An Observation Likelihood Model Integrating Edge Feature with Silhouette Feature

Bingchao Liu
School of Information Science and Engineering, University of Jinan, Jinan, 250022, China
Shandong Provincial Key Laboratary of Network based Intelligent Computing, Jinan, 250022, China
Email: 574833179@qq.com

Zhiquan Feng*, Aili Sang
School of Information Science and Engineering, University of Jinan, Jinan, 250022, China
Shandong Provincial Key Laboratary of Network based Intelligent Computing, Jinan, 250022, China
Email: ise_fengzq@ujn.edu.cn

***Abstract*—3D hand tracking in recent years has become a hot spot in Human Computer Interaction(HCI), and it also exists many difficulties. Our research roots in the observation model which is used to measure similarity between hand image and hand model for hand tracking. Firstly, using edge feature extracted from hand image and hand model projection contour to get a probability function, which measures the similarity between hand model and hand image according their chamfer distance. Secondly, using skin color segmentation algorithm to get binary hand image, at the same time we can get hand model silhouette from hand model projection points. After getting the binary hand image and hand model silhouette we then build another probability function based silhouette feature. Finally we combine the two functions to build the observation model to evaluate the probability of hand state according current observation. Applying the observation model into the weight calculation stage of particle filter can effectively increase calculation accuracy. On the other hand, we introduce hand constraints to hand model, utilize these constraints in sampling stage not only can reduce sampling degree of freedom(DOF) but also can revise the incorrect samples.**

***Index Terms*—Hand Tracking; Edge Detection; Distance Transform; Skin Color Segmentation; Observation Model; Particle Filter**

I. INTRODUCTION

Hand tracking as a natural interactive way can harmoniously express people's interactive intention. Hand tracking has important applications in many fields, such as sign language recognition, navigation, robot control, motion sensing game. Combine data glove with position tracker can accurately track hand movement, but these equipments are expensive and not convenient to wear. Vision-based hand tracking method is simple and easy to realize. Due to hand movement is complicated and non-rigid, recover three dimensional hand state according vision information still remains huge challenges.

Vision-based hand tracking methods are mainly divided into two types: appearance-based and model-based. Appearance-based methods [3, 17, 18] firstly build mappings between image feature and hand state, and then directly recover hand state from current image according these mappings. As hand movement is complex and infinite, it needs a huge database to save these mapping and not possible to build mappings for all possible hand state. Model-based methods firstly need to build three dimensional hand model, and then achieve hand tracking by comparing the similarity between hand model and hand image. In order to measure the similarity, observation model based on one or more features must be built. Stenger et al. [1] generates 2D profiles of hand model using elegant tools from projective geometry, and hand model is estimated with and Unsecented Kalman filter, which minimizes the geometry error between the profiles and edges extracted from the images. Stenger [2] utilizes edge and skin color to measure the similarity between hand template and hand image. O. B. Henia et al. [12] proposes a dissimilarity measurement function which combines the non-overlapping surface and the directed chamfer distance. Y. Wu et al. [4] analyzed natural hand movement to build a set of basis configurations, and then integrated hand constraints into the tracking algorithm which is based on Monte Carlo method. They employ both edge and silhouette feature to measure the likelihood of hypotheses. H. Zhou and T. S. Huang [20] proposed a new feature called *likelihood edge*, which combines color histogram and edge feature in the observation model. Wen-Yan Chang et al. [21] proposed an appearance-guided tracking algorithm. They treat the state space vector whose observation is known as an attractor, then theses attractors serve as prior knowledge to guide hand tracking. By applying prior appearance information, they can recover rapid motions that are difficult to solve by using simple particle filter. To get out of spurious local minima, M. Bray et al. [22] proposed a novel SMD tracker which is based on a rapid stochastic gradient descent approach with adaptive step sizes. N. Stefanov et al. [23] combines behavioural knowledge with annealed particle filtering to achieve robust tracking of hands. Distinct gestures are handled by using different models, and their transitions are learned with an variable-length Markov

doi:10.4304/jsw.8.6.1405-1410

model which represents the high-level structure and temporal ordering of gestures. They identify fingertips, palms and other parts of the hand, in order to reduce the complexity of evaluating the likelihood of a particular model configuration.

