



## Estimate 3D Arm Motion with Hierarchical Limb Model

Xuesong Yu (Corresponding author)

Department of Computer Science & Engineering

Harbin Institute of Technology

92 Xi Da Zhi Street, Harbin, 150001, China

Tel: 86-451-8641-7602-601 E-mail: yyz001@hit.edu.cn

Jiafeng Liu

Department of Computer Science & Engineering

Harbin Institute of Technology

92 Xi Da Zhi Street, Harbin, 150001, China

Tel: 86-451-8641-3631-802 E-mail: jeffery@hit.edu.cn

*The research is financed by The National Natural Science Foundation of China. No. 60672090*

### Abstract

Focusing on the problem of low computation efficiency in the process of tracking human 3D motion, an algorithm for Estimating 3D arm motion with Hierarchy Limb Model (HLM) is proposed. In our algorithm, the Hierarchy Limb Model (HLM) is proposed based on the human 3D skeleton model. Facilitated by graph decomposition, the arm motion state space, modeled by Hierarchy Limb Model (HLM), can be decomposed into low dimension subspaces. The Top-Down search strategy and the Particle Filter are used to tracking the arm motion, thus the amount of particle in tracking can be reduced. To handle server self-occlusions, the weighted color histogram and image contour are used to modeling the observation likelihood function. The result of experiment shows that our algorithm can advance the computation efficiency and handle effectively self-occlusions.

**Keywords:** 3D Arm Motion, Hierarchy Limb Model, Top-Down strategy, Particle Filter

### 1. Introduction

The human 3D motion estimation has received a significant amount of attention in recent years driven by its wide applications such as video surveillance, human activity analysis, computer animation, etc. But human 3D motion Estimation is still a challenging task because of the exponentially increased computational complexity in terms of the degrees of freedom of the object and the severe image ambiguities incurred by frequent self-occlusions.

Moeslund (Moeslund, 2001, Moeslund, 2006) *et al.* comprehensively summarized the research results of the human 3D motion estimation. They classified the existing research pose estimation algorithms into learning-based algorithm and model-based algorithm. Furthermore, learning-based algorithm for the human 3D motion estimation can be separated into two strategies. The one is the category based on human motion prior knowledge learning. The category of solution uses strong motion prior to constrain the search into the most likely region of the parameter space (Urtasun, 2006, North, 2000). One way to cope with the high-dimensional state-space is to learn low-dimensional latent variable models. In this category, many algorithm are applied to learn the latent variable models, such as *Principle Component Analysis* (Sidenbladh, 2000, Urtasun, 2005, Sidenbladh, 2002), *Relevance Vector Regression* (Agarwal, 2004), nonlinear Gaussian process dynamic models (Urtasun, 2005) and the Partial Least Square (Xinyu, 2007), etc. 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 one nonlinear filter algorithm based on the Bayesian estimation framework, the use of particle filter (Blake,1998) has been widely application (Azad, 2004, Saboune, 2005) in the area of human 3D motion

estimation. Deutsher (Deutsher, 2005) *et al.* proposed the annealed particle filter to track the human 3D motion. *Markov Chain Monte Carlo* (Sminchisescu, 2003, MunWai, 2006,) is utilized to solve the particle *degeneracy* problem. Recently, the structure graphical model (Sigal, 2004) has been used to facilitate the estimation of human 3D motion.

Although the lower computation efficiency and ambiguities have been solved effectively, the learning-based methodology can't track the random activity in natural scene but tracking the motion which has been learned. Furthermore, a great amount of samples and expensive time cost are always the challenge in learning process because of the complexity of human motion. The model-based methodology needs a great amount of sample to describe the human motion. The number of sample and time cost show exponential growth with the dimension of human motion state space.

