Generalized Two-Dimensional FLD Method for Feature Extraction: An Application to Face Recognition

Shiladitya Chowdhury[1], Jamuna Kanta Sing[2], Dipak Kumar Basu[2], and Mita Nasipuri[2]

[1] Department of Master of Computer Application, Techno India, Kolkata, India
[2] Department of Computer Science & Engineering, Jadavpur University, Kolkata, India
dityashila@yahoo.com, jksing@ieee.org, dipakkbasu@gmail.com, mitanasipuri@yahoo.com

Abstract. This paper presents a novel scheme for face feature extraction, namely, the generalized two-dimensional Fisher's linear discriminant (G-2DFLD) method. The G-2DFLD method is an extension of the 2DFLD method for feature extraction. Like 2DFLD method, G-2DFLD method is also based on the original 2D image matrix. However, unlike 2DFLD method, which maximizes class separability either from row or column direction, the G-2DFLD method maximizes class separability from both the row and column directions simultaneously. In G-2DFLD method, two alternative Fisher's criteria have been defined corresponding to row and column-wise projection directions. The principal components extracted from an image matrix in 2DFLD method are vectors; whereas, in G-2DFLD method these are scalars. Therefore, the size of the resultant image feature matrix is much smaller using G-2DFLD method than that of using 2DFLD method. The proposed G-2DFLD method was evaluated on two popular face recognition databases, the AT&T (formerly ORL) and the UMIST face databases. The experimental results show that the new G-2DFLD scheme outperforms the PCA, 2DPCA, FLD and 2DFLD schemes, not only in terms of computation times, but also for the task of face recognition using a multi-class support vector machine.

Keywords: Generalized two-dimensional FLD, Feature extraction, Face recognition.

1 Introduction

The Fisher's linear discriminant (FLD) method has been widely used for feature extraction and dimension reduction in pattern recognition and computer vision domains. The key idea of the FLD technique is to find the optimal projection that maximizes the ratio of the *between-class* and the *within-class* scatter matrices of the projected samples. A difficulty in using the FLD technique in face recognition is the "*small sample size (SSS)*" problem [1]. This problem usually arises when the number of samples is smaller than the dimension of the samples. In face recognition domain, the dimension of a face image is generally very high. Therefore, the within-class scatter matrix is almost always singular, thereby making the implementation of FLD

M.J. Zaki et al. (Eds.): PAKDD 2010, Part II, LNAI 6119, pp. 101–112, 2010.

method impossible. One direct solution of the *SSS* is to down sample the face images into a considerably small size and then perform FLD technique. However, this process is not computationally efficient as the pre-processing of images takes considerable amount of time before actual application of the FLD technique. Er *et al.* [2] proposed a PCA+FLD technique to avoid the *SSS* problem. In [2], face features are first extracted by the principal component analysis (PCA) method and then the resultant features are further processed by the FLD technique to acquire lower-dimensional discriminant features. An improved PCA technique, the two-dimensional PCA (2DPCA), was proposed by Yang *et al.* [3]. Unlike PCA, which works on the stretched image vector, the 2DPCA works directly on the original 2D image matrix. The 2DPCA is not only computationally efficient, but also superior for the task of face recognition and image reconstruction than the conventional PCA technique [3]. However, the PCA techniques yield projection directions that maximize the total scatter across all classes, i.e., across all face images. Therefore, the PCA retains unwanted variations caused by lighting, facial expression, and other factors [2], [4]. The PCA techniques do not provide any information for class discrimination but dimension reduction [2]. Recently, Xiong *et al.* [5] proposed a two-dimensional FLD (2DFLD) method, which also works directly on the original 2D image matrix and maximizes class separability either from row or column direction. The so called *SSS* problem does not arise in 2DFLD method as the size of its scatter matrices is much smaller. The 2DFLD method is found to be superior to the PCA and 2DPCA in terms of feature extraction and face recognition [5].

