http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 10 (12): 2464-2469, 2011
ISSN 1812-5638 / DOI: 10.3923/itj.2011.2464.2469

Neighborhood Preserving Fisher Discriminant Analysis

Guoqiang Wang and Weijuan Zhang
Department of Computer and Information Engineering, Luoyang Institute of Science and Technology, 471023 Luoyang, Henan, People's Republic of China

Abstract: In this study, a novel subspace learning method named Neighborhood Preserving Fisher Discriminant Analysis (NPFDA) is proposed for face recognition. Based on Fisher Discriminant Analysis (FDA), NPFDA takes into account the local geometry structure information, changes the objective function. Thus, two abilities of manifold learning and classification are combined into the proposed method. In order to improve the discriminating power, Schur-decomposition is used to get the orthogonal basis vectors. Experimental results on the Yale face database and Feret face database demonstrate the effectiveness of the proposed method.

Key words: Fisher criterion, neighborhood preserving projections, subspace learning, schur-decomposition, face recognition

INTRODUCTION

During the past two decades, subspace-based face recognition has been extensively studied and many methods have been proposed. The most popular methods include Principal Component Analysis (PCA) (Turk and Pentland, 1991) and Fisher Discriminant Analysis (FDA) (Belhumeur *et al*., 1997). PCA performs dimensionality reduction by projecting the original high-dimensional data to a low-dimensional linear subspace spanned by the leading eigenvectors of a covariance matrix. Thus, PCA builds a global linear model of the data. PCA is unsupervised learning algorithm, whereas FDA is a supervised learning method. FDA searches for the projective axes on which the data points of different classes are far from each other while constraining the data points of the same class to be close to each other as possible. While these two methods have attained reasonably good performance in face recognition, they may fail to discover the underlying nonlinear manifold structure as they seek only a compact Euclidean subspace for efficient face representation and recognition.

Recently, a number of nonlinear manifold learning algorithms have been proposed to discover the local geometry property of high-dimensional feature spaces. The most representative methods include Isomap (Tenenbaum *et al*., 2000), Locally Linear Embedding (LLE) (Saul *et al*., 2003), Laplacian Eigenmap (Belkin and Niyogi, 2002), Local Tangent Space Alignment (LTSA) (Zhang and Zha, 2004), Local Coordinates Alignment (LCA) (Zhang *et al*., 2008) and Local Spline Embedding (LSE) (Xiang *et al*., 2009). All of them attempt to embed the original data into a submanifold by preserving the local geometry structure. But they yield maps that are defined only on the training data and this issue that how to map new test data to the low dimensional space remains difficult. Linearization, kernalization and tensorization are some often used techniques to deal with the problem (Yan *et al*., 2007). For example, the most representative such algorithm is Neighborhood Preserving Projections (NPP) (Pang *et al*., 2005a, b) which is a linear approximation of LLE and the testing data can be explicitly mapped to the learned subspace. However, it deemphasizes discriminant information which may sometimes make it not suitable for recognition task.

Earlier works based on FDA suffer from not preserving the local manifold of the face structure, whereas the research works on NPP lack to preserve global features of face images. In this study, we proposed a new subspace learning method called Neighborhood Preserving Fisher Discriminant Analysis (NPFDA) which extends the original Fisher Discriminant Analysis by preserving the locality structure of the data. Our method effectively combines FDA and NPP, i.e., it can hold the strong discriminating power of FDA while preserve the intrinsic geometry relation of the data samples. In order to improve the discriminating power, Schur-decomposition is used to obtain a set of orthogonal basis eigenvectors.

A BRIEF REVIEW OF FDA AND NPP

The generic problem of linear subspace learning is the following. Let $X = [x_1, x_2, ..., x_N]$ be a set of face image vectors and $x_i \in R^D$ ($i = 1, 2, ..., N$). Each image x_i belongs to

Corresponding Author: Guoqiang Wang, Department of Computer and Information Engineering, Luoyang Institute of Science and Technology, 471023 Luoyang, Henan, People's Republic of China

one of c classes $\{X_1, X_2,..., X_c\}$. The problem is to find a projection transformation matrix W that map high-dimensional data set X to a low-dimensional vector $Y = [y_1, y_2,..., y_N]$, where $y_i = W^T x_i \in R^d$ (d<<D). Here, we briefly review how the FDA and NPP algorithms realize subspace learning.

