Application of an Improved Mean Shift Algorithm in Real-time Facial Expression Recognition

Zhao-yi PENG
School of Computer and Communication, Hunan University of Technology, Hunan, Zhuzhou, 412008，china
Email:pengzhaoyi@163.com

Yu ZHOU
School of Electrical and Information Engineering, Hunan University of Technology, Hunan, Zhuzhou, 412008，china
Email:zypzy@163.com

Yan-hui ZHU
School of Computer and Communication, Hunan University of Technology, Hunan, Zhuzhou, 412008，china
Email: swayhzhu@163.com

Zhi-qiang WEN
School of Computer and Communication, Hunan University of Technology, Hunan, Zhuzhou, 412008，china
Email: zhqwen20001@163.com

***Abstract*—In real-time facial expression recognition, accurate and fast face tracking is a very important preparatory part to obtain the image sequences of facial expressions. For this problem, an improved mean shift algorithm is proposed for real-time face tracking. Facial expression image sequences are obtained with the method. The method is based on using pixel gray value distribution as the feature as well as combination of the density distribution of the objective gradient direction. Alternating iterative operations can be carried through the iterative formula of these two features, and thus, we can make the human face rotation and translation movement tracking better. Then we used the geometric model based on human face to locate the region of facial expression features, and estimate the optical flow to calculate the Eigen-flow vectors. Finally, hidden semi-Markov model is used for facial expression recognitions. Experiments show that the proposed method can effectively track the face under rotation and translation movement of the head and it is very effective to obtain the facial expression image sequences quickly and accurately.**

***Index Terms*—improved mean shift algorithm, gradient direction, face tracking, facial expression recognition, image sequences, HSMM**

I. Introduction

Facial expression recognition is one of the important research areas of human-computer interaction. Expression recognition is divided into the expression recognition of static images and dynamic image sequence. The expression recognition based on dynamic image sequence has more value of practical application. It is widely used in public security, medical and other special departments as well as some commercial applications. For example, in the applications of video phone and video conferencing, etc, human poses and motions of the head change constantly in real time. Therefore, in order to identify the facial expression at the moment, first we must track the human face quickly and accurately and obtain the image sequence of the dynamic expression.

At present, the research of expression recognition focuses on the static image or image sequence of static images and has got much successful experience [1-3]. However, the study on the dynamic facial expression the recognition of real-time movement of the head is relatively lack of experience. In the research of dynamic real-time facial expression recognition, it is a very important preparatory work that we can obtain face image sequences accurately and rapidly.

Mean shift algorithm is brief and real-time, and able to deal with the partial occlusion and the deformation of the target, so it is widely used in real-time target tracking [4-6]. During its practical application, some scholars have improved the algorithm of mean shift to better the effect of tracking. For example, based on Lindeberg's the theory of scale space, Collins has come up with a new way used to handle the algorithm of mean shift, which could not adjust to the changes of the target scale very well as it's defects [7]. Considering the disadvantage of the classical tracking algorithm of mean shift, that's to say, many empty histogram intervals could not express the distribution of the target color accurately as a result of the evenly splitting of the whole color space, Li Peihua has put forward an improved method [8], i.e. through the cluster analysis of the color and proper division of color space of the target, the new color mode was applied to the algorithm of mean shift at last. However, in the aspects of the recognition of the real-time facial expression, these algorithms of mean shift had ineffective current effects of tracking because of the various motional head postures and the large scope of the rotate angels.

Aimed at the circumstance of the varied head postures, this paper makes use of the density distribution of gradient directions to improve mean shift algorithm and uses the improved algorithm to track and locate faces, which, on the one hand, can better the tracking of faces

doi:10.4304/jsw.6.1.100-107

when the motional postures of head are changeable; on the other hand, it can acquire the expressive information of the image sequence of face more accurately.

After real-time tracking face, we extracted facial feature regions using geometry model based on human face. We calculated the eigen-flow vectors by using optical flow estimation method. Finally, in the expression classification and recognition, we used hidden semi-Markov model to describe the model of facial expressions, then classified and recognized dynamic facial expressions.

The framework of real-time facial expression recognition system is shown in Figure 1.

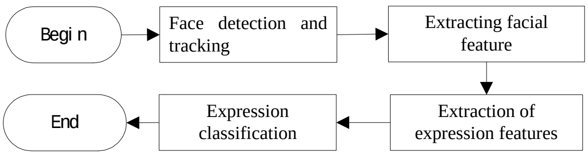


Figure1. The framework of real-time facial recognition system

II. REAL-TIME FACE DETECTION AND TRACKING

During the process of real-time expression recognition, first, it should track the face, then, extract the facial features correctly. As in the real-time environment, there is no law of movement, which leads to the varied postures of the head real-time. Mean shift algorithm is a classical algorithm with target tracking of the real-time movement, but it's effect of tracking is not good when the target is rotating intensely, so we researchers use an improved mean shift algorithm to improve the real-time target tracking while the target is rotating.

