A New Algorithm for Generalized Optimal Discriminant Vectors

WU Xiaojun (吴小俊)[1,2,4], YANG Jingyu (杨静宇)[2], WANG Shitong (王士同)[1,2], GUO Yuefei (郭跃飞)[3] and CAO Qiying (曹奇英)[1]

[1]*East China Shipbuilding Institute, Zhenjiang 212003, P.R. China*

[2]*School of Information, Nanjing University of Science & Technology, Nanjing 210094, P.R. China*

[3]*Department of Computer Science and Technology, Fudan University, Shanghai 200433, P.R. China*

[4]*Robotics Laboratory, The Chinese Academy of Sciences, Shengyang 110016, P.R. China*

E-mail: wu.xiaojun@263.net

Received February 1, 2001; revised July 5, 2001.

Abstract A study has been conducted on the algorithm of solving generalized optimal set of discriminant vectors in this paper. This paper proposes an analytical algorithm of solving generalized optimal set of discriminant vectors theoretically for the first time. A lot of computation time can be saved because all the generalized optimal sets of discriminant vectors can be obtained simultaneously with the proposed algorithm, while it needs no iterative operations. The proposed algorithm can yield a much higher recognition rate. Furthermore, the proposed algorithm overcomes the shortcomings of conventional human face recognition algorithms which were effective for small sample size problems only. These statements are supported by the numerical simulation experiments on facial database of ORL.

Keywords pattern recognition, feature extraction, discriminant analysis, generalized optimal set of discriminant vectors, face recognition

1 Introduction

Feature extraction is one of the most popular and fundamental problems in pattern recognition. For a specific problem in pattern recognition, extracting efficient features is always the key to solving the problem. A well-known problem in pattern recognition is called "the curse of dimensionality" — more features do not necessarily imply a better classification success rate. Up to now, some existing image features include visual features, moments, Fourier descriptors, and algebraic features. Visual features include edges, contours, textures and regions of an image. The algebraic features of the image are very suitable for describing the inner information of the closed boundaries. Algebraic features represent intrinsic attributions of an image. Turk and Pentland used 'eigenfaces' as the features for human face recognition. The eigenfaces are obtained by the principal component analysis technique. Z.Q. Hong and J.Y. Yang proposed an algebraic feature method in which the singular value vector was used as image features. Cheng presented an efficient recognition approach to human faces based on projective images and feature images were used for classification. Foley-Sammon transform (FST) has been considered as one of the best methods in terms of discriminant problem for linear feature extraction. FST has been applied to image classification and human facial image recognition, and the solving methods of FST under various conditions have been developed. Liu Ke proposed a new criterion of generalized optimal set of discriminant vectors for linear feature extraction, and a unified solving method is derived to solve the vectors of the generalized optimal set. The authors claim that their method is superior to several other methods, such as the Foley-Sammon method, the positive pseudoinverse method, the perturbation method, and the matrix rank decomposition method in terms

This work was supported in part by the National Natural Science Foundation of China (Grant No.60072034), Foundation of Robotics Laboratory, The Chinese Academy of Sciences (Grant No.RL200108), and University's Natural Science Research Program of Jiangsu Province, P.R. China (Grant No.01KJB520002).

of classification rate through numerical experiments. Recently, Guo proposed an iterative algorithm to solve the generalized optimal set of discriminant vectors for linear feature extraction in face recognition, and the performance of the algorithm is superior to the algorithm suggested by Liu Ke in terms of classification rate and speed. However, it is not an analytical solution of the generalized optimal set of discriminant vectors. In this paper, a study has been conducted on the essence of the generalized optimal set of discriminant vectors. We propose an analytical algorithm of solving the generalized optimal set of discriminant vectors theoretically for the first time. All the generalized optimal sets of discriminant vectors can be obtained simultaneously with the proposed algorithm. The proposed algorithm can yield a much higher recognition rate. Furthermore, the proposed algorithm overcomes the shortcomings of conventional human face recognition algorithms which were effective for small sample size problems only. A lot of experimental results have confirmed these statements.

