Hand Gesture Contour Tracking Based on Skin Color Probability and State Estimation Model

Qiu-yu Zhang, Mo-yi Zhang, Jian-qiang Hu

College of Computer and Communication,
Lanzhou University of technology, Lanzhou, 730050, China

***Abstract*—considering the deficiency of accurate hand gesture contour inaccessible and inefficiency in complex dynamic background in existing methods of hand gesture tracking, a two dimensional skin color probability forecast method is proposed. Based on this, a hand gesture segmentation method of multi-mode and a hand gesture tracking method of state estimation are extended. When hand gesture is segmented, to locate the accurate hand gesture position, this paper combines the Skin Color Probability distribution with the statistical motion information of image blocking. Then the hand region is initiated by the region growth method and the hand gesture segmentation is realized. When hand gesture is tracked, the pixel's state model is built to estimate the state of pixels after watershed computation. Then the current blocking frame is adaptive threshold segmented and the hand gesture tracking is realized. Experiments show that this method has a strong anti-noise ability in complex background. In addition, it has a better application effect in segment and tracking the hand gesture contour accurately in a real-time way.**

***Index Terms*—hand gesture segmentation, hand gesture tracking, two dimensional skin color probability distribution, state estimation method**

I. Introduction

With the development of intelligent human computer interaction and the development of visual input, hand gesture recognition becomes a hot search in human computer interaction. The problems of dynamic hand gesture recognition with appearance [1] are hand gesture segmentation, hand gesture tracking, hand gesture feature extraction and semantic recognition. For a real-time system, the hand gesture segmentation and tracking [2] are the premise and foundation of hand gesture recognition, and the result of it directly affects the recognition efficiency of the system.

The hand gestures have myriads of changes, and if the tracking is just limited on the hand gesture position, the gesture's semantic information will lost. On the contrary, tracking of the hand gesture's accurate contour will lay a good foundation for the hand gesture recognition. As the hand gesture is non-rigid, it is not enough to consider only the translation, rotation and scaling transformation, but the deformation transformation must be considered. This brings a big problem in contour tracking, and the best segmentation is not easy to find.

Mainly considered from the probability and statistics, this paper proposes an efficient method. Based on the traditional one dimensional skin color distribution, a two dimensional skin color distribution is put forward. In segmentation, the first fame is mapped into the two dimensional probability distribution, at the same time the motion information is obtained by the statistical subtraction of the blocking image. Combined the skin color distribution with the motion information, the hand gesture position can be found accurately. Then the hand region is obtained by the region growth. After the segmentation of the first frame, the hand gesture can be forecasted and tracked on the basis of the previous frame. In hand gesture tracking, each pixel can be set as an independent system, and the state estimation method is used on each pixels. Firstly, the hand gesture region of the previous frame is used to build the statistical model of the current frame. Then each pixel's probability distribution of skin color is calculated and each pixel's state is estimated. At last the states of pixels are estimated statistically on the base of the sub-blocks after the watershed of the image. The hand gesture contour tracking is realized by the adaptive threshold binary of the current blocking image. The objective of accurately tracking the hand gesture contour is realized by the combination of the hand gesture image characters with the knowledge of the probability and statistic.

The second section is to introduce the skin color probability distribution. The third section describes hand gesture segmentation method of the first frame. The fourth section describes the state estimation method of hand gesture tracking based on the previous tracked information. The fifth section is the system overview. The sixth section is the experiment and the analysis of the result. The seventh section is the conclusion.

II. Skin Color Probability Distribution

A．One dimensional skin color probability distribution

Since skin color is more sensitive to H (hue) vector of HSV color space, H vector can be used to build traditionally one dimension skin probability distribution model[3]. Firstly, each pixel of skin color samples is converted from RGB space to HSV space. $C(h)$ is the pixel number of gray scale value of H vector, and $\max[C(h)]$ is the maximum of the $C(h)$, then this pixel's

doi:10.4304/jmm.4.6.349-355

probability of skin color which has the h vector of w is $p(w) = C(w)/Max[C(h)], h \in (0,1)$

So the gray scale value of h vector is one to one correspondence with the skin color probability, and the skin color probability table can be build, which is showed in figure 1a. Skin color probability distributions image is obtained due to this table, which is showed in figure 2b. But this kind of skin color probability distribution image is built only by the H vector, and a vector's restrict is too weak to accurately locate the hand gesture position. In addition this kind of distribution is sensitive to a great lot of non-skin noise.

