Visual Comput (2008) 24: 373–380
DOI 10.1007/s00371-007-0195-7

ORIGINAL ARTICLE

Chengkai Wan
Baozong Yuan
Zhenjiang Miao

Markerless human body motion capture using Markov random field and dynamic graph cuts

Published online: 8 January 2008

Chengkai Wan (✉) · Baozong Yuan · Zhenjiang Miao
Institute of Information Science, Beijing Jiaotong University, Beijing 100044, P.R. China
wanchengkai@gmail.com

Abstract Current vision-based human body motion capture methods always use passive markers that are attached to key locations on the human body. However, such systems may confront subjects with cumbersome markers, making it difficult to convert the marker data into kinematic motion. In this paper, we propose a new algorithm for markerless computer vision-based human body motion capture. We compute volume data (voxels) representation from the images using the method of SFS (shape from silhouettes), and consider the volume data as a MRF (Markov random field). Then we match a predefined human body model with pose parameters to the volume data, and the calculation of this matching is transformed into energy function minimization. We convert the problem of energy function construction into a 3D graph construction, and get the minimal energy by the max-flow theory. Finally, we recover the human pose by Powell algorithm.

Keywords Motion capture · Dynamic graph cuts · Markov random field

1 Introduction

The purpose of human body motion capture is to detect and record the motion of a moving human body, which can be represented as poses of the human body and converted to abstract digital format. Human body motion capture is invaluable for applications such as computer animation, activity recognition, new-generation human-computer natural interaction, game production and motion analysis, etc. Existing human body motion capture technologies are centered around three main approaches: optical, magnetic and electro-mechanical. These three approaches need specific equipments and are restrictive to some degree. Markerless human body motion capture is a method that uses the images obtained from multiple cameras placed around the human body without markers to estimate the pose of the human body. Due to the advantages of vision-based markerless human body motion capture, such as non-compelling, low cost, high automaticity, it has been an increasingly hot research direction in motion capture field. [14] conducted a summary of markerless-based human motion capture. It considers motion capture as several stages: initialization, tracking, pose estimation and recognition. Each stage is divided into different types of concrete. [15] presents a method deducing the 3D pose or motion of the complete human body from a single image or a monocular sequence of images. It uses a learning-based approach to construct a probabilistic pose estimation model from a set of labeled human silhouettes. But so far, because of the high-dimension (24 degrees in our method in this paper) of the kinematic model, occlusion and self-occlusions, how to acquire robust pose information of human body from image sequences independent of special equipments and markers in the presence of image noise, loose clothing and cluttered background remains to be a challenging issue in the field of computer vision.

In recent years, research has shown that multi-camera approaches are used more and more used human motion capture. Many multi-view approaches for markerless human body motion capture have been published lately, such as [4], which tracks full human-body using markerless multi-view images as input. They use blobs attached to a kinematic model to recover joint angles in an expectation-maximization framework. [6] gives a novel approach for full body pose tracking from multiple views via stochastic sampling. The objective function definition is the sum of the distances of model vertices to the corresponding reconstruction voxels. It uses SMD (stochastic meta descent) to optimize the function considering color information. [7] introduces a shape-from-silhouette method for full body tracking from both silhouette and color information. It uses colored surface points to segment the hull into rigidly moving body parts and takes advantage of the constraint of equal motion of parts at coupling joints to estimate joint positions. [16] presents an approach for model-free markerless motion capture of articulated kinematic structures. It uses isomaps to transform the voxel space to its pose-invariant intrinsic space representation and obtains a skeleton representation.

In this paper, we present a new solution for model-based markerless human body motion capture from multiple calibrated cameras. We compute a volume data (voxel) representation from the images using the method of SFS, and consider the volume data as a MRF. Then we fit a predefined human body model with pose parameters to the volume data, the calculation of this fitting is transformed into energy function minimization. We convert the problem of energy function construction to a 3D graph construction, and get the minimal energy by the max-flow theory. Finally, we recover the human pose by the Powell algorithm.

2 Preliminaries

In this section we provide a general overview of the Markov random field (MRF). The MRF is a kind of conditional probability model that can be used to describe the correlation between adjacent areas in image processing. This method has been successfully used to resolve the problems such as image segmentation [8, 17, 20, 21] in recent years. In this paper, we consider the volume of interest as a 3D-MRF, and consider the human body reconstruction as a 3D segmentation in this 3D-MRF. For a MRF is corresponding to an energy function, we can convert the problem of human body reconstruction to an energy minimum problem.

