Research on Grasping Hand Gesture Based on Analysis of Occluded Information

Xianhui Song
School of Information Science and Engineering,
Shandong Provincial Key Laboratory of Network based Intelligent Computing. University of Jinan.
Jinan, 250022, P.R. China.
e-mail: songxianhui145@163.com

Zhiquan Feng *, Bo Yang, Wei Gai, Yan Lin
School of Information Science and Engineering,
Shandong Provincial Key Laboratory of Network based Intelligent Computing, University of Jinan.
Jinan, 250022, P.R. China.
e-mail: fzqwww@263.net

***Abstract*—A novel human hand tracking algorithm using hand occlusion information based on a single-view camera is put forward. Firstly, the moving hand is tracked by the algorithm of edge tracking. Then the hand features of boundary are used to get the maximum inscribed circle of the palm. According to the circle, the hand is divided into two parts: the palm and fingers. Finally, using the proportion of the palm and fingers in the whole hand to guide particle filter tracking. The experiments show that this algorithm can correctly predict the position of the occluded parts, and has better tracking accuracy than some of the previous work.**

***Index Terms*—edge tracing, maximum inscribed circle, palm, self-occlusion, particle filter**

I. INTRODUCTION

The hand is one of the most characteristic organs of the human. Scientists believe that the hand is one of the three most important organs which help the man having more highly wisdom. The human's fingers are very keen, which may feel the vibration amplitude of only 0.00002 millimeters. Also, people are used to use the hand gestures when they are talking with each other, or do gestures to express their feelings entirely. Therefore, The moving human-hand tracking and gesture recognition is one of the active research areas in The Human Computer Intelligent Interaction.

Vision-based gesture interfaces have a great potential in many interactive systems, with which user inputs can be easily achieved. A key issue is how to reconstruct and recognize the natural hand. Multi-purpose imaging system is being widely used now, which means using two or more cameras simultaneously from different positions or angle to observe the operators and analyzing the movements of the users' hand, then use obtained data to reconstruct three-dimensional hand [1-3]. Utsumi et al. [2] used five cameras to track moving hand at the speed of 10 frames per second, and can identify some simple hand gestures. Most systems use dual-camera as a data acquisition equipment, and use traditional vision-based theory to achieve the three dimensional reconstruction [1]. The realization of fine hand tracking should meet three requirements: real-time, robust, continuous tracking and automatic initialization [4]. The problem that multi-purpose imaging system meets is how to find corresponding feature points among different views [5]. In addition, the time cost will be proportional growth with the number of cameras, which makes it difficult to achieve real-time.

Tracking the pose of a moving hand from a monocular perspective is difficult, because self-occlusion is a not solved problem in monocular system yet. The single camera system can not recover 3D point uniquely from 2D pixel. Hand is multi-chain structure and has many degrees, these features make the occlusion of hand parts various and complex, which will lead to the same observation value matching with several different estimates. Generally, one solution is to establish gestures database [6], another way is to use hand static constraint and dynamic constraint [7-8]. S.U. Lee et al. [9] proposed a model which was based on 3D hand, to match with 2D hand edge detection. It can search and track hand effectively, and the cost of calculation was relatively reduced, but it required that the root of thumb, index finger and little finger must be visible. In addition, various hand marks are used widely. In the literature [10], Robert Y. wang et al. used cotton gloves that were printed specific pattern to achieve human-computer interaction. The operators were asked to wear these color gloves, where each color region was given a number.

In summary, the study of hand tracking, especially in monocular system, occlusion can not be avoided, and it is difficult to solve the problem. For example, track the tip point, such as fingertip, when they move rapidly or be occlusion, the tracking may be lost and initialization is required.

This paper proposes a method to track hand movements, which is use the occlusion information to solve the occlusion. The remaining sections of this paper

Manuscript received Mar. 25, 2010; revised Apr. 10, 2011.
project number: 60973093, 60873089.
* Corresponding author.

doi:10.4304/jcp.7.3.768-773

are organized as follows: in section two we give the overall framework of the algorithm. Section three describes the edge tracking. Section four describes the details of the real-time hand segmentation method and palm effective area determining method. Section five gives the tracking method of this paper. Experimental results are presented in section six to validate the performance of the proposed method. Conclusions and future work are given in the last section.

