Analysis and Simulation on Recognition Algorithm for Dynamic Facial Images

Hong-yan Zhang[1], Sheng-wei Fan[2]
1. School of Software, Henan Institute of Engineering, ZhengZhou, 451191, China
2. CHINA TOBACCO HENAN INDUSTRIAL CO., LTD., ZhengZhou, 450003, China
E-mail: zhanghongyanhong@163.com

***Abstract*—In order to realize the recognition for dynamic facial images, the paper builds a dynamic matching model. First, this paper introduces a dynamic feature extraction algorithm of feature constraint optimization which can effectively extract the 2D dynamic facial features. Then, the paper analyzes the neurons mathematical model comprehensive expressing neurons operational mechanism and applies the model to dynamic facial feature expression. Finally, we study the learning rules how BP algorithm directs neural network, establish a corresponding mathematical model, and then use facial feature dynamic rules making dynamic features to be learned quickly and complete the recognition by the error compensation. The experiment results show that the model has a higher precision of geometric feature extraction for dynamic two-dimensional face images and better recognizes characteristic with error within 0.035mm which meets the requirements of stable, reliable, high precision and anti-interference ability etc.**

***Index Terms*—face recognition; genetic algorithm; neural network; supervised learning**

I. INTRODUCTION

In the current field of biometrics, the traditional identification methods are developing to the advanced recognition technology. The information dynamic identification becomes an important trend in the development of the field of biometric identification. Relying on its unique advantages face recognition technology becomes an important part of biometric [1]. The face recognition system has been widely applied in the field of public safety, personal security, financial security, so it has significant practical value and broad prospects for development. The following types of identification methods have formed in recent research area.

The identification based on the geometric characteristics of the Gaussian characteristics of face images listed in literature [2] is one of the most common face recognition methods. It collects face images, shape of the organs and their geometric relationship to create feature vectors, and then follow the Euclidean distance and angle parameters to realize feature matching, in order to complete the recognition [3]. However, due to the lack of a unified standard of feature extraction, it is difficult to extract stable characteristic parameter from the images, especially when the target feature is blocked [4]. In addition, in the case of large changes in the expression or posture, robustness of the system decreases. Face recognition method based on a dynamic template proposed in document [5] mainly includes template matching and equal intensity lines matching method. The disadvantage of this approach is that it dependents on the template, resulting in the calculation process are more time-consuming and prone to delay which significantly limits the application [6]. Power line rule proposed in document [7] uses the equal intensity lines of multi-level gray-scale in grayscale image as feature to identify. The shortcoming is that only when the background of faces and hair are black, and the light is uniformed, can it obtain the equal intensity lines accord with real face [8]. Accordingly, the use of the method to realize the face recognition is greatly restricted in certain extent. Subspace based face recognition method presented in literature [9] establishes a face recognition system by Facial subspace method, which has excellent ability to identify local feature. However, for dynamic facial feature, the robustness of the algorithm decreases significantly, resulting in failure. Pentland et al. [10] establish an intrinsic subspace in accordance with the difference of feature space of the eyes, nose and mouth, and combine with the intrinsic subspace theory has satisfied results. However, this approach presupposes that the shape of the basis is unchanged which reduces its applicability. Shan et al. [11] adopt the specific intrinsic subspace to complete accurate identification of human faces, which greatly improves the accuracy. But this method is too time-consuming to be applied in practice. Improve KL algorithm[12] and improved hidden Markov method (HMM) [13] expand face image along row/column, treat the high-dimensional vector formed as a random vector, and then use the K-L transform to obtain orthogonal K-L basis to achieve the recognition. However, due to the randomness of feature extraction process, K-L transform process does not converge making it impossible for the dynamic face recognition. Samaria and Fallside [14] extract feature from top to bottom, from left to right, and use this characteristic value as a parameter establishing a HMM model which has very good effect on static face

doi:10.4304/jcp.8.11.3004-3011

recognition. However, when the method is applied to dynamic face recognition, there exist lots of defects.

