2D Motion Detection Bounded Hand 3D Trajectory Tracking and Gesture Recognition under Complex Background

Shuangqing Wu, Yin Zhang, Sanyuan Zhang, Xiuzi Ye, Yiyu Cai, Jianmin Zheng, Soumita Ghosh, Wenyu Chen, Jane Zhang

Abstract

In this paper, a 2D motion detection bounded hand 3D trajectory tracking and gesture recognition system is proposed for virtual reality interactions. First, the Bayes decision rule for classification of background and foreground is utilized to automatically locate the hand that bounded within a rectangle, and then the trajectory of the hand in 3D space is tracked by mean shift particle filter and stereo imaging. The skin color feature is exploited for image matting that effectively segment the hand contour in video sequence automatically. Finally the hand gesture is recognized by the connected component analysis and line approximation.

The proposed technique works without any markers or constraints, overcomes the disturbance of arms and faces in the scene, and can recognize multiple hands with different gestures under complex background.

1 Introduction

Hand motion tracking in 3D spaces and gesture recognition is an important research issue with many practical applications in the fields of virtual reality, human-computer interaction, sign language recognition, and visually assisted medical surgery and so on. The vision-based methods[Lee and Kim 1999; Ng and Ranganath 2002; Psarrou et al. 2002; Shan et al. 2004; Moeslund et al. 2006; Elmezain et al. 2008; Manders et al. 2008; Pan et al. 2010] provide a more suitable and natural human-machine interaction than the contact-based method using the mouse, keyboard, joysticks, wii remotes or sensor glove [Bowman and Hodges 1997; Pierce et al. 1999;

Jia et al. 2007]. The hand is commonly represented by its geometric features such as contours [Chang et al. 2005], fingers[Oka et al. 2002], and its distinctive features of color and texture [Soriano et al. 2000; Perez et al. 2002; Yang et al. 2005; Yuan et al. 2008]. However, every single type of features has its limitations. For example, hand contours are view-dependent and vary dramatically in natural hand motion, and skin color is not reliable due to varying illumination.

Our research objective is to design and implement a vision-based system that effectively tracks the position of the hand in 3D space and recognizes its gesture by using the clues of motion, skin color, depth and geometry shape tightly. We make three main contributions: First, the Bayes decision rule for classification of background and foreground is utilized to automatically locate the hand that bounded within a rectangle, and then the trajectory of the hand in 3D space is tracked by mean shift particle filter and stereo imaging. Second, the skin color feature is exploited for image matting that effectively segment the hand contour in video sequence automatically, the matting is carried out in the bounding box that contained the hand, so the disturbance of similar skin color regions such as arms and face is decreased. Third, and finally the hand gesture is recognized by the connected component analysis and line approximation.

The remainder of this paper is organized as follows. Section 2 describes the detection of the hand with a bounding box by the Bayes decision rule and the mean shift particle filter. The tracking of the hand motion trajectory in 3D space is demonstrated in Section 3. Hand gesture recognition based on skin color, image matting and geometry shape is then presented in Section 4. Experimental results are discussed in Section 5. Finally section 6 concludes the paper.

2 Bounding the hand in the 2D image space

2.1 Automatic hand location

In this paper, the foreground object detection is utilized to initialize the location of the hand in 2D image space. Several methods have been proposed that can work for the situations that contain a variety of background variations. We adopted the Bayes decision rule for classification of background and foreground from selected feature vectors [Li et al. 2003].

Let vt = [a0...an]T be a discrete feature vector value extracted from an image sequence at the pixel u = (x, y) at time t. By Bayes rule, the posterior probability of vt from the background b or foreground f is

$$\begin{array}{l} P(C|v_t, u) = \frac{P(v_t|C,u)P(C|u)}{P(v_t|u)}, C = borf \\ P(v_t|u) = P(v_t|b,u) \cdot P(b|u) + P(v_t|f,u) \cdot P(f|u) \end{array} \qquad \textbf{(1)}$$

[Figure 1]

Using the Bayes decision rule, the pixel is classified as background if the feature vector satisfies

$2P(v_t /b, u) \cdot P(b /u) > P(v_t /u)$ (2)

