Markerless Human Motion Extraction And Surface Estimation Based on Model Segmentation

Weilan LUO†, Toshihiko YAMASAKI†, and Kiyoharu AIZAWA††,†

† Department of Information and Communication Engineering, The University of Tokyo
†† Interfaculty Initiative in Information Studies
E-mail: †{luoweilan,yamasaki,aizawa}@hal.t.u-tokyo.ac.jp

Abstract We propose a human motion tracking method that locates the segmented body parts to the target correspondingly using synchronized multiple cameras. Our method is capable of extracting 3D articulated postures with 42 degrees of freedom through a sequence of visual hulls. We seek for the globally optimal solutions of the likelihood with the local memorization about the "fitness" of each body segment. Our method avoids the local minimum problem efficiently by mean combination and articulated combination of particles selected based on the weights of the different body segments. We deform the template surface model using the motion tracking data by linear blend skinning (LBS). The details of the surface are recovered by fitting the deformed surface to 2D silhouettes. The extracted posture and estimated surface are refined by registering the corresponding body segments. The mean distance between the deformed reference sample model to the target is about $2cm$ and the mean matching difference between the projected images of the generated surfaces and the original images is about 7% in our experiments.

Key words articulated human motion, annealed particle filter, surface estimation, model segmentation

1. Introduction

Kinematic human body motion capture and 3D spatio-temporal surfaces reconstruction, especially for fast motion clips, from synchronous multi-camera or multi-view video sequences are still the challenging and fundamental problems for many applications, including 3D animation movies and games, medical diagnostics motion analysis, or robot motion simulation. Unlike marker-based motion capture system which requires people to wear skin-tight clothing with markers, the multiple images from the marker-less system can be utilized to generate not only the human motion data but also the realistically complicated surfaces even for human wearing loose apparels.

Marker-less motion estimation has been studied in fields of computer graphics and computer vision for years. The features, such as textures, illuminations and depth informations are always utilized to calculate the correspondences between the neighboring frames. It is intuitive to represent the human motion by articulated skeleton models. Then a sequence of surfaces can be recovered easily by a skinning method such as linear blend skinning (LBS).

Vlasic et al. [1] presented a method which pulled the template skeleton to fit the visual hull by minimizing an energy function. This approach does not always work well. So if the posture is misaligned they will adjust it by hand. Gall et al. [2] also extracted the 3D articulated model which registered the contour and texture correspondences by solving an energy minimization problem in the first stage. In addition, they detected the misaligned limbs and refined the pose by particle filter to seek for global optimization. In the surface estimation stage, both Vlasic and Gall utilized the skinning method to generate the surface model and then recover the details by deforming it to match the silhouette rims.

Deutscher et al. [3] constructed an articulated body simplified by cones with elliptical cross-section and assigned the model 29 degrees of freedom. They estimated the 3D posture deforming the model to match 2D images by the annealed particle filter (APF) method. However, the oversimplified models made it difficult to recover complicated shape and motion precisely and they ignored the motion of the hands and feet. In addition, as mentioned in [4], the result of the APF relied on the quality of the initial particles. Hence, it is easy to dash into the local minimization instead of the global optimization of the likelihood especially in the higher dimensional configuration spaces. We proposed a method that defined the local weights and made new particles by mean combination in [5] to avoid the local minimiza-

tion problem. However, this method still depended on the distribution of the initial particles and not really represented the "fitness" of each body segment.

In this paper, we employ the volumetric models [6] generated from multi-view images directly for 3D pose estimation. About 2% number of voxels are chosen for motion tracking in order to decrease the computational cost as the number of voxels of each model is about 100 thousands. The segmented volumetric model is given and assigned 42 degrees of freedom (DOF) as a template volumetric model. We also capture the 3D posture for human body by annealing to generate new ones according to the survival rate of the particles. We calculate the "fitnesses" for all body segments, select several particles that matches well for different body segments and combine them to generate a new particle in order to avoid the local minimization problem of the APF method. In addition, the self-intersection detection is given to avoid to produce the unsubstantial particles.