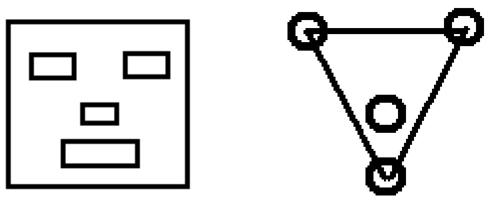
A. Detection the initial face

When using Mean shift algorithm to track face automatically, first we need to determine the initial search window, which is the initial target. In order to obtain the initial target face, we determine the initial target by face detection method based on skin-color model. The value of I in the YIQ space of face skin-color change in a particular range between 30 to 100 [9]. At the same time, in the YUV space, we find that the range of hue θ of the face skin-color is between 105˚ and 150˚. Synthesizing YIQ and YUV color systems, we establish an initial face skin-color model:

If image P can meet the two conditions simultaneously: $30<= I<=100$, $105^0<=\theta<=150^0$, the point P is a skin-color point.

Because human skin color is affected by different illumination, environment and various poses, the obtained value of face skin color will deviate. Therefore, only using a fixed threshold to segment skin color and non-skin color would lead to relatively bigger errors and the results of segmentation are unsatisfactory. We adopt an adaptive method of Gamma Correction [10], which can effectively reduce the influence of illumination and postures on skin color, thus it can greatly enhance the accuracy of skin color segmentation.

Using face skin color detection we can remove the most of the non-face skin image information. Then we can locate the face precisely according to the facial structure model on the remaining candidate face area. There are distinctive features on the human face. On the front face, two eyes are on a straight line, and the two eyes and the mouth form an isosceles triangle, in which the two eyes and the mouth are the three vertices of the triangle. In this triangle the object close to the vertex of mouth is the nose. The front face's typical structure is showed in Figure5, where Figure (a) is a typical characteristic pattern of front face, Figure(b) is a structure model of front face.



(a) (b)

Figure 2 The structure of front face

When the body poses change, the motion of the face is formed by different angel rotation of the front face. So the face structure model used in multi-pose face detection is an ordinary triangle. Because there is still a nose on the face, there should be nose features in the triangle, and such face structure model satisfied this rule is called triangular triangle. Only if such a triangular triangle is detected to satisfy the rules in skin color area could it be determined to be a face.

The target window of initial face is as shown in Figure 3.

Figure3. The target windows of the initial face

B. Using mean shift algorithm to track face translation

Once the initial tracking window is determined, the new location of the window can be tracked by using mean shift iterative calculation.

Mean shift algorithm is a traditional gradient estimation algorithm, which were put forward by Fukunaga and Hostetler in 1975 [11]. People didn't interested in the study on this algorithm until Cheng [12] published a paper in 1995, then a tide of studying and applying mean shift algorithm was set off. At present, mean shift algorithm has been widely used in the areas such as computer vision and pattern recognition, such as: target tracking [4, 13], image segmentation [5], pattern recognition and cluster analysis [6], filter [14], etc.

The iterative formula of mean shift algorithm can be expressed as [15]:

$$\mathbf{y}_{t+1} = \frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\|\frac{\mathbf{y}_t - \mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{y}_t - \mathbf{x}_i}{h}\right\|^2\right)} \quad (1)$$

Its steps are as follows:

(1) Set the initial value $\mathbf{y}_0$, the conditions of the end of iteration ε , the initial t = 0;

(2) Use formula (1) to calculate the value of $\mathbf{y}_{t+1}$;

(3) Determine whether the results meet $\left\|\mathbf{y}_{t+1} - \mathbf{y}_t\right\| \le \varepsilon$, the calculation is finished if meet; Otherwise substitute $\mathbf{y}_t$ with $\mathbf{y}_{t+1}$, and turn to step (2).

Using mean shift algorithm in the target tracking, need to first select the target model, then through the window and target candidate's targets, and window to determine the target tracking.

(1) The target model

Target window is the window obtained by face detection mentioned above. $\{\mathbf{x}_i^*\}_{i=1...n}$ is the standardized pixel position in the region of target model. And a monotone decreasing function k(x) is used, which shows that the farther away from the center it is, the smaller its pixel weight is.

