Fast Tracking 3D Arm Motion with Joint-Chain Motion Model

X.S. Yu W. Zhao J.F. Liu X.L. Tang J.H. Huang

School of Computer, Harbin Institute of Technology, Harbin 150001, China

Abstract

Focusing on the problem of low computation efficiency in the process of tracking human 3D motion, the fast tracking algorithm for 3D arm motion based on Joint-Chain Motion Model (JCMM) is proposed based on the Particle Filter. In our algorithm, via the Joint-Chain Motion Model (JCMM) is defined, the arm motion state space can be discomposed into some low dimension subspaces, and the amount of particle in tracking can be reduced. The result of experiment shows that our algorithm can advance the computational efficiency while guarantee precision of tracking.

Keywords: Particle Filter; 3D arm Motion; the Joint-Chain Motion Model; Subspace

1. Introduction

According to the intensive research in computer vision, human 3D motion tracking has been an attentive subject in recent years due to its wide applications such as virtual reality, computer animation, *etc*.

Moeslund [1] [2] *et al.* thought the human 3D motion tracking as a temporal prediction procedure. So the state transition of human motion could be represented as the first-order hidden Markov procedure, and the current state was under the constraint of last time state and current observation state. As one nonlinear filter algorithm based on the Bayesian estimation framework, the use of Particle filter [3] has been widely application [5] [6] in the area of human 3D motion tracking. The challenge of human motion tracking based on particle filter is how to advance the computational efficiency.

One category [4] [9] uses strong motion prior to constrain the search into the most likely region of the parameter space. Another solution is to learn low-dimensional latent variable models. In this way, general method is that high dimensional human state space can be projected to a nonlinear subspace using the PCA. In [12], the tracking problem is formulated as minimizing differentiable deterministic objective function. And the human 3D tracking is defined as the multi-hypothesis optimization in a Bayesian framework [10] [11]. Further more, Xinyu *et al.* [7] learn motion correlation using the Partial Least Square, and proposed the RBPF-PLS algorithm based on Rao-Blackwellised to track the walking pose.

Although these algorithms have achieved the goal, they can't track any motion in the nature scene but tracking the learned motion. And learning a general probabilistic model in full space is very difficult because of the high dimensionality and the huge amounts of training data to account for motion complexity.

Focused on the problem, the paper proposes the Joint-Chain Motion Model, in which the human motion is represented by the joint-chain. In the model, the children joints' motion are only correlated

with the father joints' motion, and the correlation of motion speed between the father joints and the children joints can be calculated by the Least Square. The high dimensionality state space of human motion can be decomposed into some joint subspace via the JCMM. As the result, the algorithm can advance the computational efficiency. The paper chooses the right arm motion video as the experimental subject.

The paper is organized as follows. Section 2 describes the Joint-Chain Motion Model, and the framwork of the tracking algorithm is proposed. The steps of the algorithm are described in Sention 3. Experimental results and analysis are shown in Section 4, and finally concludes the paper.

2. Tracking Framework

2.1. Joint-Chain Motion Model

The human 3D skeleton model [8] can be decomposed into six joint-chains. Any pose of human motion can be described as the combination of six joint-chains. Fig. 1 shows the human 3D skeleton is represented by the Joint-Chain Motion Model.

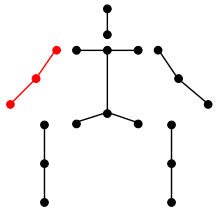


Fig. 1 the Joint-Chain Motion Model for the human 3D skeleton model.

The definition of Joint-Chain Motion Model (JCMM) includes two levels: the Joint-Chain level and the Joints level. 1) *The Joint-Chain level:* The torso Joint-Chain is at dominative level, on which other Joint-Chains motion depend. Other five Joint-Chains are at subordinate level, and their motion are independent each other. 2) *The Joints level:* The based-node is defined as the joint that controls the whole Joint-Chain motion in any Joint-Chain, while another end joint of the Joint-Chain is defined as the end-effector. If one Joint-Chain has more than three joints, the mid-level joints are defined as the mid-joints. Each joint motion only interacts with its neighbors.

