IOP Conf. Series: Materials Science and Engineering **719** (2020) 012074 doi:10.1088/1757-899X/719/1/012074

Face Recognition Method Based on Probabilistic Neural Network Optimizing Two-Dimensional Subspace Analysis

Haiyan Zhang[1,a] and Fenqi Qiao [2,b]
[1] University of Chinese Academy of Sciences, Beijing 100049, China
[2] School of Mechanical Engineering and Automation, Beihang University, No.37 Xueyuan Road, Haidian District, Beijing 10091,China
[a] 42054035@qq.com, [b] 15071005@buaa.edu.cn

Abstract. If there is noise in the original image of face recognition, the efficiency of face recognition will be affected. In this paper, a face recognition method based on probabilistic neural network optimizing two-dimensional subspace analysis was proposed. Firstly, discrete wavelet variation was used to preprocess the image, and then two-dimensional linear discriminant analysis was used for feature extraction. Finally, the probabilistic neural network was used to complete the face classification. According to the results of experiments conducted on ORL and Fei general face database and the database collected independently, the recognition rate can also be as high as 98.9% when noise is added, and compared with several new identification methods, this method can achieve better identification performance.

1. Introduction

Based on the obtained face sample database, face recognition recognizes or verify one or more faces from static or dynamic scenes with image processing and pattern recognition technology. Face recognition technology is extensively and widely used in important state organs and social security field. Face recognition technology can also be applied to video retrieval of multimedia databases and multimedia production. In recent years, facial expression information extraction based on face recognition technology became possible. It can be used to improve the human-computer interaction mode, thereby providing more humanized experience. Intelligent computer system face detection is the basis of automatic face recognition technology, affecting the accuracy and speed of subsequent face recognition, therefore, it is an important issue that must be solved in practical use. Only by detecting the face quickly and effectively can the recognition be instant and accurate.

As early as 1965, Chan and Bledsoe [1] published a technical report on Automated Face Recognition (AFR) on Panoramic Research Inc., which is the earliest academic paper on face recognition. It has been a history of more than 60 years from the discovery of face recognition technology to the application of face recognition technology. In recent years, more and more people have begun to study face recognition technology, therefore, the technology and algorithms are becoming more and more mature and the success rate of identifying faces and matching faces is getting higher and higher. At present, many Internet enterprises are studying face recognition technology and have developed many products. Researchers have proposed a number of face recognition algorithms to improve face recognition performance, such as support vector machine (SVM) [2], Hidden Markov Model (HMM) [3], Probability Method (Bayesian Network) [4] and Artificial Neural Network [5]. Among them, the artificial neural network is based on shape detection and classification. Among them, artificial neural network has attracted extensive attention because of its high operability in shape detection and classification.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

Face recognition methods based on 2D images mainly consists of three categories: global [6], local [7] and hybrid methods [8].

Local methods recognize faces mainly through the specific features of faces (corners, mouth, nose, etc.), so this method needs prior knowledge of image [9].

The global method uses the entire face surface as a source of information without considering local features [10].

The hybrid method combines the advantages of global and local methods and recognizes faces [11] by combining geometric feature detection (or structure) with local appearance feature extraction.

Compared with the one-dimensional method that one-dimensional array (vector) represents a surface, this paper focuses on a two-dimensional feature extraction method (matrix) while preserving the original shape of the face, which is mainly based on the loss of matrix vectors. The loss is geometric and temporal information related to the image pixel.

2. System Design

2.1. Preprocessing with DWT

Discrete Wavelet Transformation [12] is a signal processing tool, which is widely used for feature extraction, the compression and denoising, etc. At present, Discrete Wavelet Transformation has been applied to all kinds of face recognition studies. The main advantage of the Discrete Wavelet Transformation lies in the time scale position of the Fourier transform, according to the literature [12], DWT can be implemented with a filter group, which includes a low-pass filter and one high-pass filter.

2.2. Extracting features with 2DLDA

A two-dimensional linear discriminant analysis (2DLDA) method was proposed in literature [13]. First, the image is projected as a vector of m-dimension features.

$$Y_i = A_j X \tag{1}$$

Where, A_j is an m×n matrix, and Y is an m-dimension feature vector of the projected image A.

