Author’s Accepted Manuscript

Bagging null space locality preserving discriminant classifiers for face recognition

Liping Yang, Weiguo Gong, Xiaohua Gu, Weihong Li

PII: S0031-3203(08)00424-X
DOI: doi:10.1016/j.patcog.2008.10.014
Reference: PR 3356

To appear in: Pattern Recognition

Received date: 14 August 2008
Revised date: 10 October 2008
Accepted date: 13 October 2008

Cite this article as: Liping Yang, Weiguo Gong, Xiaohua Gu and Weihong Li, Bagging null space locality preserving discriminant classifiers for face recognition, Pattern Recognition (2008), doi:10.1016/j.patcog.2008.10.014

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Bagging null space locality preserving discriminant classifiers for face recognition

Liping Yang, Weiguo Gong*, Xiaohua Gu, Weihong Li, Yanfei Liu

Key Laboratory of Opto-Electronic Technology and Systems of Ministry of Education, Chongqing University, Chongqing 400030, China

Abstract

In this paper, we propose a novel bagging null space locality preserving discriminant analysis (bagNLPDA) method for facial feature extraction and recognition. The bagNLPDA method first projects all the training samples into the range space of a so-called locality preserving total scatter matrix without losing any discriminative information. The projected training samples are then randomly sampled using bagging to generate a set of bootstrap replicates. Null space discriminant analysis is performed in each replicate and the results of them are combined using majority voting. As a result, the proposed method aggregates a set of complementary null space locality preserving discriminant classifiers. Experiments on FERET and PIE subsets demonstrate the effectiveness of bagNLPDA.

Key words: Locality preserving, bagging, discriminant analysis, small sample size problem, face recognition

1 Introduction

Subspace methods for face recognition have been extensively studied during last two decades [1,2,3,4]. Principle components analysis (PCA) and linear discriminant analysis (LDA) are the two most popular techniques. The PCA approach, also known as eigenface method, performs dimensionality reduction

* Corresponding author. Room 1303, Main Building, College of Opto-Electronic Engineering, Chongqing University, Chongqing 400030, China. Tel.: +86 023 65112779.

Email addresses: yanglp@cqu.edu.cn (Liping Yang), wggong@cqu.edu.cn (Weiguo Gong), xhgu@cqu.edu.cn (Xiaohua Gu), weihongli@cqu.edu.cn (Weihong Li), liuyf@cqu.edu.cn (Yanfei Liu).
by projecting the original \( n \)-dimensional data onto the \( d (d \ll n) \) dimensional linear subspace spanned by the \( d \) leading eigenvectors of the data’s covariance matrix. The LDA approach performs dimensionality reduction by searching the projection 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 as close to each other as possible. Although having achieved great success in many applications, LDA still suffers from two problems when used in face recognition: (1) LDA works effectively only on the Euclidean structure, but a number of research efforts have shown that the face images possibly reside on a nonlinear submanifold \([4,5,6,7]\); (2) the training sample set is small compared with the high dimensional facial feature vector, consequently, the constructed LDA classifier is biased and has a large variance.

To address the first problem, a lot of geometrically motivated high-dimensional data analysis approaches have been presented, e.g. Isomap \([6]\), locally linear embedding (LLE) \([7]\) and Laplacian Eigenmap \([8]\). These methods have been shown to be effective in discovering the geometrical structure of the underlying manifold. However, these methods yield maps those are defined only on the training data points. How to evaluate the maps on novel test data points remains unclear. In \([9]\), a Laplacianface method, which detects the underlying nonlinear manifold structure in the manner of linear subspace learning, was proposed. Locality preserving projections (LPP), which aims to preserve the local structure of the image space, is the core of Laplacianface method. The Laplacianface method seems to have more discriminant power than PCA approach, but it is still an unsupervised learning approach. To consider the discriminant information of recognition task, several locality preserving discriminant analysis methods have been mentioned in recent years. You et al. \([10]\) proposed a neighborhood discriminant projection (NDP) method, which explicitly considers the within-class submanifold and the between-class submanifold by integrating the within-class neighboring information and the between-class neighboring information. Hu proposed an orthogonal neighborhood preserving discriminant analysis (ONPDA) method \([5]\), which effectively combines the characteristics of LDA and LPP. Yu et al. \([11]\) presented a discriminant locality preserving projections (DLPP) method to improve the classification performance of LPP. One limitation of mentioned locality preserving methods is that PCA approach, which may destroy the local structure of data points and discard some useful information for classification, is often used before locality preserving projections or discriminant locality preserving projections.

