A New Manifold-Based Feature Extraction Method

Zhongbao Liu(✉)

School of Information Engineering & Art Design,
Zhejiang University of Water Resources and Electric Power,
Hangzhou 310018, China
liu_zhongbao@hotmail.com

Abstract. Many traditional feature extraction methods takes the global or the local characteristics of training samples into consideration during the process of feature extraction. How to fully utilize the global or the local characteristics to improve the feature extraction efficiencies is worthy of research. In view of this, a new Manifold-based Feature Extraction Method (MFEM) is proposed. MFEM takes both the advantage of Linear Discriminant Analysis (LDA) in keeping the global characteristics and the advantage of Locality Preserving Projections (LPP) in keeping the local characteristics into consideration. In MFEM, Within-Class Scatter based on Manifold (WCSM) and Between-Class Scatter based on Manifold (BCSM) are introduced and the optimal projection can be obtained based on the Fisher criterion. Compared with LDA and LPP, MFEM considers the global information and local structure and improves the feature extraction efficiency.

Keywords: Feature extraction · Fisher criterion · Global characteristics · Local structure

1 Introduction

Linear Discriminant Analysis (LDA) [1] is popular in practice, in which the non-singularity problem has greatly influent its improvement of efficiencies. In view of this, many effective improvements are made by scientists: Regularized Discriminant Analysis (RDA) proposed by Friedman [2] efficiently solves the above problem; 2D-LDA is proposed to directly extract the features based on Fisher criterion [3]; Orthogonal LDA (OLDA) tries to diagonalize the scatter matrix so as to obtain the discriminant vectors [4]; Direct LDA (DLDA) [5] carries no discriminative information by modifying the simultaneous diagonalization procedure. Besides, the commonly-used improvement approach include Pseudo-inverse LDA (PLDA) [6], Two-stage LDA [7], Penalized Discriminant Analysis (PDA) [8], Enhanced Fisher Linear Discriminant Model (EFM) [9]. In recent years, as to the under-sampled problems, we proposed Scalarized LDA (SLDA) [10], and Matrix Exponential LDA (MELDA) [11].

The general strategy of above approach is to solve the singularity problem firstly and then uses the Fisher criterion to obtain the optimal projections. LDA only takes the sample global information into consideration but always neglects the local structure.

D.-S. Huang et al. (Eds.): ICIC 2019, LNAI 11645, pp. 337–346, 2019.
https://doi.org/10.1007/978-3-030-26766-7_31

On the other hand, many popular manifold learning approach such as Locality Preserving Projection (LPP) [12], only focus on the local structure.

Therefore, all the sample information including the global characteristics and local structure is taken into consideration, and propose a new Manifold-based Feature Extraction Method (MFEM). MFEM inherits the advantage of Fisher criterion and manifold learning and effectively improve the feature extraction efficiency.

2 Background Knowledge

2.1 LDA

Given a dataset matrix $X = [x_1, x_2, \ldots, x_N] = [x_1, x_2, \ldots, x_c]$ where $x_i (i = 1, \ldots, N)$, N and c are respectively the training size and the class size. N_i denoting the number of sample in the i th class.

In LDA, two scatters named between-class scatter $\boldsymbol{S_B}$ and within-class scatter $\boldsymbol{S_W}$ are defined as:

$$\boldsymbol{S_B} = \sum_{i=1}^{c} \frac{N_i}{N} (\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})^{\mathrm{T}} \tag{1}$$

$$\boldsymbol{S_W} = \sum_{i=1}^{c} \sum_{j=1}^{N_i} \frac{1}{N} (\boldsymbol{x_{ij}} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x_{ij}} - \bar{\boldsymbol{x}}_i)^{\mathrm{T}} \tag{2}$$

where $\bar{\boldsymbol{x}}_i = \frac{1}{N_i} X_i e_i$ with $e_i = [1, \ldots, 1]^T \in R^{N_i}$ is the centroid of class i and $\bar{\boldsymbol{x}} = \frac{1}{N} Xe$ with $e_i = [1, \ldots, 1]^T \in R^N$ is the global centroid.

