Bilateral Two-Dimensional Neighborhood Preserving Discriminant Embedding for Face Recognition

Jiuzhen Liang*, Member, IEEE, Chen Chen*, Member, IEEE, Yunfei Yi, Xiuxiu Xu, and Meng Ding Member, IEEE

Abstract—In this paper, we propose a novel bilateral two-dimensional neighborhood preserving discriminant embedding for supervised linear dimensionality reduction for face recognition. It directly extracts discriminative face features from images based on Graph Embedding and Fisher’s criterion. The proposed method is a manifold learning algorithm based on Graph Embedding criterion, which can effectively discover the underlying nonlinear face data structure. Both within-neighboring and between-neighboring information are taken into account to seek an optimal projection matrix by minimizing the intra-class scatter and maximizing the inter-class scatter based on Fisher’s criterion. The performance of the proposed method is evaluated and compared with other face recognition schemes on the Yale, PICS, AR and LFW databases. The experiment results demonstrate the effectiveness and superiority of the proposed method as compared to the state-of-the-art dimensionality reduction algorithms.

Index Terms—Face recognition, bilateral two-dimensional neighborhood preserving embedding, supervised learning, graph embedding, Fisher’s criterion

I. INTRODUCTION

Face recognition has drawn a great deal of attention from both biological and engineering fields in recent years. Compared with other biometric techniques such as fingerprint and iris recognition, face recognition has clear advantages of being natural and passive [1]. Therefore, many research works have been carried out in the field of face recognition [2], [3], [4], [5], [6]. As one of the most important biometric techniques, it has been applied to a wide range of applications, such as mug-shot database matching, credit card verification, security system, and scene surveillance [7].

Due to the high dimension of face data, dimensionality reduction is a key problem in face recognition [8]. Principal Component Analysis (PCA) [9] and Linear Discriminant Analysis (LDA) [10] are two classical linear techniques for dimensionality reduction. However, many research efforts have shown that the face images possibly lie on a nonlinear sub-manifold. As a result, linear approaches may fail to discover the nonlinear data structures.

Compared with linear techniques, nonlinear manifold-based approaches, such as Isomap [11], locally linear embedding (LLE) [12], local tangent space alignment (LTSA) [13], and Laplacian Eigenmaps (LE) [14], can effectively reveal the geometrical structure of the underlying manifold. These manifold learning models, however, do not provide explicit mapping functions and thus cannot project a new high-dimensional data into a low-dimensional feature space. Therefore, Ding et al. proposed a novel structure-guided manifold learning method [15] and its variants multi-layer Joint Gait-Pose Manifolds (JGP) [16], [17], [18] explicitly model the nonlinear mapping from the latent space to high dimensional data. Moreover, He et al. proposed locality preserving projections (LPP) [19] and neighborhood preserving embedding (NPE) [20], which can make the projection of new data straightforward [21]. However, there are still several disadvantages that should be noted. On one hand, images need to be transformed into vectors. The resulting image vectors often lead to a high-dimensional vector space, where it is difficult to calculate the bases to represent the original images due to the large size. On the other hand, such a matrix-to-vector transformation may lose some structural information residing in the original two-dimensional images.

Based on this observation, some two-dimensional (2D) subspace learning methods have been proposed recently, such as 2DPCA [22], 2DLDA [23], 2DLPP [24], and 2DNPE [25]. Although these 2D methods are able to solve the problem of singularity that is common in 1D methods, they perform compression only in one direction. In order to overcome this limitation, some bilateral 2D methods like B2DLPP [26], [27] and B2DNPE [28] have been presented. Different from 2DLPP and 2DNPE, B2DLPP and B2DNPE preserve the internal relationship between image rows and columns. As a result, the dimensionality reduction is performed not only in column direction but also in row direction.

Recently, a method to discover representativeness and discriminativeness by semisupervised active learning is proposed [29]. It takes advantages of both active learning and semisupervised learning. In [30], Du et al. derived a robust multi-label active learning algorithm based on an maximum correntropy criterion (MCC) by merging uncertainty and representativeness, and propose an efficient alternating optimization method.
to solve it.

However, most of these 1D and 2D methods are unsupervised methods and the unsupervised nature of these algorithms restricts their discriminating capability. In order to enhance the discriminating power of the face recognition approaches, many supervised algorithms have been developed. Ridder et al. [31] presented Supervised LLE (SLLE) by using the label information of training samples. In neighborhood selection, SLLE adopts a new Euclidean distance matrix by adding distances between samples in different classes. Besides, Han et al. [32] proposed an effective technique, called neighborhood preserving discriminant embedding (NPDE), which takes the class membership information into account based on NPE. Recently, some supervised 2D algorithms have been introduced. Jun et al. [33] proposed 2DLPPCI method that employs class information for structure of similarity matrix to obtain a subspace which preserves local neighbor structure and centralizes same class samples of training images.