For each type of feature vectors, a table of feature statistics Su,t,i,vt,

i = 1, ..., N2(N2 > N1)is maintained,N1is the number of feature vectors selected that cover a large percentage of total feature vectors,N2is the number of the most significant feature vectors

$$S_{v_t}^{u,t,i} = \begin{cases} p_u^{u,t,i} = P(v_t^i|u) \\ p_{u,b}^{u,t,i} = P(v_t^i|b,u) \\ v_t^i = [a_1^i, ..., a_n^i]^T \end{cases} \qquad (3)$$

and the conditional probabilities are obtained as

$$\begin{array}{l} P(b|u) = p_b^{u,t} \\ P(v_t|u) = \sum_{j \in M(v_t)} p_u^{u,t,j} \\ P(v_t|b,u) = \sum_{j \in M(v_t)} p_{u,b}^{u,t,j} \end{array} \qquad (4)$$

where the matched feature set is defined as

$$M(v_t) = \{k : \forall m \in \{1, ..., n\}, \left|a_m - a_m^k\right| \leq \delta\} \qquad (5)$$

In our experiment, we found that the hand detected by the foreground object detection based on Bayes decision rule is better than the result get by mixture of Gaussians method[Stauffer and

Grim-son 1999], because mixture of Gaussians tend to misclassify the foreground points when there are moving background objects.

By the foreground object detection, the motion object in the scene can be detected. So when a user wishes his or her hand to be tracked by the system, he or she just needs to held up the hand and make some movement in front of the camera, in this way multiple hands can be detected to track their motion trajectories.

2.2 Particle filter

The human hand is a highly deformable articulated object with many degrees of freedom. Due to the presence of background clutter, complex dynamics of hand motion, and varying illumination, hand tracking is typically a non-linear and non-Gaussian problem. Hence the particle filter instead of the Kalman filter is employed for hand tracking in 2D image sequences.

Particle filter (PF)[Arulampalam et al. 2002; Bray et al. 2007] is a technique for implementing a recursive Bayesian filter by Monte Carlo simulations. Managing multi-modal density allows PF to handle clutter and recover from failures in visual tracking. By incorporating the mean shift (MS)[Comaniciu et al. 2000; Yang et al. 2005] optimization into particle filter to move particles to local peaks in the likelihood, the mean shift embedded particle filter(MSPF) improves the sampling efficiency considerably, so we follow the approach of [Shan et al. 2007] to track the hand. However, in our method the stereo vision is combined for hand 3D trajectory tracking instead of only 2D image sequences tracking.

In 2D image sequence, the hand position is represented by the rectangle that bounding it. The rectangle state st is defined as (u, v, w, h), that indicates its position and size, and the observation zt is modeled as a first-order Markov process, the probability density function(pdf) p(st|z1:t) can be obtained by prediction and updated as

$$\begin{aligned} p(\mathbf{s}_t|\mathbf{z}_{1:t}) &= \kappa p(\mathbf{z}_t|\mathbf{s}_t)p(\mathbf{s}_t|\mathbf{z}_{1:t-1}) \\ p(\mathbf{s}_t|\mathbf{z}_{1:t-1}) &= \int p(\mathbf{s}_t|\mathbf{s}_{t-1})p(\mathbf{s}_{t-1}|\mathbf{z}_{1:t-1})dx_{t-1} \end{aligned} \tag{6}$$

whereis a normalizing constant, p(st|st−1) is conditional probability distribution of dynamic model, and p(st|z1:t−1) is the priori probability distribution. In MSPF, after the particles propagated by the dynamic model p(st|st−1), the MS optimization is run for each of the particles. Particles are then moved in the gradient ascent direction in the likelihood until they converge to their neighboring local peaks.

2.3 Mean shift optimization

Let {x∗ i }i=1...n be the pixel locations of the target model, and the function b: R2 → {1...m} associates the color index of the pixel x∗ i in the histogram bin b (x∗ i). The probability of the color u in the target model is derived as

$$\mathbf{q}_u = C \sum_{i=1}^{n} k\left(\|x_i^*\|^2\right) \delta\left[b\left(x_i^*\right) - u\right]$$
$$C = \frac{1}{\sum_{i=1}^{n} k\left(\|x_i^*\|^2\right)} \qquad (7)$$

where k is a kernel profile function, is the Kronecker delta function, C is the normalization constant.

