Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 715808, 8 pages
http://dx.doi.org/10.1155/2013/715808

Research Article

Optimization and Soft Constraints for Human Shape and Pose Estimation Based on a 3D Morphable Model

Dianyong Zhang,[1,2] Zhenjiang Miao,[1,2] Shengyong Chen,[3] and Lili Wan[1,2]

[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 10 April 2013; Revised 9 August 2013; Accepted 9 August 2013

Academic Editor: Yang Xu

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.

We propose an approach about multiview markerless motion capture based on a 3D morphable human model. This morphable model was learned from a database of registered 3D body scans in different shapes and poses. We implement pose variation of body shape by the defined underlying skeleton. At the initialization step, we adapt the 3D morphable model to the multi-view images by changing its shape and pose parameters. Then, for the tracking step, we implement a method of combining the local and global algorithm to do the pose estimation and surface tracking. And we add the human pose prior information as a soft constraint to the energy of a particle. When it meets an error after the local algorithm, we can fix the error using less particles and iterations. We demonstrate the improvements with estimating result from a multi-view image sequence.

1. Introduction

The detection and recovery of human shapes and their 3D poses in 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, robotics. Model-based motion capture is especially suited to markerless motion capture since it provides a way to constrain the search space by the degrees of freedom of the skeleton. Initialization of human motion capture always requires the definition of the human models approximate the shape, kinematic structure, and initial pose. Most of the approaches built a scan model of the person to be tracked. The majority of algorithms for human pose estimation are using an initialized general model with limb lengths and shape manually. Accurately detailed human model can be used for markerless motion capture to track a subject individual model, which includes information on both body shape and pose. Detailed 3D human shape and pose estimation from multi-view images is still a difficult problem that does not have a satisfactory solution.

We all know that the local optimization methods are faster, but if there are visual ambiguities, or fast motions, the tracker might fail. To get more robust result, global optimization methods can be used, like particle filter technique, because they can represent uncertainty by a rigorous Bayesian paradigm. The problem is that so many particles are needed to get the right predicted result in the dimension of human pose parameter space with usually more than 20 degrees of freedoms. Gall et al. [1] propose an approach combining local and global algorithm using skeleton and surface information. But it needs an accurate 3D scan model and it is sensitive to noise of silhouettes. As we all know that getting a 3D scan human model is very expensive, we use a 3D morphable model like Jain et al. [2] to generate an individual human model. Our parametric representation of human body is based on a 3D morphable human model with an underlying predefined skeleton. Kinds of human models can be generated based on the deformable human model that is learned from a scan database [3] of over 550 full body 3D scans taken from 114 undressed subjects. The estimated refined shape and skeleton pose from multi-view images serve as initialized model for the next frame to be tracked.

For most of the controlled environments, the local and global method can get the right result. But particle filtering costs a much higher running time to search the whole pose state space, and it needs lots of particles and iterations to predict. We add an energy function to constrain each particle by silhouettes and additional pose prior information instead of relying on only a large number of particles to search the whole pose space. It makes use of much fewer particles according to the gradient of an evaluation function.

The remaining sections of this paper are organized as follows. In Section 2, we will present the relevant previous work on model-based markerless motion capture. In Section 3, we describe the 3D morphable human model and skeleton defined information. In Section 4, we will describe the optimization and pose soft constraints algorithm in detail. In Section 5, we will show the estimation result. We will conclude this paper in Section 6.

2. Related Works

The papers [4–6] present comprehensive survey of existing related techniques in motion capture research area of computer vision. A model-based markerless motion capture system can be divided into four steps: 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 specific pose or predefined motion style. The prior 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 shapes. For improved accuracy in tracking, an articulated model which approximates the shape of a specific subject is needed. Because of few images, people cannot get the accurate body shape information, furthermore, the shape of the subject can differ from person to person. Our approach is based on a 3D morphable model of human shape and pose similar to [2]. Jain et al. using the morphable model to estimate human shape and pose simultaneously, they designed both shape particles and pose particles as the search space, its computational time is very high and it just used in reshape the human in 2D images.

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. A popular parametric model SCAPE (Shape Completion and Animation for PEople) [7], which is a data-driven method for building body shapes with different poses and individual body shape. This model has recently been adopted as morphable model to estimate human body shape from monocular or multi-view images [8–12]. Bălan et al. [8] have adopted this model closer to observed silhouettes to capture more detailed body deformations; however, it cannot capture skeleton joint parameters.

