Available online at www.sciencedirect.com

SCIENCE DIRECT®

Pattern Recognition Letters 25 (2004) 1351–1358

Pattern Recognition Letters

www.elsevier.com/locate/patrec

# Human face recognition based on multi-features using neural networks committee

Zhong-Qiu Zhao *, De-Shuang Huang, Bing-Yu Sun

*Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, P.O. Box 1130, Hefei, Anhui 230031, China*

Received 4 November 2003; received in revised form 8 March 2004
Available online 17 June 2004

**Abstract**

A novel face recognition method based on multi-features using a neural networks committee (NNC) machine is proposed in this paper. The committee consists of several independent neural networks trained by different image blocks of the original images in different feature domains. The final classification results represent a combined response of the individual networks. Then, we use the designed neural networks committee to perform human face data recognition. The experimental results show that the classification accuracy of our proposed NNC is much higher than that of single feature domain.
© 2004 Elsevier B.V. All rights reserved.

*Keywords:* Face recognition; Neural networks committee; Kernel methods; Classification; Feature domain

## 1. Introduction

It is well known that face recognition is a complex and difficult problem, but it is very important for surveillance and human–computer intelligent interaction, security, and telecommunications, etc. (Wechsler et al., 1998). To solve this problem, many approaches have been proposed, e.g., eigen-faces, geometrical features, template matching, graph matching as well as neural network and probabilistic method, etc. (Turk and Petland, 1991; Brunelli and Poggio, 1993; Pessoa and Leitao, 1999).

Fig. 1 shows a typical architecture for a face recognition system based on a single neural network. First, some significant features are extracted in order to reduce data dimension (and hence computational burden). Then, the recognition system is performed by the neural network (NN).

In order to achieve a high accuracy recognition rate, the choice of feature extractor is very crucial. At present, the existing feature extraction methods mostly include the interest operator (IO) (Wang et al., 1998), the principal component analysis (PCA) method (Turk and Petland, 1991; Er et al., 2002; Haykin, 1996), the Fisher's linear discriminant (FLD) technique (Belhumeur et al., 1997; Swet and Weng, 1996), and the Kernel Fisher linear discriminant (KFLD) method (Yang, 2002), etc. Among these feature extraction methods

* Corresponding author.
*E-mail address:* zqzhao@iim.ac.cn (Z.-Q. Zhao).

0167-8655/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2004.05.008

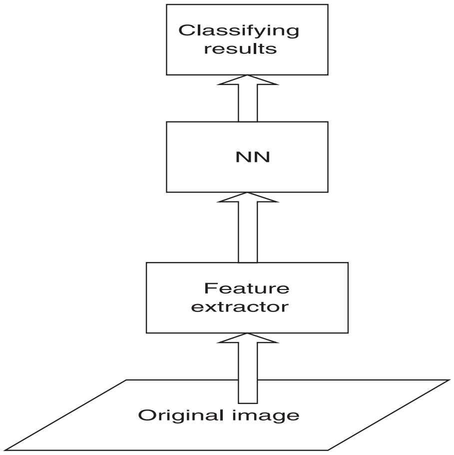


Fig. 1. System architecture for single NN.

mentioned above, the KFLD method is best for facial recognition according to the classification accuracy obtained (Yang, 2002).

Recently, a pattern recognition method based on neural networks committee machines (NNCM) has been proposed. In literature (Caleanu, 2000), the NNCM is employed to implement facial recognition. The original images are divided into several blocks, each of which is learned by a neural network module. However, in literature (Caleanu, 2000), only one feature extraction method, i.e., the interest operator, is used. Thus the recognition accuracy is relatively low.

According to the capability of the neural networks committee, the combination of an ensemble of classifiers has been proposed to achieve higher performance in comparison with the best performance achievable by employing a single classifier (Kittler et al., 1998). So, if the classification results in different feature domains are combined to achieve the final classification result, the final recognition accuracy must be higher than that of the results in the best feature domain.

In this paper, four different feature domains have been used for extracting features from input images, including the interest operator (IO), the principal component analysis (PCA), the Fisher's linear discriminant (FLD) and the Kernel Fisher linear discriminant (KFLD).