In this paper, the research subject is the 3D motion of right arm. In Fig. 1, the red chain is used to describe the Right Arm Joint-Chain Motion Model (RAJCMM). In the right arm Joint-Chain, the base-node is the right shoulder, and the end-effector is the right wrist, and the mid-node is the right elbow. Following the definition of the JCMM, the motion of right elbow joint only interact the right wrist joint, and its motion depend on the right shoulder joint.

We denote the joints sets of right arm by $J = \{j_0, j_1, j_2\}$, where the subscript of each member in sets is respectively represented for the right shoulder joint, the right elbow joint, and the right wrist joint by the ascending order. In this paper, the right arm Joint-Chain state space is formatted by the 3D coordinate triplets of each joint as Eqn. 1.

$$x = \{x_0, x_1, x_2\} \tag{1}$$

Where x_0 is right shoulder's state, x_1 is right elbow's state, and x_2 is right wrist's state. Using the RAJCMM, the problem of tracking right arm motion can be formulated as the prediction of x_t.

2.2. Tracking Framework

The state parameter x_t of right arm motion at time t is represented by the form of joint state as shown Eqn. 2:

$$x_t \ @\{x_{i,t}\}_{i=0}^{2} = \{x_{0,t}, x_{1,t}, x_{2,t}\} \tag{2}$$

We denote the father of *ith* joint as *F(i)*, and the observation state of all joints is defined as $Z_t = \{Z_{i,t}\}_{i=0}^{2}$. The *posterior* probability distribution for the right arm motion is given by:

$$P(x_t \mid z_t) = P(x_{0,t} \mid z_{0,t}) \prod_{i=1}^{2} P(x_{i,t} \mid x_{F(i),t}, z_{0,t}) \quad (3)$$

Where $x_{i,t}$ is represented for the state parameter triplet of *ith* joint at time *t*, and $x_{F(i),t}$ is described as the optimization state triplet of the *ith* joint's father joint *F(i)* at time *t*. The MAP for the right shoulder is $P(x_{0,t} \mid z_{0,t})$. The *posterior* probability distribution $P(x_{i,t} \mid x_{F(i),t}, z_{i,t})$ is represented as shown Eqn. 4.

$$P(x_{i,t} \mid x_{F(i),t}, z_{i,t}) = cP(z_{i,t} \mid x_{i,t}, x_{F(i),t}) \times \int P(x_{i,t} \mid x_{i,t-1}) P(x_{i,t-1} \mid z_{i,t-1}) dx_{i,t-1} \quad (4)$$

The likelihood $P(z_{i,t} \mid x_{i,t}, x_{F(i),t})$ is the distribution that the observation $z_{i,t}$ of current joint *i* at time *t*, which is conditionally independent of its state $x_{i,t}$ given its father joint's state $x_{F(i),t}$.

The basic idea of particle filter is to use a weighted sample set $\{x_{i,t-1}^k, w_{i,t-1}^k\}_{k=1}^N$ to estimate the posterior density $P(z_{i,t} \mid x_{i,t}, x_{F(i),t})$. So Eqn. 4 can be approximated by Eqn. 5.

$$P(x_{i,t} \mid x_{F(i),t}, z_{i,t}) \approx cP(z_{i,t} \mid x_{i,t}, x_{F(i),t}) \times \sum_k w_{i,t-1}^k P(x_{i,t} \mid x_{i,t-1}^k) \quad (5)$$

Where $x_{i,t-1}^k$ is denoted as the *kth* sample of the *ith* joint at time *t*-1, $w_{i,t-1}^k$ is the associated normalized weights updated with the following expression:

$$w_{i,t}^k \propto w_{i,t-1}^k P(z_{i,t} \mid x_{i,t}^k, x_{F(i),t}), \sum w_{i,t}^k = 1 \quad (6)$$