In this paper, a supervised extension of B2DNPE, named as bilateral two-dimensional neighborhood preserving discriminant embedding (B2DNPE), is proposed based on discriminative feature extraction for face recognition. Unlike B2DNPE, the proposed B2DNPE takes the within-neighboring information as well as between-neighboring information into account when the high-dimensional face data is projected into the embedded low-dimensional feature space. The essence of this approach is to model two scatters, called within-class scatter and between-class scatter. In neighborhood selection, the within-class scatter is formed by the reconstruction weights constructed by the samples from the same class, while the between-class scatter is formed by the reconstruction weights constructed by the samples from different classes. The proposed algorithm is an effective manifold learning algorithm based on Graph Embedding, making B2DNPE is capable of discovering the intrinsic structure hidden in the nonlinear face data. In addition, the discriminant capability of B2DNPE is further enhanced by taking advantage of Fisher’s criterion, which minimizes the intra-class scatter and maximizes the inter-class scatter. A flowchart of the proposed method is presented in Fig.1.

The rest of the paper is organized as follows. In Section 2, we provide a brief review of the related work, including 2DNPE and B2DNPE. The proposed B2DNPE method is introduced in Section 3. The experimental results on Yale, PICS and AR face databases are presented in Section 4. Finally, Section 5 concludes this study.

II. RELATED WORK

A. Two-dimensional neighborhood preserving embedding (2DNPE)

The main idea behind 2DNPE is to find the optimal projection in the column direction of images without transforming image matrices to vectors. Given a set of \(N\) training images \(X_i \in \mathbb{R}^{m \times n}(i = 1, 2, \ldots, N)\), 2DNPE attempts to seek an optimal \(n\)-dimensional transformation vector \(v\) to map the original data onto the \(m\)-dimensional space as following:\n\[Y_i = X_i v (i = 1, 2, \cdots, N).\]

The cost function of 2DNPE is defined as:
\[
\min \sum_{ij} \| Y_i - \sum_{ij} W_{ij} Y_j \|^2,
\]
where \(W_{ij}\) is the similarity between \(X_i\) and \(Y_j\).

In order to eliminate the influence of the scaling factor in the projection direction, a constraint is proposed as:
\[
Y^TY = 1 \iff v^TX^TXv = 1,
\]
where \(X = (X_1, X_2, \cdots, X_N)\) and \(Y = (Y_1, Y_2, \cdots, Y_N)\). Then the above minimization problem becomes
\[
\arg \min_{v^TX^TXv = 1} \| v^TX^T(M \otimes I_m)Xv \|,
\]
where \(M = (I_N - W)^T(I_N - W)\), and \(I_m\) is an identity matrix with the size of \(m \times m\). Operator \(\otimes\) is the Kronecker product of matrices. Here, the two matrices \(X^T(M \otimes I_m)X\) and \(X^TX\) are both symmetric and positive semi-definite. The projection matrix \(V = [v_1, v_2, \ldots, v_d]\) consists of the eigenvectors corresponding to the \(d\) smallest generalized eigenvalues of the following equation:
\[
X^T (M \otimes I_m) v = \lambda X^TX v.
\]

B. Bilateral two-dimensional neighborhood preserving embedding (B2DNPE)

The bilateral two-dimensional neighborhood preserving embedding (B2DNPE) performs dimensionality reduction in both row and column directions of image matrices. Let there be \(N\) training images \(X_i (i = 1, \ldots, N)\) all sharing the size of \((m \times n)\). The main goal of B2DNPE is to find an optimal linear transformation
\[
Y_i = U^TX_iV, i = 1, \ldots, N
\]
where \(U \in \mathbb{R}^{m \times l}(l < m)\) is the left-projection matrix and \(V \in \mathbb{R}^{n \times r}(r < n)\) is the right-projection matrix. Then there are \((l \times r)\) feature matrices \(Y_i (i = 1, \ldots, N)\). Let \(W_{ij}\) be the linear combination coefficients which are computed in the original image space. We want that \(Y_i\) can be represented using the
linear combination of its neighbors with the same coefficients for the original image $X_i$. The optimal projection matrix can be obtained by minimizing the following objective function:

$$
\min_{U,V} F(U,V) = \min_{U,V} \sum_i \|Y_i - \sum_j W_{ij} Y_j\|^2_F,
$$

(6)

where $\| \cdot \|$ is the Frobenius norm and $Y_i$ is the low dimensional representation of $X_i$ after dimensionality reduction. By simple algebra operation, we have

$$
F(U,V) = \text{tr}[U^T Q (M \otimes VV^T) Q^T U] = \text{tr}[V^T P^T (M \otimes UU^T) PV],
$$

(7)

where $Q = [X_1, X_2, ..., X_n]$ is a $(m \times nN)$ matrix generated by concatenating all the image matrices in row direction, and $P = [X_1^T, X_2^T, ..., X_n^T]^T$ is a $(mN \times n)$ matrix generated by concatenating all the image matrices in column direction. $M = (IN - W)(IN - W)^T$ is defined as a square $N \times N$ matrix, and $\text{tr}[\cdot]$ is the trace of square matrix.