Suppose L denotes the class number, M denotes the total number of training images, and the training images are represented by a matrix m×n A_j (j=1 , ….M), $\overline{A}_j(i=1,....M)$ denotes the average of all classifications, and N_i denotes the samples of each class.

The optimal vector projection is orthogonal series, which can maximize the ratio of the dispersive matrix determinants in projected images [14].

$$J_{FLD}(W_{opt}) = \arg\max_{W} \frac{\left|W^T S_b W\right|}{\left|W^T S_W W\right|} \tag{2}$$

Suppose P_b=*trace* (S_b). P_W=*trace* (S_W), where, S_b is between the classes of scattered matrices, S_w is within the class of scattered matrices, and:

$$S_b = \sum\nolimits_{i=1}^{N} N_i(\overline{Y_i} - \overline{Y})((\overline{Y_i} - \overline{Y}))^T = \sum_{i=1}^{L} N_i \left[\overline{A_i} - \overline{A}X\right]\left[(\overline{A_i} - \overline{A}X)\right]^T \tag{3}$$

$$S_W = \sum\nolimits_{i=1}^{L} \sum\nolimits_{yk \in P_i} (\overline{Y_k} - \overline{Y})((\overline{Y_k} - \overline{Y}))^T = \sum_{i=1}^{L} \sum\nolimits_{yk \in P_i} \left[\overline{A_k} - \overline{A}X\right]\left[\overline{A_k} - \overline{A}X\right]^T \tag{4}$$

Then:

$$trace(S_b) = X^T \left[\sum\nolimits_{i=1}^{L} N_i \left(\overline{A_i} - \overline{A}\right)^T (\overline{A_i} - \overline{A})\right] X = X^T S_b X \tag{5}$$

$$trace(S_W) = X^T \left[\sum_{i=1}^{L} \sum_{yk \in P_i} \left(\overline{A_i} - \overline{A} \right)^T (\overline{A_i} - \overline{A}) \right] X = X^T S_b X \tag{6}$$

The criterion can be expressed as

$$J(X) = \frac{X^T S_W X}{X^T S_b X} \tag{7}$$

In the formula, X is a column vector unit.

The unit vector *X* maximization *J*(*X*) is called the projection axis. When the X_{OPT} is maximized, the optimal projection can be obtained by the following equation:

$$X_{OPT} = \arg\max_X J(X) \tag{8}$$

If S_W is reversible, the optimal solution is suitable for solving generalized characteristic number problems.

$$S_b X_{opt} = \lambda S_W X_{opt} \tag{9}$$

Where, λ is the maximum eigenvalue of $S_{\omega}^{-1} S_b$

In general, it is not enough to have only one optimal projection axis. Usually, it is necessary to follow the following constraints to select a set of projections. x_1, x_2, ……x_d, $\{x_1, x_2, \ldots\ldots x_d\} = \arg\max_X J(X)$ where, $X_i^T Xj = 0, i \neq j, i, j = 1, 2, \ldots\ldots d$.

In fact, the optimal projection axis $x_1, x_2, \ldots\ldots x_d$ is corresponding to the best eigenvalue "d", allowing the creation of a new eigenvector of the projected orthogonal matrix *X* $S_w^{-1} S_b$, which is an n×d matrix: $X=[x_1,x_2,\ldots\ldots x_d]$

Optimum projection 2DLDA x_1, x_2 ,……x_d is used to extract image features, and the following formula can be used:

$$y_k = AX_k, k = 1, 2, \ldots .d \tag{10}$$

A set of eigenvectors can be obtained from the formula y_1, y_2……y_d, which is represented by y= [y1, y2...yd] with a size of M×D of the image feature matrix A.

2.3. PNN Classification

The research shows that compared with classification based on Euclidean Distance, probabilistic neural network classification can improve the face recognition system. The theoretical basis of solving classification problems by network is based on Bayesian Decision Theory, and is realized in feed-forward network architecture. Bayesian decision theory is the principle concept of classification, which uses statistical inference to find the maximum expectation of minimum risk.

The raw data is classified into class C, and each class has m-dimensional observations, such as $X=[x_1,x_2,\ldots\ldots x_d]$. The following Bayesian decision rules classify raw data into class C: $h_i c_i P_i(X) > h_j c_j P_j(X) \forall j \neq i$, where h_c is the prior probability of class C; The loss function is the definite value; P_c is the probability density function of Class C.

