World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:2, No:5, 2008

Using Rao-Blackwellised Particle Filter Track 3D Arm Motion based on Hierarchical Limb Model

XueSong Yu JiaFeng Liu XiangLong Tang JianHua Huang

***Abstract*—**For improving the efficiency of human 3D tracking, we present an algorithm to track 3D Arm Motion. First, the Hierarchy Limb Model (HLM) is proposed based on the human 3D skeleton model. Second, via graph decomposition, the arm motion state space, modeled by HLM, can be discomposed into two low dimension subspaces: root nodes and leaf nodes. Finally, Rao-Blackwellised Particle Filter is used to estimate the 3D arm motion. The result of experiment shows that our algorithm can advance the computation efficiency.

***Keywords*—**Hierarchy Limb Model; Rao-Blackwellised Particle Filter; 3D tracking

I. Introduction

ALTHOUGH caught the many researchers' attention because of its widely application, human 3D tracking is still a challenging task because of the exponentially increased computational complexity in terms of the freedom degrees.

Moeslund [1] [2] et al. classified the existing research pose estimation algorithms into learning-based algorithm and model-based algorithm. The model-based category builds the human motion model with human prior knowledge and the human motion constraints, and the use of stochastic sampling techniques in model-based analysis-by-synthesis to obtain the optimal estimation based on the Bayesian network framework. As the nonlinear filter algorithm based on the Bayesian estimation framework, particle filter [3] has been widely application [4] [5] in the area of human 3D motion estimation. Deutsher [6] et al. proposed the annealed particle filter to track the human 3D motion. Markov Chain Monte Carlo [7] [8] is utilized to solve the particle degeneracy problem. Recently, the structure graphical model [9] has been used to facilitate the estimation of human 3D motion. Based on the structure graphical model, Wei [13] proposed a decentralized framework. And facilitated by graph decomposition, they derived a novel Bayesian conditional density propagation rule.

In recent years, Rao-Blackwellised Particle Filter [10] [11] [12] has been widely studied in human tracking. Xinyu [12] et al. learns motion correlation using the Partial Least Square, and proposed the RBPF-PLS algorithm 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. Because of the restrictive dependency, RBPF can not be directly used for 3D human tracking.

Focused on these problems, the paper proposes the 3D arm motion tracking algorithm using Rao-Blackwellised Particle Filter based on Hierarchy Limb Model. The arm motion spaces can be divided into two parts: root variables and leaf variables by the Hierarchy Limb Model. Rao-Blackwellised Particle Filter can be used to track arm motion via the decomposition. As a result, our algorithm can advance the computational efficiency because of the lower dimensionality of the search space and the reduced amounts of particle.

The paper is organized as follows. Section 2 describes the Hierarchy Limb Model. The 3D arm tracking algorithm based on RBPF is proposed in Sention 3. Experimental results and analysis are shown in Section 4, and finally concludes the paper.

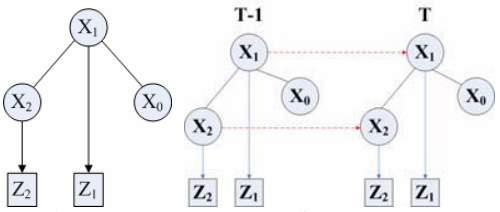
II. Hierarchy Limb Model

The arm can be represented by an arm graphical model such as shown Fig. 1 (a). The circle nodes corresponds to a part of right arm, such as the right upper arm and the right lower arm. The square nodes are the observation values assiocated with each circle nodes. The undirected links represent physical constraints among different parts of the right arm. The directed link from a part's state to its associated observation represents the local observation likelihood. In order to describe the motion of an articulated object, we accommodate the state dynamics by a dynamical graphical model such as shown in Fig. 1 (b). It contains two consecutive time frames.The directed links between consecutive states represents the dynamics translation from time t-1 to time t.

We model the arm as 2 cylinders connected at revolute joints, and denote the state of each part of arm at time t by $x_{i,t}$, where i=1,2 is the index of two parts of arm. The arm motion state space is 12-dimensional, including 6D for the global (shoulder) position and orientation, 3D for the elbow as $x_{1,t}$, 3D for the wrist as $x_{2,t}$.