To address the second problem, say the small sample size problem of LDA-based face recognition, many extended linear discriminant analysis methods have been proposed \([2,3,12,13,14,15]\). Belhumeur et al. \([2]\) proposed a two-stage PCA+LDA approach, i.e. Fisherface, which projects the high dimensional face data to a low dimensional feature space using PCA and then performs LDA in the low dimensional PCA subspace. Chen et al. \([3]\) suggested that the
null space of within-class scatter matrix contains the most discriminant information and proposed a null space LDA (N-LDA) method. Yu and Yang [15] presented a direct LDA (DLDA) algorithm, which avoids the small sample size problem by simultaneously diagonalizing both within-class scatter matrix and between-class scatter matrix. A common problem of all these methods is that they all lose some discriminative information, either in the principal or in the null space. In fact, the discriminative information resides in both subspaces. To use both subspaces, Yang et al. [14] proposed a complete kernel Fisher discriminant analysis algorithm, which extracts both the regular discriminant features from principal subspace and the irregular discriminant features from null subspace and fuses them using summed normalized-distance. Nevertheless, how to divide the space into the principal and the null subspaces and how to apportion a given number of features to the two subspaces are still open issues [16].

Recent researches indicate that when data are high-dimensional, having small training sample sizes compared to the data dimensionality, it may be difficult to construct a good and stable classifier [17,18,19]. Instead of developing a single optimal classifier, constructing and combining many weak classifiers in some way into a powerful decision rule is another appropriate choice. The examples of such combining techniques are bagging, boosting and the random subspace methods, which are originally designed for decision trees. In [20], by integrating asymmetric bagging-based SVM classifiers and random subspace SVM classifiers, an asymmetric bagging and random subspace SVM was built to solve the drawbacks of a single SVM classifier. In [21], random subspace and bagging were applied to Fisherface and N-LDA to enhance the face recognition performance. Both majority voting rule and sum rule were used to fuse the classification results of multiple LDA classifiers. In [22], an improved random subspace LDA method, which adopts Adaboost to feature selection before random sampling, was applied in face recognition.

Although it is efficient to combine both the local structure characteristic and the discriminative information of data points, DLPP also suffers from the small sample size problem when dealing with face recognition task. Therefore, the constructed DLPP classifier may be biased and has a large variance. Such a classifier may have a poor performance. In our previous paper [23], we presented a null space discriminant locality preserving projections (N-DLPP) algorithm. However, the null space is small when the training sample number is large. To alleviate the over-fitting problem, we propose a bagging null space locality preserving discriminant analysis (bagNLPDA) in this paper. Face images are first projected into the range space of a so-called locality preserving total scatter to reduce dimensionality without loss on discriminative information. Bagging strategy is then employed to generate a set of random bootstrap replicates by random sampling on the projected training set. Following bagging, null space discriminant analysis is implemented in each replicates. The
final decision is made by combining the results of all replicates using majority voting. Extensive experiments on FERET and PIE subsets show that the proposed bagNLPDA method alleviates the over-fitting problem of N-DLPP. Much discriminative information is covered in the null space locality preserving discriminant classifiers of each bootstrap replicate. Thus, the accuracies of face recognition are significantly improved.

2 Discriminant locality preserving projections

Given a set of properly normalized $w$-by-$h$ face images, we can form a training set of column image vectors $X = \{x_{ij}\}$, where $x_{ij} \in \mathbb{R}^{n=wh}$, by lexicographic ordering the pixel elements of image $j$ of person $i$. Let the training set contains $p$ persons and $q_i$ sample images for person $i$. The number of total training sample is $l = \sum_{i=1}^{p} q_i$.