B．Two dimensional skin color probability distribution

The build of the new model excludes the V vector which is directly related to light illumination, H and S vectors are used to describe a pixel's probability of skin color, and it describes the special color scope more accurately.

To statistically analysis the special skin color information, the total number of the skin color pixels is n_{totle} , the number of the pixels with H vector of H_i and S vector of S_j is

$$P(H_i, S_j), 0 \le i, j \le 255$$

Normalize it to [0,255] is

$$P'(H_i, S_j) = \frac{P(H_i, S_j)}{\max[P(H_i, S_j)]} \times 255$$
.

To set the H,S as the coordinate axis, and to draw the map using the $P'(H_i, S_j)$ as the gray scale of pixel, the (H,S) vector is in one to one correspondence with the skin color probability distribution and the two dimensional skin color search map is build, which is in figure 1b.

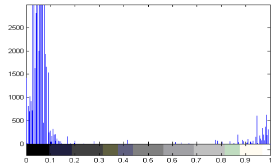
To convert the hand gesture image from RGB space into HSV space, and extract the H and S vector from it, then the pixels of the image can be mapped into the sum of the gray scale of a neighborhood of the pixel (H, S) in skin color searching map, so the corresponding skin color probability distribution image is obtained, which is in figure 2c.

Gary scale mapping function:

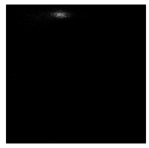
$$H(x, y) = \sum_{i,j=-n}^{n} P'(H\ (x+i, y+j), S\ (x+i, y+j))$$

$H\ (x+i, y+j)$ and $S\ (x+i, y+j)$ are the H and S vector of the pixel（x+i,y+j）respectively, and n is the size of the neighborhood.

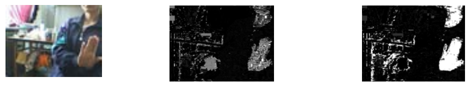
From the figure 2, the searching range of the skin color in two dimensional skin color search map is reduced obviously, so it can enhance the efficiency of skin color searching. It can see from the figure 3 that the hand color region in two dimensional skin color distribution image is more remarkably than the one dimensional and it can find the skin color region more easily.



a traditional skin color histogram of h vector



b two dimensional skin color search map
Figure 1.Skin color probability distribution



a original image b traditional skin color distribution image
c two dimensional skin color distribution image
Figure 2.Skin color distribution image of hand gesture

III. HAND GESTURE SEGMENTATION

A. motion information extraction

Skin color probability distribution image can find several skin color regions if there is a skin color background noise such as head. So the motion information must be used to locate the exact hand region in segmentation of the first frame. It is assumed that the background is relatively static and the person has moved in little scope, then it is possible for us to obtain the motion information by frame difference method. The traditional frame difference method is

$$d(x, y) = \begin{cases} 0, \left| f_1(x, y) - f_2(x, y) \right| \le \varepsilon \\ 1, other \end{cases}$$
,

where $f_1(x, y)$ and $f_2(x.y)$ represent the background pixels and current frame pixels respectively, and nonzero value represents the moving region in images. The moving extent is related to its different values, and the result is a grayscale intensity image, but, with calculating the moving information of each pixel separately, it is sensitive to noise and change of illumination. So the result is not always satisfied. A new division and statistic frame difference method is suggested (difference with the background) to find the moving information. It can avoid the effect of discrete noise successfully by statistic analysis, and make the moving information obtained stabilize and reliable.