Using the same kernel profile k with radius h, the probability of the color u in the target candidate is given by

$$\mathbf{p}_u(\mathbf{y}) = C_t \sum_{i=1}^{\kappa} k\left(\left\|\frac{\mathbf{y}-x_i}{h}\right\|^2\right) \delta\left[b(x_i) - u\right]$$
$$C_t = \frac{1}{\sum_{i=1}^{\kappa} k\left(\left\|\frac{\mathbf{y}-x_i}{h}\right\|^2\right)} \qquad (8)$$

The similarity between the target model and the target candidate is measured by the Bhattacharyya coefficient, and can be approximated using Taylor expansion

$$\rho\left[\mathbf{p}(y), \mathbf{q}\right] \approx \frac{1}{2}\sum_{u=1}^{m} \sqrt{\mathbf{p}_u\left(y_0\right)\mathbf{q}_u} + \frac{C_t}{2}\sum_{u=1}^{\kappa} w_i k\left(\left\|\frac{\mathbf{y}-x_i}{h}\right\|^2\right) \qquad (9)$$

where

$$w_i = \sum_{u=1}^{m} \delta\left[b(x_i) - u\right] \sqrt{\frac{\mathbf{q}_u}{\mathbf{p}_u\left(y_0\right)}} \qquad (10)$$

The second term in equation 9 represents the density estimate computed with kernel profile k, and the maximization can be efficiently achieved based on the mean shift iterations.

2.4 Importance sampling

The required posterior density is approximated by a weighted particle set {s(n) t , (n) t }Nn=1 at each time t. Each particle s(n) t represents one hypothetical state of the object, and is weighted by a discrete sampling probability (n) t ,The particles after running the MS optimization can be regarded as sampling from an importance function gt(st), and the weights of the particles is as follows

$$\begin{aligned} \pi_t^{(n)} &= \frac{f_t(s_t^{(n)})}{g_t(s_t^{(n)})} p(\mathbf{z}_t | \mathbf{x}_t = s_t^{(n)}) \\ f_t(s_t^{(n)}) &= p(\mathbf{x}_t = s_t^{(n)} | \mathbf{z}_{1:t-1}) \end{aligned} \qquad (11)$$

In our hand tracking process the hand is detected and tracked without taking into account any shape information, so the hand can be moved with much degree of freedom.

3 Hand 3D trajectory tracking

The bounding box that contained the hand is generated in each frame of image sequence, for the hand is represented by the rectangle state st = (u, v, w, h) that indicates its position and size during the tracking. Then the skin detection is carried out in HSV color space in this bounding box. A pixel is classified as a skin if its value is within a certain range of HSV color space, of which the thresholds are experimentally established.

We retrieve the (x, y, z) coordinate value of all the image pixels in the bounding box that have skin color, then calculate the average or mean coordinate to make it represent the 3D position of the hand, so the hand tracking is carried out in the 3D space. The movement trajectories can be applied for 3D virtual space interaction. Compared with the 3D model-based tracking which usually suffers from high computational cost, our 3D space hand tracking is much simpler and the hand movement has much degree of freedom.

The 3D point clouds are measured by the stereo camera. For the limitations of the stereo camera, not all the skin pixels in the bounding box can find its 3D coordinate value, the invalid is set to be (0, 0, 0) by the stereo camera, so those points should be excluded. What’s more, some pixels point corresponds to background may also have the similar skin color, so a disparity range is set to exclude those points for the calculation of the 3D position of the hand.

4 Hand gesture recognition

To exclude the influence of the complex background or the facial region whose color maybe similar to the skin, our hand gesture recognition is also carried out in the bounding box that generated during the tracking process in section 2.

4.1 Automatic Laplacian matting

The hand is represented by its distinctive geometric features of contours and fingers. It is possible to recognize the gesture of the hand after the hand contour is extracted. Formally, image matting methods assume the input an image I to be a composite of a foreground image f and a background image b. The color of the i-th pixel is assumed to be a linear combination of the corresponding foreground and background colors.

$$I_i = \alpha_i f_i + (1 - \alpha_i) b_i \qquad (12)$$

where ai is the pixel’s foreground opacity, or alpha. This is a severely under constrained problem, and most recent methods is interactive based or expect the user to provide a trimap[Chuang et al. 2002; Apostoloff and Fitzgibbon 2004; Sun et al. 2004] as a starting point, labeling some pixels as foreground, background, or unknown.

A cost function can be derived from the local smoothness assumptions on foreground and background colors, and in the resulting expression it is possible to analytically eliminate the foreground and background colors to obtain a quadratic cost function in alpha, in this way high quality mattes can be obtained by very few scribbles indicating background pixels or foreground pixels. For the color images with the foreground and background colors in a window satisfy the color line model, the cost function is as

$$J(\alpha) = \alpha^T L \alpha \qquad (13)$$

L is an N xN matrix and referred as the matting Laplacian, whose (i, j)-th element is

$$\sum_{k|(i,j)\in w_k} \left(\delta_{ij} - \frac{1}{|w_k|}(1 + (I_i - \mu_k)(\sum\nolimits_k + \frac{\varepsilon}{|w_k|} I_3)^{-1}) \right) \qquad (14)$$

where Ek is a covariance matrix, µk is a mean vector of the colors in the window wk,and I3 is the identity matrix.