It is proper to impose a constraint on the optimization problem to eliminate the effect of arbitrary scalability and translation:

$$
G(U,V) = \sum_i \|Y_i\|^2_F = 1 \Rightarrow \text{tr}[U^T Q (I_N \otimes VV^T) Q^T U] = 1 \Rightarrow \text{tr}[V^T P^T (I_N \otimes UU^T) PV] = 1.
$$

(8)

Now the optimization problem can be converted to solve the following generalized eigenvalue problems:

$$
Q(M \otimes VV^T)Q^T u = \lambda Q(I_N \otimes VV^T) Q^T u \quad (9)
$$

$$
P^T(M \otimes UU^T)Pv = \gamma P^T(I_N \otimes UU^T)Pv \quad (10)
$$

III. BILATERAL TWO-DIMENSIONAL NEIGHBORHOOD PRESERVING DISCRIMINANT EMBEDDING (B2DNPE)

As discussed previously, 2DPCA, 2DLPP, 2DNPE and B2DNPE are unsupervised learning methods. The class membership relation in the original data space is not considered in these methods. 2DLDA is a supervised learning algorithm and it maximizes the difference of the between-class and within-class samples. Although 2DPCA and 2DLDA are global algorithms, 2DLDA utilizes the class information to enhance its discriminant ability. That is the main reason that 2DLDA outperforms 2DPCA in most cases.

This motivates us to generate two reconstruction weights using the class information to enhance the performance of B2DNPE for face recognition. We propose B2DNPE algorithm, which has two scatter matrices named within-class scatter and between-class scatter. These two scatter matrices are formulated based on two distinct sets of reconstruction weights, which are respectively computed from the training face data of same and different persons using B2DNPE. The main objective of B2DNPE is to find an optimal projection that minimizes the ratio of the within-class scatter and the between-class scatter.

A. Finding the projection matrices by B2DNPDE

From Eq.(9) and Eq.(10), one may find that the left-projection and right-projection matrices are determined by $M$, which is derived from the reconstruction weight matrix. However, in B2DNPE, $M$ is only related to the neighborhood or the nearest neighbors. Therefore, even if two samples belong to different classes in the original space, they may close to each other in the low dimensional subspace as long as they are in neighborhood. This will lead to unfavorable results in face recognition.

To make images in the same class close to each other and those in different classes far away from each other in the dimension reduced subspace, the class information of images should be used. Therefore, our main focus is how to use class information to formulate the within-class scatter and between-class scatter.

Given a set of $N$ sample images $X_i \in \mathbb{R}^{m \times n}$, which are divided into $r$ classes. Denote $C_i$ as the set of images in the $i$-th class and $|C_i|$ is the number of images in the $i$-th class. Arrange the images such that all images in the same class appear subsequently in order. Suppose that $X = \{X_1, X_2, ..., X_N\}$ are the representations of the ordered images, then $X_1, X_2, ..., X_{|C_1|}$ are in the 1st class, $X_{|C_1|+1}, X_{|C_1|+2}, ..., X_{|C_1|+|C_2|}$ are in the 2nd class, and so on.

1) Definition of within-class scatter: In order to reflect the contribution of the $j$-th local neighbors to the reconstruction of the $i$-th face image, a within-class reconstruction weight matrix $W^w_{ij}$ is defined as: $W^w_{ij} \neq 0$ if $X_i$ and $X_j$ are from the same class; and $W^w_{ij} = 0$, otherwise. The within-class reconstruction weight matrix $W^w$ of training face data can be calculated by solving a least squares problem with a constraint:

$$
\min \sum_i \|X_i - \sum_j W^w_{ij} X_j\|^2_F \quad (11)
$$

where $X_i \in C_i$, and the constraint of Eq. (11) is defined as:

$$
\sum_j W^w_{ij} = 1. \quad \text{The resulting } W^w_{ij} \text{ of Eq. (11) will be stored in the weight matrix } W^w \text{ with the size of } N \times N.
$$

In order to keep the images from the same class close to each other in the dimension reduced subspace, the appropriate projection matrix can be obtained by minimizing the following cost function:

$$
\min_{U,V} F^w_{U,V} = \min_{U,V} \sum_i \|Y_i - \sum_j W^w_{ij} Y_j\|^2_F \quad (12)
$$

where $Y_i = U^T X_i, i = 1, ..., N$. Here $Y_i$ is the low dimensional representation of $X_i$, $U$ is the left projection matrix and $V$ is the right projection matrix when we want to project the high-dimensional data $X_i$ into a low-dimensional subspace.

Suppose $Z_i = Y_i - \sum_j W^w_{ij} Y_j, Z = [Z_1, Z_2, ..., Z_N]$ and $Y = [Y_1, Y_2, ..., Y_N]$, then the cost function $F^w_{U,V}$ can be rewritten as:

$$
F^w_{U,V} = \sum_i \|Y_i - \sum_j W^w_{ij} Y_j\|^2_F = \sum_i \|Z_i\|^2_F = \text{tr}(ZZ^T) = \text{tr}\{Y[(I_N - W^w)^T (I_N - W^w)] \otimes I_Y^T\}. \quad (13)
$$

Since $Y = U^T X_V$, the cost function in Eq.(13) can be reformulated as:

$$
F^w_{U,V} = \text{tr}[U^T Q(M^w \otimes VV^T) Q^T U]. \quad (14)
$$
Similarly, we can get another form of the cost function
\[ F(w, V) = tr[V^T P T (M^w \otimes U U^T) PV], \]
where \( Q = [X_1, X_2, \ldots, X_n] \) and \( P = [X_1^T, X_2^T, \ldots, X_n^T]^T. \)
\( M^w = (I_N - W^w)^T (I_N - W^w) \) is an \((N \times N)\) square matrix,
\( I \) is an identity matrix, and \( tr[\cdot] \) is the trace of square matrix.