A segmented template surface with an underlying skeleton model is used for surface reconstruction. In our work, we construct a mesh model by marching cubes [7] and extract the articulated skeleton model from it for the first frame. Then we segment the mesh surface into 15 parts based on the skeleton and geodesic distances similar to [8], [9]. Tadano et al. [8] proposed a motion extraction method based on Reeb graph and a geodesic function utilizing principal component analysis. Lee et al. [9] utilized the extracted skeleton chains, segmented the time-varying mesh surfaces based on skeleton models by distance calculations, and refined the 3D pose using the decomposition results. The skin attachment was estimated based on the surface labeling and heat equilibrium as described in [10]. Then the template surface is deformed based on the estimated motion and deformed to match the silhouettes. Then we extract the transformation between the template surface of the initial frame and the deformed mesh surface by solving a least squares problem. Furthermore, the motion is refined by matching the segment volumetric models by iterative closest points (ICP) method [11].

This paper is organized as follows. Section 1. explains the related work in this field, outlines our method for motion tracking and surface reconstruction. The segmentation method for the template mesh surface and volumetric models are described in in section 2.. A sampling method is talked about in section 3.. In section 5., a motion tracking method that guards to generate particles with global optimization and local memorization is introduced. Section 6. introduces the Laplacian deformation framework to recover the surface details, which enforces the reverted 2D images obtained by the estimated surface to match with the silhouettes. In Section 7., we show our experimental results, and we summarize the results and discuss the future work in Section 8..

2. Model decomposition

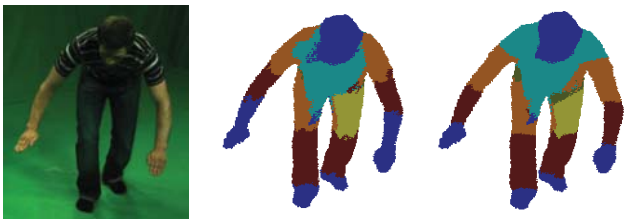
The challenging problem of human motion tracking is how to avoid the local mis-alignment while seeking for the global matching in a high dimensional configuration space. The segmented model is provided to guide the "fitness" of each body segment. In addition, the rest model is labeled by figuring out the correspondences of the deformed segmented which enforces the local registration for each limb.

2. 1 Mesh surface segmentation

It is popular to use a simplified segmented model represented by cones with joints to extract the rest postures. The complicated model such as a subject-specific model proposed in [12] can also be utilized but it is not easy to achieve from the usual maker-less system. As we sample the sequence of the visual hulls to estimate the human postures

The model decomposition work is necessary as we tend to analyze and locate each body segment for the human body and we abandon to use the simple model represented by cones with joints. Although a subject-specific model generation method is proposed in [12], it is not easy to achieve this for the usual maker-less system. However, it is hard to segment the volumetric model directly in a time-consistent manner as no explicit correspondence between frames is given in our time-varying visual hulls. So we first segment the template mesh surface then label the corresponding volumetric model.

We prefer constructing the template model for the first frame directly from images to avoid the registration problem as described in [2] while the surface is generated by the laser. We assume there are not any crossed body segments for the initial posture as the quality of the template will affect on the reconstructed time-varying sequence. The template surface and the visual hulls are all constructed using the multi-camera system. A skeleton model as shown in Fig. 2 (a) is extracted from the mesh surface. It can be obtained by hand or the methods proposed in [8], [9]. The template model is segmented into 15 body parts based on geodesic distance and the underlying skeleton (b). Then the visual hull (c)



(a) Image (b) No refinement (c) Refinement

Fig. 1 Visual hull segmentation.

is decomposed according to the segmented mesh surface. Our proposed mesh surface segmentation algorithm is conducted according to the following steps:

(1) Calculate the five start points which are the nearest one in the model to the corresponding leaf joint of the skeleton.

