The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Upper Body Tracking for Human-Machine Interaction with a Moving Camera

Yi-Ru Chen, Cheng-Ming Huang, and Li-Chen Fu, *IEEE fellow*

Abstract—**This research presents an upper body tracking method with a monocular camera. The human model is defined in a high dimensional state space. We hereby propose a hierarchical structure model to solve the tracking problem by SIR (Sampling Importance Resampling) particle filter with partitioned sampling. The image spatial and temporal information is used to track the human body and estimate the human posture. When doing the human-machine interaction, a static monocular camera may not get plenty of information from 2D images, so we must move the camera platform to a better position for acquiring more enriched image information. The proposed upper body tracking technique will then adjust to estimating the human posture during the camera moving. To validate the effectiveness of the proposed tracking approach, extensive experiments have been performed, of which the result appear to be quite promising.**

I. Introduction

With the rapid development of inexpensive and high technology machines, robots are more and more popularly seen to be equipped with different capabilities. Human-machine interaction becomes an increasingly important issue. There are many ways for humans to communicate with the robots, like using the normal computer device, such as keypad and mouse, microphone (for sound), camera (for vision), force sensor (for touch), laser and ultrasound (for depth information).

In recent years, vision-based analysis on human's motion analysis has become a popular problem due to a number of potential applications, like biomechanical analysis, computer animation, surveillance system, pedestrian alarm on approaching vehicles, human-machine interface, human-robot interaction, etc. This research field has been widely surveyed in the literature [1-5]. In particular, Moeslund *et al.* [3] presented an overview which decomposes the flow of the visual system into several processes, including the initialization, tracking procedure, pose estimation, and recognition. The first step of their proposed algorithm to analyze human's motion is human detection. In different situations, we may pay attention on the different parts of the body. The work in [5] gave an overview of human tracking via tracking of some body parts such as face, hands, fingers, etc. The results of tracking different body parts may lead to different applications. For example, tracking the full human body can be applied in security-sensitive environment, such as banks, parking lots and airports. Leg tracking can be used in gait recognition, human identification and trajectory analysis. Hand tracking has a number of applications such as gesture and sign language recognition for human-computer interaction. Face tracking can be utilized for human identification, meeting analysis, and camera's automated focus. In fact, the upper body tracking, which is also the focus of this paper, is an important subject in the field of human-robot interaction especially when home robots come into play, because the image information will be sufficient to analyze the user information, like face and arm. Thus, the upper body posture analysis becomes an instruction during the human-robot interaction.

There are many ways for human or posture detection. In tradition, the works estimate the human by marking the human body parts [6] or by asking the human to wear some special clothes [7]. However, this is inconvenient and can only be used in special situations with the complete equipments being installed. To improve the methods, many researches provide other alternative approaches without relying on markers worn or attached onto the human body. The learning-based method [8] is applied to detect the human position. Using the background subtraction or foreground segmentation techniques, the silhouette-based methods [9] can be utilized to detect humans with a static camera. Lastly, the dense disparity map is constructed for estimating human posture by using the 3D information from a multi-camera system [10]. In order to combine the posture detection algorithm with the human-robot interface, the tracking system may not be able to assume that the environment is with static background and camera since both the human and the camera mounted on the robot may move while interacting. Without such assumption, it becomes more difficult for the system to fulfill the tracking objective since now the system can not always rely on use of background subtraction to get a foreground silhouette.

This paper presents a vision-based tracking algorithm that tracks the 3D upper body posture of a human. Concerning the processing time, applying some detection technique in each frame during the course of tracking a human generally will not be efficient. Combining those tracking algorithms can reduce the complexity and save the computation cost. Besides, tracking human beings is one of the most difficult tasks, because human's motions are fast, nonlinear, unpredictable,

Manuscript received March 1, 2009. This work was supported bby National Science Council, Taiwan, under Grant NSC 97-3114-E-002-002.