II. THE OVERALL FRAMEWORK

Occlusion can not be avoided in any monocular system. In this paper we use the occlusion information to solve the occlusion. For the gesture, the palm area is generally visible. The ratio of palm area and fingers area is fixed, while hand in natural bent, the ratio will change with the occlusion between fingers and palm [11], which can be used to predict the particles in PF filter.

The proposed method has three stages viz. skin segmentation to detect the hand, edge tracking for feature extraction, and particle filter for hand tracking. The overall diagram of our method is given in Figure 1.

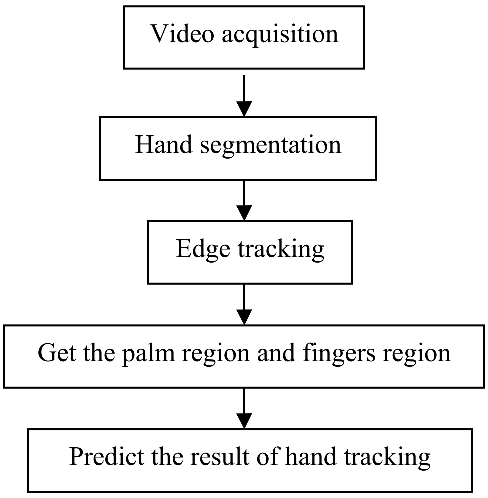


Figure 1. The overall diagram of our method

III. EDGE TRACKING

The boundary is one basic shape features of image. It refers to the set of pixels where the gray value changes with its surrounding pixels. In this paper, we use the clockwise eight connected region contour tracking method [12]. The matrix as follows:

$$F = \begin{pmatrix} 5 & 6 & 7 \\ 4 & p & 0 \\ 3 & 2 & 1 \end{pmatrix}$$

(1)

where p is the target pixel.

The edge tracking includes the following steps:

a) Scanning the image from top to bottom and left to right,when meet the first hand pixel, save it as the starting point, and record it as p_0*;*

b) From p_0 *point, in the direction of clockwise to study the eight points value in* p_0 *neighborhood. When meet the first hand pixel, save it as the starting point, and record it as* p_1*. In this method* p_0 *is named the first neighborhood target point of* p_1*.*

c) Searching from the next point of the first target point, to look for the next target point in the neighborhood;

d) Cycle step c until the first goal of neighborhood of p_n *is* p_0 *and the first goal of neighborhood of* p_{n+1} *is* p_1*, the edge tracking ends.*

IV. GET THE PALM AND FINGERS REGION

The result of hand edge detection is the contour of palm and fingers. If we want to get the palm region from the map, we must distinguish between the palm and fingers. In fact, most hand tracking algorithms do not separate the palm from fingers, while treating them as a whole to study. The size of palm is a key attribute of hand, and it can be used as an identification feature. In this article, we get the location and size of the palm by seeking the maximum inscribed circle of palm, and define the part which is inside circle as palm's coverage, another region that belongs to hand but outside the inscribed circle is named as finger's region [13-16]. As shown in Figure 2:

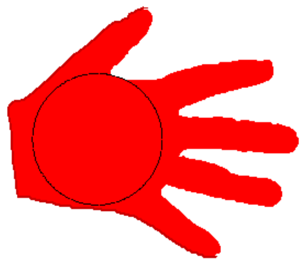


Figure 2. The maximum inscribed circle of the palm

Before seeking the maximum inscribed circle, we need to get the hand direction vector, the center point of the hand, and the wrist center point firstly. Since the previous work has extracted the hand shape completely, the hand-centric is calculated as:

$$\overline{x} = \frac{\sum_i \sum_j i \times f(i,j)}{\sum_i \sum_j f(i,j)} \tag{2}$$

$$\overline{y} = \frac{\sum_i \sum_j j \times f(i,j)}{\sum_i \sum_j f(i,j)} \tag{3}$$

$$f(i,j) = \begin{cases} 1 & if\,(i,j) \in R \\ 0 & other \end{cases} \tag{4}$$

where R is the region inside the maximum inscribed circle.

