2014 22nd International Conference on Pattern Recognition

Automatic Gait Motion Capture with Missing-marker Fillings

Xiaoming Deng*, Shihong Xia†, Wenzhong Wang‡, Zhaoqi Wang†, Liang Chang§, Hongan Wang*
*Beijing Key Lab of Human-Computer Interaction, Institute of Software, Chinese Academy of Sciences, China
†Institute of Computing Technology, Chinese Academy of Sciences, China
‡Department of Computer Science and Technology, Anhui University, China
§College of Information Science and Technology, Beijing Normal University, China
Email: {xiaoming,hongan}@iscas.ac.cn,{xsh,zqwang}@ict.ac.cn,wenzhong@ahu.edu.cn,changliang@bnu.edu.cn

***Abstract*—Although marker-based optical motion capture has been a useful method for computer animation during the past decades, automatic and robust motion tracking from multiple video sequences is still very challenging. Several critical issues in practical implementations are not adequately addressed. For example, how to track and identify the reconstructed 3D points after image matching process? How to handle the heavy occlusion problem? This paper gives a careful investigation of the above issues. In particular, we propose a novel way to track and identify proper markers, and a new method of filling missing markers by taking account of the human model constraints. Experiments are presented to show its accuracy and robustness.**

***Keywords*—motion capture; gait analysis; visual tracking;**

I. Introduction

Marker-based motion capture has been an important tool for collecting and analyzing the human motions in clinical gait analysis and sports training. However, even with high-fidelity and expensive motion capture equipment, motion capture data may still contain noise, missing data and outliers that must be removed manually prior to further processing[1][2][3][4]. It is still challenging to capture reliable and clean 3D human motion data automatically.

In marker-based optical motion capture, multi-view image data is used to compute the motion parameters of moving objects. This method uses a calibrated multi-camera system to reconstruct the motions of moving subjects by measuring the 3D trajectories of passive reflective markers attached to the subjects. To use the recorded data, information such as joint angles, skeletal parameters and the topology of the captured subjects should be extracted. Key issues are that markers can be ambiguous, occluded or missing from certain cameras, and thus the 3D reconstruction and tracking may fail. All these conditions reduce the capability of automatic motion capture. Most professional systems(e.g. Vicon [5] et al.) provide a tool called labeling to identify each marker based on the predefined human topological model. However, even in those systems lots of user's labor-intensive editing work is required, which may lead further errors [6][1][7]. Several researchers have focused on model-based motion capture data processing, including tracking and filling [7], [8]. Herda [7] presented a skeleton-based tracking method. The method needs initializing the first tracking frame manually, which is an error prone process due to the fact that there are often lots of erroneous reconstructed markers. Its major drawback is its weak robustness. Their labeling criterion is the smoothness of the marker accelerations within a sliding window [7], which is not true in most real scenarios due to abrupt limb motions. Li et al. [8] proposed a hierarchical search strategy to reconstruct articulated poses with sparse feature points, while the time complexity is high, up to 5 seconds for identifying markers per frame. Yu et al. [6] proposed a marker labeling method for multi-articulated targets. Data-driven approaches with a motion database are also used to fill the missing markers[1].

In this paper, we focus on marker-based gait motion capture, and propose robust model-based method of tracking and identifying markers as well as method of filling missing markers. Here, 18 markers are attached to human lower body as Vicon [5] does (See Fig.1), each body segment consists of several markers, all markers on the same body segment are assumed to form a rigid-body system, thus a body segment is also called as a rigid body. Our method automatically identify body markers. To fill missing markers, our method setups a candidate set for each lost marker, and then reliably identify a proper filling from this set. Our contributions lie in two aspects: 1) We present a new label likelihood, which is based on the assumption that the distance between every two markers on the same rigid body is constant during gait motions. With this likelihood, we can efficiently obtain robust marker labels using maximum likelihood criterion. 2) We propose a reliable marker filling method, which can fill missing markers automatically in gait motions. The basic idea is to infer possible positions of missing markers by exploring the rigidity constraints of body segments and the knowledge of human body structure.

(a) (b)

Fig. 1. (a) 18 markers on human lower body. (b) Human model. Red balls are joints, blue balls are markers, and body segments are shown as black segments. The markers within an ellipse belong to the same body segment. This figure shows four body segments, whereas the experiments in this paper use seven body segments due to the symmetry of human model.

II. Algorithm

A. Assumptions and Outline

We aim at marker-based lower body motion capture. The lower body consists of a set of connected rigid bodies, and n-earby bodies are linked by joints (See Fig. 1(b)). Our algorithm consists of tracking initialization, marker tracking(also called as marker labeling) and filling, which can be summarized in Algorithm 1 and Fig. 2. We made four assumptions on the input data: 1) the multi-camera system has been calibrated[9], thus the projection matrices of all cameras in the same coordinate system and epipolar geometry between cameras are known. In our experiments, we use a multi-camera autocalibration method [10][11]; 2) there is at least one frame in which all markers can be reconstructed; 3) most of markers on each rigid body can be reconstructed at some point in time; 4) the geometric model of markers on lower body of human(called

The second author is with Beijing Key Lab of Mobile Computing and Pervasive Devices. The work is supported by National Key Technology R&D Program(No.2013BAK03B07)

