coordinates. We should transform the human body model local coordinates $(x, y, \text{ and } z)$ into the image coordinates (x', y') , this can be done by three steps: the first step is to transform the local coordinates into the world coordinates; the second is to

transform the world coordinates into the camera coordinates and the last is to project the camera coordinates, into the image coordinates. The initial location parameters $o = (x_0, y_0, z_0)$, the center of the local human body coordinate in the world coordinates, which are estimated from the centroid of the visual hull generated from multiviews.

Let $S^k, k = 1 \dots K$ be an input silhouette body image captured by camera C^k . $S_m^k(\alpha)$ is a silhouette image generated by projecting $M(\alpha)$ onto an image plane using the camera C^k calibration parameters.

5.2. Shape Deformation. If we define a cost function f that measures the difference of two silhouette images, our goal is to find α that minimizes this total penalty as follows:

$$E_{\text{total}}(\alpha) = \sum_{k=1}^K f(S^k, S_m^k(\alpha)) \quad (3)$$

for a suitable cost function f .

The solution of our problem can be expressed as a minimization of an error function depending on the correspondences shape parameters. We map the surface manifold of the vertex space of the body mesh onto the extract silhouette. And enable those pixels which correspond to contour vertices. We consider the problem of morphable model fitting body silhouette with image silhouettes.

We measure how close a given body shape hypothesis is to the input foreground silhouettes using the distance function between silhouette S and T like papers [7, 19], an asymmetric distance between them in this case,

$$\tilde{d}(S, T) = \frac{\sum_{i,j} (S_{i,j} D_{i,j}(T))}{(\sum_{i,j} S_{i,j})^\delta} \quad (4)$$

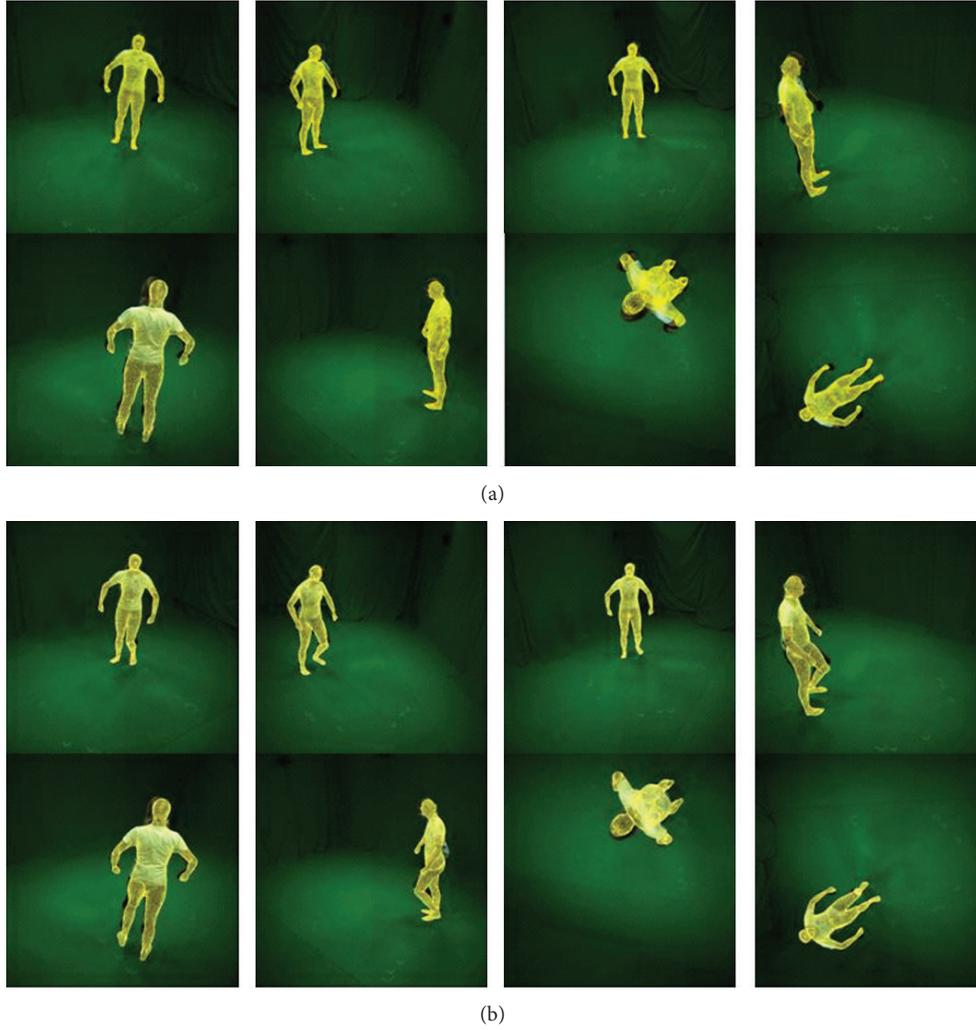


FIGURE 5: Shape estimation results based eight cameras: (a) fitting result of the first frame, (b) fitting result of the 25th frame.