The optimal function of LDA is:

$$J(\boldsymbol{W_{opt}}) = \max_{W} \frac{\boldsymbol{W}^{\mathrm{T}} \boldsymbol{S_B} \boldsymbol{W}}{\boldsymbol{W}^{\mathrm{T}} \boldsymbol{S_W} \boldsymbol{W}} \tag{3}$$

The Eq. (3) is equivalent to:

$$\max_{\boldsymbol{W}} \ \boldsymbol{W}^{\mathrm{T}} \boldsymbol{S_B} \boldsymbol{W} \tag{4}$$

and

$$\min_{\boldsymbol{W}} \ \boldsymbol{W}^{\mathrm{T}} \boldsymbol{S_W} \boldsymbol{W} \tag{5}$$

where $\boldsymbol{W}$ is the optimal projection.

The projection $\boldsymbol{W}$ can be obtained by calculating the eigenvectors.

It can be seen from the above analysis, LDA tries to preserve the global characteristics invariant before and after feature extraction. Its efficiency can not be improved because it neglects the local structure of each class.

2.2 LPP

The optimal problem of LPP is:

$$\min_{W} \sum_{i,j} (\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x_i} - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x_j})^2 S_{ij} \tag{6}$$

$$s.t \sum_{i} \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x_i}\boldsymbol{D_{ii}}\boldsymbol{x_i}^{\mathrm{T}}\boldsymbol{W} = 1 \tag{7}$$

where $\boldsymbol{W}$ is the optimal projection, S_{ij} is the weight function which reflects the similarity of samples, $\boldsymbol{D_{ii}} = \sum_j S_{ij}$.

The above optimization problem can be transformed as follows based on the linear algebra theory:

$$\min_{W} \boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W} \tag{8}$$

$$s.t.\ \boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}=1 \tag{9}$$

where $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{S}$.

The optional projection matrix is obtained by computing all the nonzero eigenvectors of $\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W} = \boldsymbol{\lambda}\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}$.

In conclusion, LPP tries to preserve the local characteristic but does not take the global characteristics into consideration, especially, when encountering noise, the feature extraction efficiency of LPP is greatly influenced.

3 MFEM

Feature extraction is a classical preprocessed approach in dealing with the high-dimensional samples. Though they are widely-used in practice, the feature extraction efficiency is limited due to neglecting the global characteristics and local structure. In order to take all the characteristics of the training samples, a new Manifold-based Feature Extraction Method (MFEM) is proposed. In MFEM, Within-Class Scatter based on Manifold (WCSM) and Between-Class Scatter based on Manifold (BCSM) are introduced and the optimal projection can be obtained based on the Fisher criterion.

3.1 Between-Class Scatter Based on Manifold

Inspired by manifold learning, we firstly construct the adjacency graph $\boldsymbol{G_D} = \{\boldsymbol{X}, \boldsymbol{D}\}$ where $\boldsymbol{G_D}$ donates a graph with different classes, X and D donate the dataset and the weight of different classes respectively. The different-class weight function of two random samples $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ can be defined:

$$D_{ij} = \begin{cases} \exp(-d \Big/ \left\| \boldsymbol{x_i} - \boldsymbol{x_j} \right\|^2), & l_i \neq l_j \\ 0, & l_i = l_j \end{cases} \tag{10}$$

where l_i (i = 1, 2, …, N) donates the class label and d is a constant.

The different-class weight function verifies that if the samples $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ belong to different classes, the weight of them is large; or else, the weight is zero.