Function mapping $b: R^2 \to \{1...m\}$ quantifies the pixel color in the location of $\mathbf{x}_i^*$ into level m, which can be represented by function $b(\mathbf{x}_i^*)$. So the probability of the characteristics q_u （$u = 1...m$） in target model is as follows:

$$q_u = C\sum_{i=1}^{n} k(\| \mathbf{x}_i^* \|^2)\delta[b(\mathbf{x}_i^*) - u] \quad (2)$$

Where δ is the Kronecker delta function, C is a standard factor, in order to ensure $\sum_{u=1}^{m} q_u = 1$, so:

$$C = 1 \Big/ \sum_{i=1}^{n} k(\| \mathbf{x}_i^* \|^2) \quad (3)$$

(2) The candidate target

$\{\mathbf{x}_i\}_{i=1...n_h}$ represents the pixel location of the candidate target and the center of the current frame is y. The same kernel function $k(x)$ is used, and bandwidth is h, so the probability $p_u(\mathbf{y})$ of the characteristics in candidate target $u = 1,...,m$ is expressed as:

$$p_u(\mathbf{y}) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{\mathbf{y} - \mathbf{x}_i}{h}\right\|^2\right)\delta[b(\mathbf{x}_i) - u] \quad (4)$$

where：

$$C_h = 1 \Big/ \sum_{i=1}^{n_h} k\left(\left\|\frac{\mathbf{y} - \mathbf{x}_i}{h}\right\|^2\right) \quad (5)$$

C_h is a standardized constant that not dependent on y. x_i is the pixel location and y is one of the pixels, so for a fixed nuclear and different value of h, C_h can be pre-calculated. Bandwidth defines the scalability of the candidate target. n_h is the number of pixels that is considered in the locating process.

(3) Target tracking

Bhattacharyya coefficient $\rho(\mathbf{y})$ is used to measure the similarities of the two targets, where:

$$\rho(\mathbf{y}) = \rho[\mathbf{p}(\mathbf{y}), \mathbf{q}] = \sum_{u=1}^{m} \sqrt{p_u(\mathbf{y}) q_u} \quad (6)$$

Iterative formula of Mean shift algorithm is used, just like formula (7) shows.

$$\mathbf{y}_1 = \frac{\sum_{i=1}^{n_h} \mathbf{x}_i w_i g\left(\left\|\frac{\mathbf{y}_0 - \mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} w_i g\left(\left\|\frac{\mathbf{y}_0 - \mathbf{x}_i}{h}\right\|^2\right)} \quad (7)$$

Where:

$$w_i = \sum_{u=1}^{m} \sqrt{\frac{q_u}{p_u(\mathbf{y}_0)}} \delta[b(\mathbf{x}_i) - u] \quad (8)$$

Therefore face tracking can be simplified as: through repeated iteration of the mean shift vector, the target constantly moves to y1 from the current frame position y0, and look for the location that make the candidates target's features most similar to the target's. The complete algorithm is as follows:

(1) Initialize the target location y_0 of the current frame, calculate $\{p_u(\mathbf{y}_0)\}_{u=1...m}$, and evaluate the following formula:

$$\rho[\mathbf{p}(\mathbf{y}), \mathbf{q}] = \sum_{u=1}^{m} \sqrt{p_u(\mathbf{y}_0) q_u} \quad (9)$$

(2) Calculate the weight $\{w_i\}_{i=1...n_h}$ through formula (8).

(3) Find out the next the location of candidate target through the iterative formula (7).

(4) Calculate the value of $\{p_u(\mathbf{y}_1)\}_{u=1...m}$, then evaluate the following formula:

$$\rho[\mathbf{p}(\mathbf{y}_1), \mathbf{q}] = \sum_{u=1}^{m} \sqrt{p_u(\mathbf{y}_1) q_u} \quad (10)$$

(5) When $\rho[\mathbf{p}(\mathbf{y}_0), \mathbf{q}] < \rho[\mathbf{p}(\mathbf{y}_1), \mathbf{q}]$, so $\mathbf{y}_1 \leftarrow \frac{1}{2}(\mathbf{y}_0 + \mathbf{y}_1)$, calculate the value of $\rho[\mathbf{p}(\mathbf{y}_1), \mathbf{q}]$.

(6) If $\left\|\mathbf{y}_1 - \mathbf{y}_0\right\| < \varepsilon$, so the calculation is suspended; otherwise: $\mathbf{y}_0 \leftarrow \mathbf{y}_1$, and turn to Step 2.

Tracking results is as shown in Figure 4.