Fisher discriminant analysis: FDA is one of the most popular subspace learning techniques which tries to find the subspace that best discriminates different classes by maximizing the between-class scatter matrix S_b while minimizing the within-class scatter matrix S_w in the projective subspace. S_b and S_w can be defined as:

$$S_b = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (1)$$

$$S_w = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \quad (2)$$

where, μ is the mean image of all images, μ_i is the mean image of class X_i and N_i is the number of samples in class X_i.

The objective function of FDA is as follows:

$$W_{opt} = \underset{W}{\arg\max} \frac{|W^T S_b W|}{|W^T S_w W|} = [w_1 \ w_2 \ \cdots \ w_d] \quad (3)$$

where, $\{w_i | i = 1, 2,..., d\}$ is the set of generalized eigenvectors of S_b and S_w corresponding to the d largest generalized eigenvalues $\{\lambda_i | i = 1, 2,..., d\}$, i.e.:

$$S_b w_i = \lambda_i S_w w_i, \ i = 1, 2,..., m \quad (4)$$

Note that there are at most c-1 nonzero generalized eigenvalues and so an upper bound on d is c-1, where c is the number of classes.

Neighborhood preserving projections: NPP is a linear approximation of the nonlinear Local Linear Embedding for learning a locality preserving subspace which preserves the intrinsic geometry relation of the data and local structure.

Simliarly to LLE (Saul *et al*., 2003), NPP first constructs a neighborhood graph for X based on either the K nearest neighbor or ε neighborhood criterion. Then, the affinity matrix S can be obtained by minimizing the reconstruction error:

$$\phi(S) = \sum_{i=1}^{N} \left\| x_i - \sum_{j=1}^{K} s_{ij} x_j \right\|^2 \quad (5)$$

With constrains:

$$\sum_j s_{ij} = 1$$

and $S_{ij} = 0$ if x_{ij} is not of the k nearest neighbor x_i.

The basis idea behind NPP is that the same weight S_{ij} that reconstructs the point x_i in D-dimensional space should also reconstruct its image y_i in d-dimensional space. The projection tansformation matrix W can be obtained by solving the following minimizing problem:

$$\min_W \sum_{i=1}^{N} \left\| y_i - \sum_{j=1}^{K} s_{ij} y_j \right\|^2 \quad (6)$$

With the constaint:

$$\frac{1}{N} \sum_{i=1}^{N} y_i y_i^T = I,$$

where, I is a I identy matrix.

By simple algebra formulation, the minimization problem reduces to finding:

$$\begin{aligned} &\arg\min W^T XMX^T W \\ &W^T XX^T W = I \end{aligned} \quad (7)$$

where,

$$M = (1-S)(1-S)^T$$

Finally, the column vector of transformation matrix W that minimizes the objective functions is given by the minimum eigenvalues solution to the following generalized eigenvalue problem:

$$XMX^T W = \lambda XX^T W \quad (8)$$

NEIGHBORHOOD PRESERVING FISHER DISCRIMINANT ANALYSIS

Here, we describe Neighborhood Preserving Fisher Discriminant Analysis (NPFDA) algorithm that learns a locality preserving and global discriminating subspace. The proposed NPFDA approach combines global feature preservation technique FDA and local feature preservation technique NPP to form the high-quality feature subspace.

Justification: Since NPFDA is designed to hold both the linear global character and the nonlinear local character and also possess the discriminant ability, we can innovate a heuristic object function by combination the FDA and NPP:

$$J(W) = \max \frac{W^T S_b W}{W^T S_w W} - W^T XMX^T W \quad (9)$$

where, S_b is the between-class scatter matrix, S_w is the within-class scatter matrix. X is the image set and W is the projection direction. The definition of matrix M is the same as in neighbourhood presering projection. Since NPP wants to find a projection direction W to make W^T as small as possible, we can instead here choose $-W^T XMX^T W$ and make it as large as possible in the low dimension space.

Alternatively, we can reformulate Eq. 9 as:

$$\begin{aligned} J(W) &= \max W^T S_b W - W^T XMX^T W \\ \text{s.t. } & W^T S_w W = I \end{aligned} \quad (10)$$

The constrained maximization problem, Eq. 10 can be done using the method of Lagrange multipliers:

$$L(W, \lambda) = (W^T (S_b - XMX^T)W) - \lambda (I - W^T S_w W)$$

where, λ is the Lagrange multiplier. Setting the gradients with respect to W to zero, thus we can get:

$$(S_b - XMX^T)W = \lambda S_w W \quad (11)$$

By defining $L = S_b - XMX^T$, $C = S_w$, Eq. 11 can be rewritten in the form of a generalized eigenvalue problem:

$$LW = \lambda CW \quad (12)$$

Obtaining orthogonal eigenvectors: The generalized eigenvectors obtained by solving Eq. 12 are nonorthogonal. This makes it difficult to faithfully represent the data. Here, we introduce Schur decomposition to get orthogonal basis vectors.