for the numerator, where $S_{i,j} = 1$ for the pixels inside silhouette S otherwise $S_{i,j} = 0$. $D_{i,j}(T)$ is a distance transform function and when the pixel (i, j) inside T , the function $D_{i,j}(T) = 0$, for the points outside T , $D_{i,j}(T)$ is defined as the Euclidean distance to the closest point on the boundary of T . For the denominator, it is a normalization term based on the size of the silhouette S . As the Balan and Black [19], we set the $\delta = 1.5$ to achieve the effect of cameras depth invariance.

Then, the distance between the projection of the 3D model points and the image observation is minimized with the model PCA weights parameters α . We define the objective function like Balan and Black [19] as follows:

$$E_{\text{total}}(\alpha) = \sum_{k=1}^K \tilde{d}(S_m^k(\alpha), S^k) + \tilde{d}(S^k, S_m^k(\alpha)) \quad (5)$$

this function uses a symmetric distance to match the estimated and observed silhouettes over the K camera views, but we only consider the shape parameters not including the pose parameters. We minimize it just to optimize a set of 20 principal shape components from all camera perspectives.

6. Experimental Results

In this section, we provide experimental results for our 3D morphable model silhouette fitting process described in Section 5. We test our system using a database of MPI08 [21, 28] (hb data set) provided by the University of Hannover, Germany, a person is captured with 8 HD cameras with a resolution of $1004 * 1004$ pixels. The initial position of human model is computed from the visual hull centroid, and the human skeleton pose can be computed by the visual hull mesh. Our goal is to match the boundary contours between input and morphable models silhouettes. For computational efficiency, we use the distance between the outline of one silhouette and the outline of the other and vice-versa can be used. Therefore, we have to measure the distance between the model contours and image contours. We choose the error function to be minimized, we fit our morphable model to images by iteratively minimizing the error in all views, and we estimate 20 PCA components for the 3D morphable human model. Figure 5 shows the final estimated results, we choose the first frame and the 25th frame to test.