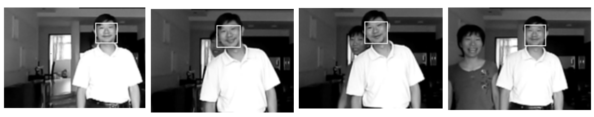


Figure 4 Using mean shift algorithm for real-time face tracking

C. The improvement of mean shift algorithm

In real-time face tracking process, there is the head of the movement of translational motion, and there will be a variety of rotary motion. Introduction of the mean shift in front of face tracking algorithm for tracking the head of translational motion has a better effect, but for a variety of posture, such as bow or head rotation, tracking the candidate window will contain more background pixels, resulting in tracking error, so tracking is not effective. For example: on the face shown in Figure 5, tracking results do not all cover the face area. Therefore, they easily lead to a loss of tracking targets. Peng Ningsong et al proposed a kernel function bandwidth tracking algorithm automatically selected gradient tracking window, and it is very effective [16], but the head movement in the rotation, on the face of the tracking results is not good. Therefore, in considering the target pixel value distribution characteristics and geometry information, and on the basis of using the density gradient target distribution characteristics, to improve the mean shift algorithm, more can be done to improve the attitude motion of the head when the human face is tracking.

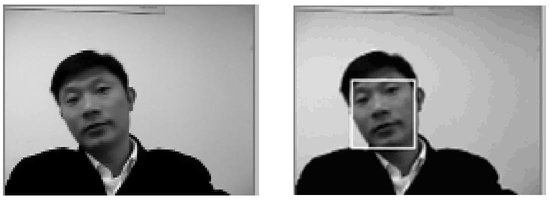


Figure5 Tracking performance chart of the head rotation

Let $\{\theta_u\}_{u=1,2,...,m}$ be target pixels gradient angle of discrete values, which is equivalent to $(-\pi, \pi]$ to be divided into m angles, its density distribution is $B(\theta_u)_{u=1,2,...,m}$, the formula of histogram-weighted based on kernel is formula (11) and (12) [17]:

$$\hat{q}_u = CB(\theta_u)k\left(\left\|\frac{\theta_u}{\pi}\right\|^2\right) \tag{11}$$

$$\hat{p}_u(\theta) = C_p B(\theta_u)k\left(\left\|\frac{\theta-\theta_u}{\pi}\right\|^2\right) \tag{12}$$

$\hat{q}_u$ is the density distribution of target, $\hat{p}_u(\theta)$ is the density distribution of the candidate target.

Where:

$$C = 1\Big/\sum_{u=1}^{m} k\left(\left\|\frac{\theta_u}{\pi}\right\|^2\right) \tag{13}$$

$$C_p = 1\Big/\sum_{u=1}^{m} k\left(\left\|\frac{\theta-\theta_u}{\pi}\right\|^2\right) \tag{14}$$

C and C_p are the normalization factors.

Reference to formula (6), the similarity of target is defined as:

$$\rho(\theta) = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\theta)\hat{q}_u} \tag{15}$$

Reference to iterative formula (7), we can get the estimates of θ when $\rho(\theta)$ is made significant, that is $\hat{\theta}_j$, the iterative formula is as follows:

$$\hat{\theta}_{j+1} = \frac{\sum_{u=1}^{m} \theta_u w_u g\left(\left\|\frac{\hat{\theta}_j-\theta_u}{\pi}\right\|^2\right)}{\sum_{u=1}^{m} w_u g\left(\left\|\frac{\hat{\theta}_j-\theta_u}{\pi}\right\|^2\right)} \tag{16}$$

Formula (16) In fact, is from the angle of direction to track the target, namely, can track the rotation of the target.

The Improved algorithm, which is based on the track of target's translation, combines the tracking of the rotation of target .The algorithm steps are as follows:

(1) Using the method above making the location of initial face, gain access to the initial search window, that is set the value of the initial position y_0 and orientation $\hat{\theta}_0$ in the iterative formula (7) and formula (16).

(2) Calculate the current position of $p_u(\mathbf{y})$, and calculate y_1 using the formula (7) .

(3) Calculate $\hat{p}_u(\theta)$, and calculate $\hat{\theta}_{j+1}$ using the formula (16) .

(4) If $\|\mathbf{y}_1 - \mathbf{y}_0\| < \varepsilon$ and $\|\hat{\theta}_{j+1} - \hat{\theta}_j\| < \varepsilon$, it suspends; otherwise: $\mathbf{y}_0 \leftarrow \mathbf{y}_1$, $\hat{\theta}_j \leftarrow \hat{\theta}_{j+1}$, and go to step 2.