Wu (Wu, 2003) *et al.* proposed a mean field Monte Carlo algorithm based on a dynamic Markov network for 2D articulated body tracking, and decomposes the human motion state space into multiple linear subspaces via the MFMC. Wei (Wei, 2007) *et al.* proposed decentralized articulated graphical model to improve computation efficiency based on 2D human motion tracking. With the improvement of decentralized articulated graphical model, we propose the particle filter based on the hierarchy limb model for estimating the arm 3D motion. Facilitated by graph decomposition, the hierarchy limb model decomposes the right arm motion state space into two linear subspaces. Based on the hierarchy limb model, our algorithm searches each subspace with the particle filter. 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. To handle efficiently handle the severe self-occlusion problem, our algorithm propose the angle relation model between upper arm and lower arm in the arm motion process. To build the angle relation model, the least square is used to fitting linear of the arm contour.

The paper is organized as follows. Section 2 describes the hierarchy limb model, and the tracking algorithm based on particle filter is proposed. The iamge likelihood function are described in Sention 3. Experimental results and analysis are shown in Section 4, and finally concludes the paper.

## 2. Frameworks

In this section, we describe the key components of the arm motion estimation framework, namely, the hierarchy limb model, and the estimation framework based on the particle filter.

### 2.1 Arm Hierarchy Limb Model

Our algorithm uses a generic model that represents the arm structure. The arm model, as illustrated in Figure. 1, consists of two components: *kinematics model* and *structure graphical model*.

#### 2.1.1 Arm Kinematics Model

Each limb of arm kinematics model includes two components: kinematics vector and shape vector. Kinematics vector consists of six parameters and is used to the prediction of arm 3D motion. Shape vector is used to describe the approximated arm 3D shape, including seven parameters.

To represent kinematics state of each limb, we define the kinematics vector as  $x = \{T, \theta_x, \theta_y, \theta_z\}$ , where  $T$  is global translation vector, and the rotation vecotr,  $\theta = \{\theta_x, \theta_y, \theta_z\}$ , presents the angles that the limb rotate around three coordinate axes as shown Figure. 1(c). The limb shape vector include three 3D cylinder constants and four shape constants in image plane. Three cylinder constants include the height  $l$  of cylinder, the radius  $r$  of cylinder, and the origin  $O$  of local coordinate system where  $O \in SO(3)$ . The shape vector  $e$  is equal to the vertex set of the quadrilateral that is the approximation of the cylinder on the image plane, as liustrated in Figure. 2. We define the shape vector  $e$  as  $e = \{e_1, e_2, e_3, e_4\}$ , where  $e_i = (x_i, y_i)$ ,  $i = 1, 2, 3, 4$ . As a result, the shape vector can be defined as  $\Phi = \{l, r, O, e_1, e_2, e_3, e_4\}$ .

The arm state space is represented as  $x = \{x_0, x_1, x_2\}$ . Where  $x_0$  is formatted by the 3D coordinate triplets that is the ground truth of the right shoulder,  $x_1$  is presented for the right upper arm, and  $x_2$  is presented for the right lower arm. We denote model of each limb as follows:

$$x_i = \{T_i, \theta_{i,x}, \theta_{i,y}, \theta_{i,z}\}, i = 1, 2 \quad (1)$$

$$\Phi_i = \{l_i, r_i, O_i, e_{i,1}, e_{i,2}, e_{i,3}, e_{i,4}\}, i = 1, 2 \quad (2)$$

#### 2.1.2 Arm Structure Graphical Model

The right arm can be represented by an arm graphical model such as shown Figure. 3 (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 associated 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 Figure. 3 (b). It contains two consecutive time frames. The directed links between consecutive states represents the dynamics translation from time  $t-1$  to time  $t$ .

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. Figure. 4 (a) show the decomposition result for the right arm in Figure. 3 (b), and Figure. 4 (b) is the associated *moral graph* via the *separation theorem* and the characteristics of the dynamic Markov network.

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

## 2.2 Tracking with Particle Filter

Via the arm hierarchy limb model, we propose the tracking framework based on particle filter and the particle generation.

### 2.2.1 Tracking framework

Using the arm hierarchy limb model, the right arm motion can be decomposed into the motion of two parts, while the state space is decomposed into three low-dimensionality state spaces in the tracking process. Based on the decomposition, the overall state space optimization process can be formulated as the state subspace optimization of each limb following by the top-down search strategy via the Particle Filter.