2 F-S Transform and the Optimal Discriminant Criterion

Let $w_1, w_2, \ldots, w_m$ be m known pattern classes, and $X = \{x_i\}$, $i = 1, 2, \ldots, N$, be the set of n-dimensional samples. Each x_i in X belongs to a class w_j, i.e., $x_i \in w_j$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, m$.

The Fisher criterion can be defined as follows:

$$J_f(\varphi) = \frac{\varphi^T S_b \varphi}{\varphi^T S_w \varphi} \tag{1}$$

where φ is an arbitrary n-dimensional vector, S_b and S_w are the between-class scatter matrix and the within-class scatter matrix respectively. Let φ_1 be the unit vector which maximizes $J_f(\varphi)$, then φ_1 is the first vector of Foley-Sammon optimal set of discriminant vectors (the between-class distance in the direction of φ_1 will be maximum while the within-class distance will be minimum), the i-th vector of Foley-Sammon optimal discriminant vectors will be calculated by optimizing the following problem:

$$\max_{\varphi_j^T \varphi_i = 0, \|\varphi_i\| = 1} \{J_f(\varphi_i)\}, \quad j = 1, 2, \ldots, i-1 \tag{2}$$

Let $S = \{\varphi_i\}$, $i = 1, 2, \ldots, r$, then the following linear transform is called FST:

$$y = \Phi^T x \tag{3}$$

where $\Phi = (\varphi_1, \varphi_2, \ldots, \varphi_r)$.

Let Y be the transformed version of X by (3), then the optimal discriminant criterion can be defined as follows:

$$J(\Phi) = \frac{tr(\Phi^T S_b \Phi)}{tr(\Phi^T S_w \Phi)} = \frac{\sum_{i=1}^{r} \varphi_i^T S_b \varphi_i}{\sum_{i=1}^{r} \varphi_i^T S_w \varphi_i} \tag{4}$$

3 Theory of the Analytical Algorithm of Generalized Optimal Set of Discriminant Vectors

Definition 1. *Let*

$$J(\tilde{\Phi}) = \max_{\Phi} J(\Phi) \tag{5}$$

where $\Phi = (\varphi_1, \varphi_2, \ldots, \varphi_r)$, $\widetilde{\Phi} = (\tilde{\varphi}_1, \tilde{\varphi}_2, \ldots, \tilde{\varphi}_r)$, $\varphi_1, \varphi_2, \ldots, \varphi_r$ *and* $\tilde{\varphi}_1, \tilde{\varphi}_2, \ldots, \tilde{\varphi}_r$ *are unit orthogonal column vectors in an n-dimensional space. Then* $J(\Phi)$ *is called the generalized Fisher discriminant function, and* $\tilde{\varphi}_1, \tilde{\varphi}_2, \ldots, \tilde{\varphi}_r$ *are the generalized optimal discriminant vectors.*

Liu[19] provided a solution to the generalized optimal discriminant vectors:

(1) $\tilde{\varphi}_1$ is the unit vector in the n-dimensional space which maximizes $J_f(\varphi)$, i.e., $\tilde{\varphi}_1$ is the first vector of the Foley-Sammon optimal set of discriminant vectors.

(2) The i-th generalized optimal discriminant vector can be obtained by solving the following problem:

$$\max_{\tilde{\varphi}_j^T \tilde{\varphi}_i=0, \|\tilde{\varphi}_i\|=1} \{J_i(\tilde{\varphi}_i)\}, \quad j=1,2,\ldots,i-1 \tag{6}$$

where

$$J_i(\varphi)=\frac{\sum_{j=1}^{i-1}\tilde{\varphi}_j^T S_b \tilde{\varphi}_j+\varphi^T S_b \varphi/\|\varphi\|^2}{\sum_{j=1}^{i-1}\tilde{\varphi}_j^T S_w \tilde{\varphi}_j+\varphi^T S_w \varphi/\|\varphi\|^2} \tag{7}$$

Theorem 1. *$J(\Phi)$ in Definition 1 may be replaced by the following:*

$$\widetilde{J}(\Phi)=\frac{tr(\Phi^T S_b \Phi)}{tr(\Phi^T S_t \Phi)}=\frac{\sum_{i=1}^{r}\varphi_i^T S_b \varphi_i}{\sum_{i=1}^{r}\varphi_i^T S_t \varphi_i} \tag{8}$$

The proof procedure is omitted since it is the same as that of the corollary in [19].