Let $y_{ij}$ be the low dimensional feature space projection of $x_{ij}$. Discriminant locality preserving projections tries to maximize an objective function as follows [23]:

$$J = \frac{\sum_{m,n=1}^{p} (\bar{y}_m - \bar{y}_n) B_{mn} (\bar{y}_m - \bar{y}_n)^T}{\sum_{i=1}^{p} \sum_{n=1}^{q_i} (y_{im} - y_{in}) W_{mn}^{(i)} (y_{im} - y_{in})^T}$$

(1)

where $T$ represents the transpose of matrix or vector, $\bar{y}_i$ is the projected mean vector of person $i$, i.e. $\bar{y}_i = (1/q_i) \sum_{j=1}^{q_i} y_{ij}$, $W_{mn}^{(i)}$ and $B_{mn}$ separately represent the elements of within-class and between-class weight matrices.

$$W_{mn}^{(i)} = \begin{cases} \exp(-\frac{||x_{im} - x_{in}||^2}{2\sigma^2}), & \text{if } x_{im} \text{ is among } k \text{ nearest neighbor of } x_{in} \text{ or } x_{in} \text{ is among } k \text{ nearest neighbor of } x_{im} \\ 0, & \text{otherwise} \end{cases}$$

(2)

$$B_{mn} = \begin{cases} \exp(-\frac{||x_m - x_n||^2}{2\sigma^2}), & \text{if } x_m \text{ is among } k \text{ nearest neighbor of } x_n \text{ or } x_n \text{ is among } k \text{ nearest neighbor of } x_m \\ 0, & \text{otherwise} \end{cases}$$

(3)

where $\sigma$ is an empirically determined parameter, $\bar{x}_i$ is the mean vector of person $i$, i.e. $\bar{x}_i = (1/q_i) \sum_{j=1}^{q_i} x_{ij}$. Thus, the between-class weight matrix is $B = [B_{mn}], (m,n = 1,2,\cdots,p)$, the within-class weight matrix is $W = \text{diag}\{W^{(1)}, \cdots, W^{(p)}\}$, where $W^{(i)} = [W_{mn}^{(i)}], (m,n = 1,2,\cdots,q_i)$. It is clear that both $B$ and $W$ are symmetric positive semi-definite matrices.

Suppose that the mapping from $x_{ij}$ to $y_{ij}$ is $\Phi$, i.e. $y_{ij} = \Phi^T x_{ij}$, then, the
objective function (1) can be rewritten as

\[ J(\Phi) = \frac{\Phi^T \bar{X} \bar{H} \bar{X} \Phi}{\Phi^T X L X^T \Phi} \]  \tag{4} \]

where \( L \) and \( H \) are Laplacian matrices. \( L = \mathbf{D} - \mathbf{W}, \mathbf{D} = \text{diag}\{\mathbf{D}_1, \ldots, \mathbf{D}_p\}, \)
\( \mathbf{D}_i \) is a diagonal matrix and its elements are column (or row) sum of \( \mathbf{W}^{(i)} \), i.e. \( D_{mn}^{(i)} = \sum_n W_{mn}^{(i)} \); \( \mathbf{H} = \mathbf{E} - \mathbf{B} \), \( \mathbf{E} \) is a diagonal matrix and its elements are column (or row) sum of \( \mathbf{B} \), i.e. \( E_{mn} = \sum_n B_{mn} \). \( \bar{X} = [\bar{x}_1, \ldots, \bar{x}_p] \).

Before continue our discussion, we would like to give following definitions:

- **locality preserving within-class scatter**: \( S_w = X L X^T \);
- **locality preserving between-class scatter**: \( S_b = \bar{X} \bar{H} \bar{X}^T \);
- **locality preserving total scatter**: \( S_t = S_b + S_w \).

It is easy to prove that \( S_w, S_b \) and \( S_t \) are symmetric semi-positive definite matrices.

