Voxel Data Based Marker-less Human Motion Capture Using Geometry Model

Wuqin Lei[*] and Jiangbin Zheng[*]
[*] Shaanxi Provincial Key Laboratory of Speech and Image Information Processing,
School of Computer Science, Northwestern Polytechnical University, Xi'an
E-mail: leiwuqin@mail.nwpu.edu.cn, zhengjb@nwpu.edu.cn

***Abstract* — In this paper, we propose a novel approach for human posture estimation using geometry model. A volumetric reconstruction of a participant is obtained from multi-camera images. After definition of body model, the geometry model is fitted into the 3D human reconstructed volume. The gray theory, which is applicable to the prediction problem of a time-varying nonlinear system, is utilized to perform the forecasting job during tracking. Moreover, a hierarchical estimation is applied to avoid local optimum. Finally the posture is gained from the geometric parameter. Experimental results show the efficiency of the proposed algorithm and precision of posture estimation.**

I. Introduction

With the demands and potential economic values of motion capture techniques in various fields such as manufacture of character animation in film, video games development, virtual reality applications, athletic training and biomechanical analysis, many studies on human body posture estimation have been undertaken [1- 4]. It has become one of the hottest research aspects.

Many approaches have been proposed for human motion capture, which can be categorized into two major classes: mark-based and marker-less. Matthias Weber [5] introduces a passive mark-based optic motion capture system with tree-based optimization and a neural network, which is used for fitting and mapping markers. Although, such a hybrid approach achieves quite robust tracking results, its wiring of markers causes inconvenience of natural motion. Moreover, the marker may disappear when obstacles occur, and thus the tracking may be interrupted. Besides, as markers are attached on the clothes of participant, the extracted data cannot represent the real skeleton of human body. Therefore, marker-less motion capture is highly in demand.

With respect to marker-less motion capture, some kinds of visual cues from video images have been employed for human posture estimation [6 - 12]. There are some approaches on video images and voxel data. Several convincing results have been achieved especially in 2D estimation [6]. But for wider application, instead of processing directly with 2D images, human posture estimation method has been increasingly focused on using voxel data reconstructed from multi-view images.

The model-free approaches obtain a body posture from voxel data directly. T. Tung et al. [7,8] utilize a Reeb Graph based method to get topology of objects including human body. It is established using a μ function defined over the object surface considered as a compact manifold. The surface of the object is divided into regions according to the value interval of μ, and the node is associated to each connected component of the regions. However, results obtained by Reeb Graph often depend on human body postures when it applies to the reconstructed voxel data, and because of the heavy computation, it is not feasible for real-time processing.

Most marker-less motion capture approaches are model-based, either on skeleton model or geometric model. The former uses stick to represent body skeleton and ignores the dimension. J.Gall et al. [10] recover the movement of the skeleton and deformation of the 3D surface through an articulated template model and multi-view silhouettes. The methods based on geometric model adopt blobs, spheres, cylinders, ellipsoids, or other geometrical primitives to denote human body volume [11, 12]. Mikic et al. [4] propose a human body model that consists of ellipsoids and cylinders and use the twists framework to describe it, resulting in a non-redundant set of parameters. F.Caillette[9] presents the use of 3-D Gaussian blobs for tracking individual body parts. They design the blob-fitting procedure to exploit the hierarchical structure of the reconstructed volume using both position and color. In [3, 12], a human body model flashed out by 11 cylinders is involved and the framework of PEA is utilized during tracking process.

After comparing with these aforementioned methods, we present a novel voxel based marker-less motion capture approach. It contributes with precise estimation and lower computation as we only use the cloud points of body surface.