The center of wrist (x_1, y_1) is obtained by getting the average value of all points from the image which is minus the background and the hand region. Both the wrist central point and hand's central point decide the direction vector of hand (figure 3), which can describe the yaw angle of hand and improve the speed of gestures matching. It is shown in equation 5.

$$X = (x_1 - \overline{x}, y_1 - \overline{y}) \tag{5}$$

Figure 3. The direction vector of hand

Because of the presence of fingers, the hand center point is close to the finger region, which lead to the following question: the inscribed circle is not the biggest one [17], so we cannot regard hand's center as palm's center. Now we propose a modified method: connect the center point of hand with the wrist center point, and seek the maximum inscribed circle of each point on the straight line, then record the point whose radius is the biggest as the center point of the palm (figure 4).

Figure 4. The progress of seeking the maximum inscribed circle of each point on the straight line

V. TRACKING METHOD

According to the physiological characteristics of the hand, hand is a hinge - pole structure which is composed by the joint and section unites. The movement of fingers and palm decide the shape of hand together. Hand movements are mainly to rotate around the joints, so we can use the values of upper joints to describe the rotation angle of lower sections.

The particle filter is a sequential Monte Carlo algorithm, i.e. a sampling method for approximating a distribution that makes use of its temporal structure. A "particle representation" of distributions is used. In particular, we will be concerned with the distribution $P(x_t|z_{0:t})$ where x_t is the unobserved state at time t, and $z_{0:t}$ is the sequence of observations from time 0 to time t [18-19].

When we use particle filter to track hand, first, it is to determine the system state and observation state. In addition, the position, size and direction of hand have been got, so we define the state vector as M (x, y, r, d), where, x, y is the center of the hand; r is the ratio of the palm and fingers; d is the direction vector of hand. Then seek the largest inscribed circle of palm and obtain the size of the palm area A, and the size of fingers area B, regarding the radius R of the maximum inscribed circle as systematic observation value. Finally define the size of the palm area $A = \pi R^2$, finger regional area B = The total area of the hand - A, and $r = B / A$.

The next step is to establish the state transition equation:

$$X_k = \overline{X} + A(X_{k-1} - \overline{X}) + Bw \quad (6)$$

where $\overline{X}$ is the average estimate value in the state of k -1 , X_k is the value of some particle in the state k; Bw is the Gaussian process noise with zero mean, and coefficient A = 1, B = 1. This model is based on the current position and velocity of target particle, and some random disturbance is used to predict the next state motion. Then in order to get the new state of the particle, we need calculate the weight of each particle, such as the equation (7) below:

$$w_{k,i} = e^{-E_{k,i}} \quad (7)$$

where $E_{k,i}$ is the hausdorff distance, which indicates the difference between intending state and actual state. Next step is to normalize all the particle weights and calculate the weight sum of N particles. According to the static constraint and dynamic constraint, we can speculate the occluded part of the hand from the visible part.

VI. EXPRIMENTS

The proposed hand tracking method has been implemented using Visual C++ and OpenGL, and several experiments have been carried out to verify this method. The image size in our testing is 300*400.

In our virtual environment, the operator is asked to complete the task of fixing the screw and nut which are on the desk. The specific requirements are as follows: the movement of the operator's hand must be natural. When the screw and nut are fixed, the tracking ends.

In the progress of hand segmentation due to using the minus background method alone the region of arm is extracted, and segmentation using skin color only will include the objects whose color is similar to the hand region. Therefore, in this paper we segment the image based on the two methods together, the results are shown in figure 5. The user is first required to specify a region on the user's hand to obtain the training data in the HSL color space. These results show that the hand can be accurately segmented in real-time under large rotation and scale changes. The above segmentation method can be used to track the hand in real-time.

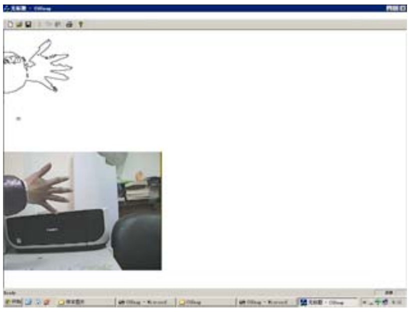
The above segmentation method can be used to get the hand in real-time, and we can assume continuity of the position of the hand during tracking. Figure 6 shows some hand tracking results in a live video. We have tested the accuracy and speed of tracking the hand among three different algorithms: our method (MICT), PF filter, and Simulate Anneal Arithmetic, which is shown in Figure 7, 8. The average accuracy for tracking the hand is 0.677, which is better than the other method: the PF filter (PF) whose accuracy is 0.599 and the Simulate Anneal Arithmetic (GPAPF) whose accuracy is 0.551. In addition, in this method we can get good result by using fewer particles, so the average elapsed time for tracking the hand is less than the other method, which can meet the real time performance for real applications.