Theorem 2[24]. *Suppose A is a real symmetric matrix of n order, B is a positive-definite matrix of n order, then:*

$$\lambda_0=\frac{\sum_{l=1}^{r}\tilde{\varphi}_l^T A \tilde{\varphi}_l}{\sum_{l=1}^{r}\tilde{\varphi}_l^T B \tilde{\varphi}_l}=\max_{\substack{\varphi_i^T\varphi_j=0\\ \|\varphi_i\|=1}}\left(\frac{\sum_{l=1}^{r}\varphi_l^T A \varphi_l}{\sum_{l=1}^{r}\varphi_l^T B \varphi_l}\right), \quad i,j=1,\ldots,r, \quad i\neq j \tag{9}$$

iff

$$\sum_{l=1}^{r}\tilde{\varphi}_l^T(A-\lambda_0 B)\tilde{\varphi}_l=\max_{\substack{\varphi_i^T\varphi_j=0\\ \|\varphi_i\|=1}}\left(\sum_{l=1}^{r}\varphi_l^T(A-\lambda_0 B)\varphi_l\right)=0, \quad i,j=1,\ldots,r, \quad i\neq j \tag{10}$$

Theorem 3[24]. *Under the assumption of Theorem 2, it holds that:*

$$(1)\ \lambda<\lambda_0 \ \textit{iff} \max_{\substack{\varphi_i^T\varphi_j=0\\ \|\varphi_i\|=1}}\left(\sum_{l=1}^{r}\varphi_l^T(A-\lambda B)\varphi_l\right)>0.$$

$$(2)\ \lambda>\lambda_0 \ \textit{iff} \max_{\substack{\varphi_i^T\varphi_j=0\\ \|\varphi_i\|=1}}\left(\sum_{l=1}^{r}\varphi_l^T(A-\lambda B)\varphi_l\right)<0.$$

Theorem 4[24]. *Let*

$$\sum_{l=1}^{r}\tilde{\varphi}_l^T(\lambda)(A-\lambda B)\tilde{\varphi}_l(\lambda)=\max_{\substack{\varphi_i^T\varphi_j=0\\ \|\varphi_i\|=1}}\left(\sum_{l=1}^{r}\varphi_l^T(A-\lambda B)\varphi_l\right),$$

then

$$\lim_{\lambda\to\lambda_0}\frac{\sum_{l=1}^{r}\tilde{\varphi}_l^T(\lambda)A\tilde{\varphi}_l(\lambda)}{\sum_{l=1}^{r}\tilde{\varphi}_l^T(\lambda)B\tilde{\varphi}_l(\lambda)}=\lambda_0,$$

where λ_0, A, B are the same as those in Theorem 3, λ is a variable and $\tilde{\varphi}_i(\lambda)$ is the i-th vector corresponding to λ.

Corollary. $\left|\frac{\sum_{l=1}^{r}\tilde{\varphi}_l^T(\lambda)A\tilde{\varphi}_l(\lambda)}{\sum_{l=1}^{r}\tilde{\varphi}_l^T(\lambda)B\tilde{\varphi}_l(\lambda)}-\lambda_0\right|\leq\left(1+\frac{r\mu}{\Delta}\right)|\lambda-\lambda_0|,$
where $\Delta=\lambda_{n-r+1}+\cdots+\lambda_n$, $\mu=\max\{|\lambda_1|,|\lambda_n|\}$, $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_n$ are the eigenvalues of matrix B.

Guo designed an iterative algorithm to solve the generalized optimal discriminant vectors based on the above theorems.