XueSong Yu is with the Harbin Institute of Technology, Harbin, Heilongjiang China (corresponding author to provide phone: 086-013304819068;; e-mail: yyz001@hit.edu.cn).

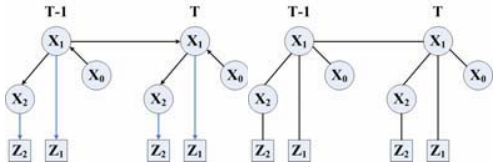
JiaFeng Liu, XiangLong Tang and JianHua Huang were with the Harbin Institute of Technology, Harbin, Heilongjiang China.



(a) (b)

Fig. 1 Arm structure graphical model, (a) Graphical model for arm, (b) Dynamical graphical model for arm motion analysis.

According to the characteristics of arm motion, the motion of any node of the arm only interacts with its children nodes. For example, the motion of lower arm is not constrained by any limbs but only the motion of corresponding upper arm. Fig. 2 (a) show the descomposion result for the right arm in Fig. 1 (b), and Fig. 2 (b) is the associated *moral graph* via the *separation theorem* and the charactics of the dynamic Markov network [13].



(a) (b)

Fig. 2 Arm graphical model decomposition, (a) The decomposition of dynamical graphical model; (b) Corresponding moral graph.

Using the arm hierarchy model, the problem of tracking right arm motion can be formulated as the prediction of x at time t.

III. 3D Arm Tracking with RBPF

By the HLM, arm motion can be decomposed into two parts: $x_{1,t}$ (upper arm) and $x_{2,t}$ (lower arm). We denote $x_{1,t}$ as *root variable* R_t, and $x_{2,t}$ as *leaf variable* L_t. In the algorithm, we marginalize the lower arm pose variable by using the motion speed correlation matrix V. The strategy of Rao-Blackwellised Particle Filter is to partition the full state space into two parts, R_t (*root variable*) and L_t (*leaf variable*), so the $P(L_t \mid R_t, Z_{1:t})$ is a distribution that can be computed exactly conditional on R_t, and the distribution $P(X_t \mid Z_{1:t})$ can be estimated using sample methods, such as Markov Chain Monte Carlo.

The posterior probability distribution for the right arm motion can be factorized by Equation (1).

$$P(R_t, L_t \mid Z_{1:t}) = P(L_t \mid R_t, Z_{1:t}) P(R_t \mid Z_{1:t}) \tag{1}$$

Given the speed correlation matrix V, and the image measurements $Z_{1:t}$, the Equation (1) can be approximated by the following expression:

$$\begin{aligned} P(R_t, L_t \mid Z_{1:t}, V) &= P(L_t \mid R_t, Z_{1:t}, V) P(R_t \mid Z_{1:t}) \\ &\propto P(L_t \mid R_t, Z_{1:t}, V) P(Z_t \mid R_t) \int P(R_t \mid R_{t-1}) P(R_{t-1} \mid Z_{t-1}) dR_{t-1} \\ &\propto P(Z_t \mid R_t, L_t) P(L_t \mid R_t, Z_{1:t}, V) \int P(R_t \mid R_{t-1}) P(R_{t-1} \mid Z_{t-1}) dR_{t-1} \end{aligned} \tag{2}$$

In Equation (2), the integral parts is propagated the *root variables* R_t utilizing the state transition model according to the standard Particle Filter. The *leaf variables* L_t can be predicted analytically using Kalman Filtering prediction. The image likelihood model, as $P(Z_t \mid R_t, L_t)$, is computed in the state space, including *root* and *leaf variable*.

A. State Transition Model

In particle filter theoretical framework, the state transition model, by which particle is generated, is described as shown Equation 3.

$$x_t = x_{t-1} + v_t, \quad v_t \sim N(\mu, \Sigma) \tag{3}$$

Where v_t is drawn from the Gaussian noise that the expectation μ is a 3×1 scalar, which is defined as the motion speed of *root variables*, and the variance Σ is the 3×3 diagonal matrix. Utilizing hard prior, we can eliminates the particles corresponding to implausible arm pose to reduce the search space.