A MRF comprises of an undirected graph $G = \langle V, E\rangle$ where V is a finite set of vertices and $E \subset V \times V$ is a set of edges. The vertices consist of a set of discrete random variables $S = \{s_1, s_2, ..., s_n\}$ defined on the index set V and a label set $L = \{l_1, l_2, ..., l_n\}$ of all possible labels. Each variable s_v takes a value l_v from the label set. C is the clique set of the MRF. Clique is the special set, in which each vertex is the neighbor of others or just one vertex. Then, $y = \{s_v = l_v | v \in V, l_v \in L\}$ will represent the configuration to the MRF. [18] gives the probability density of the MRF, which can be written in terms of a Gibbs distribution as:

$$p(y) = \frac{1}{Z} \exp\left(-\frac{1}{T} E(y)\right) \tag{1}$$

$$E(y) = \sum_{c \in C} E_c(y), \tag{2}$$

where $E(y)$ is the energy function, $E_c(y)$ is the potential energy function of the clique c. T is a constant, and Z is a normalized constant, where $E(y)$ can be written in terms of unary and pair-wise energy terms in the simplest interesting case as:

$$E(y) = \sum_{v \in V} h(s_v) + \sum_{(u,v) \in E} g(s_u, s_v). \tag{3}$$

In the context of the 3D human body reconstruction, the set S corresponds to the set of all voxels in the volume of interest, the variable s_v denotes the labeling of the voxel $v \in V$. The label set L comprises of two labels ('obj', 'bkg') representing whether the voxel belongs to the human body or not. Therefore each configuration y will represent a result of 3D reconstruction of the human body. Supposing z denote the set of observed data in current frame and taking a Bayesian perspective, we wish to find the best configuration y (i.e., the optimal labels for the voxel in current frame) which maximize the posterior probability $p(y|z)$, or in other words, to solve MAP-MRF problem. This can be done by finding the configuration with the minimum energy:

$$y_{\mathrm{opt}} = \arg\min_y E(y). \tag{4}$$

Then inference on the optimal labels corresponding to the MRF is seen as an energy minimization problem.

3 Algorithm details

3.1 System overview

In [14], Moelund and Granum describe the problem of motion capture into four stages: initialization, tracking, pose estimation and recognition. But different researchers have different opinions on whether the human body motion capture should include the person's behavior and identity recognition. Owing to recognition is a very big research direction, we think that the human body motion capture can halt at "pose estimation" and may not include the stage of recognition.

Figure 1 shows the flowchart of our algorithm. The input of our algorithm consists of successive frames from