We use the improved method to execute face translation and face tracking of the depth rotation by implementing the alternative and iterative calculations of formula (7) and (16). Because using traditional mean shift algorithm to track the mobile target, once the target is lost, need to re-detection and re-location the target, especially when head spins, the targets are lost more easily. So the algorithm leads to spend a lot of time on face detection of initial position. The introduction of improved method described above simplify the process of re-detection and location greatly, so it tracks the human face fast and efficiently for it considers the direction angle of the target and doesn't lose the target too easily.

D. Experimental results

To verity the effectiveness of the algorithm, we shoot a video about various postures of movements of head. Then, we intercepted series of bitmaps from each piece of video to form the image sequence. Before testing these bitmaps, we made filtering and other preprocessing. Then, detected the face of the first bitmap, and marked the initial position and orientation, which were shown in the first picture of Figure 4.

Throughout the experiment, we tested 120 images sequence of different pose. Table 1 is statistics of tracking results before and after the algorithm is improved.

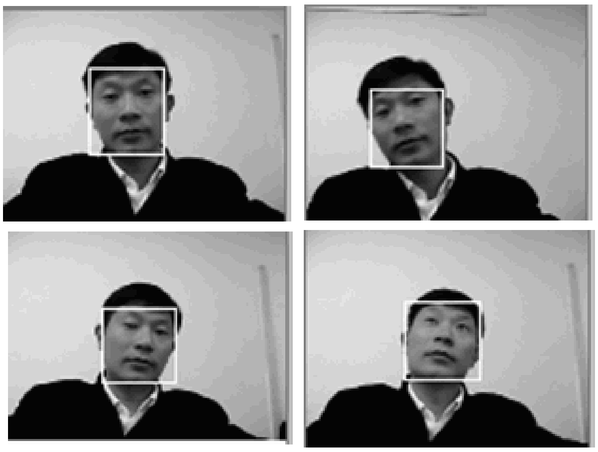
Table 1 Comparison of two algorithms for tracking results

Algorithm	Total testing	Number of successful tracking	Success rate (%)
Traditional algorithm of mean shift	120	99	82.5
Improved algorithm of mean shift	120	118	98.33

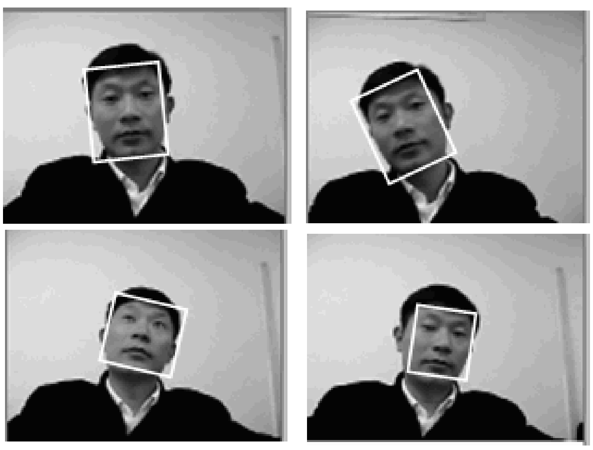
In table 1, the unsuccessful track means the target lost. Through the analysis, we found the most of unsuccessful track situations happen when the head deflection in the case of large angles. when head had a large deflection angle,

Therefore, in the process of different head poses, for example, improved tracking algorithm is more accurate than unimproved ones, and it is rarely to lose targets at the time when the head in deep rotation, and using this algorithm, the case of missing targets also rarely happen, so it saves a lot of time for the detection re-positioning of face, and greatly improves the tracking efficiency.

Figure 6 (a) and (b) shows the tracking's experimental results. Figure 6 (b) also shows: it can ensure the stability of tracking that target area covers nearly all the target pixels in tracking performance chart after algorithm improved.



(a) Tracking performance chart before using the improved method



(b) Tracking performance chart after using the improved method
Figure6. The target windows of the initial face

III. EXTRACT EXPRESSION FEATURES OF DYNAMIC FACIAL IMAGE SEQUENCE

After the face tracking of mean shift algorithm mentioned above, we can get the image sequence of the human face. Next we shall extract facial expression from these image sequences.

A. Locate the region of the facial expression features

Before extracting the facial expression, we should locate the regions of the facial expression features. Because the facial expression is mainly expressed in the areas of eyes, mouth, nose and so on, we must locate the feature regions first in order to extract these features. This paper adopts the method of eye location based on hybrid projection function while locating the eyes [18].