2) Definition of between-class scatter. We define a between-class reconstruction weight matrix by using the information between classes. The between-class reconstruction weight matrix \( W^b \) is constructed as: suppose \( X_i \in C_l \), if \( X_j \notin C_l \) and \( X_j \) is one of the \( k \) nearest neighbors of \( X_i \), \( W^b_{ij} \neq 0 \); otherwise, \( W^b_{ij} = 0 \). The between-class reconstruction weight matrix \( W^b \) of training face data can be obtained by minimizing the following objective function:
\[ \min \sum_i \|X_i - \sum_j W^b_{ij} X_j\|^2. \]

Without loss of generality, an additional constraint needs to be added to Eq.(16): \( \sum_j W^b_{ij} = 1 \). Then \( W^b_{ij} \) will be stored in the weight matrix \( W^b \), which is an \( N \times N \) matrix.

In order to make the face images from different classes far away from each other after projection, we maximize the following cost function:
\[ F^b(U, V) = \max_{U, V} \sum_i \|Y_i - \sum_j W^b_{ij} Y_j\|^2, \]
where \( M^b = (I_N - W^b)^T (I_N - W^b) \).

The objective function of B2DNPDE is to minimize \( F^w(U, V) \) and maximize \( F^b(U, V) \) at the same time by
\[ \arg \min_{U, V} \frac{F^w(U, V)}{F^b(U, V)} = \arg \min_{U, V} \frac{tr[U^T Q(M^w \otimes V V^T) Q^T U]}{tr[U^T Q(M^b \otimes V V^T) Q^T U]}, \]

or
\[ \arg \min_{U, V} \frac{F^w(U, V)}{F^b(U, V)} = \arg \min_{U, V} \frac{tr[V^T P T (M^w \otimes U U^T) PV]}{tr[V^T P T (M^b \otimes U U^T) PV]} . \]

From the objective functions in Eq.(19) and Eq.(20), we can see that the proposed B2DNPDE is a supervised learning method which models the within-class scatter and between-class scatter by graph embedding in a manner of Fisher’s criterion [34], [35]. By using Fisher’s criterion in B2DNPDE, the within locality is preserved while the separability of the between locality is increased. Then we can choose an iterative procedure to get the left-projection and the right-projection matrices by solving the following generalized eigenvalue problems:
\[ Q(M^w \otimes V V^T) Q^T u = \lambda Q(M^b \otimes V V^T) Q^T u, \]
\[ P^T (M^w \otimes U U^T) P v = \gamma P^T (M^b \otimes U U^T) P v. \]

Note that the four matrices \( Q(M^w \otimes V V^T) Q, Q(M^b \otimes V V^T) Q^T, P^T (M^w \otimes U U^T) P \) and \( P^T (M^b \otimes U U^T) P \) are all symmetric and positive semi-definite. The column vectors \( u_i (i = 1, \cdots, l) \) and \( v_i (i = 1, \cdots, r) \) are the eigenvectors corresponding to the \( l \) and \( r \) smallest eigenvalues of Eq.(21) and Eq.(22), respectively. Therefore, the projection matrices as follows: \( U = [u_1, u_2, \ldots, u_l] \) and \( V = [v_1, v_2, \cdots, v_r] \).

Algorithm 1: The algorithm pseudo-code of B2DNPDE

**Input:** the original training samples are \( X_1, X_1, \ldots, X_N \) and their labels, a testing image \( T \), the dimension of reduced subspace \( d \).

**Output:** the label of the testing image \( T \).

**begin**

Recorder \( X_1, X_1, \ldots, X_N \) according to their labels

Construct within-class scatter \( W^w \)

\( Q = [X_1, X_2, \ldots, X_n] \)

\( P = [X_1^T, X_2^T, \ldots, X_n^T]^T \)

\( M^w = (I_N - W^w)^T (I_N - W^w) \)

\( M^b = (I_N - W^b)^T (I_N - W^b) \)

Initialization: \( U = I_m \)

**while** \( U, V \) do not converge **do**

find eigenvectors of \( P^T (M^w \otimes U U^T) PV = \gamma P^T (M^b \otimes U U^T) PV \)

find eigenvectors of \( Q(M^w \otimes V V^T) Q^T u = \lambda Q(M^b \otimes V V^T) Q^T u \)

for \( i = 1 : N \) do

\( Y_i = U^T X_i \)

\( T_{proj} = U^T TV \)

neighborindex = 0

\( \min_{dist} = INFINITY \)

for \( i = 1 : N \) do

\( d(Y_i, T_{proj}) = \sum_{x=1}^{m} \| Y_i(x,y) - T_{proj}(x,y) \|^2 \)

if \( d(Y_i, T_{proj}) < \min_{dist} \) then

\( \min_{dist} = d(Y_i, T_{proj}) \)

neighborindex = \( i \)