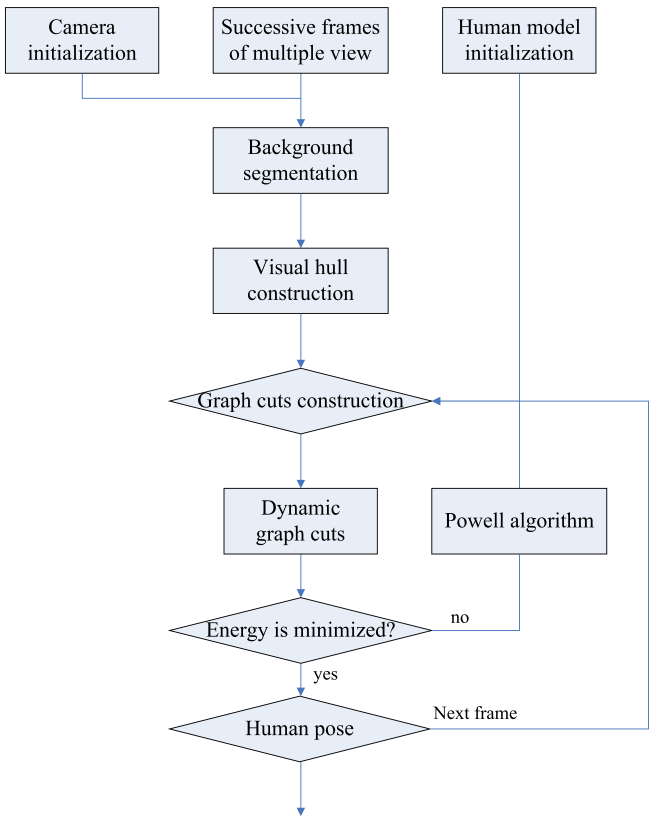


Fig. 1. Illustrated flowchart of our approach

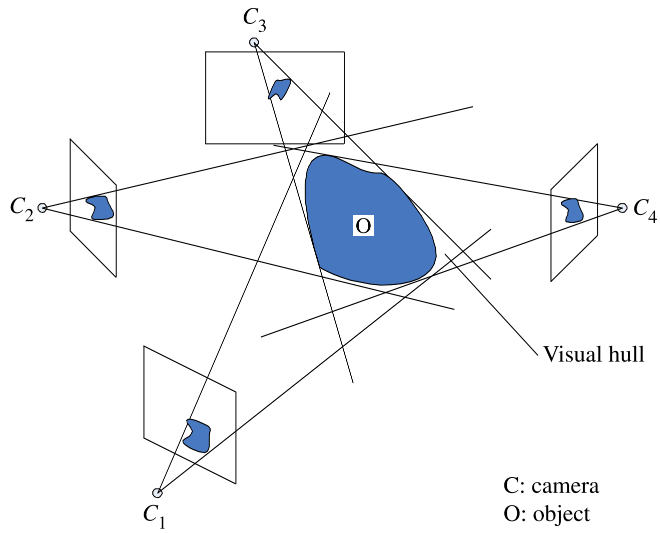


Fig. 2. Sketch map of visual hull [23]

multiple cameras which are static and synchronized by a control system. The output consists of the pose parameters of human body in the image sequence.

We use five conventional fixed USB cameras to input the images without any other special hardware. We put the cameras all around the subject, pointing towards the centre of the capture space. These cameras are synchronized by our program and the images are recorded at a frame rate of 20 fps. As for other computer vision systems, the parameters of the cameras need to be known in the initialization. So the first step of our algorithm is camera calibration. We use the algorithm proposed by [2]. In this step, we can get the intrinsic and extrinsic calibration parameters of cameras under the same world coordinate reference frame.

We use the visual hull of the human body as a basis of our algorithm. The visual-hull of a human is the maximal portion of the space which, projected into the camera image planes, lies totally inside all the silhouettes of the human. So, after acquisition, all images undergo a foreground/background segmentation as a preliminary step before constructing the human body visual hull. We have proposed a new foreground segmentation method for applications using static cameras. It formulates background segmentation as an energy minimization problem. [1] describes the details of the algorithm which eliminates shadows and produces good quality silhouettes.

Next, we recover the human pose from the visual hull. Although more and more researchers estimate the human pose from the 3D visual hull data, most of them take no account of the impact of the errors in 3D data reconstructed from the images using SFS. As we know, because of the occlusion, self-occlusion and image segmentation error, visual hull generated from SFS often involves some inaccurate parts, this problem is particularly serious when the camera number is less. Unfortunately, the 3D matching process largely depends on the accuracy of the 3D data. If the 3D data contain serious errors, it will greatly influence the accuracy of the estimated posture. We take full account of the impact of the SFS reconstruction error to the motion capture and propose a new method which is based on 3D-MRF and 3D dynamic graph cuts [13]. We regard the volume of interest as a 3D-MRF, and consider the human body reconstruction as a 3D segmentation in this 3D-MRF, then using dynamic graph cuts algorithm we can get a min-cut of this MRF. Considering the cost of this min-cut as energy, we can get the cost of this reconstruction. Minimizing this energy, we can recover the human pose.

3.2 Human body model

Human body models can greatly simplify the pose estimation and also make the pose estimation more accurate and robust. Although complicated model can improve the calculation accuracy a little bit. We design a new kinematic human body model to balance complexity and the degree of reality. In our model, each node represents a stylized joint position and the lines between nodes represent bones.

Our model is parameterized with 24 degrees of freedom. Figure 3 shows the human model used in our algorithm.

We use the size of statistical average as the default size of each part in the model. In addition to make the model applicable to various performers, we can use an interactive graphical user interface to adjust the size parameters.

In addition to geometry, the more important parameters of the model are the pose parameter x. In our algorithm, the pose parameter x is a 24-dimensional vector, it contains the three-dimensional spatial location or towards of the components of the human body.

3.3 Human body motion capture via MRF

If we use visual hull of a human and a human model as input I to reconstruct the human body, and consider each voxel in the volume of interest as a node in the 3D-MRF, and each vertice corresponds to a discrete random variable with two labels (‘obj’, ‘bkg’), representing whether the voxel belongs to the human body or not, then any three-dimensional reconstruction results generated from the input data I correspond to the incident of the maximum posteriori probability in a certain MRF. In other words, it corresponds to the incident of energy function getting smallest result.

$$y_{\text{opt}} = \arg\max_y p(y|I) = \arg\min_y E(y). \tag{5}$$

We add the hidden pose variable x to the energy function $E(y)$ of the MRF. For different x, although the forms of energy function $E_x(y)$ unchanged, coefficient changes with x.