From several experiments of Frame difference, the S vector of the HSV model for Frame difference method is decided to use to eliminate the effect of illumination, since S vector is not sensitive to illumination. After obtaining the frame difference image on S vector, it is divided evenly into parts, each of which is $m \times m$ in size and the moving information of each of which is analyzed in statistic. A part is set as a sample; it is proposed to calculate its mean $E_{x,y}$ and variance $S_{x,y}$, and combine the two information as

$$D_{x,y} = \sqrt{E_{x,y}{}^2 + S_{x,y}{}^2}$$

which is the obtained statistic moving information. The moving information $D_{x,y}$ of each pixel, which is owned by the part that it belongs to, is the statistic value that will be stored, so that the same value will be found in a part. Normalizing $D_{x,y}$ to [0, 1] can obtain its statistic moving information image. From the figure 3, it is shown that the statistic moving information image is a gray scale image and the position of the moving object can be found. The more the moving information, the whiter it will be.

B. hand gesture region location

In the hand gesture segmentation of the first frame, the statistic moving information and skin probability distribution image must be combined to locate the position of the hand gesture. The seed point is calculated in order to gain the clustering pixels to segment the hand gesture.

The traditional way of calculating [4] the middle point is following:

$$w_u = \frac{\sum_x \sum_y xB(x,y)}{\sum_x \sum_y B(x,y)} \quad w_v = \frac{\sum_x \sum_y yB(x,y)}{\sum_x \sum_y B(x,y)} \tag{1}$$

$\sigma(w_u, w_v)$ is the middle point, and $B(x, y)$ is the value of pixels on skin probability distribution image. The improved method of calculating the middle point combined with the statistic moving information is following:

$$w_u = \frac{\sum_x \sum_y xD(x,y)B(x,y)}{\sum_x \sum_y B(x,y)D(x,y)} \quad w_v = \frac{\sum_x \sum_y yD(x,y)B(x,y)}{\sum_x \sum_y B(x,y)D(x,y)} \tag{2}$$

$D(x, y)$ is the value of statistic moving information. The improved method adds the statistic moving information value. In addition, the moving information can be modified in some extremely complicated background. It can increase the value of the moving information of the actual moving regions and decrease the value of the moving information of the noise points. As a result the improved method can increase their difference to enhance the impact of the moving information toward the calculating of the middle point of the hand gesture. The feasible method is to have the obtained moving information with n power, and an element α can be multiplied to in case the result of the power operation is not fit to store and calculate. The value of α can be decided under the real condition.

$$D_{x,y}' = (\sqrt{E_{x,y}{}^2 + S_{x,y}{}^2})^n \times \alpha \tag{3}$$

On the condition that the hand gesture is the mainly moving object with the skin color, and that the large region of the skin color noise is relatively moved small, because of the add of the statistic moving value, the calculated middle point will be near the area of the more active object, namely near the hand gesture, to abandon the head information to find the hand gesture position.

C. Hand gesture region extraction

In the hand gesture region of the skin probability distribution image, the hand pixel value is not distributed evenly. Further more, the value is not a continuous variable, so it is hard to handle by the ordinarily method. The skin probability distribution image shows that the gray scale value is larger and more collective in the region of the hand gesture, and the value of non-hand region shows darker. So it is better to use a small threshold to binary the intensity image of the skin probability distribution, then to segment the hand gesture by the use of density information.

With using the density information of the discrete value, the evenly division of the image is also needed. This division is the same as the division when it obtains the statistic moving information. When the size of a part is set to be $m \times m$ and the density of a part is $q/(m \times m)$, there are q (0<q<= $m \times m$) hand pixels in every parts. If q is larger than the threshold, then the number of hand pixels is enlarged and all the pixels in this region are set as hand pixels (value 1). So it can avoid the disadvantage of the discrete pixel value, and map the hand into a connected region to realize a simple, reliable region growth algorithm. The threshold can be selected by experience or by tests. Then the hand region was extracted by region growth which starts from the selected seed point.

The steps in detail are to search the surrounding area connectivity (4-connected neighborhood or 8-connected neighborhood), then to merge the hand region of 1 value. Not until no extra hand area exists with the expanding pixels, is the hand area segmented accurately. The result of the region growth is in fig.4.

On the segment process mentioned above, the ROI (the Region of interest [5]) is introduced in order to decrease the computational complexity, since the skin color probability distribution model, moving information analysis, and region growth are three united processing progress. The first ROI is the search window $W_1(x, y)$ of the skin color area obtained by skin color distribution images. Then, based on the $W_1(x, y)$, rather than the whole image, the statistic moving information is calculated to obtain the window $W_2(x, y)$ as the second ROI. The experiments show that this method can decrease the computational complexity substantially in hand gesture segmentation.