Theorem 5[16]. *Suppose A is a real symmetric matrix, then it holds that:*

$$\max_{\substack{\varphi_i^T\varphi_j=0\\ \|\varphi_i\|=1}} \sum_{l=1}^{r} \varphi_l^T A\varphi_l = \lambda_1 + \cdots + \lambda_r, \quad \min_{\substack{\varphi_i^T\varphi_j=0\\ \|\varphi_i\|=1}} \sum_{l=1}^{r} \varphi_l^T A\varphi_l = \lambda_{n-r+1} + \cdots + \lambda_n,$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the n eigenvalues of matrix A.

And suppose $\tilde{\varphi}_1, \tilde{\varphi}_2, \ldots, \tilde{\varphi}_n$ are the orthogonal unit eigenvectors corresponding to $\lambda_1, \lambda_2, \ldots, \lambda_n$, then $\sum_{l=1}^{r} \tilde{\varphi}_l^T A\tilde{\varphi}_l = \lambda_1 + \cdots + \lambda_r$, $\sum_{l=n-r+1}^{n} \tilde{\varphi}_l^T A\tilde{\varphi}_l = \lambda_{n-r+1} + \cdots + \lambda_n$.

Theorem 6[16]. *Let A, E be Hermite matrices of n order respectively. $\tilde{A} = A + E$, $\alpha_i, \beta_i, \tilde{\alpha}_i$, $i = 1, 2, \ldots, n$, are the eigenvalues of $A, E, \tilde{A}$ in decreasing order, then $\alpha_i + \beta_n \leq \tilde{\alpha}_i \leq \alpha_i + \beta_1$.*

Definition 2[16]. *Let A be a matrix of n order, then the summation of the diagonal elements of matrix A is called the trace of matrix A, which is represented by trA.*

So, we have $trA = \sum_{i=1}^{n} a_{ii}$.

Theorem 7[16]. *Let A be a matrix of n order, $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all the eigenvalues of A respectively, then*

$$tr(A) = \sum_{i=1}^{n} \lambda_i$$

Theorem 8. *Let $\varepsilon_1 = \sum_{l=1}^{r} \tilde{\varphi}_l^T(\lambda)(A - \lambda B)\tilde{\varphi}(\lambda) = \max_{\substack{\varphi_i^T\varphi_j=0\\ \|\varphi_i\|=1}} \left(\sum_{l=1}^{r} \varphi_l^T(A - \lambda B)\varphi_l\right)$. $\tilde{\sigma}_i$, $i = 1, 2, \ldots, n$, are the eigenvalues of $A - \lambda B$ in descending order respectively, then*

$$\varepsilon_1 = \sum_{i=1}^{r} \tilde{\sigma}_i \text{ and } \lim_{\lambda \to \lambda_0} \varepsilon_1 = 0$$

where λ_0, A, B are the same as those in Theorem 3.

Proof. Let $\lambda = \lambda_0 + \varepsilon$, then we have $A - \lambda B = A - \lambda_0 B + (-\varepsilon)B$. And let $\sigma_i, \tilde{\sigma}_i$, $i = 1, 2, \ldots, n$, be the eigenvalues of $A - \lambda_0 B$, $A - \lambda B$ in decreasing order respectively, then, according to Theorem 2, we have $\varepsilon_1 = \sum_{i=1}^{r} \tilde{\sigma}_i$.

Let $\varepsilon > 0$, according to Theorem 6, $\sigma_i + (-\varepsilon)\lambda_1 \leq \tilde{\sigma}_i \leq \sigma_i + (-\varepsilon)\lambda_n$ holds. So $\sum_{i=1}^{r} \sigma_i + r(-\varepsilon)\lambda_1 \leq \varepsilon_1 = \sum_{i=1}^{r} \tilde{\sigma}_i \leq \sum_{i=1}^{r} \sigma_i + r(-\varepsilon)\lambda_n$. From the definition of λ_0, we have $\sum_{i=1}^{r} \sigma_i = 0$. And because $\lim_{\lambda\to\lambda_0} r(-\varepsilon)\lambda_i = \lim_{\varepsilon\to 0} r(-\varepsilon)\lambda_i = 0$, $i = 1, 2, \ldots, n$, so $\lim_{\lambda\to\lambda_0} \varepsilon_1 = 0$. In the case of $\varepsilon < 0$, $\sum_{i=1}^{r} \sigma_i + r(-\varepsilon)\lambda_n \leq \varepsilon_1 = \sum_{i=1}^{r} \tilde{\sigma}_i \leq \sum_{i=1}^{r} \sigma_i + r(-\varepsilon)\lambda_1$, $\lim_{\lambda\to\lambda_0} \varepsilon_1 = 0$ also holds.