The state parameter  $x_t$  of right arm motion at time  $t$  is represented by the form of joint state as shown Eqn. 3:

$$x_t \square \{x_{i,t}\}_{i=1}^2 = \{x_{1,t}, x_{2,t}\} \quad (3)$$

Where  $i$  is the index of parts. Assmued the father node of  $i$ th node defined as  $F(i)$ , the observation state of all limbs is respresented as  $z_t = \{z_i^j\}_{i=1}^2$ . The posterior probability distribution for the right arm motion is given by:

$$P(x_t | z_t) = P(x_{0,t} | z_{0,t}) \prod_{i=1}^2 P(x_{i,t} | x_{F(i),t}, z_{0,t}) \quad (4)$$

Where  $x_{i,t}$  is defined as the prediction value of  $i$ th parts at time t,  $x_{F(i),t}$  is the observation value of the father associated with  $x_{i,t}$ ,  $P(x_{0,t} | z_{0,t})$  is defined as the maximum a posterior (MAP) for the right shoulder where is the constant. As a result,  $P(x_{i,t} | x_{F(i),t}, z_{i,t})$  can be approximated by the following expression:

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

Where  $N$  is count of particle,  $K$  is the index of particles,  $x_{i,t-1}^k$  is the  $k$ th particle of  $i$ th part at time  $t-1$ ,  $w_{i,t-1}^k$  is the weight value associated with  $x_{i,t-1}^k$  and can be modified as Eqn. 6.

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

Substituting Eqn. 6 into Eqn. 3, the joint state  $x_t$  can be approximated as shown Eqn. 7:

$$x_t = \{x_{i,t}\} \approx \left\{ \sum_k w_{i,t}^k \times x_{i,t}^k \right\}_{i=1}^2 \quad (7)$$

### 2.2.2 Particle Generation

In this section, we describe in details the praticle generation based on the arm hierarchy limb model and tracking framework.

In particle filter theoretical framework, the state transition model, by which particle is generated, is described as shown Eqn. 8.

$$x_t = x_{t-1} + v_t, \quad v_t \square N(\mu, \Sigma) \quad (8)$$

Where  $v_t$  is the Gaussian noise that the expectation  $\mu$  is a  $3 \times 1$  scalar, which is defined as the motion speed of part, and the variance  $\Sigma$  is the  $3 \times 3$  diagonal matrix.

The motion speed of part  $i$  depends on the speed of part  $i$  at time  $t-1$  and the motion speed of its father part  $F(i)$ . We represent the motion speed of part  $i$  at time  $t$  as the row vector  $v_{i,t} = (v_{i,t}^x, v_{i,t}^y, v_{i,t}^z)$ , where superscript of each element of vector is defined as the rotation angle speed of X axis, Y axis, and Z axis while each element of the row vector is

independent. The row vector  $v_{F(i),t} = (v_{F(i),t}^x, v_{F(i),t}^y, v_{F(i),t}^z)$  is defined as the motion speed of  $F(i)$ . If  $t < 3$ ,  $v_{i,t}$  is confirmed as following equation:

$$v_{i,t} = \begin{cases} (0, 0, 0) & t = 0 \\ (x_{i,t} - x_{i,t-1}, y_{i,t} - y_{i,t-1}, z_{i,t} - z_{i,t-1}) & t = 1, 2 \end{cases} \quad (9)$$

If  $t \geq 3$ ,  $v_{i,t}^x, v_{i,t}^y, v_{i,t}^z$  can be calculated independently by Eqn. 10.

$$\begin{aligned} v_{i,t}^x &= \alpha_{i,t-1} \times (v_{i,t-1}^x, v_{F(i),t}^x)' \\ v_{i,t}^y &= \beta_{i,t-1} \times (v_{i,t-1}^y, v_{F(i),t}^y)' \quad t \geq 3 \\ v_{i,t}^z &= \gamma_{i,t-1} \times (v_{i,t-1}^z, v_{F(i),t}^z)' \end{aligned} \quad (10)$$

Where, the coefficient  $\alpha_{i,t-1}, \beta_{i,t-1}, \gamma_{i,t-1}$  are the  $2 \times 1$  scalar obtained by least squares method, which is represented as the speed coefficient vector of part  $i$  at time  $t-1$  in the X axis, Y axis, and Z axis.