The skin color image pixels in the bounding box are labeled as the foreground pixels, while the pixels on the four edges with a certain width in the bounded image that do not have the skin colors can be labeled as the background pixels. In the experiment, considering the matting efficiency, only the skin color pixels in a small region with a certain width and height are labeled as the foreground pixels. The central of this small region can be set to be the mean of all the pixels in this bounding box that have skin color. [Figure 2]

4.2 Geometry analysis

We solve the segmentation problem by performing connected component analysis, and selecting the largest remaining connected component to be the input for the gesture recognition. Then an appropriate contour approximation is applied by compress horizontal, vertical, and diagonal segments and leaves only their end points. The approximated contours now can be used to find fingers. For this a line approximation technique was devised. It has the following steps:

Step1: If the contour has three points (the largest contour left is unlikely to have less than three points), find the smallest angle. If the smallest angle is below the angle threshold value, calculate the length of the line from the point of the smallest angle to the midpoint of its opposite side. If the length is greater than the threshold, add the line to the list of detected lines;

Step2: If a contour has more than three points. For each point pn do the following: (a) Calculate angle between the lines formed by the points pn−1 and pn, pn and pn+1; (b) Calculate the length of the line from pn to the mid-point of the line joining pn−1 and pn+1; (c) If the angle in step a, is less than the threshold angle and the length of the line in step b is greater than the threshold length then add the line found in step b, to the list of detected lines and set the next point to be processed as pn+2, else set the next point for processing as pn+1.

The detected line numbers then can be employed to recognize the hand gestures. If one hand is in the scene, then 0-5 gestures can be recognized, while there are two hands, maximum 6 × 6 combined gestures can be used to send different commands to the virtual reality interaction system.

The framework of our hand motion tracking and gesture recognition system is shown in Figure 1. The automatic hand localization, hand 3D motion trajectories tracking, hand contour extraction, and gesture recognition are the key components of the system. The hand is continuously tracked in image space while the calculation of the 3D position of the hand and the hand gesture recognition steps are performed once every few frames so as to speed up the system performance.

5 Experimental results

We performed hand tracking and gesture recognition experiments on more than 20 video sequences of hand movement with gesture transformation. The input image sequence is captured by the stereo camera Bumblebee 2. Each sequence contains about 600 frames, and the image resolution is 320 × 240 pixels. No markers or any special constraints about illumination and background were set in capturing. The proposed algorithms were implemented on a computer (Intel(R) Core(TM) 2, 1.86GHz, and 2.0GB of Ram) with MS Visual C++.

Figure 2 presents tracking and recognition results from one test image sequence (frame 219) with the arms and face appeared in the scene. The green bounding box indicates the tracked hand, and Figure 2(a) is the detected skin regions with the blue color, (b) is the automatically generated scribble image, (c) is the image matting, (d) is the binary image of the matting, (e) is the approximated hand contour,and (f) is the recognition result. The detected skin region in this frame image is shown in Figure 2(g), the detection and recognition of the hand that based on only the skin color would face difficulty for the arms and face have similar colors, and the detected skin region on the hand is too sparse and scattered to extracted its connected component contour. However, in our method, this problem can be settled to some extent for the hand is encircled in the bounding box by the motion tracking process, and the skin color is only used to assist the automatic matting. [Figure 3] [Figure 4]

Figure 3 gives another experiment result of tracking and recognition from one test image sequence (frame 219) with different background and with different gesture pose. Figure 3(a) is the detected skin regions with blue color, (b) is the automatically generated scribble image, (c) is the image matting, and (d) is the recognition result. The regions that marked by red color in Figure 3(e) indicate the pixels with skin color that have valid (x, y, z) coordinate value calculated by the stereo camera.

Figure 4 is one recognition result when some background pixels in the bounding box that also include some skin color region. Figure 4(a) is the detected hand, (b) is the detected skin regions with the blue color, (c) is the automatically generated scribble image, (d) is the image matting, and (e) is the recognition result. Figure 5 is the example of the tracking and recognition result from one test image sequence (frame 58) under light illumination that is different with the frame image shown in Figure 4. Figure 5(a) is the detected hand, (b) is the detected skin regions with blue color, (c) is the automatically generated scribble image, (d) is the image matting, (e) is the approximated hand contour, and (f) is the recognition result.