Then, the optimal mapping \( \Phi_{opt} \) that maximizes the objective function (4) can be eigenvectors corresponding to the set of largest eigenvalues of \( S_w^{-1} S_b \). However, in face recognition, where the number of samples is smaller than the dimensionality of samples, the matrix \( S_w \) is always singular, and the computation of \( S_w^{-1} \) becomes complex and difficult. To overcome this drawback, Yu et al. [11] applied PCA on face data to reduce the dimensionality before DLPP. It is obvious that the null space of \( S_w \), which contains important discriminative information [3], was discarded.

### 3 Bagging null space locality preserving discriminant analysis

In [23], we have presented a null space discriminant locality preserving projections algorithm. It gets the null space of **locality preserving within-class** scatter matrix by performing singular value decomposition (SVD) on the \( n \times n \) matrix \( S_w \), which is time consuming or even infeasible due to that \( n \) is usually large for face recognition. To overcome this drawback, PCA approach is applied before N-DLPP. However, the PCA approach may discards some useful discriminant information. Here, we propose an new null space locality preserving discriminant analysis (NLPDA) method, which performs locality preserving discriminant analysis without PCA.

Another problem of null space discriminant analysis is that the null space becomes small when the training sample number is large. A lot of discriminative information outside the null space will be lost. To alleviate this problem,
bagging strategy is employed to construct multiple complementary NLPDA classifiers. The final decision is made by fusing results of these NLPDA classifiers using majority voting.

### 3.1 Null space locality preserving discriminant analysis

Null space locality preserving discriminant analysis method tries to maximize the same objective function as that of DLPP presented in Section 2. The Laplacian matrices $L$ and $H$ are always real symmetric positive semi-definite. According to linear algebra, $L$ and $H$ can be decomposed as follows:

$$L = P_L \Lambda_L P_L^T \quad H = P_H \Lambda_H P_H^T$$

where $\Lambda_L$ is the eigenvalue matrix of $L$, i.e. $\Lambda_L = \text{diag}\{\lambda_L^1, \lambda_L^2, \cdots, \lambda_L^l\}$, the columns of $P_L$ are the orthogonal eigenvectors of corresponding eigenvalues of $L$; $\Lambda_H$ is the eigenvalue matrix of $H$, i.e. $\Lambda_H = \text{diag}\{\lambda_H^1, \lambda_H^2, \cdots, \lambda_H^p\}$, the columns of $P_H$ are the orthogonal eigenvectors of corresponding eigenvalues of $H$.

Based on the properties of real symmetric semi-positive definite matrix, it is easy to know that all the eigenvalues of both $L$ and $H$ are non-negative, i.e. $\lambda_m^L \geq 0 (m = 1, 2, \cdots, l)$ and $\lambda_m^H \geq 0 (m = 1, 2, \cdots, p)$. Consequently, the three scatter matrices, called locality preserving within-class, locality preserving between-class and locality preserving total scatter matrices, can be rewritten as:

$$S_w = XLX^T = H_wH_w^T$$

$$S_b = XLX^T = H_bH_b^T$$

$$S_t = S_w + S_b = H_tH_t^T$$

where the precursors $H_w$, $H_b$ and $H_t$ in (6), (7) and (8) are

$$H_w = XP_L\Lambda_L^{1/2}$$

$$H_b = XP_H\Lambda_H^{1/2}$$

$$H_t = [H_w \quad H_b]$$

Since $S_w$, $S_b$ and $S_t$ are all semi-positive definite and $S_t = S_w + S_b$, it is easy to prove that the null space of $S_t$ does not include any discriminative information [24]. In such a manner, we can perform discriminant analysis only in the range subspace of $S_t$. Suppose that $\text{rank}(S_t) = r$, then the range subspace of $S_t$ is $\mathbb{R}^r = \text{span}\{q_1, q_2, \cdots, q_r\}$. The projected locality preserving within-class locality preserving between-class and locality preserving total scatter matrices are

$$\tilde{S}_w = Q^T S_w Q, \quad \tilde{S}_b = Q^T S_b Q, \quad \tilde{S}_t = Q^T S_t Q$$
where \( Q = \{q_1, q_2, \cdots, q_r\} \) can be obtained by performing singular value decomposition of \( H_t \).