The probability density function of each category can be expressed as follows:

$$P_i(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \sigma^d} \frac{1}{N_i} \sum_{j=1}^{N} \exp\left[\frac{-(x - x_{ij})^t (x - x_{ij})}{2\delta^2} \right] \tag{11}$$

Where, $P_i(x)$ denotes the dimension of training vectors, σ is a smoothing parameter, *d* is the dimension of training vectors, N_i denotes the total number of training vectors of class *i*, while x_{ij} is a neuron vector and *x* is a test vector.

PNN mathematical expressions are as follows:

$$a = rad\ \ bas(\|IW - x\|b) \tag{12}$$

$$y = compet(LW\alpha) \tag{13}$$

Where Q represents the number of input targets/pairs;

R represents the number of types of input data; IW represents the input weight.

The PNN structure consists of two layers [15]:

Layer One: Calculating the distance from the input vector to the input weight (IW) and producing a vector whose elements indicate the distance from input to IW.

Layer Two: Comparing various net output probability vectors, and finally selecting the maximum value of these probabilities in the competition transfer function of the second layer output, and display the number 1 for the class of the maximum class and the number 0 of the other class. The structure of the system is shown in Figure 1.

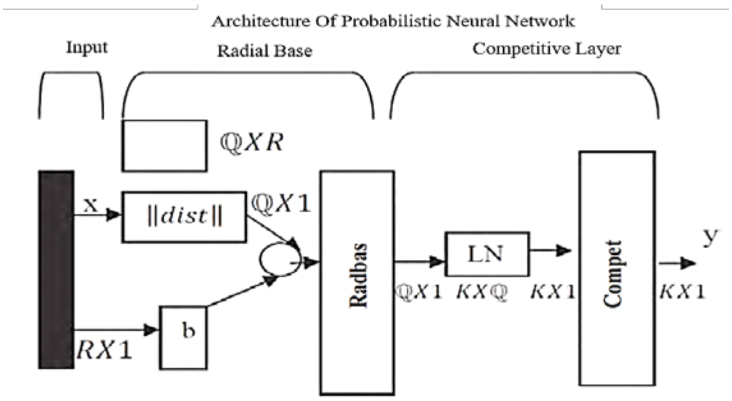

Figure 1. Architecture of probabilistic neural networks

Back-propagation neural network is used, and the probability of each hidden unit in the network can approximate any continuous nonlinear function, with a Gaussian function as the activation function:

$$rad\ \ bas = \exp\left[(-n^2)\right] \tag{14}$$

Finally, one or more larger values are selected as output unit, and the data points used for indication in the same category are contended to convert the output end of the summation unit into a function, namely

$$compet(n) = e_{i}\left[0000_{1}0......0_{i}\right], n(I) = MAX(n) \tag{15}$$

For the task of face recognition, the subspace of the face is obtained through 2DLDA, which is defined as the orthogonal basis containing the most relevant information of the face. The vector is the feature vector of the distributed covariance matrix. The recognition process is realized by forwarding the feature matrix to the constructed classifier and discriminator, where the feature proof must be converted into a vector before it is provided to the PNN classifier.

3. Experiment

In order to evaluate and test the described in the face recognition system discussed in this paper, three databases were selected: ORL face database [16], FEI [17] and self-constructed database. All experiments are carried out in MATLAB, which is installed on personal computers with the dual-processor T5870 of 2.03GHz and 2GBRAM.

3.1. Data Set

a) ORL Face Database: The changes of light on faces generated by the environment are uniform, and facial expressions may be different. It includes 40 subjects and each has 10 different images, with an angle from 20 degrees to the top of the head, tilting left or right. The image is a gray image with 92*112 pixels. An image example is shown in Figure 2.

Figure 2. Four images of the same person in ORL face database

b) FEI database: The database contains 200 objects, 14 different images for each person, and the head is not tilted. The image is a gray scale image with 200*180 pixels. An image example is shown in Figure 3.

Figure 3. An example of someone's image in the FEI database

c) Self-constructed database: the capture device (camera) is used to collect face images at different times to form a database. The database selects 100 different objects, each with N=10 images, and each image is shot from a different angle, namely, 1000 face images in JPG format, an image example is shown in Figure 4.

Figure 4. Example image in self-constructed database

3.2. Effect of Parameters on Recognition Performance

ORL database was used in the experiment to study the effect change of characteristic number on recognition rate. In this experiment, Euclidean distance was used, and the recognition result is shown in figure 5.

