Research Article

Block-Wise Two-Dimensional Maximum Margin Criterion for Face Recognition

Xiao-Zhang Liu¹ and Guan Yang²

¹ School of Computer Science, Dongguan University of Technology, Dongguan 523808, China

² School of Computer Science, Zhongyuan University of Technology, Zhengzhou 450007, China

Correspondence should be addressed to Xiao-Zhang Liu; liuxiaozhang@gmail.com

Received 13 August 2013; Accepted 7 November 2013; Published 22 January 2014

Academic Editors: J. H. Sossa and C. Zhang

Copyright © 2014 X.-Z. Liu and G. Yang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Maximum margin criterion (MMC) is a well-known method for feature extraction and dimensionality reduction. However, MMC is based on vector data and fails to exploit local characteristics of image data. In this paper, we propose a two-dimensional generalized framework based on a block-wise approach for MMC, to deal with matrix representation data, that is, images. The proposed method, namely, block-wise two-dimensional maximum margin criterion (B2D-MMC), aims to find local subspace projections using unilateral matrix multiplication in each block set, such that in the subspace a block is close to those belonging to the same class but far from those belonging to different classes. B2D-MMC avoids iterations and alternations as in current bilateral projection based two-dimensional feature extraction techniques by seeking a closed form solution of one-side projection matrix for each block set. Theoretical analysis and experiments on benchmark face databases illustrate that the proposed method is effective and efficient.

1. Introduction

Most well-known appearance-based face recognition methods are based on subspace techniques for feature extraction, such as principal component analysis (PCA) [1], linear discriminant analysis (LDA) [2], and maximum margin criterion (MMC) [3]. These conventional appearance-based techniques are based on the so-called vector-space model. Under this model, the original two-dimensional (2D in short) image data are reshaped into a one-dimensional (1D in short) long vector by stacking either rows or columns of the image. This vector-space model makes pattern recognition and analysis techniques be conveniently applied to image domain, and numerous successes have been achieved. However, it also introduces the following problems in practical applications. First, the intrinsic 2D structure of image matrix is removed. As a result, the spatial information stored in the 2D image is discarded and not effectively utilized for representation and recognition. Second, each image sample is modeled as a point in a high-dimensional space; for example, for an image of size

112×92 , the commonly used image size in face recognition, the dimension of the vector space is 10304, and the size of the scatter matrices is 10304×10304 . Obviously, a large number of training samples are needed to get a reliable and robust estimation of data statistics. This problem, known as curse of dimensionality, is often confronted in real applications. Third, a very limited number of data are usually available in real applications such that the small sample size (SSS) problem [4] comes forth frequently in practice.

To overcome the above drawbacks, efforts have been made to seek to extract the features directly without vectorization of image samples; that is, the representation of an image sample is retained in matrix form [5]. With this consideration, some bilateral projection based 2D feature extraction techniques have been proposed for seeking transforms on both sides of the image matrix, such as GLRAM (generalized low-rank approximation of matrices) [6], which can be seen as a kind of two-dimensional PCA, and 2DLDA (two dimensional LDA) [7], which can implicitly resolve the SSS problem suffered by LDA. These 2D methods are more

computationally efficient than their 1D counterparts, respectively. And, GLRAM and 2DLDA are evaluated empirically to be more effective than PCA and LDA, respectively [6, 7], due to preserving the intrinsic spatial information of data matrix.

Furthermore, two dimensional MMC (2DMMC) has been proposed [8], which aims to find two orthogonal projection matrices to project the original image matrices to a low-dimensional matrix subspace. In the projected subspace, a sample is close to those in the same class but far from those in different classes. Both theoretical analysis and experiments on benchmark face recognition datasets illustrate that 2DMMC is more effective and more efficient than GLRAM and 2DLDA. However, like GLRAM and 2DLDA, the algorithm of 2DMMC involves iterations and alternations of computing two-side projection matrices, which are time-consuming, and an arbitrary initial value before iterations cannot guarantee the global optimum.