**return** the label of the image neighborindex

**end**

B. Testing stage of the B2DNPDE method

Once the left-projection and right-projection matrix \( U \) and \( V \) are obtained, any testing image can be mapped into the reduced subspace, then a nearest-neighbor classifier is used for face classification. Assume that \( \{Y_i|i=1,2,\ldots,N\} \) are the representations of the training samples \( \{X_i|i=1,2,\ldots,N\} \) after left and right projections using the linear transformation \( Y_i = U^T X_i V \). The distance between two face images in the dimension reduced subspace is defined as
\[ d(X_i, Y_j) = \sum_{x=1}^{l} \sum_{y=1}^{r} [Y_i(x,y) - Y_j(x,y)]^2 . \]

Given a testing sample \( T \), suppose its matrix representation in the dimension reduced subspace is \( T_{proj} \). If \( d(Y_i, T_{proj}) = \min_{j} d(Y_j, T_{proj}) \), then the testing face image \( T \) will be classified into the same class as the \( l \)-th image in the training data set.

C. The algorithm pseudo-code of B2DNPDE

The main steps of B2DNPDE method are summarized in Algorithm 1.
Table I: Comparisons between B2DNPDE and other methods in terms of top recognition accuracy(%) on Yale database
(the parameter \( p \) denotes the number of samples from each person in the training set)

<table>
<thead>
<tr>
<th>Methods</th>
<th>( p = 5 )</th>
<th>( p = 6 )</th>
<th>( p = 7 )</th>
<th>( p = 8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSLPP</td>
<td>85.61(5)</td>
<td>84.93(22)</td>
<td>82.67(24)</td>
<td>89.11(23)</td>
</tr>
<tr>
<td>GILPP</td>
<td>83.39(35)</td>
<td>81.87(38)</td>
<td>77.92(40)</td>
<td>85.11(43)</td>
</tr>
<tr>
<td>2DPCA</td>
<td>84.39(100 × 11)</td>
<td>85.87(100 × 11)</td>
<td>81.17(100 × 11)</td>
<td>84.89(100 × 10)</td>
</tr>
<tr>
<td>2DLPP</td>
<td>84.11(100 × 27)</td>
<td>84.73(100 × 32)</td>
<td>80.92(100 × 15)</td>
<td>85.89(100 × 17)</td>
</tr>
<tr>
<td>2DNPE</td>
<td>83.94(100 × 31)</td>
<td>84.60(100 × 30)</td>
<td>81.00(100 × 40)</td>
<td>83.22(100 × 51)</td>
</tr>
<tr>
<td>S2DNPE</td>
<td><strong>87.11</strong>(100 × 17)</td>
<td><strong>88.00</strong>(100 × 35)</td>
<td><strong>85.08</strong>(100 × 26)</td>
<td><strong>89.22</strong>(100 × 11)</td>
</tr>
<tr>
<td>B2DLPP</td>
<td>84.22(30 × 30)</td>
<td>85.53(25 × 25)</td>
<td>81.33(32 × 32)</td>
<td>84.89(28 × 28)</td>
</tr>
<tr>
<td>B2DNPE</td>
<td>84.11(31 × 31)</td>
<td>85.73(31 × 31)</td>
<td>81.83(33 × 33)</td>
<td>85.00(25 × 25)</td>
</tr>
<tr>
<td>B2DNPDE</td>
<td><strong>86.28</strong>(21 × 21)</td>
<td><strong>88.73</strong>(18 × 18)</td>
<td><strong>87.00</strong>(17 × 17)</td>
<td><strong>93.22</strong>(69 × 69)</td>
</tr>
</tbody>
</table>

IV. EXPERIMENTS AND DISCUSSION

To evaluate the performance of the proposed B2DNPDE method for face recognition, experiments are conducted on four well-known face image databases, Yale [36], AR [38] and LFW [39]. The performance of the proposed B2DNPDE method is compared with the state-of-the-art dimension reduction algorithms for face recognition, including 2DPCA, 2DLPP, 2DNPE, S2DNPE, B2DLPP, B2DNPE, GSLPP [40] and GILPP [41]. In the experiments, the data set is divided into two parts: one for training, the other for testing. We adopt a nearest neighbor classifier for recognition. Note that 2DLPP, 2DNPE, S2DNPE, GSLPP and GILPP on the Yale database have 7 images and these 7 images demonstrate variations in neutral, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink (see Fig. 2).

![Figure 2: Some sample images of the Yale face database.](image)

In order to compare the performance of the proposed method with other methods with varying number of training samples, \( p \) images are randomly selected from each person to construct the training set, while the remaining images are used as testing set. There is no overlap in the sample images between the training and testing sets. The value of \( p \) is set to 5, 6, 7 and 8 in the experiments on the Yale database. For each given \( p \), we vary the dimension of feature subspace and record the best recognition accuracy of each algorithm independently. And the recognition accuracies are averaged over 20 random tests. The size of feature matrices in the dimension reduced subspace is \( d \times d \) for the proposed method, B2DNPDE and B2DLPP, \( 100 \times d \) for 2DPCA, 2DLPP, 2DNPE and S2DNPE, \( d \) for GSLPP and GILPP. Table I lists the average top recognition accuracies and the corresponding dimensions of feature matrices for different methods with different training samples per person. The dimension of feature matrix corresponding to the top recognition accuracy is listed in the parenthesis next to the accuracy. It can be found that when \( p \) is equal to 6, 7 and 8, the proposed method outperforms the other methods in terms of recognition accuracy.