In this paper, we build a dynamic matching model. First, this paper introduces a dynamic feature extraction algorithm of feature constraint optimization which can effectively extract the 2D dynamic facial features. Then, the paper analyzes the neurons mathematical model comprehensive expressing neurons operational mechanism and applies the model to dynamic facial feature expression. Finally, we study the learning rules how BP algorithm directs neural network, establish a corresponding mathematical model, and then use facial feature dynamic rules making dynamic features to be learned quickly and complete the recognition by the error compensation. The experiment results show that the model has a higher precision of geometric feature extraction for dynamic two-dimensional face images and better recognizes characteristic with error within 0.035mm which meets the requirements of stable, reliable, high precision and anti-interference ability etc.

The rest of this paper is organized as follows：First, we analyze the proposed algorithm for dynamic facial feature extraction, and then analyze the dynamic studying process of facial recognition by using neural network algorithm and optimize the extraction process. Finally, we test the performance of the algorithm and give out the results.

II. Analysis on Dynamic Facial Feature Extraction Algorithm

A.Calibration for the Dynamic Facial Feature Points

Due to great changes in the dynamic characteristics and lack of constraints, this paper proposes a distribution model for facial feature points' extraction. The steps can be briefly described as follows: First calibrate feature points for the training set artificially, and then aligned all shapes in training set in a unify coordinates in order to facilitate statistical analysis. Suppose there are two similar shape x_1 and x_2, to get the minimum E when convert x_2 to $M(x_2)+t$, the value of rotation angle θ, scaling factor s and translation vector $t=\begin{bmatrix} t_x \\ t_y \end{bmatrix}$.

$$E=[x_1-(M(s,\theta)x_2+t)]^T W[x_1-(M(s,\theta)x_2+t)]$$

Wherein

$$\begin{cases} M(s,\theta)\begin{bmatrix} x_{jk} \\ y_{jk} \end{bmatrix}=\begin{pmatrix} (s\cos\theta)x_{jk}-(s\sin\theta)y_{jk} \\ (s\sin\theta)x_{jk}+(s\cos\theta)y_{jk} \end{pmatrix} \\ t=(t_x,t_{y,\ldots\ldots}t_x,t_y)^T \end{cases} \quad (1)$$

W is a diagonal matrix, wherein the matrix elements w_k is the value of the weight corresponding to each point.

$$w_k=(\sum_{i=1}^{n} Variance(k,l))^{-1}$$

Suppose Distance(k,l) represents the distance from k-th point to the 1st point, then Variance(k,l) describes the variance of Distance(k,l) in the training set.

For convenience of calculation, we assume that

$$\begin{cases} a_x=s\cos\theta \\ a_y=s\sin\theta \end{cases}$$

Then you can get the following linear equation

$$\begin{pmatrix} X_2 & -Y_2 & W & 0 \\ Y_2 & X_2 & 0 & W \\ Z & 0 & X_2 & Y_2 \\ 0 & Z & -Y_2 & X_2 \end{pmatrix}\begin{bmatrix} a_x \\ a_y \\ t_x \\ t_y \end{bmatrix}=\begin{bmatrix} X_1 \\ Y_1 \\ C_1 \\ C_2 \end{bmatrix} \quad (2)$$

Wherein

$$\begin{cases} X_i=\sum_{k=1}^{n} w_k x_{ik} \\ Y_i=\sum_{k=1}^{n} w_k y_{ik} \\ Z=\sum_{k=1}^{n} w_k(x_{2k}^2+y_{2k}^2) \\ W=\sum_{k=1}^{n} w_k \\ C_1=\sum_{k=1}^{n} w_k(x_{1k}x_{2k}+y_{1k}y_{2k}) \\ C_2=\sum_{k=1}^{n} w_k(y_{1k}x_{2k}-x_{1k}y_{2k}) \end{cases} \quad (3)$$

Through the equation above, the values of α_x, α_y, t_x and t_y can be calculated from which we can get s, θ and t to align shape x_2 with x_1. Aligning instance of images in the face database shown in Figure 1

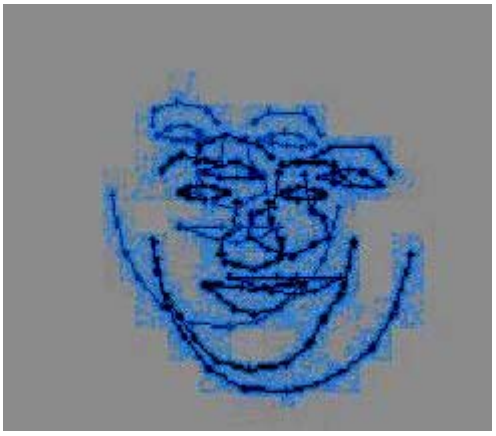

Figure 1 The Shape aligned result

Figure 2 shows the five positive photo alignment results in the training set. The blue dots indicate the distribution of the shape points after aligned and purple line represents the shape of the median.