### 3. Image Likelihood Function

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 theoretical 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.

#### 3.1 Color Distribution Likelihood

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  $m=8 \times 8 \times 8=512$  bins, is chosen and calculated in HSV color space to decrease the effect of the illumination.

The projection quadrilateral of the set  $e$  of the limb shape vector 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:

$$p_c^{y,u} = C \sum_{j=1}^n k \left( \left\| \frac{y - \tilde{x}^i}{S_{Dr}} \right\| \right) \delta[h(\tilde{x}^i) - u] \quad (11)$$

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 similarity between two weighted color histograms.

#### 3.2 Edge Likelihood

We split the arm from the background with the method combined with the background subtract and skin detector. The least square is used to fit the image edge points obtained from the edge to calculate the slope of long edge of contour.

(1) Human contour is subtracted from the background by background difference, and mathematical morphology is used to distill the whole human contour.

(2) Split the right arm contour from the human contour using the ground truth of initial frame.

(3) The point set of contour  $E_C$  can be detected from the arm contour via the contour detection methods. The point set of contour  $E_C$  can be divided into two subsets by the skin detector, including the point set of right upper arm contour,  $E_C^U$ , and the point set of right lower arm contour,  $E_C^L$ .

(4)  $L_1$  is the slope of the long edge of the right upper arm by the fitting linear using least square method, and  $L_2$  associated with the right lower arm.

Assumed at time  $t$ ,  $l_{i,t}^k$  is the slope of the long edge of the quadrilateral, which is the projection of the set  $e_{i,t}^k$  of the limb shape vector of part  $i$ . Then  $l_{i,t}^k$  can be modeled as a Gaussian distribution as following:

$$p(l_{i,t}^k) = \exp\left(-\frac{(l_{i,t}^k - \mu)^2}{2\sigma^2}\right) / (\sigma\sqrt{2\pi}) \quad (12)$$

Where  $i$  is the index of limbs,  $k$  is the index of particle,  $\mu$  is equality to  $l_{i,t}$  that is the slope the long edge associated with limb  $i$ ,  $\sigma$  is the covariance of the slope set  $l_{i,t}$ .

### 3.3 Self-occlusion

To handle with the server self-occlusion, we proposed the algorithm based on the angle relation model for two intersecting lines modeled by the method described in section 3.2.

The angle between the upper arm and the lower arm is defined as  $\omega$ , and  $\omega_{th}$  is the constant threshold determined by empirical value. So the posterior probability  $P(z_{i,t}^k | x_{i,t}^k, x_{F(i),t})$  can be represented as shown Eqn. 13.

$$P(z_{i,t}^k | x_{i,t}^k, x_{F(i),t}) = \begin{cases} p_{i,t}^{k(u)} \times p(l_{i,t}^k) & \omega > \omega_{th} \\ p(l_{i,t}^k) & \omega \leq \omega_{th} \end{cases} \quad (13)$$

## 4. Experimental Results and Analysis

### 4.1 Experimental Design

We have done experiments to track the right arm motion using the HumanEva data sets (Sigal, 2006), which were captured at 25 fps by Leonid et al. of American Brown University using 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 u 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 experimental video. Then the Eqn. 5 can be simplified as the following equation:

$$P(x_t | z_t) = \prod_{i=1}^2 P(x_t^i | x_t^{F(i)}, z_t^i) \quad (14)$$

In subsection 3.3, the angle threshold is the empirical value:  $w_{th} = \pi / 6$ .

### 4.2 Experimental Result

Based on the parameters set in the previous subsection, we track the right arm motion using particle filter based on arm hierarchy limb model. In each experiment, the count of particle for tracking each limb 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 lower arm under different count of particle using our algorithm in X direction, Y direction and Z direction. The Eqn. 15 is represented for *mean error*. The Eqn. 16 is represented for *error variance*.

$$Mean = \frac{\sum_{t=1}^T (x_t - X_t)}{T} \quad (15)$$

$$Std = \sqrt{\frac{\sum_{t=1}^T (x_t - Mean)^2}{T}} \quad (16)$$

In Eqn. 15 and Eqn. 16, 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 1, 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. 5 shows the tracking results of 3D arm motion by our algorithm as the count of particle for limbs 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 arm hierarchy limb model (AHLMPF), need  $KN$  particles for all limbs 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, AHLMPF, 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 show the comparison of average time for tracking one frame image between two algorithms. Table. 3 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.