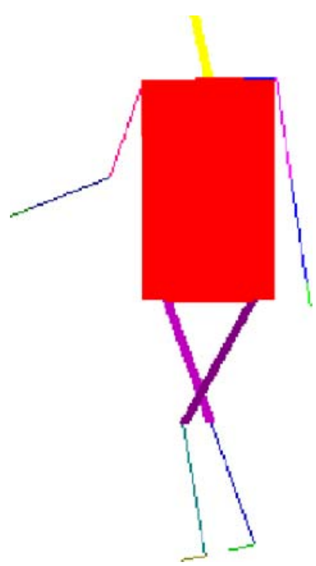


Fig. 3. A simple stick-model of human body

For a given posture x, we can get the best posture reconstruction y, and it corresponds to the minimum of the $E_x(y)$. If we keep adjusting the posture human model x to find the x_{opt} that corresponds to the minimum of the energy function $E_x(y)$, the posture x_{opt} is the optimal posture.

$$y_{\text{opt}} = \arg\min_y E_x(y) \tag{6}$$

$$x_{\text{opt}} = \arg\min_x E_x(y_{\text{opt}}) \tag{7}$$

$$x_{\text{opt}} = \arg\min_x \min_y [E_x(y)]. \tag{8}$$

3.4 Energy function construction

After the analysis of the MRF model method, we convert the problem to an energy minimization. So we will construct the energy function $E_x(y)$ of the MRF. In frame k, the energy function will be decided by the visual hull of the frame h_k, the hidden pose variable x_k and the configuration y. So the energy can be written as $\psi(x_k, h_k, y)$.

According to (3) and in terms of individual and pairwise interaction function, the energy function can be written as:

$$\psi(x_k, h_k, y) = \sum_{i \in V} (\phi_1(s_i|x_k, h_k) + \phi_2(s_i|x_k, h_k) + \sum_{j \in V, j \neq i} \phi_3(s_i, s_j|x_k, h_k)), \tag{9}$$

where $\phi_1(s_i|x_k, h_k)$ and $\phi_2(s_i|x_k, h_k)$ are the unary terms specifying the cost for assigning the label s_i to the corresponding voxel. $\phi_3(s_i, s_j|x_k, h_k)$ is the interaction term that add the smoothness restriction to the space.

Given a visual hull h_k which is decided by the observed data of current frame z_k, we can define $\phi_1(s_i|x_k, h_k)$ as:

$$\phi_1(s_i|x_k, h_k) = \begin{cases} C_1 * (1 - u(i)) & \text{if } s_i = \text{‘}obj\text{’} \\ C_1 * u(i) & \text{if } s_i = \text{‘}bkg\text{’} \end{cases}, \tag{10}$$

where C_1 is a constant, and $u(i)$ can be calculated by:

$$u(i) = \begin{cases} 1 & \text{if } s_i \in h_k \\ 0 & \text{otherwise} \end{cases}. \tag{11}$$

To the human model with certain pose parameter x_k, we define $\phi_2(s_i|x_k, h_k)$ as:

$$\phi_2(s_i|x_k, h_k) = \begin{cases} \max\{C_2 * (d_i - C_3), 0\} & \text{if } s_i = \text{‘}obj\text{’} \\ \max\{C_2 * (C_3 - d_i), 0\} & \text{if } s_i = \text{‘}bkg\text{’} \end{cases}, \tag{12}$$

where d_i is the distance from s_i to the human model with pose parameter x_k. C_3 is the average distance from the human model to the surface of real human body. C_2 is a constant.

The pair-wise interaction term $\phi_3(s_i, s_j|x_k, h_k)$ can be defined as:

$$\phi_3(s_i, s_j|x_k, h_k) = \begin{cases} C_4 & \text{if } s_i \neq s_j \\ 0 & \text{if } s_i = s_j \end{cases} . \tag{13}$$

3.5 Solve the inner optimization

In this step, we will solve the inner optimization of (8). Given a visual hull h_k (which is decided by the observation vector z_k) and certain pose x_k, we have converted a 3D human body reconstruction to energy minimization. This process can be optimized by simulated annealing or Gibbs sampling algorithm. But they are usually slow convergence. In this work, we minimize the energy using dynamic graph cuts.

To a certain pose x and visual hull h, (9) is a energy of general 3D-MRF. It can be solved using graph cuts if they are sub-modular [9]. The condition for sub-modularity is given as:

$$E(0, 0) + E(1, 1) \leq E(0, 1) + E(1, 0). \tag{14}$$

This implies that the energy for two labels with similar values should be less than the energy for them with different values. In our case, this is indeed the case.

We construct a 3D graph for the energy function. Each common node corresponds to a voxel in the volume of interest, and we construct other two special nodes corresponding to the label 'obj' and 'bkg'.