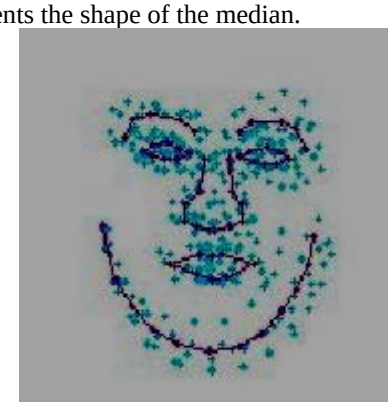

Figure 2 Training set image aligned

Using method mentioned above to align the shapes in training set into a unify coordinate, and then extract statistical information. Supposed N shapes have been aligned (each shape contains n manual calibration points), first find the median shape represented by the green line in Figure 2

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad (4)$$

It is very easy to use PCA to weed out redundant data, and the covariance matrix is ($2n \times 2n$):

$$S = \frac{1}{N}\sum_{i=1}^{N} dX_i dX_i^T$$

Wherein $dX_i = X_i - \bar{X}_i$.

Then decompose the singular value of matrix according to $Sp_k = \lambda_k p_k$, $p_k^T p_k = 1$, we can find the eigenvalues and eigenvectors of S.

Get the first t λ_k in $\lambda = [\lambda_1, \lambda_2, ...\lambda_{2n}]$:

$$\frac{\sum_{k=1}^{t}\lambda_k}{\sum_{k=1}^{2n}\lambda_k} \geq Explained - rate$$

(Explained-rate is a constant)

The positions of the feature points in the training set are interrelated, so a few variables can be used to represent most of the shape in the training set.

$$x = \bar{x} + Pb$$

Wherein $P = (p_1, p_2, ...p_t)$, $b = (b_1, b_2, ...b_t)^\tau$.

B.The Optimization Process of Dynamic Characteristics Constraints

Point distribution model gives a series of parameters such as the average of all the shapes of the training set, and the vector used to control the shape model. By a specific algorithm, all the shapes in the training set are aligned and principal component analysis is carried out to establish the model of shape.

In the searching process, Local Appearance Models are adopted. Local Appearance Models are used to describe the contour feature around manual calibration points of the training set. It assumes that the model is multivariate Gaussian distribution. For example, to the j-th manual calibration points in training images, we can first get its median contour the covariance matrix, and then search the optimal matching point in the new image according to Mahalanobis Distance.

$$f_j(g_s) = (g_s - \bar{g}_j)^t S_j^{-1}(g_s - \bar{g}_j) \qquad (5)$$

As can be seen from the above formula, an initialization position determines the speed and time of the search function. Therefore, a reasonable initial position of the image is possible to significantly reduce the matching time and improve efficiency of the algorithm

C.Acial Feature Extraction

Initial model positioning and search strategy of local feature points are the two key factors that affect the effect of dynamic face localization. To make face matching fast and accurate, the algorithm is made the following improvements in this paper:

Based on triangular features' automatic positioning method, the paper locates the facial feature with precise key feature.

The main process of the triangular positioning method used in this paper is as follows:

(1) First, in the region of the face image, the approximate location of the lips, (x_3, y_3) , is determined by PCA feature extraction method as the initial position of a point to determine the position of the face.