This work is also in the frame of model-based hand tracking, particle filter technology is used as tracking tool. On the one hand, an observation model combining edge with silhouette is built, and then apply it into calculating particle's weight. On the other hand, hand constraints are introduced into 3D hand model so as to reduce degree of freedom(DOF). The combination of these two parts can increase particle's accuracy and effectiveness so as to increase tracking accuracy.

II. 3D Hand Model

Model-based hand tracking firstly needs to construct a 3D hand model built by simulating real human hand structure [8], as depicted in Fig. 1(a). It defines each finger's name and each joint angle's name. The 3D model used in this paper was proposed by Feng [6], as displayed in Fig. 1(b). The whole hand model has 26 DOFs: 20 local DOFs and 6 global DOFs. The joint angles connecting palm and finger have two DOFs with flexision/extension and abduction/adduction movements, other joint angles only have one DOF with flexision/extension movement.

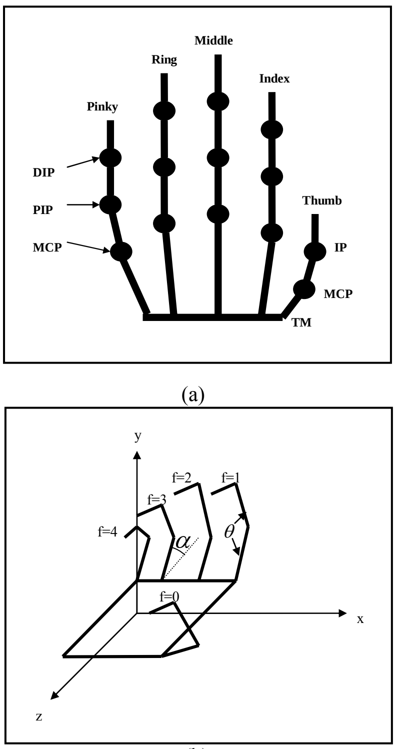


Figure 1. (a)The structure of hand. (b)The adopted hand model in this paper.

A. Hand Constraints

As hand movements are not arbitrary, the hand model must be constrainted according hand constraints. A lot of work has been done in hand constraints [4, 5, 8, 10, 11]. These work are mainly about the relationship between joint angles. As hands between different people exist many differences, we apply some universal hand constraints into hand model. The hand constraints used here are divided into two types: static hand constraints and dynamic hand constraints. Static hand constrains [8, 10] means the scope of joint angle without extra force. Dynamic hand constraints are the relationships between joint angles when hand grasps or opens, they are also divided into relationships in the same finger and relationships beweent different fingers. The hand constrains are listed in table 1.

TABLE I.
Hand Constraints

	Static Constraints	Dynamic Constraints
Thumb	$0^{\circ} \le \theta_{TM} \le 90^{\circ}$, $0^{\circ} \le \theta_{IP} \le 60^{\circ}$, $0^{\circ} \le \theta_{MCP} \le 80^{\circ}$, $0^{\circ} \le \alpha_{TM} \le 60^{\circ}$	$\theta_{IP} = \theta_{MCP}$, $\theta_{TM} = \frac{1}{3}\theta_{MCP}$
Index Middle Ring Pinky	$0^{\circ} \le \theta_{MCP} \le 90^{\circ}$, $0^{\circ} \le \theta_{PIP} \le 90^{\circ}$, $0^{\circ} \le \theta_{DIP} \le 90^{\circ}$	$\theta_{DIP} = \frac{2}{3}\theta_{PIP}$
Index Ring Pinky	$-15 \le \alpha_{MCP} \le 15^{\circ}$	$\alpha_{Index-MCP} = 15\times(1-\theta_{Middle-PIP}/90)$ $\alpha_{Ring-MCP} = 15\times(\theta_{Middle-PIP}/90-1)$ $\alpha_{Pinky-MCP} = 15\times(\theta_{Ring-PIP}/90-1)$
Middle	$\alpha_{MCP} = 0^{\circ}$	

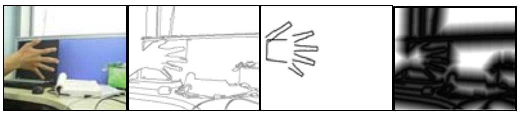
θ represents angle of finger's flexion/extention movements, and α represents angle of finger's abduction/adduction movements. These hand constraints not only can limit the joint angle in a reasonable scope but also can reduce the tracking DOFs, the tracking DOFs can be reduced from 26 to 16.