We set the weights of the links of common nodes by the pair-wise interaction term $\phi_3(s_i, s_j|x_k, h_k)$, and set the weights of the links of common and special nodes according to the unary terms $\phi_1(s_i|x_k, h_k)$ and $\phi_2(s_i|x_k, h_k)$.

Minimizing the energy using dynamic graph cuts, we can get the optimal configuration y_{opt} to a certain pose x and visual hull h.

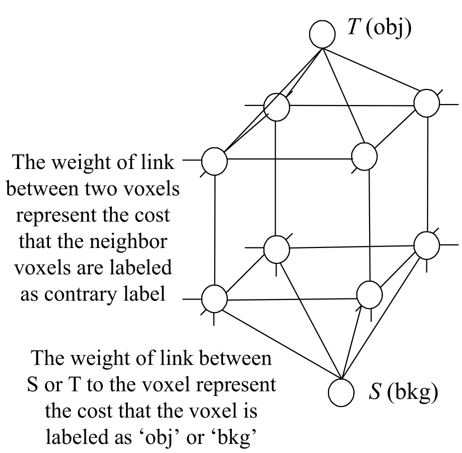


Fig. 4. Construction of 3D graph [8]

3.6 Dynamic graph cuts

Graph cuts were developed by [8, 9] in recent years and several new algorithms based on it have been developed to solve the problem of energy minimization and image segmentation. The main idea of the graph cuts is to construct a special graph for the energy function that will be optimized, so the min-cut of the graph can optimize the energy

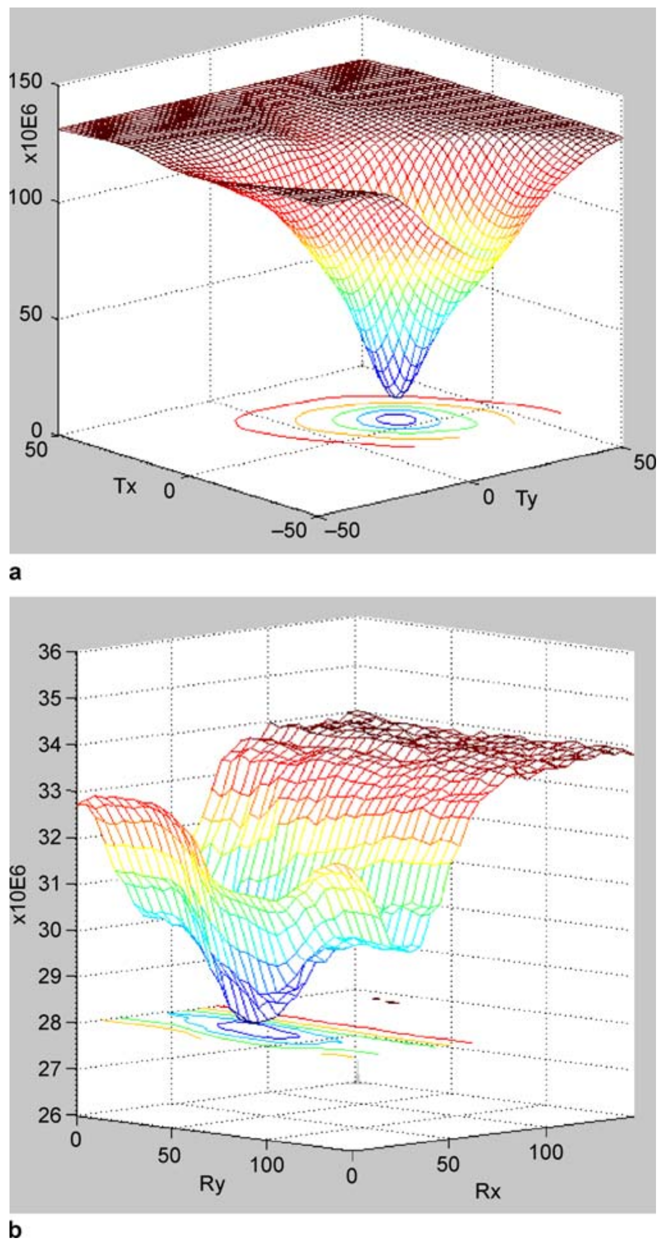


Fig. 5. a $E_x(y_{opt})$ when only changing the global translation in x and y axis, **b** $E_x(y_{opt})$ when only changing the joint angles of the left shoulder in x and y axis

function. On the other hand, the problem of min-cut can be solved by max-flow algorithm.

