Journal of Computational Information Systems 9: 8 (2013) 2993–3000
Available at http://www.Jofcis.com

Fisher Locality Preserving Projections for Face Recognition ⋆

Guoqiang WANG [1], Yunxing SHU [1,∗], Dianting LIU [2], Yanling SHAO [3]

[1] *Department of Computer and Information Engineering, Luoyang Institute of Science and Technology, Luoyang 471023, China*

[2] *Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33124, USA*

[3] *Henan University of Science and Technology, Luoyang 471003, China*

Abstract

In this paper, a novel dimensionality reduction method termed Fisher Locality Preserving Projections (FLPP) is proposed by introducing the maximum scatter difference criterion (MSDC) to the objective function of Locality Preserving Projections (LPP). FLPP not only inherits the advantages of LPP which attempts to preserve the local structure, but also makes full use of class information and orthogonal subspace. After being embedded into a low-dimensional subspace, the samples of the same class maintain their intrinsic neighbor relations, whereas the samples of the different classes are far from each other. In addition, the small sample size problem (SSS) is avoided. As a result, the most discriminative feature is extracted. Experiment results on the ORL and FERET face databases demonstrate the effectiveness of the proposed FLPP method.

Keywords: Locality Preserving Projections; Dimensionality Reduction; Manifold Learning; Maximum Scatter Difference Criterion; Face Recognition

1 Introduction

Face recognition [1] is one of the most active and challenging research topic in computer vision and pattern recognition, in which dimensionality reduction is a crucial step. In the past several decades, many dimensionality reduction methods have been proposed, in which the most well-known ones are principal component analysis (PCA) [2] and linear discriminant analysis (LDA) [3]. PCA aims to project the data along the directions of maximal variances so that the reconstruction error can be minimized. Different from PCA, LDA attempts to better discriminate patterns of different classes. Though extracting most discriminative information from high-dimensional data,

⋆Project supported by Young Core Instructor of colleges and universities in Henan province Funding scheme (No. 2011GGJS-173), the Science and Technology Project of Henan province (No. 122102210138).

∗Corresponding author.
Email address: `wgq2211@163.com` (Yunxing SHU).

April 15, 2013

LDA suffers the small sample size (SSS) problem in practice. To address this problem, extensive methods have been proposed in the literatures [4, 5].

Recently, manifold learning methods have drawn much attention [6, 7]. Unlike PCA and LDA, which aim to preserve the global Euclidean structure of the data space, manifold learning algorithms aim to preserve the inherent manifold structure. Among these original manifold learning methods [8, 9, 10], the most representative is Laplacian Eigenmap (LE) [8]. However, LE is defined only on the training samples, and the issue of how to map new test data remains difficult. As a linear version of LE , Locality Preserving Projections (LPP) [11] is proposed, which can be defined everywhere, rather than only on the training sample. LPP also preserves the local structure of the sample data. Although LPP is effective in many domains, it nevertheless suffers from some limitations: it deemphasizes discriminant information, which makes it suitable for recognition task; it suffers from the SSS problem too.

To address these problems, we present a new dimensionality reduction method, called Fisher Locality Preserving Projections (FLPP). Based on LPP, FLPP adds the maximum Scatter difference criterion (MSDC) [4] to its objective function. It not only holds the strong discriminant power of MSDC, but also preserves the intrinsic geometry of the sample data. It makes the samples of the same class maintain their intrinsic neighbor relations, whereas the samples of different classes are far from each other. In addition, a set of orthogonal basis eigenvectors are obtained to further improve the discriminant power. The inverse matrix operation is not necessarily constructed, and the SSS problem encountered in traditional LPP algorithm is avoided.

2 Related Works

Let $X = [x_1, x_2, \ldots, x_N]$ be a set of face image vectors, and $x_i \in R^D (i = 1, 2, \ldots, N)$. Linear dimensionality reduction techniques try to find a transformation matrix $A = [\alpha_1, \alpha_2, \ldots, \alpha_d] \in R^{D\times d}$ that maps these D points to be new points $Y = [y_1, y_2, \ldots, y_N]$ in a dimensional space, where $y_i = A^T x_i \in R^d (d << D)$, T denotes the transposition of matrix.