III. Observation Model

In model-based hand tracking, the similarity between hand state and current observation is measured so as to search the optimal hand state. The observation model has a direct effect on tracking accuracy. The observation is usually built according one or more features. Consider the robustness and simplicity of feature we choose edge and silhouette to build the observation model.

A. Edge-Based Similarity Measurement

Edge is an important feature, it is robust to lighting change. There are many edge detection operators, we choose canndy detection operator. As canndy operator has a smooth process, it is not sensitive to image noise. Besides the edge obtained from canndy operator is relatively complete and continuous. The resulted edge image is displayed in Fig. 2(b). On the other hand, the contour of hand model can be obtained by projecting 3D hand model into 2D image plane, as displayed in Fig. 2(c). After getting edge and contour we measure their similarity by calculating their chamfer distance.



(a) (b) (c) (d)

Figure 2. (a) is the original image. (b) is the resulted edge image. (c) represents the contour of hand model. (d) represents b's corresponding distance map after distance transform.

Assume that $A=\{a_i\}_{i=1}^{Na}$ represents point set of contour, $B=\{b_i\}_{i=1}^{N_b}$ represents point set of edge. Chamfer distance from A to B is defined as [2]:

$$d_{chamfer}(A,B)=\frac{1}{N_a}\sum_{a\in A}\min_{b\in B}\|| a-b \||^2 \tag{1}$$

To effectively calculate chamfer distance between contour point set and edge point set we use two pass algorithm [19] to obtain edge image's distance image. Fig. 2(d) displays the resulted distance image. It is equal to smooth edge image, the more the pixel closes to edge point the smaller the distance is. According their chamfer distance a similarity probability function can be defined as:

$$p_{edge}=\exp(-d_{chamfer}(A,B)) \tag{2}$$

B. Silhouette-Based Similarity Measurement

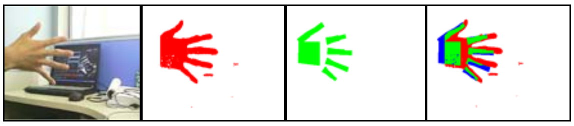
The similarity between two objects can also be measured by comparing their silhouette. Here we compare the binary hand image with model projection to measure the similarity between hand model and hand image. Firstly, we use skin color segmentation rules in [15] to segment out hand region from frame image. The segmentation rules are defined as:

$$\begin{aligned}&R>95\,\&\&G>40\,\&\&\,B>20\,\&\&\\&(\max\{R,G,B\}-\min\{R,G,B\}>15)\,\&\&\\&|R\text{-}G|>15\,\&\&R>G\,\&\&R>B\end{aligned}$$

R, G, B respectively represent red, green and blue component of 24 color image. The merit of this segmentation method is that it can directly operate on the captured color image to realize fast segmentation, Fig. 3(b) displays the segmented binary hand image. Secondly, each part of fingers and palm of hand model are treated as rigid, and their projection can be approximated as a set of quadrilaterals, as displayed in Fig. 3(c). After getting binary hand image and model projection we define a silhouette-based similarity measurement probability function:

$$p_{silhouette}=\exp\{-[(S_{image}\cup S_{model})-(S_{image}\cap S_{model})]\} \tag{3}$$

S_{image}, S_{model} respectively represent binary hand image and model projection, Fig. 3(d) displays their overlap, green region represents $S_{image}\cap S_{model}$, blue region represents $(S_{image}\cup S_{model})-(S_{image}\cap S_{model})$.



(a) (b) (c) (d)

Figure 3. (a) is the original image. (b) represents binary hand image obtained through skin color segmentaion algorithm. (c) is hand model's silhouette. (d) represents overlap of b and c.

Finally, we combine the above two similarity measurement probability functions to build a total observation likelihood model $p_{observation}=p_{edge}p_{silhouette}$, and then use it in PF to calculate the similarity between hand state hand image.