Most recently, the approach has been used to infer pose and shape from a single image. Guan et al. [9] have considered more visual cues, shading cues, and internal edges as well as silhouettes to fit the SCAPE model to an uncalibrated single image with the body height constrained. Sigal et al. [10] describe a discriminative model based on mixture of experts to estimate SCAPE model parameters from monocular and multicameras image silhouettes. Chen et al. [11] proposed a probabilistic generative method that models 3D deformable shape variations and infers 3D shapes from a single silhouette image. 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. [13] proposed a hybrid tracker approach that combined correspondence based local optimization with five inertial sensors placed at human body; although they can obtain a much accurate and detailed human tracker, they need additional sensors. The existing research about model-based tracking approaches the problem using a Bayesian filtering formulation or as an optimization problem. Gall et al. [1, 14, 15] introduce an approach for global optimization that is for human motion capturing called interacting simulated annealing (ISA), which is based on a particle filter and simulated annealing. While global optimizations are capable of running fully automatic, the computation time is very high. Recently, several papers [16, 17] implement it and improve it. 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. 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 data sets. Our morphable model is based on scan database [3] of over 550 full body 3D scans taken from 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. They also defined semantic correspondence between the scans. We learn what the PCA model contains for each subject and the modeling shape variations of human body via PCA method. Therefore, a human model is given by

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

where the human shape parameter is $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, m_i is the ith eigen human and m_0 is the mean or average human model. Similar to [1, 2], the morphable model is a combination of bone skeletons and joints. Like Jain et al. [2] and Gall et al. [1], we drive the body pose by a defined underlying skeleton. Shows in Figure 2. And the shape parameters can be described by PCA parameters. In our paper, we define 20 human PCA components like [2]. We define a kinematic chain; the motion of body model can be parameterized by the joint angles. For many years, kinematic chains are widely used human tracking and motion capture systems. The mesh deformation can be controlled by linear

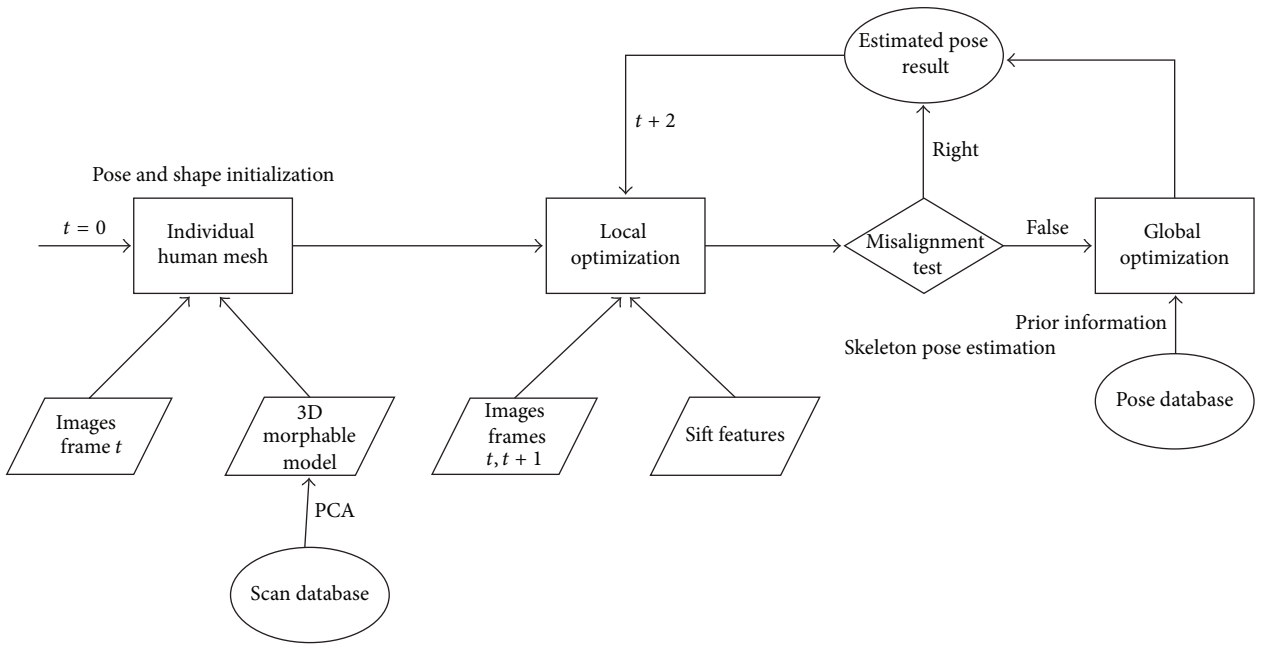


Figure 1: Pipeline of our method.