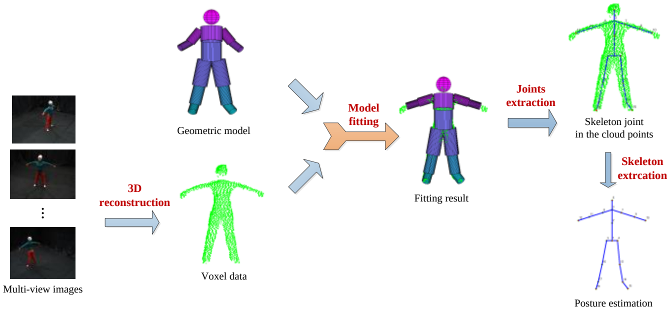

Fig.1 The method flowchart

Our method flowchart is shown in Fig.1. The main components are 3D voxel reconstruction, model definition, model fitting and human body tracking, which would be discussed in the following. The skeleton parameters can be recovered from the proposed geometry model.

The rest of the paper is organized as follows. Section II refers to the definition of body model. Section III describes the proposed model-based human tracking algorithm. Section IV presents the experimental results. Finally, the conclusion follows in section V.

II. DEFINITION OF BODY MODEL

As the human body is very complicated, it is necessary to simplify it. In our experiments, we define two kinds of body models: a skeleton model that represents the articulated structure and a geometrical model that represents the surface of the body, as shown in Fig.2.

The geometrical model consists of 12 cylinders which represent the torso and limbs, and one sphere that represents the head. We define three parameters for each geometrical component, including the radius, angles of horizontal and vertical rotation.

The skeleton model includes one root and 17 joints: head, neck, breast, left and right shoulders, elbows, hands, hips, knees, ankles, feet. All these joints are organized into a hierarchical structure, as shown in Fig.3. The root is at the top and it has three child nodes: left hip, breast and right hip. Each non-leave node has its own child nodes. When the parent joint angles are changed, their child joints change accordingly.

Once the affiliations of joints, the skeleton length and radius of cylinders and sphere are initialized, they will not be changed during motion.

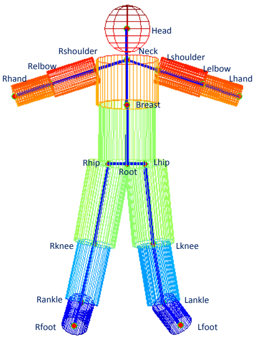

Fig. 2 Skeleton model and geometry model

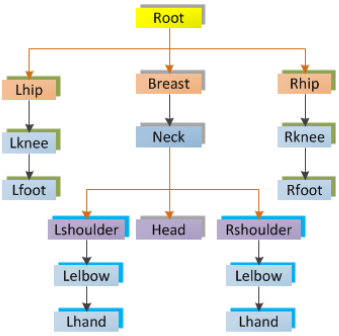

Fig. 3 The joints hierarchical structure

III. 3D SKELETON ESTIMATION

A. Model Fitting

We consider the parent node as the center of a sphere, the skeleton length as the radius. θ ($0<\theta<\pi$) is the angle between radius and positive Z-axis. ψ ($0<\psi<2\pi$) is the angle between the projection of radius in X-Y plane and positive X-axis. So the child node is a point at the angle $<\theta,\psi>$ of the sphere. Each joint location $p(x_i,y_i,z_i)$ is obtained by (1) (2) (3).

$$x_i = r*\sin\theta*\cos\psi \tag{1}$$

$$y_i = r*\sin\theta*\sin\psi \tag{2}$$

$$z_i = r*\cos\theta \tag{3}$$

Where r is the skeleton length.

We denote the skeleton model by $\boldsymbol{X}$. The mathematical expression of $\boldsymbol{X}$ can be written as follows:

$$X=\{x,y,z,\theta_1,\theta_2,\ldots\theta_{17},\psi_1,\psi_2,\ldots\psi_{17}\} \tag{4}$$

Here, x,y,z are the coordinates of the root, $\theta_1,\theta_2,\ldots\theta_{17}$ and $\psi_1,\psi_2,\ldots\psi_{17}$ are the child angles defined above. Assume that X^t is the model of time instance t , the model at time instance $t+1$ X^{t+1} can be written as:

$$X^{t+1}=X^t+\Delta X \tag{5}$$

Fitting the cylindrical model to the reconstructed human body in 3D volume that is a set of voxels can be transformed into an optimization problem. The skeleton joints can be obtained as centers of upper and lower bases of cylinders. The angles of cylindrical model and skeleton model satisfy the following constraints:

$$\theta=\beta+\pi/2 \tag{6}$$

$$\psi=2\pi-\alpha \tag{7}$$

α , β are the cylinder parameters for horizontal and vertical rotation angles respectively.