In this paper, we have extended the 2DFLD algorithm [5] and present a novel generalized two-dimensional FLD (G-2DFLD) technique, which maximizes class separability from both the row and column directions simultaneously. Like 2DFLD method, G-2DFLD method is also based on the original 2D image matrix. In G-2DFLD method, two alternative Fisher's criteria have been defined corresponding to row and column-wise projection directions. Unlike 2DFLD method, the principal components extracted from an image matrix by the G-2DFLD method are scalars. Therefore, the size of the resultant image feature matrix is much smaller using the G-2DFLD method than that of using the 2DFLD method. The experimental results on the AT&T and the UMIST databases show that the new G-2DFLD scheme outperforms the PCA, 2DPCA, FLD and 2DFLD schemes, not only in terms of computation time, but also for the task of face recognition using a multi-class support vector machine (SVM).

The remaining part of the paper is organized as follows. Section 2 describes the procedure of extracting face features using 2DFLD technique. Section 3 presents the key idea and algorithm of the proposed G-2DFLD method for feature extraction. The experimental results on the AT&T and the UMIST face databases are presented in Section 4. Finally, Section 5 draws the conclusion remarks.

2 Two-Dimensional FLD (2DFLD) Method for Feature Extraction

The 2DFLD [5] method is based on the 2D image matrix. It does not need to form a stretched large image vector from the 2D image matrix. The key idea is to project an image matrix **X**, an $m \times n$ random matrix, onto a projection matrix **A** of dimension $n \times k$ $(k \leq n)$ to get an image feature matrix **Y** of dimension $m \times k$ by the following linear transformation [5]:

$$\boldsymbol{Y} = \boldsymbol{XA} \tag{1}$$

Let there are N training images, each one is denoted by $m \times n$ image matrix $\mathbf{X}_i$ (i=1, 2, …, N). The training images contain C classes (subjects), and the c^{th} class C_c has N_c samples ($\sum_{c=1}^{C} N_c = N$). Let the mean image of the training samples is denoted by $\boldsymbol{\mu}$ and the mean image of the c^{th} class is denoted by $\boldsymbol{\mu}_c$. The between-class and within-class scatter matrices $\boldsymbol{G}_b$ and $\boldsymbol{G}_w$, respectively are defined as follows:

$$G_b = \sum_{c}^{C} N_c (\mu_c - \mu)^T (\mu_c - \mu) \tag{2}$$

$$G_w = \sum_{c}^{C} \sum_{i \in c}^{N} (X_i - \mu_c)^T (X_i - \mu_c) \tag{3}$$

Then the two-dimensional Fisher's criterion J(Q) is defined as follows:

$$J(Q) = \frac{\left| Q^T G_b Q \right|}{\left| Q^T G_w Q \right|} \tag{4}$$

where Q is the projection matrix.

It may be noted that the size of both the $\boldsymbol{G}_b$ and $\boldsymbol{G}_w$ is $n \times n$. If $\boldsymbol{G}_w$ is a nonsingular matrix, the ratio in (4) is maximized when the column vectors of the projection matrix **Q**, are the eigenvectors of $G_b G_w^{-1}$. The optimal projection matrix $\mathbf{Q}_{\mathrm{opt}}$ is defined as follows:

$$\begin{aligned} Q_{opt} &= \arg\max_{Q} \left| G_b G_w^{-1} \right| \\ &= [q_1, q_2, \ldots, q_k] \end{aligned} \tag{5}$$

where $\{q_i \mid i=1, 2, \ldots, k\}$ is the set of normalized eigenvectors of $G_b G_w^{-1}$ corresponding to k largest eigenvalues $\{\lambda_i \mid i=1, 2, \ldots, k\}$.