(2) Label the body segments of head, hands and feet in respectively. We assign a plane which passes the joint of two articulated bones and splits them in two parts. It is easy to detect the vertices which does not belong the corresponding segment. The geodesic distances to the start point are computed. Assume that the start point is labeled as i, the minimal geodesic distance of the vertex that does not belong to the same body segment according to the plane is k, then the vertices whose geodesic distances are less than k will be labeled as i. We add a virtual point which is assumed to be neighboring to all the vertices whose geodesic distances are k. The virtual point is used as the start one for the next segmentation process.

(3) Repeat Step 2 to segment the lower parts of the legs and arms, the upper parts, and the body in turn.

This method is robust for the uniform model while it represents each body segment clearly. We recover the time-varying surfaces based on the template model, the volumetric sequence is able to be labeled according to the corresponding one. Each volumetric model is partitioned into 15 limbs in accordance with the minimum Euclidean distances to the corresponding deformed template model. However, it will be time-consuming if we calculate all the distances between points in the mesh surface and the visual hull. We replace the vertices with the nearest volume data and represent the visual hull to be labeled by boolean values. For each volume, the $k-$neighboring data are detected in turn until it meets with the surface data. Then the segmented body is utilized to extract the articulated pose for the next frame.

2. 2 Visual hull segmentation

The visual hull can be labeled by searching the nearest vertex in the corresponding mesh surface. The segmented model is then utilized in the motion tracking process for the next frame. It is intuitive to use the reconstructed surface to decompose the volumetric model but usually the surface generated by the motion tracking data always exists noise. If we do not remove the noise, the segmentation result will have effect on the motion tracking results and then make the segmentation worse for the next frame. The error will be accumulated. It is easy to locate the positions for hands, feet and the head. At first we relabel these parts according to the bone length and the volume, locate the bone joints and label the father segments again. This method is simple and effective to avoid the accumulative problem. It is always harder to locate the right position for hands and feet then we can learn from Fig. 1(b) that the parts of the hands occupy the ones of the forearms. It is because of the error of motion estimation and in return this result will cause the mis-aligned problem for motion extraction. The error will be removed by refining the segmentation based on the skeleton and volume as shown in Fig. 1(c). We also can see from that parts of the noise data in front of the chest occur as the volume intersection method are not split into the body part of the chest but this kind of noise can be removed in the process of surface reconstruction using the template surface.

3. Sampling

The voxel size is set $1cm$ in this paper. The numbers of the voxels in the visual hulls for an adult are usually over 80 thousands therefore it is time-consuming to use all the data to estimate the 3D posture sequence while tracking. Our proposed method for sampling makes sure that the selected data spread around randomly in the human body segments and the number of the voxels in the sampling model is

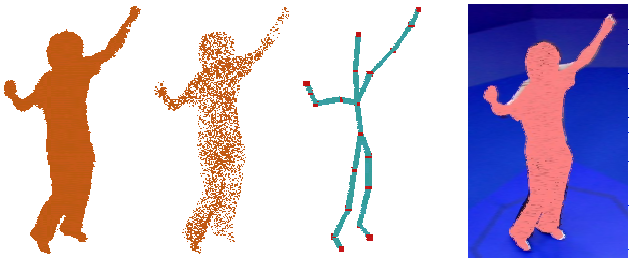
It is easy to obtain the bounding box of a volumetric model Z and we enlarge it to be a cube as it can be divided equally into eight cubes. For each cube C_i, the volume ratio between the voxels in the cube and the cube is used to determine the following process: split or pick samples. In this paper, the process of division will be repeated until the ratio is above 75%. About 2% percent of the voxels in the intersection of Z and the cube are selected while the division of the cube is stopped.

4. Pose extraction

In this section, we describe the motion tracking ap-

(a) (b) (c) (d)

Fig. 2 The template models used for motion tracking and surface estimation. From left to right: the segmented visual hull, the sample of (a), the mesh surface and the skeleton model.