In addition, in our experiments, radial basis function neural networks (RBFNN) are adopted to implement the committee members. And all the experiments are conducted by averaging 10 experimental results, so they are authentic.

This paper is organized as follows. Section 2 presents four face feature domains; Section 3 introduces radial basis function neural networks (RBFNN) and neural networks committee machines; Section 4 presents several experimental results on the ORL database of faces; finally, several conclusions are given in Section 5.

## 2. Face feature domains

### 2.1. Interest operator (IO)

A feature extraction technique referred to as the interest operator (Wang et al., 1998) is used to find the directional variances in the horizontal, vertical, and both diagonal directions for each block in target images. These directional variances show the local activity in a block. After obtaining these directional variance features, we present them to a neural-network-based classifier.

The equations for calculating the block variance, represented as $\sigma$, and the directional variances in horizontal, vertical, diagonal-135, and diagonal-45 directions, respectively, represented as $\sigma_{\mathrm{H}}, \sigma_{\mathrm{V}}, \sigma_{\mathrm{D}_{135^\circ}}$ and $\sigma_{\mathrm{D}_{45^\circ}}$, are given below:

$$\mu = \frac{1}{N \times N} \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} p(y,x) \tag{1}$$

$$\sigma = \frac{1}{N \times N} \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} [p(y,x) - \mu]^2 \tag{2}$$

$$\sigma_{\mathrm{H}} = \frac{1}{N \times N} \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} [p(y,x) - p(y,x-1)]^2 \tag{3}$$

$$\sigma_{\mathrm{V}} = \frac{1}{N \times N} \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} [p(y,x) - p(y-1,x)]^2 \tag{4}$$

$$\sigma_{\mathrm{D}_{135^\circ}} = \frac{1}{N \times N} \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} [p(y,x) - p(y-1,x-1)]^2 \tag{5}$$

$$\sigma_{\mathrm{D}_{45^\circ}} = \frac{1}{N \times N} \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} [p(y, x-1) - p(y-1, x)]^2 \tag{6}$$

where $\{p(y,x), 0 \leqslant y, x \leqslant N-1\}$ represents the pixels in an $N \times N$ block (note that $N = 5$ in our experiments). We represent this variance feature set for an input image block as $\Theta$, where $\Theta = \{\sigma, \sigma_{\mathrm{H}}, \sigma_{\mathrm{V}}, \sigma_{\mathrm{D}_{135^\circ}}, \sigma_{\mathrm{D}_{45^\circ}}\}$.

### 2.2. Principal component analysis (PCA)

Let a face image $X_i$ be a two-dimensional $m \times m$ array of intensity values. Thus, an image may also be considered as a vector of $m^2$ dimension. Denote the training set of $n$ face images by $X = (X_1, X_2, \ldots, X_n) \subset \Re^{m^2 \times n}$, and we assume that each image belongs to one of $c$ classes. Thus, define the covariance matrix as follows:

$$\Gamma = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})(X_i - \overline{X})^{\mathrm{T}} = \Phi\Phi^{\mathrm{T}} \tag{7}$$

where $\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_n) \subset \Re^{m^2 \times n}$ and $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, the eigenvalues and eigenvectors of the covariance $\Gamma$ can be calculated. Let $U = (U_1, U_2, \ldots, U_r) \subset \Re^{m^2 \times n}$ $(r < n)$ be the $r$ eigenvectors corresponding to the $r$ largest eigenvalues. Thus, for a set of original face images $X \subset \Re^{m^2 \times n}$, their corresponding eigenface-based feature $Y \subset \Re^{r \times n}$ can be obtained by projecting $X$ into the eigenface space as follows:

$$Y = U^{\mathrm{T}} X \tag{8}$$

### 2.3. Fisher's linear discriminant (FLD)

In general, the FLD is used to find an optimal subspace for classification in which the ratio of the between-class scatter and the within-class scatter is maximized. Let the between-class scatter matrix be defined as

$$S_{\mathrm{B}} = \sum_{i=1}^{c} n^i (\overline{X}^i - \overline{X})(\overline{X}^i - \overline{X})^{\mathrm{T}} \tag{9}$$