Theorem 9. *Suppose A is a real symmetric matrix of n order, B is a positive-definite matrix of n order, then:*

$$\lambda_0 = \frac{\sum_{l=1}^{n} \tilde{\varphi}_l^T A\tilde{\varphi}_l}{\sum_{l=1}^{n} \tilde{\varphi}_l^T B\tilde{\varphi}_l} = \max_{\substack{\varphi_i^T\varphi_j=0\\ \|\varphi_i\|=1}} \left(\frac{\sum_{l=1}^{n} \varphi_l^T A\varphi_l}{\sum_{l=1}^{n} \varphi_l^T B\varphi_l}\right), \quad i, j = 1, \ldots, n, \ i \neq j \tag{11}$$

iff

$$\lambda_0 = \frac{trA}{trB}$$

Proof. Let $\tilde{\sigma}_i$, $i = 1, 2, \ldots, n$, be the eigenvalues of $A - \lambda B$ in decreasing order, then, according to Theorem 8, we have $\varepsilon_1 = \sum_{i=1}^{n} \tilde{\sigma}_i$. On the other hand, according to Theorem 7, we have $tr(A - \lambda B) = \sum_{i=1}^{n} \tilde{\sigma}_i$. According to Theorem 8, the optimal value of λ, i.e., λ_0, can be obtained when $\varepsilon_1 = 0$. Therefore, from $\varepsilon_1 = tr(A - \lambda B) = trA - \lambda trB = 0$, we can solve the optimal value of $\lambda = \lambda_0 = \dfrac{trA}{trB}$.

Theorem 10. *For a classification problem with c classes, there are $c - 1$ effective optimal discriminant vectors at most.*

According to Theorem 10, we usually choose $c - 1$ optimal discriminant vectors for a classification problem with c classes. We have found the eigenvalues of $A - \lambda B$ are decreasing exponentially, so we can choose some eigenvectors corresponding to the biggest eigenvalues of $A - \lambda B$, while still keeping high accuracy. That is to say, even a few eigenvectors corresponding to the biggest eigenvalues of $A - \lambda B$ are adopted, we still have $\lambda = \lambda_0 \approx \frac{trA}{trB}$ with high accuracy.

4 The Proposed Algorithm of Generalized Optimal Set of Discriminant Vectors

The following algorithm is designed to solve the generalized optimal discriminant vectors based on the above theorems.

4.1 Case 1: S_t is Nonsingular

In this case, $S_t^{-1}(0) = \{\alpha | S_t\alpha = 0\} = \phi$, $\overline{S_t^{-1}(0)} = R^n$, and S_t is a positive-definite matrix.

(1) Compute the optimal value of λ, i.e., λ_0. According to Theorem 9, $\lambda_0 = \frac{trS_b}{trS_t}$.

(2) Compute $\tilde{\varphi}_1(\lambda), \ldots, \tilde{\varphi}_r(\lambda)$, which are the eigenvectors of $S_b - \lambda_0 S_t$ corresponding to the n eigenvalues of it. Then $\tilde{\varphi}_1(\lambda), \ldots, \tilde{\varphi}_n(\lambda)$ are the generalized optimal discriminant vectors exactly according to the above analysis.

4.2 Case 2: S_t is Singular

Suppose $S_t^{-1}(0) = \text{span}\{\alpha_1, \ldots, \alpha_k\}$, $\overline{S_t^{-1}(0)} = \text{span}\{\beta_1, \ldots, \beta_{n-k}\}$, where $\alpha_1, \ldots, \alpha_k$ and $\beta_1, \ldots, \beta_{n-k}$ are both orthogonal unit vectors.