B. Prediction for the Leaf Variables

The leaf variables can be predicted by the use of the motion speed correlation matrix V using the Kalman Filter while the root variables are propagated. Condition on the root variables and the correlation matrix V, the *leaf variable* L_t are described as the following expression:

$$\begin{aligned} L_t &= AL_t + V(R_t^- - \hat{R}_{t-1}) + \xi_{t-1}, p(\xi) \sim N(0, \Phi) \\ Z_{L,t} &= HL_t + \zeta_t, p(\zeta) \sim N(0, \Psi) \end{aligned} \tag{4}$$

Where ξ represent the process noise and ζ represent the measurement noise. V is defined as the motion speed correlation matrix. Then the leaf variables are predicted by Equation (5).

$$\begin{aligned} L_t^{i-} &= AL_{t-1}^i + V(R_t^{i-} - R_{t-1}^i) \\ P_t^{i-} &= AP_{t-1}^i A^T + \Phi \end{aligned} \tag{5}$$

In Equation (5), i is defined as the particle index. In Equation (4) and Equation (5), A and H are assumed to be the diagonal matrix.

According to the Equation (4) and Equation (5), Kalman Updates can be computed as the following expression:

$$K_t^i = P_t^{i-} H^T (H P_t^{i-} H^T + \Psi)$$
$$L_t^i = L_t^{i-} + K_t^i (Z_{L,t} - H L_t^{i-}) \quad (6)$$
$$P_t^i = (1 - K_t^i H) P_t^{i-}$$

C. Image Likelihood

The observation likelihood model is represented for the matching relationship between the human appearance model and the features subtracted from the image among the particle filter theoritical framework. In this section, color distribution and image edge information are used to calculate the matching similarity between the human appearance model and the features subtracted from the image.

Color distributions are used as target models as they achieve robustness against non-rigidity, rotation and partial occlusion. The weighted color histogram, which consists of m=8×8×8=512 bins, is choosen and calculated in HSV color space to decrease the effection of the illumination.

The projection quadrilateral of the limb shape is defined as Dr, y is the point which is the projection of the origin of the local coordinate system on image plane, and color distribution is defined as $p_c^y = \{p_c^{y,u}\}_{u=1}^m$. For any pixel point $\tilde{x}^i \in Dr$, $p_c^{y,u}$ can be calculated as following expression:

$$p_c^{y,u} = C \sum_{j=1}^{n} k\left(\left\| \frac{y - \tilde{x}^i}{S_{Dr}} \right\|\right) \delta[h(\tilde{x}^i) - u] \quad (7)$$

Where $\delta(x)$ is the *Delta* function, S_{Dr} is the area of Dr, C is the normalized constant. The Bhattacharyya distance is used to calculate the simility between two weigthed color histograms.

IV. Experimental Result and Analysis

We have done experiments to track the right arm motion using the HumanEva data sets [14]. The experiment chooses the right arm motion color video made in the front to reduce the self-occlusions.

A. Experimental Result

Table 1 is the comparison of *mean error*, *Mean*, and *error variance*, *Std.* between the ground truth and the prediction value of the right lower arm under different count of particle using our algorithm in X direction, Y direction and Z direction. The Equation (8) is represented for *mean error*. The Equation (9) is represented for *error variance*.

$$Mean = \frac{\sum_{t=1}^{T} (x_t - X_t)}{T} \quad (8)$$

$$Std = \sqrt{\frac{\sum_{t=1}^{T} (x_t - Mean)^2}{T}} \quad (9)$$

In Equation (8) and Equation (9) the frames of test video is described as T, and T=796. x_t is the prediction value, X_t is the ground truth at frame t.

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

TABLE I
THE ERROR MEAN AND STD. OF TRACKING RESULT USING OUR ALGORITHM

Direction	X	Y	Z
Mean	13.4550	11.4206	10.0029
Std.	10.4712	9.1102	10.3215

B. Experimental Analysis

The count of joints, which need be tracked in each tracking process, is defined as K. Each joint needs N_i particles to track the joint,i=1,..K. HLMRBPF nees $\prod_{i=1}^{K} N_i$ particles for all limbs and its computational complexity is $E(\prod_{i=1}^{K} N_i)$. 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, N_1=200, N_2= N_1/10, N=200. Then time-cost of HLMRBPF tracking one every frame image is 5908ms, while the standard Particle Filter (SPF) needs 14768ms.

Table II show the comparison of *mean error*, *Mean*, *error variance,* and *Std.* between the prediction values using two algorithms and the ground truth in X direction, Y direction, and Z direction.