In general, the face recognition rates vary with the dimension of the subspace. In order to explore the relationship between the recognition accuracy and dimension of subspace, experiments are performed under a series of different reduced dimensions. In this experiment, all the algorithms are repeated 20 times in order to report the average performance. Fig. 3 shows the plot of recognition accuracy versus the reduced dimensionality \( d \) for B2DNPDE, B2DNPE, B2DLPP, 2DPCA, 2DLPP, 2DNPE, S2DNPE, GSLPP and GILPP on the Yale database (for \( p = 5, 6, 7, 8 \)). From Fig. 3, we can see that when the parameter \( p \) is 6, 7 and 8, respectively, the proposed method B2DNPDE yields the highest performance in terms of accuracy, while S2DNPE obtains the highest recognition accuracy when \( p \) equals to 5.

B. Experiments on PICS database

The PICS [37] database was constructed by University of Stirling for computational vision and control. This database consists of 84 grayscale images of 12 women. Each person has 7 images and these 7 images demonstrate variations in lighting condition and facial expression (e.g. surprise, pain, neutral, happy, fear, disgust and angry). The size of each image is \( 241 \times 181 \), with 256 grey levels per pixel. In our experiments, each image in the PICS database is resized to \( 60 \times 45 \). Some sample images are shown in Fig. 4.

Similar to the experiment on the Yale database, two experiments are presented on the PICS database. In the first
experiment, we compare the performance of different algorithms under conditions where the number of training samples varies. Here, we conduct four tests with a varying number of training samples on the PICS database. Since there are only 7 images for each individual, the number of training samples per person, $p$, is set to 3, 4, 5 and 6. For each given $p$, we run each algorithm 20 times to randomly choose the training set and compute the average face recognition accuracy for different dimensions of feature subspace. The best recognition accuracies obtained by different methods as well as the corresponding dimensions of feature subspace (the numbers in parentheses) on the PICS database are given in Table II.

In the second experiment, we investigate the performance of the proposed method over a range of reduced dimensions, and provide comparisons with other methods. For each subject in the PICS face database, we randomly select $p$ samples for training, and the remaining ones for testing. All the algorithms are repeated 20 times for each $p$. Fig. 5 demonstrates the average recognition accuracies of various methods (including B2DNPDE, B2DNPE, B2DLPP, 2DPCA, 2DLPP, 2DNPE, S2DNPE, GILPP, GSLPP and GILPP) under the
cases of different dimensions of feature subspace. It is evident from this figure that the proposed approach always has better recognition rate than the competing methods under the four groups of tests.

C. Experiments on AR database

The AR face database [38] contains over 4000 color face images of 126 people (70 men and 56 women). Images feature frontal view faces with different facial expressions, lighting conditions and occlusions (sun glasses and scarf). The pictures of most persons were taken in two sessions (separated by two weeks). Each section contains 13 color images and 120 individuals (65 men and 55 women) participated in both sessions. The details of all frontal views in the first session are: neutral expression, smile, anger, scream, left light on, right light on, all sides lights on, wearing sun glasses, wearing sun glasses and left light on, wearing sun glasses and right light on, wearing scarf, wearing scarf and left light on, wearing scarf and right light on. The second session is repeated under the same conditions. The images of 60 persons (30 men and 30 women) from these 120 individuals are selected and used in our experiment. Each individual has 26 images (14 full facial images and 12 facial images with occlusions) and we manually crop the face portion of image and then normalize it to $25 \times 20$. Some sample images of one subject are shown in Fig. 6.

We follow the same two experiments as conducted on the Yale and PICS databases. In the first experiment, we also conduct four groups of experiments by varying the size of the training set. The value of $p$, the number of training samples per person, is set to 5, 9, 13 and 17 on the AR face database. Table III lists the average best face recognition rate of each method as well as the corresponding dimensionality of reduced subspace (the numbers in parentheses) on this database.

The performance of the proposed method with varying reduced dimension is also evaluated on the AR database. Firstly, $p(p=5, 9, 13, 17)$ samples of each individual are randomly selected for training, and the rest for testing. Secondly, we take the average after repeating 20 trials for each given $p$. The relationship between the recognition accuracies of different algorithms and the dimensionality of feature subspace is depicted in Fig. 7, where we can conclude that the proposed approach achieves superior performance over other algorithms under the four groups of tests.
Table III: Comparisons between B2DNPDE and other methods in terms of top recognition accuracy(%) on the AR database (the parameter $p$ denotes the number of samples from each person in the training set.)