Y. R. Chen and C. M. Huang are with the Department of Electrical Engineering, National Taiwan University, Taiwan, ROC.

L. C. Fu is with the Department of Electrical Engineering and Computer Science and Information Engineering, National Taiwan University, Taiwan, ROC (e-mail: lichen@ntu.edu.tw).

and there are so many kinds of possible human posture. So far in the literature, the particle filter, which is a popular tracking algorithm, can successfully solve the non-Gaussian state estimation problems in nonlinear systems. As a consequence, this paper deals with the human tracking problem by employing the particle filter as a basic tracking technique. Besides that, we also adopt the optical flow to find the motion part and the movement in the image so that the human model can be robustly identified, which in turn will facilitate accomplishment of tracking.

This paper is organized as follows. In Section II, we first introduce our human model and particle filter. And then in Section III, we describe the tracking algorithm and the feature we used in this paper. In Section IV, we describe the scenario of human-robot interaction with monocular camera. In Section V, we demonstrate some experimental result to validate the effectiveness of the proposed tracking approach. Finally, we conclude this paper in Section VI with some relevant discussion.

II. Human Model and Tracking Filter

A. Human Model Definition

As discussed in the previous work [5], there are a few kinds of human model which can be categorized into 2D models and 3D models. In 2D models, they usually try to fit the body contour by ellipse and rectangle. There are two kinds of 3D models, namely, stick model and volumetric model. The stick model also named skeleton model in which skeletons are connected by the joints. The volumetric model defines a 3D model to represent the 2D contour, such as a cylinder to represent a rectangle, a cone to represent a triangle, and a sphere to represent a circle.

Here we use 3D stick model to determine the joint positions. As shown in Fig. 1, $\mathbf{p}_0$ is the origin of the coordinate system. In general, the size of every part is proportional to the face size. We model the head as an ellipse whose ratio between length of the minor axis to that of the major axis is 1:1.2. But the shoulder width and arm length are more different from one person to another. In order to improve the accuracy of the model, we need an initial pose as shown in Fig. 1 to decide the shoulder width and the arm length of one specific user. After taking the proper size from the user, we can start the tracking system. The human model consists of head, shoulders, elbows and wrists. Next, we use the particle filter to estimate the particle state which comprises joint angles of each body part. The shoulder joint (θ, ϕ, ρ) can be represented in the spherical coordinate system S which is centered at shoulder position. We perform coordinate transformation through the following operations. First, we rotate the x-axis in S with angle θ about the y-axis, and then rotate x-axis with angle ϕ about y-axis, finally rotate y-axis with angle ρ about z-axis. After that, we translate the current coordinate system to elbow coordinate system H, and rotate y-axis about z-axis with angle γ. The details of the transformation mentioned above all are illustrated in Fig. 2 and 3, with all the relationships and limitations of angles provided.

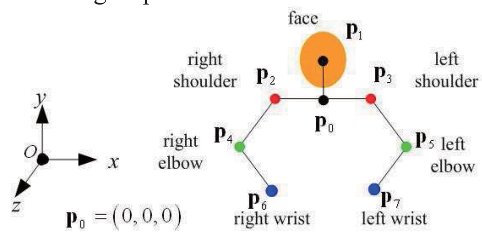


Fig. 1. Human model with the initial posture.

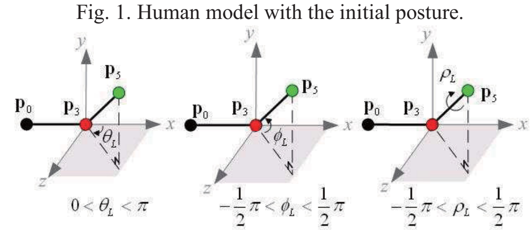


Fig. 2. Joint angle for the shoulder with limited range.

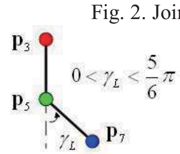


Fig. 3. Joint angle for the elbow with limited range. The stick between $\mathbf{p}_5$ and $\mathbf{p}_7$ lies on the plane decided by the rotation angle ρ_L.