Now, each face image $\mathbf{X}_i$ (i=1, 2, …, N) is projected into the optimal projection matrix $\mathbf{Q}_{\mathrm{opt}}$ to obtain its $(m \times k)$-dimensional 2DFLD-based features $\mathbf{Y}_i$, which is defined as follows:

$$Y_i = \overline{X}_i Q_{opt} \; ; i = 1, 2, \ldots, N \tag{6}$$

where $\overline{X_i}$ is mean-subtracted image of $\mathbf{X}_i$

3 Generalized Two-Dimensional FLD (G-2DFLD) Method for Feature Extraction

3.1 Key Idea and the Algorithm

Like 2DFLD method, the generalized two-dimensional FLD (G-2DFLD) method is also based on 2D image matrix. The only difference is that, it maximizes class separability from both the row and column directions simultaneously by the following linear transformation:

$$\boldsymbol{Z} = \boldsymbol{U}^{T} \boldsymbol{X}\boldsymbol{V} \tag{7}$$

where **U** and **V** are two projection matrices of dimension $m{\times}p$ $(p{\le}m)$ and $n{\times}q$ $(q{\le}n)$, respectively. Therefore, our goal is to find the optimal projection directions **U** and **V** so that the projected vector in the $(p{\times}q)$-dimensional space reaches its maximum class separability.

3.1.1 Alternate Fisher's Criteria

We have defined two alternative Fisher's criteria $J(U)$ and $J(V)$ corresponding to row and column-wise projection directions as follows:

$$J(U) = \frac{\left|U^{T} G_{br} U\right|}{\left|U^{T} G_{wr} U\right|} \tag{8}$$

and

$$J(V) = \frac{\left|V^{T} G_{bc} V\right|}{\left|V^{T} G_{wc} V\right|} \tag{9}$$

where

$$\mathrm{G}_{\mathrm{br}} = \sum_{\mathrm{c}}^{\mathrm{C}} \mathrm{N}_{\mathrm{c}} (\mu_{\mathrm{c}} - \mu)(\mu_{\mathrm{c}} - \mu)^{\mathrm{T}} \tag{10}$$

$$\mathrm{G}_{\mathrm{wr}} = \sum_{\mathrm{c}}^{\mathrm{C}} \sum_{\mathrm{i} \in \mathrm{c}}^{\mathrm{N}} (\mathrm{X}_{\mathrm{i}} - \mu_{\mathrm{c}})(\mathrm{X}_{\mathrm{i}} - \mu_{\mathrm{c}})^{\mathrm{T}} \tag{11}$$

$$\mathrm{G}_{\mathrm{bc}} = \sum_{\mathrm{c}}^{\mathrm{C}} \mathrm{N}_{\mathrm{c}} (\mu_{\mathrm{c}} - \mu)^{\mathrm{T}} (\mu_{\mathrm{c}} - \mu) \tag{12}$$

$$\mathrm{G}_{\mathrm{wc}} = \sum_{\mathrm{c}}^{\mathrm{C}} \sum_{\mathrm{i} \in \mathrm{c}}^{\mathrm{N}} (\mathrm{X}_{\mathrm{i}} - \mu_{\mathrm{c}})^{\mathrm{T}} (\mathrm{X}_{\mathrm{i}} - \mu_{\mathrm{c}}) \tag{13}$$

We call the matrices $\boldsymbol{G}_{br}$, $\boldsymbol{G}_{wr}$, $\boldsymbol{G}_{bc}$ and $\boldsymbol{G}_{wc}$, as *image row between-class scatter matrix*, *image row within-class scatter matrix*, *image column between-class scatter matrix* and *image column within-class scatter matrix*, respectively. It may be noted that size of the scatter matrices $\boldsymbol{G}_{br}$ and $\boldsymbol{G}_{wr}$ is $m\times m$, whereas, for $\boldsymbol{G}_{bc}$ and $\boldsymbol{G}_{wc}$ the size is $n\times n$. The sizes of these scatter matrices are much smaller than that of the conventional FLD algorithm, whose scatter matrices are $mn\times mn$ in size. For a square image, $m=n$ and we have $\mathbf{G}_{\mathrm{br}} = G_{bc}^T$ and $\mathbf{G}_{\mathrm{wr}} = G_{wc}^T$ and vice-versa.