An un-directional weighted graph $G = \langle V, E\rangle$ consists of a set of nodes (vertices V) and a set of undirected edges (E) that connect them. Two terminal s and t are called source and sink, respectively. The edges connecting each node to the source or sink are called t-links. The edges connecting to two neighboring nodes are

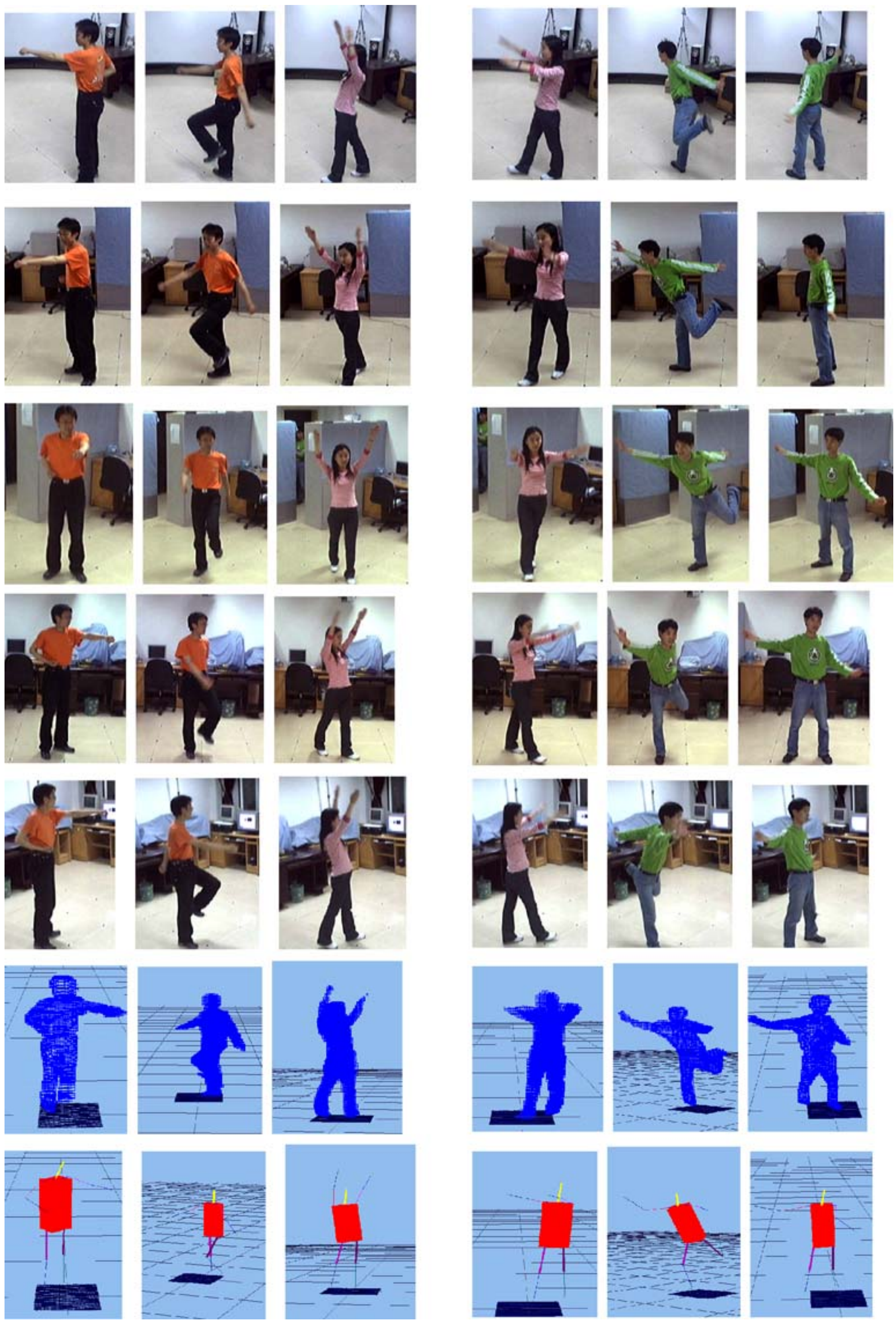


Fig. 6. Visual hull and experiment results

called n-links. Each edge in the graph is assigned a certain weight.

Graph cuts have been proved to be well suited for segmentation of images in many articles [8,17]. In our system, graph cuts can be regarded as 3D volume segmentation. In the context of 3D volume segmentation, V corresponds to the set of all voxels in current time. Two additional terminal S and T represent background terminal and object terminal. Then the edges of E contain all the links between neighboring voxels and the links between the S or T and other nodes. Each edge in the graph is assigned a certain weight depending on the visual hull and human model with pose parameters. The weight of t-link represents the cost that the voxel is labeled as ‘obj’ or ‘bkg’, and the weight of n-link represents the cost that the neighbor voxels are labeled as contrary label. We set the n-link and t-link of the graph according to the energy function (9) and n-link is decided by the unary terms, t-link decided by the interaction term. Each voxel has two t-links and six n-links except the voxel on the edge.