IV. Apply Hand Constraints and Observation Model into Hand Tracking

To continuously evaluate optimal hand state from hand state space according observation, we choose PF as tracking tool. PF, based on Monto Carlo method, randomly samples in state space, and then gives each sample a corresponding weight. The integration operation can be replaced by the mean of these particles so as to realize state update. PF can provide a robust tracking framework [7], it has a good performance on the situation where posterior distribution $p(x\mid \bar{z}_t)$ and observation process $p(x_t\mid z)$ are non-gaussian or non-linear(x represents state vector, $\bar{z}_t=\{z_1,\cdots,z_t\}$ represents historical observations at time t). The propagation of state is assumed as first-order Markove state model $p(x_t\mid x_{t-1})$ that is current state x_t is only determined by last state x_{t-1}, the particles' weights at time t are equal to $p(x_t\mid z_t)\propto p(x_t\mid x_{t-1},z_t)p(x_{t-1}\mid z_{t-1})$, where $p(x_{t-1}\mid z_{t-1})$ represents particles' weights at time t-1, so the state can be recursively updated each time. In practice, we apply above hand constraints into sampling stage, and apply observation model $p_{observation}$ into calculating particles' weights. The tracking algorithm is as follow:

(1) Initialization. At time $t=0$, get initial hand state x_0 according the initial method proposed in [6]. Then the initial particle set $\{(x_0^i,1/N),i=1,\dots,N\}$ is obtained from initial hand state according prior distribution $p(x_0)$, N represents the particle number and x_0^i is the

*i*th initial particle.

(2) Prediction. Transfer particles of last time to get new particle set according state transform equation $p(x_t \mid x_{t-1})$. When time t = 0, the transform equation can be expressed as $x_t^i = x_0^i + rand$; when time t > 0, the transform equation can be expressed as: $x_t^i = \overline{x}_{t-1} + A(x_{t-1}^i - \overline{x}_{t-1}) + Bw_t$ [9], $\overline{x}_{t-1}$ represents the mean of hand state at time t-1, A controls transform speed, Bw_t is stochastic noise. Here we apply dynamic hand constraints into state vector to build relationship bwteen different variables, this only needs to sample for limited variables so as to reduce the sampling freedom of degree. After sampling, the variables of samples are checked that whether or not the variable beyonds the maximize $\theta_{\max}$ or minimize $\theta_{\min}$ defined in static hand constraints, if so revise variable to $\theta_{\max}$ or $\theta_{\min}$.

(3) Weight Caculation. Each particle's weight w_t^i is calculated according the above built observation model $p_{observation}(z_t \mid x_t^i)$, and then obtain a set of weighted particles $\{(x_t^i, w_t^i), i = 1,..., N\}$. Finally, the weight of each particle is normalized through normalization equation $w_t^i = w_t^i / (\sum_{j=1}^{N} w_t^j), i = 1, ..., N$.

(4) Output Mean State. After normalization, the mean hand state at time t is expressed as: $\varepsilon[x_t] = \sum_{i=1}^{n} x_t^i w_t^i$, and then reconstruct 3D hand according $\varepsilon[x_t]$.

(5) Update: t = t+1, go to step(2).

V. Experimental Result

The frame image is captured by an USB camera, and the tracking system is implemented on VC++6.0 platform. We respectively track only index finger moves(#1), fingers move excepting thumb(#2), hand translates in x, y plane(#3), hand rotates along z-axis or x-axis(#4, #5), index finger moves while hand translates in x, y plane(#6), index finger moves while hand rotates along z-axis(#7), fingers move while hand roates along z-axis(#8), the tracking result is displayed as Fig. 4.