Since the 3D human voxel data has no other information as texture or color, only the coordinate positions of points can be used. We count the number of voxel points that fall into the corresponding cylinder, and the marching function is defined as below:

$$f(\Delta x)= fitness(\alpha,\beta)=\max(\sum_{i=1}^{N}Vi) \tag{8}$$

Here N is the number of voxel data. If the point fall into the cylinder, let $V_i=1$, otherwise, $V_i=0$.

B. Gray Prediction for Human Body Tracking

There are many branches of gray system theory such as gray evaluation theory, gray cluster theory and etc. In this paper, we adopt GM (1, 1)[13] which is a kind of single variable and first-order linear dynamic gray model to perform the motion prediction.

Suppose that the original series of data about samples is defined as:

$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots x^{(0)}(n)) \tag{9}$$

The new sequence $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \ldots x^{(1)}(n))$ can be generated with the gray prediction algorithm [13].

During the human body tracking, if we try to estimate all the motion parameters simultaneously, it may be trapped into local optimum. According to the hierarchical structure, the estimation begins with the root, then its child nodes and their descendants. The estimation procedure can be described as follows:

```
Repeat
    Step 1) Utilize GM (1, 1) to obtain the current joint's prediction sequence $X^{(t)} = (x^{(t)}(1), x^{(t)}(2), \ldots x^{(t)}(n))$.
    Step 2) Search in the predictable range to optimize (8) to find the appropriate parameters $\alpha, \beta$ of corresponding cylinder.
    Step 3) Convert cylinder parameters $\alpha, \beta$ to skeleton angles $\theta, \psi$ by (6) (7).
    Step 4) Compute the current joint coordinates and its descendants through (1) (2) (3).
    Step 5) Update the gray prediction model's original series with the new data.
Until
    Accomplish all the joints traversal.
```

C. *3D Skeleton extraction*

We obtain each joint 3D position through the algorithm mentioned above. Each node has been connected with its parent node or child node. According to the articulated information, the joint nodes are linked by sticks to form a 3D skeleton of the human body.

IV. Experiments

A. *Experimental setup*

To evaluate the efficiency of the proposed human body posture estimation method, 3D reconstruction and 3D skeleton extraction experiments are carried out.

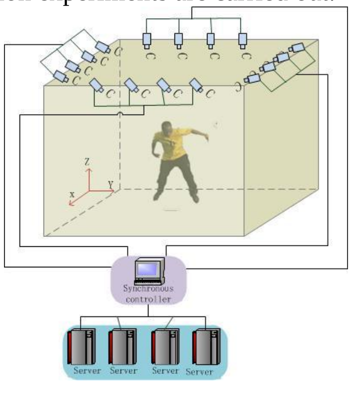

Fig. 4 Composition of multi-camera system

In order to acquire multi-view images for 3D reconstruction, we establish a multi-camera system as shown in Fig.4. The experimental environment is a 7m *7m indoor scene with 16 cameras surrounded on the ceiling. The image from each camera is digitized into a computer connected to it. A special controller is employed to ensure the camera synchronous acquisition. Experiments are conducted on a PC with a 2.93GHz CPU and 2G memory. The 3D reconstruction and 3D skeleton extraction algorithm are implemented by c++ and matlab respectively.

B. *3D reconstruction of human volume*

In our experiments, the voxel data is obtained by an improved photo-consistency method [14]. Synchronized 16 images with a 648 by 490 pixel resolution via IEEE 1394 interface are stored in four servers. The frame rate is 30fps. 3D volume of the actor in front of cameras is reconstructed from multi-view images. The reconstructed size of voxel is 5mm*5mm*5mm.Only the surface data are kept. Fig.5 shows the results of reconstructed voxel data from three viewpoints.

