3D Motion Parameters Fusion under a Multi-Vision Motion Capture System

Jiangbin Zheng, Huanhuan Zhang, Xiuxiu Li, Yanning Zhang

School of Computer, Northwestern Ploytechnical University

Xi'an Shaanxi China 710072

zhengjb@nwpu.edu.cn

huanhuanzero@163.com

lixiuiu1013@126.com

ynzhang@nwpu.edu.cn

***Abstract*—In order to track multiple markers in a multi-vision motion capture system, it is pivotal to fuse multiple 3D positions of each marker because these 3D positions produced by the different binocular visions are different from each other. A novel fusing method is proposed in this paper. Firstly, an improved nearest neighbour data association method is used to associate each track with the marker which it belongs to. After that an extended 3D Kalman filter is implemented to track each marker. In the association procedure, the associated correct rate is improved by using of an adaptive association threshold and space-time constraint. Experiments show that the proposed method can obtain the all markers' 3D motion parameters accurately.**

I. INTRODUCTION

The marker-based multi-vision motion capture system [1] has been widely applied in movie industry, computer game, animated cartoon, biomechanics and simulating training. In this system, the tracked multiple markers images are quite similar and their calculated 3D positions in the different binocular visions are often different. These make tracking become very difficult. The data fusion technology is one of the pivotal technologies to solve this problem, which determines which local tracks obtained in all the binocular visions to the same marker and then calculate a fused global track with these local tracks.

A data fusion procedure involves the track association and the track fusion. The track association determines whether the local tracks obtained in different cameras belong to the same target. The nearest neighbour algorithm (NN) [2] is a straightforward approach for track associating. However, the NN method could cause incorrect association when the targets are dense. In order to solve this problem, many approaches are proposed such as joint probabilistic data association (JPDA) [3], fuzzy data association (FDA) [4], multiple hypothesis tracking (MHT) [5], etc. This paper gives also a data association method based on NN for its simpleness and effectiveness, and this method improves the associated correct rate because the threshold can be changed adaptively by considering the track error.

The track fusion integrates the associated local tracks of one target into a global track. Many algorithms have been proposed, such as hybrid fusion (HF) [6], covariance intersection (CI) [7], an approximated correlated tracks fusion method [8], etc. In these algorithms, the correlativity of the track errors of the local tracks of one target should be considered because of the common process noise. In this paper, the track errors could be considered as uncorrelated because the common process noise can be ignored in the predigesting indoor environment, and a filter is designed to predict and track each marker in the track fusion procedure.

A novel fusing algorithm is proposed in this paper. In the association procedure, a data association method based on NN is implemented to associate the local tracks with the marker which they belong to, in which the adaptive association threshold and space-time constraint is used to reduce the associated incorrect rate. In the track fusion stage, an extended 3D Kalman Filter [9] is used to track each marker.

II. THE FUSION ARCHITECTURE

There are three kinds of fusion architectures [10]: the centralized fusion, the distributed fusion and the hybrid fusion. The distributed fusion is preferred because it reduces the load of the fusion centre greatly through separate the fusion centre from the local tracking. The fusion architecture of this paper is showed in Fig. 1.

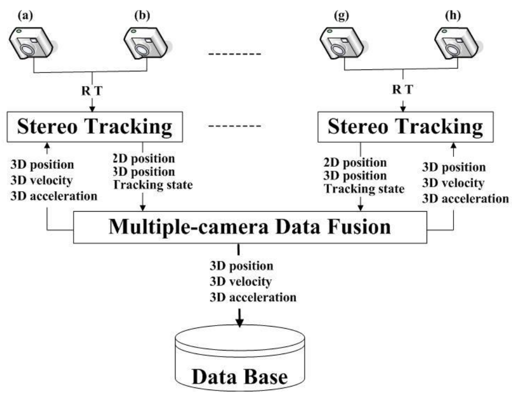


Fig. 1. The fusion architecture

The scene is covered with eight cameras (a)-(h) in Fig.1, and the neighbouring two cameras are composed one binocular vision subsystem because of their larger overlapped region than others. Each subsystem can detect and track markers independently. The fusion module receives the local tracks of each marker produced by all the subsystems and its goal is to compute a global track. The global tracks obtained in the fusion module are feed back every subsystem to improve its tracking performance and stored in a database to drive a human model.

In the proposed fusion procedure, each local track which is received from the subsystem includes each marker's 2D position, 3D position and tracking state, and the global track is one maker's 3D position, 3D velocity and 3D acceleration. The 3D positions of one marker in each local track are fused primarily to reduce the quantity of the fusing data, and the 2D positions of one marker in each local track are used to supply more information when an occlusion happens.