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.433

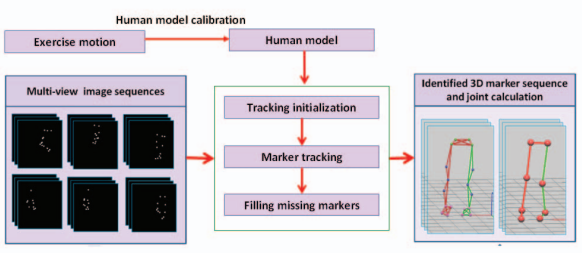
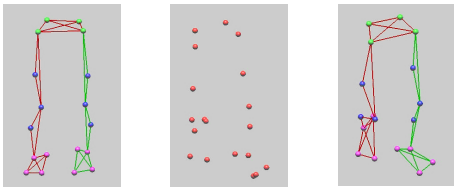


Fig. 2. Overview of the gait motion capture method.



(a) (b) (c)

Fig. 3. Tracking initialization. (a) Human model. (b) Reconstructed markers. (c) Labeled markers.

as human model) are precalibrated with a separate exercise motion [5] via a nearly automatic method like [12](human model calibration), with which the distances between markers on the same rigid body are obtained.

Algorithm 1 Marker tracking with automatic missing-marker fillings (See C and D of Section II for explanations)

Input: labels at the previous frames, marker coordinates at frame k by triangulation, and calibrated human model
Output: labels of markers at the current frame k
Create a set of candidates for each markers (Step 1 and 2 in C of Section II), *matching candidate set*.
Set $NeedRecomp(r) = true$ and $LabelRound(r) = 0$ for all rigid bodies R_b.
for each rigid body $r \in R_b$ **do**
  Search for the maximum likelihood labeling (3) within r via exhaustive search in the matching candidate set (C of Section II);
  Update matching candidate set for labeled markers of r;
  if rigid body r is fully labeled **then**
    $NeedRecomp(r) = false$ and $LabelRound(r) = 2$;
  end if
end for
while $\min_{r \in R_b} LabelRound(r) < 2$ **do**
  Find rigid body r with the most labeled markers from rigid bodies satisfying $NeedRecomp(r) = true$ and $LabelRound(r) = \min_{r \in R_b} LabelRound(r)$;
  Create *filling candidate set* for missing markers on r (D of Section II);
  Search for the maximum likelihood marker labels (See Eq. (3)) within r via exhaustive search of the matching candidate set and the filling candidate set;
  Update the matching candidate set and remove filling candidate set for labeled markers of r;
  if rigid body r is fully labeled **then**
    $NeedRecomp(r) = false$ and $LabelRound(r) = 2$;
  else
    $LabelRound(r) = LabelRound(r) + 1$;
  end if
end while

B. Tracking Initialization

Before tracking, we should label the reconstructed 3D markers (See Fig.3). Human motion contains local geometric invariance in rigid segments [6][8], and this allows affine matching with the model at the segment level. Firstly, for each frame, a set of 3D points can be reconstructed by stereo matching with the calibrated camera parameters. Secondly, a marker-labeling method is used to label the reconstructed 3D markers. We use a hierarchical tracking initialization approach. The waist segment is regarded as root. The marker-labeling process begins at the root, and then continues to label the rigid markers along the hierarchical human body models (See Fig.1 (b)). If all the markers on the subject can be labeled, then we start the marker tracking process; otherwise, go to the next frames for tracking initialization until all the markers can be labeled. There are four steps to accomplish the tracking initialization task.

1) The initialization starts with the first frame, which has at least as many reconstructed markers as the attached markers on the human body. In the reconstructed markers, there exist both true markers and reconstruction outliers due to ambiguity of stereo matching [9].

2) Remove infeasible marker configurations with the distance constraint. For each segment in the body model, there are a large number of potential correspondences among the reconstructed markers. We eliminate most of the infeasible correspondences with the distance constraint. Since the distances of between-markers in a rigid segment keep almost constant in motion, we remove all those marker configurations, whose between-marker distances deviate from those of the segment of the human model by a user defined threshold.
Let $\mathbf{S} = \{\mathbf{P}_i\}_{i=1}^N$ be the N markers on a segment of the body model and $\mathbf{S}' = \{\mathbf{Q}_i\}_{i=1}^N$ be a group of ordered N reconstructed markers. We identify $\mathbf{S}'$ as a proper marker labeling candidate of segment $\mathbf{S}$ only if the following criteria is satisfied:
$$|d(\mathbf{P}_i, \mathbf{P}_j) - d(\mathbf{Q}_i, \mathbf{Q}_j)| < \tau_1, \forall i, j \in [1, N] \quad (1)$$
where $d(\cdot, \cdot)$ denotes the Euclidean distance, and τ_1 is a predefined threshold. Note that different order of the same marker set should be treated as different labeling.