The 3D human model will be projected onto the image plane, and then we evaluate the joint angles to fit the human model. We have to decide the relationships of position, scale and orientation between the human model and the image plane, as shown in Fig. 4. The position of $\mathbf{p}_0$ on the image coordinate and the scale r of the human model are decided by the position $\mathbf{p}_1 = (x, y)$ and the size of the head, because the head detector is much more reliable than other limb detector. And then, the human body will be predicted with an orientation α. After determining the relationships of position, scale and orientation, we can project the human model onto the image plane.

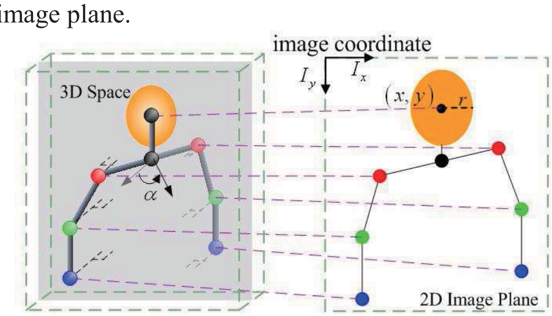


Fig. 4. Parameters (x, y, r, α) of upper body projection.

B. Sampling Importance Resampling (SIR) Particle Filter

Particle filter, which is based on Bayesian filter with sequential Monte Carlo method, is a solution of tracking problem based on statistical estimation. According to Bayesian Theorem, the conditional probability function is

$$p(\mathbf{x}_t | \mathbf{z}_{0:t}) \propto p(\mathbf{z}_t | \mathbf{x}_t) \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) \, p(\mathbf{x}_{t-1} | \mathbf{z}_{0:t-1}) \, d\mathbf{x}_{t-1}, \quad (1)$$

where $\mathbf{x}_t$ denote the state vector in $\mathcal{R}^n$, and $\mathbf{z}_t$ denote the observation vector in $\mathcal{R}^m$. In the Bayesian framework, we want to recursively calculate the belief in the state $\mathbf{x}_t$ with all past observations $\mathbf{z}_{0:t} = \{\mathbf{z}_i,\ i = 1, \cdots t\}$, so it is required to construct the posterior $p\left(\mathbf{x}_t \middle| \mathbf{z}_{0:t}\right)$. Particle filter approximates the Bayesian filter by sequential Monte Carlo using a set of N weighted particles $\left\{\mathbf{x}_t^{(i)}, w_t^{(i)}\right\}_{i=1}^{N}$. The weights of the samples are normalized, i.e. $\sum_{i=1}^{N} w_t^{(i)} = 1$, then the posterior at time t can be approximated as

$$p\left(\mathbf{x}_t \middle| \mathbf{z}_{0:t-1}\right) \approx \sum_{i=1}^{N} w_t^{(i)} \delta\left(x_t - x_t^{(i)}\right), \tag{2}$$

where $\delta(\cdot)$ is the kernel function, such as Dirac's delta function or Gaussian kernel.

One of the general particle filters is sampling importance resampling (SIR) particle filter [11], the particles are generated from the posterior as a proposal importance density function from previous frame. The resampling step solves the degeneracy problem by the temporal information and draws the particles with better observation interpretation but it causes lots of particles with the same state and losses the diversity of the particles. This problem is known as impoverishment phenomenon. To avoid the impoverishment phenomenon and improve the degeneracy problem, some algorithms employ complex sampling strategies or specific prior knowledge about the objects, such as partitioned sampling [12], layered sampling [13], annealed importance sampling [14], particle swarm optimization [15], and scatter search technique [16].

The impoverishment phenomenon is more obviously in high dimensional state space. Furthermore, in high dimension state space, the complexity and inefficiency grows exponentially. In order to cover the high dimensional state space, the number of particles has sufficiently large. In this paper, we combine the SIR particle filter with partitioned sampling not only to overcome the degeneracy problem and impoverishment phenomenon, but also deal with the high dimensional state space.