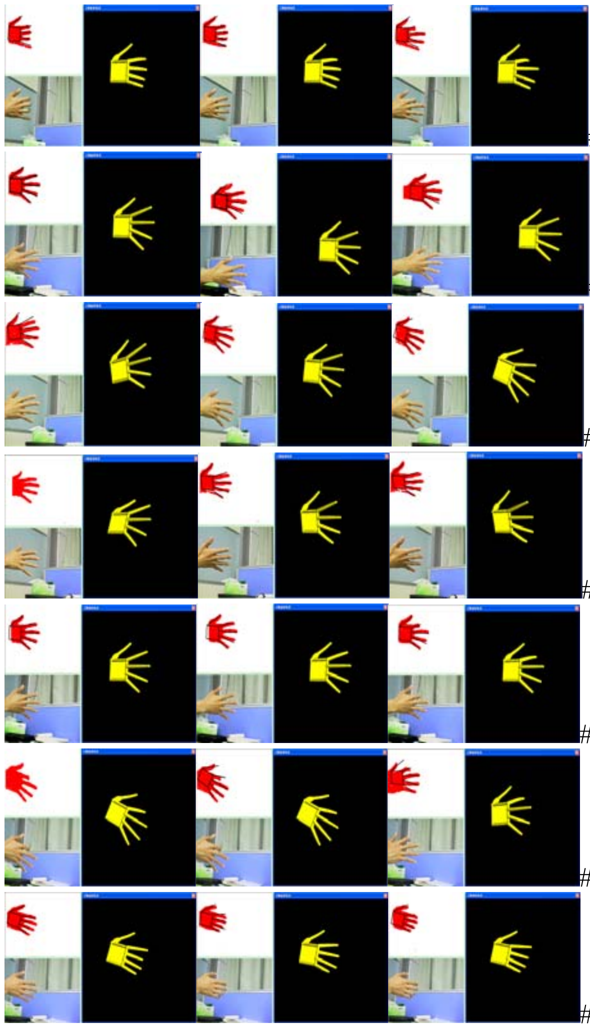


Figure 4. Tracking result.

The bottom left of the image is the observation image, upper left is the skin region segmented from observation image using the method in [15]. As can be seen the segmentation algorithm has a good performance when the light is stable. At the same time, we found the segmentation algorithm is sensitive to red object, it mistakes red pixel for skin color. The right of the image displays the recoverd 3 D hand drawed by openGL. For testing the accuracy of the tracking system, on the one side we track index finger moves and record its angle change tracked by PF; on the other side we record the angle change collected by 5DT data glove with the same action. The compare of tracking result and ground data is showed as Fig. 5.

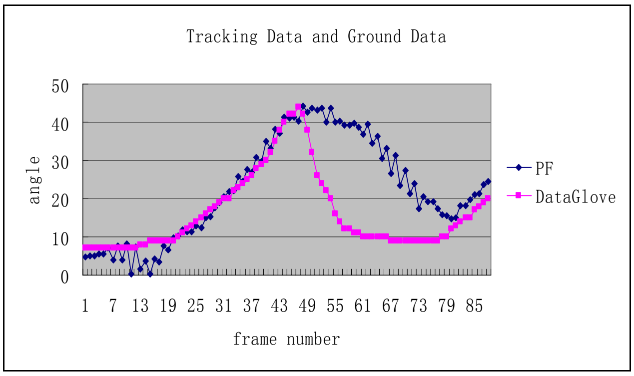


Figure 5. The compare between tracking data and ground data(DIP angle of index finger).

From Fig. 5 we can see that our tracking system can basically track hand movements. As sampling is random, the evaluated optimal hand state may fluctuate around ground data. For testing the accuracy of particle's weight calculated by our observation model, we compare our observation model with Hausdorff distance-based observation model. Hausdorff distance-based observation measures the similarity between hand model and hand image by calculating the Hausdorff distance between two point sets which obtained respectively from hand model and hand image. We choose particles' weights in 50 frames to analyze, each frame uses 10 particles, totally 500 particles. As displayed in Fig. 6 and Fig. 7.

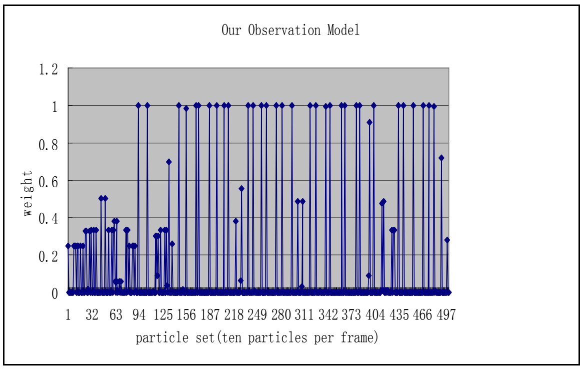


Figure 6. Particle weight calculated by our observation model.