In order to preserve the local characteristics of different classes, the samples $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ belonged to different classes will be far away after feature extraction. The optimization problem can be described as follows.

$$\max_{\boldsymbol{W}} \sum_{i,j} (\boldsymbol{y_i} - \boldsymbol{y_j})^2 D_{ij} \tag{11}$$

Where $\boldsymbol{y_i} = \boldsymbol{W}^{\mathrm{T}} \boldsymbol{x_i}$, $\boldsymbol{W}$ donates the projection matrix and $\boldsymbol{x_i} \in \boldsymbol{X}$.

$\sum_{i,j} (\boldsymbol{y_i} - \boldsymbol{y_j})^2 D_{ij}$ is reformulated to the following equations based on the algebraic transformation.

$$\begin{aligned} &\tfrac{1}{2} \sum_{i,j} (\boldsymbol{y_i} - \boldsymbol{y_j})^2 D_{ij} \\ &= \tfrac{1}{2} \sum_{i,j} (\boldsymbol{W}^{\mathrm{T}} \boldsymbol{x_i} - \boldsymbol{W}^{\mathrm{T}} \boldsymbol{x_j})^2 D_{ij} \\ &= \sum_{i,j} (\boldsymbol{W}^{\mathrm{T}} \boldsymbol{x_i} D_{ii} \boldsymbol{x_i}^{\mathrm{T}} \boldsymbol{W} - \boldsymbol{W}^{\mathrm{T}} \boldsymbol{x_i} D_{ij} \boldsymbol{x_j}^{\mathrm{T}} \boldsymbol{W}) \\ &= \boldsymbol{W}^{\mathrm{T}} \boldsymbol{X} \boldsymbol{D}' \boldsymbol{X}^{\mathrm{T}} \boldsymbol{W} - \boldsymbol{W}^{\mathrm{T}} \boldsymbol{X} \boldsymbol{D} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{W} \\ &= \boldsymbol{W}^{\mathrm{T}} \boldsymbol{X} (\boldsymbol{D}' - \boldsymbol{D}) \boldsymbol{X}^{\mathrm{T}} \boldsymbol{W} \\ &= \boldsymbol{W}^{T} \boldsymbol{S_D} \boldsymbol{W} \end{aligned} \tag{12}$$

where $\boldsymbol{S_D} = \boldsymbol{X}(\boldsymbol{D}' - \boldsymbol{D})\boldsymbol{X}^{\mathrm{T}}$, $\boldsymbol{D}'$ is a diagonal matrix and $\boldsymbol{D}' = \sum_j D_{ij}$.

By taking (12) to (11), (11) is reformulated to

$$\max_{\boldsymbol{W}} \ \boldsymbol{W}^{\mathrm{T}} \boldsymbol{S_D} \boldsymbol{W} \tag{13}$$

Based on the above analysis, we can see Eqs. (4) and (13) reflect the global characteristics of different classes and local structure of each class respectively. In order to fully utilize all the above information, we can obtain the following optimization expression based on (4) and (13).

$$\begin{aligned} &\max_{\boldsymbol{W}} \ \alpha \boldsymbol{W}^{\mathrm{T}} \boldsymbol{S_B} \boldsymbol{W} + (1 - \alpha) \boldsymbol{W}^{\mathrm{T}} \boldsymbol{S_D} \boldsymbol{W} \\ &= \max_{\boldsymbol{W}} \ \boldsymbol{W}^{\mathrm{T}} [\alpha \boldsymbol{S_B} + (1 - \alpha) \boldsymbol{S_D}] \boldsymbol{W} \\ &= \max_{\boldsymbol{W}} \ \boldsymbol{W}^{\mathrm{T}} \boldsymbol{M_B} \boldsymbol{W} \end{aligned} \tag{14}$$

where $\boldsymbol{M_B} = \alpha \boldsymbol{S_B} + (1 - \alpha) \boldsymbol{S_D}$ and α is a parameter balancing $\boldsymbol{S_B}$ and $\boldsymbol{S}_D$. $\boldsymbol{M_B}$ is called Between-Class Scatter based on Manifold (BCSM).