The ratios in (8) and (9) are maximized when the column vectors of the projection matrix **U** and **V**, are the eigenvectors of $G_{br}G_{wr}^{-1}$ and $G_{bc}G_{wc}^{-1}$, respectively. The optimal projection (eigenvector) matrix $\mathbf{U}_{opt}$ and $\mathbf{V}_{opt}$ are defined as follows:

$$\begin{aligned} U_{opt} &= \arg\max_{U} \left| G_{br} G_{wr}^{-1} \right| \\ &= [\mathrm{u}_1, \mathrm{u}_2, \ldots, \mathrm{u}_p] \end{aligned} \tag{14}$$

$$\begin{aligned} V_{opt} &= \arg\max_{V} \left| G_{bc} G_{wc}^{-1} \right| \\ &= [\mathrm{v}_1, \mathrm{v}_2, \ldots, \mathrm{v}_{\mathrm{q}}] \end{aligned} \tag{15}$$

where $\{\mathrm{u}_\mathrm{i} \mid \mathrm{i}=1, 2, \ldots, p\}$ is the set of normalized eigenvectors of $G_{br}G_{wr}^{-1}$ corresponding to p largest eigenvalues $\{\lambda_\mathrm{i} \mid \mathrm{i}=1, 2, \ldots, p\}$ and $\{\mathrm{v}_\mathrm{j} \mid \mathrm{j}=1, 2, \ldots, q\}$ is the set of normalized eigenvectors of $G_{bc}G_{wc}^{-1}$ corresponding to q largest eigenvalues $\{\alpha_\mathrm{j} \mid \mathrm{j}=1, 2, \ldots, q\}$.

3.1.2 Feature Extraction

The optimal projection matrices $\mathbf{U}_{\mathrm{opt}}$ and $\mathbf{V}_{\mathrm{opt}}$ are used for feature extraction. For a given image sample **X**, an image feature is obtained by the following linear projection:

$$z_{ij} = \boldsymbol{u}_i^{\boldsymbol{T}} \boldsymbol{X}\boldsymbol{v}_j, i = 1,2,\ldots,p; j = 1,2,\ldots,q \tag{16}$$

The z_{ij} (i=1, 2, …, p; j=1, 2, …, q) is called a *principal component* of the sample image **X**. It should be noted that each *principal component* of the 2DFLD method is a vector, whereas, the *principal component* of the G-2DFLD method is a scalar. The principal components thus obtained are used to form a G-2DFLD-based *image feature matrix* **Z** of dimension $p\times q$ $(p\leq m,\ q\leq n)$, which is much smaller than the 2DFLD-based *image feature matrix* **Y** of dimension $m\times k$ $(k\leq n)$. Therefore, in this case an image matrix is reduced considerably in both the row and column directions simultaneously.

4 Experimental Results

The performance of the proposed method has been evaluated on the AT&T Laboratories Cambridge database (formerly ORL database) [6] and the UMIST face database [7]. The AT&T database is used to test performance of the proposed method under the condition of minor variations of rotation and scaling and the UMIST database is used to examine the performance of the method when the angle of rotation of the facial images is quite large. The experiments were carried out in two different strategies; randomly partitioning the database and n-fold cross validation test.

We have designed a multi-class support vector machine (SVM) using Gaussian kernels for classification of the images to test the effectiveness of the G-2DFLD algorithm. The SVM has been recently proposed by Vapnik *et al.* [8] for binary classification and found to be very effective for pattern recognition. A SVM finds the hyperplane that separates the samples of two classes while maximizing the distance from either class to the hyperplane. This hyperplane is called Optimal Separating Hyperplane (OSH), which minimizes the risk of misclassification of both the training and test samples. A multi-class SVM has been designed by combining two class SVMs. In particular, we have adopted the *one-against-all* strategy to classify samples between each class and all the remaining classes. The one-against-all strategy is discussed as follows:

Let the training set ($\mathbf{X_i}$, c_i) consists of N samples of M classes, where c_i $(c_i \in 1,2,\ldots,M)$ represents the class label of the sample $\mathbf{X}_i$. An SVM is constructed for each class by discriminating that class from the remaining (M-1) classes. Thus the number of SVMs used in this approach is M. A test pattern $\mathbf{X}$ is classified by using the *winner-takes-all* decision strategy, i.e., the class with the maximum value of the discriminant function f($\mathbf{X}$) is assigned to it. All the N training samples are used in constructing an SVM for a class. The SVM for class k is constructed using the set of training samples and their desired outputs, ($\mathbf{X_i}$, y_i). The desired output y_i for a training sample $\mathbf{X}_i$ is defined as follows:

$$y_i = \begin{cases} +1 & \text{if} \quad c_i = k \\ -1 & \text{if} \quad c_i \neq k \end{cases} \tag{17}$$

The samples with the desired output y_i = +1 are called *positive* samples and the samples with the desired output y_i = -1 are called *negative* samples.

4.1 Experiments on the AT and T Face Database

The AT&T database contains 400 gray-scale images of 40 persons. Each person has 10 gray-scale images, having a resolution of 112×92 pixels. Images of the individuals have been taken by varying light intensity, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses) and against a dark homogeneous background, with tilt and rotation up to 20^o and scale variation up to 10%. Sample face images of a person are shown in Fig. 1.

4.1.1 Randomly Partitioning the Database

In this experimental strategy, we randomly select d images from each subject to form the training set and the remaining images are included in the test set. To ensure sufficient training and to test the effectiveness of the proposed technique for different sizes of the training sets, we choose the value of d as 3, 4, 5, 6 and 7. It should be noted that there is no overlap between the training and test images. To reduce the influence of performance on the training and test sets, for each value of d, experiment is repeated 20 times with different training and test sets. Since the numbers of projection vectors p and q have a considerable impact on the performance of the G-2DFLD algorithm, we perform several experiments by varying the values of p and q.

Fig. 1. Sample images of a subject from the AT&T database

Fig. 2 shows the recognition rates of the G-2DFLD algorithm using a multi-class support vector machine (SVM). For each value d, average recognition rates are plotted by varying the values of p and q. For d=3, 4, 5, 6 and 7 the best average recognition rates are found to be 92.82%, 95.94%, 97.68%, 98.72% and 98.42%, respectively and the dimension $(p \times q)$ of the corresponding image feature matrices are (16×16), (16×16), (14×14), (14×14) and (8×8), respectively.

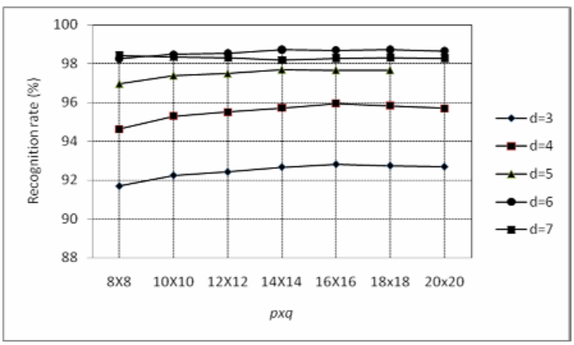

Fig. 2. Average recognition rates of the G-2DFLD algorithm on the AT&T database for different values d by varying the values of p and q

4.1.2 N-Fold Cross Validation Test

In this experiment, we divide the AT&T database (formerly ORL database) into ten-folds randomly, taking one image of a person into a fold. Therefore, each fold consists of 40 images, each one corresponding to a different person. For ten-folds cross validation test, in each experimental run, nine folds are used to train the multi-class SVM and remaining one fold for testing. Therefore, training and test sets consist of 360 and 40 images, respectively. The average recognition rates by varying the image feature matrix (i.e. $p \times q$) are shown in Fig. 3. The best average recognition rate is found to be 99.75% using image feature matrix of size (8×8).