and the within-class scatter matrix be defined as

$$S_{\mathrm{W}} = \sum_{i=1}^{c} \sum_{X_k \in n^i} (X_k - \overline{X}^i)(X_k - \overline{X}^i)^{\mathrm{T}} \tag{10}$$

where $\overline{X} = (1/n)\sum_{j=1}^{n} X_j$ is the mean image of the ensemble, and $\overline{X}^i = (1/n^i)\sum_{j=1}^{n^i} X_j^i$ is the mean image of the $i$th class, $n^i$ is the number of samples in the $i$th class, and $c$ the number of classes. As a result, the optimal subspace, $E_{\mathrm{optimal}}$ by the FLD can be determined as follows:

$$E_{\mathrm{optimal}} = \arg\max_{E} \frac{|E^{\mathrm{T}} S_{\mathrm{B}} E|}{|E^{\mathrm{T}} S_{\mathrm{W}} E|} = [c_1, c_2, \ldots, c_{c-1}] \tag{11}$$

where $[c_1, c_2, \ldots, c_{c-1}]$ is the set of generalized eigenvectors of $S_{\mathrm{B}}$ and $S_{\mathrm{W}}$ corresponding to the largest generalized eigenvalues $\lambda_i, i = 1, 2, \ldots, c-1$, i.e.,

$$S_{\mathrm{B}} E_i = \lambda_i S_{\mathrm{W}} E_i \quad i = 1, 2, \ldots, c-1 \tag{12}$$

Thus, the feature vector, $P$, for any query face images, $X$, in the most discriminant sense can be calculated as follows:

$$P = E_{\mathrm{optimal}}^{\mathrm{T}} U^{\mathrm{T}} X \tag{13}$$

### 2.4. Kernel Fisher linear discriminant (KFLD)

To derive the KFLD features, we first map the data $(X_1, X_2, \ldots, X_n)$ by $\Phi$, a non-linear mapping, into the feature space $F$. Then in $F$, the optimal subspace, $W_{\mathrm{optimal}}$, that we will find, can be determined as follows:

$$W_{\mathrm{optimal}} = \arg\max_{W} \frac{|W^{\mathrm{T}} S_{\mathrm{B}}^{\Phi} W|}{|W^{\mathrm{T}} S_{\mathrm{W}}^{\Phi} W|} = (w_1, w_2, \ldots, w_m) \tag{14}$$

where $S_{\mathrm{B}}^{\Phi}$ and $S_{\mathrm{W}}^{\Phi}$ are the corresponding between-class scatter and within-class scatter matrices in $F$, i.e.,

$$S_{\mathrm{B}}^{\Phi} = \sum_{i=1}^{c} n^i (\mu_i^{\Phi} - \mu^{\Phi})(\mu_i^{\Phi} - \mu^{\Phi})^{\mathrm{T}} \tag{15}$$

$$S_{\mathrm{W}}^{\Phi} = \sum_{i=1}^{c} \sum_{X_k \in n^i} (\Phi(X_k) - \mu_i^{\Phi})(\Phi(X_k) - \mu_i^{\Phi})^{\mathrm{T}} \tag{16}$$

with

$$\mu_i^{\Phi} = (1/n^i)\sum_{X_k \in n^i} \Phi(X_k),$$

$$\mu^{\Phi} = (1/n)\sum_{i=1}^{n} \Phi(X_i).$$

Each $w_i$ in Eq. (14) can be computed by solving the following generalized eigenvalue problems:

$$S_{\mathrm{B}}^{\Phi} w_i = \lambda_i S_{\mathrm{W}}^{\Phi} w_i \tag{17}$$

From the theory of reproducing kernels we know that any solution $w_i \in F$ must be in the spanning space of the mapped data, i.e., $w_i \in \mathrm{span}\{\Phi(X_1), \Phi(X_2), \ldots, \Phi(X_n)\}$, which can be written as

$$w_i = \sum_{j=1}^{n} \alpha_i^j \Phi(X_j) \tag{18}$$

Using this expansion, the numerator of Eq. (14) can be rewritten as:

$$w_i^{\mathrm{T}} S_{\mathrm{B}}^{\Phi} = \alpha_i^{\mathrm{T}} M \alpha_i \tag{19}$$