3.2 Within-Class Scatter Based on Manifold

Similarity with BCSM, the same-class weight function of two random samples $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ is defined:

$$S_{ij} = \begin{cases} \exp(-\left\|\boldsymbol{x_i} - \boldsymbol{x_j}\right\|^2 \Big/ s), & l_i = l_j \\ 0, & l_i \neq l_j \end{cases} \tag{15}$$

where l_i (i = 1, 2, .., N) donates the class label and s is a constant.

The same-class weight function verifies that if the samples $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ with the same class label, the weight of them is large; or else, the weight is zero.

In order to keep the neighborhood close, it can be described as:

$$\min_{\boldsymbol{W}} \sum_{i,j} (\boldsymbol{y_i} - \boldsymbol{y_j})^2 S_{ij} \tag{16}$$

where $\boldsymbol{y_i} = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x_i}$, W donates the projection matrix and $\boldsymbol{x_i} \in \boldsymbol{X}$.

$\sum_{i,j} (\boldsymbol{y_i} - \boldsymbol{y_j})^2 S_{ij}$ is reformulated to the following equations based on the algebraic transformation.

$$\begin{aligned} &\tfrac{1}{2}\sum_{i,j} (\boldsymbol{y_i} - \boldsymbol{y_j})^2 S_{ij} \\ &= \tfrac{1}{2}\sum_{i,j} (\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x_i} - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x_j})^2 S_{ij} \\ &= \sum_{i,j} (\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x_i} S_{ii} \boldsymbol{x_i}^{\mathrm{T}}\boldsymbol{W} - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x_i} S_{ij} \boldsymbol{x_j}^{\mathrm{T}}\boldsymbol{W}) \\ &= \boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{S}'\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W} - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{S}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W} \\ &= \boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{S}' - \boldsymbol{S})\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W} \\ &= \boldsymbol{W}^{\mathrm{T}}\boldsymbol{S_S}\boldsymbol{W} \end{aligned} \tag{17}$$

where $\boldsymbol{S_S} = \boldsymbol{X}(\boldsymbol{S}' - \boldsymbol{S})\boldsymbol{X}^T$, $\boldsymbol{S}'$ is a diagonal matrix and $\boldsymbol{S}^{'} = \sum_j S_{ij}$.

By taking (17) to (16), (16) is reformulated to

$$\max_{\boldsymbol{W}} \ \boldsymbol{W}^{\mathrm{T}}\boldsymbol{S_S}\boldsymbol{W} \tag{18}$$

The following optimization expression based on (5) and (18).

$$\begin{aligned} &\max_{\boldsymbol{W}} \ \beta\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S_W}\boldsymbol{W} + (1-\beta)\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S_S}\boldsymbol{W} \\ &= \max_{\boldsymbol{W}} \ \boldsymbol{W}^{\mathrm{T}}[\beta\boldsymbol{S_W} + (1-\beta)\boldsymbol{S_S}]\boldsymbol{W} \\ &= \max_{\boldsymbol{W}} \ \boldsymbol{W}^{T}\boldsymbol{M_W}\boldsymbol{W} \end{aligned} \tag{19}$$

where $\boldsymbol{M_W} = \beta\boldsymbol{S_W} + (1-\beta)\boldsymbol{S_S}$ and β is a parameter balancing $\boldsymbol{S}_W$ and $\boldsymbol{S}_S$. $\boldsymbol{M}_W$ is called Within-Class Scatter based on Manifold (WCSM).

3.3 The Optimization Problem

Inspired by the Fisher criterion, the above optimization problem can be described as follows.

$$J = \max_{W} \frac{\boldsymbol{M_B}}{\boldsymbol{M_W}} = \max_{W} \frac{\boldsymbol{W}^{\mathrm{T}}(\alpha \boldsymbol{S_B} + (1-\alpha)\boldsymbol{S_D})\boldsymbol{W}}{\boldsymbol{W}^{\mathrm{T}}(\beta \boldsymbol{S_W} + (1-\beta)\boldsymbol{S_S})\boldsymbol{W}} \tag{20}$$

The solution of the maximization of (20) is given by computing all the nonzero eigenvectors of $\boldsymbol{M_B W} = \lambda \boldsymbol{M_W W}$.