Then the state $x_{i,t}$ can be estimated as shown Eqn. 7:

$$x_{i,t} \approx \sum_k w_{i,t}^k \times x_{i,t}^k \quad (7)$$

Substituting (7) into (2), the state x_t is represented as shown Eqn. 8.

$$x_t = \{x_{i,t}\} \approx \{\sum_k w_{i,t}^k \times x_{i,t}^k\}_{i=0}^2 \quad (8)$$

3. Tracking Arm Motion with JCMM

In this section we present particle generation algorithm and weighted color histogram algorithm based target area.

3.1. Particle Generation

In particle filter theoretical framework, the state transition model is described as shown Eqn. 9.

$$x_t = x_{t-1} + v_t, v_t : N(\mu, \Sigma) \quad (9)$$

Where v_t is the Gaussian noise and μ is a 3×1 scalar, defined as the motion speed of current joint, and the variance Σ is the 3×3 diagonal matrix.

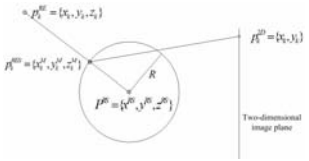


Fig. 2 One 3D Particle is projected to the image plane via the constraint sphere model.

Before projecting 3D particle to two dimensional image planes, the particle need be transferred as the following equation:

$$
\begin{aligned}
&R = \sqrt{(x_{1,t} - x_{0,t})^2 + (y_{1,t} - y_{0,t})^2 + (z_{1,t} - z_{0,t})^2} \\
&l = x^k_{1,t} - x_{0,t}; m = y^k_{1,t} - y_{0,t}; n = z^k_{1,t} - z_{0,t}; \\
&x'^k_{1,t} = x_{0,t} + l \times R / \sqrt{l^2 + m^2 + n^2} \\
&y'^k_{1,t} = y_{0,t} + m \times R / \sqrt{l^2 + m^2 + n^2} \\
&z'^k_{1,t} = z_{0,t} + n \times R / \sqrt{l^2 + m^2 + n^2}
\end{aligned} \tag{10}
$$

In Eqn.10, we denote the right elbow as j_1 and its father, the right shoulder, as j_2. The Eqn. 10 is the projection equation in our algorithm. N is the count of particles, and $1 \le k \le N$. The sphere center of the constraint model of the right elbow is defined as $x_{0,t} = (x_{0,t}, y_{0,t}, z_{0,t})$, which is the 3D coordinate of right shoulder, and $x_{1,t} = (x_{1,t}, y_{1,t}, z_{1,t})$ is defined as the right elbow. The radius R of sphere is represented by the 3D distance between the right elbow and right shoulder. $x'^k_{1,t} = (x'^k_{1,t}, y'^k_{1,t}, z'^k_{1,t})$ is the projection point of particle $x^k_{1,t}$ on the sphere of the constraint model. Using the intrinsic and extrinsic parameter matrixes of the camera, we can get the projection point $\hat{x}^k_{1,t}$ of $x'^k_{1,t}$ on the image plane.

3.2. Weighted Color Histogram

The observation likelihood model is represented for the matching relationship between the human appearance model and the features subtracted from the image.

The appearance model of right arm is confirmed by the weighted color histogram of target rectangles, and these target rectangles are form of initial frame ground truth of each joint, including right shoulder, right elbow and right wrist and so on. In our algorithm, the target area is represented as rectangle, while the length of rectangle is the Euclidean distance between the 2D projection point of particle and 2D projection point of the particle's father joint, and the height of rectangle is confirmed by the experience value *Arm-Height*.

3.3. Motion Speed Update

The motion speed of any joint j_i depends on the speed of the joint j_i at time t-1 and the motion speed of its father joint $j_{F(i)}$.