<table>
<thead>
<tr>
<th>Methods</th>
<th>$p = 5$</th>
<th>$p = 9$</th>
<th>$p = 13$</th>
<th>$p = 17$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSLPP</td>
<td>80.67(40)</td>
<td>88.35(43)</td>
<td>93.83(46)</td>
<td>95.47(43)</td>
</tr>
<tr>
<td>GILPP</td>
<td>80.55(58)</td>
<td>84.54(60)</td>
<td>93.89(59)</td>
<td>94.30(62)</td>
</tr>
<tr>
<td>2DPCA</td>
<td>80.48(25 × 20)</td>
<td>87.62(25 × 20)</td>
<td>90.22(25 × 20)</td>
<td>91.29(25 × 20)</td>
</tr>
<tr>
<td>2DLPP</td>
<td>81.77(25 × 20)</td>
<td>88.43(25 × 20)</td>
<td>92.03(25 × 19)</td>
<td>95.33(25 × 4)</td>
</tr>
<tr>
<td>2DNPE</td>
<td>82.78(25 × 19)</td>
<td>89.31(25 × 19)</td>
<td>92.06(25 × 20)</td>
<td>94.23(25 × 3)</td>
</tr>
<tr>
<td>S2DNPE</td>
<td>88.60(25 × 19)</td>
<td>92.20(25 × 19)</td>
<td>92.55(25 × 20)</td>
<td>93.61(25 × 20)</td>
</tr>
<tr>
<td>B2DLPP</td>
<td>82.25(20 × 20)</td>
<td>92.42(20 × 20)</td>
<td>95.45(20 × 20)</td>
<td>97.96(20 × 20)</td>
</tr>
<tr>
<td>B2DNPE</td>
<td>84.89(20 × 20)</td>
<td>93.07(20 × 20)</td>
<td>95.57(18 × 18)</td>
<td>97.67(20 × 20)</td>
</tr>
<tr>
<td>B2DNPDE</td>
<td>91.23(20 × 20)</td>
<td>96.92(18 × 18)</td>
<td>96.52(17 × 17)</td>
<td>98.03(20 × 20)</td>
</tr>
</tbody>
</table>

Figure 7: Recognition accuracy vs. dimension of feature subspace using $p$ images of individuals for training on the AR face database.

D. Experiments on LFW database

The LFW database [39] contains images of 5729 different individuals in unconstrained environment. LFW-a is a version of LFW after alignment using commercial face alignment software [42]. We gather the subjects including no less than twenty-five samples and then get a dataset with 42 subjects of LFW after alignment using commercial face alignment software [42]. We gather the subjects including no less than twenty-five samples and then get a dataset with 42 subjects in LFW-a. We directly crop the face images and then resize them to $(32 \times 32)$. A few sample images of one subject are shown in Fig. 8.

For each subject, 6–9 samples are randomly chosen for training and the remaining samples for test. In other words, the value of $p$ is set to 6, 7, 8 and 9 in the first experiment. The average top face recognition accuracy of each algorithm and the corresponding dimensionality of reduced subspace (the numbers in parentheses) on the LFW dataset are presented in Table IV based on 20 random realizations of the training set.

Similarly, on the LFW database, we show the relationship between the recognition accuracies of different algorithms and the dimensionality of feature subspace in Fig. 9. The experimental results in Fig. 9 show that B2DNPDE outperforms the state-of-the-art approaches by considerable margins when $p$ equals to 6,7,8,9. This demonstrates the superiority of our method for face recognition in unconstrained environment.
Table IV: Comparisons between B2DNPDE and other methods in terms of top recognition accuracy(%) on the LFW database (the parameter $p$ denotes the number of samples from each person in the training set)

<table>
<thead>
<tr>
<th>Methods</th>
<th>$p = 6$</th>
<th>$p = 7$</th>
<th>$p = 8$</th>
<th>$p = 9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSLPP</td>
<td>47.70(49)</td>
<td>49.57(43)</td>
<td>51.28(49)</td>
<td>48.65(35)</td>
</tr>
<tr>
<td>GILPP</td>
<td>48.45(76)</td>
<td>46.33(79)</td>
<td>46.64(80)</td>
<td>47.46(80)</td>
</tr>
<tr>
<td>2DLPP</td>
<td>30.16(32 × 11)</td>
<td>32.12(32 × 5)</td>
<td>33.81(32 × 1)</td>
<td>35.44(32 × 7)</td>
</tr>
<tr>
<td>2DNPPE</td>
<td>32.44(32 × 11)</td>
<td>33.69(32 × 8)</td>
<td>35.54(32 × 5)</td>
<td>37.86(32 × 5)</td>
</tr>
<tr>
<td>S2DNPE</td>
<td>40.87(32 × 3)</td>
<td>43.60(32 × 3)</td>
<td>45.33(32 × 3)</td>
<td>46.39(32 × 3)</td>
</tr>
<tr>
<td>B2DLPP</td>
<td>33.87(18 × 18)</td>
<td>38.02(21 × 21)</td>
<td>38.04(18 × 18)</td>
<td>40.87(16 × 16)</td>
</tr>
<tr>
<td>B2DNPE</td>
<td>34.05(21 × 21)</td>
<td>36.26(21 × 21)</td>
<td>39.58(20 × 20)</td>
<td>39.68(21 × 21)</td>
</tr>
<tr>
<td>B2DNPDE</td>
<td><strong>53.39</strong>(12 × 12)</td>
<td><strong>57.83</strong>(10 × 10)</td>
<td><strong>60.15</strong>(10 × 10)</td>
<td><strong>62.42</strong>(9 × 9)</td>
</tr>
</tbody>
</table>

Therefore, our method is more practical for real world applications.