C. Partitioned Sampling

Partitioned sampling is a strategy to alleviate the intense need for a large number of particles in a high dimensional state space. This strategy decomposes the state space into two or more partitions, and sequentially applies appropriate weighted resampling operation for each partition. It has been used in the tracking of articulated objects [12], fusing the sensors providing information about different state space dimensions [17], and human motion capture [18]. Partitioned sampling is suitable for objects with hierarchical structure. [19] proposes a hierarchical part-based detection, based on the higher reliability of the head-and-torso detector. This paper follows the idea, so we decompose the high dimensional state space of the upper body into two parts, that is, for the upper body posture tracking, we first localize the head and torso position, next estimate the arm posture. We don't decompose the arm posture into upper arm and forearm, because the connection between upper arm and forearm is tight. Even though we don't model upper arm and forearm separately, the number of particles based on our method is still much less than the number of particles without decomposing the high dimensionality of the whole upper body.

III. Upper Body Tracking

Particle filter with the partitioned sampling is used here to solve the upper body tracking problem. For starting the system, we set an initial pose as shown in Fig. 1 to construct the body characteristic of the individual user, such as shoulder width and arm length. While the system detects the human face, it will try to fit the initial pose. If the system has well fitting, then it starts the upper body tracking procedure. Otherwise, the system determines the person does not want to interact with the robot. After setting these parameters, the 3D human posture which has mentioned in Section II.A can be evaluated for each particle. Then measure the weight of each particle by likelihood function, also called fitness and weighting function which influences the accuracy and complexity most. There are kind of common features, such as edge, contour, symmetry, texture, color, silhouette, depth, motion. If we can combine several features to get the high accuracy rate, we will loss the efficiency of the system. Here, we use edge, color and optical flow information to evaluate the likelihood function. Knowing the weight of each particle, we can evaluate the posterior probability density function and estimate the target position.

In the high dimensional state space of human model, we use partitioned sampling to draw particles more efficiently so the particles number can be reduced. The most reliable part detector in human body is the head detector [19]. We first decide the state of head position (x, y), scale r, and orientation α of human body. And then we can use partitioned sampling to generate other state $(\theta_R, \phi_R, \rho_R, \gamma_R)$ for right arm and $(\theta_L, \phi_L, \rho_L, \gamma_L)$ for left arm. The state

$$\mathbf{x} = \left[(x, y, r, \alpha), (\theta_R, \phi_R, \rho_R, \gamma_R), (\theta_L, \phi_L, \rho_L, \gamma_L)\right], \tag{3}$$

can model the upper body posture. By projecting the human model on the image plane, we can get the image observation for each state $\mathbf{x}$ and evaluate the weight of each particle in full state space.

But in clutter environment, there should contain much edge information in the background. The likelihood function is not enough to make the accurate decision, even with the human body constrains. So in the different situation, we use different method to detect the motion body part. We separate the scenario into static camera and motion camera which can decide by the camera motion feedback information. According to the camera movement, we use several ways to capture motion region and enhance the edge in the motion region.

A. Color Histogram

The color histogram is computed in YCrCb color space. The reference color model is constructed with Cr-Cb channel to exclude illumination influence which is mainly contained in Y channel. The 2D $N_1 \times N_2$ bins reference color histogram denoted as $\mathbf{h}_{ref}$

$$\mathbf{h}_{ref} = \left\{ h_{ref}\left(i, j\right) \right\}_{i=1, j=1}^{i=N_1, j=N_2} . \tag{4}$$

The color likelihood function is measured with the Bhattacharyya distance between the reference histogram and the histogram corresponding to each particle:

$$f_{color}\left(\mathbf{h}_{ref}, \mathbf{h}_{\mathbf{x}'}\right) = \sqrt{1 - \sum_{i=1}^{i=N_1} \sum_{j=1}^{j=N_2} h_{ref}\left(i, j\right) h_{\mathbf{x}'}\left(i, j\right)} , \tag{5}$$

where $\mathbf{h}_{\mathbf{x}'}$ is the histogram corresponding to each particle, $\mathbf{x}'$ is the substate about the measurement part. For example, if now we estimate the right arm, then $\mathbf{x}'$ will contain $\left(\theta_R, \phi_R, \rho_R, \gamma_R\right)$. The likelihood function of color histogram is defined as the following

$$p_{color}\left(\mathbf{z}_t \mid \mathbf{x}'\right) \propto \exp\left[-f_{color}^{\ 2}\left(\mathbf{h}_{ref}, \mathbf{h}_{\mathbf{x}'}\right)\right] . \tag{6}$$