(a) (b) (c) (d)

Fig. 3 The sample model (b) of the visual hull (a) to be tracked is compared with the template sample to extract the posture (c). The deformed surface (d) is generated based on the motion data.

proach that seeks for 3D optimal kinematic chains fitting to each body segment. Various of particles are produced to seek for the optimal one with global optimization and local matching. Self-collision detection is used to guide to produce "real-life" particles.

4. 1 Motion representation

Twists representation and exponential coordinates as given in [13], [14] are employed for expressing the rigid motion of the sample model. It is then restricted to move with articulated constraints in high dimensional configuration spaces of 42 DOF. DOF of the global translation and rotation are treated as six. Wrist, knee and ankle joints are defined with two degrees of freedom. Shoulder, hip, neck and upper body joints are given three degrees of freedom. Then we define the state of the sample by a vector $\chi = (t_1, t_2, t_3, \theta_1, \theta_1, \ldots, \theta_{39})$ that consists of the three parameters of the global translation and 39 rotate angles. The global translation is expressed by the following 4×4 matrix

$$\mathbf{T_G} = \begin{pmatrix} 0 & 0 & 0 & t_1 \\ 0 & 0 & 0 & t_2 \\ 0 & 0 & 0 & t_3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{1}$$

For the joint angle θ_i, the rotate axis ω_i and corresponding joint J_i are known from the skeleton model. The rotation $R(\theta_i)$ is given by

$$\mathbf{R}(\theta_\mathbf{i}) = \begin{pmatrix} e^{\widehat{\omega_i}\theta_i} & (I - e^{\widehat{\omega_i}\theta_i})(\omega_i \times J_i) + \omega_i \omega_i^T J_i \theta_i \\ \mathbf{0} & 1 \end{pmatrix} \tag{2}$$

where $\widehat{\omega_i}$ is the matrix representation of ω_i as described in [15]. The rigid transformation of body segment i is represented by

$$T_i = \prod_{j \in k(i)} R(\theta_j) + T_G \tag{3}$$

4. 2 Self-intersection detection

If a segmented mesh surface with an underlying skeleton model is given, any human motions can be represented by a configuration vector $\chi = \{\mathbf{t}; \theta\}$ that consists of the three parameters of the global translation $\mathbf{t}$ and the rotate angles for all the human body segments. In this paper, wrist, knee and ankle joints are defined with two degrees of freedom. Shoulder, hip, neck and upper body joints are given three degrees of freedom therefore the total degrees of freedom of the human body is 42. The human pose will be decided if the value of χ is given. It is theoretically right that every random chosen of χ is corresponding to a 3D posture but not always a "real" human motion as there is no self-intersection constraints for the state. In this section, a human self-intersection detection method is proposed to move the "bad" particles in which some body segments collide with each other.

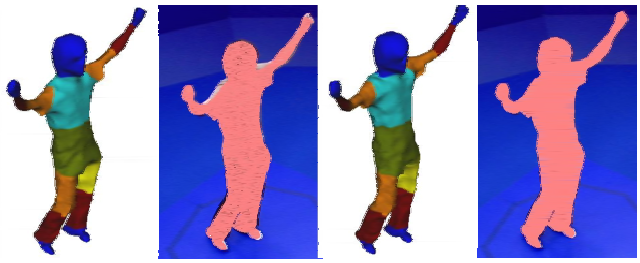
A simple approach is presented to eliminate the unexpected human motion with self-intersection. The sample template volumetric model and the skeleton are utilized to remove the unexpected particles. We take the cylinder to represent each body limb based on the skeleton. The self-intersection weight is defined as following

$$w_{self-inter}(S(\chi)) = \min_i \frac{1}{N} \sum_{vox_j \notin S_i(\chi)} p(vox_j, C_i), \tag{4}$$

where $S_i(\chi)$ is the body segment i and N is the number of the volumetric model $S(\chi)$. The value of $p(vox_j, C_i)$ is 1 if vox_j is in the cylinder C_i, otherwise 0. In the process of producing new particles, we remove the self-crossed ones according to $w_{self-inter}$.