From the optimization expression of MFEM, it can be seen MFEM not only takes the global characteristics into consideration, but also preserves the local structure. MFEM inherits the advantages of LDA and LPP and improves the feature extraction efficiency to some extent. When $\alpha = \beta = 1$ or $d = s = \infty$, MFEM is equivalent to LDA; When $\alpha = \beta = 0$, $d = \infty$ and $s < \infty$, MFEM is equivalent to LPP.

In practice, $\boldsymbol{M_W}$ maybe singular and the optimal projection can not be obtained by the above approach. For the sake of convenience, the singular value perturbation by adding a little positive number to the diagonal of $\boldsymbol{M_W}$ is introduced to solve the singular problem.

3.4 Optimization Algorithm

Input: the original dataset $\boldsymbol{X}$ and the reduced dimension $\boldsymbol{d}$
Output: the corresponding lower dimensional dataset $\boldsymbol{Y} = [\boldsymbol{y_1}, \boldsymbol{y_2}, \ldots, \boldsymbol{y_d}]$
Step1: Construct the adjacency graph $\boldsymbol{G_D} = \{\boldsymbol{X}, \boldsymbol{D}\}$ and $\boldsymbol{G_S} = \{\boldsymbol{X}, \boldsymbol{S}\}$ where $\boldsymbol{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_N}\}$ donates the original dataset, $\boldsymbol{D}$ and $\boldsymbol{S}$ respectively donate the weights of different classes and the same class. We put an edge between $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ if they are in different classes in the $\boldsymbol{G_D}$, or else, put an edge between them in the $\boldsymbol{G_s}$.
Step2: Compute the different-class weights and the same-class weights. If different-class samples $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ are connected, utilize Eq. (10) to compute the different-class weights; else utilize Eq. (15) to compute the same-class weights.
Step3: Compute $\boldsymbol{S_W}$, $\boldsymbol{S_B}$, $\boldsymbol{M_W}$ and $\boldsymbol{M_B}$.
Step4: Solving the singular problem of $\boldsymbol{M_W}$. The singular value perturbation is introduced to solve the singular problem. Let $\boldsymbol{M_W}$ transform to $\boldsymbol{M'_W}$ after perturbation.
Step5: Compute the optimal projection $\boldsymbol{W}$. The solution of the optimal projection $\boldsymbol{W}$ is given by computing all the nonzero eigenvectors of $\boldsymbol{M_W^{-1} M_B W} = \lambda \boldsymbol{W}$ or $\boldsymbol{M'^{-1}_W M_B W} = \lambda \boldsymbol{W}$. The nonzero eigenvectors corresponding to the biggest d eigen-values are combine to form the optimal projection $\boldsymbol{W} = [\boldsymbol{w_1}, \ldots, \boldsymbol{w_d}]$.
Step6: As to a certain sample $\boldsymbol{x}_i \in X$, the corresponding lower dimensional sample can be obtained by $\boldsymbol{y_i} = \boldsymbol{W}^{\mathrm{T}} \boldsymbol{x_i}$.

4 Experimental Analysis

4.1 UCI Two-Dimensional Visualization

Wine dataset, including 178 samples with 3 classes, in UCI machine learning repository is used in the experiment. Set the reduced dimension is two, successively run PCA, LPP, LDA, MFEM on the wine dataset, and we can obtain the experimental results, shown in Table 1.