III. DATA FUSION ALGORITHM

A. Extended Kalman Filter

An extended 3D Kalman Filter is used to track each marker, which is consisted by 10 parameters:

$$M_{kj} = (p_k, \hat{p}_k, \Delta p_k, v_k, \hat{v}_k, \Delta v_k, a_k, \hat{a}_k, \Delta a_k) \quad (1)$$

Where p_k, $\hat{p}_k$, Δp_k is 3D calculated position, 3D predictive position and the predictive error of the marker M_{kj} respectively at time *k*; v_k, $\hat{v}_k$, Δv_k is 3D calculated velocity, 3D predictive velocity and the predictive error respectively; a_k, $\hat{a}_k$, Δa_k is 3D calculated acceleration, 3D predictive acceleration and the predictive error respectively.

The equations (2, 3) are used to predict the maker's position. The equations (4, 5) are used to update the maker's velocity. The equations (6, 7) are used to update the maker's acceleration.

$$\hat{p}_{k+1} = p_k + v_k \cdot \Delta t + 0.5 \cdot a_k \cdot \Delta t^2 \quad (2)$$

$$\Delta p_{k+1} = \Delta p_k + \Delta v_k \cdot \Delta t + 0.5 \cdot \Delta a_k \cdot \Delta t^2 \quad (3)$$

$$\hat{v}_{k+1} = \alpha \cdot v_{k+1} + (1-\alpha) \cdot \hat{v}_k \quad (4)$$

$$\Delta v_{k+1} = \alpha \cdot \left|\hat{v}_{k+1} - v_{k+1}\right| + (1-\alpha) \cdot \Delta v_k \quad (5)$$

$$\hat{a}_{k+1} = \beta \cdot a_{k+1} + (1-\beta) \cdot \hat{a}_k \quad (6)$$

$$\Delta a_{k+1} = \beta \cdot \left|\hat{a}_{k+1} - a_{k+1}\right| + (1-\beta) \cdot \Delta a_k \quad (7)$$

where $0 \le \alpha, \beta \le 1$.

B. Principle of the Fusion Algorithm

1) Space-Time Constraint: The space-time constraint is used to measure the 3D matches of two tracks. One marker's 3D positions and velocities from the two-camera subsystems should be identical at the same time, because they are view-independent entities. In reality, they are not identical because of image noise, calibration error and 2D matching error. If one marker is visible in *h* subsystems, it will have *h* 3D local tracks.

2) Classification of Markers: In the fusion process, the markers are classified into five classes according to the occlusive degree and the visible degree in the view of one camera. Different fusion strategies are applied to fuse the corresponding markers. We take four cameras for example showed in Fig. 2.

The first class: Be visible in one subsystem at least, and on right tracking state currently. Because these markers have been tracked correctly in the subsystems, only the track fusion process is needed to be done by using their 3D positions.

The second class: Be invisible in any subsystem, but visible in two cameras at least and on right tracking state currently. To this kind of maker, their 2D positions are used in track fusion process.

The third class: Neither visible in any subsystem nor visible in any two cameras. Because they are out of 3D view, the current tracks are replaced by the predictive tracks temporarily.

The fourth class: Be visible in any subsystem, and on new tracking state. Track association is needed to identify their IDs before the track fusion process.

The fifth class: Be visible in the subsystem, and on occlusive tracking state. To this kind of maker, we only feed back the fused track to the subsystem to modify the occlusive tracking.

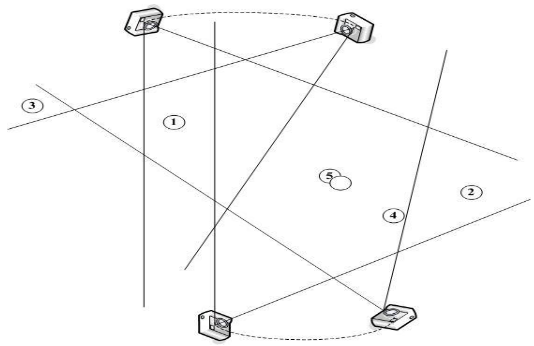


Fig. 2. Five classes of the tracked markers

3) Identifying New Markers' ID: There is no new marker in the motion capture system because the total number of markers has been settled ahead. But for the subsystem the new markers probably exist. At time *k*, *n* markers' tracks $\{M_{k1}, M_{k2}, \ldots M_{kn}\}$ are needed to obtain, where *n* is the total number of markers in the system and kj is the global ID of the marker M_{kj}. Assume *s* global tracks $\{M_{k1}, M_{k2}, \ldots M_{ks}\}$ are obtained currently, and *n-s* tracks $\{M_{ks+1}, M_{ks+2}, \ldots M_{kn}\}$ are unresolved. Temporality m local new tracks $\{UM_1, UM_2, \ldots UM_m\}$ are given. The problem is to compute an optimal *ID* d_i, $d_i \in \{M_{k1}, M_{k2}, \ldots M_{kn}\}$ for each marker UM_i, the formulation is:

$$d_i = \begin{cases} \arg\max L_1(M_{kj}, UM_i),\ j = 1\ldots s, i = 1\ldots m \\ \arg\max L_2(M_{kl}, UM_i), l = s+1\ldots n, i = 1\ldots m \end{cases} \quad (8)$$

L_1 is used to measure the likelihood of UM_i associating with $\{M_{k1}, M_{k2}, \ldots M_{ks}\}$, and L_2 is used to measure the likelihood of UM_i associating with $\{M_{ks+1}, M_{ks+2}, \ldots M_{kn}\}$. If d_i is identified, the marker UM_i will be added to the track which it belongs to.

The improved NN method is used. The shorter distance between UM_i and M_{kj} (M_{kl}), the more interrelated they are. UM_i will be associated with M_{kj} (M_{kl}) when their distance is the shortest and in the threshold ε_1 (ε_2), where ε_1 is constant and ε_2 is variable.

C. Procedure of Fusion Algorithm

Step1. Use Eq.(2) to predict all the markers' positions at next time $k+1$.

Step2. Fuse the first marker M_{kj}, $M_{kj} \in \{M_{k1},...M_{ks}\}$.

Assume the marker M_{kj} is visible in h subsystems. Its track can be expressed as $M_{kj} = \{M_{kj1}, M_{kj2}, ...M_{kjh}\}$, $1 < h < n$, where M_{kjh} is a local track obtained in a subsystem. Finally s markers' tracks $\{M_{k1}, M_{k2}, ...M_{ks}\}$ can be obtained. If $s=n$ is true, turn to Step5.

Step3. Fuse the second marker M_{kj}, $M_{kj} \in \{M_{ks+1}, ...M_{kn}\}$.

Two cameras, in which the marker M_{kj} is tracked on right tracking state, are randomly combined to a binocular subsystem, and its 3D position is calculated by public cross line algorithm. Add M_{kj} to the tracks $\{M_{k1}, M_{k2}, ...M_{ks}\}$. If $s=n$ is true, turn to Step5.

Step4. Fuse the third marker M_{kj}, $M_{kj} \in \{M_{ks+1}, ...M_{kn}\}$.

The 3D track of the marker M_{kj} will be replaced by the predictive track because it is out of 3D view. The predictions of $\{M_{ks+1}, M_{ks+2}, ...M_{kn}\}$ are expressed as $\{\hat{M}_{ks+1}, \hat{M}_{ks+2}, ...\hat{M}_{kn}\}$.

Step5. Associate the new marker UM_i with the marker M_{kj}, $M_{kj} \in \{M_{k1},...M_{ks}\}$.

$$\begin{aligned} \tilde{L}_1\left(UM_i, M_{kj}\right) &= \left|UM_i - \tfrac{1}{h}\sum_{l=1}^{h} M_{kil}\right|, M_{kjl} \in M_{kj} \\ L_1\left(UM_i, M_{kj}\right) &= \frac{1}{\tilde{L}_1\left(UM_i, M_{kj}\right)} \end{aligned} \quad (9)$$

Assume $kj = \arg\max L_1\left(UM_i, M_{kj}\right)$ according to Eq.(9), only $L_1\left(UM_i, M_{kj}\right) < 1/\varepsilon_1$ is true, can UM_i be associated with M_{kj}. The parameter ε_1 is 30mm, equals the marker's radius. If UM_i is associated, turn to Step7.

Step6. Associate the new marker UM_i with the marker $\hat{M}_{kj}$, $\hat{M}_{kj} \in \{\hat{M}_{ks+1},..\hat{M}_{kn}\}$.

$$\begin{aligned} \tilde{L}_2\left(UM_i, \hat{M}_{kj}\right) &= \left|UM_i - \hat{M}_{kj}\right| \\ L_2\left(UM_i, \hat{M}_{kj}\right) &= \frac{1}{\tilde{L}_2\left(UM_i, \hat{M}_{kj}\right)} \end{aligned} \quad (10)$$

Assume $kj = \arg\max L_2\left(UM_i, \hat{M}_{kj}\right)$ according to Eq.(10), only $L_2\left(UM_i, \hat{M}_{kj}\right) < 1/\varepsilon_2$ is true, can UM_i be matched with $\hat{M}_{kj}$. The parameter ε_2 is variational along with the predictive error. It is determined by the marker's radius and the predictive error. If UM_i is not associated, it is an invalid marker, whereas add it to the suited track M_{kj}.