Figure 6 is one example of the tracking and recognition result from one test image sequence (frame 80) with multi-hands. Figure 6(a) are the detected hands, (b) is the image matting, and (e) is the recognition result. The result shows that the system is capable of robustly tracking multiple hands with different gestures under complex background. [Table 1]

Table 1 show the running time of the each module in our hand 3D trajectory tracking and gesture recognition system for one test image sequence, the image sequence capture by the stereo camera contains 600 frames, the stereo imaging calculation and the recognition is carried out every 5 frames, so its running on total 120 frames. As we can see from the table that recognition process is somehow time consuming, for the image matting rely on solving large sparse matrices, the algorithm about fast matting using large kernel matting Laplacian matrices would be applied to speed up the system performance. And we should point out that the average running time for each frame also affects the performance of the tracking the hand with the bounding box in 2D image sequence, and finally would influences the recognition. [Figure 5] [Figure 6]

6 Conclusions

Vision-based hand tracking and gesture recognition is a task of great importance for intelligent human-computer interaction. In this paper, an integrated framework that contains automatic hand localization, hand 3D motion trajectories tracking, hand contour extraction, and gesture recognition is proposed. The capability of the system to handle complex background, illumination variation, and multiple hands without any markers or constraints is demonstrated by experiment results. Further investigation will be carried out to utilize the disparity data to enhance the recognition when some skin like background pixels in the bounding box that just connected with the hand part. Exploit a 3D virtual reality interaction system based on our hand tracking and gesture recognition will also be our future work.

Acknowledgements

This work is supported by the National Basic Research (973) Program of China (No. 2009CB320804), National Hi-Tech Research and Development Program (863) of China (No. 2007AA04Z1A5), Zhejiang Provincial Natural Science Foundation of China(Y1090597), and ARC 9/09 Grant (MOE2008-T2-1-075) of Singapore.

References

APOSTOLOFF, N., AND FITZGIBBON, A. 2004. Bayesian video matting using learnt image priors. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1, I407 – I414.

ARULAMPALAM, M. S., MASKELL, S., GORDON, N., AND CLAPP, T. 2002. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. IEEE Transactions on Signal Processing 50, 2, 174 – 188.

BOWMAN, D. A., AND HODGES, L. F. 1997. Evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. Proceedings of the Symposium on Interactive 3D Graphics, 35 – 38.

BRAY, M., KOLLER-MEIER, E., AND VAN GOOL, L. 2007. Smart particle filtering for high-dimensional tracking. Computer Vision and Image Understanding 106, 1, 116 – 129.

CHANG, W.-Y., CHEN, C.-S., AND HUNG, Y.-P. 2005. Appearance-guided particle filtering for articulated hand tracking. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1, 235 – 242.

CHUANG, Y.-Y., AGARWALA, A., CURLESS, B., SALESIN, D. H., AND SZELISKI, R. 2002. Video matting of complex scenes. ACM Transactions on Graphics 21, 3, 243 – 248.

COMANICIU, D., RAMESH, V., AND MEER, P. 2000. Real-time tracking of non-rigid objects using mean shift. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2, 142 – 149.

ELMEZAIN, M., AL-HAMADI, A., APPENRODT, J., AND MICHAELIS, B. 2008. A hidden markov model-based continuous gesture recognition system for hand motion trajectory. Proceedings of International Conference on Pattern Recognition.

JIA, Y., LI, S., AND LIU, Y. 2007. Tracking pointing gesture in 3d space for wearable visual interfaces. Proceedings of ACM International Multimedia Conference and Exhibition, 23 – 29.

LEE, H.-K., AND KIM, J. H. 1999. Hmm-based threshold model approach for gesture recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 21, 10, 961 – 973.

LI, L., HUANG, W., GU, I. Y., AND TIAN, Q. 2003. Foreground object detection from videos containing complex background. Proceedings of ACM International Multimedia Conference and Exhibition, 2 – 10. technique. Proceedings of ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry.

MANDERS, C., FARBIZ, F., YIN, T. K., MIAOLONG, Y., CHONG, B., AND GUAN, C. G. 2008. Interacting with 3d objects in a virtual environment using an intuitive gesture system. Proceedings ofACMSIGGRAPHInternational Conference on Virtual-Reality Continuum and Its Applications in Industry.

MOESLUND, T. B., HILTON, A., AND KRUGER, V. 2006. A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding 104, 2-3 SPEC. ISS., 90 – 126.

NG, C. W., AND RANGANATH, S. 2002. Real-time gesture recognition system and application. Image and Vision Computing 20, 13-14, 993 – 1007.