Table 1. Recognition rates of PCA, LPP, LDA, MFEM on the face datasets

Data sets	k	PCA	LPP	LDA	MFEM
ORL	3	0.711(28)	0.789(28)	0.814(30)	**0.875(20)**
	4	0.808(28)	0.867(30)	0.875(30)	**0.954(18)**
	5	0.845(22)	0.890(24)	0.905(30)	**0.950(21)**
	6	0.863(22)	0.906(30)	0.950(30)	**0.963(25)**
	7	0.892(20)	0.917(22)	0.925(26)	**0.958(20)**
	8	0.873(20)	0.925(30)	0.938(26)	**0.963(23)**
Yale	4	0.619(12)	0.733(14)	0.667(14)	**0.733(12)**
	5	0.667(14)	0.763(14)	0.767(14)	**0.767(13)**
	6	0.653(12)	0.770(14)	0.747(10)	**0.787(14)**
	7	0.750(12)	0.833(12)	0.833(14)	**0.900(14)**
	8	0.800(10)	**0.899(14)**	0.822(14)	0.867(14)

It can be seen from Fig. 1, some different-class samples are overlapped after feature extraction by PCA. LPP, LDA and MFEM can mainly fulfill the feature extraction task but the efficiencies are different. After feature extraction by LPP, some samples lying near the three-class boundary are overlapped. Therefore, compared with LDA and MFEM, the efficiency of LPP is lowest. The efficiencies of LDA and MFEM are both high, but in the respect of distribution, MFEM shows much more perfect than LDA. This is because MFEM tries to preserve the original distribution by taking both the global and local characteristics, while LDA only focus on the global characteristics based on the Fisher criterion so as to make the different-class samples far and the same-class samples close. Although the within-class scatter reflects the closeness of the same-class samples, yet it does not take the relationship of adjacent samples before and after feature exaction.

4.2 Experiments on Face Datasets

Experiment datasets include ORL face dataset and Yale face dataset. We will discuss the relationship between the sizes of training samples and recognition rates as well as the relationship between the reduced dimensions and recognition rates.

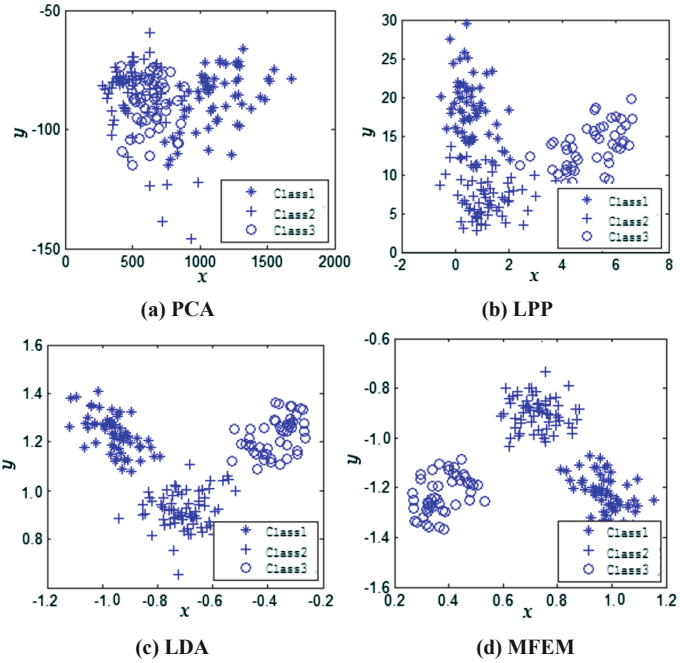

Fig. 1. The experiment results of 2-dimensional visualization

Relationship Between the Size of Training Samples and the Recognition Rate. The training dataset consists of the first k images of each subject, and the remainders are used for test. The values of k in ORL and Yale dataset are selected from 3, 4, 5, 6, 7 and from 3, 5, 7, 9 respectively. The comparative experimental results are show in Table 1. In order to overcome the small size problem in LDA, we utilize PCA + LDA instead for LDA in the experiment.