where

$$M = \sum_{i=1}^{c} (M_i - \overline{M})(M_i - \overline{M})^{\mathrm{T}} \tag{20}$$

$$(M_i)_j = \frac{1}{n^i} \sum_{X_k \in n^i} k(X_j, X_k) \tag{21}$$

$$\overline{M}_j = \frac{1}{n} \sum_{i=1}^{n} k(X_j, X_i) \tag{22}$$

$$k(X_j, X_i) = \Phi(X_j)\Phi(X_i) \tag{23}$$

$$\alpha_i = (\alpha_i^1, \alpha_i^2, \ldots, \alpha_i^n)^{\mathrm{T}} \tag{24}$$

Now, considering the denominator of Eq. (14), and using similar transformation, we have:

$$w_i^{\mathrm{T}} S_{\mathrm{W}}^{\Phi} w_i = \alpha_i^{\mathrm{T}} L \alpha_i \tag{25}$$

where $L = \sum_{j=1}^{c} K_j (I - 1_{n^j}) K_j^{\mathrm{T}}$; $K_j$ is an $n \times n^j$ matrix with $(K_j)_{nm} = k(X_n, X_m), X_m \in n^j$; $I$ is the identity matrix and $1_{n^j}$ the matrix with all entries $1/n^j$.

Combining Eq. (19) with Eq. (25), the optimal subspace, $W_{\mathrm{optimal}}$, can be determined by following two formula:

$$\alpha_{\mathrm{opt}} = \arg\max_{\alpha} \frac{|\alpha^{\mathrm{T}} M \alpha|}{|\alpha^{\mathrm{T}} L \alpha|} = (\alpha_1, \alpha_2, \ldots, \alpha_3) \tag{26}$$

and

$$\begin{aligned} W_{\mathrm{optimal}} &= (w_1, w_2, \ldots, w_m) \\ &= \left( \sum_{i=1}^{n} \alpha_1^i \Phi(X_i), \sum_{i=1}^{n} \alpha_2^i \Phi(X_i), \ldots, \sum_{i=1}^{n} \alpha_m^i \Phi(X_i) \right) \end{aligned} \tag{27}$$

To extract the features of a new pattern $X_j$ with the KFLD, we can simply project the mapped pattern $\Phi(X_j)$ onto this subspace, and the corresponding result can be described as follows:

$$\begin{aligned} \Phi(X_j) W_{\mathrm{optimal}} &= \left( \sum_{i=1}^{n} \alpha_1^i \Phi(X_i)\Phi(X_j), \right. \\ &\qquad \left. \sum_{i=1}^{n} \alpha_2^i \Phi(X_i)\Phi(X_j), \ldots, \sum_{i=1}^{n} \alpha_m^i \Phi(X_i)\Phi(X_j) \right) \\ &= \left( \sum_{i=1}^{n} \alpha_1^i k(X_i, X_j), \right. \\ &\qquad \left. \sum_{i=1}^{n} \alpha_2^i k(X_i, X_j), \ldots, \sum_{i=1}^{n} \alpha_m^i k(X_i, X_j) \right) \end{aligned} \tag{28}$$

## 3. Radial basis function neural networks committee machines

### *3.1. Radial basis function neural networks (RBFNN)*

A RBF neural network (RBFNN) is a universal approximator that is of the best approximation property, and its learning speed is very fast because of locally tuned neurons (Park and Wsandberg, 1991; Girosi and Poggio, 1990; Huang, 1999a; Huang, 1999b). So, RBFNNs have been widely used for function approximation and pattern recognition.

A RBFNN can be considered as a mapping: $\Re^r \rightarrow \Re^s$. Let $P \in \Re^r$ be the input vector and $C_i \in \Re^r (1 \leqslant i \leqslant u)$ be the prototype of the input vectors. The output of each RBF unit can be written as:

$$R_i(P) = R_i(\|P - C_i\|) \quad i = 1, \ldots, u \tag{29}$$

where $\|\cdot\|$ indicates the Euclidean norm on the input space. Usually, the Gaussian function is preferred among all possible radial basis function due to the fact that it is factorizable. Hence

$$R_i(P) = \exp\left(-\frac{\|P - C_i\|^2}{\sigma_i^2}\right) \tag{30}$$

where $\sigma_i$ is the width of the $i$th RBF unit. The $j$th output $y_j(P)$ of a RBFNN is

$$y_j(P) = \sum_{i=1}^{u} R_i(P) \times w(j, i) \tag{31}$$

where $w(j, i)$ is the weight of the $i$th receptive field to the $j$th output.