In the range space of \( S_t, \tilde{S}_w \) and \( \tilde{S}_b \) are semi-positive definite and \( \tilde{S}_t \) is positive definite. Extensive researches indicated that the null space of \( \tilde{S}_w \) contains important discriminative information \([3,14,23]\). Due to that it is symmetric, \( \tilde{S}_w \) can be decomposed as
\[
\tilde{S}_w = U \Sigma U^T
\]
where the \( r \times r \) orthogonal matrix \( \Sigma = \text{diag}\{\lambda_1^w, \lambda_2^w, \cdots, \lambda_q^w, 0, \cdots, 0\} \) is the eigenvalue matrix of \( \tilde{S}_w \) (\( q \) is the rank of \( \tilde{S}_w \)). \( U = [u_1, u_2, \cdots, u_q, u_{q+1}, \cdots, u_r] \), \( u_m \) is the corresponding eigenvector of \( \lambda_m^w \) (\( m = 1, 2, \cdots, r \)). Consequently, the null space of \( \tilde{S}_w \) is the span of \( Q_w = [u_{q+1}, \cdots, u_r] \). The projection of \( \tilde{S}_w \) and \( \tilde{S}_b \) on the null space of \( \tilde{S}_w \) are
\[
\hat{S}_w = Q_w^T \tilde{S}_w Q_w = 0, \quad \hat{S}_b = Q_w^T \tilde{S}_b Q_w
\]

In the null space of \( \tilde{S}_w \), \( \hat{S}_b \) is positive definite, but \( \hat{S}_w \) is always zero, which leads the denominator of Eq.(4) to zero. According to [3], we modify the objective function to
\[
J(\Psi) = \Psi^T \hat{S}_b \Psi
\]
The optimal mapping \( \Psi_{opt} \) that maximizes Eq. (15) is the eigenvectors corresponding to the set of largest eigenvalues of \( \hat{S}_b \). Thus, the transformation matrix of NLPDA is
\[
G = QQ_w \Psi_{opt}
\]

### 3.2 Bagging null space locality preserving discriminant classifiers

Bagging is a method for generating multiple versions of a classifier and using these to get an aggregated classifier \([25,26]\). The aggregation does a majority voting when predicting a class. The multiple versions are formed by making bootstrap replicates of the training samples and using these as new training sets.

In NLPDA, the null space of \( \tilde{S}_w \) becomes small when the training sample number is large, which leads to the over-fitting problem that the null space can not contain enough discriminative information. To alleviate this problem, bagging strategy, which generates random bootstrap replicates by sampling the training set, is employed. Each generated replicate contains partial samples of the training set. Null space discriminant analysis is then performed on each replicate. The results of all NLPDA classifiers are finally aggregated using majority voting. Due to that the null space of \( S_t \) does not include any discriminative information and the Laplacian matrices \( L \) and \( H \) need to be constructed using

\[

"
all the samples in the training set, we first project all the training samples on the range space of $S_t$ before bagging. The detailed procedures of bagNLPDA is as follows.

At the training stage,

**Step 1.** Construct $H_t$ based on Eq. (11), (10) and (9). Perform singular value decomposition on $H_t$ to get the mapping $Q$ from the image space to the range space of $S_t$. Project all the samples of training set from the image space to the range space of $S_t$, i.e. $Z = Q^T X$, where $Z = \{z_{ij}\}$, $z_{ij} \in \mathbb{R}^r$ is the projection of $x_{ij}$.

**Step 2.** Generate $K$ bootstrap replicates $\{Z_k\}_{k=1}^K$. Each replicate contains the training samples of $p'$ individuals randomly selected from the $p$ classes.

**Step 3.** Perform null space locality preserving discriminant analysis on each replicate based on Eq. (12), (13), (14) and (15). Construct $K$ transformation matrices $\{G_k\}_{k=1}^K$ based on Eq. (16).