The method determined the center of the eye accurately through the hybrid projection function. Hybrid function combines the integral projection function and the variance projection function. The function comprehensively considered the mean and variance of the image in a certain direction. The function can not only reflect the different images gray-scale, but also reflect the different grayscale change rate, with strong flexibility and accuracy. So it can reflect comprehensive changes of the mean gray and variance in a direction of the image.

After having identified the center of the eye, we scan around the eye to get the feature area of the eyebrows.

Then scan the area around the eyes, and we can obtain the region of eyebrows' features. Then basing on facial geometric features [19], we find the regions of mouth and nose, and locate the feature regions of them. After locating these feature regions, standardize them, which mean to normalize their size, and we normalize them into the size of 20 * 30.

B. Optical flow calculation and feature extraction

Because optical flow contains the information of target's motions, we can use it to determine the movement of the target. In the expression image sequence, optical flow calculation method is used to extract the features of dynamic expression [20], because it can reflect the substance of expression changes and the features are less affected by illumination.

In optical flow estimation , we adopt gradient-based method proposed by Horn and Schunck [21]. This method is more suitable for the calculation of facial deformation. Its calculation is also relatively simple, just estimating the position's instantaneous velocity field point by point.

Optical flow is the movement speed of a certain gray pixel in the image plane. Described as below:

$$\frac{\mathrm{d}l}{dt} = \left(\frac{dx}{dt}, \frac{dy}{dt}\right) \quad (17)$$

Suppose there are two pictures of continuous projections in time *t* and time *t* + *dt*. In picture t there is the pixel whose gray scale is f (x, y). The coordinates of the pixel is x, y. The pixel shifted from the coordinate t + dt to the coordinate f (x + dx, y + dy), f (x, y, t) = f (x + dx, y + dy, t + dt). When changes are continuous and dt

→ 0 (dt is the time interval between two frames), By the Taylor series being launched, the result can be got:

$$\frac{\partial f}{\partial x}\frac{dx}{dt}+\frac{\partial f}{\partial y}\frac{dy}{dt}+\frac{\partial f}{\partial t}=0, u=\frac{dx}{dt}, v=\frac{dy}{dt} \tag{18}$$

Finally, we have to solve the optical flow by using the formula (19) [20].

$$\begin{cases} u_{kl}^{n+1}=\bar{u}_{kl}^{n}-\dfrac{f_x\bar{u}_{kl}^{n}+f_y\bar{u}_{kl}^{n}+f_t}{1+\lambda(f_x^2+f_y^2)}f_x \\ v_{kl}^{n+1}=\bar{v}_{kl}^{n}-\dfrac{f_x\bar{u}_{kl}^{n}+f_y\bar{u}_{kl}^{n}+f_t}{1+\lambda(f_x^2+f_y^2)}f_y \end{cases} \tag{19}$$

When the eigen-sequence would be obtained, we projected the playing field of the characteristics region in each frame of a set of image sequences on their corresponding substrate. Components u and v were combined serially as a eigen-flow vectors. Then a group of the eigen-flow vectors was made up of a eigen-sequence.

Then the eigen-flow vectors of a group of sequences is formed to a feature sequence.

Figure 7 is the optical flow field of eye region when a sadness facial expression is extracted.

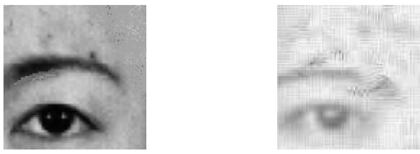


Figure7. Sadness expression image and its optical flow field

IV. EXPRESSION RECOGNITION

Expression recognition is to classify the expressions into different categories based on the different facial expressions. This paper adopts six kinds of expressions defined by Ekman, which are: surprise, fear, disgust, anger, Happy, sadness. Before the classification we establish the facial expression model first to describe the features of facial expression sequence. A facial expression model based on hidden semi-Markov model (HSMM) is adopted for facial expression recognition, because it has good results in the expression recognition for partly hidden face [21].

In HSMM, a facial expression feature sequence is expressed by a number of states and each state generates a section of feature vector. The facial expression recognition of feature sequence is to find the matching state sequences under the condition of known observed value sequence.

The expression model of HSMM is defined as: $\lambda=(A, B_1, B_2, ..., B_N, D, \pi)$, where: A is the probability distribution matrix state transition, $A=\{a_{ij}\}, 1\le i, j\le M$, $\sum a_{ij}=1$, M is the state number of the model .

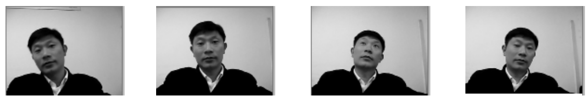
B_k is the probability of observed value, D is the probability distribution of the time of state dwell, π is the probability distribution of the initial state.