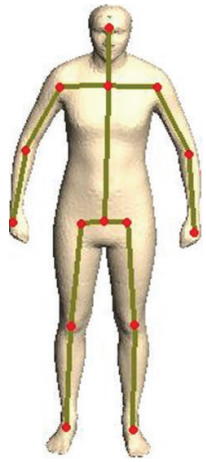


Figure 2: The template human model with the underlying skeleton.

blend skinning (LBS) technique. If $\omega_{i,j}$ 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:

$$v_i' = \sum_j \omega_{i,j} \left(T_j v_i\right). \tag{2}$$

The bone weights for the vertices mean how much each bone transformation affects each vertex. These weights are normalized such that $\sum \omega_{i,j} = 1$. We rig the skeleton with 22 degrees of freedom including the six degrees for the global position and orientation of the model. And we rig the model using the autorig method of Baran and Popović [18], attaching the weights value in Maya software.

3.1. Human Adaptation for Multiview. We change the human pose according to the real pose from the visual hull model which is reconstructed from multi-view silhouettes. Then, we estimated the shape parameters from multisilhouettes. For the detailed information, we refer to our former paper [19]. And the adaptation results shown in Figures 3 and 5.

3.2. Human Kinematic Chain. The position of 3D vertex v_i which is associated with kinematic chain k_i and influenced by n_{k_i}, the rigid body motion was represented as a twist. A joint of a body limb can be modeled by a twist $\theta\widehat{\xi}$. Every 3D rigid motion can be represented in an exponential form of the homogeneous matrix as follows:

$$M = \exp\left(\theta\widehat{\xi}\right) = \exp\begin{pmatrix} \widehat{\omega} & v \\ 0_{3\times 1} & 0 \end{pmatrix}, \tag{3}$$

where $\theta\widehat{\xi}$ is the matrix representation of a twist $\widehat{\xi} \in \mathrm{se}(3) = \{(v, \widehat{\omega}) \mid v \in R^3,\ \widehat{\omega} \in \mathrm{so}(3)\}$ with $\mathrm{so}(3) = \{A \in R^{3\times 3} \mid A = -A^T\}$.

The coordinates of the transformed point can be described as

$$T_\chi V_i = \prod_{j=0}^{n_{k_i}} \exp\left(\theta_{l_{k_i}(j)} \widehat{\xi}_{l_{k_i}(j)}\right) V_i, \tag{4}$$

where l_{k_i} is a mapping that represents the order of the joints in the kinematic chain, k_i is the limb associated with V_i, n_{k_i} are the joints influencing the position and rotation of limb k_i. For further details we refer to [20]. We denote by the joint angles state vector $\theta = (\theta_1, \ldots, \theta_n)$ and the 6 parameters of the twist ξ_0 associated with the model reference system. We define a vector that represents the state of the human model as

$$\chi = \left(\theta_0 \widehat{\xi}, \theta\right). \tag{5}$$

(a) (b) (c)

Figure 3: The morphable model adaptation result for the first frame. (a) the result of morphable model projected to image, (b) the scan model overlaid with the estimated model fit in, and (c) the difference between two model with corresponding mesh points,the unit is millimeter, the average distance is 0.075 mm and standard deviation is 16.16 mm.

In Section 4, we will describe how to compute vector χ that makes the 3D morphable model pose to fit the pose of the person in images.

4. Optimization and Surface Tracking

In the multiple camera way, each camera has its own coordinates. We should transform the human body model local coordinates (x, y, 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.

4.1. Local Optimization. For the local optimization, we use contour correspondences and the texture correspondences to get the right estimation result.

According to the extracted image body contour and the silhouette of the projected surface mesh, the closest point correspondences between these two contours can be used to define a set of corresponding 3D rays and 3D points in order to minimize the error between the 2D and 3D point of a correspondence. For the texture correspondences, we use SIFT features between two frames taken from the same camera (Figure 4). 3D point-line based pose estimation is modeled as a 3D plücker line $L = (d, l)$ with the direction vector d of the line and moment l. The error of a pair of 3D-2D points can be given by the norm vector between the transformed 3D point $T_\chi V_i$ and the 3D ray line $L_i = (d_i, l_i)$,

$$\left\| \prod (T_\chi V_i) \times d_i - l_i \right\|_2 \tag{6}$$

while the transformed point is in homogeneous coordinates and the plücker coordinates are not, $\prod$ is the projection from the homogeneous to nonhomogeneous coordinates.