Figure 3.Original hand gesture images and statistic moving information images

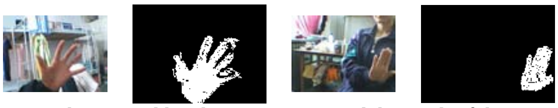


Figure 4.The original hand gesture image and the result of the region growth

IV. HAND GESTURE TRACKING

On the condition that the rate of hand gesture moving is not very fast, and that the hand gesture moves smoothly, the pre-frame information and the color distribution of the current frame can be used in tracking the successive hand gesture images.

A. Shape based Statistical model

The statistical model is built on the basis of the object shape tracked in previous frame. The method in reference [6] is to build the exponential distribution statistical model of the object. In the middle of the region, the probability of moving object is 1, and in the region which has the distance to the middle point larger than the threshold, the probability of moving object is exponentially decay. The object's shape is not considered in this method but only the distance between the pixel and the middle point is used as the parameter of the statistical model. Considering this deficiency, a shape based statistical model of the moving object is built here.

The deformation of the hand gesture shape is smooth in continuous frames, and the difference between the two continuous fames is not obviously. In order to control the deformation of the structure, the idea of deformation template based tracking [7] can be used. Take the biometric based deformation template [8] as a reference, the model that can control the deformation is designed. And the deformation range and probability are set to make the model has the ability to control the deformation.

As the figure 5 shows, the inside curve is defined as the kernel curve $C_{\ker}$ of the model, the outside curve is defined as the borderline curve C_{bord} of the model, the middle curve is the hand gesture contour, and the region between the $C_{\ker}$ and C_{bord} is the deformation region. The pixel probability of hand is 1 in region inside $C_{\ker}$, the pixel probability of hand is 0 in region outside C_{bord}, and in the region between the $C_{\ker}$ and C_{bord} the probability is decline from the $C_{\ker}$. The equation of the probability model is the equation 4, the $P_{pos}(x, y)$ is the position of the pixel (x,y), and the image of the object shape based model is showed in figure 7c.

Figure 5.the hand gesture template

$$P_{hand}(x,y) = \begin{cases} 1 & P_{pos}(x,y) \textit{ is inside the } C_{\ker} \\ \dfrac{1}{\min(\| P_{pos}(x,y) - C_{\ker} \|)} & P_{pos}(x,y) \textit{ is between the } C_{\ker} \textit{ and } C_{bord} \\ 0 & P_{pos}(x,y) \textit{ is outside the } C_{bord} \end{cases} \tag{4}$$

The question is how to create the $C_{\ker}$, C_{bord} in model structure, and how to get $\min \| P_{pos}(x,y) - C_{\ker} \|$? The clue to solve this is: The deformation region is defined using the previous hand region tracked or segmented. And the deformation region is directly related to the maximum rate of the hand deformation between two successive frames, which is assumed as N. To erode and dilate the binary image of the previous frame N times, the region contour of the Nth dilate is the borderline curve C_{bord}, and the region contour of the Nth erode is the kernel curve $C_{\ker}$. Set the binary image of the Nth dilate and the Nth erode as A and B respectively, the original hand binary image as C, the A-B image is the deformation region, and A-C and C-B is the dilate region and erode region of the model respectively. The minimum of the distance between $P_{pos}(x,y)$ and $C_{\ker}$ is,

$$\min \| P_{pos}(x,y) - C_{\ker} \|$$

and it meet the equation 5:

$$\min \| P_{pos}(x,y) - C_{\ker} \| = \begin{cases} N + n_{dilate} & P_{pos}(x,y) \in A - C \\ N - n_{erode} & P_{pos}(x,y) \in C - B \end{cases} \tag{5}$$

n_{dilate} and n_{erode} are the dilate number and erode number that the original hand binary image dilate and erode to reach to the pixel $P_{pos}(x,y)$.

To build the statistical model of all the pixels of the image, every pixel's probability of hand gesture is obtained, it is related to the shape of the object, and it is more accordance with the object's deformation and transformation character.