3) Label markers by minimizing the residue error of rigid transformation. Let $\mathbf{S}_k = \{\mathbf{Q}_{k,i}\}_{i=1}^N$ be the remaining K marker groups corresponding to the segment $\mathbf{S} = \{\mathbf{P}_i\}_{i=1}^N$. We calculate the rigid transformation $(\mathbf{R}_k, \mathbf{t}_k)$ between $\mathbf{S}_k$ and $\mathbf{S}$ using the absolute orientation algorithm [13]. Among these $\mathbf{S}_k$s, we select the one that minimizes the following residue error:
$$residue_k = \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{P}_i - \mathbf{R}_k \mathbf{Q}_{k,i} - \mathbf{t}_k\|^2 \quad (2)$$
Once the most probable S_k is found and $residue_k < \tau_2$, all markers in this set are immediately labeled as the corresponding markers in the model, and we remove the labeled markers from the reconstructed markers of the current frame, and go to label the next unlabeled segment along the hierarchical human model. τ_2 is a predefined threshold. If $residue_k \geq \tau_2$, the label results of the segment is not reliable, and thus go to the next frames for tracking initialization until all the markers can be labeled.

4) If all the markers on the subject can be labeled, then we start the marker tracking process (C of Section II).

We illustrate Step 2 and Step 3 by a 2D example in Fig. 4.

Remark 1: Because the distance constraint (Step 3) cannot discriminate symmetric body segments (for example left foot and right foot), we label markers by minimizing the residue error of rigid transformation (Step 4). The distance constraint is useful to remove most of infeasible groups of markers with less computation cost than those with rigid transformation.

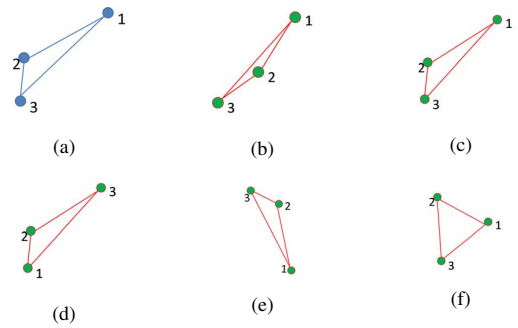


Fig. 4. Illustration of tracking initialization by a 2D example. (a) is a segment in the body model with three attached markers(blue dots). Each marker is numbered with its label. By assumption, the distances between each two of these markers are constant when the segment is in motion. (b)-(f) are five potential configurations of the reconstructed markers. Note that (c) and (d) are actually the same marker sets with different labels. In Step 2, we eliminate all marker configurations that violate criteria Eq. (1), these include (b),(d) and (f). In Step 3, we estimate the rigid transformations between (c) and (a), (e) and (a). Based on these transformations, we calculate the residues Eq. (2) and select (c) as the correct configuration (which has smaller residue than (e)). The marker labels are then determined by (c).

C. Marker Tracking

After the tracking is initialized, all the markers are labeled, and these markers can be tracked from one frame to the next, resulting in marker trajectories over the entire sequence. Hereafter, marker tracking is also called as marker labeling. The marker labeling is carried out along the hierarchical human body models. We explore two constraints on the marker labeling, one is temporal and the other is structural. The temporal constraint enforces: 1) the marker positions are close to those in the previous frame; 2) the distances between markers on a rigid body are almost constant during human motion. The structural constraint states that the between-marker distances on a rigid body do not deviate much from those in the human body model (which are known from a calibration process).

The basic idea of the tracking process is as follows (Fig. 5 is an illustration).

1) After all 3D body markers are labeled in the previous frame $k-1$, we project the 3D coordinate of each marker onto all the camera views with calibrated projective matrices, and obtain the 2D positions of the marker in all the views. In the current frame k, for each 3D marker we setup a circular neighborhood for each camera view, which is centered at the 2D position of the marker in frame $k-1$, and put the detected 2D markers within this neighborhood in a 2D marker candidate table. The radius of circular neighborhood is predefined, which depends on the velocity of markers, the image resolution, camera pose and marker-camera distance. In the experiments, the radius is set to 50 pixels.
2) For each 3D marker, we carry out marker image matching using the 2D marker candidate table in all the camera views, and find a set of geometrically consistent marker matches between each image pair with epipolar geometry[9]. We organize the matches for each 3D marker into a list, where the list is a set of matching 2D markers across multiple images. Then we calculate the 3D positions of the matching markers using triangulation[9]. For each 3D marker, if the distance between the 3D position of matching markers and the 3D marker's position in frame $k-1$ is larger than a threshold t_0, then the matching markers are removed from the list based on the fact that the markers does not move very fast. Then we store the list and the corresponding 3D positions of each 3D marker in a *matching candidate set*. The threshold t_0 is predefined, which is larger as markers moves faster. In our gait tracking experiments, we set t_0 as 50 mm.
3) After the matching candidate sets of all 3D markers are constructed, we use the label likelihood (See Eq.(3)) to label the markers. If there still exist unlabeled markers, we can fill missing markers by the method in D of Section II.

In the following, we describe our method to identify markers from the matching candidate sets. We introduce a criterion to evaluate how the label result fits with the human model. Since the distances between different markers in the rigid body do not change much during gait motions, the geometric structure of each rigid body provides a proper criterion for labeling evaluation.