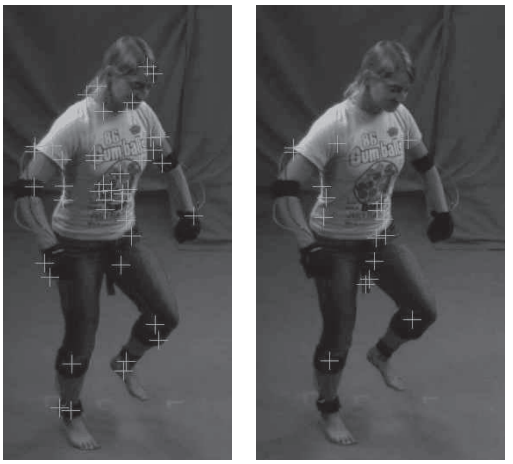
For the accurate result, we have to minimize the alignment error between the body contour of the image and the projected surface mesh.

To find the vector χ, the sum of errors over all correspondences should be minimized. Assume that have N correspondences, the error to be minimized can be described by the following equation:

$$\operatorname{argmin} \frac{1}{2} \sum_i^N w_i \left\| \prod \left(T_\chi V_i \right) \times d_i - l_i \right\|_2^2, \tag{7}$$

where w_i is a weight parameter of each correspondence. We linearise the model using the first order Taylor approximation as follows:

$$\exp\left(\theta \widehat{\xi}\right) = \sum_{k=0}^{\infty} \frac{\left(\theta \widehat{\xi}\right)^k}{k!} \approx I + \theta \widehat{\xi} = I + \theta_0 \widehat{\xi}_0 + \cdots + \theta_j \widehat{\xi}_j, \tag{8}$$



(a) (b)

Figure 4: The sift feature of neighbour frames.

Figure 5: The result of shape and pose initialization from 8 view images.

where I is the identity matrix, and if we insert the Taylor approximation into (7), we can get the following equation:

$$\operatorname{argmin}\frac{1}{2}\sum_{i}^{N} w_i \left\| \prod \left(\left(I + \theta_0 \widehat{\xi}_0 + \sum_{j=1}^{n_{k_i}} \theta_{l_{k_i}(j)} \widehat{\xi}_{l_{k_i}(j)} \right) V_i \right) \times d_i - l_i \right\|_2^2. \tag{9}$$

Although the local optimization has converged to a minimum value, it is not guaranteed that the result is right. For example, we have a low error, but some limbs are still misaligned. We calculate this error for each limb; if a misaligned limb is detected, then all subsequent limbs will be labelled as misaligned. Then, we continue the global optimization step to fix it.

4.2. Filter Particle with Soft Constraints. The local optimization methods are faster and they can get accurate results, but if there are visual ambiguities or fast motions, the motion track will fail. To deal with these problems of local optimization system, we use particle filters to represent uncertainty through a rigorous Bayesian paradigm. The global optimization methods use a set of particles to estimate the pose. If whole body pose has to be estimated, the computation time will be very large. The problem with the global optimization is the distribution of the optima in search space. Every particle has its own state and a weight. After each iteration, the particle generates a new state. It is constructed by a linear interpolation between the predicted pose and the estimated pose from the previous frame. Usually, the compute time is very large; it is determined by the number of particles and iterations. When all particles have been updated, they will be resampled. The particles that are far from the correct solution will be discarded. The weights of particles are evaluated according to the following equation and normalised to a sum of 1. For reducing the number of particles and iterations, we add the pose prior information to constrain the particle. So, in order to find the optimal value for pose χ, we define the energy function as follows:

$$E(\chi) = E_S(\chi) + \lambda E_R(\widetilde{\chi}) + E_p(\chi), \tag{10}$$

where the first term measures the silhouette error between the projected surface model and the silhouette image. The second term is a penalty for strong deviations from the predicted pose, and λ is the weight factor of the penalty; we set it to 0.01. The third term is a human pose prior constraint of the predicted pose. The silhouette error for pose χ for view camera C calculates the pixelwise differences between the projected surface model and the silhouette image. It is generated by projecting the surface model according to the pose of the particle; the error of a particle pose for view C can be given by

$$E_S^C(\chi) = \frac{1}{\left|S_C^p\right|} \sum_{p \in S_C^p} \left| D_C^p(\chi)(p) - D_C(p) \right| + \frac{1}{|S_C|} \sum_{q \in S_C} \left| D_C^p(\chi) - D_C(q) \right|, \tag{11}$$

where S_C^p is the estimated projected surface and S_C is the correct binary silhouette images, D_C^p and D_C are their Chamfer distance transforms, and the sums over p and q are to show that the differences are only the pixels located inside the silhouette area of projected surface model S_C^p and the silhouette S_C, not the background areas. So, it is very expensive for every particle. We set each pixel inside the projected surface to zero; the silhouette energy term $E_S(\chi)$ is defined as the average of $E_S^C(\chi)$ over all views.