The row vector $v_{i,t} = (v^x_{i,t}, v^y_{i,t}, v^z_{i,t})$ is represented as the motion speed of the joint j_i at time t. The motion speed of father joint $j_{F(i)}$ is defined as $v_{F(i),t} = (v^x_{F(i),t}, v^y_{F(i),t}, v^z_{F(i),t})$. If $t<3$, $v_{i,t}$ is confirmed as following equation:

$$
v_{i,t} = \begin{cases} (0,0,0) & t = 0 \\ (x_{i,t} - x_{i,t-1}, y_{i,t} - y_{i,t-1}, z_{i,t} - z_{i,t-1}) & t = 1,2 \end{cases} \tag{11}
$$

If $t \ge 3$, $v^x_{i,t}$, $v^y_{i,t}$, $v^z_{i,t}$ must be calculated independently by Eqn. 12.

$$
\begin{aligned}
v^x_{i,t} &= \alpha_{i,t-1} \times (v^x_{i,t-1}\ v^x_{F(i),t})' \\
v^y_{i,t} &= \beta_{i,t-1} \times (v^y_{i,t-1}\ v^y_{F(i),t})' \qquad t \ge 3 \\
v^z_{i,t} &= \gamma_{i,t-1} \times (v^z_{i,t-1}\ v^z_{F(i),t})'
\end{aligned} \tag{12}
$$

Where, the coefficient $\alpha_{i,t-1}$, $\beta_{i,t-1}$, $\gamma_{i,t-1}$ are the 2×1 scalar obtained by least squares method.

4. Experimental Results and Analysis

4.1. Experimental Design

We have done experiments to track the right arm motion using the HumanEva data sets [13], which were captured at 25 fps by Leonid *et al*. of American Brown University via the VICON system. The experiment chooses the right arm motion color video made in the front to reduce the self-occlusions. The tracking experiments have done by Visual Studio .NET 2003 with dual-core 1.8GHz and 1G DDR memory PC. The video has 796 frames image sequence and image resolution is the 640×480.

Spatial position of the right shoulder joint has not evidently change in experi-

mental video. Then the Eqn. 3 can be simplified as the following equation:

$$P(x_t \mid z_t) = \prod_{i=1}^{2} P(x_t^i \mid x_t^{F(i)}, z_t^i) \quad (13)$$

In Eqn. (9), the main diagonal elements of diagonal matrix $\sum$ are equivalent to a constant, and the value of constant is 40. In subsection 3.2, the height of target area is experimental value: *ArmHeight*=10.

4.2. Experimental Result

Based on the parameters set in the previous subsection, we track the right arm motion using the tracking algorithm based on JCMM. In each experiment, the count of particle for tracking each joint is 50, 100, 150, and 200; respectively, the count of particle for all joints is 100, 200, 300, and 400.

Table 1 is the comparison of *mean error, Mean,* and *error variance, Std.,* between the ground truth and the prediction value of the right wrist joint under different count of particle using our algorithm in X direction, Y direction and Z direction. The Eqn. 14 is represented for *mean error*. The Eqn. 15 is represented for *error variance*.

$$Mean = \sum_{i=1}^{T} (x_t - X_t) / T \quad (14)$$

$$std = \sqrt{\sum_{i=1}^{T} (x_t - Mean)^2 / T} \quad (15)$$

In Eqn. 14 and Eqn. 15, the frames of test video is described as T, and T=796. x_t is the prediction value and X_t is the ground truth at frame t.

From Table 1, the *mean error* and *error variance* between the prediction and ground truth have not evidently changes as the particle count of all joint increasing. Then we can draw the conclusion that the count of particle for all joints can not affect the tracking result of our algorithm. Fig. 3 shows the tracking results of 3D arm motion by our algorithm as the count of particle for all joints is 400. It is no evidently different between the tracking results of our algorithm and the real pose of arm motion.