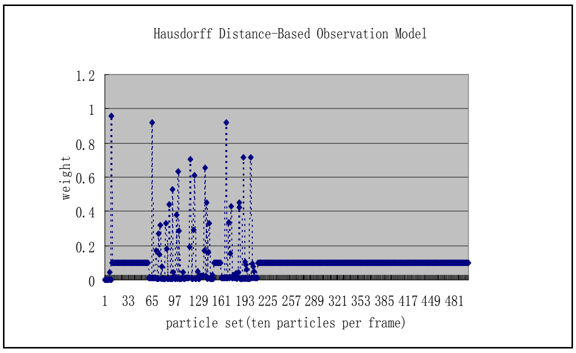


Figure 7. Particle weight calculated by hausdorff distance-based observation model.

From the Fig. 6 and Fig. 7 we can see that the difference of particle weight calculated by our observation model in each frame are greater than by Hausdorff distance-based observation model. The particle weight calculated by Hausdorff distance-based observation model don't have difference after a period of time, all the weights are 0.1, this is because Hausdorff distance is sensitive to image noise, in the presense of image noise Hausdorff-distance between two point set are easily effected by image noise. On the other hand when the difference between two point sets is not obvious, Hausdorff distance cannot distinguish this difference. Our observation model not only can raise the weight of particle that is close to real state, but also can low invalid particle's weight. In the case of without hand constraints the finger may bend toward opposite direction, which violates the fact. This is because in the monocular condition, evaluating hand state according single image refers singularity problem that is an observation may correspond to multiple optimal solutions. Fig. 8 gives out the time cost of our tracking system, the average running time is about 280ms/frame, most of the time is consumed in PF stage, about 230ms/frame, because it needs to recursively calculate weight for each particle which refers complex image processing. The image size in the paper is 400×300. Computer has an Intel(R) Core (TM) Quad CPU 2.66Hz processor and 4G memory.

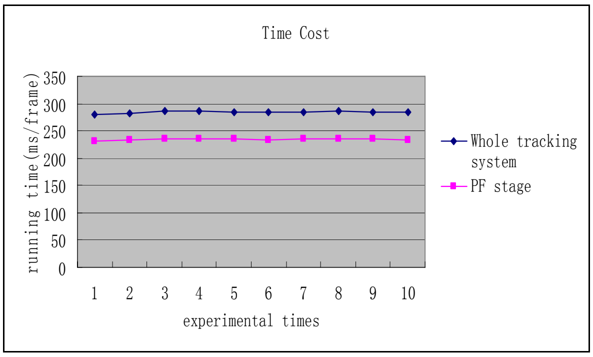


Figure 8. Time cost of our tracking system.

VI. Conclusion

This paper firstly utilizes edge detection algortihm and skin color segmentation algorithm to get edge and binary image from hand image. On the other side get contour and silhouette from hand model by projection transform. Then an observation model based on these obtained features is built to measure the similarity between hand state and observation. We use PF technology to track hand movement and apply the built observation model into calculating particle's weight. This realizes tracking both local and globle hand movements. For testing the accuracy of our observation model, we compare it with Hausdorff distance-based observation model, experimental result shows that the accuracy of our observation model is higher than Hausdorff distance-based observation model. At the same time we introduce hand constraints to hand model to make sure the hand model is a reasonable gesture and reduce tracking degree of freedom. As the remainder DOFs of hand model are still very high(16 DOFs), to evaluate optimal hand state still needs a large number of samples, at the same time our observation model needs complex image processing, the time cost will increase with the increasing of sample number, take into account the real-time we can only achieve tracking with limited DOFs.

ACKNOWLEDGMENT

This paper is supported by National Natural Science Foundation of China (No. 61173079 and No.60973093), Natural Science Foundation for Distinguished Youth Scholar of Shandong Province (No.JQ200820) and Key Project of Natural Science Foundation of Shandong Province (ZR2011FZ003).