2.1 Outline of locality preserving projections (LPP)

As a recently promising method, LPP which derives from Laplacian Eigenmap (LE) inherits the properties of LE and owns more merits. The objective function of LPP is defined as follows [11]:

$$\min \sum_{i,j} (y_i - y_j)^2 S_{ij} \tag{1}$$

where the matrix S is a weight matrix to ensure that the projected vectors y_i and y_j are close to each other if the original vectors x_i and x_j are close. A reasonable criterion to decide S is the nearest neighbor function, $S_{ij} = \exp(-\|x_i - x_j\|^2/t)$ if x_i is among k nearest neighbor of x_j or x_j is among k nearest neighbor of x_i; otherwise, $S_{ij} = 0$. Note that S is a symmetric and positive semi-definite matrix.

Suppose A is a transformation matrix, that is $Y = A^T X$. By simple algebra formulation, the objective function can be reduced to:

$$\frac{1}{2}\sum_{i,j} (y_i - y_j)^2 S_{ij} = \frac{1}{2}\sum_{i,j} (A^T x_i - A^T x_j)^2 S_{ij} = A^T X (D - S) X^T A = A^T X L X^T A \tag{2}$$

where D is a diagonal matrix with its elements being the row (or column since S is symmetric) sums of S, $D_{ii} = \sum_j S_{ij}$. $L = D - S$ is the Laplacian matrix. To eliminate the arbitrary scaling factor, a constraint, that is $YDY^T = 1, i.eA^TXDX^TA = 1$ is imposed.

Finally, the minimization problem reduces to finding:

$$\arg \min_{A^TXDX^T=1} A^TXLX^TA \tag{3}$$

The transformation matrix A that minimizes the objective function can be obtained by solving the following generalized eigenvalues problem:

$$XLX^T\alpha = \lambda XDX^T\alpha \tag{4}$$

2.2 Outline of maximum scatter difference criterion (MSDC)

To avoid the singularity problem of Fisher discriminant criterion (FDC), a novel linear discriminant criterion maximum scatter difference [4] is proposed. MSDC tries to seek the optimal projection directions on which the data points of different classes are projected as far from each other as possible, while requiring the data points of same classes to be close to each other. D-ifferent from FDC, MSDC adopts the generalized scatter difference instead of the generalized Rayleigh quotient as the measure of separability of the projected sample data.

Let $A = [a_1, a_2, \ldots, a_d] \in R^{D\times d}$ be the projection matrix. The projection of a sample X on the projection vectors $a_1, a_2, \ldots, a_d$ is A^Tx. The between-class and within-class scatter of the projected training samples $A^Tx_1, A^Tx_2, \ldots, A^Tx_N$ are defined respectively as:

$$\tilde{S}_b = A^TS_bA \tag{5}$$

$$\tilde{S}_w = A^TS_wA \tag{6}$$

The objective function of MSDC is defined as:

$$\max J(A) = \tilde{S}_b - C \cdot \tilde{S}_w \tag{7}$$

where the parameter C is a nonnegative constant which balances the relative merits of maximizing the between-class scatter to the minimization of the within-class scatter, the larger the parameter C is, the more important the within-class scatter (relative to the between-class scatter).

With Eq.(5) and Eq.(6), Eq.(7) can be transformed into

$$\max J(A) = A^TS_bA - C \cdot A^TS_bA = A^T(S_b - C \cdot S_w)A \tag{8}$$

here, $S_b - C \cdot S_w$ is a generalized scatter difference matrix, and C is a nonnegative parameter. To make the optimal discriminant vectors $a_1, a_2, \ldots, a_d$ be orthonormal, we can use theroem 1.

Theorem 1 The unit eigenvectors of the generalized scatter difference matrix $S_b - C \cdot S_w$ corresponding to the first d largest eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_d$ is the optimal discriminant vectors of MSDC. The optimal discriminant vectors $a_1, a_2, \ldots, a_d$ satisfy:

$$(S_b - C \cdot S_w)a_i = \lambda_j a_i \tag{9}$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$

Comparing the MSDC with the classical FDC, we find that the former avoids calculation of the inverse within-class scatter, i.e.$S_w^{-1}S_b$ is substituted by $S_b - S_w$ this can not only make computationally more efficient but also avoid the singular problem of the within-class scatter.

3 Fisher Locality Preserving Projections (FLPP)

LPP only puts stress on preserving the geometrical structure of the underlying manifold. However, discriminant information is ignored. In order to improve the discriminability of the original LPP, a new dimensionality reduction algorithm called Fisher locality preserving projections (FLPP) is proposed through integrating MSDC into the objective function of LPP, which indicates further application of class label information. FLPP not only preserves the local structure relationship but considers the discriminant information.