Denote $S_l = \{\mathbf{C}_p\}_{p=1}^{R_l}$ to be the set of R_l markers attached to the l-th rigid body. The candidates of $\mathbf{C}_p$ are obtained by the matching candidate set of marker p. Our goal is to find the most probable marker assignment of each $\mathbf{C}_p$ of l-th rigid body from the matching candidate sets. The marker labeling can be solved by maximizing the probability $P(S_l)$ of S_l, which is the proposed *label likelihood* and can be calculated as follows:

$$P(S_l) = \prod_{\forall (s,t)\in H} P(\mathbf{C}_s, \mathbf{C}_t) \tag{3}$$

where $H = \{(s,t)|h(s,t)=1\}$, $h(s,t)=1$ holds if and only if there is a rigid body segment link between the markers s,t.

$P(\mathbf{C}_s, \mathbf{C}_t)$ is defined as:

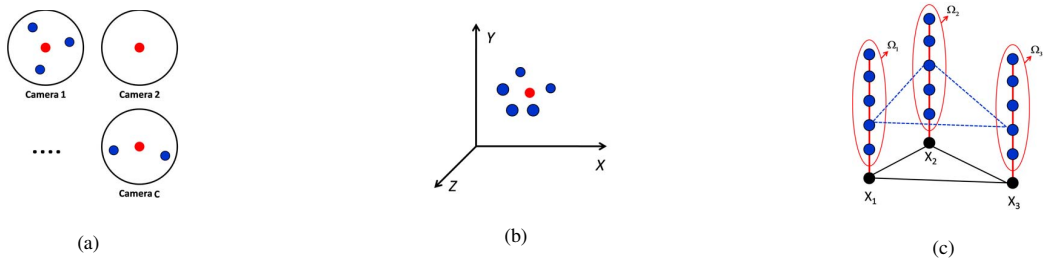
$$P(\mathbf{C}_s, \mathbf{C}_t) \propto \begin{cases} 0 & (\mathbf{C}_s, \mathbf{C}_t) \notin U(s,t) \\ \exp(-\frac{(d(\mathbf{C}_s,\mathbf{C}_t)-d_{s,t})^2}{2\sigma^2(s,t)}) & (\mathbf{C}_s, \mathbf{C}_t) \in U(s,t) \end{cases} \tag{4}$$

where $U(s,t)$ is a feasible set of marker s and t with their corresponding matching candidate sets, which is defined as follows

$$\begin{aligned} U(s,t) = \{&(\mathbf{C}_s, \mathbf{C}_t)|d(\mathbf{C}_s, \mathbf{C}_t) > t_1 \cap\ d(\mathbf{C}_s, \mathbf{C}_t) < t_2 \\ &\cap\ |d(\mathbf{C}_s, \mathbf{C}_t) - d'(\mathbf{C}_s, \mathbf{C}_t)| < t_3 \cap\ |d(\mathbf{C}_s, \mathbf{C}_t) - \\ &d_{s,t}| < t_3\}(t_2 > t_1 > 0, t_3 > 0) \end{aligned} \tag{5}$$

$d(\mathbf{C}_s, \mathbf{C}_t)$ and $d'(\mathbf{C}_s, \mathbf{C}_t)$ are the distances between the markers s,t in the current and previous frames respectively. $d_{s,t}$ is the distance between the markers s,t in the human model, and it is computed in the human model calibration step(See Section 2.1). $\sigma^2(s,t)$ is the deviation of the distance between the markers s,t, while for simplicity we use $\sigma(s,t)$ as a positive constant for all s,t.

In $U(s,t)$, $d(\mathbf{C}_s, \mathbf{C}_t) > t_1, d(\mathbf{C}_s, \mathbf{C}_t) < t_2$ means that the distance between the markers s,t cannot be smaller than a threshold t_1 or larger than a threshold t_2, $|d(\mathbf{C}_s, \mathbf{C}_t) - d'(\mathbf{C}_s, \mathbf{C}_t)| < t_3$ means that the distances of markers s,t between the current and previous frames should not change more than a threshold t_3, and $|d(\mathbf{C}_s, \mathbf{C}_t) - d_{s,t}| < t_3$ means that the the distances of markers s,t between the current frame and the human model should not change more than a threshold t_3. In the experiments, the thresholds t_1, t_2, t_3 are predefined and fixed. Eq. (3) is an evaluation of the rigid body temporal and structural coherence, and for each rigid body, the markers can be labeled by maximizing (3) from the matching candidate sets. If a 3D body marker is labeled, the labeled 3D marker is kept in its matching candidate set, and the other candidates in the set of this body marker are removed. This improves the efficiency of labeling process significantly. If only a part of markers (at least three) in a rigid body have matching candidates, these markers can also be labeled with the label likelihood.



(a) (b) (c)

Fig. 5. Illustration of our marker labeling process. (a) Find probable projections of marker j onto the C camera views. The red dots are projections of marker j in the previous frame. Blue dots in the vicinity of red dots are detected 2D marker positions in the current frame. These detections constitute the 2d marker candidate table Γ_j. Note that there may be no detections in the vicinity of red dot. (b) Reconstruct probable 3D positions of marker j. Here the red dot is the position of marker j in the previous frame, and the blue dots are reconstructed from Γ_j in the current frame. These reconstructions form the set Ω_j. (c) Label markers with the label likelihood. The black dots represent three markers $\{X_1, X_2, X_3\}$ attached to the same rigid body. The blue dots are probable positions of these markers $(\Omega_1, \Omega_2, \Omega_3)$. Our goal is to select for each X_i a particular position from Ω_i, and these selected positions form the most probable configuration subject to temporal and structural constraints (See Eq. (3)). The three positions linked with blue dash lines may be such an optimal configuration.