OKA, K., SATO, Y., AND KOIKE, H. 2002. Real-time fingertip tracking and gesture recognition. IEEE Computer Graphics and Applications 22, 6, 64 – 71.

PAN, Z., LI, Y., ZHANG, M., SUN, C., GUO, K., TANG, X., AND ZHOU, S. Z. 2010. A real-time multi-cue hand tracking algorithm based on computer vision. Proceedings - IEEE Virtual Reality, 219 – 222.

PEREZ, P., HUE, C., VERMAAK, J., AND GANGNET, M. 2002. Color-based probabilistic tracking. Proceedings of European Conference on Computer Vision, 661 – 675.

PIERCE, J. S., STEARNS, B. C., AND PAUSCH, R. 1999. Voodoo dolls: Seamless interaction at multiple scales in virtual environments. Proceedings of Symposium on Interactive 3D Graphics, 141 – 145.

PSARROU, A., GONG, S., AND WALTER, M. 2002. Recognition of human gestures and behaviour based on motion trajectories. Image and Vision Computing 20, 5-6, 349 – 358.

SHAN, C., WEI, Y., QIU, X., AND TAN, T. 2004. Gesture recognition using temporal template based trajectories. Proceedings of International Conference on Pattern Recognition 3, 954 – 957.

SHAN, C., TAN, T., AND WEI, Y. 2007. Real-time hand tracking using a mean shift embedded particle filter. Pattern Recognition 40, 7, 1958 – 1970.

SORIANO, M., MARTINKAUPPI, B., HUOVINEN, S., AND LAAKSONEN, M. 2000. Skin detection in video under changing illumination conditions. Proceedings of International Conference on Pattern Recognition vol.1, 839 – 842.

STAUFFER, C., AND GRIMSON, W. 1999. Adaptive background mixture models for real-time tracking. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2, 246 – 252.

SUN, J., JIA, J., TANG, C.-K., AND SHUM, H.-Y. 2004. Poisson matting. ACM Transactions on Graphics 23, 3, 315 – 321.

YANG, C., DURAISWAMI, R., AND DAVIS, L. 2005. Efficient mean-shift tracking via a new similarity measure. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition I, 176 – 183.

YUAN, M., FARBIZ, F., MANDERS, C. M., AND TANG, K. Y. 2008. Robust hand tracking using a simple color classification

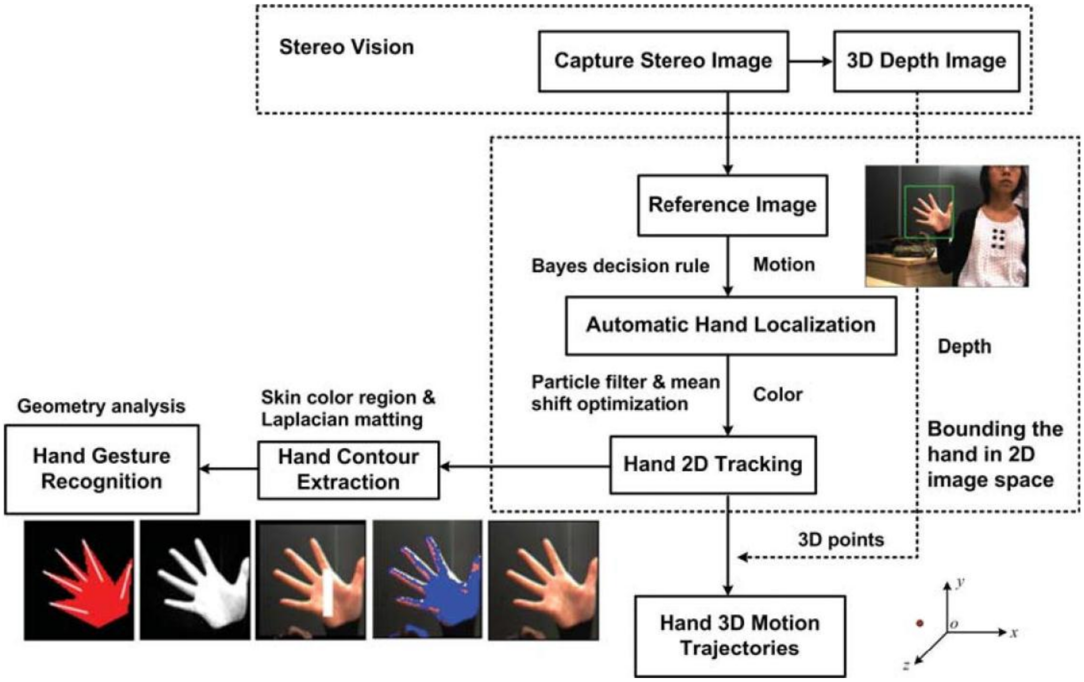

Figure 1: The framework of the 2D motion detection bounded hand 3D trajectory tracking and gesture recognition system.