Once the graph is constructed, dynamic graph cuts based upon graph flow will find an optimal (minimum cost) cut. Some voxels are labeled as ‘obj’, and the others are labeled as ‘bkg’. Then the separation of the object from background is completed, and the cost of this min-cut can be calculated. A minimum cost cut generates a segmentation that is optimal in terms of properties that are built into the edge weights, so we can consider this min-cut as a reconstruction of a 3D object based on the visual hull and the human model. We consider this cost as the energy of this reconstruction of the pose and the visual hull. If the pose is more consistent with the visual hull, the 3D reconstructed object will be more accurate with the human body, and the cost of this min-cut will be lower. In this way, we can get the effect of the pose parameters to the visual hull.

3.7 Pose recovery

In this step, we will solve the outer optimization of (8). That is to get the optimal pose x_{opt} of the human. The minimum of the energy function $E_x(y_{\mathrm{opt}})$ reflects the accuracy of the human pose x. So, the task of the pose recovery is to find x_{opt}, which makes the $E_x(y_{\mathrm{opt}})$ minimum.

Figure 5 shows how $E_x(y_{\mathrm{opt}})$ changes with the change of the pose parameters x. It can be clearly seen that the energy surface is locally uni-modal nearby x_{opt}. Thus we can optimize it using standard optimization algorithm. As the calculation of the gradient of energy function is very inconvenient, we adopt the Powell optimization method [22] to solve this problem.

In order to solve this problem, we should give an initial value to the optimization of each frame. In the first frame, we use an interactive graphical user interface to set initial pose parameters. In the following frames, we use the optimal pose parameters of previous frame x^*_{opt} as the initial value of this frame directly, and use Powell algorithm to minimize the function, we can get the optimal pose parameters x_{opt} of current frame. The experiments demonstrate that this method is feasible when the frame rate larger than 15 fps (the movement between two frames is small).

4 Experimental results

We used real video sequences of a human from five synchronized cameras as input, and tested our algorithm on it. Figure 6 presents a series of results of seven frames that we obtained from the experiment. Each column in the figure is the images from five cameras, visual hull and the pose recognition result. All experimentswere conducted with 640×480 images on a P4 3GHz PC offline, taking about one minute per frame.

5 Conclusions

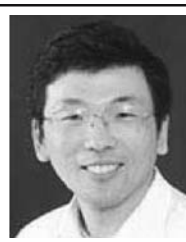
In this paper, we present a new solution for model-based markerless human body motion capture from multiple calibrated cameras. We compute a volume data (voxel) representation from the images using SFS. Different from other methods, we consider the compact of the error data in the visual hull, combine motion capture with 3D reconstruction in MRF-MAP framework and solve the two issues at the same time. We consider the 3D volume data mentioned above as a MRF, and fitting human model to the visual hull is transformed into energy function minimization. We convert the problem of energy function construction into a 3D graph construction, and get the minimal energy by the max-flow theory. Finally, we can recover the human pose by the Powell algorithm. Since the compact of the error messages in the visual hull to the motion capture is considered, our method has higher estimation accuracy than those of other general methods. Several experimental results illustrate the promising performance of this algorithm.

Acknowledgement This work is supported by the National Natural Science Foundation of China (No. 60441002), national Basic Research Program of China - 2006CB303105, 2006CB303104 and 2004CB318110 and University Key Research Project (No. 2003SE002).

This work is done based on the preliminary research made by Sun Yunda, a member of our lab, collaborating with Computer Vision Group, Oxford Brookes University, UK. We greatly appreciate their work.