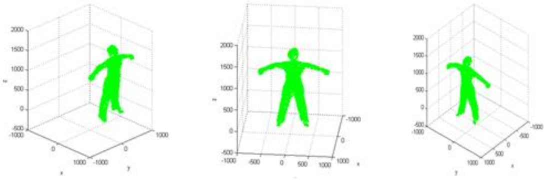

Fig. 5 Reconstructed volume from different viewpoints

C. *Skeleton extraction results*

In the proposed method, we use gray model for prediction, as it is flexible to deal with motion change. We do the comparison between prediction data and actual data. We list part of the joint results in table I. The second and third columns are the actual and prediction data of each node's x y coordinates respectively. The relative error is presented in the last column. From the table, we can see that the minimal value is 0.0268, and the maximal is 0.1638. Most of the relative errors are on the verge of zero. And the gray model fits well in the algorithm.

TABLE I

Comparison between prediction data and actual data

Joint node	Actual data	Prediction data	Relative error
Breast	578.5097	551.6524	0.0464
	79.9979	69.6654	0.1292
Neck	607.8455	576.5353	0.0515
	105.8609	103.0236	0.0268
Lshoulder	900.4588	1039.3	0.1542
	-151.8889	-167.8286	0.1049
Lelbow	511.4226	486.0674	0.0496
	319.4167	334.6454	0.0477
Lhip	851.4484	712.0201	0.1638
	89.5509	104.0041	0.1389
Lknee	511.4226	486.0674	0.0496
	-111.4040	-97.3393	0.1262

The human posture estimation is a major task in our experiment. The subject's action is limb movements. We take out several frames of a sequence estimation results as shown in Fig.6. The first row is the front view images of the multi-cameras. The second row shows the geometry model after fitting to the voxel data. And the next row is the joint