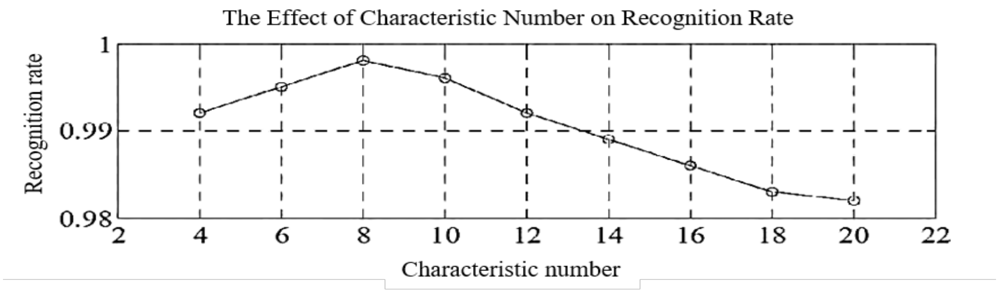

Figure 5. The Effect of Characteristic Number on Recognition Rate

It can be seen from Figure 5 that the method mentioned in this paper has the best recognition rate based on the seventh clean face (K=8) on the ORL, which can be as high as 99.8%. It can be seen from the fourteenth feature vector that the recognition rate fell to 98.9%. There is a similar trend in the FEI database that if the characteristic number is too large, it will cause over-fitting, resulting in the reduction of recognition rate.

To get the best recognition rate, the λ variance should be as small as possible; however, the characteristic number should not be too large. Therefore, it is necessary to find an optimal parameter setting. The influence curve of characteristic number changes on λ variance is shown in Figure 6.

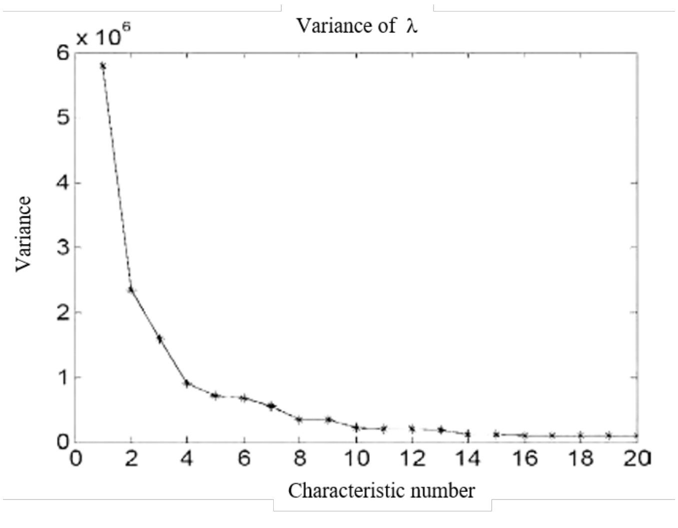

Figure 6. The effect of characteristic number on λ variance

It can be seen from Fig. 6 that as the characteristic number increases, the λ variance decreases. When the characteristic number is 8, the λ variance tends to be stable. It is also found that when the characteristic number is 8, the best recognition rate can be realized with the method discussed in this paper. Therefore, characteristic number in all subsequent experiments was set to 8.

The best recognition results of this method in various database are shown in Table 1.

Table 1. The best face recognition rate with the method in this paper (%)

Database	ORL	FEI	Self-constructed database
Recognition rate	99.8	99.5	98.5

It can be seen from Table 1 that the recognition rate of this method in ORL and FEI can be as high as 99.8% and 99.5%, respectively, and the recognition rate can be as high as 98.5% in the self-constructed database, which verifies the effectiveness of the proposed method.

3.3. Comparison and Analysis

In order to better show the excellence of this method, an experiment is conducted to test the noise and noise-free images separately. Three kinds of noise were used in the experiment: 1) salt and pepper noise with noise density a=0.06; 2) Gaussian noise, noise average m=0, variance v=0.04; 3) Gaussian noise, noise average m=0, variance v=0.06. The image example of three noises are shown in Fig. 7.