5. Motion estimation with local memorization

In this section, we describe the motion tracking ap-



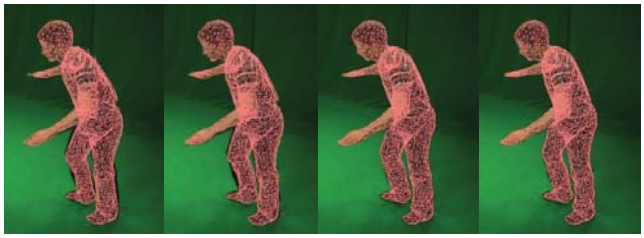
(a) Deformed surface by LBS (b) Estimated surface

Fig. 4 The template is deformed as an initial surface, and the deformed one is projected to compare with the images to estimate the details.

proach that seeks for 3D optimal kinematic chains fitting to each body segment. As introduced in section 1., it is easy to dive into local minimization while tracking human motion in high dimensional spaces, especially for fast motion clips. There are two main reasons. One is that the solution relies on the distribution of the initial particles. The other one is that it produces the new particles for the next layer according the fitness of the whole human body so that it may cause the local misaligned problem. We propose an approach not only annealing to generate the new particles according to their distribution but also combining some particles with local optimization for different limbs that avoids the local minimization problem effectively.

We utilize the 3D data directly to estimate the human postures. As shown in Fig. 2, the template volumetric model (a) is decomposed into 15 body segments and about 2% number of voxels (b) are selected for motion estimation. The mesh model (c) is deformed to generate a sequence of surfaces by the extracted motion data and the silhouettes constraints that will be introduced in section 6..

The voxel in the target sample are labeled due to the index of the nearest voxel in the template sample. The reliability of the correspondences rely on the similarity of the two model. A distance function is provided here to measure the "fitness" for each body segment. The self-intersection problem has been avoid as introduced in section 4. 2. Now the problem is that we want the body segment i of the reference is also near to the same part of the target but usually it cannot be confirmed. Therefore we use a punitive way to calculate the distance. Assume Vox_1 and Vox_2 are the same body segment in the reference and the target respectively, $Vox_1 = vox_{1,1}, vox_{1,2}, \ldots, vox_{1,n_1}$ and $Vox_2 = vox_{2,1}, vox_{2,2}, \ldots, vox_{2,n_2}$, a punitive value is given in the function



(a) (b) (c) (d)

Fig. 5 Surface estimation.

$$Dst(Vox_1, Vox_2) = \alpha + \frac{1}{n_1}\sum_{i=1}^{n_1} \min_j \parallel vox_{1,i} - vox_{2,j} \parallel \quad (5)$$

$$\alpha = \begin{cases} c_1 & n_2 = 0 \\ c_2 \frac{\|n_1 - n_2\|}{n_1 n_2} \sum_{j=1}^{n_2} min_i \parallel vox_{1,i} - vox_{2,j} \parallel & else \end{cases} \quad (6)$$

In our program, we choose $c_1 = 30(cm)$ and $c_2 = 2$ to punish the wrong labeling.Then we deform the template sample in Fig. 2(a) to match the target measuring by a difference function $D(S(\chi), S^*)$ between the deformed sample $S(\chi)$ and the target model S^* as shown in Fig. 3(a)

$$D(S(\chi), S^*) = \frac{1}{15}\sum_{j=1}^{15} Dst(S_j(\chi), S_j^*), \quad (7)$$

where $S_j(\chi)$ is the body segment j and N_j is the number of voxels of it.

The measurements of the quality of particles are combined in a simple way by

$$w(X, Z) = exp - D(S(\chi), S^*), \quad (8)$$

where $S(\chi)$ and S^* are the sample models of X and Z. We also take the idea of annealing particle filter to guarantee the particles to the global optimal solution of the likelihood according to the distribution of the weight function. The modification is that we produce new particles by taking local fitness into consideration in the process. We use $Dst(S_j(\chi), S_j^*)$ as the value of local fitness for each body limb in equation (7), sort them for each body segment and produce new particles by mean combination and articulated combination method.