4.3. Experimental Analysis

The count of joints, which need be tracked in each tracking process, is defined as K. Each joint needs N particles to track the joint. Then our algorithm, particle filter based on JCMM, need KN particle for all joints and its computational complexity is E (KN). While standard particle filter generates N^K kinds of combination patterns of particle in whole state space, which is formulated as N^K kinds of motion states and the computational complexity of the standard particle filter is E (N^K). In our experiment, K is 2, and N will be 200, 150, 100, and 50. To track the right arm motion, the particle count of our algorithm, JCMMPF, is 100, 200, 300, and 400, while the standard particle filter will generate 40000, 22500, 10000, and 2500 kinds of combination pattern in state space.

Based on the parameters set in subsection 4.1, Table. 2 is the comparison of average time for tracking one frame image between two algorithms. Table. 3 shows the comparison of *mean error, Mean,* and *error variance, Std.,* between the prediction values using two algorithms and the ground truth in X direction, Y direction, and Z direction.

Following Table 2, the time-cost of JCMMPF is less than SPF as the particle count increasing, and the computational efficiency is improved obviously. As Shown in Table 3, the *Mean* and *Std* have not evident difference compared the ground truth with the tracking result JCMMPF, SPF.

5. Conclusion

There always is the computational efficiency problem for a large amount particle by the method of human 3D motion tracking based on Particle Filter.

The paper proposes 3D arm motion fast tracking algorithm. Based on the JCMM, the algorithm can transfer the global optimal search of the whole state space to the top-bottom search based on the joints under the case that the dimension of state space is unchangeable. In the process of tracking, the particle count is reduced by the prediction of each joint of JCMM. The experiment shows that the tracking result using our algorithm is not evident difference compared with the standard particle filter under the same dimension of state space. The algorithm can effectively apply to track 3D arm motion based on Particle Filter.

6. References

[1] Thomas B. Moeslund and E. Granum. A survey of computer vision-based human motion capture [J]. *Computer Visual and Image Understand*, 2001, vol. 81, pp. 231–268

[2] Thomas B. Moeslund, Adrian Hilton, Volker Kru¨ger. A survey of advances in vision-based human motion capture and analysis [J]. *Computer Vision and Image Understanding*. Vol.104, No.2, 2006, pp.90-126

[3] A. Blake and M. Isard. Condensation—Conditional Density Propagation for Visual Tracking [C]. *Int'l J. Computer Vision,* vol. 29, No. 1, pp. 5-28, 1998

[4] R. Urtasun, D. J. Fleet, P. Fua. 3D people tracking with Gaussian process dynamical models [C]. *Proc. Computer Vision and Pattern Recognition*, pp.238-245, Vol 1, 2006.

[5] P. Azad, A. Ude, R. Dillmann, G. Cheng. A full body human motion capture system using particle filtering and on-the-fly edge detection [C]. *4th IEEE/RAS International Conference on Humanoid Robots*, 2004, pp. 941 – 959

[6] Jamal Saboune, Francois Charpillet. Using Interval Particle Filtering for Marker less 3D Human Motion Capture [C]. *Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence*, page(s):7 pp, 2005

[7] Xinyu Xu, Baoxin Li. Learning Motion Correlation for Tracking Articulated Human Body with a Rao-Blackwellised Particle Filter [C]. *Proc. International Conference on Computer Vision*. Page(s):1-8, Vol 1, 2007

[8] J.K. Aggarwal, Q. Cai. Human motion analysis: a review [C]. *Proc of IEEE Nonrigid and Articulated Motion Workshop*. 16 June 1997 Page(s):90 – 102

[9] B. North, A. Blake, M. Isard, and J. Rittscher. Learning and classification of complex dynamics [J]. *IEEE Trans. PAMI*,25(9):1016-1034, 2000

[10] H. Sidenbladh, M. J. Black, L. Sigal. Implicit probabilistic models of human motion for synthesis and tracking [C]. *Proc. European Conference on Computer Vision*, pp. 784-800, Vol 1, 2002