B. Enhanced Edge Contour

The edge information can be utilized to distinguish the outlier of each body part. The contour of head is modeled by ellipse template, and the contour of torso and arms are modeled by rectangle template. Define the contour matching function as

$$f_{contour}\left(\mathbf{x}'\right) = \frac{\sum_{i=1}^{N} d_i}{N} , \tag{7}$$

where d_i is the distance between maximum edge and contour template, and N is the number of sample points along the contour. The smaller d_i represents that the edge approaches the contour template more closely. As a result, the likelihood of enhanced edge contour is measured by

$$p_{contour}\left(\mathbf{z}_t \mid \mathbf{x}'\right) \propto \exp\left[-f_{contour}^{\ 2}\left(\mathbf{x}'\right)\right] . \tag{8}$$

The edge obtained from single image frame is obvious distinguish each human part and background in the simple environment. However, in the complex environment, there may exist strong edge in background. The strong edge is easy to attract the concentrate of particles. In order to enhance the edge around the human, the motion detection is used. Two kinds of methods are used for motion detection with static or motion camera. With the static camera, the background image can not sure be taken, so the moving edge is proposed. Two continuous images are compared to find the different part. With the motion camera, the optical flow can evaluate the different movement object.

C. Part Movement

The color and edge only fit the model in 2D, but the depth information is lost. To improve the 3D model, the optical flow is a general feature in a motion video. The feature point $\mathbf{f}_t^{(k)}$ [20] can extracted from the image and compute the displacement by tracking these feature points at frame t. Here, the feature points are tracked by KLT algorithm [20], which evaluates the sum-of-squared-difference between image frames. The movement of optical flow and the particle state variation for different body part can be estimate the likelihood function.

First, by the result of initialization, the feature points is identified into different body part, such as head, torso, right arm, and left arm. Next, we can generate a feature point set B_i for each body part:

$$B_i = \left\{ \mathbf{f}_t^{(k)} = \left(x_t^{(k)}, y_t^{(k)}\right) \middle| k \in \text{part } i \right\} , \tag{9}$$

where $i = \text{H}$ (head), T (torso), LA (left arm), RA (right arm) or B (background). By knowing the position $\mathbf{f}_t^{(k)}$ of each feature points, we can approximate the motion model. To simplify our motion model, we do not consider the depth between objects in the background and the camera. We only use a simple translation model. Because we have the information of which feature point belongs to which body part, so we first evaluate the distance $l_{t-1}^{(k)}$ and angle $\varphi_{t-1}^{(k)}$ between feature points $\mathbf{f}_{t-1}^{(k)}$ belong to part i and the center $\mathbf{O}_{t-1,i}$ of the part i with

$$l_{t-1}^{(k)} = \left| \mathbf{f}_{t-1}^{(k)} - \mathbf{O}_{t-1,i} \right| , \tag{10}$$

$$\varphi_{t-1}^{(k)} = \angle\left(\mathbf{f}_{t-1}^{(k)} - \mathbf{O}_{t-1,i}\right) . \tag{11}$$

And then, we can evaluate the similarity $d_{t,i}\left(\mathbf{x}'\right)$ between the feature points and the particle state of each part:

$$d_{t,i}\left(\mathbf{x}'\right) = \frac{1}{2K_i} \sum_{k \in i} \left(\frac{l_t^{(k)}\left(\mathbf{x}'\right) - l_{t-1}^{(k)}}{l_{t-1}^{(k)}} + \frac{\varphi_t^{(k)}\left(\mathbf{x}'\right) - \varphi_{t-1}^{(k)}}{\varphi_{t-1}^{(k)}} \right) , \tag{12}$$

where i =H, T, LA and RA. So we can utilize the movement similarity to generate the likelihood function $p_{movement}\left(\mathbf{z}_t \mid \mathbf{x}\right)$ for the particle filter:

$$p_{movement}\left(\mathbf{z}_t \mid \mathbf{x}'\right) \propto \exp\left[-f_{movement}^{\ 2}\left(\mathbf{x}'\right)\right] , \tag{13}$$

in which

$$f_{ovement}\left(\mathbf{x}'\right) = \frac{1}{4} \sum_{i=\text{H, T, LA, RA}} d_{t,i}\left(\mathbf{x}'\right) . \tag{14}$$