(2) This position as the lower left vertex of a rectangular frame, in accordance with the dimensions given by knowledge of a priori probabilities of the human face, we set the aspect ratio of the rectangular frame and its corresponding length (as small rectangles shown in Figure 3). Then smoothing graphics, I_{mean} the average gray value of pixels within the scope of the initial rectangle is calculated

Figure 3 The small rectangle in the face image

$$\begin{cases} E = \sum\limits_{\substack{x_0 \le x \le x_0 + w_0 \\ y_0 \le y < y_0 + h_0}} (x, y) \\ I_{mean(0)} = \dfrac{1}{w_0 h_0} E_0 \end{cases} \quad (6)$$

w_0 and h_0 respectively are the length and width of the initial rectangle

(3) In the case of unchanged the aspect ratio of the rectangular frame, we extend one pixel step along the top right. Find the minimum gradation value of the pixel point in the right side of the rectangle and the upper edge line in the image smooth processed. The minimum value is expressed as $I_{m.h}$.

(4) The calculate the difference between average grayscale value of the rectangular frame and the minimum gradation value of the right side of the rectangle and the upper side of the line, then compare it with statistical specific threshold D. If $I_{mean} - I_{m.h} < D$, the rectangular frame has not yet reached the rectangular area and should continue the rectangular extension. Otherwise, the edge of the rectangle has reached pupil.

(5) In accordance with the position of the pixel point with the minimum gray value to analyze the approximate area where the pupil is.

(6) In this region, Gaussian convolution is carried out to all pixels within a fixed rectangular area. The points with minimum gray value in the results are selected as the center of the pupil with coordinates of (x_1, y_1), (x_2, y_2).

In this way, the approximate location of lips (x_3, y_3) and pupil center position (x_1, y_1) (x_2, y_2), three points form an inverted triangle facial feature.(as shown in Figure4)

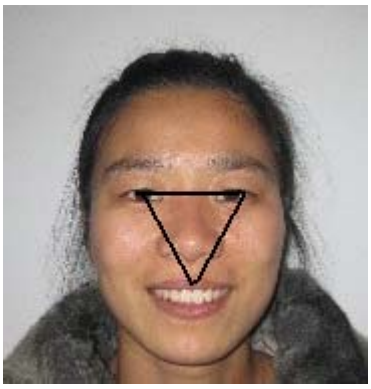

Figure 4 inverted triangle features

Experimental results show that the application of this characteristic can well initialize the template of ASM algorithm. In most cases, initialization carried out by triangular characterized model can obtain the satisfactory initial state.

After selecting the initial position, the average shape model of active shape model is initialized.

The coordinates of the left and right pupil in average model are $(x_1^{'}, y_1^{'})$, $(x_2^{'}, y_2^{'})$, lips center coordinate is $(x_3^{'}, y_3^{'})$ and $y_1^{'} = y_2^{'}$. Rotation angle θ and scale changes s can be calculated:

$$\theta = \arctan \frac{y_2 - y_1}{x_2 - x_1}$$

$$s = \frac{\sqrt{(y_3 - y_2 - y_1)^2 + (x_3 - x_2 - x_1)^2}}{\sqrt{(y_3^{'} - y_2^{'} - y_1^{'})^2 + (x_3^{'} - x_2^{'} - x_1^{'})^2}} \quad (7)$$

Angle rotate and scale the average model according to θ and s, and then calculate the translational coordinate values. In this way, the average model can be moved to corresponding position locating the inverted triangle points in the image by the same proportion as scaling and rotation (Figure 5).

Figure 5 the inverted triangle Positioning and initialization results

It can be inferred from the experimental results that even in a few cases that the triangulation is inaccurate, compared with the result of direct estimation the recognition result's deviation degree is very small. Therefore, the result for ASM average model initial positioning is ideal by proposed triangulation method.

III. Analysis on Dynamic Learning Process of Facial Features

A. Neural Network Analysis

The study of neural network, also called training, constantly get close to the target output via autonomously adjusting its internal connection weights of the threshold value according to the response to external signals. This adaptive learning ability is one of the very important characteristics of the neural network. In accordance with the learning styles, neural network can be divided into:

(1) Supervised learning: In this mode, comparing the expect output with actual output to get error from which can we adjust the interior weights of the threshold value in order to narrow the gap between actual output and expect one.