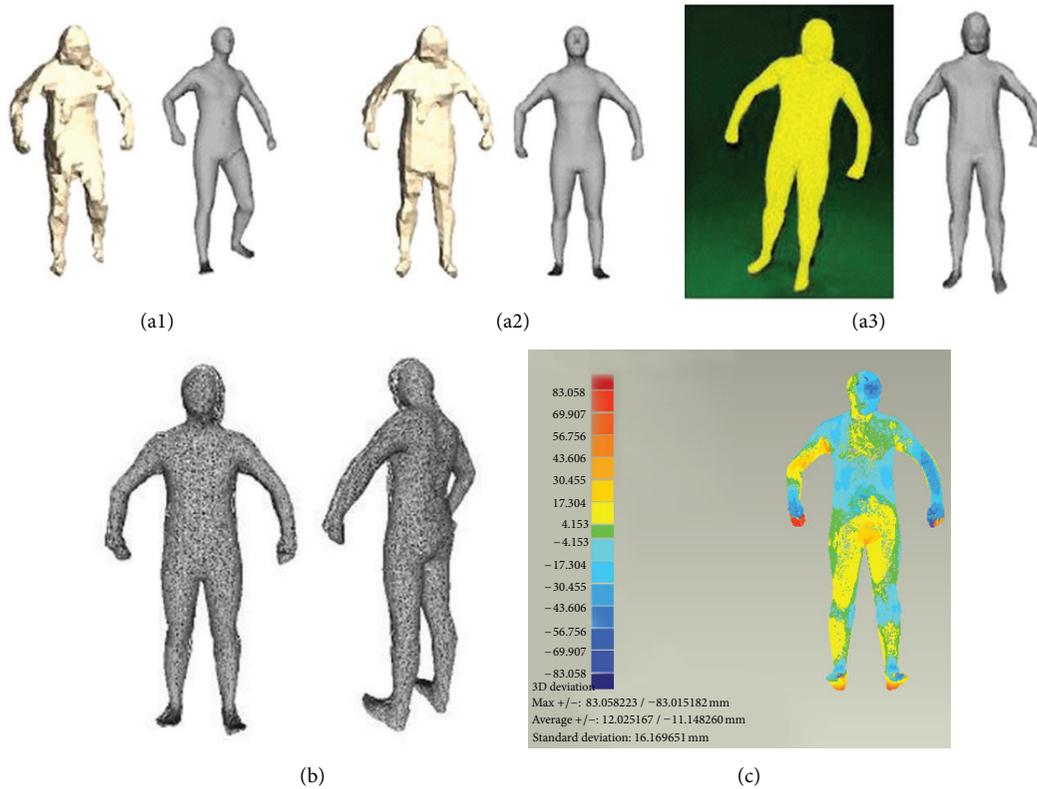


FIGURE 6: The compare result with visual hull and scan model: (a1) and (a2) are compared result with visual hull mesh, (a3) is the 3D scan model, (b) the scan model overlaid with the estimated model fit in, and (c) the compare difference with the scan model using corresponding mesh points, the unit is millimeter.

As we know, it is a difficult problem that estimates biometric parameters from a given 3D human model, and we do not know the ground truth values of this subject. We just compare it to the 3D scan model provided by Pons-Moll et al. [21, 28]. Figure 6 presents the compare result with the visual hull mesh and the scan model: the average distance is 0.075 mm, standard deviation is 16.16 mm, we can see that estimated human shape is very similar to the scan model except the hands part, for most of application, the hands parameter is not important.

7. Conclusion

Our system implements automatic initialization and gives right estimates. We have presented a method for estimating 3D human pose and shape from multi-view imagery. The approach based a learned 3D morphable human model using PCA method. The pose variation is due to the underlying skeleton. When we get the approximate body shape and pose we take into account not only the skeleton pose but also the nonrigid deformations of the human body shape. Then, we can obtain detailed human model shapes with full correspondence. The shape we computed can replace the 3D scan model in motion capture areas. However, in these cases the human body silhouette should be easy to extract from the image data. The subject should be in tight clothes in multi-view cameras (more than 5 cameras) and have no occlusion. And we do

not use tracking algorithm, we fit the 3D morphable model for every frame independently. In the future work, we will consider more constraints on body shape, for the exposed parts of the body, the skin color can be constrained to add the fitting precision. And surface tracking will also be considered in the future.

Acknowledgments

The authors would like to thank Hasler [6], Gall [25], and Pons-Moll [21, 28] for providing their database for research purpose. This work is supported by the National Key Technology R&D Program of China (2012BAH01F03), National Natural Science Foundation of China (60973061, 61173096, R1110679), National 973 Key Research Program of China (2011CB302203), Doctoral Fund of Ministry of Education of China (20100009110004, 20113317110001), Beijing Natural Science Foundation (4123104), and Tsinghua-Tencent Joint Lab for IIT.

References

- [1] F. Moreno-Noguer and P. Fua, "Stochastic exploration of ambiguities for non-rigid shape recovery," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.
- [2] M. Carlini and S. Castellucci, "Modelling and simulation for energy production parametric dependence in greenhouses,"