Following Table 2, the time-cost of AHLMPF 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 of AHLMPF and SPF.

## 5. Conclusions

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

- A, Blake and M, Isard. (1998). Condensation—Conditional Density Propagation for Visual Tracking [C]. *Int'l J. Computer Vision*, vol. 29, No. 1, pp. 5-28, 1998.
- A. Agarwal and B. Triggs. (2004). 3D Human Pose from Silhouettes by Relevance Vector Regression, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 882-888, 2004.
- B. North, A. Blake, M. Isard, and J. Rittscher. (2000) . Learning and classification of complex dynamics [J]. *IEEE Trans. PAMI*,25(9):1016-1034, 2000.
- C. Sminchisescu, and B. Triggs. (2003). Kinematic Jump Processes for Monocular 3D Human Tracking. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 69-76, 2003.
- H. Sidenbladh, M. J. Black, and D. J. Fleet. (2000). Stochastic Tracking of 3D Human Figures Using 2D Image Motion [C]. *Proc. European Conference on Computer Vision*, pp. 702-718, Vol 2, 2000.
- H. Sidenbladh, M. J. Black, and L. Sigal. (2002). Implicit probabilistic models of human motion for synthesis and tracking [C]. *Proc. European Conference on Computer Vision*, pp. 784-800, Vol 1, 2002.
- J. Deutscher and I. Reid. (2005). Articulated body motion capture by stochastic search. *Proc. International Journal of Computer Vision*, pp 185- 205, Vol 2, 2005.
- Jamal, Saboune and Francois, Charpillet. (2005). 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.
- L. Sigal and M. J. Black. (2006). HumanEva: Synchronized video and motion capture dataset for evaluation of articulated human motion. TR CS-06-08, Brown University, 2006.
- L. Sigal, S. Bhatia, S. Roth, M. J. Black, and M. Isard. (2004). Tracking loose-limbed people. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 421-428, Vol 1, 2004.
- MunWai, Lee and Isaac, Cohen. (2006). 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.
- P. Azad, A. Ude, R. Dillmann, and G. Cheng. (2004). 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.
- R. Urtasun, D. Fleet and P. Fua, Monocular. (2005). 3D tracking of the golf swing [C]. *Proc. Computer Vision and Pattern Recognition*, pp. 932-938, Vol 2, 2005.
- R. Urtasun, D. J. Fleet, and P. Fua. (2006). 3D people tracking with Gaussian process dynamical models [C]. *Proc. Computer Vision and Pattern Recognition*, pp.238-245, Vol 1, 2006.
- Thomas, B. Moeslund, Adrian, Hilton, and Volker, Krüger. (2006). 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.
- Thomas, B. Moeslund, and E, Granum. (2001). A survey of computer vision-based human motion capture [J]. *Computer Visual and Image Understand*, 2001, vol. 81, pp. 231–268.
- Wei Qu, and Dan Schonfeld. (2007). Real-time decentralized articulated motion analysis and object tracking from

videos. *IEEE Trans Image Processing*, VOL. 16, NO. 8, Aug. 2007.

Xinyu, Xu, and Baoxin, Li. (2007). 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.

Y. Wu, G. Hua, and T. Yu. (2003). Tracking articulated body by dynamic Markov network. *IEEE Conf. Computer Vision*, pp 1094 - 1101, Vol.2, 2003.