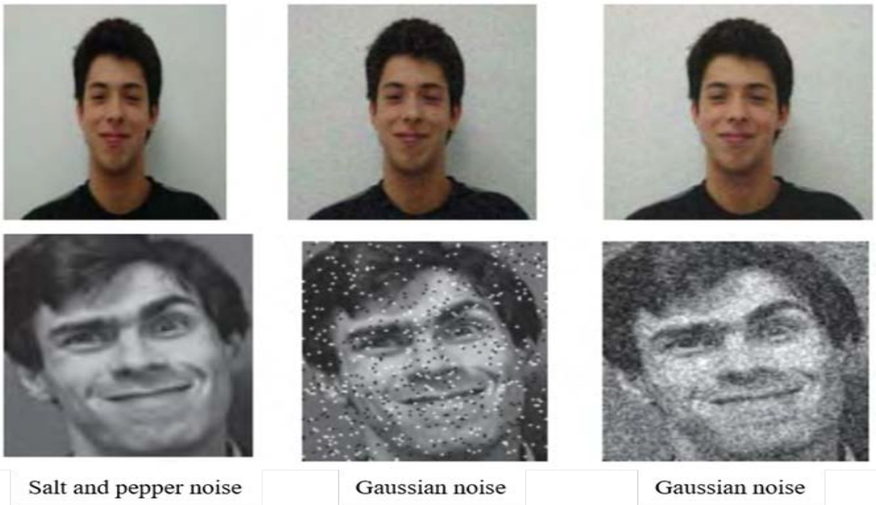

Figure 7. Image example with noise

The first half of the images of each object in the database were selected for training and the rest were used for testing. And the recognition result of the method in this paper is compared with that of several new methods, including the method based on HMM and SVD coefficients (HMM-SVDC) [3], tensor local Fisher discriminant analysis (TL-FLDA) [7], kernel fusion multiple descriptors-based multi-scale local phase quantization method (KFMD-MLPQ) [16], 2DPCA method based on wavelet transformation and SVM classification (WT-SVM-2DPCA) [10]. The recognition rates of each method with noise and without noise respectively are shown in Table 2 and Table 3. Table 4 shows the execution time of each method.

Table 2. Face recognition rate of each method without noise (%)

Method	ORL	FEI	Self-constructed database
HMM-SVDC	98.6	98.7	97.4
TL-FLDA	99.2	99.0	97.9
KFMD-MLPQ	99.4	99.2	99.2
WT-SVM-2DPCA	99.6	99.2	98.3
Method in this paper	99.8	99.5	98.5

Table 3. Face recognition rate of each method with noise (%)

Method	ORL	FEI	Self-constructed database
HMM-SVDC	96.5	96.1	94.5
TL-FLDA	97.0	96.5	95.1
KFMD-MLPQ	97.5	97.2	95.5
WT-SVM-2DPCA	97.8	97.6	96.4
Method in this paper	98.9	98.3	97.6

Table 4. Total time of each method (seconds)

Method	ORL	FEI	Self-constructed database
HMM-SVDC	25.593	30.459	14.582
TL-FLDA	21.322	28.347	12.495
KFMD-MLPQ	19.549	27.190	11.587
WT-SVM-2DPCA	17.480	25.039	9.733
Method in this paper	18.292	26.520	10.304

It can be seen from Table 2 that in most cases, the recognition rate of the method in this paper in three databases is higher than that of any other method. However, the recognition rate is not obviously improved. It can be seen from Table 3 that for images with noise, the recognition rate of this method is obviously higher than that of any other methods, indicating that the proposed method has better performance when processing images with noise. It can be seen from Table 4 that the execution time of this method is slightly higher than that of WT-SVM-2DPCA method. It is mainly because the WT-SVM-2DPCA method extract features with 2DPCA, which is quicker than the 2DLDA of this method. Compared with other methods, the execution time of this method is the shortest, indicating that this method excels in computational cost.

In summary, the following conclusion can be drawn that the method used in this paper reduces the computation time of the two-dimensional subspace method training and improves the recognition rate, indicating the excellence of this proposed method.