Let $X = [x_1, x_2, \ldots, x_N]$ be a set of vectors of N face images taking values in a D dimensional image space, and each face image x_i belongs to one of C classes $\{X_1, X_2, \ldots, X_c\}$. The objective function of LPP is as follows:

$$\min J(Y) = \sum_{i,j} (y_i - y_j)^2 S_{ij} \Rightarrow \min J(A) = A^T XLX^T A \tag{10}$$

The objective function of MSDC is as follows:

$$\max J(A) = A^T (S_b - C \cdot S_w) A \tag{11}$$

Thus, it is easy to know that the objective function of MSDC is equivalent to $\min J(A) = A^T (C \cdot S_w - S_b) A$.

The transformed function of MSDC is introduced into the objective function of LPP, the objective function of FLPP can be obtained:

$$\min J(A) = \theta A^T XLX^T A + (1 - \theta) A^T (C \cdot S_w - S_b) A \tag{12}$$

where θ is a adjustable factor which determines the weight of the locality structure information of face manifold and the discriminant information of samples. $L = D - S$, L has the same solving method and meaning as in LPP algorithm. The parameter C has the same meaning as in MSDC algorithm. S_w is called the within-class scatter, and S_b is called the between-class scatter matrix. According to [11], S_t, S_w, and S_b can be formulated, respectively as follows:

$$S_t = (1/N) \sum\nolimits_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T = (1/N) X (I - (1/N) ee^T) X^T = XGX^T \tag{13}$$

$$S_w = \sum\nolimits_{i=1}^{C} \sum\nolimits_{x \in X_i} (x - \mu_i)(x - \mu_i)^T = X(I - E)X^T = XMX^T \tag{14}$$

$$S_b = S_t - S_w = X(G - M)X^T = XBX^T \tag{15}$$

where $G = I - (1 - N)ee^T$, I is an identity matrix, $e = (1, 1, ..., 1)^T$,$M = I - E$,$E_{ij} = 1/N_c$ if x_i and x_j belong to class X_c; otherwise, $E_{ij} = 0$, $B = G - M$. Thus, Eq. (12) can be rewritten as:

$$\min J(A) = A^T X (\theta L + (1 - \theta)(C \cdot M - B)) X^T A \tag{16}$$

The optimal projection matrix $A = (\alpha_1, \alpha_2, \ldots, \alpha_d)$ can be obtained by computing Eq.(16). To eliminate the linear correlation for the extracted features, the column vectors of matrix A are always required to be orthogonal, that is $A^T A = I$, I is the identity matrix. By using Lagrange multiplier method, the constrained minimization problem can be converted to the following generalized eigenvalue problem:

$$X(\theta L + (1-\theta)(C \cdot M - B))X^T A = \lambda A \tag{17}$$

This means that the optimal projection matrix $A = (\alpha_1, \alpha_2, \ldots, \alpha_d)$ is given by d orthogonal eigenvectors associated with d smallest eigenvalues of $X(\theta L + (1-\theta)(C \cdot M - B))X^T$.

In face recognition application, the dimension D of the vector samples is usually large, so performing FLPP by directly solving the eigenvectors of the $D \times D$ matrix $X(\theta L + (1-\theta)(C \cdot M - B))X^T$ is still computationally intensive. To reduce the computational demand, in this subsection, we present an efficient algorithm for performing FLPP.

Suppose $P = [\beta_1, \beta_2, \ldots, \beta_m]$ to be the matrix of all unit eigenvectors of total scatter matrix S_t corresponding to nonzero eigenvalues and $u \in R^{m\times 1}$ to be the eigenvector of the matrix $P^T(X(\theta L + (1-\theta)(C \cdot M - B))X^T)P$ corresponding to the eigenvalue λ, we have:

$$P^T(X(\theta L + (1-\theta)(C \cdot M - B))X^T)Pu = \lambda u \tag{18}$$

$$PP^T(X(\theta L + (1-\theta)(C \cdot M - B))X^T)Pu = \lambda Pu \tag{19}$$

Let $Q = [\beta_{m+1}, \beta_{m+2}, \ldots, \beta_D]$, where $\beta_{m+1}, \beta_{m+2}, \ldots, \beta_D$ are eigenvectors of S_t corresponding to zero eigenvalue, it follows that $Q^T x_i = 0 (1 = 1, 2, \ldots, N)$ and $Q^T X = 0$.