In our experiments, the weights $w(j, i)$, the hidden center $C_i$, and the shape parameter of Gaussian kernel function $\sigma_i$ are all adjusted according to a hybrid learning algorithm (HLA) (Er et al., 2002), which combines the gradient paradigm and the linear least square (LLS) paradigm.

### 3.2. Neural networks committee machines

The idea of committee machines is based on a simple engineering principle, namely, the idea referred to as "divide and conquer" (Su and Basu, 2001; Fun and Hagan, 1996; Rahman and Fairhurst, 2000; Hanaen and Salamon, 1990). So called "divide and conquer" is meant that a complex computational task is divided into a set of less complex tasks so that the divided subtasks can be readily solved. The solutions corresponding to these subtasks are then combined at a later stage to produce the final results for the original complex problem.

Recently, many combination strategies of committee machines have been developed. In our experiments, a plurality voting strategy (Hanaen and Salamon, 1990) is adopted for combining the committee members. In this combination strategy, the final decision is the classification result reached by more classifier members than any other, and the class label can be got by the following formula:

$$i = \arg\max_{j=1}^{c}(K_j) \tag{32}$$

where $K_j$ denotes the number of the classifiers which support class $j$.

### 3.3. System architecture of the proposed NNCM

When performing classification using a neural networks committee adopting voting integrating strategy, it is inefficient if the committee contains too few members.

So, in our experiments, the original images are firstly divided into several blocks to increase the number of the committee members. Each image block can produce four committee members' samples, respectively, in four different feature domains.

The system architecture of the proposed committee machine is a feedforward structure, as shown in Fig. 2.

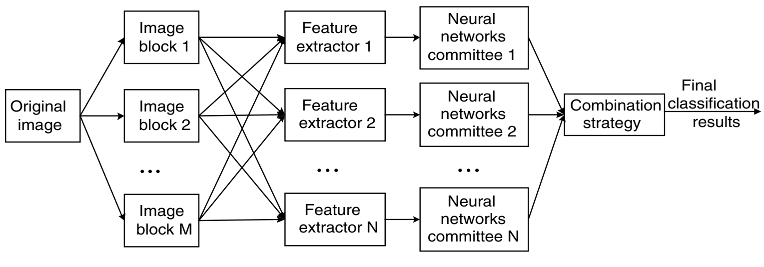


Fig. 2. System architecture of neural networks committee based on multi-features.

## 4. Experiments

### 4.1. Data set

In our experiments, AT&T laboratories database of faces (formerly, also referred to as 'The ORL database of faces') was used. There are 10 different images per subject for 40 distinct subjects. For some of the subjects, the images were taken at different time periods, by varying lighting slightly, changing facial expressions (open/closed eyes, smiling/non-smiling) and facial details (glasses/no-glasses). All the images were taken against a dark homogeneous background and the subjects are in up-right, frontal position (with tolerance for some side movement). Fig. 3 shows an example of face database of one subject of AT&T laboratories.

In all cases, five training images per person (thus 200 total training images) were randomly taken for training, and the remaining images (200 total images) are taken for testing.

### 4.2. Experimental results

We divided the original images into $2*2$, $4*2$, $4*3$, $4*4$ blocks, and then extracted the features using the extractors of the interest operator (IO) method, the principal component analysis (PCA) method, the Fisher's linear discriminant (FLD) technique and the Kernel Fisher linear discriminant (KFLD) method, respectively.

In our experiments, the IO method turns an original image of size of $112*92$ into a $(112 \bmod 5+1)*(92 \bmod 5+1)*5 = 2185$ dimensional vector, the PCA method turns each image block into a 180 dimensional vector, the Fisher's linear discriminant (FLD) turns each original image into a 30 dimensional vector, and the KFLD method turns each original image into a 40 dimensional vector.

Fig. 3. Images of a certain subject.