Instead of performing eigenanalysis on the matrix $C^{-1} L$, Schur-decomposition is performed on $C^{-1} L$. Suppose the Schur decomposition of $C^{-1} L$ is $C^{-1} L = UTU^T$, where $U = [u_1, u_2,..., u_D]$ is an orthogonal matrix, T is a quasi-upper-diagonal matrix with the real eigenvalues of the matrix $C^{-1} L$ on the diagonal. Assume $u_1, u_2,..., u_d$ to be Schur vectors of $C^{-1} L$ corresponding to the first d largest real eigenvalues. It is obvious that $u_1, u_2,..., u_d$ are orthogonal to each other. $U = [u_1, u_2,..., u_d]$ is an optimal solution of the maximization problem Eq. 10. The following theorem reveals the fact:

Theorem 1: Suppose $u_1, u_2,..., u_d$ to be discriminant Schur vectors of NPFDA. Thus, we have:

$$\begin{aligned} J([u_1, u_2, \cdots, u_d]) &= tr([u_1, u_2, \cdots, u_d]^T C^{-1} L [u_1, u_2, \cdots, u_d]) \\ &= \max_{W \in R^{D \times d}} J(W) \end{aligned} \quad (13)$$

Proof: Since $u_1, u_2,..., u_d$ are Schur vectors of the matrix $C^{-1} L$ corresponding to the first d largest real eigenvalues, we have:

$$C^{-1} L u_i = \lambda_i u_i \; i = 1, 2,..., d \quad (14)$$

where, λ_i is the ith largest real eigenvalue of the matrix $C^{-1} L$. From the formula (14), it follows that:

$$[u_1, u_2,..., u_d]^T C^{-1} L [u_1, u_2,..., u_d] = \text{diag} [\lambda_1, \lambda_2,..., \lambda_j] \quad (15)$$

Thus, we have:

$$\begin{aligned} J([u_1, u_2, \cdots, u_d]) &= tr([u_1, u_2, \cdots, u_d]^T C^{-1} L [u_1, u_2, \cdots, u_d]) \\ &= \prod_{j=1}^{d} \lambda_j = \min_{W \in R^{D \times d}} J(W) \end{aligned} \quad (16)$$

The outline of NPFDA: The algorithm procedure of NPFDA can be summarized as follows:

Step 1: Compute the matrix M

- For each data point x_i, determine its k nearest neighbors by KNN or ε-ball algorithm
- Compute the weights S_{ij} that best linearly reconstruct x_i from its neighbors, solving the constrained least-squares problem in Eq. 5
- Let $M = (1-S)(1-S)^T$

Step 2: Compute the between-class scatter matrix S_b and the within-class scatter matrix S_w

Step 3: Let $L = S_b - XMX^T$, $C = S_w$, Schur decomposition is performed on the matrix $C^{-1} L$ and obtain othogonal basis vectors $U = [u_1, u_2,..., u_d]$

Step 4: Obtain the embedding Y in R^d using $Y = U^T X$

Step 5: Classify the embedding results using a suitable classifier

EXPERIMENTS AND DISCUSSION

Here, we conduct several experiments on different databases (YALE face database and FERET face database) to demonstrate the effective and robustness of our proposed method named Neighborhood Preserving Fisher Discriminant Analysis (NPFDA). We have also compared the proposed method with four popular methods including PCA (Turk and Pentland, 1991), FDA (Belhumeur *et al*., 1997), NPP (Yan *et al*., 2007; Pang *et al*., 2005b), LPP (He and Niyogi, 2003; He *et al*., 2005). We applied the nearest neighborhood classifier with Euclidean metric for recognition. To have a fair comparison, all the results reported here are based on the best tuned parameters of all the compared method.