Step7. Calculate the mean of each M_{kj}, $M_{kj} \in \{M_{k1},...M_{kn}\}$.

Assume every local track's weight of the marker M_{kj} is uniform.

$$M_{kj} = \tfrac{1}{h}\sum_{l=1}^{h} M_{kjl} \quad (11)$$

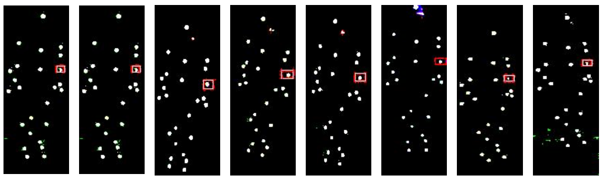
Step8. Use Eq.(3)-Eq.(7) to update all the markers' position predictive errors, velocities and accelerations.

Step9. Feed back the 3D tracks of the fourth and fifth markers to the subsystems which they belong to.

IV. EXPERIMENTS

Our experiment is done in 5m*5m indoor scene surrounded by the black cloth. The 8 cameras are located in front of an actor, and the resolution of the images is 752×480 and the frame rate is 30F/s. The angle between two neighbouring cameras isn't zero. The 25 white markers with 25mm diameter are attached on the actor. The performance of the proposed fusion algorithm is presented by tracking the marker in the rectangle.

The images from eight cameras at the 1st frame are shown in Fig. 3, which shows that the marker is visible in all the cameras. The fusing results of the fist frame are shown in Table1. The "subi" means the no.ith stereo tracing subsystem, where i=1,...,4. Their tracking state is expressed as (x_1, x_2), where x_1 is the 2D tracking state in the left camera and x_2 is the 2D tracking state in the right camera. "Fusion" means the data fusion result, and its tracking state is expressed as x_3. $x_1(x_2, x_3)$ equals one of the numbers $\{0,1,2,3\}$, where 0 means the invisible tracking state, 1 means the right tracking state, 2 means the occlusive tracking state, 3 means the new tracking state.



(a) (b) (c) (d) (e) (f) (g) (h)

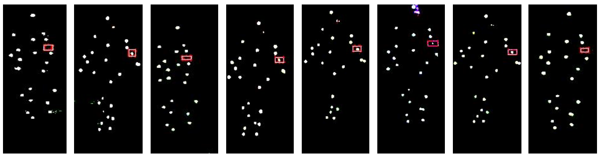
Fig. 3. Eight images at 1st frame: the marker is visible in all the cameras

TABLE I
RESULT IN THE 1ST FRAME

Sub No.	Tracking state	3D tracking position (mm)
Sub1	(1,1)	(528.385,-232.728,358.509)
Sub2	(1,1)	(522.454,-222.163,349.096)
Sub3	(1,1)	(536.571,-231.568,342.852)
Sub4	(1,1)	(535.798,-226.015,342.539)
Fusion	1	(530.802,-228.119,348.249)

The images from eight cameras at the 15th frame are shown in Fig. 4, which shows that the marker is lost in camera (a), (c), (h). The results are shown in Table2. "Prediction" means the predictive position by the extended 3D Kalman Filter, "Error" means the position predictive error and "unused" means that the local 3D positions are not used in fusion procedure.

The four "unused" show that all the subsystems are invalid at the 15th frame because the marker is not tracked correctly in any subsystem. We combine the right camera of sub1 and the left camera of sub3 to a temporary binocular subsystem. The predictive error in Table2 implies that the proposed algorithm could track the marker accurately when it is not tracked correctly or lost in any subsystem.