In this paper, we propose a novel framework for 2D generalization of conventional MMC to extract discriminating features directly from 2D face images. The proposed algorithm, namely block-wise two-dimensional maximum margin criterion (B2D-MMC), aims to find local subspace projections by obtaining one-side projection matrix in each block set, such that in the subspace a block is close to those belonging to the same class but far from those belonging to different classes. B2D-MMC introduces a block-wise dividing method for face images as in [9], and the dividing method has been proven to be reliable. Based on one-side projection and block-wise learning, B2D-MMC eludes seeking iterative and alternating two projection matrices, as in GLRAM, 2DLDA and 2DMMC, and has more power of learning local characteristics of images.

The rest of this paper is organized as follows. Section 2 provides background information on 2DMMC. In Section 3, our Block-wise Two Dimensional Maximum Margin Criterion is proposed. The experiments on standard face recognition datasets are demonstrated in Section 4. Finally, we draw our conclusions in Section 5.

2. Review on MMC and 2DMMC

2.1. LDA and MMC. The most popular unsupervised feature extraction method is principal component analysis (PCA). It aims to find a subspace in which the variance of the projected data is a maximum. But PCA does not take into account the class information, so the features extracted are not very suitable for classification [2]. Linear discriminant analysis (LDA) is a well-known supervised method which has been shown to be more effective than PCA in face recognition tasks [2].

As supervised feature extraction methods, MMC and LDA share the notations of between-class scatter matrix and within-class scatter matrix as follows.

Given a set of N sample images $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ taking values in the d -dimensional vector form, each belonging to one of C classes. Assume the i th class contains N_i sample vectors $\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_{N_i}^{(i)}$, $i = 1, 2, \dots, C$, so $N = \sum_{i=1}^C N_i$.

The mean vector of the i th class and that of the sample set are, respectively, given by

$$\mathbf{m}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{x}_j^{(i)}, \quad \mathbf{m} = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^{N_i} \mathbf{x}_j^{(i)}. \quad (1)$$

The between-class scatter matrix $\mathbf{S}_b$ and within-class scatter matrix $\mathbf{S}_w$ are, respectively, defined as

$$\begin{aligned} \mathbf{S}_b &= \frac{1}{N} \sum_{i=1}^C N_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T, \\ \mathbf{S}_w &= \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^{N_i} (\mathbf{x}_j^{(i)} - \mathbf{m}_i)(\mathbf{x}_j^{(i)} - \mathbf{m}_i)^T. \end{aligned} \quad (2)$$

LDA is based on Fisher criterion, which aims to maximize the between-class distance and minimize the within-class distance as follows:

$$\begin{aligned} \mathbf{W}_{\text{opt}} &= \arg \max_{\mathbf{W}} \frac{|\mathbf{W}^T \mathbf{S}_b \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_w \mathbf{W}|} \\ &= [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p], \end{aligned} \quad (3)$$

where $|\cdot|$ denotes the determinant of matrix and $\mathbf{w}_i$ is the generalized eigenvector of $\mathbf{S}_b$ and $\mathbf{S}_w$ corresponding to the i th largest generalized eigenvalue λ_i , that is,

$$\mathbf{S}_b \mathbf{w}_i = \lambda_i \mathbf{S}_w \mathbf{w}_i, \quad i = 1, 2, \dots, p. \quad (4)$$

If $\mathbf{S}_w$ is nonsingular, the solution can be obtained by applying an eigendecomposition to matrix $\mathbf{S}_w^{-1} \mathbf{S}_b$. However, in face recognition applications, where generally the number of training images N is much smaller than that of pixels in each image d , one is confronted with the difficulty that the within-class scatter matrix $\mathbf{S}_w$ is always singular [2], since the rank of $\mathbf{S}_w$ is at most $N - C$. This is so-called the Small Sample Size (SSS) problem which the LDA method suffers from.