Remark 1: In Line 12 to 22 of Algorithm 1, we use iterations to fill the missing marker, and the reason is as follows. If a segment S_a has several missing markers, a missing markers on S_a could not be filled due to the fact that some other markers on S_a cannot be filled, while the common markers of S_a and its nearby rigid segment S_b can be filled in the following filling process for S_b. Thus, another filling iteration could be helpful to fill the missing marker on S_a.

D. Filling Missing Markers

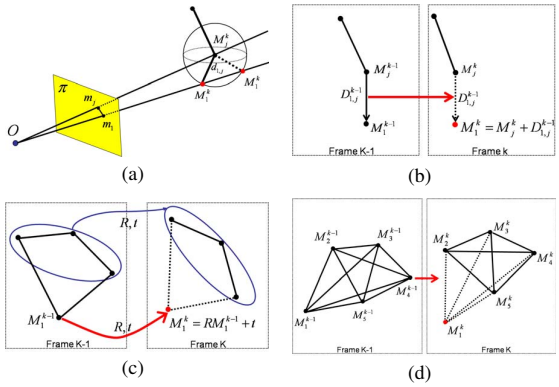
The labeling of all markers is necessary in order to associate the human model to the cloud of markers [7]. However, even in professional motion capture systems, it is often the case that the number of reconstructed markers is smaller than the actual markers on the human, and lots of manually filling work is required (See *Fill Tools* in Vicon iQ [5]. Therefore, marker labeling alone is not sufficient for automatic motion capture due to tracking limitations, and a further automatic filling process is required. In this section, we setup the *filling candidate set* of each missing marker by all the four filling methods (See Fig.6), and then select the proper filling marker from the candidate sets.

1) Setup of Filling Candidate Set: In this subsection, we use four methods to setup a *filling candidate set* of each missing marker, and the filling candidates obtained by all the methods will be selected later with *label likelihood* as shown in Eq.(3).

Denote k to be the current frame number, m_1 to be a missing marker, m_j to be the labeled marker that is connected with m_1 by a rigid body, $\mathbf{M}_j^k$ to be the 3D coordinate of the marker m_j in the frame k. In the following, we use all the following four methods to setup a filling candidate set of each missing marker.

(1) **With Monocular Marker Reconstruction** If a body segment containing the missing marker has at least one labeled marker, then we can proceed with this method (See Fig. 6(a)).

We check the 2D markers in all camera views for those that are not used to reconstruct any labeled 3D marker. The feasible 3D marker $\mathbf{M}_1$ corresponding to m_1 should locate on the 3D ray passing through the camera center $\mathbf{O}$ and the 2D image position m_1[9]. Given that m_j is connected to m_1 by a rigid link with length $d_{1,j}$, $\mathbf{M}_1$ can be recovered by intersecting this ray with a sphere centered at $\mathbf{M}_j$ with radius $d_{1,j}$ by Proposition 1. In our experiment, we use each unlabeled 2D marker as m_1 and setup the set of monocular marker reconstruction. Here, $\mathbf{P}_{3\times4} = [\mathbf{H}_{3\times3}, \mathbf{p}_4]$ is the projection matrix of a camera, and it is recovered by multi-camera calibration[9].



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

Fig. 6. Four methods to setup the missing candidate set. (a) With monocular marker reconstruction. (b) With displacement vectors between markers on rigid body. (c) With rigid transformation of rigid body. (d) With five points constraint.

Proposition 1 Denote $\mathbf{m}_1$ to be the image of a missing marker m_1 in a view, $d_{1,j}$ to be the distance between the missing marker m_1 to another connected labeled marker m_j. With the coordinate of m_j, $\mathbf{M}_j^k$, the filling candidates of the missing marker m_1 can be computed by $\mathbf{M}_1^k = l_1\mathbf{X}_1 + \mathbf{X}_2$ and $\mathbf{M}_1^k = l_2\mathbf{X}_1 + \mathbf{X}_2$, in which $\mathbf{X}_1 = \mathbf{H}^{-1}\mathbf{m}_1$, $\mathbf{X}_2 = -\mathbf{H}^{-1}\mathbf{p}_4$, and $l_{1,2}$ can be computed by $l_{1,2} = \frac{-b\pm\sqrt{\Delta}}{2a}$ (if $\Delta \geq 0$), where $\Delta = b^2 - 4ac, a = \|\mathbf{X}_1\|_2^2, b = 2(\mathbf{X}_2 - \mathbf{M}_j^k)^T\mathbf{X}_1, c = \|\mathbf{X}_2 - \mathbf{M}_j^k\|_2^2 - d_{1,j}^2$. (The proof is in the appendix of this paper.)

(2) **With Displacement Vectors between Markers on Rigid Body** If a body segment containing the missing marker has at least one labeled markers, then we can proceed with this method(See Fig. 6(b)).