B. State estimation method

The biological model of PCNN set one pixel as one neuron in image processing, and each neuron is the coupled oscillator [9] that has connected with other neurons. This test exactly proceeds from this angle, it set each pixel as a system, and each system's state translation is analyzed statistically. The state translation relationship is showed in figure 6. Each system is defined to have two states, which is S_1 and S_2. S_1 denotes that the current pixel is in hand region; S_2 denotes that the current pixel is not in hand region; the input signal x_i denotes the current pixel color's probability of hand gesture. It is known that on the moment of t-1 the pixel's probability of hand region is p(S(t-1)), then the signal x_i is inputted, then on the moment of t the pixel's probability of state S(t) can easily obtained. It is showed in equation 6.

$$p(S(t),S(t-1)\,|\,x_i)=p(S(t-1))\times p(S(t)\,|\,S(t-1),x_i) \qquad (6)$$

After the obtaining of the tracked hand gesture of the previous frame, the skin color in hand region is combined with the skin color sample to update the current skin color sample. It can make the method has the ability of self learning and reach the special people's skin color. The current hand color sample is used to calculate the two dimensional skin color probability distribution of the current frame, and it is substituted in function (6), then the pixel probability of hand gesture in current frame can be estimated.

Considering the state statistical model of the image, in calculation of the two dimensional skin color probability distribution of the current frame, the deformation region in image can be set as the ROI, and the skin color probability inside the kernel curve set as 1, and the skin color probability outside the borderline curve set as 0. It can decrease the computational complexity substantially.

To compute on the independent pixel, it is the skin color information that can be used to update the pixel probability of hand gesture. Not considering the pixel's distribution in the region, the state of pixel is easy to be disturbed by noise. Considering this aspect, the image is blocked by the single scale watershed transform [10], which is showed in figure 7b. As the same sub-block have the same color and motion character, so the pixels in the same sub-block are definitely belong to the same object. Based on this, the pixel's probability of hand gesture is statistically analyzed after the image blocking. The pixel's probability of hand gesture is set as the average of the pixel's probability of hand gesture of the block.

After the blocking, the histogram of the pixel's probability of hand gesture is built. Set the first valley value of the histogram as the threshold, so the binary hand image can be binarized from the current hand gesture probability distribution image and the tracked object can be obtained. The number of the hand pixels tracked can be counted and compared with the number of the pixels tracked from the previous frame. As the moving of the hand gesture is smooth, the number of the hand pixels of the two successive frames must be approximately the same or the difference is in certain scale. If the difference is too obviously, it means that the region tracked is incomplete or the region have some noise included, so the threshold must fine tuning toward the corresponding axis direction to make the pixel number difference in certain scale and the area change is in a certain scope also.

Figure 7d shows that the tracked hand gesture contour is very smooth and is approximately the accurate segmentation, so the tracked result of the hand gesture can meet the practical need.

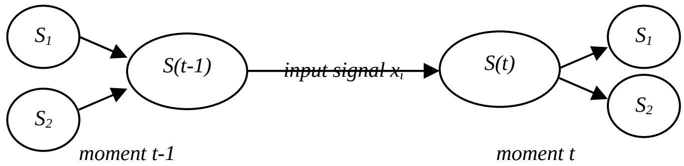
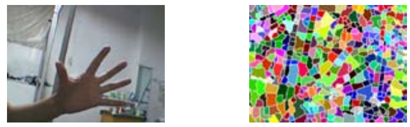


Figure 6.the relationship between t moment state S(t) and t-1 moment state S(t-1)



a original hand gesture image b watershed image

c the shape based statistical model map d the tracked hand region

Figure 7.the steps of the hand gesture tracking

V. SYSTEM OVERVIEW

The framework of our hand gesture tracking system is as follows:

(1)Skin color information handling. The two dimensional skin color probability distribution map is got by the skin color collection of the hand. And the two dimensional skin probability distribution image of the first frame is obtained.

(2) Motion information obtaining. The blocked image of the first frame was differentiating with the background, and the sub-block is set as 5×5. The motion information of the first frame is calculated by equation (2) and modified by equation(3), where α =100, n=4.

(3) Segmentation. Skin color probability distribution was combined with the statistic motion information value to find the seed point for region growth. So the hand region is segmented by sub-block based region growth from the seed point, the threshold is set as 0.1, q=2, and a sub-block is set as 2×2.