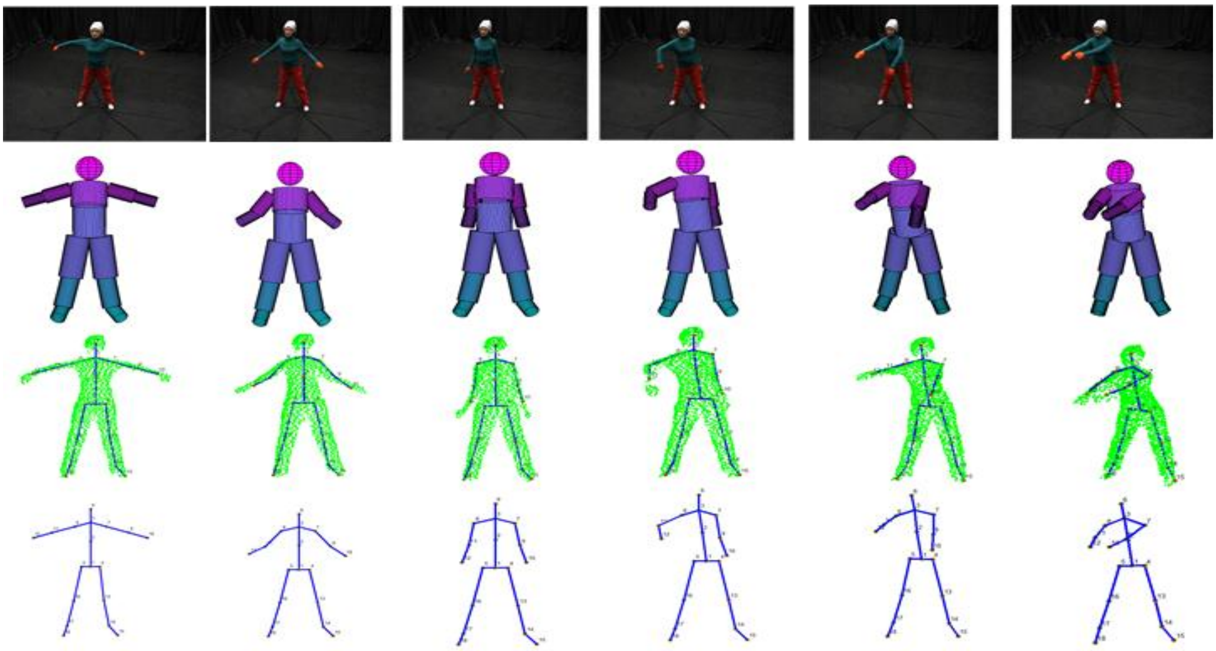

Fig.6 Results of motion estimation at frame 287, 308, 313, 332, 361

extraction from the 3D points cloud. In the fourth row, we give out the final human body motion estimation. For convenience, there is a sign beside each joint that numbered from 1 to 18. From the results, we can see that the algorithm can do a good estimation.

V. Conclusions

This paper investigates a novel marker-less motion capture algorithm based on voxel data reconstructed from multi-camera images. The simple geometry model is utilized to fit the 3D volume. The prediction is performed using GM (1, 1). Furthermore, hierarchical estimation is used to avoid local optimum efficiently and enables accurate 3D skeleton extraction. The presented experimental results illustrate the validity of the approach.

In the future, we plan to pay more attention to automatic initial model establishment, motion prediction and the accuracy for the localization of joints. Last but not least, a better model may facilitate the tracking process greatly. Application of the extracted skeleton data to drive CG model may be another important work.

Acknowledgment

This paper is supported by the National Natural Science Foundation of China (No. 60970069) and the graduate starting seed fund of Northwestern Polytechnical University (No.Z2011135).

References

[1] J. Gall, C. Stoll, E. de Aguiar, et al. “Motion capture using joint skeleton tracking and surface estimation”, the *Proc. of the CVPR*, 2009, pp. 1746-1753.

[2] R. Kehl, M. Bray, L. Van Gool, “Full body tracking from multiple views using stochastic sampling”, in *Proc. of the CVPR*, 2005, pp. 129-136.

[3] J. Yan, Y. Li, E. Zheng, Y. Liu, “An accelerated human motion tracking system based on voxel reconstruction under complex environments”, in *Proc. of ACCV*, 2009, pp. 313-324.

[4] I. Mikic, M. Trivedi, E. Hunter, “Human body model acquisition and tracking using voxel data”, *IJCV*, 2003, Vol.53, pp.199-223.

[5] M. Weber, “A hybrid approach towards fully automatic 3D marker tracking”, in *Proc. of the ACM symposium on VRST*, 2008, pp.243-244.

[6] X. Zhang, C. Li, X. Tong, ”Efficient human pose estimation via parsing a tree structure based human model”, in *Proc. of the ICCV*, 2009, pp.1349-1356.

[7] T. Tung, F. Schmitt, T. Matsuyama, “Topology matching for 3D video compression”, in *Proc. of the CVPR*,2007, pp.1-8.

[8] T. Tung, F. Schmitt, “Augmented reeb graphs for content-based retrieval of 3D mesh models”, in *Proc. of the SMI*, 2004, pp.157-166.

[9] F. Caillette,”Real-time markerless 3-D human body tracking”, *PHD thesis, University of Manchester*,2007

[10] K. Ogawara X. Li, K. Ikeuchi, “Marker-less human motion estimation using articulated deformation model”, in *Proc. of ICRA*, 2007, pp. 46-51.

[11] K. Takahashi, Y. Nagasawa, M. Hashimoto, “Remark on markerless human motion capture from voxel reconstruction with simple human model”, in *Proc. of the IROS 2008*, pp. 755-760

[12] S. Shen, H. Deng, Y. Liu, “Probability evolutionary algorithm based human motion tracking using voxel data”, in *Proc. of the CEC*, 2008, pp.44-49.

[13] J. M. Jou, Y. Shiau, P. Chen, S. Kuang,”A low-cost gray prediction search chip for motion estimation”, *IEEE Trans. on circuits and systems*, 2002,vol.49,No.7,pp.928-938

[14] C. Ning, L. Xiuxiu, Z. Jiangbin, “3D voxel reconstruction under multiple constraints in multi-view environment”, *Journal of Computer Application*, 2011,Vol.31,No.2,pp.344-346