We can obtain:

$$((PP^T + QQ^T)X(\theta L + (1-\theta)(C \cdot M - B))X^T)Pu = \lambda Pu \tag{20}$$

Let$A = [P\ Q]$, we have

$$AA^T = PP^T + QQ^T \tag{21}$$

Then we can obtain

$$(AA^T X(\theta L + (1-\theta)(C \cdot M - B))X^T)Pu = \lambda Pu \tag{22}$$

It is obvious that A is a unitary matrix, so we have

$$X(\theta L + (1-\theta)(C \cdot M - B))X^T Pu = \lambda Pu \tag{23}$$

i.e.,Pu is the eigenvector of the matrix $X(\theta L + (1-\theta)(C \cdot M - B))X^T$ corresponding to the eigenvalue λ. Note that $P^T X(\theta L + (1-\theta)(C \cdot M - B))X^T P$ is of size $m \times m$, which has much smaller size than that of $X(\theta L + (1-\theta)(C \cdot M - B))X^T$, since usually $m = rank(S_t) << D$.

4 Experimental Results

Two face image databases, namely, ORL database and FERET database, are used to evaluate the effectiveness of the proposed FLPP method. We have also compared the proposed method with other five methods such as PCA [2], LDA [3], MSDC [4], NPP [12] and LPP [11]. We applied the nearest neighborhood classifier with Euclidean metric for classification.

4.1 Experiments on ORL database

The ORL database contains 400 images from 40 individuals, each providing ten different images. The images are captured at different times and have different variations, including expressions and facial details. For the purpose of computational efficiency, all images were in grayscale and cropped and resized to 32× 32 pixels. Ten sample images of one person are displayed in Fig.1.

Fig. 1: Sample face images from the ORL database

4.1.1 Parameter selection

As there is no theory to guide the setting of the parameter θ and C, we can only perform some experiments to roughly estimate the proper value of the parameter θ and C. To get the optimal parameter θ and C with the FLPP algorithm, 5 samples of each individual are randomly selected from the ORL database for training and the remaining are used for testing. The parameter C is set to 1, 10 and 100 respectively. The parameter θ is changed from 0.1 to 1.0 with step 0.1. The reduced dimension is 39. The recognition rates of FLPP under different parameter θ and C are given in Table 1. As can be seen, when $\theta = 0.5$,$C = 10$, the recognition rate is the best.

Table 1: The recognition rate of different parameter θ and C on ORL database

θ / C	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1	0.957	0.953	0.958	0.956	0.966	0.952	0.959	0.968	0.963	0.960
10	0.962	0.970	0.961	0.967	0.974	0.962	0.966	0.967	0.964	0.961
100	0.962	0.959	0.962	0.963	0.966	0.967	0.968	0.955	0.962	0.953

4.1.2 Face recognition

In the experiments,we randomly selected l(l varies from 3 to 5) samples of each individual for training and the others for testing. For each given l , We repeat the classification process 10 times by different splits and calculate the average recognition rates. Table 2 reports the average best performance of different methods. As can be seen, our proposed FLPP method has better recognition rate than other methods in almost all cases, and its performance improves significantly as the number of training samples increases.

4.2 Experiments on FERET database

The proposed algorithm is tested on a subset of the FERET database which contains 200 individuals and seven images for each person. It is composed of images whose names are marked with two-character strings: bd, bj, bf, be, bc, ba, bk. The subset involves variations in facial

Table 2: Comparison of different methods for ORL database (mean ± std (dim))

Method	3Train	4Train	5Train
PCA	78.34± 2.88(119)	85.76± 2.16(159)	88.35 ± 1.85(199)
LDA	87.51± 3.05(39)	91.29±2.55(39)	93.56±1.48(39)
MSDC	87.60± 2.81(39)	91.40±2.34(38)	94.02±2.61(37)
NPP	87.72±2.40(39)	91.35±2.21(39)	93.55±1.28(39)
LPP	87.42±2.46(39)	91.53±2.15(40)	94.20±1.53(39)
FLPP	90.68±2.05(39)	92.70±1.81(38)	95.12±2.02(39)

expression, illumination and pose. All the images in the subset are automatically cropped based on the location of eyes and the cropped image was resized to 40×40 pixels. Fig.2 shows sample images of one person. Similarly to the strategy adopted on ORL, l(=3,4,5)images per person are randomly selected for training and the rest are used for testing. All the results are 10 times averaged. The best performances of these algorithms as well as the reduced dimensions are shown in Table 3. We can draw a similar conclusion as before.