(4) Statistical model built. In hand gesture tracking, the previous hand binary image is used to build the shape based statistical model, N=7, and the 3×3 rectangle model is used in morphological operation.

(5) Skin color probability distribution obtaining. The hand color of the previous tracked hand region combined with the hand color sample is used to update the current special hand color which is used to calculate the two dimensional skin color probability distribution.

(6) State estimation. The pixel probability of hand gesture in current frame can be calculated by the equation (6). The current frame is blocked by watershed computation and the probability of the pixels of the hand is reset as the average probability of the sub-block. Then the histogram based on pixel's probability of hand can be build. The first valley value of the histogram is set as the

threshold to segment the hand statistical image to obtain the hand region.

(7) Adjustment. The area of the hand gesture region segmented is counted to adjust the result of tracking. If the area is obviously different with the previous one, it is means that there has a noise disturbance, so the threshold must be tuning towards the certain axis direction and the image must be segmented another time until the region area is in a certain scope. Then the system returns to the step (4) to track the next hand gesture frame.

VI. THE EXPERIMENTAL RESULTS AND ANALYSIS

The VC++ is used to realize the hand gesture segmentation and hand gesture tracking method proposed here and two groups of experiments are done in order to verify the effectiveness of the method. An ordinary USB camera is used to collect a group of unmarked hand gesture video stream and a group of face video stream indoors. The method ability of tracking the moving skin color object is tested on an ordinary computer with the Pentium IV 1.6GHz CPU and the 256M memory. The resolution of the camera used in this system is 320×240, and the captured image is 24 truecolor (RGB). The rate of the hand gesture tracking is 17 frames per second. It can meet the real-time and accuracy requirement in hand gesture contour tracking in human computer interaction.

Figure 8 shows the result of the hand gesture tracking indoors using the method of state estimation while hand gesture is in translation, rotation, scaling transformation and deformation. The hand’s shape and deformation in figure is relatively complex, but by the use of the state estimation method with the partition based adaptive threshold segmentation, the hand gesture contour can be tracked accurately. And the method is robust in complex background. In some frames, as there is a drastic change in color, the contour tracked is incomplete, and the efficiency of the method is reduced, such as the image 4 of the figure 8.

Figure 9 shows that when there is a person walking in the background which makes a moving skin color disturbance, the result of the face tracking by this method. It can see from the figure that when the walking person makes environment and light illumination change drastically, the face contour can be tracked accurately in the video stream. It indicates that the statistical state estimation method combined with the skin color model have a good anti-noise ability. So this method can solve the problem of ordinary background change and skin color disturbance.

Figure 10 shows that when 1% pepper and salt noise and gauss noise is added respectively, the result of the tracked face. It can see from it that when noise is added with no skin color disturbance, the tracked contour is as accurate as before. When noise is added with skin color background disturbance, the tracked contour of the object is not as accurate as before. This is mainly because the noise added make the information that obtained from the skin color not enough to distinguish the exact contour. So the motion analysis should be considered to add to update the method in the future to make it more adaptive to complex background.

Table 1 shows the result of the comparison between this method and the other method of the hand gesture tracking based on single video. Compared with the other hand gesture tracking method, this method realized the accurately hand gesture contour tracking in dynamic background and has some kind of anti-noise ability. In aspect of the tracking rate, this method is obviously more efficient than the Ivan Laptev[11]’s multi-mode hand gesture tracking method (10f/s). The rate of the single hand gesture tracking method of Yoichi Sato[12] is 25~30 f/s, which is a little better than the method here, but it can only track the hand gesture of a single form and can not solve the problem of hand gesture deformation. The deformation method of Wang Xi-Ying[13] can just track the limited 6 kind of special forms. All of the other methods don’t consider the complex background and skin color disturbance.