Assume χ^i captures well for the body segment $i, i = 1, 2, \ldots, 15$, the mean combination method just take the mean values of the useful part of all the particles as described in [5]. In addition, we proposed the articulated combination method by taking the positions of the joints into account. We can get a new skeleton in

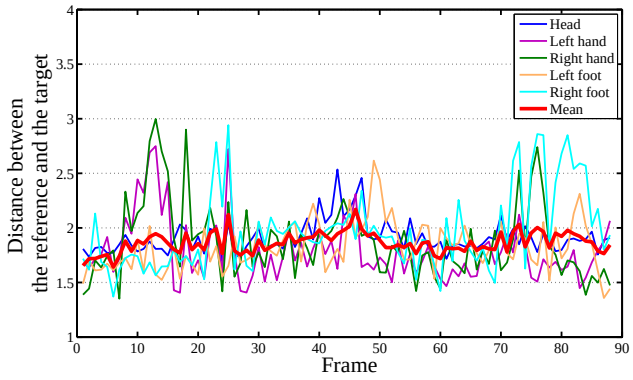


Fig. 6 The distance between the deformed reference model and the target.

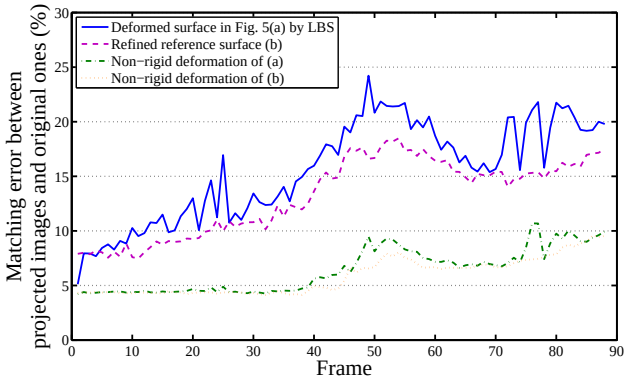


Fig. 7 The matching error between the project images of the reconstructed surface and the original images.

which the joint belongs to the body segment i based on the value of χ^i, and the χ^* is obtained by approaching to the skeleton approximatively. In our experiments, it was found that setting the layer $M = 10$ with the particle number $N = 100$ worked well for human motion tracking.

6. Surface estimation

Blend skinning approaches such as linear blend skinning are widely used in shape reconstruction when 3D skeleton poses are provided. We also use the method proposed by Baran et al. [10] to estimate the skin weights for the vertices of the surface.

6.1 Linear blend skinning

It makes the skin attachment estimation easier as the template model has been segmented into 15 body parts. We replace the heat contribution weight of the nearest bone to the vertex i belonging to the body limb j by the nearest distance from the vertex to the nearest joint in the body segment j as we have segmented the human body into 15 parts. The weights are given by solving the following equation

$$(H - L)w^i = Hp^i \tag{9}$$

p^i is a vector with $p^i_j = 1$ if the vertex j is in the body segment i and $p^i_j = 0$ otherwise. H is the diagonal matrix with $H_{jj} = 1/d(j)^2$ where $d(j)$ is the nearest distance from the vertex j to the bone in the body segment i. The vertex j can be updated as following

$$v^*_j = \sum_{i=1}^{15} w^i_j T_i v_j \tag{10}$$

where T_i is the rigid transformation of the body segment i. It linearly interpolates the vertices if the skin attachment is calculated. The picture shown in Fig.5(a) is the deformed surface using the extracted human motion.