Due to high frame rate of the motion capture system, the displacement vector between two markers in a rigid body can be approximated by the displacement in the previous frame $k-1$. Therefore, if a marker in a rigid body is missing, given the other labeled markers in the rigid body and the displacement vectors in the previous frame, the filling candidate $\mathbf{M}_1^k$ can be obtained. Firstly, we calculate $\mathbf{D}_{1,j}^{k-1} = \mathbf{M}_1^{k-1} - \mathbf{M}_j^{k-1}$, which is the displacement vector between 3D coordinates of marker m_1 and marker m_j. And then, the filling candidate of the marker m_1 can be computed by $\mathbf{M}_1^k = \mathbf{M}_j^k + \mathbf{D}_{1,j}^{k-1}$.

(3) **With Rigid Transformation of Rigid Body** If a body segment containing the missing marker has at least three labeled markers, then we can proceed with this method(See Fig. 6(c)).

Due to rigid constraint of rigid body, the rigid body structure is almost fixed within the capture process. If a body segment, which has at least three labeled markers, has a marker m_1 missing, we can use the labeled markers in the previous and current frame to compute the rigid transformation $\mathbf{R}, \mathbf{t}$ with a state-of-the-art method [13], then the filling candidate of the marker m_1 in the current frame k can be computed by $\mathbf{M}_1^k = \mathbf{R}\mathbf{M}_1^{k-1} + \mathbf{t}$.

(4) **With Five Points Constraint** If at least four non-coplanar rigid neighbor markers of the missing marker have been labeled, and the distances between all the five markers are known, then we can proceed with this method(See Fig. 6(d)). Before introducing the method, we introduce the following proposition.

Proposition 2 With four labeled non-coplanar 3D markers $\{\mathbf{M}_j^k = (X_j, Y_j, Z_j)^T\}_{j=2}^5$ and a missing 3D marker $\mathbf{M}_1^k = (X_1, Y_1, Z_1)^T$, if the distances between each pair of the five markers $\{d_{i,j}\}$ are known, then the filling candidate of the missing marker $\mathbf{M}_1^k = (X_1, Y_1, Z_1)^T$ can be computed with the following equations:

$$\begin{vmatrix} W_2 & W_3 & W_4 & W_5 & 0 & W_1 \\ 0 & d_{2,3}^2 & d_{2,4}^2 & d_{2,5}^2 & 1 & d_{2,1}^2 \\ d_{2,3}^2 & 0 & d_{3,4}^2 & d_{3,5}^2 & 1 & d_{3,1}^2 \\ d_{2,4}^2 & d_{3,4}^2 & 0 & d_{4,5}^2 & 1 & d_{4,1}^2 \\ d_{2,5}^2 & d_{3,5}^2 & d_{4,5}^2 & 0 & 1 & d_{5,1}^2 \\ 1 & 1 & 1 & 1 & 0 & 1 \end{vmatrix} = 0 \qquad (6)$$

by which we can get three linear equations of X_1, Y_1, Z_1 by replacing $(W_1, ..., W_5)$ with $(X_1, ..., X_5)$, $(Y_1, ..., Y_5)$ and $(Z_1, ..., Z_5)$ respectively. Since we have three unknown variables X_1, Y_1, Z_1 and three linear equations, then we can compute $\mathbf{M}_1^k = (X_1, Y_1, Z_1)^T$. This proposition was given and proved in [14].

If at least four rigid neighbor markers of the missing marker m_1 have been labeled, and the distances between m_1 and its labeled neighbor markers can be computed in the previous frame or human model, then the filling candidate of the marker m_1 in the k-th frame, $\mathbf{M}_1^k$, can be obtained with Proposition 2.

2) Selection of Filling Candidate Set: For each 3D marker, if the distance between the 3D position of filling candidate marker and the 3D body marker's position in frame $k-1$ is larger than a threshold t_0(set to 50mm), then the filling marker is removed from the filling candidate set. We start with the rigid body which has the most labeled markers among all the rigid bodies with missing markers, and identify the missing markers using the label likelihood (See Eq.(3)) with the aid of labeled markers. The candidates of each 3D missing marker are obtained with the combined set of its filling and matching candidate sets, and the candidate of each labeled 3D marker is the labeled marker obtained in C of Section II. If a 3D marker is filled, its filling candidate set is removed. We store the filled 3D marker in its matching candidate set and remove other candidates in this set. This step will be iterated until no more unlabeled markers can be filled (See Algorithm 1). If a marker can not be filled, then we go to the tracking initialization step to restart tracking.

E. Model-Based Postprocessing

In order to ensure the rigid body constraint(i.e. the distance between different markers in a rigid body keeps almost constant), we refine the marker coordinates by minimizing the following cost functions in the sequence of 3D marker positions $\{\tilde{\mathbf{M}}^k, k = 1, .., K_0\}$:

$$L(\tilde{\mathbf{M}}^k) \quad = \quad E_1^k + \eta E_2^k, k = 1, .., K_0 \qquad (7)$$

where $E_1^k = \frac{1}{N}\sum_{i=1}^{N} ||\mathbf{M}_i^k - \tilde{\mathbf{M}}_i^k||^2$ and $E_2^k = \frac{1}{R}\sum_{(s,t)\in H}(||\tilde{\mathbf{M}}_s^k - \tilde{\mathbf{M}}_t^k||^2 - d_{s,t}^2)^2$. $\tilde{\mathbf{M}}_i^k$ is initialize with $\mathbf{M}_i^k$ by our method in C and D of Section II. E_1^k means that the refined coordinates of markers in frame k, $\tilde{\mathbf{M}}_i^k$, could not deviate much from initial estimations $\mathbf{M}_i^k$, and E_2^k is the marker distance constraint on the same rigid body of frame k. R in E_2^k is the total number of the rigid marker-pair of the lower body, and $d_{s,t}$ is the distance of markers s and t in the human model. η is a predefined weight. The above optimization problems are solved with Levenberg-Marquardt method[9]. The optimization process converges fast because of known skeletal lengths.