4. Conclusion

In order to solve the problem of face recognition with strong noise image, this paper puts forward a grid method based on DWT-2DLDA and probabilistic neural classifier. According to the experimental results, the method used in this paper has better recognition performance than other methods. It has been verified through experiments that DWT can be used as a preprocessing technology. With the help of this technology, the system mentioned in this paper has advantages of better recognition rate, high computing speed and reducing PNN memory calculation. Since this method can show better robust performance in this kind of uncontrolled environment, it will be applied to video monitoring in the future

It has been 60 years since the face recognition technology was first put forward. The face recognition technology becomes more and more mature and various algorithms come out continuously. Although great achievements have been gained, since different algorithms require different implementation environments, there are still so many difficulties in facial recognition technology that need to be solved, including face detection and key point location under complex conditions, illumination change, facial gestures, facial expressions, facial shade, changes in a person's age, large-scale face recognition problem, lack of samples, a large amount of data learning problems, collection equipment for face information and so on.

With the continuous development of technology, face recognition technology will develop in the following directions in the future: face recognition research based on multi-data and multi-method integration; feature acquisition research with dynamic tracking; face nonlinear modeling research; 3D modeling research; automatic face recognition research. It is believed that with the continuous advancement of related disciplines such as pattern recognition, computer vision, image processing and machine learning, the face recognition technology will be continuously improved and optimized. Face recognition technology will be more widely applied to daily life more conveniently and more safely.

5. References

[1] BLEDSOE W W. The model method in facial recognition [J]. Panoramic Research Inc Palo Alto, 1964(1): 55-64.

[2] Wang Yang, Yan Yunyang, and Wang Hongyuan. Face Recognition Based on Residual Space and SVM with Bidirectional Two Dimensions PCA [J]. Computer Science, 2012, 39(12): 268-271.

[3] Anand C, Lawrance R. Algorithm for Face Recognition Using HIM and SVD Coefficients[J]. Artificial Intelligent Systems and Machine Learning, 2013, 5(3):125-130.
[4] Zhang Ruyan, Wang Shitong. Multi-classification Recogition Method Applied to Facial Image on Distribution Characteristic of Heavy-tailed Noise [J]. Journal of Electronics and Information Technology, 2012, 34 (12): 523-528.
[5] Whither Biometrics Committee. Biometric recognition: challenges and opportunities [M]. National Academies Press, 2010.
[6] Chan C H, Tahir M A, Kittler J, et al. Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors[J].Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2013, 35(5): 1164-1177.
[7] Zheng Jianwei, Wang Wanliang, Yao Xiaomin et al. Face Recognition Using Tensor Local Fisher Discriminant Analysis [J]. Acta Automatica Sinica, 2012, 38 (9): 1485-1495.
[8] Su Yu, Shan Shiguang, Chen Xilin et al. Integration of Global and Local Feature for Face Recognition [J]. Journal of Software, 2010, 21 (8): 1849-1862.
[9] Mashhoori A, Zolghadri Jahromi M. Block-wise two-directional 2DPCA with ensemble leaming for face recognition [J]. Neurocomputing, 2013, 10(8):111-117.
[10] Xu W, Lee E J. Face Recognition Using Wavelets Transform and 2D PCA by SVM Classifier [J]. International Journal of Multimedia & Ubiquitous Engineering, 2014, 9(3): 243-249.
[11] Zhu Yulian, Peng Xing. Two-dimensional PCA Based on Random Sampling for Face Recogniton [J]. Journal of Chinese Mini-Micro Computer Systems, 2011, 32 (12): 2461-2465.
[12] Li D, Tang X, Pedrycz W. Face recognition using decimated redundant discrete wavelet transforms[J].Machine Vision and Applications, 2012, 23(2): 391-401.
[13] Zhang H, Jonathan Wu Q M, Chow T W S, et al. A two-dimensional Neighborhood Preserving Proiection for appearance-based face recognition [J]. Patterm Recognition, 2012, 45(5): 1866-1876.
[14] Guo Jinxin, Chen Wei. Face Recognition Based on HOG Multi-feature Fusion and Random Forest [J]. Computer Science, 2013, 40 (10): 279-282.
[15] Eslamloueyan R. Designing a hierarchical neural network based on fuzzy clustering for fault diagnosis of the Tennessee-Eastman process [J].Applied Soft Computing, 2011, 11(1): 1407-1415.
[16] Chan CH, Tahir M A, Kittler J, et al. Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors[J].Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2013; 35(5):1164-1177.
[17] Pandya J M, Rathod D, Jadav J J.A survey of face recognition approach [J].International Journal of Engineering Research and Applications (IJERA), 2013, 3(1):632-635.