The training based on HSMM facial expression model can be referred to [22].

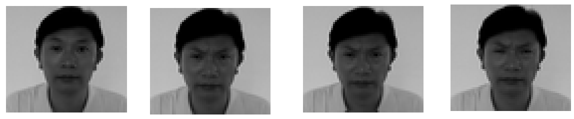
After the model training is complete, it's time to do testing. In recognition, Viterbi algorithm is used [19], which not only can find a good state transition path but also calculate the corresponding probability. Regarding to the facial expression model based on HSMM, the problem is to find a best state sequence for N observation sequences $O^k=O_1^kO_2^k...O_T^k$ generated at the same time, 1≤k≤N. Then respectively calculate the probability $P(O|\lambda_j)$ of this sequence in the six HSMM-based models, in which j∈[1, 6]. The largest value of probability $P(O|\lambda_j)$ is taken to be the number of the recognized expressions.

V. EXPERIMENTAL RESULT

The cohn-kanade [23] expression database and some of our own shooting expression data is used in this experiment. Train and recognize the six basic facial expressions defined by Ekman and establish an HSMM for each expression, then train respectively these six expressions. 200 expression image sequences are selected for training and 100 sequences are used for reorganization tests. Part of facial expression images sequence is shown in Figure 8. Only two places of face region are chosen: the eye region and the mouth region. In the training and recognition process, the HSMM states number M = 3, the expression features sub-region N = 2. The reorganization has achieved good results. Table 2 shows the results of real-time facial expression recognition.



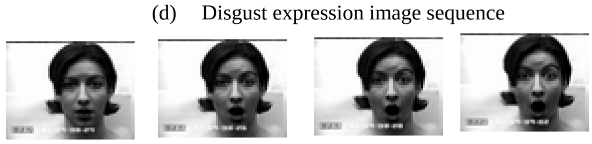
(a) Neutral expression image sequence



(b) Anger expression image sequence

(c) Happy expression image sequence

(d) Disgust expression image sequence



(e) Surprise expression image sequence

(f) Sadness expression image sequence

Figure8. Part of facial expression images sequence

TABLE 2 THE RESULTS OF REAL-TIME FACIAL EXPRESSION RECOGNITION

Emotion	**Hap.**	**Ang.**	**Dis.**	**Fear**	**Sad.**	**Sur.**
Recognition accuracy (%)	89.5	82.4	90.5	72.5	77.4	96.2

VI. CONCLUSION

This essay discusses the improved mean shift algorithm was applied successfully in real-time facial expression recognition. We used the improved mean shift algorithm for real-time face tracking to solve the problems that the bias window is larger and easy to lose objects in the traditional mean shift tracking algorithm, when the movements of head postures change frequently. Based on using pixel value distribution characteristics, the method that combines the density distribution of the target gradient direction improved the traditional mean shift method. Improved method realized the face tracking of head gesture multi changed, by making alternative and iterative calculations on the iterative formula of two characteristics. Thereby it reduces the face re-testing and re-locating process , resulting in by losing target detection and the tracking process. The method greatly improves the tracking efficiency. During determining the initial target, we used face detection method based on skin color model and used an adaptive gamma correction, so that we can solve the impact of complexion which is made by the illumination and pose. Thus the positioning accuracy of the initial face window is improved. After the facial expression image sequences were obtained, we adopted optical flow calculation method to extract the dynamic expression features, then built HSMM model to carry out dynamic facial expression recognition.

Experiments show that the proposed method in this paper is that the improved mean shift algorithm is better applied to real-time recognitions facial expressions. It provides a new method about dynamic real-time facial expression recognition. However, when sudden large-scale violent head movements happen, we still need to make face re-location. This situation affects the efficiency of the tracking, which is the problem we have to continue to study in the future.

ACKNOWLEDGMENT

This work was supported by Hunan Provincial Key Laboratory of Network and Information Security (No.NISL201001) and supported by the Education Department Foundation of Hunan Province (No. 09C324).

REFERENCES

[1] Pantic M, Rothkrantz L, “Facial action recognition for facial expression analysis from static face images” , *IEEE Transactions on Systems,Man,and Cybernetics-Part B,* vol.14,no.3,pp.1449-1461, 2004.

[2] Gao Wen,Jin Hui,“Analysis and recognition of facial expression image”, *Chinese journal of computers,* vol.20, no.9, pp.782-789, Sept,1997.

[3] PanticM, RothkrantzL, “Expert system for automatic analysis of facial expression”, *Image Vision Computing,*vol.18, no.11,pp.881-905,2000.