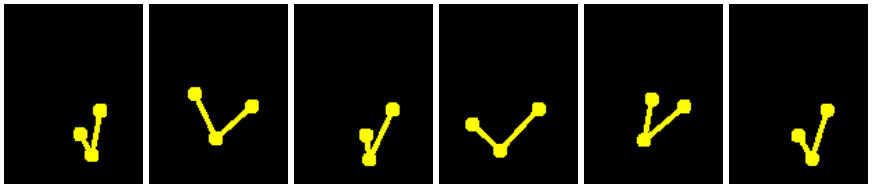
[11] R. Urtasun, D. Fleet and P. Fua. Monocular 3D tracking of the golf swing [C]. *Proc. Computer Vision and Pattern Recognition*, pp. 932-938, Vol 2, 2005

[12] H. Sidenbladh, M. J. Black, D. J. Fleet. Stochastic Tracking of 3D Human Figures Using 2D Image Motion [C]. *Proc. European Conference on Computer Vision*, pp. 702-718, Vol 2, 2000

[13] L. Sigal and M. J. Black. HumanEva: Synchronized video and motion capture dataset for evaluation of articulated human motion. TR CS-06-08, Brown University, 2006

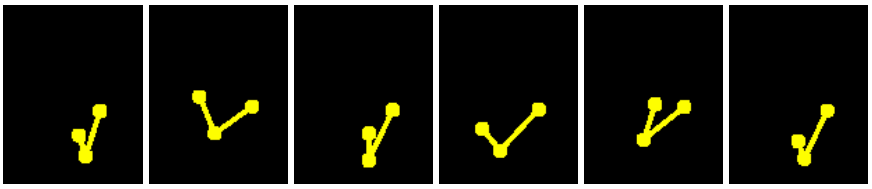
Table. 1 The *Mean Error* and *Std.* under different count of particle in our algorithm

		the count of particle for all joints			
		400	**300**	**200**	**100**
X	***Mean***	15.5854	14.8932	14.3304	15.8442
	Std.	13.5604	12.9412	12.7845	13.2104
Y	***Mean***	14.6420	14.8668	13.3681	14.1709
	Std.	13.0121	12.9268	12.0737	12.2031
Z	***Mean***	11.0992	11.3492	10.5854	11.6533
	Std.	11.1532	11.9329	10.8260	11.3366



Frame 100 Frame 200 Frame 300 Frame 400 Frame 500 Frame 600

Fig. 3(a) 3D animation for the tracking value of our algorithm



Frame 100 Frame 200 Frame 300 Frame 400 Frame 500 Frame 600

Fig. 3(b) 3D animation for the ground truth

Fig. 3 3D animation Comparison between the tracking result by our algorithm and ground truth

Table. 2 the time-cost comparison between two algorithms under different particle counts

		Time-Cost per frame image (ms)			
		***N*=200**	***N*=150**	***N*=100**	***N*=50**
JCMMPF	**Time (ms)**	3029	2066	1882	878
	Particle Count	400	300	200	100
SPF	**Time (ms)**	14653	8650	5253	2830
	Particle Count	40000	22500	10000	2500

Table. 3 the comparison of *Mean* and *Std.* for tracking right wrist between two algorithms

		N=200		**N=150**		**N=100**		**N=50**	
		JCMMPF	**SPF**	**JCMMPF**	**SPF**	**JCMMPF**	**SPF**	**JCMMPF**	**SPF**
X	***Mean***	15.5854	14.6005	14.8932	15.5477	14.3304	14.4146	15.8442	15.0641
	Std.	13.5604	13.2656	12.9412	13.5994	12.7845	12.7099	13.2104	13.2622
Y	***Mean***	14.6420	11.9950	14.8668	12.1771	13.3681	12.0101	14.1709	12.3643
	Std.	13.0121	10.8131	12.9268	10.8082	12.0737	10.8933	12.2031	11.1384
Z	***Mean***	11.0992	13.7927	11.3492	13.0867	10.5854	13.6143	11.6533	14.1985
	Std.	11.1532	12.4430	11.9329	11.7670	10.8260	12.5962	11.3366	12.5667