References

1. Sun, Y.D., Li, B., Yuan, B.Z., Miao, Z.J., Wan, C.K.: Better foreground segmentation for static cameras via new energy form and dynamic graph-cut. In: The 18th IEEE International Conference on Pattern Recognition (ICPR '06), vol. 4, pp. 49–52 (2006)
2. Tsai, R.Y.: A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE J. Robot. Automat. **3**(4), 323–344 (1987)
3. Deutscher, J., Blake, A., North, B., Bascle, B.: Tracking throught singularities and discontinuities by random sampling. In: Proc. 7th Int. Conf. on Computer Vision, vol. 2, pp. 1144–1149 (1999)
4. Caillette, F., Howard, T.: Real-Time Markerless Human Body Tracking with Multi-View 3-D Voxel Reconstruction. BMVC, vol. II, pp. 596–606 (2004)
5. Mikic, I., Trivedi, M., Hunter, E., Cosman, P.: Human body model acquisition and tracking using voxel data. Int. J. Comput. Vision **53**(3), 199–223 (2003)
6. Kehl, R., Bray, M., Van Gool, L.: Full body tracking from multiple views using stochastic sampling. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2005), vol. 2, pp. 129–136 (2005)
7. Cheung, G.K.M., Baker, S., Kanade, T.: Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture. In: The Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. I-77–I-84 (2003)
8. Boykov, Y., Jolly, M.: Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. In: The Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV 2001), vol. 1, pp. 105–112 (2001)
9. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. **26**(2), 147–159 (2004)
10. Saboune, J., Charpillet, F.: Using interval particle filtering for marker less 3D human motion capture. In: International Conference on Tools with Artificial Intelligence (ICTAI 2005), pp. 621–627. IEEE Computer Society (2005)
11. Sundaresan, A., Chellappa, R.: Markerless motion capture using multiple cameras. In: Computer Vision for Interactive and Intelligent Environment (CVIIE 2005), pp. 15–26. IEEE Computer Society, Los Alamitos, CA (2005)
12. Niskanen, M., Boyer, E., Horaud, R.: Articulated motion capture from 3d points and normals. In: Proc. the 16th British Machine Vision Conference (BMVC 2005), vol. I, pp. 439–448. British Machine Vision Association, Oxford (2005)
13. Kohli, P., Torr, P.H.S.: Efficiently solving dynamic markov random fields using graph cuts. In: 10th IEEE Int. Conf. on Computer Vision, vol. 2, pp. 922–929 (2005)
14. Moeslund, T.B., Granum, E.: A survey of computer vision-based human motion capture. Computer Vision and Image Understanding **81**, 231–268 (2001)
15. Ankur, A., Bill, T.: Monocular human motion capture with a mixture of regressors. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 3, p. 72 (2005)
16. Chu, C.W., Jenkins, O.C., Mataric, M.J.: Markerless kinematic model and motion capture from volume sequences. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. II-475–II-482 (2003)
17. Boykov, Y., Kolmogorov, V.: Computing Geodesics and Minimal Surfaces via Graph Cuts. In: Proc. Int. Conf. Computer Vision, vol. 1, pp. 26–33 (2003)
18. Kinderman, R., Snell, J.L.: Markov Random Fields and Their Applications. Am. Math. Soc., Providence, RI (1980)
19. Isard, M.A., Blake, A.: CONDENSATION conditional density propagation for visual tracking. Int. J. Comput. Vision **29**(1), 5–28 (1998)
20. Kohli, P., Torr, P.H.S.: Efficiently solving dynamic Markov random fields using graph cuts. In: 10th IEEE International Conference on Computer Vision (ICCV 2005), vol. 2, pp. 922–929 (2005)
21. Kumar, M., Torr, P.H.S., Obj, Z.A.: Cut. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 18–25 (2005)
22. Press, W., Flannery, B., Teukolsky, S., Vetterling, W.: Numerical Recipes in C. Cambridge Uni. Press, (1988)
23. Cheung, G.K.M., Baker, S., Kanade, T.: Visual hull alignment and refinement across time: a 3D reconstruction algorithm combining shape-from-silhouette with stereo. In: The Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (18–20 June 2003), vol. 2, pp. 375–382 (2003)



CHENGKAI WAN male, born in 1976, Ph.D. of the Institute of Information Science, Beijing Jiaotong University, P.R. China. His fields of research include image processing, pattern recognition and computer vision.

BAOZONG YUAN received the Ph.D. degree in electrical engineering from USSR in 1960. He has joined the Beijing Jiaotong University in 1953. He was a visiting professor at the University of Pittsburgh, USA, and the University of Wales, UK in 1982, 1983 and 1988, respectively. His research interests include computer vision, virtual reality, image processing, computer graphics, speech signal processing, and multimedia information processing and data communication.

ZHENJIANG MIAO professor, Ph.D. supervisor, senior member of Chinese Institute of Computers. In 1987, received his BS degree in radio electronics from Tsinghua University of China, and obtained MS and Ph.D. degrees in signal and information processing from Northern Jiaotong University in 1990 and 1994, respectively. From 1994 to 2004 he worked in France and Canada at French National research Institute of Agronomy, National Research Council of Canada, Nortel Networks, etc. Since April 2004 he has jointed Beijing Jiaotong University. His main research field includes Pervasive Computing, HCI and VR, Network and Distributed Computing, Machine Perception.