By using the movement information from feature points, we can improve the accuracy of the human posture.

D. Overall Likelihood

The overall likelihood function is the multiply of each cues mentioned above

$$p\left(\mathbf{z}_t \middle| \mathbf{x}'\right) \propto p_{color}\left(\mathbf{z}_t \middle| \mathbf{x}'\right)^{C_{color}} \cdot p_{contour}\left(\mathbf{z}_t \middle| \mathbf{x}'\right)^{C_{contour}} \cdot p_{movement}\left(\mathbf{z}_t \middle| \mathbf{x}'\right)^{C_{movement}} , \tag{15}$$

where the order C_{color}, $C_{contour}$ and $C_{movement}$ is the weight of each cue. Usually the weight are the same, only in special situation, the weight should be modified, such as, $C_{contour}$ should be larger in the complex environment, and C_{color} should be larger if the background contains the object with similar color.

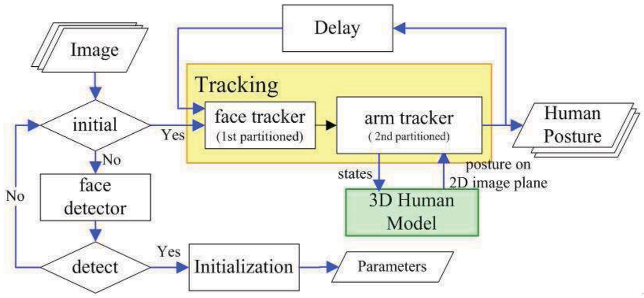


Fig. 5. The system flow char.

The overall tracking framework is summarized in Fig. 5. The system is triggered with an initial pose as shown in Fig. 1, and characteristic parameters of the human model are evaluated from the detected pose. Partitioned sampling decomposes the state space into three subspaces. The first space covers the state describing the head and torso orientation, and the second and third subspaces cover the states for the left and right arms. The particles in the first partition are drawn from the likelihood function with the color and contour. The particles in the second and third partitions are drawn from the likelihood function with the color, contour and movement.

IV. Human Robot Interaction

The natural human-robot interaction should be followed the ease body language for the user in face-to-face interaction. Using the upper body tracking algorithm, the human posture can be analyzed. Only with the monocular camera, the loss of the depth information is difficult to reconstruct the 3D model. However the robot can modify the 3D posture estimation by change the view angle. The motion information changes the human orientation and the position, so the result can analyze from the well understanding position.

A general instruction is pointing gesture as the movement of the arm towards a pointing target. The robot can estimate the human posture and move to the target at the same time by correcting the posture with 3D human model. After starting the tracking system, the robot can also track this user and estimate his posture until the system over. Without the assumption for static camera, the user can interact with robot at the comfortable position, such as sit on sofa, or spacious room. This upper body tracking algorithm is a technique which can combine with other gesture recognition and decide the robot behavior.

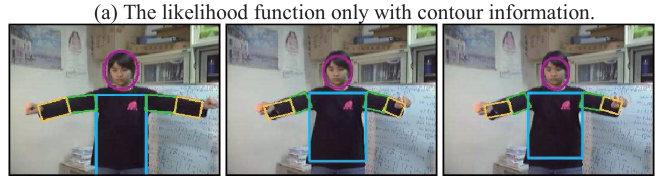
V. Experimental Result

The upper body tracking algorithm has been tested on several image sequences, including static camera and moving camera platform. In the initialization, 4 parameters, which are the length and width for upper arm and forearm, for human model will be determined. Then the 12 states in (3) for modeling the 3D human posture will be estimated hierarchically. When doing the tracking with partitioned sampling, 50 particles are used in the first space, and 80 particles are employed in the second and third subspaces. In the following experimental results, we present some snapshots during the upper body tracking in different scenarios.