VII. CONCLUSION

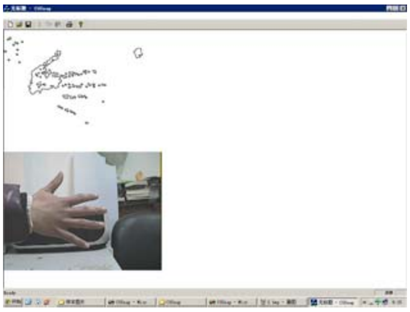
In this paper, an effective hand tracking method is proposed based on hand occlusion information. This method is simple to operate by moving the user's hand freely for interactive operations. Some experiments have been conducted to validate that the proposed method can correctly predict the position of the occluded parts, and have better tracking accuracy. The future research will address to get the information of each finger and hand gesture spotting for sign recognition via single camera system.

ACKNOWLEDGMENT

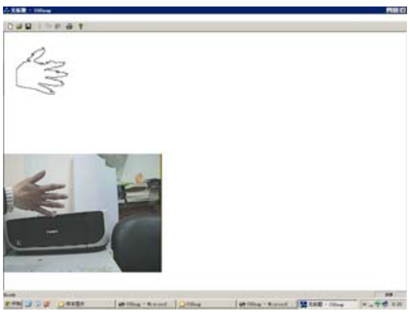
This paper is supported by National Natural Science Foundation of China (No.60973093, No. 60873089), Natural Science Foundation for Distinguished Youth Scholar of Shandong Province (No.JQ200820), Key Project of Natural Science Foundation of Shandong Province (2006G03), Science and Technology Plan of Shandong Province Education Department (J07YJ18).



(a)

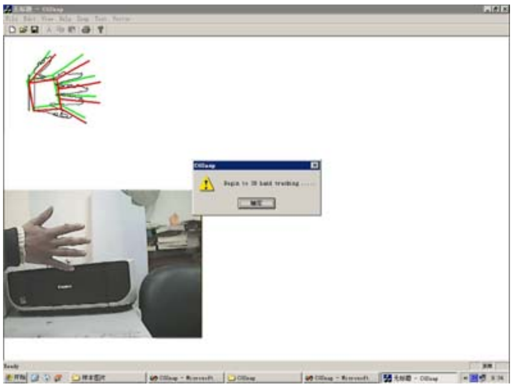


(b)

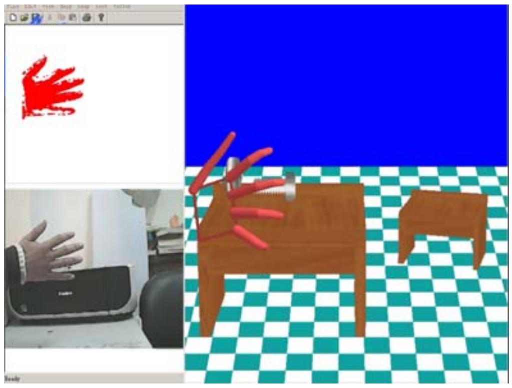


(c)

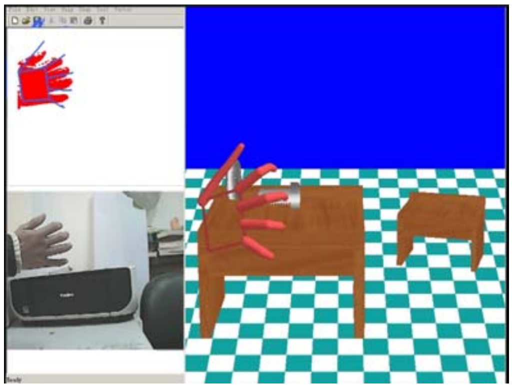
Figure 5. (a). The result of minus background. (b). The result of skin color. (c). The result of method in this paper