VII. CONCLUSION

A method based on two dimensional skin color model, statistical segmentation, and state estimation is proposed to segment and track the hand gesture contour in video stream. This method overcomes the deficiency of accurate hand gesture contour inaccessible and inefficiency in complex dynamic background in existing hand gesture tracking method. This method firstly uses the skin color distribution and the motion information to segment the first frame of the hand gesture, and then uses the state estimation to track the hand gesture contour on the basis of the previous frame. In the method, the skin color, region partition, shape information, and especially the statistic state estimation method are used to reduce the deformation range to control the deformation to track better. Experiments show that this method can track the hand gesture contour accurately in video stream in real-time. Especially when noise is added, a good tracking result will also be achieved.

This method can also be used to track the other objects with special color. But when the background is extremely complex or skin color disturbance is severely, it may reduce the efficiency in tracking the contour. This is mainly due to the fact that when hand skin color is dramatically changed in a short time or severely disturbed by the skin color noise, it is not easy to track accurately. This deficiency should be overcome in the future.

Figure 8.the hand gesture tracking experiment

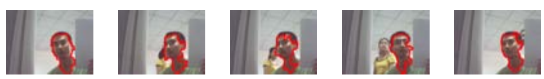
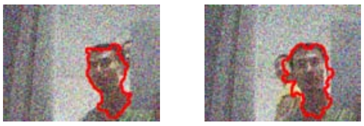


Figure 9.the hand gesture tracking experiment



a the tracking result when 1% gauss noise is added

b the tracking result when 1% pepper and salt noise is added

Figure 10.the tracking result of the face when the noise is added

Table 1.
The comparison of tracking methods

	Our method	IvanLaptev method	Yoichi Sato method	Wang Xi-Ying method
Average tracking rate(f/s)	17	10	>=25	18
Deformable gesture tracking	Yes	No	No	limited
Gesture contour tacking	Yes	No	No	No
Anti-noise ability	Yes	No	Yes	No
Dynamic background	Yes	No	No	No

REFERENCES

[1] Hongmo Je, Jiman Kim, Vision-Based Hand Gesture Recognition for Understanding Musical Time Pattern and Tempo. The 33rd Annual Conference of the IEEE Industrial Electronics Society, Nov. 5-8, 2007, Taipei, Taiwan.

[2] Thomas Gumpp, Pedram Azad, Unconstrained Real-time Markerless Hand Tracking for Humanoid Interaction, Proc. IEEE–RAS Int. Conf. on Humanoid Robots, 2006.

[3] Haiting Zhai,Xiaojuan Wu, Research of a real-time Hand Tracking Algorithm, ICNN&B’05,International Conference, Vol2, 13-15 Oct,2005.

[4] Mohamed Rizon, Haniza Yazid, Object Detection using Geometric Invariant Moment, American Journal of Applied Sciences,2(6):1876-1878,2006.

[5] G.McAllister, S.J.McKenna, Hand tracking for behavior understanding, Image and Vision Computing, 20(12):827-840,2002.

[6] Zha Yu-fei, Bi Du-yan, An adaptive Particle Filter for Moving Objects Tracking[J], Journal of Electronic& Information Technology, 2007,29(1):92-95.

[7] Yu Zhong,Anil K.Jain, Object Tracking Using Deformable Templates, IEEE Trans, Pattern Analysis and Machine Intelligence,vol.22,No.5,May 2000.

[8] Tang Ze-ying, Lu Han-qing, A Novel Analytical-form-based Parametric Deformable Template[J], Journal of Image and Graphics, 2004,9(7):775-780.

[9] T.Lindblad and J.M.Kinser, Image Processing Using Pulse-Coupled Neural Networksp, 2005.

[10] Shen Yun-tao, Guo Lei, Moving object tracking algorithm based on Hausdorff distance [J], Computer Applications, 2005,Sept,Vol25,No9.

[11] Laptev I, Lindeberg T. Tracking of multi-state hand models using particle filtering and a hierarchy of multi-scale image features, In: Scale-Space 2001, LNCS 2106,2001,63－74.

[12] Sato Y, Kobayashi Y, Koike H, Fast tracking of hands and fingertips in infrared images for augmented desk interface, In: Proc. of the IEEE Int’l Conf. on Automatic Face and Gesture Recognition, 2000, 462－468.

[13] Wang Xi-Ying, Zhang Xi-Wen, An Approach to Tracking Deformable Hand Gesture for Real-Time Interaction, Journal of Software,Vol18,No10,October 2007.