As an efficient and robust alternative to LDA, Maximum Margin Criterion (MMC) [3] is defined as

$$\begin{aligned} \max_{\mathbf{W}} \quad & \text{tr}(\mathbf{W}^T (\mathbf{S}_b - \mu \mathbf{S}_w) \mathbf{W}), \\ \text{s.t.} \quad & \mathbf{W}^T \mathbf{W} = \mathbf{I}, \end{aligned} \quad (5)$$

where $\text{tr}(\cdot)$ denotes the matrix trace and μ is a weighted parameter which is set to 1 in [3]. MMC is to find the optimal projection matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_q]$, which is composed of the q eigenvectors corresponding to the largest q eigenvalues of $\mathbf{S}_b - \lambda \mathbf{S}_w$.

The constraint $\mathbf{W}^T \mathbf{W} = \mathbf{I}$ allows MMC to avoid calculating the inverse of $\mathbf{S}_w$ and thus to elude the potential SSS problem.

2.2. 2DLDA and 2DMMC. 2DLDA [7] and 2DMMC [8] consider data with matrix representation and share the notations of between-class scatter and within-class scatter as follows.

Let $\mathbf{X}_j^{(i)} \in \mathbb{R}^{n \times m}$, $j = 1, 2, \dots, N_i$, be the images in the sample set belonging to the i th class, $i = 1, 2, \dots, C$ ($N = \sum_{i=1}^C N_i$). Both 2DLDA and 2DMMC aim to find two orthogonal projection matrices, $\mathbf{U} \in \mathbb{R}^{n \times l_1}$ and $\mathbf{V} \in \mathbb{R}^{m \times l_2}$, that map each image matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ to $\mathbf{Y} \in \mathbb{R}^{l_1 \times l_2}$, such that $\mathbf{Y} = \mathbf{U}^T \mathbf{X} \mathbf{V}$. The mean matrix of the i th class and that of the sample set are respectively given by

$$\mathbf{M}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{X}_j^{(i)}, \quad \mathbf{M} = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^{N_i} \mathbf{X}_j^{(i)}. \quad (6)$$

In the low dimensional matrix space resulting from the linear transformation $\mathbf{U}$ and $\mathbf{V}$, the between-class scatter $\tilde{\mathbf{S}}_b$ and within-class scatter $\tilde{\mathbf{S}}_w$ are, respectively, defined as

$$\begin{aligned} \tilde{\mathbf{S}}_b &= \text{tr} \left(\sum_{i=1}^C N_i \mathbf{U}^T (\mathbf{M}_i - \mathbf{M}) \mathbf{V} \mathbf{V}^T (\mathbf{M}_i - \mathbf{M})^T \mathbf{U} \right), \\ \tilde{\mathbf{S}}_w &= \text{tr} \left(\sum_{i=1}^C \sum_{j=1}^{N_i} \mathbf{U}^T (\mathbf{X}_j^{(i)} - \mathbf{M}_i) \mathbf{V} \mathbf{V}^T (\mathbf{X}_j^{(i)} - \mathbf{M}_i)^T \mathbf{U} \right). \end{aligned} \quad (7)$$

For both 2DLDA and 2DMMC, the optimal transformations $\mathbf{U}$ and $\mathbf{V}$ would maximize $\tilde{\mathbf{S}}_b$ and minimize $\tilde{\mathbf{S}}_w$.

2DLDA proposed in [7] can be formulated as

$$(\mathbf{U}, \mathbf{V}) = \arg \max_{\mathbf{U}, \mathbf{V}} \frac{\tilde{\mathbf{S}}_b}{\tilde{\mathbf{S}}_w}. \quad (8)$$

The optimization (8) is with respect to $\mathbf{U}$ and $\mathbf{V}$, and a closed form solution cannot be obtained.

