2009 Second International Conference on Intelligent Computation Technology and Automation

Orthogonal Discriminant Neighborhood Preserving Projections for Face Recognition

Guoqiang Wang, Xiaojing Hou
Department of Computer and Information Engineering
Luoyang Institute of Science and Technology
Luoyang, China
wgq2211@163.com

***Abstract*—Subspace learning is one of the main directions for face recognition. In this paper, a novel subspace learning approach, called Orthogonal Discriminant Neighborhood Preserving Projections (ODNPP), is proposed for robust face recognition. The aim of ODNPP is to preserve the within-class geometric structure, while maximizing the between-class scatter. In order to improve the discriminating power, Schur decomposition is used to obtain the orthogonal basis eigenvectors. Experiment results on ORL face database and Yale face database demonstrate the effectiveness and robustness of the proposed method.**

Keywords- face recognition; subspace learning; within-class geometric structure; between-class scatter; Schur decomposition

I. INTRODUCTION

Face recognition has become an active research area in computer vision and pattern recognition, due to military, commercial and law enforcement applications. Many face recognition methods have been developed over the past few decades. Among these methods, the so-called subspace-based approach is one of the most effective and well-studied techniques to face recognition. Principal Component Analysis (PCA) [1] and Linear Discriminant Analysis (LDA) [2] are two of the most popular linear subspace methods that can extract useful features with low computational complexity.

Recently, many research efforts have shown that the face images possibly reside on a nonlinear submanifold [3,4,5]. But both PCA and LDA see only the linear manifold that based on the Euclidean structure. They fail to capture the underlying structure which lies on a nonlinear submanifold hidden in the image space. Several nonlinear techniques were proposed to discover the nonlinear structure of the manifold such as Locally Linear Embedding (LLE) [4], Isometric Mapping (ISOMAP) [5], Laplacian Eigenmap (LE) [6]and Local Tangent Space Alignment (LTSA)[7]. These methods are defined only on training data, and the issue of how to map new test data remains difficult. Therefore, they cannot be applied directly to face recognition task. But their properties of locality preserving are fairly significative. In order to overcome this drawback, Pang et al. proposed a linear subspace method, named Neighborhood Preserving Projection (NPP) [8] which aims at preserving the local manifold structure. Although NPP is successful in many domains, it nevertheless suffers from disadvantages. Firstly, it deemphasizes discriminant information that is very important for recognition task. Secondly, the basis functions obtained by NPP are not orthogonal. This makes it difficult to reconstruct the data.

In this paper, a new subspace method, called Orthogonal Discriminant Neighborhood Preserving Projections (ODNPP). Based on NPP, ODNPP takes into account the class labels to preserve the within-class neighboring geometry of the image space and introduces the between-class scatter into the objective function. On the other hand, ODNPP uses Schur decomposition to get the orthogonal bases of the face space.

II. NEIGHBORHOOD PRESERVING PROJECTIONS

Let $X=[x_1,x_2,\cdots,x_N]$ be a set of face image vectors, and $x_i \in R^D$ $(i=1,2,\cdots,N)$. Linear dimensionality reduction techniques try to find a transformation matrix $A=[a_1,a_2,\cdots,a_d]\in R^{D\times d}$ that maps the set X of these N samples to the set $Y=[y_1,y_2,\cdots,y_N]$, where $y_i = A^{\mathrm{T}} x_i \in R^d (d << D)$.

NPP is a linear approximation to LLE[4,8]. NPP finds a linear transformation matrix A by minimizing the following objective function

$$\min_A \sum_{i=1}^{N} \left\| y_i - \sum_{j=1}^{k} W_{ij} y_j \right\|^2 \tag{1}$$

with the constraint

$$(1/N)\sum_{i=1}^{N} y_i y_i^T = I \tag{2}$$

where the W is a weight matrix, I is a $(d\times d)$ identity matrix. Since LLE seeks to preserve the intrinsic geometric properties of the local neighborhoods, it assumes that the same weights which reconstruct the point x_i by its neighbors in the high dimensional space, will also reconstruct is image y_i, by its corresponding neighbors, in the low dimensional

978-0-7695-3804-4/09 $26.00 © 2009 IEEE
DOI 10.1109/ICICTA.2009.162

space. Then, the weights can be computed by minimizing the following objective function

$$\min \sum_{i=1}^{N} \left\| x_i - \sum_{j=1}^{k} W_{ij} x_j \right\|^2 \tag{3}$$

with the constraints $\sum_{j=1}^{k} W_{ij} = 1$, and $W_{ij} = 0$ if x_i is not one of the k nearest neighbors of x_i.