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

		the count of particle for all limbs			
		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

Table 2. the time-cost comparison between AHLMPF and SPF under different particle counts

		Time-Cost per frame image (ms)			
		N=200	N=150	N=100	N=50
AHLMPF	Time (ms)	2908	2014	1543	782
	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 using AHLMPF and SPF

		N=200		N=150		N=100		N=50	
		AHLMPF	SPF	AHLMPF	SPF	AHLMPF	SPF	AHLMPF	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

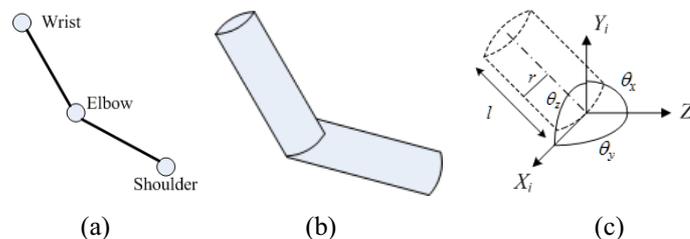


Figure 1. Arm hierarchical limb model represents the articulated structure of the arm.

- (a) Kinematics description characterized by the joints position,
- (b) arm shape approximated by two cylinders,
- (c) Local coordinate system of limb  $i$ .

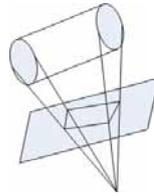


Figure 2. Arm shape image planar approximation via the projection equation.

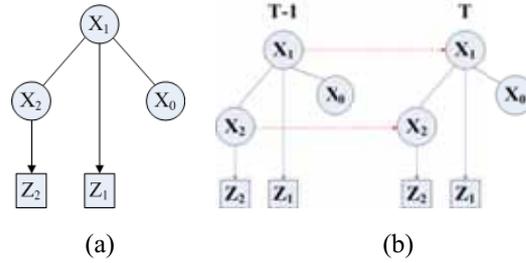


Figure 3. Arm structure graphical model, (a) Graphical model for arm, (b) Dynamical graphical model for arm motion analysis.

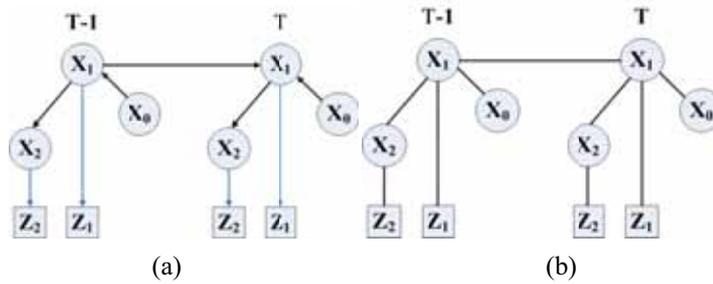


Figure 4. Arm graphical model decomposition, (a) The decomposition of dynamical graphical model; (b) Corresponding moral graph.

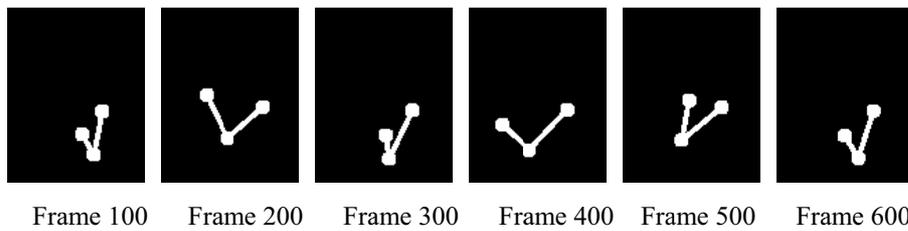


Figure. 5(a) 3D animation for the tracking value of our algorithm

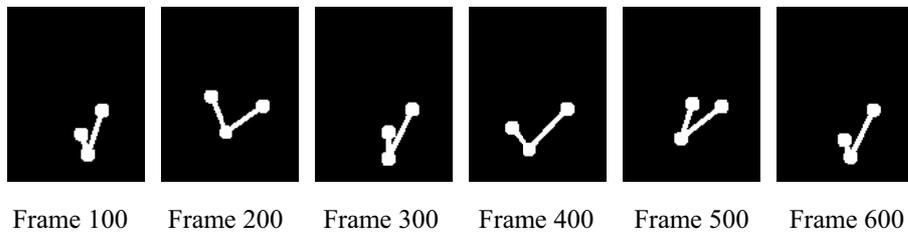


Figure. 5(b) 3D animation for the ground truth

Figure 5. 3D animation Comparison between the tracking result by our algorithm and ground truth