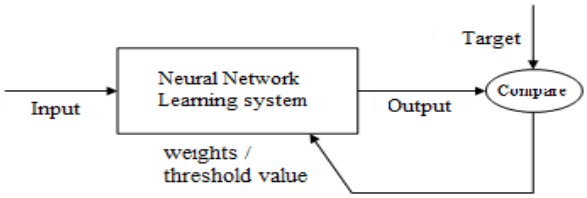

Figure 6 (a) Supervised learning principle diagram

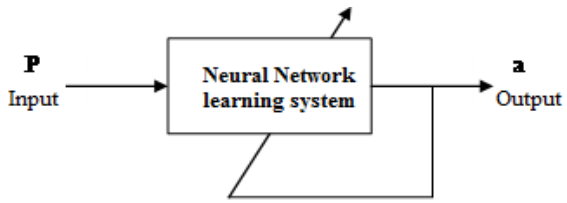

Figure6 (b) Unsupervised learning principle diagrams

(2) Unsupervised learning: In this initial network state, the connection weight is a small value. This method of learning has a certain memory capacity, which will produce a corresponding output when input similar stimulation.

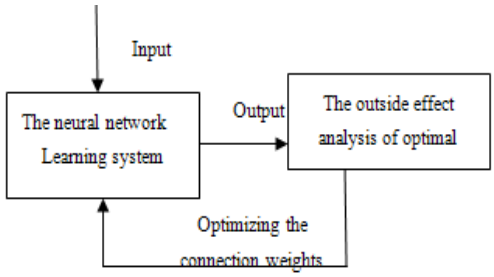

Figure7. Reinforcement learning principle diagram

(3) Strengthen learning: This learning method is between the two ways above. External stimuli can only give evaluation, but not exact correct answer. The learning system improves its connection weight by strengthening these stimuli.

B. The Model of Optimal Genetic Neural Network

As a guided learning rule, BP algorithm's basic idea is that the learning process consists of two processes including signal forward propagation and error back propagation. In forward propagation, input samples from the input layer transmit to the output layer via hidden layers. Based on the difference between the actual and the expected output of the output layer, the connection weights and bias coefficient between the layers are adjusted reversing from the output layer to the hidden layer then to the input layer. The process will be repeated until the actual output meet the accuracy requirements of the error or run the maximize algebra calculation.

The BP algorithm multiple sensors are the most widely used neural network model currently. Multilayer perception applications generally use three-layer neural network model: the input layer, single hidden layer and output layer. The three layers model of BP is described as Figure8:

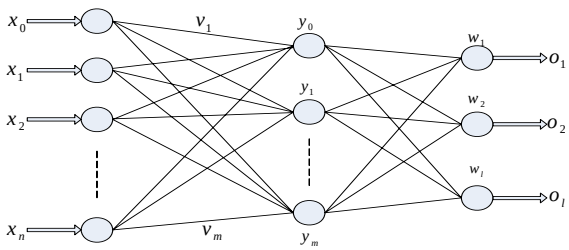

Figure 8. The model of BP network

In the model, $X=(x_1,x_2,\cdots,x_n)^T$ is the input vector, $Y=(y_1,y_2,\cdots,y_m)^T$ represents the input vector of hidden layer, $O=(o_1,o_2,\cdots o_l)^T$ is the output vector of output layer, $d=(d_1,d_2,\cdots d_l)^T$ is the expected vector. The weight matrix from input layer to hidden layer is expressed as V, $V=(v_1,v_2,\cdots,v_m)^T$. The weight from hidden layer to output layer is described with W, $W=(w_1,w_2,\cdots,w_i)^T$.

The input-output relationships between the layers of the BP neural network are as follows:

To hidden layer:

$$y_j = f(net_j) \qquad \text{j=1,2,}\cdots\text{,m}$$

$$net_j = \sum_{i=0}^{n} v_{ij}x_i \qquad \text{j=1,2,}\cdots\text{,m}$$

To output layer:

$$o_k = f(net_k) \qquad k=1,2,\cdots,l$$

$$net_k = \sum_{j=0}^{m} w_{jk}y_j \qquad k=1,2,\cdots,l$$

$f(x)$ is unipolar Sigmoid function:

$$f(x)=\frac{1}{1+e^{-x}}$$

$f(x)$ is continuous and differentiable and

$$f'(x)=f(x)\left[1-f(x)\right] \qquad (8)$$

When the output is not equal to the expected value, the output error E can be expressed as follow:

$$E=\frac{1}{2}(d-o)^2=\frac{1}{2}\sum_{k=1}^{l}(d_k-o_k)^2$$

Expanding the above error definition to the hidden layer, equation (9) can be obtained:

$$E=\frac{1}{2}\sum_{k=1}^{l}\left[d_k-f\left(\sum_{j=0}^{m}w_{jk}y_j\right)\right]^2 \qquad (9)$$