2DMMC is defined in [8] as

$$\begin{aligned} \max_{\mathbf{U}, \mathbf{V}} \quad & \tilde{\mathbf{S}}_b - \mu \tilde{\mathbf{S}}_w, \\ \text{s.t.} \quad & \mathbf{U}^T \mathbf{U} = \mathbf{I}, \quad \mathbf{V}^T \mathbf{V} = \mathbf{I}, \end{aligned} \quad (9)$$

where μ is a weighted parameter. Also, a closed form solution can not be obtained due to bilateral unknown projections.

Due to the difficulty of computing the optimal $\mathbf{U}$ and $\mathbf{V}$ simultaneously, 2DLDA and 2DMMC both utilize iterative alternating schemes; in each iteration, first they optimize the objective with respect to $\mathbf{U}$ when fixing $\mathbf{V}$ ($\mathbf{V}$ is initialized as any orthogonal matrix before iterations) and then optimize the objective with respect to $\mathbf{V}$ when fixing $\mathbf{U}$. The alternating computation framework in each iteration is reviewed below.

Computation of $\mathbf{U}$. For a fixed $\mathbf{V}$, $\tilde{\mathbf{S}}_b$ and $\tilde{\mathbf{S}}_w$ can be rewritten as

$$\tilde{\mathbf{S}}_b = \text{tr} (\mathbf{U}^T \mathbf{S}_b^V \mathbf{U}), \quad \tilde{\mathbf{S}}_w = \text{tr} (\mathbf{U}^T \mathbf{S}_w^V \mathbf{U}), \quad (10)$$

where

$$\begin{aligned} \mathbf{S}_b^V &= \sum_{i=1}^C N_i (\mathbf{M}_i - \mathbf{M}) \mathbf{V} \mathbf{V}^T (\mathbf{M}_i - \mathbf{M})^T, \\ \mathbf{S}_w^V &= \sum_{i=1}^C \sum_{j=1}^{N_i} (\mathbf{X}_j^{(i)} - \mathbf{M}_i) \mathbf{V} \mathbf{V}^T (\mathbf{X}_j^{(i)} - \mathbf{M}_i)^T. \end{aligned} \quad (11)$$

For 2DLDA, similar to the optimization problem in (3), the optimal $\mathbf{U}$ can be obtained by computing an eigendecomposition on $(\mathbf{S}_w^V)^{-1} \mathbf{S}_b^V$ that is composed of the l_1 eigenvectors corresponding to the largest l_1 eigenvalues of $(\mathbf{S}_w^V)^{-1} \mathbf{S}_b^V$.

For 2DMMC, similar to the optimization problem in (5), the optimal $\mathbf{U}$ can be obtained by computing an eigendecomposition on $\mathbf{S}_b^V - \mu \mathbf{S}_w^V$, that is composed of the l_1 eigenvectors corresponding to the largest l_1 eigenvalues of $\mathbf{S}_b^V - \mu \mathbf{S}_w^V$.

Computation of $\mathbf{V}$. From the property $\text{tr}(\mathbf{A}\mathbf{A}^T) = \text{tr}(\mathbf{A}^T\mathbf{A})$ for any matrix $\mathbf{A}$, when $\mathbf{U}$ is fixed, a key observation is that $\tilde{\mathbf{S}}_b$ and $\tilde{\mathbf{S}}_w$ can be rewritten as

$$\tilde{\mathbf{S}}_b = \text{tr} (\mathbf{V}^T \mathbf{S}_b^U \mathbf{V}), \quad \tilde{\mathbf{S}}_w = \text{tr} (\mathbf{V}^T \mathbf{S}_w^U \mathbf{V}), \quad (12)$$

where

$$\begin{aligned} \mathbf{S}_b^U &= \sum_{i=1}^C N_i (\mathbf{M}_i - \mathbf{M})^T \mathbf{U} \mathbf{U}^T (\mathbf{M}_i - \mathbf{M}), \\ \mathbf{S}_w^U &= \sum_{i=1}^C \sum_{j=1}^{N_i} (\mathbf{X}_j^{(i)} - \mathbf{M}_i)^T \mathbf{U} \mathbf{U}^T (\mathbf{X}_j^{(i)} - \mathbf{M}_i). \end{aligned} \quad (13)$$