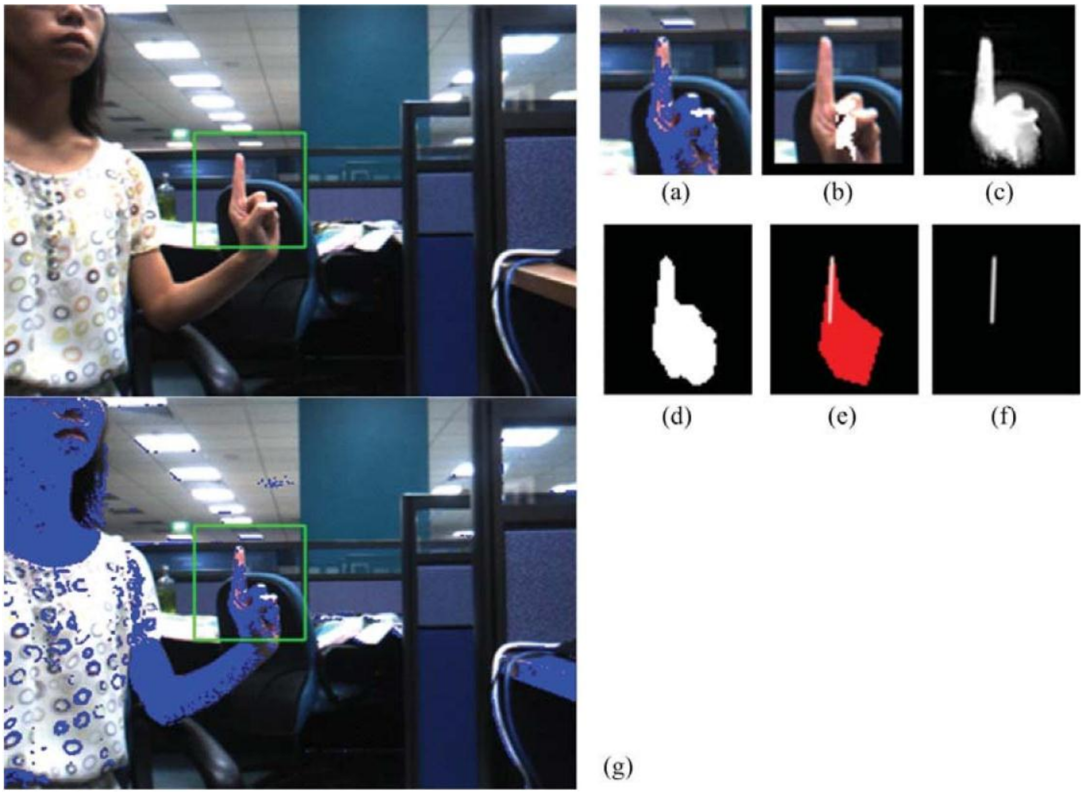

Figure 2: The tracking and recognition results from one test image sequence (frame 219) with the arms and face appeared in the scene.

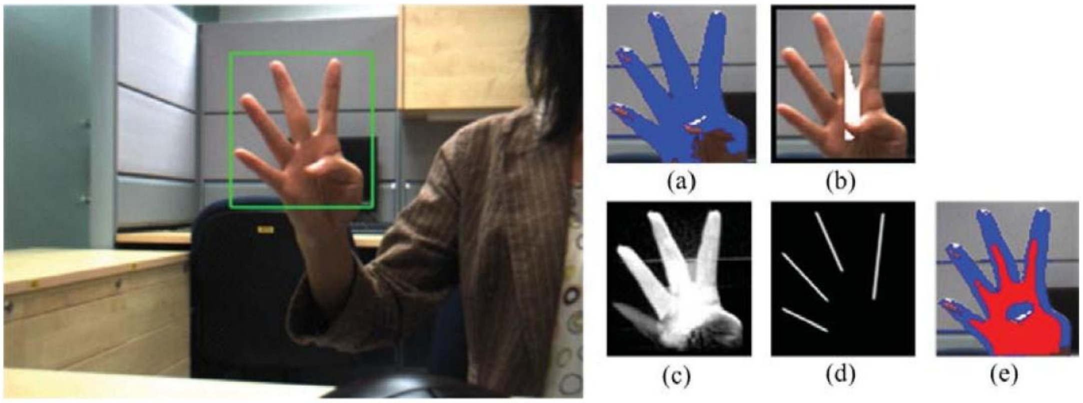

Figure 3: The tracking and recognition results from one test image sequence (frame 280) with different background and different gesture pose.