The second term of (10) describes a smoothness constraint in the lower dimensional parameter space as follows:

$$E_R(\widetilde{\chi}) = \left\| \widetilde{\chi} - P(\widehat{\chi}) \right\|_2^2. \tag{12}$$

The third term is a soft human pose constraint of the energy function. It contains the human anatomical constraints and the pose probability density from the training samples:

$$E_P(\chi) = E_{\text{prior}}(\chi) + E_{\text{learned}}(\chi). \tag{13}$$

For the human motion, the joint angle should abide to human anatomical rule, so

$$E_{\text{prior}}(\chi) = \sum_i \frac{\max^2\left(0, \chi_i^{\min} - \chi_i, \chi_i - \chi_i^{\max}\right)}{\sigma^2}, \tag{14}$$

where the joint angle bounds $(\chi_i^{\min}, \chi_i^{\max})$, like papers [21], we learning the various pose probabilities from a set of training samples. We use a nonparametric density estimate by Parzen-Rosenblatt estimator [22, 23] with a Gaussian kernel like Brox et al. [15, 20] as follows:

$$p_{\text{learned}}(\chi) = \frac{1}{\sqrt{2\pi}\sigma T} \sum_{j=1}^{T} \exp\left(-\frac{\left(\chi_j - \chi_i\right)^2}{2\sigma^2} \right), \tag{15}$$

where T is the number of training samples and σ is the kernel width parameter, which is learned from motion capture data, and is the maximum nearest neighbor distance between all training samples. The pose prior information can provide human anatomical constraints and right pose parameters of normal degrees of freedom (DOF). We use about 100 samples from different motions for the physical constraints by

$$E_{\text{learned}}(\chi) = -\log\left(p_{\text{learned}}(\chi)\right). \tag{16}$$

5. Experimental Results

We test our system using a database of MPI08 [13, 24] (hb data set) provided by the University of Hannover Germany; the person is captured with 8 HD cameras with a resolution of $1004 * 1004$ pixels. For the 3D mesh model, considering the difficulty to get the laser scan model, we use our 3D morphable model generated by PCA method according to the multi-view image silhouettes. Due to complex motion, the correspondences between the model and the silhouettes cannot provide enough information to estimate the correct pose.

The local optimization is capable of tracking the person, but it cannot recover from errors. These errors usually happen

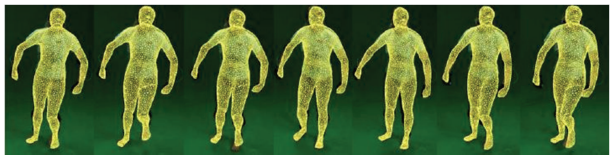


Figure 6: The estimation result from camera view 1.

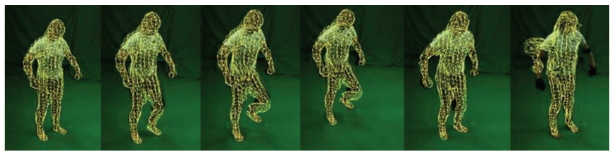
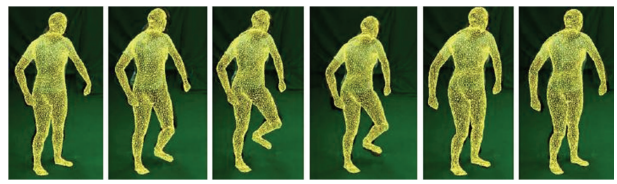


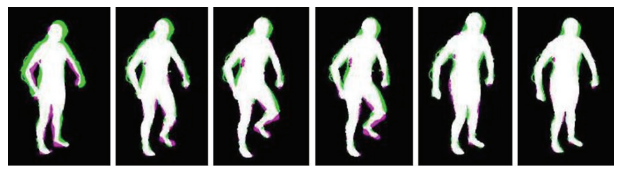
Figure 7: The estimation result using visual hull mesh mode from view camera 2: for the limit of camera numbers, the visual hull model is not very exact and the irregular triangle mesh especially in the shoulder parts, when the mesh deforms the pose, there will generate error like the frame 45, and the following frames will be all error. The morphable model can get right estimation results.

for smaller parts of the body, while bigger parts are less prone to error. The error's number highly depends on the visibility of the body parts; the frame rate and the speed of the moving body parts are the reason too. The global optimization is only initiated after misalignment. But global optimization algorithms are hard to use pose estimation because of the high computational cost of approaches. When we add the pose prior information, the particles can be optimized. Gall et al. [1], 15 iterations and $(20 * 15 = 300)$ a maximum of 300 particles are sufficient to estimate the pose correctly. We just use 25 pose particles and 10 iterations by pose prior information constraint. In the MPI08 data set, no ground truth data is available. For the experiments carried out using this data set, there is no concrete evaluation to be given. Therefore, we will use our own vision to determine whether an estimation result is visually correct or incorrect. Show as Figures 5, 6, 7, 8, and 9. Figure 5 show the result of pose and shape initialization. For testing the validity of the morphable, we use the visual hull mesh model to track. The visual hull model is very rough, and the skeleton cannot get well rig with mesh vertices especially in shoulder parts. When the skeleton deforms, the mesh will make an error, as shown in Figure 7. Figure 8 shows the estimation result using morphable model from view camera 2. Figure 9 shows the local and global estimation result.