REFERENCES

[1] B. Stenger, P. R.S. Mendonc¸a, and R. Cipolla. “Model based 3D tracking of an articulated hand,” Computer Vision and Pattern Recognition, Vol. 2, pp. 310–315, 2001.
[2] B.Stenger. “Model-Based Hand Tracking Using A Hierarchical Filter,”. Cambridge: University of Cambridge, 2004.
[3] V. Athitsos and S. Sclaroff. “Estimating 3D hand pose from a cluttered image,” Computer Vision and Pattern Recognition, Vol. 2, pp. 432–439, June 2003.
[4] Y. Wu, J. Y. Lin, and T. S. Huang. “Capturing natural hand articulation,” Eighth IEEE International Conference on Computer Vision, Vol. 2, pp. 426–432, 2001.
[5] J. Lee and T. L. Kunii. “Model-Based analysis of hand posture,”. IEEE Computer Graphics and Applications, Vol. 15, pp. 77–86, September 1995.
[6] Zhiquan Feng et al. “3D-Freehand-Pose Initialization Based on Operator’s Cognitive Behavior Models,” The Visual Computer, Vol. 26, pp. 607-617, 2010.
[7] J. Deutscher, A. Blake, I. Reid. “Articulated body motion capture by annealed particle ltering,” International Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 126–133, 2000.
[8] J. Lin, Y. Wu, T.S. Huang. “Modeling the constraints of human hand motion,” IEEE Human Motion Workshop, pp. 121–126, 2000.
[9] M. Isard and A. Blake. “Condensation-Conditional density propagation for visual tracking,” International Journal of Computer Vision, Vol. 29, pp. 5-28, 1998.
[10] M Ferre, S Cobos, J Ortego, MA Sanchez-Urán. “D 4.4. 1-First prototype of hand model, sufficient for mimicking realistic hand motions, but not full hand manipulation of objects yet,” 2007.
[11] E. Brown and A. Vardy. “Articulated Human Hand Model with Inter-Joint Dependency Constraints,” Computer Science 6752, Computer Graphics Project Report, 1998.
[12] Henia, O. B., Hariti, M., Bouakaz, S. “A two-step minimization algorithm for model-based hand tracking,” 18th International Conference on Computer Graphics, Visualization and Computer Vision (WSCG), 2010.
[13] Liu Tangli et al. “3D Articulated Hand Tracking by Nonparametric Belief Propagation on Feasible Configuration Space,” Journal of Computer-Aided Design & Computer Graphics, Vol. 20, pp. 476-481, 2008.
[14] FENG Zhi-quan et al. “Research on Hand Gestures Tracking Based on Particle Filtering Aiming at Optimizing Time Cost,” Acta Electronica Sinica, Vol. 37, pp. 1989-1995, 2009.
[15] V. Vezhnevets, V. Sazonov, A. Andreeva. “A survey on pixel-based skin color detection techniques,” Proceedings of the International Conference on Computer Graphics, pp. 85–92, 2003.
[16] FU Yi-li, LIU Cheng. “Hand modeling and motion controlling based on lay figure in virtual assembly,” Computer Integrated Manufacturing Systems, Vol. 15, pp. 681-684, 2009.
[17] R. Rosales, S. Sclaroff, and V. Athitsos. “3D hand pose reconstruction using specialized mappings,” Eighth IEEE International Conference on Computer Vision, Vol. 1, pp. 378–385, 2001.
[18] Y. Wu, T. Huang, “View-independent recognition of hand postures,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 88–94, 2000.
[19] http://www.cs.auckland.ac.nz/~rklette/Books/MK2004/pdf-LectureNotes/08slides.pdf.
[20] H. Zhou and T. S. Huang. “Tracking Articulated Hand Motion with Eigen Dynamics Analysis,” Ninth IEEE International Conference on Computer Vision, Vol. 2, pp. 1102–1109, 2003.
[21] Wen-Yan Chang, Chu-Song Chen and Yi-Ping Hung. “Appearance-Guided Particle Filtering for Articulated Hand Tracking,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.
[22] M. Bray, E. Koller-Meier, N. N. Schraudolph and L. Van Gool. “Smart particle filtering for high-dimensional tracking,” Computer Vision and Image Understanding, Vol. 106, pp. 116–129, 2007.
[23] N. Stefanov, A. Galata and R. Hubbold. “A real-time hand tracker using variable-length Markov models of behaviour,” Computer Vision and Image Understanding, Vol. 108, pp. 98–115, 2007.

Bingchao Liu, born in 1987, enrolled postgraduate. His main research interests in human-computer interaction(HCI).

Zhiquan Feng, born in 1964. Ph.D, Professor, and maste supervisor. His main research interests include image processing, moving human hand tracking, and human-computer interaction.

Aili Sang, born in 1987, enrolled postgraduate. Her main research interests in human-computer interation(HCI).