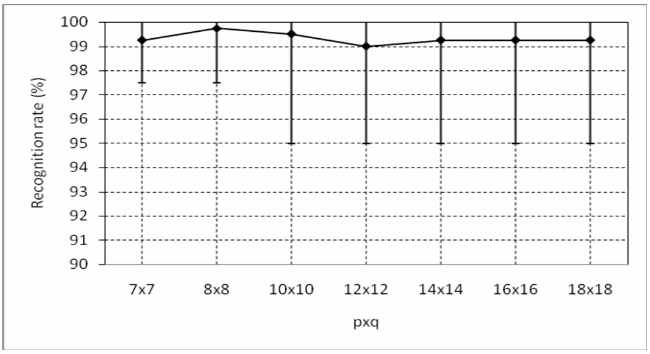

Fig. 3. Average recognition rates of the G-2DFLD algorithm on the AT&T database for 10-folds cross validation test by varying the values of p and q. The upper and lower extrema of the error bars represent the maximum and minimum values, respectively.

4.1.3 Comparison with Other Methods

For a fair comparison, we have implemented the PCA, 2DPCA, PCA+FLD and 2DFLD algorithms and used the same multi-class SVM and parameters for classification. The comparisons of the best average recognition rates of the PCA, 2DPCA, PCA+FLD and 2DFLD algorithms along with the proposed G-2DFLD algorithm using the two different experimental strategies are shown in Table 1. It may be noted that in all the cases the performance of the G-2DFLD method is better than the PCA, 2DPCA, PCA+FLD and 2DFLD methods.

Table 1. Comparison of different methods in terms of average recognition rates (%) on the AT&T database. Figures within the parentheses denote the number of features.

Experiment	Method	d=3	d=4	d=5	d=6	d=7
Randomly partition, d images/subject	G-2DFLD	92.82 (16×16)	95.94 (16×16)	97.68 (14×14)	98.72 (14×14)	98.42 (8×8)
	PCA	85.58 (60)	89.42 (60)	93.10 (60)	95.28 (60)	96.01 (60)
	2DPCA	91.27 (112×16)	94.33 (112×16)	96.83 (112×14)	97.72 (112×14)	97.79 (112×8)
	PCA+FLD	83.65 (25)	88.65 (25)	92.60 (25)	95.30 (25)	95.83 (25)
	2DFLD	92.30 (112×16)	95.08 (112×16)	97.50 (112×14)	98.26 (112×14)	97.88 (112×8)
10-folds cross validation test	G-2DFLD	99.75 (8×8)				
	PCA	97.00 (60)				
	2DPCA	99.25 (112×8)				
	PCA+FLD	98.25 (25)				
	2DFLD	99.00 (112×8)				

Table 2. Comparison of average feature extraction time (in seconds) using 200 training and 200 test images on the AT&T database

Method	# of features	Time (seconds)
G-2DFLD	14×14 = 196	12.95
PCA	60	55.10
2DPCA	112×14 = 1568	32.55
PCA+FLD	25	55.75
2DFLD	112×14 = 1568	22.35

Table 2 shows the average time (in seconds) taken by the G-2DFLD, PCA, 2DPCA, PCA+FLD and 2DFLD methods for feature extraction on the AT&T database using an IBM Intel Pentium 4 Hyper-Threading technology, 3.0 GHz, 2 GB DDR-II RAM computer running on Fedora 9 Linux Operating Systems. It may be again noted that the proposed G-2DFLD method is more efficient than the PCA, 2DPCA, PCA+FLD and 2DFLD methods in term of computation time.

4.2 Experiments on the UMIST Face Database

The UMIST[1] face database is a multi-view database, consisting of 575 gray-scale images of 20 people (subject), each covering a wide range of poses from profile to frontal views. Each image has a resolution of 112×92 pixel. Each subject also covers a range of race, sex and appearance. Unlike the ORL database, the number of images per people is not fixed; it varies from 19 to 48. Fig. 4 shows some of the sample images of a subject from the database.