At the recognition stage,

**Step 1.** The input face data $x$ is projected to $K$ subspaces using the $K$ transformation matrices $\{G_k\}_{k=1}^K$.

**Step 2.** Perform nearest neighbor classification (NNC) on each subspace to generate $K$ classification results $\{C_k(x)\}_{k=1}^K$.

**Step 3.** Fuse the $K$ classification results using majority voting, the final class of $x$ is chosen as

$$C(x) = \arg \max_i \sum_{k=1}^K T_k(x \in i).$$

where

$$T_k(x \in i) = \begin{cases} 1, & C_k(x) = i \\ 0, & \text{otherwise} \end{cases}.$$

This algorithm has several novelty features. First, we construct Laplacian matrices $L$ and $H$ directly in the image space, where the local structure of face data points is exactly described. The Laplacian matrices $L$ and $H$ are constructed using all the samples in training set, bagging strategy has no impact on the geometrical structure shape of all the training samples. Second, bagging on training samples enlarges the null space of locality preserving within-class scatter matrix, therefore, the over-fitting problem of NLPDA is alleviated. Third, the dimension of feature space is first greatly reduced without loss on discriminative information by removing the null space of locality preserving total scatter matrix.
4 Experiments

We conduct experiments on subsets of FERET [27] and PIE [28] face databases. The FERET face database, which is a standard database for testing and evaluating state-of-art face recognition algorithms, contains 14 126 images from 1199 individuals. In our experiments, we select a subset which contains 1131 frontal images from 229 individuals. The PIE face database contains 41 368 images with different poses, different illuminations and different expressions from 68 individuals. We select 1425 frontal images with different illuminations from the 68 individuals to form the subset.

All the images in the two subsets are rotated and scaled so that the centers of the eyes are placed on specific pixels and then they are cropped to $32 \times 32$ pixels. Thus, each image can be represented by a 1024-dimensional vector in image space. Figure 1 and 2 show the sample images from the two subsets. All the experiments in this section are conducted 20 times and averaged. The simple nearest neighbor classifier is employed for classification.

Fig. 1. Sample images for one individual of the FERET subset.

Fig. 2. Sample images for one individual of the PIE subset.

4.1 Performance analysis on NLPDA, DLPP and Fisherface

We compare the recognition performances of NLPDA, DLPP and Fisherface on both FERET and PIE subsets.

In the FERET subset, two images of each individual are randomly selected for training, while the remaining images are used for testing. Thus, we have 468 training samples and 731 testing samples. In the PIE subset, for each individual four images are randomly selected for training and the remaining images are used for testing. So the training sample set size is 272 and the testing sample set size is 1153.
Figure 3 and 4 report the accuracies of different methods with the variance of the dimensionality of subspaces. As expected, both on FERET and PIE subsets, the locality preserving discriminant methods, i.e. NLPDA and DLPP, outperform the Fisherface method, which indicates that combining both the local structure characteristic and the discriminative information of data points is efficient for face recognition.

![Graph showing recognition accuracies (%) of NLPDA, DLPP and Fisherface on FERET subset (2 training samples).](image)

Fig. 3. Recognition accuracies (%) of NLPDA, DLPP and Fisherface on FERET subset (2 training samples).

In Fig. 3, the performance of NLPDA increases with the increasing of dimensionality. However, DLPP and Fisherface methods attain the highest accuracies when the dimensionality is between 80 and 120. When the dimensionality keeps on increasing, the performances of these two methods are gradually decreasing, which illustrates that the eigenvectors corresponding small eigenvalues are sensitive to noise and make the classifiers unstable.

#### 4.2 Bagging NLPDA

In this section, we discuss the experiments on bagging NLPDA. FERET subset is employed in the experiments. Two images of each individual are randomly selected for training and the remaining images are used for testing.
Figure 5 reports the performance of bagNLPDA. We generate 20 replicates and each replicate contains 120 individuals for training. It is obvious that the single NLPDA classifier constructed from each replicate is less effective than the original NLPDA classifier constructed on the full training set. However, the bagNLPDA method, which combines the multiple NLPDA classifiers in each replicate, significantly outperforms the original NLPDA classifier.