Further expanding the equation (9) to the input layer, (10) can be obtained

$$E=\frac{1}{2}\sum_{k=1}^{l}\left\{d_k-f\left[\sum_{j=0}^{m}w_{jk}f\left(\sum_{i=0}^{n}v_{ij}x_i\right)\right]\right\}^2 \qquad (10)$$

From equation above, we can see that the network error is the function of V and W. Therefore error adjusting would change these two parameters, further more adjust the whole system. Adjusting weights value may continuously decrease the error; therefore the fastest adjustment method is to make the adjustment value of weights proportional to descent of error gradient, that is to say:

$$\Delta w_{jk}=-\eta\frac{\partial E}{\partial w_{jk}} \qquad j=0,1,2,\cdots,m\ ;\quad k=1,2,\cdots,l$$

$$\Delta v_{ij}=-\eta\frac{\partial E}{\partial v_{ij}} \qquad i=0,1,2,\cdots,n\ ;\quad j=1,2,\cdots,m$$

The negative sign indicates descent gradient, constant $\eta\in(0,1)$ represents the proportion coefficient reflection learning velocity in training. BP algorithm commonly refers to the error gradient descent method to ensure the identification of facial dynamic feature.

IV. Algorithm Verification

A.The Establishment of Facial Image Database

In this paper, we adopt ORL (Olivetti Research Laboratory) facial database consisting of 400 grayscale images with 112*92 pixels. Each face has varying degrees of change, such as facial expressions, face pose, and wearing or not wearing glasses, as shown in Figure9. ORL has been widely used today, so we choose the database as a test object. We randomly select four images of 10 individuals in the database as training images, the remaining six images of each person as test images. Preprocessing all images, the training image will be obtained. PCA feature extraction in order to reduce the processing data compute, and all the images are mapped to the feature space got, then the data obtained is actually used for face recognition. In this article we select eigenvectors corresponding to the 20 largest eigenvalue in PCA feature space to composite characteristics space, so we get the input matrix of quantum neural network is 20 * 40.

Figure9. All images of individual in the ORL face database

B. Improved Genetic Algorithm is used to Optimize the Quantum BP Neural Network Face Recognition

The genetic algorithm is applied to the threshold of the initial weight optimization of quantum BP neural network face recognition; simulation results obtained are as follows:

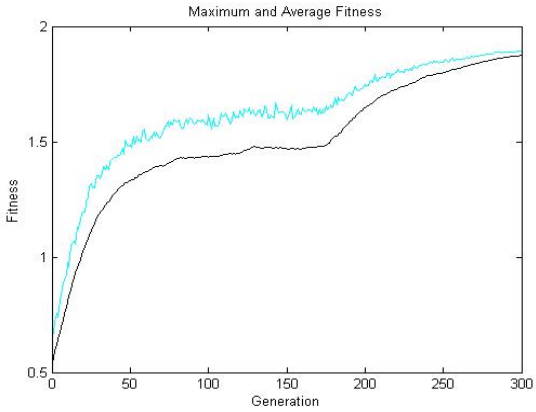

Figure10. Improved genetic algorithm to optimize the initial weights threshold of neural network

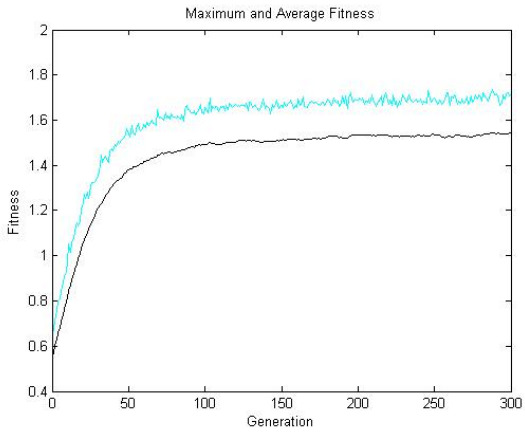

Figure11. Basic genetic algorithm to optimize the initial weights threshold of neural network