Because $\forall\alpha \in S_t^{-1}(0)$, $\alpha^T S_b \alpha = \alpha^T S_t \alpha = 0$, so, the vectors in $S_t^{-1}(0)$ contribute nothing to classifying, hence the generalized optimal discriminant vectors should be selected from $\overline{S_t^{-1}(0)}$.

$\forall\beta \in \overline{S_t^{-1}(0)}$, $\beta = a_1\beta_1 + a_2\beta_2 + \cdots + a_{n-k}\beta_{n-k} = P\hat{\beta}$, where $P = (\beta_1, \beta_2, \ldots, \beta_{n-k})$, $\hat{\beta} = (a_1, a_2, \ldots, a_{n-k})^T$, and in the equation of $\tilde{J}(\Phi)$, let $\varphi_l = P\hat{\varphi}_l$, $l = 1, 2, \ldots, n$, then in the subspace of $\overline{S_t^{-1}(0)}$, we have $\tilde{J}(\Phi) = \frac{\sum_{l=1}^{n} \hat{\varphi}_l^T (P^T S_b P)\hat{\varphi}_l}{\sum_{l=1}^{n} \hat{\varphi}_l^T (P^T S_t P)\hat{\varphi}_l} \equiv \tilde{J}(\widehat{\Phi})$, where $\widehat{\Phi} = (\hat{\varphi}_1, \ldots, \hat{\varphi}_n)$, and it is obvious that $P^T S_t P$ is a positive-definite matrix. Analogous to the case of 4.1, $\tilde{\widehat{\Phi}} = (\tilde{\hat{\varphi}}_1, \tilde{\hat{\varphi}}_2, \ldots, \tilde{\hat{\varphi}}_n)$ can be calculated. It is easy to prove the following two relations:

$$\|\varphi_l\| = \|P\hat{\varphi}_l\| = 1 \text{ iff } \|\hat{\varphi}_l\| = 1,$$
$$\varphi_i^T \varphi_j = 0 \quad i \neq j \text{ iff } \hat{\varphi}_i^T \hat{\varphi}_j = 0, \ i \neq j.$$

So the generalized optimal discriminant vectors are $\tilde{\varphi}_l = P\tilde{\hat{\varphi}}_l$, $l = 1, 2, \ldots, n$.

5 Experimental Results

In order to test the performance of the proposed algorithm in this paper, numerical simulation experiments have been done on the facial database of ORL. The sample set is transformed into an r-dimensional space ($r \leq n$) by the proposed methods, Liu's method and Guo's method respectively. Each transformed sample set is tested by the minimum distance classifier designed on the subspace spanned by the discriminant vectors calculated with the relevant method. In each experiment, we first take a part of the sample set as training samples to calculate the optimal discriminant vectors and to design the minimum distance classifier, then use all samples of the sample set to test the classifier. We did our experiments on the ORL face database (http://www.cam-orl.co.uk/facedatabase.html) which can be used freely for academic research. The Cambridge ORL database contains 40 distinct persons, each person having ten different images, taken at different times, varying lighting slightly, facial expressions (open/closed eyes, smiling/nonsmiling), and facial details (glasses/no glasses). All the images are taken against a dark homogeneous background and the persons are in the upright, frontal position (with tolerence for some side movement). Some of the ORL face images are shown in Fig.1. A subset of the ORL images is used as the training data for computing the optimal set of discriminant vectors. All the 400 images are taken as test data. The number of features is $c-1$ in all the experiments, where c is the number of classes. Table 1 shows the experimental results with three methods. A lot of experimental results show that the present method is more efficient than other methods mentioned above in terms of both classification rate and time of computation.