6.2 Silhouette constraints

Vlasic [1] and Gall [2] took the 3D models generated by skinning methods as initial model and iteratively deformed it to recover the surface details. Our process of surface estimation is similar to theirs. We deform the template surface by linear blend skinning method as shown in Fig. 4(a). Non-rigid transformation should also be taken into consideration to recover the surface details. It is intuitive to take the surface points from the visual hulls as constraints. Unfortunately, the visual hull suffer from the noises caused by not only the qualities of the silhouettes but also the reconstruction method, therefore, we turn to deform the mesh surface to match the 2D silhouettes instead of the visual hull. We project the reconstructed model to the original images from different views, estimated the corresponding vertices by silhouette constraints. The refined surface are generated by solving a least-squares optimization problem as follows

$$\underset{v}{argmin}\{\parallel LV - \delta \parallel + \alpha \parallel C_{sil}V - q_{sil} \parallel, \tag{11}$$

where L is the Laplacian matrix and δ are the differential values of the deformed mesh surface with vertices V. C_{sil} is a parameter matrix to express the constraints of the silhouette rims and q_{sil} are the corresponding confined points as described in [1], [2]. In our work, we iterate eight times to refine the mesh surface and set the parameter of $\alpha = 0.01$. We use the deformed surface by linear blend skinning for 3D motion estimation of the next frame. The process is repeated and the time-varying sequence with underlying skeleton chains can be reconstructed.

6.3 Pose refinement and the reference surface estimation

It is hard to ensure the accuracy of the extracted hu-

Fig. 8 The top: the generated surfaces projected to the original images. The bottom: the segmented surfaces.

man posture for each frame for some reasons. Noise exists in the visual hull that is reconstructed by volume intersection method as shown in Fig. In addition, only near 2% of the number of the voxels in the visual hull is selected so that the result of the extracted motion is due to the quality of the selected sample models. In other hand, the extracted motion will be utilized for the next frame so it is important to remove the error.

7. Experimental results

The initial template surface G is utilized to estimate the reference surface G_t of the frame t in order to avoid the accumulative problem. In the first step we calculate the transformation between G and the deformed model G_t^* generated as introduced in section 6. 2. For each body segment i, the problem is to get the optimal solution T_i between the template body segment G^i and G_t^{*i} that minimize the following least squares criterion which has been described in [16] [17]

$$\sum_{v_j^i \in G_i, v_j^{*i} \in G_i} \|v_j^{*i} - T_i v_j^i\|^2. \qquad (12)$$

We take the segmented volumetric sample model to match the target again by iterative closest points (ICP) registration to smooth the error caused by mesh deformation and the optimal solution of equation 12. The extracted transformation for each body limb provides a good initial estimation for the ICP algorithm. The template surface is deformed to generate the reference surface for the frame t by the LBS method. It is seen in Fig. 5 that the refined surface (b) matches better with the original image than the deformed surface (a) by LBS. Fig. 5 (c)(d) show the generated surfaces by non-rigid deformation of (a)(c) respectively.

We use the public dataset provided by Gall et al. [2] and NHK lab to carry out our experiments. The purpose of our approach is to extract 3D articulated kinematic chains directly from a time-varying volume sequence. In our program, the property of binary representation of volumetric model makes it easy to compare models and compute distances.

The distance for each body segment and the mean value of the whole body between the deformed reference model and the target is estimated due to equation 57. We show the distances of the head, hands and feet as it is harder to track the pose for these body segments than others. All the distances are less than $3cm$ and the mean distance is about $2cm$ so that the error can be erased by non-rigid deformation effectively. This result shows the ability of our proposed method for ensuring the global optimization and local matching.

We project the reconstructed surfaces to generate 2D images and compare them with the original images. The matching error for the deformed surface and the refined reference model increase as shown in Fig. 7 since these two kinds of models are all generated based on the template model of the initial frame and the extracted human motion by LBS method. The non-rigid changes will increase. The mean matching error is about 7% after non-rigid deformation and do not increase obviously for about 45 frames. The refined surfaces fit better with the original images. Unfortunately, the matching error

is still increase. The reason is that the error between the reference model and the images increase although we refine the reference surface which is utilized for the next frame. The quality of the reference model have effect on the final results. In Fig. 8 we show the reconstructed surfaces.