Fig. 1: Sample face images from the FERET database

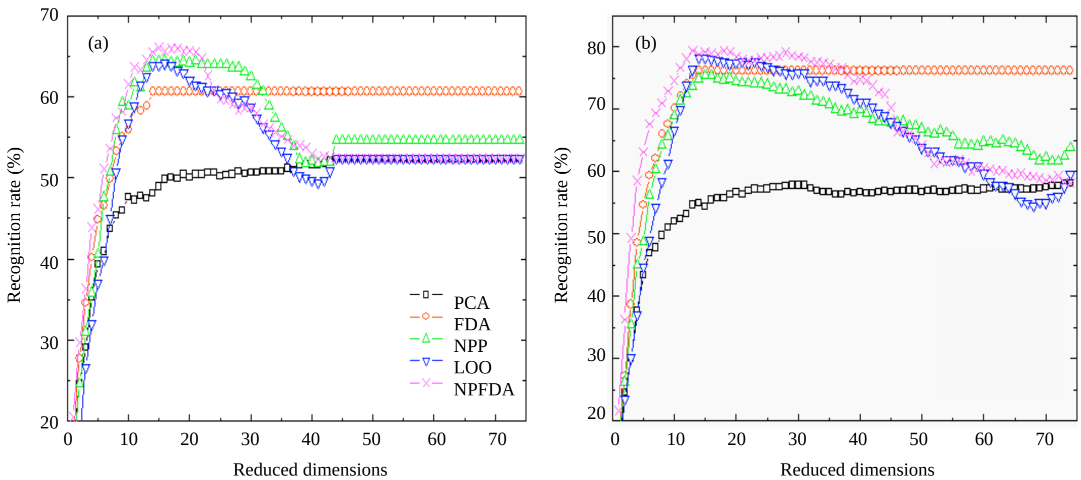

Fig. 2 (a-b): Recognition rate vs. dimensionality reduction on Yale face database: the left sub-figure is achieved by selecting 3 images person for training and the right sub-figure is achieved by selecting 5 images person for training

Table 1: Recognition accuracy (%) comparison on Yale face database

Method	3Train	5Train
PCA	52.25±2.78 (44)	58.11±3.36 (74)
FDA	60.58±3.67 (14)	76.11±2.47 (14)
NPP	64.50±3.15 (16)	75.33±3.05 (15)
LPP	64.08±3.37 (16)	78.22±3.11 (14)
NPFDA	66.00±3.74 (15)	79.33±3.06 (18)

YALE: The YALE face database contains 165 grayscale images of 15 individuals. There are 11 images per subject and these images demonstrate variation in lighting condition (center-light, left-light, right-light) and facial expression (happy, surprised, wink, sad, sleepy and normal). All face images are resized to 32×32 for computation efficiency in our experiments. Figure 1 shows the sample images of two individuals. For each people, l (= 3, 5)images are randomly selected for training and the rest are used for testing. The random selection is repeated 10 times. The plot of the average recognition rate versus subspace dimensions of all methods are shown in Fig. 2 and the best recognition results and the corresponding reduced dimensions obtained by each method is listed in Table 1. As can be seen, NPFDA algorithm outperforms the other algorithms involved in this experiment. This reason is that NPFDA considers the class label information and preserves the intrinsic structure from the raw face images, in addition, NPFDA obtains the orthogonal basis eigenvectors. Thus, it can produce more discriminative embedding results.

FERET: The subset of FERET face database contains 100 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”. This subset involves two facial expression images, two left pose images, two right pose images and an illumination image. All the images in the subset are to be the size of 40×40. Figure 3 shows sample images of two persons. Simliarly to the strategy adopted on Yale, l = (= 3, 5) images per person are randomly selected for training and the rest are used for testing. All the tests are repeated over 10 random splits independently and then the average recognition results are calculated. The recognition results are shown in