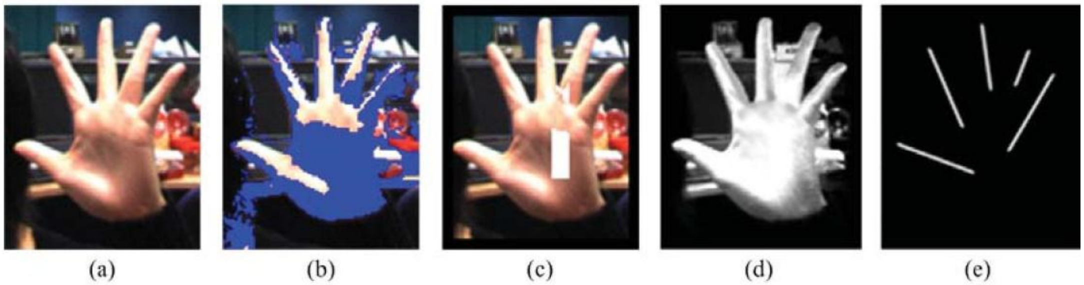

Figure 4: The recognition result from one test image sequence (frame 106) when the background in the bounding box has the similar skin color.

	Stereo imaging	Bounding the hand	Recognition
1	3782	59713	54467
2	3773	54267	63076
3	3783	53279	65920
4	3791	53176	60608

Table 1: The running times(ms) of the each module of the hand 3D trajectory tracking and gesture recognition system for one test image sequence that contained 600 frames.

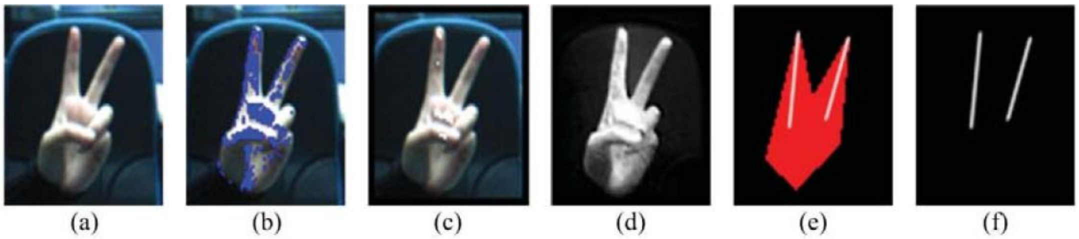

Figure 5: The example of the tracking and recognition result from one test image sequence (frame 58) under different light illumination.

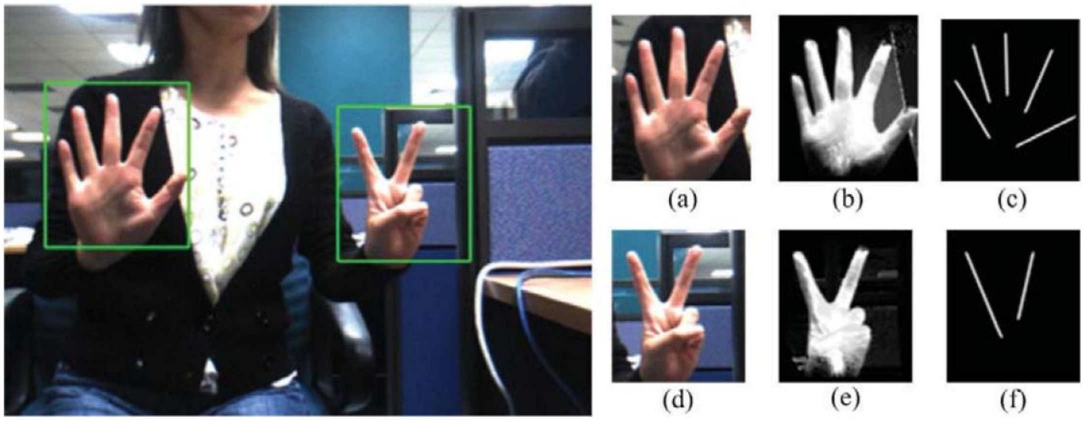

Figure 6: An example of the tracking and recognition result from one test image sequence (frame 80) with multi-hands.