E. Analysis of the neighborhood size $k$

In the proposed B2DNPDE method, an important parameter is $k$, i.e., the $k$-nearest neighbors of a sample in between-class scatter. In order to analyze the effect of the neighborhood size $k$ on classification result, we design four tests to evaluate the average error rate under varying size of $k$. The parameter $k$ is chosen from 2, 3, ..., 10 on the PICS and LFW databases. For the AR database, it is selected from 2, 3, ..., 20. Fig. 10 plots the relationship between the performance of B2DNPDE and the neighborhood size $k$ of one individual in between-class scatter. The experiment is carried out by randomly splitting the image samples so that $p$ images for each person are used for training and the remaining for testing. For each fixed combination of $p$ and $k$, we run the classification test 20 times.

From Fig. 10(a)-(c), one may observe that the performance of B2DNPDE is a little sensitive to the parameter $k$. When the value of $p$ is fixed in each test, the error rate of B2DNPDE varies with the neighborhood size $k$. Based on the parameter tuning results, we can find that the value of $k$ should not be too large, otherwise the linear structure of the local neighborhood will be affected. In most cases, when the value of $k$ is a little smaller than the number of samples from each person in the training set, the error rate will be minimal.

F. Discussions

From the experiments on the Yale, PICS, AR and LFW databases, we summarize the following conclusions:

1) The experimental results show that the proposed method B2DNPDE outperforms the other methods in terms of recognition performance except for the first test on Yale when $p = 5$. In B2DNPDE, the within-class scatter, which is constructed by the samples from the same class, is used to reveal the underlying data structure. If we select a small number of samples from one class for training, there is not adequate information for B2DNPDE to find the optimal projection matrix. Moreover, increasing the number of training samples enhances the recognition rate of B2DNPDE on all databases.

2) The proposed B2DNPDE is consistently better than B2DNE in all experiments on the Yale, PICS, AR and LFW face databases. B2DNE is an unsupervised method and it builds reconstruction weight matrix by selecting the neighboring points with $k$ nearest distances. Therefore, even if the two points come from different classes, they might be in neighborhood. To some extent, the real underlying data structure may be lost and this will affect the discriminative ability of B2DNPDE. However, the proposed B2DNPDE is able to find the optimal projection that preserves the within-neighborhood scatter, while the samples from different classes are kept away from each other after projection via between-class scatter maximization. Consequently, the discriminative capability of B2DNPDE is better than B2DNE, which has been verified in the experiments.

3) 2DNE is an unsupervised algorithm, while S2DNE is the supervised extension of 2DNE. The neighborhood selection of S2DNE is restricted in the same class by using class membership information. Although the classification ability of S2DNE is obviously improved as compared with 2DNE, its performance is not comparable to the proposed B2DNPDE. This is because that B2DNPDE uses both within-class and between-class information to construct the projection matrix, while S2DNE utilizes the within-class information only.

4) GSLPP and GILPP use Gabor wavelet [43] to extract the Gabor feature representation of face images, which is more robust to varying external factors (such as illumination, face expression and pose). However, the recognition performance of these two methods are not better than the other methods in most cases. This is due to the fact that they use vector-based approaches to extract the low dimensional face feature. Vector-based approaches need to transform image matrices into vectors which may lose some useful structure information. In addition, GSLPP, a supervised method, only considers the within-class information like S2DNE. Therefore, the performance of our approach (B2DNPDE) is superior to GSLPP.

V. Conclusion

In this paper, we propose a new supervised manifold learning method for face recognition, dubbed bilateral two-dimensional neighborhood preserving discriminant embedding (B2DNPDE). In order to reduce the feature space and find an optimal projection subspace, B2DNPDE utilizes the class information of training images for constructing the projection matrix. The key of B2DNPDE algorithm is to seek a projection matrix that preserves the intra-class neighborhood geometry and at the same time keeps the samples from different classes.
Figure 9: Recognition accuracy vs. dimension of feature subspace using $p$ images of individuals for training on the LFW face database.

Figure 10: The performance of B2DNPDE vs. the neighborhood size $k$ on the different databases. (the parameter $p$ denotes the number of samples from each person in the training set)

far away from each other after dimensionality reduction. Therefore, the discriminant capability is much improved as compared to considering the intra-class information only. Extensive experimental evaluations on the Yale, PICS, AR and LFW face databases show that the proposed B2DNPDE method consistently outperforms the state-of-the-art dimensionality reduction approaches for face recognition. In future research, we will focus on finding an appropriate dimension of feature subspace to balance the computation cost and recognition accuracy.

ACKNOWLEDGMENTS

This work is supported by Blue Project of Universities in Jiangsu Province Training Young Academic Leaders Object, and National Natural Science Foundation of China(No.61170121). The authors would like to thank the online face database providers including Yale, PICS, AR and LFW.

REFERENCES