Here, our experimental results show that the classification accuracy can reach the average value of 94.9% and 95.7%, respectively, when we didn't divide the original images and implemented the classification task using a single RBF network in the FLD or KFLD feature domain. However, when the original images are divided into different blocks, it can be found that the FLD and the KFLD extraction methods are both inefficient by our experiments. So, the neural networks committee in our experiments contains at most one FLD or one KFLD member. For example, if the original images are divided into $2*2$ blocks, the committee has eight committee members when only the IO method and the PCA method are combined (shortly referred to as the IO + PCA). However, when the PCA method and the FLD method are combined (shortly called as the PCA + FLD), it has five committee members (4 PCA members and only 1 FLD member). When voting for the final results, the FLD member should hold a higher weight than other members (because it represents the whole image while other members represent only a certain block of the original image). For the case of dividing original images into $2*2$ blocks, if the weights of the PCA members are set to 1, the weight of FLD member should be set to 2–3.

Consequently, the averaged results over 10 experiments are shown in Table 1. The table shows that the lowest test error rate of face recognition can reach 2.3%.

### 4.3. Discussions

In addition, we also drew the bar chart of the classification error rates vs different number of blocks in different feature domains, as shown in Fig. 4. It can be seen from this chart that in any feature domains dividing images into blocks can always attain lower test error than not dividing images, and the more blocks the images are divided into, the better classification performance can be gained.

Table 1
The experimental results in different feature domains

| Test error % | 1 * 1 Block | 2 * 2 Block | 4 * 2 Block | 4 * 3 Block | 4 * 4 Block |
|---|---|---|---|---|---|
| IO | 13.0 | 13.5 | 11.0 | 10.8 | 11.2 |
| PCA | 8.9 | 9.8 | 8.5 | 7.6 | 7.8 |
| FLD | 5.1 | – | – | – | – |
| KFLD | 4.3 | – | – | – | – |
| IO + PCA | – | 6.7 | 6.3 | 5.2 | 5.4 |
| PCA + FLD | – | 4.2 | 4.5 | 4.2 | 3.8 |
| PCA + KFLD | – | 3.5 | 3.6 | 2.4 | 2.3 |
| IO + PCA + FLD | 4.9 | 4.0 | 4.4 | 3.7 | 3.6 |
| IO + PCA + KFLD | 3.2 | 3.4 | 3.4 | 2.3 | 2.3 |

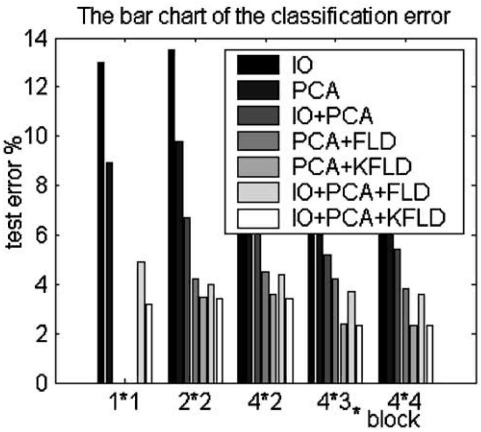


Fig. 4. The bar chart of the classification error rates vs different number of blocks in different feature domains.

Moreover, It can also be seen from these bar curves that the 'IO + PCA' method, the 'PCA + FLD' method, and the 'PCA + KFLD' method all have better performance than the single feature methods, and the 'IO + PCA + FLD' method and the 'IO + PCA + KFLD' method have better performance than those of any two-features methods. In other words, the more the features are adopted, the better the performance attained is.

In Fig. 4, the columns of the 'PCA + FLD' and the 'IO + PCA + FLD' methods are very close to each other, which indicate that the IO members contribute little to the final classification results. The same instance appears in the case of the 'PCA + KFLD' method and the 'IO + PCA + KFLD' method. Both of these instances are caused by the poor classification performance of the IO classifier members. However, the IO classifier members inevitably consume lots of computation time. So, when multiple features are adopted in practical application, both of classification performance and computation complexity should be considered, and a balance should be found.

At the same time, the 'PCA + FLD' has a better performance than the 'IO + PCA' because the 'FLD' has a better performance than the 'IO', and the 'PCA + KFLD' has a better performance than the 'PCA + FLD' because the 'KFLD' has a better performance than the 'FLD'.