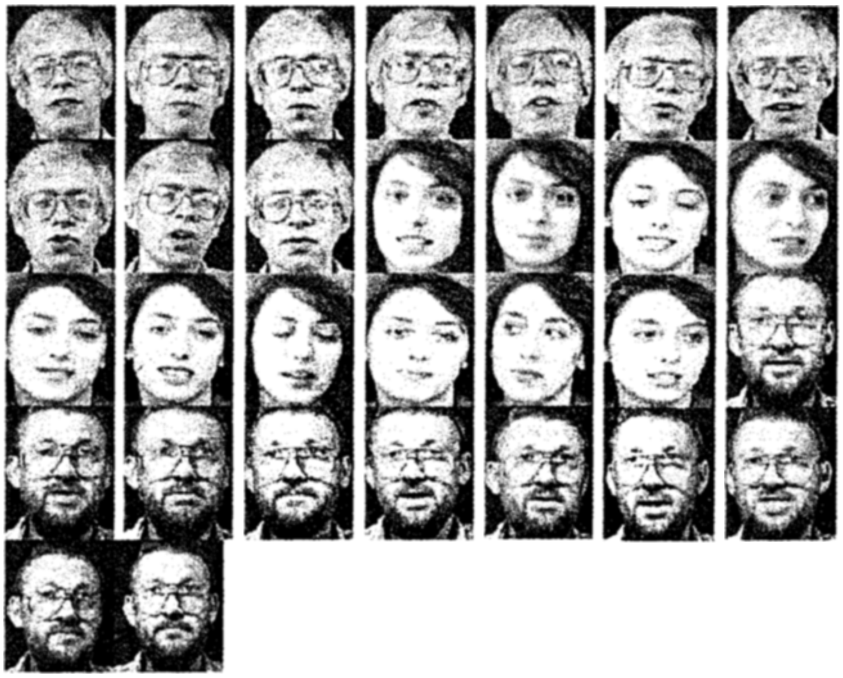

Fig.1. Some faces in our experiments from ORL face database.

Table 1. Classification Results of Human Faces (ORL Face Database)

Number of classes	Number of optimal discriminant vectors	Number of training samples	Number of erroneous classification samples and computation time in seconds					
			Liu's method		Guo's method		Presented method	
8	7	4	2	132.43	0	53.38	0	47.73
19	18	4	16	517.23	24	71.29	8	42.24
30	29	4	49	636.76	29	98.76	18	41.03
39	38	4	100	856.68	69	72.51	32	15.66

References

[1] Rama Chellappa, Wilson C L, Sirohey S. Human and machine recognition of faces: A survey. In *Proceedings of the IEEE,* 1995, 83(5): 705–740.

[2] Rosenfeld A. Survey: Image analysis and computer vision: 1996. *Computer Vision and Image Understanding,* 1997, 62(1): 33–93.

[3] Turk M, Pentland A. Eigenfaces for recognition. *J. Cognitive Neuroscience,* 1991, 3(1): 71–86.

[4] Swets D, Weng J, Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence,* 1996, 18(8): 831–836.

[5] Moghaddam B, Pentland A. Probabilistic visual learning for object representation. *IEEE Trans. Pattern Analysis and Machine Intelligence,* 1997, 19(7): 696–710.

[6] Foley D H, Sammon J W. An optimal set of discriminant vectors. *IEEE Trans. Computer,* 1975, 24(3): 281–289.

[7] Fukunaga K. Introduction to Statistical Pattern Recognition. Academic Press, New York, 1990.

[8] Rowley H A, Baluja S, Kanade T. Neural network-based face detection. *IEEE Trans. Pattern Analysis and Machine Intelligence,* 1998, 20(1): 25–38.