5.1. Computation Time. The computation time depends on several factors, such as the quality of the model, number of camera views, and the quality of the original images. The global optimization costs the most computation time of the whole system. The computation time depends on the number of particles and iterations. The computer is used with an Intel Core 2 Duo processor at 3.0 GHz, and a 2.0 G memory. We make the program by multi-threaded. For each frame, local optimization of our system may cost 8–10 seconds, after misalignment, the global optimization part may cost 180–300 seconds.



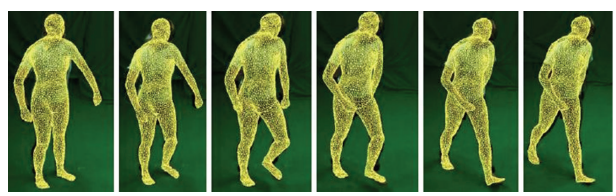
(a)



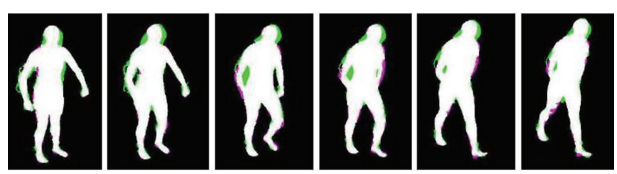
(b)

Frame 19 Frame 24 Frame 29 Frame 31 Frame 37 Frame 45

(c)



(d)



(e)

Frame 49 Frame 57 Frame 65 Frame 70 Frame 75 Frame 80

(f)

Figure 8: The estimation result using morphable model: (d, f) the estimation result projection to original image; (b, e) the silhouette difference of the estimated projected view and the original view; the green parts belong to the original silhouette, and purple parts belong to the estimated projected silhouette. We can see that we almost get the right estimation result; (c, f) the output model mesh with estimated pose.

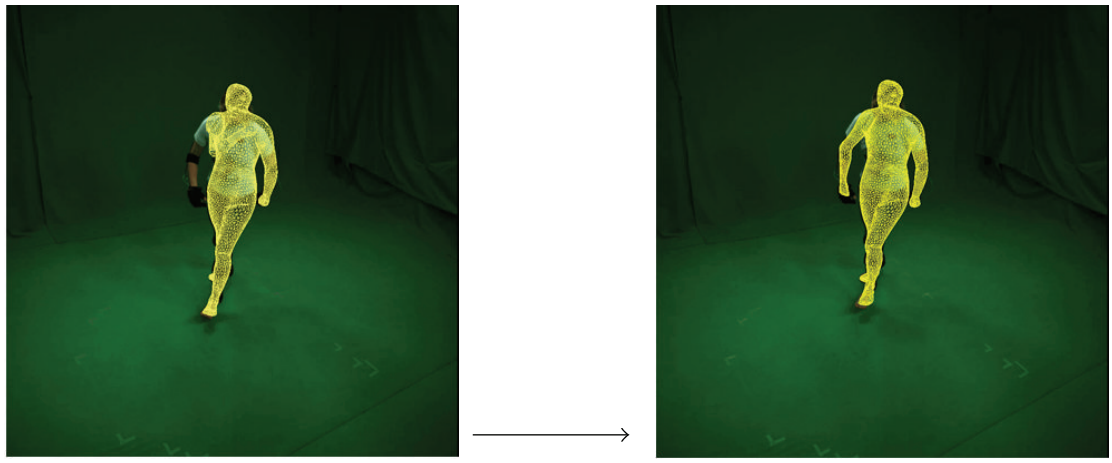


Figure 9: The local estimation Result. When the motion is slow, the estimation result using local optimization method is correct (frame no. 1–60), Once it changes faster, the tracking will make error no. 70 frame, this will need global optimization to fix. Using annealed particle filtering, we need 100 particles and 15 iterations, after we add the pose prior constraint for the particles, we just need 25 particles and 10 iterations.