Fig. 4. Some sample images of a subject from the UMIST database

4.2.1 Randomly Partitioning the Database

Like AT&T database, we randomly select d images from each subject to form the training set and the remaining images are included in the test set. We choose the value of d as 4, 6, 8 and 10. It should be again noted that there is no overlap between the training and test images. For each value of d, experiment is repeated 20 times with

[1] At present UMIST database contains 475 images. However, we have used the earlier version of the UMIST database to test with more number of images.

different training and test sets. Fig. 5 shows the recognition rates of the G-2DFLD algorithm using a multi-class SVM. For each value *d*, average recognition rates are plotted by varying the values of *p* and *q*. For *d*=4, 6, 8 and 10 the best average recognition rates are found to be 86.22%, 92.28%, 95.54% and 96.92%, respectively and the dimension *(p×q)* of the corresponding image feature matrices are (14×14), (14×14), (14×14) and (18×18), respectively.

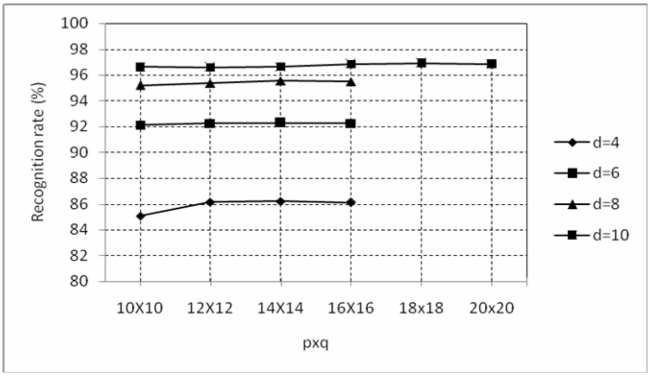

Fig. 5. Average recognition rates of the G-2DFLD algorithm on the UMIST database for different values *d* by varying the values of *p* and *q*

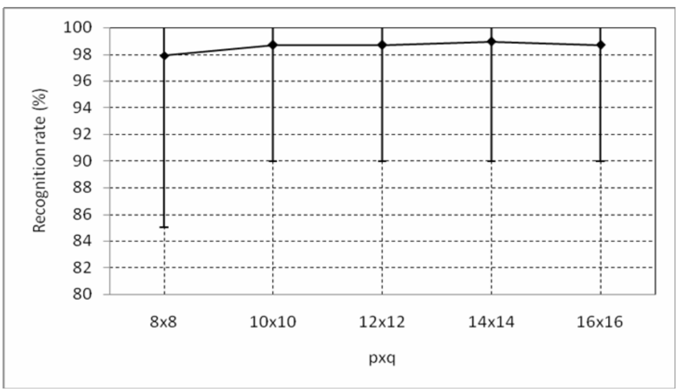

Fig. 6. Average recognition rates of the G-2DFLD algorithm on the UMIST database for 19-folds cross validation test by varying the values of *p* and *q*. The upper and lower extrema of the error bars represent the maximum and minimum values, respectively.

4.2.2 N-Fold Cross Validation Test

Since the number of images per subject varies from 19 to 48, we have randomly divided the database into 19 folds, taking one image of a subject into a fold. Therefore, in each fold there are 20 images, each one corresponding to a different subject. For 19-folds cross validation test, in each experimental run, 18 folds are used to train the multi-class SVM and remaining one fold is used for testing. Therefore, training and test sets consist of 360 and 20 images, respectively in a particular experimental run. The average recognition rates by varying the image feature matrix (i.e. $p \times q$) are shown in Fig. 6. The best average recognition rate is found to be 98.95% using image feature matrix of size (14×14).

4.2.3 Comparison with Other Methods

For a fair comparison, like AT&T database, we have implemented the PCA, 2DPCA, PCA+FLD and 2DFLD algorithms and used the same multi-class SVM and parameters for classification. The comparisons between the best average recognition rates of the PCA, 2DPCA, PCA+FLD and 2DFLD algorithms along with the propose G-2DFLD method using the two different experimental strategies are shown in Table 3. It may be again noted that in all the cases the performance of the G-2DFLD method is better than the PCA, 2DPCA, PCA+FLD and 2DFLD methods, excepts in 19-folds cross validation test, where the performance of the 2DPCA method matches with that of the proposed G-2DFLD method.