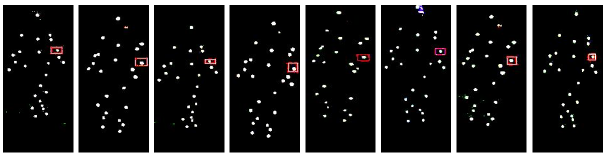
(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 4. Eight images at 15th frame: the marker is lost in camera (a), (c), (h)

TABLE III
RESULT IN THE 15TH FRAME

Sub No.	Tracking state	Tracking position (mm)
Sub1	(0,1)	unused
Sub2	(0,1)	unused
Sub3	(1,2)	unused
Sub4	(2,0)	unused
Fusion	1	(552.661,-354.104, 400.122)
Prediction(mm)		(553.342,-357.69,412.559)
Error(mm)		(0.68121,3.5864,12.4366)

The images from eight cameras at the 33rd frame are shown in Fig. 5, which shows that the marker appears again at camera (a), (c), (h). The results are shown in Table3. "ID" means the id of the marker, and it equals d_{12}. The stress *IDs* in Table3 imply that the marker is identified correctly at sub1, sub2 and sub3 using the proposed algorithm.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 5. Eight images at 33rd frame: the marker appears again at camera (a), (c), (h)

TABLE IIIII
RESULT IN THE 33RD FRAME

Sub No.	Tracking state	Tracking position (mm)	ID
Sub1	(3,1)	unused	$\mathbf{d_{12}}$
Sub2	(3,1)	unused	$\mathbf{d_{12}}$
Sub3	(1,2)	unused	d_{12}
Sub4	(1,3)	unused	$\mathbf{d_{12}}$
Fusion	1	(585.175,-160.672,191.02)	d_{12}

V. CONCLUSIONS

In this paper, a novel fusion algorithm is proposed to obtain the marker's accurate 3D track.

In future, the detection and elimination of the wrong position for a certain marker will be studied when the 3D positions in some subsystems are different from each other obviously, and the factor of the cameras' viewpoint is considered to improve the accuracy of 3D positions in a binocular subsystem.

ACKNOWLEDGMENT

The paper is supported by Hi-Tech Research and Development Program of China (Program No. 2006AA01Z324), and National Natural Science Grants of China.

REFERENCES

[1] Barca, J.C.; Rumantir, G.; Koon Li, R.; "A New Illuminated Contour-Based Marker System for Optical Motion Capture", Innovations in Information Technology, 2006. Nov. 2006 Page(s):1 - 5 Digital Object Identifier 10.1109/INNOVATIONS.2006.301980.

[2] Bar-Shalom Y. "Multitarget-Multisensor tracking: applications and advances", Norwood, MA:Artech House, 1990. 65-68.

[3] Puranik, S.P.; Tugnait, J.K. "Multisensor tracking of multiple maneuvering targets using multiscan JPDA and IMM filtering", American Control Conference, 2005. Proceedings of the 2005 8-10 June 2005 Page(s): 2058-2063 vol.3. Digital Object Identifier 10.1109/ACC.2005.1470273.

[4] Ashraf M. Aziz,Tummala M,Cristi R. Fuzzy logic data correlation approach in multisensor-multitarget tracking systems. Signal Processing,1999,76(2):195-209.

[5] Blackman, S.S. "Multiple hypothesis tracking for multiple target tracking", Aerospace and Electronic Systems Magazine, IEEE Volume 19, Issue 1, Part 2, Jan 2004 Page(s): 5-18 Digital Object Identifier 10.1109/MAES.2004.1263228.

[6] Liang Jimin; Wang Changming; Yin Tingrui; "A hybrid measurement fusion algorithm for multisensor target tracking", Intelligent Control and Automation, 2004. WCICA 2004. Fifth World Congress on Volume 6, 15-19 June 2004 Page(s): 5422-5425 Vol.6. Digital Object Identifier 10.1109/WCICA.2004.1343765.

[7] Ren C. Luo, Ying Chih Chou, Ogst Chen. "Multisensor Fusion and Integration Algorithms Applications, and Future Research Directions", Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation August 5 - 8, 2007, Harbin, China.

[8] Kaplan, L.M.; Blair, W.D.; Bar-Shalom, Y. "Simulations Studies of Multisensor Track Association and Fusion Methods", Aerospace Conference, 2006 IEEE 4-11 March 2006 Page(s): 16pp. Digital Object Identifier 10.1109/AERO.2006.1655913.

[9] Jiangbin Zheng, David Dagan Feng, Wanchi siu, Yanning Zhang, XiuYing Wang, Rongchun Zhao. "The accurate extraction and tracking of moving objects or video surveillance", Proceedings of International Conference on Machine Learning and Cybemetics, pp.1909-1913, 2002.

[10] C.-Y. Chong, S. Moni, W. H. Barker, and K.-C. Chang, "Architectures and algorithms for track association and fusion," IEEE Aerospace and Electronic Systems Magazine, vol. 15, no. 1, pp. 5-13, Jan. 2000.