- Mathematical Problems in Engineering*, vol. 2010, Article ID 590943, 28 pages, 2010.
- [3] D. Vazquez, A. M. Lopez, and D. Ponsa, "Unsupervised domain adaptation of virtual and real worlds for pedestrian detection," in *Proceedings of the 21st International Conference on Pattern Recognition*, pp. 3492–3495, Barcelona, Spain, November 2012.
 - [4] P. Fan, S.-Y. Chen, S.-J. Lu et al., "High dynamic range color image rendering with human visual adaptation simulation," in *Proceedings of the 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 234–237, Taipei, Taiwan, September 2009.
 - [5] D. Zhang, Z. Miao, and M. Wang, "Images based human volumetric model reconstruction and animation," in *Proceedings of the 2nd International Conference on Image Analysis and Signal Processing (IASP '10)*, pp. 210–213, April 2010.
 - [6] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H. P. Seidel, "A statistical model of human pose and body shape," *Computer Graphics Forum*, vol. 28, no. 2, pp. 337–346, 2009.
 - [7] A. Balan, L. Sigal, M. Black, J. Davis, and H. Haussecker, "Detailed human shape and pose from images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, June 2007.
 - [8] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, 2001.
 - [9] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, no. 2-3, pp. 90–126, 2006.
 - [10] M. Sunkel, B. Rosenhahn, and H. P. Seidel, "Silhouette-based generic model adaptation for marker-less motion capturing," in *Proceeding of the 2nd Conference on Human Motion: Understanding, Modeling, Capture and Animation*, pp. 119–135, 2007.
 - [11] J. Gall, B. Rosenhahn, and H. P. Seidel, "An introduction to interacting simulated annealing," in *Human Motion-Understanding, Modeling, Capture and Animation, Computational Imaging and Vision*, vol. 36, pp. 319–345, Springer, Berlin, Germany, 2008.
 - [12] N. Hasler, H. Ackermann, B. Rosenhahn, T. Thormählen, and H. P. Seidel, "Multilinear pose and body shape estimation of dressed subjects from image sets," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 1823–1830, June 2010.
 - [13] R. Benjamin, C. Michel, and N. Vincent, "Generic initialization for motion capture from 3D shape," in *Proceedings of the International Conference on Image Analysis and Recognition (ICIAR '10)*, pp. 306–315, June 2010.
 - [14] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis, "SCAPE: shape completion and animation of people," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 241–253, 2005.
 - [15] L. Sigal, A. O. Balan, and M. J. Black, "Combined discriminative and generative articulated pose and nonrigid shape estimation," in *Proceedings of the 21st Annual Conference on Neural Information Processing Systems (NIPS '07)*, December 2007.
 - [16] P. Guan, A. Weiss, A. O. Balan, and M. J. Black, "Estimating human shape and pose from a single image," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '09)*, pp. 1381–1388, September 2009.
 - [17] S. Zhou, H. Fu, L. Liu, D. Cohen-Or, and X. Han, "Parameter reshaping of human bodies in images," *ACM Transactions on Graphics*, vol. 29, no. 4, article 126, 2010.
 - [18] A. O. Balan and M. J. Black, "The naked truth: estimating body shape under clothing," in *Proceedings of the 10th European Conference on Computer Vision (ECCV '08)*, D. A. Forsyth, P. H. S. Torr, and A. Zisserman, Eds., vol. 5303 of *Lecture Notes in Computer Science*, pp. 15–29, Springer, Marseille, France, 2008.
 - [19] Y. Chen, T. Kim, and R. Cipolla, "Inferring 3D shapes and deformations from single views," in *Computer Vision—ECCV*, vol. 6313, pp. 300–313, Springer, Berlin, Germany, 2010.
 - [20] G. Pons-Moll, A. Baak, T. Helten, M. Müller, H. P. Seidel, and B. Rosenhahn, "Multisensor-fusion for 3D full-body human motion capture," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 663–670, June 2010.
 - [21] J. Lee, B. Moghaddam, H. Pfister, and R. Machiraju, "Silhouette-based 3D face shape recovery," in *Graphics Interface Proceedings*, pp. 21–30, Canadian Computer-Human Communications Society, 2003.
 - [22] A. Jain, T. Thormählen, H. P. Seidel, and C. Theobalt, "Movie re-shape: tracking and reshaping of humans in videos," *ACM Transactions on Graphics*, vol. 29, no. 5, article 148, 2010.
 - [23] S. Y. Chen, J. H. Zhang, Y. F. Li, and J. W. Zhang, "A hierarchical model incorporating segmented regions and pixel descriptors for video background subtraction," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 118–127, 2012.
 - [24] J. Gall, C. Stoll, E. De Aguiar, C. Theobalt, B. Rosenhahn, and H. P. Seidel, "Motion capture using joint skeleton tracking and surface estimation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 1746–1753, June 2009.
 - [25] S. Corazza, L. Mündermann, A. M. Chaudhari, T. Demattio, C. Cobelli, and T. P. Andriacchi, "A markerless motion capture system to study musculoskeletal biomechanics: visual hull and simulated annealing approach," *Annals of Biomedical Engineering*, vol. 34, no. 6, pp. 1019–1029, 2006.
 - [26] I. Baran and J. Popovi, "Automatic rigging and animation of 3D characters," *ACM Transactions on Graphics*, vol. 26, no. 3, article 72, 2007.
 - [27] L. Pishchulin, A. Jain, M. Andriluka, T. Thormählen, and B. Schiele, "Articulated people detection and pose estimation: reshaping the future," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 3178–3185, June 2012.
 - [28] A. Baak, T. Helten, M. Mueller, G. Pons-Moll, H. P. Seidel, and B. Rosenhahn, "Analyzing and evaluating markerless motion tracking using inertial sensors," in *Proceedings of the 11th European Conference of Computer Vision (ECCV)*, Crete, Greece, September 2010.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