(a) Result of initialization (b) Frame #361

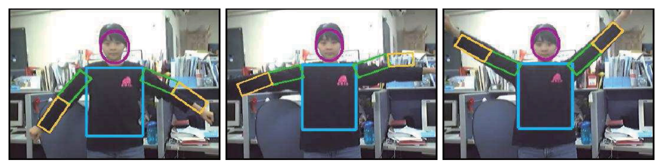
Fig. 6. The likelihood function only with contour information.

(a) The likelihood function only with contour information.

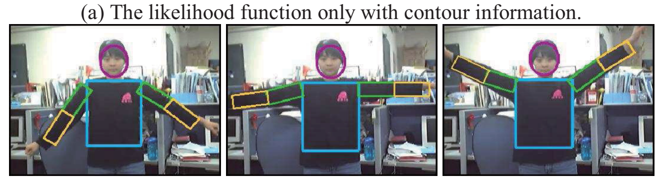


(b) The likelihood function with contour, color, and movement information.

Fig. 7. Effect of the motion information for the movement in 3D.



(a) The likelihood function only with contour information.



(b) The likelihood function with contour, color, and movement information.

(c) Edge Image for this scenario.

Fig. 8. Effect of the movement information in complex environment with noisy edge.

First, we show the result only use the contour information as described in Section III.B with static camera in the simple environment in Fig. 6. The result is good enough for the simple action without orientation and movement in 3D. To demonstrate the effect of the movement of feature points with static camera in the simple environment, the result is shown is Fig. 7. In Fig. 8(a), the joint angle between upper arm and forearm is drift. The joint angle is decreasing due to the contour is equally strong in the wrong posture. In fact, the joint angle is still 180 degree which shown in Fig. 8(b). The displacement obtained from feature points has another benefit. In the test sequence, there have much noise edge in the complex environment, so the tracker without feature points is easily disturbed by the background. With the assistance of the movement information from feature points, the proposed tracker can successfully estimate the arm position.

The human size can not evaluate from the distance between the human to camera, but can be modified by the scale of the head, because the size of every part is in proportion to the face size. As shown in Fig. 9, even the human close or move away, the tracker can capture the human with right size.

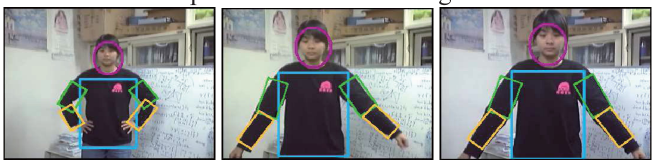


Fig. 9. The face scale influence.

The human posture is projected on the image plane. There have some similar image observation by different posture according to the projection, even the movement of feature points can not solve the projection problem. Here, we can move the camera platform to a better position, and then correct the human posture. For example, when the shoulder angle $\theta < 60^{\circ}$, the tracker can fit the right joint position, but over this range, even though the tracker still match with the arm, the joint position drifted away. By moving the camera platform, the joint position can be modified to the right position. Figure 10 presents this situation, and the proposed tracker gradually obtain the correct posture of right arm.

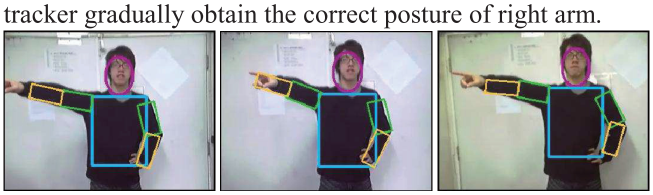


Fig. 10. Moving camera platform to modified the human posture.