In bagNLPDA, two parameters $p'$ and $K$, say the number of individuals each replicate containing and the number of replicates, should be chosen. Table 1 reports performances of bagNLPDA, where the replicates contain different number of individuals for training. The number of replicates is fixed as 20. Bagging NLPDA attains the highest accuracy when the number of individuals is between 120 and 140. When the number of individuals keeps on increasing, the recognition accuracy of bagNLPDA is gradually decreasing. This result is consistent with the analysis in [18] that when bootstrapping the training set in bagging, in average only $1 - 1/e \approx 63.2\%$ of the training samples is used in each bootstrap replicate. Table 2 reports recognition accuracies of bagNLPDA with different number of bagging replicates. It is obvious that the recognition accuracies of bagNLPDA are more stable when using a relatively large number of replicates.
Fig. 5. Recognition accuracies (%) of combining 20 NLPDA classifiers constructed from bagging replicates. Each replicate contains 120 training individuals

Table 1
Recognition accuracies (%) of bagNLPDA containing different number ($p^\prime$) of individuals for training. The number of replicates is 20.

<table>
<thead>
<tr>
<th>$p^\prime$</th>
<th>100</th>
<th>120</th>
<th>140</th>
<th>160</th>
<th>180</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>72.8</td>
<td>73.2</td>
<td>73.0</td>
<td>72.8</td>
<td>72.2</td>
<td>71.3</td>
</tr>
<tr>
<td>Variance</td>
<td>2.25</td>
<td>2.40</td>
<td>2.48</td>
<td>2.30</td>
<td>2.39</td>
<td>2.23</td>
</tr>
</tbody>
</table>

Table 2
Recognition accuracies (%) of bagNLPDA with different number ($K$) of bagging replicates. Each replicate contains 120 training individuals.

<table>
<thead>
<tr>
<th>$K$</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>71.4</td>
<td>72.3</td>
<td>72.9</td>
<td>73.2</td>
<td>73.2</td>
<td>73.3</td>
</tr>
<tr>
<td>Variance</td>
<td>1.35</td>
<td>1.41</td>
<td>1.59</td>
<td>1.31</td>
<td>1.27</td>
<td>1.32</td>
</tr>
</tbody>
</table>

4.3 Comparison bagNLPDA with NLPDA, DLPP and Fisherface

Experiments in this section are performed to evaluate the performances of bagNLPDA, NLPDA, DLPP and Fisherface methods with varying number of
training samples of each individual. Both the FERET and PIE subsets are employed.

We randomly select $i (i = 2, 3, 4$ for the FERET subset and $i = 4, 8, 12$ for the PIE subset) images of each individual for training and the remaining images for testing. The two parameters of bagNLPDA are set as follows: $p' = 120$ and $K = 20$ for the FERET subset; $p' = 43$ and $K = 20$ for the PIE subset. Table 3 and 4 report the recognition accuracies of the four methods. The proposed bagNLPDA method significantly outperforms the other three subspace-based methods in terms of the recognition accuracies at different training sample size. As illustrated in Table 3, the recognition accuracies of locality preserving-based discriminant analysis methods, i.e. NLPDA and DLPP, are higher than those of Fisherface, which again demonstrates that combining both the local structure characteristic and the discriminative information of data points is efficient for face recognition. As illustrated in Table 3 and 4, the performances of NLPDA, DLPP and Fisherface methods are not always increase with the increasing of training sample size, which indicates that single classifiers of these three methods are unstable for face recognition. The proposed bagNLPDA method can efficiently alleviate this problem and its performance gradually increases with the increasing of training sample size.

Table 3
Recognition accuracies (%) on FERET subset.