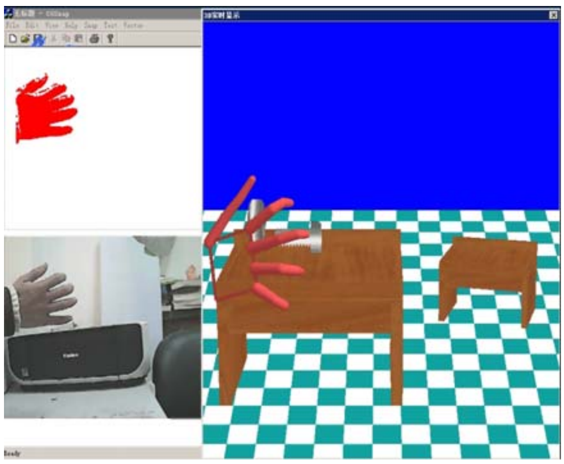
(a) Begin to 3D hand tracking



(b)



(c)



(d)

Figure 6. Examples of our tracking method

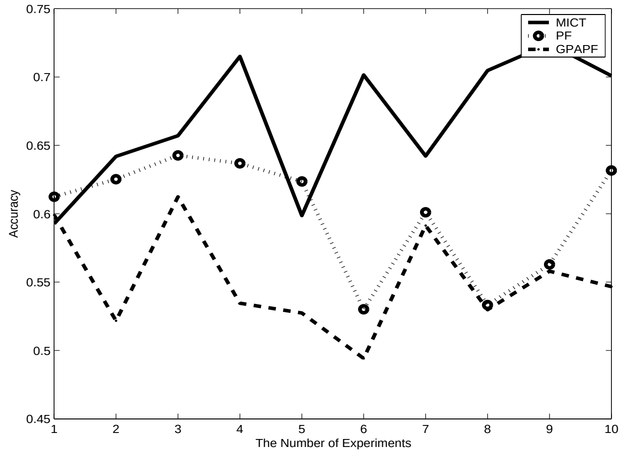


Figure 7. the accuracy of three algorithms

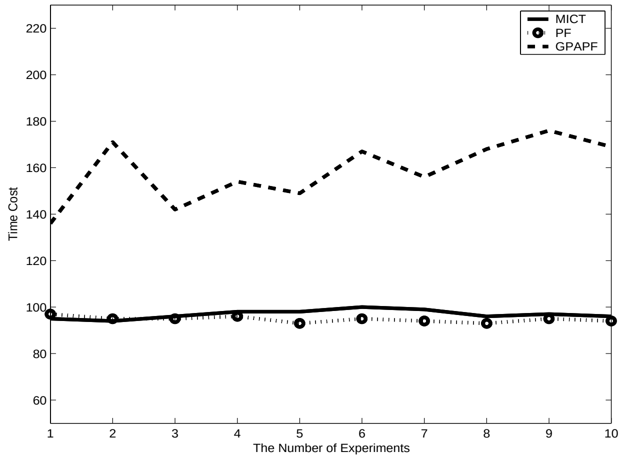


Figure 8. the average elapsed time of three algorithms

REFERENCES

[1] Delamarre Q.and Faugeras O.. Finding pose of hand in video images: A stereo-based approach [J]. In: Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998, pp. 585-590.

[2] Utsumi A.,and Ohya J.. Multiple-hand-gesture tracking using multiple cameras [J]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition'99, Piscataway, New Jersey, 1999, pp. 1473-1478.

[3] Isidoro J.and Sclaroff S.. Stochastic mesh-based multiview reconstruction [J]. In: Proceedings of the 1st International Symposium on 3D Data Processing Visualization and Transmission (3DPVT 2002), Padova, Italy, 2002, pp. 568-577.

[4] Li Weiqun. A general survey and research of hand tracking method [J]. Computer and Digital Engineerin. 2010, 38(9): pp. 79-182.

[5] FENG Zhi-Quan,MENG Xiang-Xu1, LI Xue-Qing, and WANG Ru. Research on 3D reconstruction for visible hand based on single View [J]. Chinese journal of computer. 2006, 29(3): pp. 448-456.

[6] Nobutaka Shimada, Kousuke Kimura, Yoshiaki Shirai, and Yoshinori Kuno. Hand posture estimation by combining 2-D appearance-based and 3-D model-based approaches [J]. Proceedings of 15th International Conference on Pattern Recognition,Barcelona,Spain,2000,.3: pp. 709-712.