8. Conclusion

We proposed a model-based motion tracking method to estimate the articulated motion property for a volume sequence directly in the 3D space. Several postures that match well for different human body segments are selected to produce a new particle which captures well with the whole human body. Although the accuracy relies on the quality of visual hulls, our tracking method works well for fast human motion tracking in high dimensional configuration spaces as compared to other methods such as the APF algorithm. We generate the deformed surface using the motion capture data by LBS method and then match it to 2D silhouettes constraints to produce a sequence of surfaces with the same topology. The posture and surface are refined again by registration method.

References

[1] D. Vlasic, I. Baran, W. Matusik and J. Popović: "Articulated mesh animation from multi-view silhouettes", ACM Trans. Graph., **27**, pp. 97:1–97:9 (2008).

[2] J. Gall, C. Stoll, E. de Aguiar, C. Theobalt, B. Rosenhahn and H. P. Seidel: "Motion capture using joint skeleton tracking and surface estimation", IEEE Conf. Computer Vision and Pattern Recognition, pp. 1746–1753 (2009).

[3] J. Deutscher and I. Reid: "Articulated body motion capture by stochastic search", Int. J. Comput. Vision, **61**, 2, pp. 185–205 (2005).

[4] J. Gall, B. Rosenhahn, T. Brox and H.-P. Seidel: "Optimization and filtering for human motion capture", Int. J. Comput. Vision, **87**, pp. 75–92 (2010).

[5] W. Luo, T. Yamasaki and K. Aizawa: "Articulated human motion capture from segmented visual hull and surface reconstruction", APSIPA ASC 2010, pp. 109–116 (2010).

[6] I. Mikić, M. Trivedi, E. Hunter and P. Cosman: "Human body model acquisition and tracking using voxel data", Int. J. Comput. Vision, **53**, 3, pp. 199–223 (2003).

[7] W. E. Lorensen and H. E. Cline: "Marching cubes: A high resolution 3d surface construction algorithm", SIGGRAPH '87, pp. 163–169 (1987).

[8] R. Tadano, T. Yamasaki and K. Aizawa: "Fast and robust motion tracking for time-varying mesh featuring reeb-graph-based skeleton fitting and its application to motion retrieval", IEEE International Conference on Multimedia & Expo, pp. 2010–2013 (2007).

[9] N. S. Lee, T. Yamasaki and K. Aizawa: "Hierarchical mesh decomposition and motion tracking for time-varying-meshes", IEEE International Conference on Multimedia & Expo, pp. 1565–1568 (2008).

[10] I. Baran and J. Popović: "Automatic rigging and animation of 3d characters", SIGGRAPH '07, New York, NY, USA (2007).

[11] P. J. Besl and H. D. Mckay: "A method for registration of 3-d shapes", Pattern Analysis and Machine Intelligence, IEEE Transactions on, **14**, 2, pp. 239–256 (1992).

[12] S. Corazza, L. Mündermann, E. Gambaretto and T. P. Andriacchi: "Automatic generation of a subject specific model for accurate markerless motion capture and biomechanical applications", IEEE Trans. on Biomedical Engineering, **57**, 4, pp. 806–812 (2010).

[13] C. Bregler, J. Malik and K. Pullen: "Twist based acquisition and tracking of animal and human kinematics", Int. J. Comput. Vision, **56**, pp. 179–194 (2004).

[14] C. Bregler and J. Malik: "Tracking people with twists and exponential maps.", CVPR'98, pp. 8–15 (1998).

[15] R. M. Murray, Z. Li and S. S. Sastry: "A Mathematical Introduction to Robotic Manipulation", CRC Press, Ann Arbor (1994).

[16] D. W. Eggert, A. Lorusso and R. B. Fisher: "Estimating 3-d rigid body transformations: a comparison of four major algorithms", Mach. Vision Appl., **9**, pp. 272–290 (1997).

[17] S. Pellegrini, K. Schindler and D. Nardi: "A generalisation of the icp algorithm for articulated bodies.", BMVC (Eds. by M. Everingham, C. J. Needham and R. Fraile), British Machine Vision Association (2008).