<table>
<thead>
<tr>
<th>Size</th>
<th>bagNLPDA</th>
<th>NLPDA</th>
<th>DLPP</th>
<th>Fisherface</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>73.6±1.34</td>
<td>70.3±1.53</td>
<td>68.7±1.96</td>
<td>61.6±1.54</td>
</tr>
<tr>
<td>3</td>
<td>83.6±1.56</td>
<td>78.2±1.62</td>
<td>78.6±1.92</td>
<td>69.1±1.72</td>
</tr>
<tr>
<td>4</td>
<td>85.6±2.29</td>
<td>75.7±2.81</td>
<td>75.5±3.25</td>
<td>68.9±3.73</td>
</tr>
</tbody>
</table>

Table 4
Recognition accuracies (%) on PIE subset.

<table>
<thead>
<tr>
<th>Size</th>
<th>bagNLPDA</th>
<th>NLPDA</th>
<th>DLPP</th>
<th>Fisherface</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>93.8±1.58</td>
<td>91.2±2.01</td>
<td>91.9±1.79</td>
<td>91.1±2.11</td>
</tr>
<tr>
<td>8</td>
<td>99.1±0.47</td>
<td>98.5±0.65</td>
<td>98.5±0.72</td>
<td>98.6±0.54</td>
</tr>
<tr>
<td>12</td>
<td>99.9±0.08</td>
<td>98.5±0.63</td>
<td>99.4±0.32</td>
<td>98.9±0.62</td>
</tr>
</tbody>
</table>

5 Conclusions

In this paper, we present a novel bagging null space locality preserving discriminant analysis method for feature extraction and recognition. All training samples are first projected into the range space of locality preserving total scatter to reduce dimensionality without loss on discriminative information
and then randomly sampled to a set of bootstrap replicates. Null space discriminant analysis is performed in each replicate and the results of them are combined using majority voting. The proposed method can efficiently alleviate the over-fitting problem of single NDLPP classifier and improve the performance of face recognition. Experimental results on FERET and PIE subsets indicate that bagNLPDA performs significantly better than NDLPP, DLPP and Fisherface methods in terms of recognition accuracy.

In this study, we only fuse the discriminative information in the null space of locality preserving within-class scatter of each replicate. The range space of locality preserving within-class scatter also contains much discriminative information. It may further improve the performance. Investigating a method of combining discriminative information in these two subspaces is a direction for our further study.

Acknowledgments

This work is supported by the National High-Tech Research and Development Plan of China under grant No. 2007AA01Z423, the Defense Basic Research Project of the 'Eleventh Five-Year-Plan' of China under grant No. C10020060355, and the Key Research Project of the Natural Science Foundation of Chongqing Municipality of China under grant No. CSTC2007AC2018. The authors would like to thank the anonymous reviewers for their constructive advice.

References


LI-PING YANG received his B.S. degree from Chongqing University, China, in 2003. From October 2003 to September 2004, he studied in the University of Electro-Communication of Tokyo, Japan, as an exchange student. From 2005, he is a Ph.D. candidate in Chongqing University. His research interests include pattern recognition and image processing.

WEIGUO GONG received his Ph.D. degree in computer science from the Tokyo Institute of Technology of Japan in March 1996 as a scholarship gainer of Japan Government. From April 1996 to March 2002, he served as a researcher or senior researcher in NEC Labs of Japan. Now he is a professor of Chongqing University, China. He has published over 60 research papers in international journals and conferences and two books as an author or co-author. His current interests are in the areas of pattern recognition and intelligent information technology.

XIAOHUA GU received her B.S. degree in information engineering from Chongqing University, China, in 2005. Currently, she is a Ph.D. candidate in the College of Opto-Electronic Engineering, Chongqing University. Her research interests include pattern recognition, face recognition and computer vision.

WEIHONG LI received her Ph.D. degree from Chongqing University in 2006. Now she is an associate professor in Chongqing University. Her current research interests are in the areas of pattern analysis, kernel methods and machine learning.

YANFEI LIU received her B. S. degree in information engineering from Chongqing University, China, in 2007. Currently, she is a Master student in the College of Opto-Electronic Engineering, Chongqing University. Her research interests are in pattern recognition and information acquiring and processing.