Finally, the minimization problem can be converted to solving a generalized eigenvalue problem as follows:

$$XMX^{\mathrm{T}} A = \lambda XX^{\mathrm{T}} A \tag{4}$$

where $M = (I - W)^{\mathrm{T}} (I - W)$, $I = diag(1,1,\cdots,1)$.

XX^{T} is non-singular after some pre-processing steps on X in NPP, thus, the basis function of NPP can also be regarded as the eigenvectors of the matrix $(XX^{\mathrm{T}})^{-1} XMX^{\mathrm{T}}$ associated with the smallest eigenvalues. Since $(XX^{\mathrm{T}})^{-1} XMX^{\mathrm{T}}$ is not symmetric in general, the basis function of NPP are non-orthogonal. Therefore, NPP can not preserve the metric structure of the high-dimensional space.

Once the eigenvectors are computed, let $A_d = [a_1, a_2, \cdots, a_d]$ be the transformation matrix. Thus, the Euclidean distance between two data points in the reduced space can be computed as follows:

$$\begin{aligned} dist(y_i, y_j) &= \left\| y_i - y_j \right\| \\ &= \left\| A^{\mathrm{T}} x_i - A^{\mathrm{T}} x_j \right\| \\ &= \left\| A^{\mathrm{T}} (x_i - x_j) \right\| \\ &= \sqrt{(x_i - x_j)^{\mathrm{T}} A A^{\mathrm{T}} (x_i - x_j)} \end{aligned} \tag{5}$$

If A is an orthogonal matrix, $AA^{\mathrm{T}} = I$ and the metric structure is preserved.

III. Orthogonal Discriminant Neighborhood Preserving Projections

NPP only put stress on preserving neighbor geometric structure of the raw data samples. However, it ignores the crucial discriminant class information. In addition, the projection matrix obtained by NPP algorithm is non-orthogonal matrix, NPP fails to preserve the metric structure of the nonlinear submanifold space. In order to improve recognition performance of NPP, we proposed a novel subspace method for face recognition, i.e., ODNPP. In our framework, neighboring geometric relations are preserved according to prior class label information, between-class scatter constraint is added into the objective function of NPP, and the orthogonal basis vectors of the face subspace are obtained by Schur decomposition. Therefore, the objection function of ODNPP is defined as follows:

$$J = \frac{\sum_{i=1}^{N} \left\| y_i - \sum_{j=1}^{N} W_{ij} y_j \right\|^2}{\sum_{i=1}^{C} n_i \left\| m_i - m \right\|^2} \tag{6}$$

where W is a weight matrix. Since class labels are available, each data point is reconstructed by the linear combination of other points which belong to the same class. In other words, the weights w_{ij} is computed by minimizing (3) with constraints $\sum_{j=1}^{k} W_{ij} = 1$, and $W_{ij} = 0$ if x_i and x_j are from different classes. It means that the within-class geometric relation is emphasized. As a result, the weight matrix W not only reflects local geometry relation but also carries discriminative information. Note that in ONPDP, one does not need to set the parameter k, the number of nearest neighbors. C is the number of face classes, n_i is the number of samples in the ith class, $m_i = (1/n_i) \sum_{y_j \in X_i} y_j$, $m = (1/N) \sum_{i=1}^{N} y_i$.

Recall that $y_i = A^T x_i$, the objective function can be reduced to:

$$\begin{aligned} J(A) &= \frac{\sum_{i=1}^{N} \left\| A^T x_i - \sum_{i=1}^{N} W_{ij} A^T x_j \right\|^2}{\sum_{i=1}^{C} \left\| (1/n_i) \sum_{x_k} A^T x_k - (1/N) \sum_{i=1}^{N} A^T x_i \right\|^2} \\ &= \frac{tr(A^T XMX^T A)}{tr(A^T \sum_{i}^{C} n_i \left\| (1/n_i) \sum_{x_k \in X_i} x_k - (1/N) \sum_{i=1}^{N} x_i \right\|^2 A)} \\ &= \frac{tr(A^T S_M A)}{tr(A^T S_B A)} \end{aligned} \tag{7}$$

where the symbol "tr" denotes the operation of trace, $S_M = XMX^T$, $S_B = \sum_{i}^{C} n_i \left\| (1/n_i) \sum_{x_k \in X_i} x_k - (1/N) \sum_{i=1}^{N} x_i \right\|^2$.