Table 3. Comparison of different methods in terms of average recognition rates (%) on the UMIST database. Figures within the parentheses denote the number of features.

Experiment	Method	d=4	d=6	d=8	d=10
Randomly partition, d images/subject	G-2DFLD	86.22 (14×14)	92.28 (14×14)	95.54 (14×14)	96.92 (18×18)
	PCA	80.72 (60)	86.53 (60)	94.01 (60)	95.11 (60)
	2DPCA	85.70 (112×14)	91.91 (112×14)	95.07 (112×14)	96.60 (112×18)
	PCA+FLD	76.31 (25)	85.69 (25)	90.93 (25)	93.72 (25)
	2DFLD	86.12 (112×14)	92.16 (112×14)	95.25 (112×14)	96.55 (112×18)
19-folds cross validation test	G-2DFLD	98.95 (14×14)			
	PCA	98.68 (60)			
	2DPCA	98.95 (112×14)			
	PCA+FLD	96.36 (25)			
	2DFLD	98.68 (112×14)			

5 Conclusion

In this paper, we have presented a novel scheme for face feature extraction, namely, generalized two-dimensional FLD (G-2DFLD) method, which is based on the original 2D image matrix. The G-2DFLD algorithm maximizes class separability from both the row and column directions simultaneously, resulting in a smaller image feature matrix. To realize this, we have defined two alternative Fisher's criteria corresponding to row and column-wise projection directions. The principal components extracted from an image matrix by the G-2DFLD method are scalars. Since the size of the scatter matrices in the proposed G-2DFLD algorithm is much smaller than those in the conventional PCA and FLD schemes, the computational time for feature extraction is much less. The experimental results on the AT&T and UMIST databases show that the G-2DFLD method is more efficient than the PCA, 2DPCA, PCA+FLD, and 2DFLD methods, not only in terms of computation times, but also for the task of face recognition using a multi-class support vector machine (SVM).

Acknowledgment. This work was partially supported by the UGC major research project (F. No.: 37-218/2009(SR), dated: 12-01-2010), CMATER and the SRUVM projects of the Department of Computer Science & Engineering, Jadavpur University, Kolkata, India. The author, Shiladitya Chowdhury would like to thank Techno India, Kolkata for providing computing facilities and allowing time for conducting research works. The author, D. K. Basu would also like to thank the AICTE, New Delhi for providing him the Emeritus Fellowship (F.No.: 1-51/RID/EF(13)/2007-08, dated 28-02-2008).

References

1. Fukunaga, K.: Introduction to Statistical Pattern Recognition. Academic Press, New York (1990)
2. Er, M.J., Wu, S., Lu, J., Toh, H.L.: Face recognition with radial basis function (RBF) neural networks. IEEE Trans. Neural Networks 13, 697–710 (2002)
3. Yang, J., Zhang, D., Frangi, A.F., Yang, J.-Y.: Two-dimensional PCA: A new approach to appearance-based face representation and recognition. IEEE Trans. Pattern Anal. Mach. Intell. 26, 131–137 (2004)
4. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces versus fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Machine Intell. 19, 711–720 (1997)
5. Xiong, H., Swamy, M.N.S., Ahmad, M.O.: Two-dimensional FLD for face recognition. Pattern Recognition 38, 1121–1124 (2005)
6. ORL face database. AT&T Laboratories, Cambridge, U. K., `http://www.uk.research.att.com/facedatabase.html`
7. Graham, D.B., Allinson, N.M.: Characterizing Virtual Eigensignatures for General Purpose Face Recognition: From Theory to Applications. In: Wechsler, H., Phillips, P.J., Bruce, V., Fogelman-Soulie, F., Huang, T.S. (eds.) Computer and Systems Sciences. NATO ASI Series F, vol. 163, pp. 446–456 (1998)
8. Vapnik, V.N.: Statistical learning theory. John Wiley & Sons, New York (1998)