III. EXPERIMENTS

We conduct comparison experiments to evaluate the performance of the method without filling missing markers, the seminal method in [7] and our method. The thresholds in the method are set as $\tau_1 = t_3 = 30$ mm, $\tau_2 = 10$ mm, $t_0 = 50$ mm, $t_1 = 20$ mm, $t_2 = 600$ mm. The thresholds are fixed for all experiments.

We tested the algorithm performance with a variety of human motion sequences, including running and jumping. The data set used in our experiments consists of a motion capture sequences, which is with 1211 frames captured by six Photonfocus video cameras at 120 FPS. The input data is a jumping sequence of unlabeled 2D markers with outlier and missing markers. We carry out comparison experiments by down-sampling the capture rates to 60 and 40 FPS. The down sampling is meaningful, because 60 and 40 FPS are usually enough for smooth motion capture and the expense of used camera can be at a lower price. If a frame is with a unfilled marker and the tracking initialization cannot be done, we regard the frame as a failure frame. Comparisons are shown in Tab. 1. The results without filling missing markers have many failure frame, and the seminal model-based method [7] does improve the tracking results significantly. As shown in Tab. 1, the tracking results of our method are fairly better than the results obtained with the other two methods, and our method can track all the markers successfully. [1]

TABLE I. FAILURE FRAME NUMBER BY OUR METHOD, [7] AND WITHOUT FILLING.

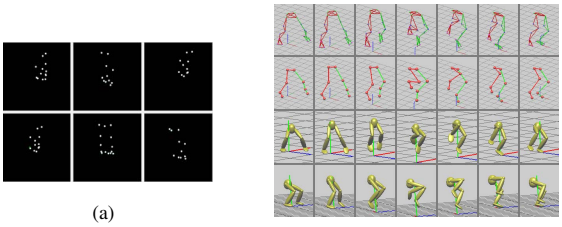
frame number	FPS	our method	method[7]	without filling
1211	120	0	159	171
605	60	0	49	91
403	40	0	45	50

Illustrative results of a jumping sequence is given in Fig. 7, in which we show the motion data of six frames in 3D position space and the joints are estimated by the method [15]. The RMS distance errors of all rigid mark-pairs by our method and [7] are 6.52 mm and 13.50 mm. The ground truth marker-pair distances of each rigid segments are measured with a ruler before the gait motion tracking. The average distance of rigid marker-pairs in the lower body is about 210 mm. Thus, our method can get higher accuracy. The capture sequences were subject to intermittent missing points, while our method can identify the markers throughout the motion capture sequences. With our method, a motion capture frame costs about 0.03 second on an average.

Illustrative results of a walking sequence are shown in Fig. 8. The RMS distance errors of all rigid mark-pairs by our

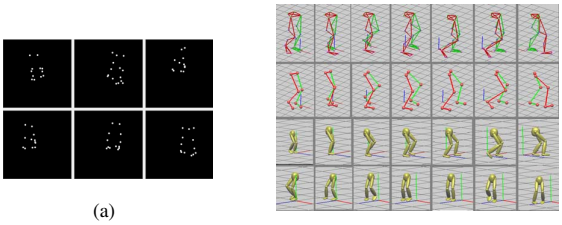
[1] Please see the supplementary video for more comparisons.

method and [7] are 7.49 mm and 9.57 mm. Thus, our method can get higher accuracy. With the RMS errors in walking and jumping and Fig. 9(a)(b), we know that the error with our method is almost at the same level (smaller than 10mm) when the motion type changes. The error with [7] increases if the motion type changes from walking to jumping, which could be caused by that the smoothness assumption of the marker trajectory in [7] does not hold in fast motions like jumping.



(a) (b)

Fig. 7. The results of jump motion sequence. (a) 2D markers in six camera views. (b) Gait motion data of seven frames in 3D position space. (first row) The markers in position space(marker are shown as green, pink and blue balls). (second row) The joints in position space(joints are shown as red balls). (third row) and (fourth row) are the motion data in joint-angle space from two viewpoints.



(a) (b)

Fig. 8. The results of walk motion sequence. The items (a)(b) have the same meanings as in Fig. 7.