Fig. 2: Sample face images from the FERET database

Table 3: Comparison of different methods for FERET database (mean ± std (dim))

Method	3Train	4Train	5Train
PCA	46.53±1.64 (599)	50.42±2.14(799)	53.86±3.02(999)
LDA	66.62±1.95(80)	70.93±2.55(92)	74.28±2.24(150)
MSDC	70.52±1.82(199)	75.36±2.61(196)	79.41±2.13(198)
NPP	74.17±2.42(194)	77.28±2.09(197)	80.83±2.34(201)
LPP	75.52±1.82(167)	78.36±2.14(196)	82.17±2.56(198)
FLPP	85.94±2.14(189)	87.45±2.39(199)	90.85±1.98(197)

4.3 Discussion

From the above experiments, several apspects are worthwhile to emphasize:

(1) Compared to PCA, LDA, and MSDC which attempt to preserve the global Euclidean structure of the data space, FLPP attempts to preserve local neighboring geometry and discriminant structure of the data manifold. Therefore, when applied to complex nonlinear data, the proposed FLPP can efficiently extract intrinsic features and achieve better performance.

(2) FLPP significantly outperforms NPP and LPP, irrespective of the variations in training sample size. There are two reasons contributing to this phenomenon. On the one hand, FLP-P takes into account the class discriminant information. On the other hand, the orthogonal eigenvectors basis contributes to more locality preserving power.

(3) Our FLPP method is the top performer in all the experimental cases, which shows that FLPP can handle the SSS problem very well.

5 Conclusion

A novel dimensionality reduction algorithm called Fisher Locality Preserving Projections algorithm (FLPP) is proposed in this paper. FLPP is designed to achieve good discrimination ability by introducing MSDC to the objective function of LPP. It can preserve both global discriminant and local geometrical structure of the data. In addition, the inverse matrix operation and the SSS problem are avoided. Experimental results on ORL and FERET face databases indicate FLPP works well and has a promising performance.

References

[1] W. Zhao, R. Chellappa, A. Rosenfeld, P. J. Phillips. Face recognition: a literature survey, ACM Computing Surveys, 35(4), 2003, pp. 399-458.

[2] M. Turk, A. Pentland, Eigenfaces for Recognition, Journal of Cognitive Neuoscience, 3(1), 1991, pp. 71-86.

[3] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, Eigenfaces vs. Fisherfaces. Recognition Using Class Specific Linear Projection, IEEE Trans.Pattern Analysis and Machine Intelligence, 19(7), 1997, pp. 711-720.

[4] F. X. Song, K. Cheng, J. Y. Yang, et al., Maximum scatter difference, large margin linear projection and support vector machines, Acta Automat. Sinica, 30(6), 2004, pp. 890-896.

[5] L. F. Chen, H. Y. M. Liao, M. T. Ko, et al., A new LDA-based face recognition system which can sovle the small sample size problem, Pattern Recognition Letter, 33(10), 2000, pp. 1713-1726.

[6] G. Q. Wang, W. J. Zhang, D. Liu, Discriminant Uncorrelated Neighborhood Preserving Projections, Journal of Information & Computational Science, 8(14), 2011, 3019-3026.

[7] D. T. Liu, Z. Y. Ou, et al., Face Recognition using Hierarchical Isomap, In: Proceeding of Fifth IEEE Workshop on Automatic Identification Advanced Technologies, 2007, pp. 103-106.

[8] M. Belkin, P. Niyogi, Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering, In. Advances in Neural Information Processing Systems, 2002.

[9] L. Saul, S. Roweis, Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds, Journal of Machine Learning Research, 2003, 4, pp. 119-155.

[10] J. B. Tenenbaum, V. de Silva, J. C. Langford, A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science, 290(5500), 2000, 2319-2323.

[11] X. He, S. Yan, Y. Hu, P. Niyogi, H. J. Zhang, Face Recognition Using Laplacianfaces, IEEE.Trans. Pattern Analysis and Machine Intelligence, 27(3), 2005, pp. 328-340.

[12] Y. W. Pang, N. H. Yu, et al., Face Recognition Using Neighborhood Prserving Projections, In: Proceeding. of Pacific-Rim Conference on Multimedia, 2005, 3768, pp. 854-864.