6. Conclusion

In this work, a robust and accurate human motion capture method [1] has been investigated and improved. Both the local optimization and global optimization methods are all based on image silhouettes and so a proper background subtraction is required. For most model-based methods, they all need a 3D scan model. We have presented a method for estimating human pose and shape from multi-view imagery. The approach based on a learned 3D morphable human model using PCA method and a pose prior information as a soft constraint concludes anatomical constraints and the pose probability density from the training samples.

Good initial pose and shape are very important point to get the right pose estimation result. It is not suitable in applications in which a real-time pose estimation is required. Beside the body pose, we can also provide the human mesh model instead of the 3D scan model. This gives additional information about the pose and the person. When we get the approximate body shape and pose, 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. The subject should be in tight clothes in multi-view cameras. In the future work, we will consider single view-person pose estimation.

Acknowledgments

The authors would like to thank the anonymous reviewers for their constructive comments. they would like to thank Hasler [3], Gall [1] and Pons-Moll [13, 24] for providing their database for research purpose. This work is supported by the National Key Technology R&D Program of China (2012BAH01F03), the National Natural Science Foundation of China (60973061), the National 973 Key Research Program of China (2011CB302203), the Ph.D. Programs Foundation of the Ministry of Education of China (20100009110004), and Beijing Natural Science Foundation (4123104).

References

[1] 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, Miami, Fla, USA, June 2009.

[2] A. Jain, T. Thormählen, H.-P. Seidel, and C. Theobalt, "MovieReshape: tracking and reshaping of humans in videos," *ACM Transactions on Graphics*, vol. 29, no. 6, Article ID 1866174, 2010.

[3] 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.

[4] 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.

[5] 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.

[6] G. Pons-Moll and B. Rosenhahn, *Book Chapter on Model Based Pose Estimation to Appear in Guide to Visual Analysis of Humans: Looking at People*, Springer, New York, NY, USA, 2011.

[7] 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.

[8] A. O. Bălan, L. Sigal, M. J. Black, J. E. Davis, and H. W. Haussecker, "Detailed human shape and pose from images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, Minneapolis, Minn, USA, June 2007.

[9] 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 12th International Conference on Computer Vision (ICCV '09), 2009.

[10] L. Sigal, A. Balan, and M. J. Black, "Combined discriminative and generative articulated pose and non-rigid shape estimation," in *Proceedings of the 21st Annual Conference on Neural Information Processing Systems (NIPS '07)*, December 2007.

[11] Y. Chen, T.-K. Kim, and R. Cipolla, "Inferring 3D shapes and deformations from single views," in *Proceedings of the 11th European Conference on Computer Vision*, pp. 300–313, 2010.

[12] S. Zhou, H. Fu, L. Liu, D. Cohen-Or, and X. Han, "Parameter reshaping of human bodies in images," *ACM Transactions on Graphics (TOG)*, vol. 29, no. 4, 2010.

[13] 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, San Francisco, Calif, USA, June 2010.

[14] J. Gall, B. Rosenhahn, and H. P. Seidel, "An introduction to interacting simulated annealing," in *Human Motion*, vol. 36 of *Understanding, Modeling, Capture and Animation, Computational Imaging and Vision*, pp. 319–345, Springer, 2008.

[15] J. Gall, B. Rosenhahn, T. Brox, and H.-P. Seidel, "Optimization and filtering for human motion capture: AAA multi-layer framework," *International Journal of Computer Vision*, vol. 87, no. 1-2, pp. 75–92, 2010.

[16] D. Olga, *Motion capture in uncontrolled environments [M.S. thesis]*, Eidgenossische Technische Hochschule Zurich; INFK; Computer Vision and Geometry Group, 2010.

[17] Y. Liu, C. Stoll, J. Gall, H.-P. Seidel, and C. Theobalt, "Markerless motion capture of interacting characters using multi-view image segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 1249–1256, Providence, RI, USA, June 2011.

[18] I. Baran and J. Popović, "Automatic rigging and animation of 3D characters," *ACM Transactions on Graphics*, vol. 26, no. 3, 2007.

[19] D. Y. Zhang, Z. J. Miao, and S. Y. Chen, "Human model adaptation for multi-view markerless motion capture," *Mathematical Problems in Engineering*, vol. 2013, Article ID 564214, 7 pages, 2013.

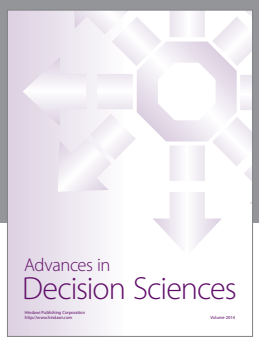
[20] T. Brox, B. Rosenhahn, and D. Cremers, "Contours, optic flow, and prior knowledge:cues for capturing 3D human motion in videos," in *Human Motion*, vol. 36 of *Understanding, Modeling, Capture and Animation*, pp. 265–293, Spring, Computational Imaging and Vision, 2007.