IV. Conclusions

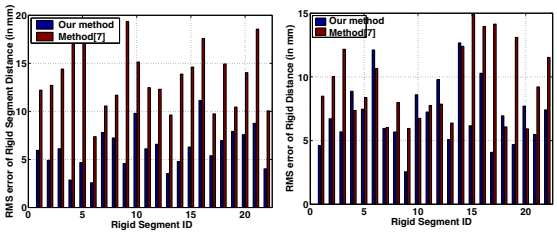
In this paper, we propose a new gait motion tracking and marker filling method in passive optical motion capture. Our method can identify body markers automatically. For missing markers, our method can set up a filling candidate set with available rigid body constraints, and then reliably identify the missing markers from the candidate sets. The robustness and accuracy have been demonstrated by experiments. The method is automatic and an online algorithm, which requires no user interaction once the algorithm starts, thus it is very suitable for applications.

Acknowledgment

This work is supported by National Natural Science Foundation of China (Nos. 61232013, 61173055, 61170182, 61005039), National 973 Project (No.2013CB329305), Beijing Higher Education Young Elite Teacher Project(No. YETP0231), NLPR open project (No. 20090096). We are grateful to Wu Huang for his help in the algorithm implementations, all the reviewers for their inspiring suggestions.

References

[1] H. Lou and J. Chai, "Example-based human motion denoising," *IEEE Trans. on Visualization and Computer Graphics*, vol. 16, pp. 870–879, 2010.



(a) (b)

Fig. 9. RMS distance errors of every rigid mark-pairs by our method and [7]. (a) Jumping. (b) Walking.

[2] W. Zhao, J. Chai, and Y. Xu, "Combining marker-based mocap and rgb-d camera for acquiring high-fidelity hand motion data," *Proceedings of Eurographics/ACM SIGGRAPH Symposium on Computer Animation*, pp. 33–42, 2012.
[3] A. Kirk, J. O'Brien, and D. A. Forsyth, "Skeletal parameter estimation from optical motion capture data," *Proceedings of CVPR*, pp. 782–788, 2005.
[4] W. Wang, X. Deng, X. Qiu, S. Xia, and Z. Wang, "Learning local models for 2d human motion tracking," *Proceedings of IEEE ICIP*, 2009.
[5] OMG, "Vicon motion capture system," *www.vicon.com*.
[6] Q. Yu, Q. Li, and Z. Deng, "Online motion capture marker labeling for multiple interacting articulated targets," *Proceedings of Eurographics*, pp. 477–483, 2007.
[7] L. Herda, P. Fua, R. Plankers, and et. al., "Using skeleton-based tracking to increase the reliability of optical motion capture," *Human Movement Science*, pp. 313–341, 2001.
[8] B. Li, Q. Meng, and H. Holstein, "Articulated pose identification with sparse point features," *IEEE Trans. on SMC-B*, vol. 34, pp. 1412–1422, 2004.
[9] R. Hartley and A. Zisserman, "Multiple view geometry in computer vision," *Cambridge University Press*, 2004.
[10] T. Svoboda, D. Martinec, and T. Pajdla, "A convenient multi-camera self-calibration for virtual environments," *PRESENCE: Teleoperators and Virtual Environments*, vol. 14, pp. 407–422, 2005.
[11] X. Deng, F. Wu, Y. Wu, F. Duan, L. Chang, and H. Wang, "Self-calibration of hybrid central catadioptric and perspective cameras," *Computer Vision and Image Understanding*, vol. 116, 2012.
[12] D. Ross, D. Tarlow, and R. Zemel, "Learning articulated structure and motion," *International Journal of Computer Vision*, vol. 88, 2010.
[13] K. Arun, T. Huang, and S. Blostein, "Least-squares fitting of two 3-d point sets," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 9, pp. 698–700, 1987.
[14] L. Yang, "Solving spatial constraints with global distance coordinate systems," *International Journal of Computational Geometry and Applications*, pp. 553–548, 2006.
[15] M. Silaghi, R. Plankers, R. Boulic, P. Fua, and D. Thalmann, "Local and global skeleton fitting techniques for optical motion capture," *Modeling and Motion Capture Techniques for Virtual Environments, Lecture Notes in Artificial Intelligence, No. 1537*, pp. 26–40, 1998.

Appendix

Proof of Proposition 1: The back-projected line of $\mathbf{m}_1$ can be expressed by $\mathbf{X}(l) = l\mathbf{X}_1 + \mathbf{X}_2$[9], in which $\mathbf{X}_1 = \mathbf{H}^{-1}\mathbf{m}_1$, $\mathbf{X}_2 = -\mathbf{H}^{-1}\mathbf{p}_4$. Since the 3D body marker is on the back-projected line, we can obtain $\mathbf{M}_1^k = l\mathbf{X}_1 + \mathbf{X}_2$. Since the distance between $\mathbf{M}_1^k$ and $\mathbf{M}_j^k$ is $d_{1,j}$, we have $(\mathbf{M}_1^k - \mathbf{M}_j^k)^T(\mathbf{M}_1^k - \mathbf{M}_j^k) = d_{1,j}^2$. Therefore, we get an equation about l

$$\|\mathbf{X}_1\|_2^2 l^2 + 2(\mathbf{X}_1 - \mathbf{M}_j^k)^T\mathbf{X}_1 l + \left\|\mathbf{X}_2 - \mathbf{M}_j^k\right\|_2^2 - d_{1,j}^2 = 0 \quad (8)$$

We use the denotation in Proposition 1. If $\Delta \geq 0$, the roots of l are $l_{1,2} = \frac{-b \pm \sqrt{\Delta}}{2a}$.