[7] Tang Haoxuan, Hong Bingrong, and Liu We. Motion simulation of hand and knowledge-based control metho [J]. High Technology Letters, 2001, (10): pp. 61-65.

[8] FU Yi-li and LIU Cheng. Hand modeling and motion controlling based on lay figure in virtual assembly [J]. Computer Integrated Manufacturing Systems, 2009, 15(4): pp. 681-684.

[9] S.U. Lee and I. Cohen. 3-D Hand reconstruction from monocular view [J]. International Conference on Pattern Recognition. Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04). Cambridge, UK, 2004, 3: pp. 310-313.

[10] Robert Y. Wang and Jovan Popović. Real-time hand-tracking with a color glove [J]. ACM Transactions on Graphics (SIGGRAPH 2009), 2009, 28(3): pp. 1-8.

[11] BurneHogarth. Drawing dynamic hands [M]. Guang Xi Art publishing company. 1997.

[12] TANGSh-wei, FENGYong-qiang, and CHENJing. Reseamh of image pre-proeessing of well-Ioggin gbased on wavelet transform [J]. Joumal of Qiqihsr University. 2006, 22(5): pp. 54-60.

[13] LI Wen Xin, XIA Sheng Xiong, ZHANG Da Peng ,and XU Zhuo Qun. A new palmprint identification method using bi-directional matching based on major line [J]. Journal of Computer Research and Development. 2004, 41(6): pp. 996-1002.

[14] YUE Feng, ZUO Wang-Meng, and ZHANG Da-Peng. Survey of palm print recognition algorithms [J]. Acta Automatica Sinica. 2010, 36(3): pp. 353-365.

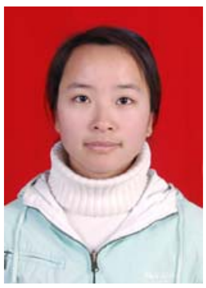
[15] HUO Sheng-jun, FAN Song-qing, and ZHAO Chen-yin, Measurement of fingers' width and length of every segment in human [J]. Computer Integrated Manufacturing Systems,2003, 9(4): pp. 326-328.

[16] LING Jing and YAO Ming-hai. The palm image capture & positioning RIO based on the maximum inscribed circle [J], Journal of Jinhua College of Vocation and Technology, 2010, 10(3): pp. 73-76.

[17] Dai Lei. Non-constraint palm image acquisition system and corresponding feature locating algorithm [J]. Journal of Data Acquisition and Processing. 2010, 25(2): pp. 183-187.

[18] LU Yi-jun, CHEN Yi-min, HUANG Shi-hua, CHEN Ming, and YAO Zheng-wei. Hand tracking based on particle filter and improved level set [J]. Computer Engineering. 2010, 36(13): pp. 159-161.

[19] HU Shi-qiang and JING Zhong-liang, Overview of particle filter algorithm [J], Control and Decision. 2005, 20(4): pp. 361-371.



Xianhui Song, born in1986, is a postgraduate of University of Jinan, Jinan. Her main research interests include image processing, moving human hand tracking, and human-computer interaction.

Zhiquan Feng is a professor of School of Information Science and Engineering, Jinan University. He got the Master degree from northwestern polytechnical university, china in 1995, and PhD degree from Computer Science and Engineering Department, shandong university in 2006. He has published more than 50 papers on international journals, national journals, and conferences in recent years. His research interests include: human hand tracking/recognition/interaction, virtual reality, human-computer interaction and image processing.

Bo Yang is a professor and vice-president of University of Jinan, Jinan, China. He is the Director of the Provincial Key Laboratory for Network-based Intelligent Computing and also acts as the Associate Director of Shandong Computer Federation, and Member of the Technical Committee of Intelligent Control of Chinese Association of Automation. His main research interests include computer networks, artificial intelligence, machine learning, knowledge discovery, and data mining. He has published numerous papers and gotten some of important scientific awards in this area.

Wei Gai, born in 1987, Shandong Province, China. She is an enrolled postgraduate of university of Jinan. And research interests on human-computer interaction and moving human tracking.

Yan Lin, born in 1989, enrolled postgraduate. Her main research interests in moving hand tracking and hunman-computer interaction.