The projection matrix $A = [a_1, a_2, \cdots, a_d]$ that optimizes the objective function can be obtained by solving the generalized eigenvalue problem:

$$S_M a_i = \lambda_i S_B a_i, \lambda_1, \lambda_2, \cdots, \lambda_d \quad (8)$$

The basis vectors of (8) are non-orthogonal. Here, we introduce Schur decomposition to get orthogonal basis vectors. Suppose the Schur decomposition of $S_B^{-1}S_M$ is $S_B^{-1}S_M = UTU^{\mathrm{T}}$, where $U = [u_1, u_2, \cdots, u_D]$ is an orthogonal matrix, T is a quasi-upper-diagonal matrix with the real eigenvalues of the matrix $S_B^{-1}S_M$ on the diagonal. Assume $u_1, u_2, \cdots, u_d$ to be Schur vectors of $S_B^{-1}S_M$ corresponding to the first d smallest real eigenvalues. It is obvious that $u_1, u_2, \cdots, u_d$ are orthogonal to each other.

Theorem 1. Suppose $u_1, u_2, \cdots, u_d$ to be discriminant vectors of ODNPP. Thus, we have

$$J([u_1, u_2, \cdots, u_d]) = \frac{\left|[u_1, u_2, \cdots, u_d]^{\mathrm{T}} S_M [u_1, u_2, \cdots, u_d]\right|}{\left|[u_1, u_2, \cdots, u_d]^{\mathrm{T}} S_B [u_1, u_2, \cdots, u_d]\right|}$$

$$= \min_{A \in R^{D \times d}} J(A) \quad (9)$$

Proof. Since $u_1, u_2, \cdots, u_d$ are Schur vectors of the matrix $S_B^{-1}S_M$ corresponding to the first d smallest real eigenvalues, we have

$$S_B^{-1} S_M u_i = \lambda_i u_i \; i = 1, 2, \cdots, d \quad (10)$$

where λ_i is the ith smallest real eigenvalue of the matrix $S_B^{-1}S_M$. From the formula (10), it follows that

$$[u_1, u_2, \cdots, u_d]^{\mathrm{T}} S_M [u_1, u_2, \cdots, u_d]$$

$$= [u_1, u_2, \cdots, u_d]^{\mathrm{T}} S_B [u_1, u_2, \cdots, u_d] diag(\lambda_1, \lambda_2, \cdots, \lambda_d) \quad (11)$$

Thus, we have

$$J([u_1, u_2, \cdots, u_d]) = \frac{\left|[u_1, u_2, \cdots, u_d]^{\mathrm{T}} S_M [u_1, u_2, \cdots, u_d]\right|}{\left|[u_1, u_2, \cdots, u_d]^{\mathrm{T}} S_B [u_1, u_2, \cdots, u_d]\right|}$$

$$= \prod_{j=1}^{d} \lambda_j = \min_{A \in R^{D \times d}} J(A) \quad (12)$$

IV. EXPERIMENTS

In this section, experiments are conducted on ORL face database [9] and Yale face database [10] to evaluate the performance of the proposed ODNPP algorithm. ODNPP algorithm is compared with Eigenface, Fisherface and NPP.

For all the experiments, the images are cropped based on the centers of eyes, and the cropped images are normalized to 32×32 pixel arrays with 256 gray levels per pixel. Each image can be represented by 1024-dimensional vector in image space. For its simplicity, the nearest-neighbor classifier using Euclidean metric is employed.

A. Experiments on the ORL database

The ORL database contains 400 images of 40 individuals. Each individual has 10 images, which are captured at different times, different lighting conditions, different facial expressions and facial accessories (glasses/no glasses). The images were taken with a tolerance for some tilting and rotation of the face up to 20 degrees. 10 sample images of one individual in the ORL database are displayed in Fig. 1.

For each individual, we randomly select i $(i = 3, 4, 5, 6)$ samples for training, and the rest are used for testing. All the results are 10 times averaged. The best performances obtained by these algorithms as well as the corresponding dimensionality of the optimal subspace are showed in Table I.

B. Experiments on the Yale database

The Yale face database contains 165 images, with variations in lighting conditions (left-light, center-light, right-light), facial expressions (normal, happy, sad, sleepy, surprised, and wink), and with/without glasses, from 15 individuals. Fig. 2 shows 11 sample images of one individual.

In the Yale database, i $(i = 3, 4, 5, 6)$ images of each individual are randomly selected for training, while the remaining images are used for testing. We repeat this process 10 times and calculate the average recognition rates. The best performances of these algorithms as well as the corresponding dimensionality of the optimal subspace are shown in Table II.

C. Discussions

From Table I and Table II, it is very obvious that the ODNPP method outperforms the other methods. In addition, we can see that for all methods, the recognition accuracy increases as the number i of training samples per individual increase. From the above two experiments, several points are worthwhile to emphasize:

- Compared to PCA, LDA which attempt to preserve the global Euclidean structure, ODNPP aims at preserving the local neighboring geometry structure, which is more important than the global Euclidean structure in many practical classification problems. Moreover, ODNPP utilizes neighboring class information to enhance the discriminative ability.