TABLE II
THE COMPARISON OF ERROR MEAN AND STD. OF TRACKING RESULT USING HLMRBPF AND SPF

	X Direction		Y Direction		Z Direction	
	Mean	***Std.***	***Mean***	***Std.***	***Mean***	***Std.***
HLM-RBPF	12.87	10.25	13.10	11.88	12.30	11.99
SPF	15.55	13.60	12.18	10.81	13.09	11.77

V. Conclusions

The paper proposes 3D arm motion fast tracking algorithm. Based on the HLM, 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 HLMRBPF. 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.

REFERENCES

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

[2] Thomas. B. Moeslund, Adrian. Hilton and Volker. Kru¨ger. "A survey of advances in vision-based human motion capture and analysis". *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". *Int'l J. Computer Vision,* vol. 29, No. 1, pp. 5-28, 1998.

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

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

[6] J. Deutscher and I. Reid. "Articulated body motion capture by stochastic search". *Proc. International Journal of Computer Vision*, pp 185- 205, Vol 2, 2005.

[7] C. Sminchisescu and B. Triggs. "Kinematic Jump Processes for Monocular 3D Human Tracking". *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 69-76, 2003.

[8] MunWai. Lee and Isaac. Cohen. "A Model-Based Approach for Estimating Human 3D Poses in Static Images". *IEEE Transactions on Pattern Analysis And Machine Intelligence*, Vol. 28, NO. 6, June 2006.

[9] L. Sigal, S. Bhatia, S. Roth, M. J. Black, and M. Isard. "Tracking loose-limbed people". *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 421-428, Vol 1, 2004.

[10] X. Xu, B. Li. "Adaptive Rao-Blackwellised Particle Filter and its evaluation for tracking in surveillance". *IEEE Trans. Image Processing*, 16(3):838-849, 2007.

[11] K. Murphy and S. Russell. *Rao-Blackwellised particle filtering for dynamic Bayesian networks*. In *Sequential Monte Carlo Methods in Practice*, A. Doucet, et al Eds. New York: Springer-Verlag, 2001, ch. 24.

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

[13] Wei Qu and Dan Schonfeld. "Real-time decentralized articulated motion analysis and object tracking from videos". *IEEE Trans Image Processing*, VOL. 16, NO. 8, Aug. 2007.

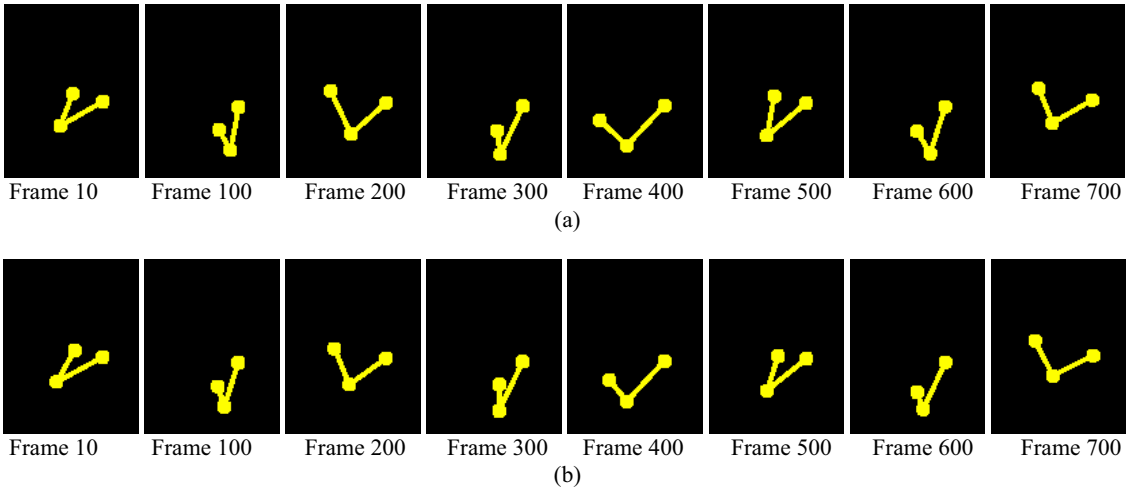


Fig. 3 3D animation Comparison between the tracking result by our algorithm and ground truth; (a) 3D animation for the tracking value of our algorithm; (b) 3D animation for the ground truth.