It can be seen from Table 1, compared with PCA, LPP, LDA, MFEM performs best on the ORL dataset and except $k = 8$, the efficiency of MFEM is highest on the Yale dataset.

Relationship Between the Reduced Dimensions and Recognition Rates. The training dataset consists of the first 5 images of each subject, and the remainders are used for test. We can obtain the experiment results shown in Fig. 2.

It can be seen from Fig. 2, as the reduced dimension becomes higher, the corresponding recognition rate mainly has an upward tendency. Compared with PCA, LPP, LDA, MFEM performs best.

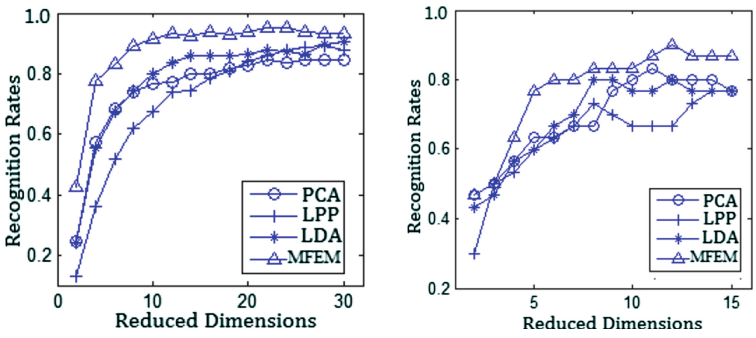

Fig. 2. Relationship between reduced dimensions and recognition rates

5 Conclusions

Researches on current feature extraction approaches can be reduced to two ways, one pays more attention on the global structure, and the other originates from local structure and tries to make the relationship between samples before and after feature extraction be invariant. In view of shortages of classical feature extraction approaches, MFEM is proposed. MFEM considers all the information and improves the feature extraction efficiency. Experiments on some standard datasets verify the effectiveness of MFEM. In practice, linear inseparability is a quite common problem and how to solve it is attracting more and more researchers' interest. MFEM proposed in this paper is suitable to the linear separability situation, how to expand MFEM to linear inseparability is our next work.

References

1. Martinez, A.M., Kak, A.C.: PCA versus LDA. IEEE Trans. Pattern Anal. Mach. Intell. **23** (2), 228–233 (2001)
2. Friedman, H.: Regularized discriminant analysis. J. Am. Stat. Assoc. **84**(405), 165–175 (1989)
3. Li, M., Yuan, B.: 2D-LDA: a novel statistical linear discriminant analysis for image matrix. Pattern Recogn. Lett. **26**(5), 527–532 (2005)
4. Ye, J.P., Xiong, T.: Computational and theoretical analysis of null space and orthogonal linear discriminant analysis. J. Mach. Learn. Res. **7**, 1183–1204 (2006)
5. Yu, H., Yang, J.: A direct LDA algorithm for high-dimensional data with application to face recognition. Pattern Recogn. **34**(11), 2067–2070 (2001)
6. Ji, S.W., Ye, J.P.: Generalized linear discriminant analysis: a unified framework and efficient model selection. IEEE Trans. Neural Netw. **19**(10), 1768–1782 (2008)
7. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eiegnfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997)

8. Hastle, T., Ruja, A., Tibshirani, R.: Penalized discriminant analysis. Ann. Stat. **23**(1), 73–102 (1995)
9. Liu, C.J., Wechsler, H.: Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Trans. Image Proc. **11**(4), 267–276 (2002)
10. Liu, Z., Wang, S.: Improved linear discriminant analysis method. J. Comput. Appl. **31**(1), 250–253 (2011)
11. Liu, Z., Wang, S.: An improved LDA algorithm and its application to face recognition. Comput. Eng. Sci. **33**(7), 89–93 (2011)
12. He, X.F., Niyogi, P.: Locality preserving projections. In: Advances in Neural Information Processing Systems (NIPS), Vancouver, Canada, pp. 153–160 (2003)