The curve below represents the average recognition rate of the population, while the curve above represents the maximum recognition rate of the population. From the figure, we can see that the 150 generation improved genetic algorithm and the unimproved basic genetic algorithm does not make any difference. This is due to that the variation comes into play only after a loci fixed more than 150 generations. On the other hand, the variability of the genetic algorithm is to solve the problem that basic genetic algorithm is easy to fall into local extreme and the evolutionary algebra gradually increasing is not sensitive to the fitness function, so when convergence curve gradually becomes smooth, it is impossible to have outstanding individual for basic genetic algorithm in150 generation. But through fixing long-term unchanged gene locus, the improved genetic algorithm shrinks the space genetic manipulation required, which is more for the genetic manipulation.

TABLE 1

THE COMPARISON OF BASIC AND IMPROVED GENETIC ALGORITHM

Gene algebraic	0	75	150	225	300
Fitness of basic genetic algorithm	0.53	1.61	1.63	1.63	1.63
Fitness of improved gen algorithm	0.52	1.63	1.64	1.76	1.92
Number of loci marked	0	56	72	85	101
Number of loci fixed	0	0	0	45	68

In the table, we can see that the maximum fitness and average fitness of the entire population has been greatly improved with the increase in the fixed loci.

V. CONCLUSIONS

The paper analyzes the improved genetic neural network and then rationally optimizes the feature extraction process in the facial recognition, finally face uses grayscale images providing by ORL face database system to achieve simulation of improved genetic neural network algorithm. It is proved that the proposed improved genetic algorithm is particularly suitable for populations of large individual's length. With the increasingly high demand for face recognition applications, the recognition of unconstrained facial images in random state will be the main research contents in the future.

REFERENCES

[1] D. Voth, “Face recognition technology”, *Intelligent Systems, IEEE*, vol.18, NO. 3, 2003, pp. 4-7.

[2] M. H. Yang, D. J. Kriegman, N. Ahuja, “Detecting faces in images: A survey”, *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 2002, vol. 24, NO. 1, pp. 34-58.

[3] R. Brunelli, T. Poggio, "Face recognition: Features versus templates", *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, vol. 15, NO. 10, 1993, pp. 1042-1052.

[4] R. Chellappa, C. L. Wilson, S. Sirohey, "Human and machine recognition of faces: A survey", *Proceedings of the IEEE*, vol. 83, NO.5, 1995, pp. 705-741.

[5] G. Zhao, M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, NO.6, 2007, pp. 915-928.

[6] A. F. Bobick, J. W. Davis, "The recognition of human movement using temporal templates", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, NO. 3, 2001, pp. 257-267.

[7] W. Zhao , R. Chellappa , P. J. Phillips , A. Rosenfeld. "Face recognition: A literature survey", *Acm Computing Surveys (CSUR)*, 2003, 35(4): 399-458.

[8] M. A. Grudin, "On internal representations in face recognition systems", *Pattern recognition*, vol. 33, NO.7, 2000, pp. 1161-1177.

[9] R. L. Hsu, M. Abdel-Mottaleb, A. K. Jain, "Face detection in color images", *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 2002, vol. 24, NO.5, pp. 696-706.

[10] A. P. Pentland, "Fractal-based description of natural scenes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, 1984, pp. 661-674.

[11] C. Shan, S. Gong, P. W. McOwan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study", *Image and Vision Computing*, vol. 27, NO. 6, 2009, pp. 803-816.

[12] A. Levy, M. Lindenbaum, "Sequential Karhunen-Loeve basis extraction and its application to images", *International Conference on Image Processing*, IEEE, vol. 2, 1998, pp. 456-460.

[13] M. Bicego, U. Castellani, V. Murino, "Using Hidden Markov Models and wavelets for face recognition", *12th International Conference on Image Analysis and Processing*, 2003, pp. 52-56.

[14] F. Samaria, F. Fallside, "Face identification and feature extraction using hidden markov models", *Olivetti Research Limited*, 1993.