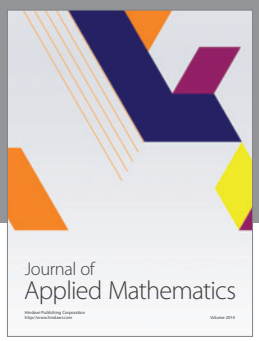
[21] L. Sigal, A. O. Balan, and M. J. Black, "HumanEva: synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *International Journal of Computer Vision*, vol. 87, no. 1-2, pp. 4–27, 2010.

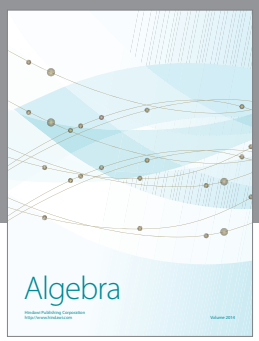
[22] E. Parzen, "On estimation of a probability density function and mode," *Annals of Mathematical Statistics*, vol. 33, pp. 1065–1076, 1962.

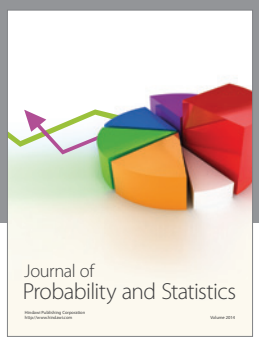
[23] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *Annals of Mathematical Statistics*, vol. 27, pp. 832–837, 1956.

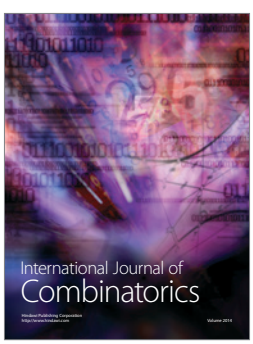
[24] A. Baak, T. Helten, M. Mueller, G. Pons-Moll, H. P. Seidel, and B. Rosenhahn, "Analyzing and evaluating markerless motion tracking using inertial sensors," in *Proceedings of the 11th European conference on Trends and Topics in Computer Vision (ECCV '10)*, pp. 139–152, 2010.

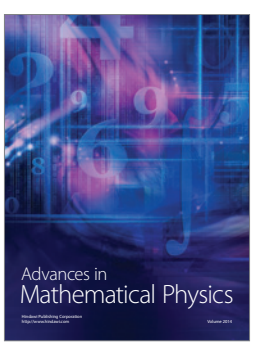


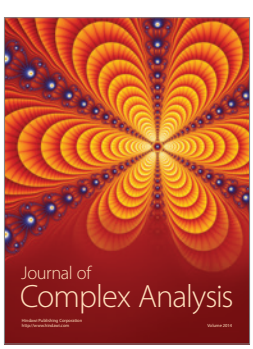


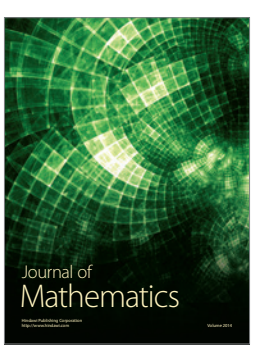


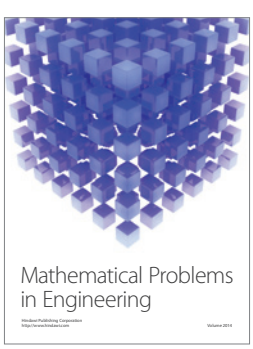


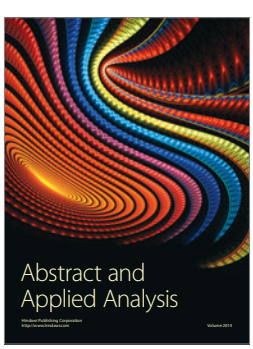


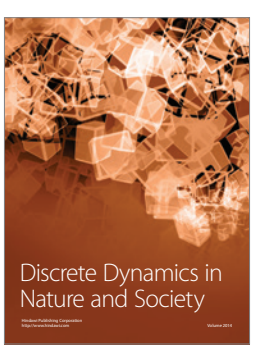


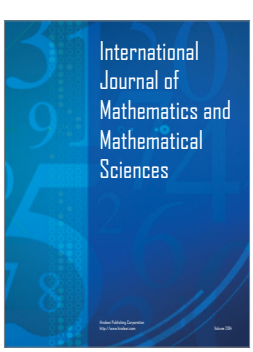












Hindawi

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