- Different from NPP which aim to preserving the local manifold structure, ODNPP preserves the within-class neighboring geometry while maximizing the between-class scatter. Furthmore, ODNPP introduces Schur decomposition to obtain a set of orthogonal basis vectors and does not suffer from the problem of metric distortion. Therefore, ODNPP has stronger recognition ability.
- Like LDA and NPP, ODNPP applies PCA pre-processing to perform dimensionality reduction and avoid small-sample-size (SSS) problem. Moreover, the most noise can be removed by PCA algorithm.

V. Conclusions

In this paper, we present a novel subspace learning method, called orthogonal discriminant neighborhood preserving projections (ODNPP). ODNPP has two prominent characteristics. First, it integrates the within-class neighboring information and the between-class scatter information, which enables it to achieve strong discriminative ability. Second, it uses Schur decomposition to get orthogonal subspace, which further improves recognition performance. Experimental results on two well-known face databases demonstrate the effectiveness and robustness of the proposed ODNPP algorithm. ODNPP is a linear method in nature. In future study, we will generalize the ODNPP to kernel ODNPP by introducing a kernel skill [11].

References

[1] M. Turk and A. Pentland, "Eigenface for recognition," Journal of Cognitive Neuoscience, vol. 3, winter 1991, pp. 71-86, doi:10.1162/jocn.1991.3.1.71.

[2] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," IEEE Trans. on PAMI, vol. 19, Jul. 1997, pp. 711-720, doi:10.1109/34.598228.

[3] Y. Chang, C. Hu, and M. Turk, "Manifold of Facial Expression," Pro. IEEE Int'l Workshop Analysis and Modeling of Faces and Gestures, IEEE Press, Oct. 2003, pp. 28-35, doi:10.1109/AMFG.2003.1240820.

[4] S. L. Roweis, L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science, vol. 290, Dec. 2000, pp. 2323-2326, doi: 10.1126/science.290.5500.2323.

[5] J. B. Tenenbaum, V. de Silva, J. C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," Science, vol. 290, Dec. 2000, pp. 2319-2323, 10.1126/science.290.5500.2319.

[6] M. Belkin, P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representaiton," Neural Computation, vol. 15, jun. 2003, pp.1373-1396, doi: 10.1162/089976603321780317.

[7] Z. Y. Zhang, H. Y. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment," SIAM Journal of Scientific Computing, vol. 26, 2005, pp. 313-338, doi: 10.1137/S1064827502419154.

[8] Y. W. Pang, L. Zhang, Z. K. Liu, "Neighborhood Preserving Projections (NPP): A Novel Linear Dimension Reduction Method," Proc. of International Conference on Intelligent Computing, Springer-Verlag Press, Sep. 2005, pp.117-125, doi:10.1007/11538059.

[9] Olivetti Research Laboratory (ORL) Database, http:// www.cl.cam.ac.uk/ research/dtg/attarchive/facedatabase.html, 2002.

[10] Yale Univ. Face Database, http://cvc.yale.edu/ projects/yalefaces/ yalefaces.html, 2002.

[11] V. Vapnik, The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.

Figure 1. Sample face images for one individual of the ORL database

TABLE I. Recognition Accuracy (%) of Different Algorithms on The ORL Database

size	Eigenface	Fisherface	NPP	ODNPP
3	79.1±3.29(119)	86.4±2.81(39)	87.3±2.74(41)	91.4±2.51(39)
4	81.4±2.65(159)	90.8±2.04(39)	90.6±2.13(39)	93.9±1.00(39)
5	89.7±1.84(199)	93.5±1.48(39)	94.1±1.17(39)	96.7±1.00(39)
6	89.6±2.35(239)	94.6±1.25(39)	95.2±1.41(41)	96.9±1.65(39)

Figure 2. Sample face images for one individual of the Yale database

TABLE II. Recognition Accuracy (%) of Different Algorithms on The Yale Database

size	Eigenface	Fisherface	NPP	ODNPP
3	51.5±3.99(44)	62.0±4.30(14)	67.4±4.63(16)	68.8±6.88(14)
4	56.0±3.83(59)	70.8±4.88(14)	74.0±5.17(15)	74.8±5.29(16)
5	57.3±4.60(74)	71.4±5.18(14)	74.1±3.17(14)	75.1±2.29(14)
6	62.2±4.17(89)	77.2±3.46(14)	80.6±3.22(17)	81.7±3.55(14)