In a word, the more efficient the single member is, the more it contributes to the final classification. So, when multiple features are adopted in practical application, the feature domains of relatively better performance should be sought out.

## 5. Conclusions

This paper proposed a new face recognition method based on multiple feature domains using neural networks committee machines. The experiments about the multiple feature domains method were compared with the single feature domain methods. And the experimental results for recognizing the AT&T laboratories database of faces showed that the face recognition accuracy of the multiple feature domains method is higher than that of any one single feature domain. In particular, when the original images are divided into blocks, the highest classification accuracy can be attained. Obviously, our proposed method indeed improves the classification accuracy.

## Acknowledgements

The authors acknowledge a Grant from NSF of China and the Grant of "100 Persons Program" of Chinese Academy of Science of China.

## References

Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J., 1997. Eigenfaces versus fisher faces. IEEE Trans. Pattern Anal. Machine Intell. 19, 711–720.

Brunelli, R., Poggio, T., 1993. Face recognition: Features versus templates. IEEE Trans. PAMI 15, 1042–1052.

Caleanu, C.-D., Facial recognition using committee of neural networks. 5th Seminar on Neural Network Applications in Electrical Engineering, *NEUREL-2000*.

Er, M.J., Wu, S., Lu, J., Toh, H.L., 2002. Face recognition with radial basis function (RBF) neural networks. IEEE Trans. Neural Networks 13 (3).

Fun, M.-H., Hagan, M.T., 1996. Levenberg–Marquardt training for modular networks. IEEE Int. Conf. on Neural Networks 1(3–6), 468–473.

Girosi, F., Poggio, T., 1990. Networks and the best approximation property. Biological Cybernet. 63, 169–176.

Hanaen, L.K., Salamon, P., 1990. Neural network ensembles. IEEE Trans. Pattern Anal. Machine Intell. 12 (10), 993–1001.

Haykin, S., 1996. Adaptive Filter Theory, third ed. Prentice-Hall, Upper Saddle River, NJ.

Huang, D.S., 1999a. Application of generalized radial basis function networks to recognition of radar targets. Int. J. Pattern Recogn. Artif. Intell. 13 (6), 945–962.

Huang, D.S., 1999b. Radial basis probabilistic neural networks: Model and application. Int. J. Pattern Recogn. Artif. Intell. 13 (7), 1083–1101.

Kittler, J., Hatef, M., Duin, R.P.W., Matas, J., 1998. On combining classifier. IEEE Trans. Pattern Anal. Machine Intell. 20 (March), 226–239.

Park, J., Wsandberg, J., 1991. Universal approximation using radial basis functions network. Neural Comput. 3 (2), 246–257.

Pessoa, L., Leitao, A.P., 1999. Complex cell prototype representation for face recognition. IEEE Trans. Neural Networks 10 (6), 1528–1531.

Rahman, A., Fairhurst, M., 2000. Decision combination of multiple classifiers for pattern classification: Hybridisation of majority voting and divide and conquer techniques. The Fifth IEEE Workshop on Applications of Computer Vision 2000.

Su, M., Basu, M., 2001. Gating improves neural network performance. Proc. IJCNN'01 3, 2159–2164.

Swet, D.L., Weng, J., 1996. Using discriminant eigenfeatures for image retrieval. IEEE Trans. Pattern Anal. Machine Intell. 18, 831–836.

Turk, M., Petland, A., 1991. Face recognition using eigenfaces. J. Cognitive Neurosci. 3, 71–86.

Wang, L.C., Der, S.Z., Nasrabadi, N.M., 1998. Automatic target recognition using a feature decomposition and data-decomposition modular neural network. IEEE Trans. Image Process. 7 (8), 1113–1122.

Wechsler, H. et al., 1998. Face recognition. In: From Theory to Applications Proc. of the NATO ASI. Springer-Verlag, Germany.

Yang, M.-H., 2002. Kernel eigenfaces vs. Kernel fisher faces: Face recognition using Kernel methods. Proc. 5th IEEE Internat. Conf. on Automatic Face and Gesture Recognition, 2002.