[9] Yoon K S *et al.* Hybrid approaches to frontal view face recognition using the hidden Markov model and neural network. *Pattern Recognition,* 1998, 31(3): 283–293.
[10] Sammon J W. An optimal discriminant plane. *IEEE Trans. Computer,* 1970, 19(3): 826–829.
[11] Tian Q *et al.* Image classification by the Foley-Sammon transform. *Optical Engineering,* 1986, 25(7): 834–839.
[12] Hong Z Q. Algebraic feature extraction of images for recognition. *Pattern Recognition,* 1991, 24(3): 211–219.
[13] Okada T *et al.* Theory of feature extraction by orthogonal discriminant vectors. *Trans. IECE (A),* 1982, 65(8): 767–771. (in Japanese)
[14] Hamamoto Y *et al.* A note on the orthogonal discriminant vector of pattern recognition. *Trans. IECE (A),* 1989, 72(2): 414–419. (in Japanese)
[15] Kittler J. On the discriminant vector method of feature selection. *IEEE Trans. Comput.,* 1977, 26(6): 604–606.
[16] Hong Z Q, Yang J Y. Optimal discriminant plane for a small number of samples and design method of classifier on the plane. *Pattern Recognition,* 1991, 24(4): 317–324.
[17] Cheng Y Q, Zhuang Y M, Yang J Y. Optimal fisher discriminant analysis using the rank decomposition. *Pattern Recognition,* 1992, 25(1): 101–111.
[18] Liu K, Cheng Y Q, Yang J Y. An efficient algorithm for Foley-Sammon optimal set of discriminant vectors by algebraic method. *International Journal of Pattern Recognition and Artificial Intelligence,* 1992, 6(5): 817–829.
[19] Liu K, Cheng Y Q, Yang J Y. A generalized optimal set of discriminant vectors. *Pattern Recognition,* 1992, 25(1): 731–739.
[20] Wang Genglu, Shi Rongchang. Theory of Matrix. National Defense Industry Press, Beijing, 1988.
[21] Sun Jiguang. Perturbation Analysis of Matrix. Science Press, Beijing, 1987.
[22] Liu K, Cheng Y Q, Yang J Y. Algebraic feature extraction for image recognition based on an optimal discriminant criterion. *Pattern Recognition,* 1993, 26(1): 903–911.
[23] Guo Yuefei, Yang Jingyu. An iterative algorithm for the generalized optimal set of discriminant vectors and its application to face recognition. *Chinese J. Computers,* 2000, 23(11): 1189–1195. (in Chinese)
[24] Guo Yuefei. Algebraic feature extraction of face images and research on the optimal discriminant vectors [Dissertation]. Nanjing University of Science and Technology, June, 2000. (in Chinese)

WU Xiaojun received his B.S. degree in mathematics from Nanjing Normal University, Nanjing, P.R. China in 1991 and M.S. degree in engineering from Nanjing University of Science and Technology, Nanjing, P.R. China in 1996 respectively. He is currently working toward the Ph.D. degree in the Department of Computers at the same university. He is also an associate professor in East China Shipbuilding Institute. Since 1996, he has been teaching in the Department of Electronics and Information, East China Shipbuilding Institute where he is an assistant dean of the Department and director of the Center of Computer Fundamental Education of East China Shipbuilding Institute. His current research interests are pattern recognition, fuzzy systems, neural networks and intelligent systems. He has published more than 30 papers in the above areas. Dr. Wu was a fellow of United Nations University, International Institute for Software Technology (UNU/IIST) from 1999 to 2000. He won the most outstanding postgraduate award from Nanjing University of Science and Technology in 1996, and outstanding paper award from the Committee of Military System Engineering, China System Engineering Society in 1999.

YANG Jingyu received his B.S. degree in computer science from Harbin Institute of Military Engineering, Harbin, China. From 1982 to 1984 he was a visiting scholar at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. From 1993 to 1994 he was a visiting professor at the Department of Computer Science, Missouri University. He was a visiting professor at Concordia University in Canada in 1998. He is currently a professor and Chairman in the Faculty of Information at Nanjing University of Science and Technology. He is the author of over 200 scientific papers in computer vision, pattern recognition, and artificial intelligence. He has won more than 20 provincial awards and national awards. His current research interests are in the areas of pattern recognition, robot vision, image processing, information fusion, and artificial intelligence.

WANG Shitong received his B.S. and M.S. degrees both in computer science from Nanjing University of Aeronautics and Astronautics, Nanjing, P.R. China in 1984 and 1987 respectively. He is a professor in East China Shipbuilding Institute, where he is the dean of the Department of Electronics and Information. His current research interests are pattern recognition, fuzzy systems, neural networks and intelligent systems. He has published more than 120 papers in the above areas. Prof. Wang was a visiting fellow in London University from 1995 to 1996. He was a postdoctoral fellow in Bristol University which belongs to Royal Society of Great Bratain from 1998 to 1999. Since the beginning of this year, he has been a visiting professor at Hong Kong University of Science and Technology.