VI. CONCLUSION

This paper presented an upper body tracking algorithm based on particle filter with partitioned sampling. The model uses color, edge and movement information. Unlike other tracking algorithm which requires the background information to cut a human silhouette, we do not need a static background. In order to apply the algorithm to human-machine interaction, our approach can handle the situation with moving camera platform and utilize the different view angle to correct the 3D human posture. The computational time for the overall system now is about 15 fps, which achieves the near real-time performance.

In this paper, the human model only limited the joint angle, so sometimes it may generate impossible posture. In the future, we plan to combine learning based human model, such as trained by SVM (support vector machine) or RVM (relevance vector machine), to prove the human posture estimate. Moreover, we will examine and analysis our system in more complicated scenarios of human-robot interaction and aim to reduce the overall computational time.

ACKNOWLEDGMENT

This work was supported by the National Science Council of the Republic of China under Grant NSC 97-3114-E-002-002.

REFERENCES

[1] J. K. Aggarwal and Q. Cai, "Human motion analysis: a review," in *Nonrigid and Articulated Motion Workshop, 1997. Proceedings., IEEE*, 1997, pp. 90-102.

[2] D. M. Gavrila, "The visual analysis of human movement: A survey," *Computer Vision and Image Understanding,* vol. 73, pp. 82-98, 1999.

[3] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding,* vol. 81, pp. 231-268, 2001.

[4] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding,* vol. 104, pp. 90-126, 2006.

[5] J. J. Wang and S. Singh, "Video analysis of human dynamics--a survey," *Real-Time Imaging,* vol. 9, pp. 321-346, 2003.

[6] J. Lee, J. Chai, P. S. A. Reitsma, J. K. Hodgins, and N. S. Pollard, "Interactive control of avatars animated with human motion data," *ACM Transactions on Graphics (TOG) archive,* vol. 21, pp. 491-500, 2002.

[7] K. Dong-Wan and J. Ohya, "Estimating Postures of a human wearing a multiple-colored suit based on color information processing," in *Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on*, 2003, pp. I-261-4 vol.1.

[8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, pp. 886-893 vol. 1.

[9] W. Liang, T. Tieniu, N. Huazhong, and H. Weiming, "Silhouette analysis-based gait recognition for human identification," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 25, pp. 1505-1518, 2003.

[10] N. Jojic, M. Turk, and T. S. Huang, "Tracking self-occluding articulated objects in dense disparity maps," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 1999, pp. 123-130 vol.1.

[11] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *Signal Processing, IEEE Transactions on,* vol. 50, pp. 174-188, 2002.

[12] J. MacCormick and M. Isard, "Partitioned sampling, articulated objects and interface-quality hand tracking," *Lecture Notes in Computer Science,* vol. 1843, pp. 3-19, 2000.

[13] J. MacCormick and A. Blake, "A probabilistic exclusion principle for tracking multiple objects," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 1999, pp. 572-578 vol.1.

[14] J. Deutscher, A. Blake, and I. Reid, "Articulated body motion capture by annealed particle filtering," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, 2000, pp. 126-133 vol.2.

[15] Q. Wang, L. Xie, J. Liu, and Z. Xiang, "Enhancing particle swarm optimization based particle filter tracker," *Computational Intelligence,* pp. 1216-1221, 2006.

[16] J. J. Pantrigo, A. S. Montemayor, and R. Cabido, "Scatter Search Particle Filter for 2D Real-Time Hands and Face Tracking," *Image Analysis and Processing,* pp. 953-960, 2005.

[17] P. Perez, J. Vermaak, and A. Blake, "Data fusion for visual tracking with particles," *Proceedings of the IEEE,* vol. 92, pp. 495-513, 2004.

[18] J. Deutscher, A. Davison, and I. Reid, "Automatic partitioning of high dimensional search spaces associated with articulated body motion capture," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, pp. II-669-II-676 vol.2.

[19] R. Navaratnam, A. Thayananthan, P. Torr, and R. Cipolla, "Hierarchical part-based human body pose estimation," in *Pattern Analysis, Statistical Modelling and Computational Learning* Oxford, UK, 2005.

[20] S. Jianbo and C. Tomasi, "Good features to track," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, 1994, pp. 593-600.