[4] Comaniciu D, Ramesh V, “Mean shift and optimal prediction for efficient object tracking”, *In Proc. of the IEEE Int’l Conf. on Image Processing,* New York: IEEE Press, pp.70-73, 2000.

[5] Comaniciu D, “Image segmentation using clustering with saddle point detection”, *In Proc. of the IEEE Int’l Conf. on Image Processing,* New York: IEEE Press,vol.3,pp. 297-300, 2002.

[6] Wua K, Yang M, “Mean shift-based clustering”, *Pattern Recognition,*vol.40,no.11, pp. 3035- 3052, 2007.

[7] R.T. Collins. “Mean-shift blob tracking through scale space”, *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp:234-241,2003.

[8] LI Pei-Hua, “An Improved Mean shift Algorithm for Object Tracking”, *ACTA Automatica sinica*.vol.33,no.4,pp:347-354,2007.

[9] Zhang H.M.,Zhao D.B.and Gao W., “ Face detection under rotation in image plane using skin color model, neural network and feature-based face mode”, *Chinese journal of computers,*vol.25,no.11,pp.1250-1256,2002.

[10] Peng Z.y.,Zhou Y. and Wang P., “Multi-pose face detection based on adaptive skin color and structure model” , *In Proceedings of the fifth International Conference on Computational Intelligence and Security(CIS2009), Beijing, China* ,pp.325-329,2009.

[11] Fukunaga K, Hostetler L D, “The estimation of the gradient of a density function, with applications in pattern recognition”, *IEEE Trans. Information Theory,* vol.21,pp: 32-40, 1975.

[12] Cheng Y Z. “Mean shift, mode seeking, and clustering”, *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol.17,no.8,PP. 790-799, 1995.

[13] Collins R T. “Mean Shift blob tracking through scale space”, *Proc. of the Conf. on Computer Vision and Pattern Recognition, Washington, DC: IEEE Computer Society,* vol.2,pp. 234- 240, 2003.

[14] Comaniciu D, Meer P, “Mean shift analysis and applications”, *In Proc. of the IEEE Int’l Conf. on Computer Vision*. New York: IEEE Press,vol.2,pp.1197-1203, 1999.

[15] Wen Z.Q., “Research on Moving Object Tracking Based on Mean Shift”, Doctoral thesis, Central South University,Changsha,China, Sept,2008.

[16] PENG Ning-Song, YANG Jie, LIU Zhi, “Automatic selection of kernel-bandwidth for mean-shift object tracking”, *Journal of Software,*vol.16,no.9,pp.1542-1549, 2005.

[17] WANG Chang-jun, ZHU Shan-an, “Mean shift based targets ’ rotation and translation tracking”, *Journal of Image and Graphics*, vol.12, no.8,pp. 1367-1371, 2007.

[18] Geng x.,Zhou Z.H. and Chen S.F., “Eye location based on hybrid projection function”, *Journal of Software,* vol.14, no.8,pp.1394-1399,2003.

[19] Yuille A, Hallinan P and Cohen D, “Feature extraction from faces using deformable templates”,*Computer Vision*, ,vol.8,no.2pp.99-111,1992.
[20] JIN Hui,GAO Wen. “Analysis and Application of the Facial Expression Motions Based on Eigen-Flow”.*Journal of Software*, vol.14,no.12,pp.2098-2105, 2003.
[21] Horn B, Schunck B, “Determining optical flow” ,*Artificial Intelligence*, vol.17,pp.185-203, 1981.
[22] Tian Jian-hua,Zhang Jian-ming and Shi Yi-bin. “Facial expression recognition based on hidden semi-Markov models”,*Computer Engineering and Design*,vol.28,no.23,pp.5564-5567,2007.
[23] Kanade, T., Cohn, J. F. and Tian, Y. (2000). Comprehensive database for facial expression analysis. *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00),* Grenoble, France, 46-53.

Zhaoyi PENG was born in 1968.He is an associate professor of Hunan University of Technology, Zhuzhou, China. His research interests are in digital image processing and pattern recognition.

Email:pengzhaoyi@163.com.

Yu ZHOU was born in 1974.She is an associate professor of Hunan University of Technology, Zhuzhou, china. Her research interests are in digital image processing and pattern recognition.

Yanhui ZHU was born in 1968. She is an professor of Hunan University of Technology, Zhuzhou, china. Her research interests are in intelligent information processing and text classification.

Zhiqiang WEN was born in 1973.He is an associate professor of Hunan University of Technology, Zhuzhou, China. His research interests are in computer vision, object detecting, and object tracking .