Fig. 3: Sample face images from the FERET database

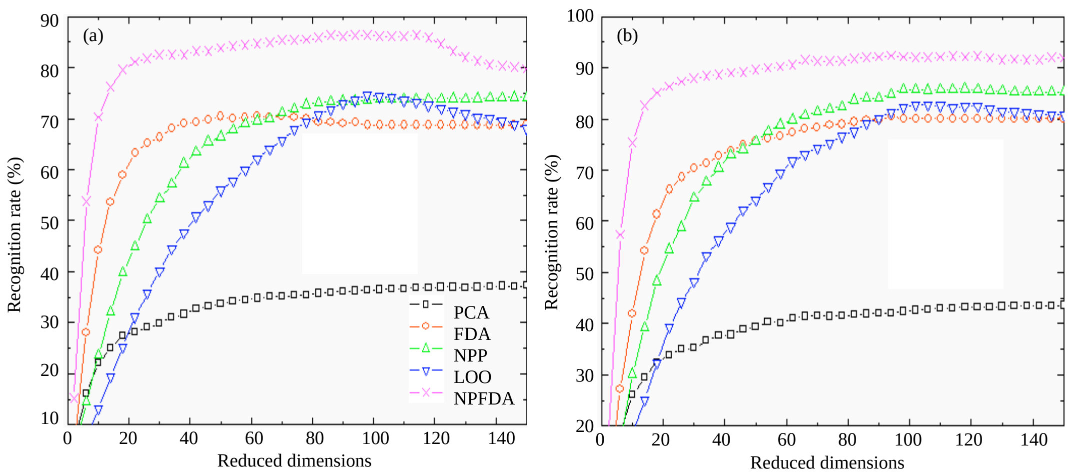

Fig. 4 (a-b): Recognition rate vs. dimensionality reduction on FERET face database: the left sub-figure is achieved by selecting 3 images person for training and the right sub-figure is achieved by selecting 5 images person for training

Table 2: Recognition accuracy (%) comparison on FERET face database

Method	3Train	5Train
PCA	38.00±1.64 (299)	44.85±3.02 (499)
FDA	70.58±1.79 (70)	80.45±2.89 (94)
NPP	74.17±2.43 (146)	85.85±3.51 (118)
LPP	74.52±1.82 (98)	82.75±2.98 (102)
NPFDA	86.35±1.82 (94)	92.25±1.87 (94)

Fig. 4 and Table 2. We can draw a similar conclusion as before.

CONCLUSIONS AND FUTURE WORK

In this study, a novel subspace learning algorithm named Neighborhood Preserving Fisher Discriminant Analysis (NPFDA) is proposed. NPFDA makes use of local manifold structure information and discriminant information. In some sense, the proposed method can be regarded as a combination of NPP and FDA. On the other hand, NPFDA uses Schur decomposition to get othogonal bases of the face subspace. Experimental results show that the proposed method is indeed effective and efficient. Our future work is to extend NPFDA to nonlinear form by kernel trick.

ACKNOWLEDGMENTS

This study is supported by Foundation of He'nan Educational Committee (number 2010A520028).

REFERENCES

Belhumeur, P.N., J.P. Hespanha and D.J. Kriegman, 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intel., 19: 711-720.

Belkin, M. and P. Niyogi, 2002. Laplacian eigenmaps and spectral techniques for embedding and clustering. Adv. Neural Inform. Process. Syst., 14: 485-491.

He, X.F. and P. Niyogi, 2003. Locality preserving projections. Proc. Adv. Neural Inform. Process. Syst. Conf., 16: 153-160.

He, X., S. Yan, Y. Hu P. Niyogi and H.J. Zhang, 2005. Face recognition using laplacianfaces. IEEE. Trans. Pattern Anal. Machine Intell., 27: 328-340.

Pang, Y., L. Zhang, Z. Liu, N. Yu and H. Li, 2005a. Neighborhood Preserving Projections (NPP): A novel linear dimension reduction method. Lecture Notes Comput. Sci., 3644: 117-125.

Pang, Y., N. Yu, H. Li, R. Zhang and Z. Liu, 2005b. Face recognition using neighborhood prserving projections. Lecture Notes Comput. Sci., 3768: 854-864.

Saul, L.K., S.T. Roweis and Y. Singer, 2003. Think globally, fit locally: Unsupervised learning of low dimensional manifolds. J. Machine Learn. Res., 4: 119-155.

Tenenbaum, J., V. de Silva and J. Langford, 2000. A global geometric framework for nonlinear dimensionality reduction. Science, 290: 2319-2323.

Turk, M. and A. Pentland, 1991. Eigenfaces for recognition. J. Cognitive Neurosci., 3: 71-86.

Xiang, S., F. Nie, C. Zhang and C. Zhang, 2009. Nonlinear dimensionality reduction with local spline embedding. IEEE Trans. Knowledge Data Eng., 21: 1285-1298.

Yan, S., D. Xu, B. Zhang, H.J. Zhang, Q. Yang and S. Lin, 2007. Graph embedding and extensions: A general framework for dimensionality reduction. IEEE Trans. Pattern Anal. Machine Intell., 29: 40-51.

Zhang, T., X. Li, D. Tao and J. Yang, 2008. Local Coordinates Alignment (LCA): A novel method for manifold learning. J. Pattern Recogn. Artif. Intell., 22: 667-690.

Zhang, Z. and H. Zha, 2004. Principle manifolds and nonlinear dimensionality reduction via tangent space aligenment. Siam J. Sci. Comput., 26: 313-338.