For 2DLDA, similar to the optimization problem in (3), the optimal $\mathbf{V}$ can be obtained by computing an eigendecomposition on $(\mathbf{S}_w^U)^{-1} \mathbf{S}_b^U$ that is composed of the l_2 eigenvectors corresponding to the largest l_2 eigenvalues of $(\mathbf{S}_w^U)^{-1} \mathbf{S}_b^U$.

For 2DMMC, similar to the optimization problem in (5), the optimal $\mathbf{V}$ can be obtained by computing an eigendecomposition on $\mathbf{S}_b^U - \mu \mathbf{S}_w^U$, that is composed of the l_2 eigenvectors corresponding to the largest l_2 eigenvalues of $\mathbf{S}_b^U - \mu \mathbf{S}_w^U$.

In contrast to 2DLDA, 2DMMC has the following advantages which makes it stable and efficient: (1) the objective of (9) increases monotonically through iterations; hence the convergence of 2DMMC is rigorously guaranteed [8]; (2) 2DMMC avoids computing inverse matrices in each iteration.

However, as bilateral projection based 2D feature extraction techniques, 2DMMC, 2DLDA, and GLRAM share such shortcomings: The iterations and alternations are time-consuming, and an arbitrary initial value of $\mathbf{V}$ cannot guarantee the global optimum.

3. Proposed Framework

Bilateral projection based 2D feature extraction techniques, such as 2DMMC, 2DLDA, and GLRAM, consider seeking transforms on both sides of image matrices, that is, both left and right projections are taken, but the computation of two-side projection matrices involves time-consuming iterations and alternations, and the initialization before iterations may lead to local optimum. In our study, in order to overcome forgoing shortcomings, we propose a framework that only takes right multiplication of each block to extract the inter-row spatial information. Our block-wise approach to face recognition, namely, Block-wise Two Dimensional Maximum Margin Criterion (B2D-MMC), is described as follows.

FIGURE 1: Images and their blocks of the first subject from ORL database: (a) 10 images of size 112×92 ; (b)–(e) 4 image blocks of size 28×92 for each image in (a).

3.1. Block-Wise Model for Face Recognition. Since we deal with images cropped either manually or by a face detection procedure, our block-wise model divides the face image into nonoverlapping groups of rows, which are called image blocks. Let $\mathbf{X} \in \mathbb{R}^{n \times m}$ denote a face image, where n, m are the numbers of rows and columns of $\mathbf{X}$, respectively. $\mathbf{X}$ is divided into n_B nonoverlapping image blocks $\mathbf{X}(k) \in \mathbb{R}^{r_B \times m}$, $k = 1, 2, \dots, n_B$, each including r_B rows of image $\mathbf{X}$. Figure 1 shows an example of image blocks. In the example, images of the first subject from the ORL database, which have the size of 112×92 , are partitioned into four blocks of size 28×92 , that is, $n = 112$, $m = 92$, $n_B = 4$, and $r_B = 28$.

For all sample images, the set of k th image blocks is referred to as the k th block set $\mathbb{BS}_k$, which spans a subspace referred to as the k th block manifold, $k = 1, 2, \dots, n_B$. The advocated B2D-MMC algorithm attempts to find a local subspace projection, that is, unilateral projection matrix, in each block set.

3.2. B2D-MMC. Considering a C -class problem, the i th class contains N_i training image matrices $\mathbf{X}_j^{(i)} \in \mathbb{R}^{n \times m}$, $j = 1, 2, \dots, N_i$, where $\mathbf{X}_j^{(i)}$ is the j th training image in class i , $i = 1, 2, \dots, C$, and n, m are the numbers of rows and columns of face images, respectively. Let N be the total number of training images, that is, $N = \sum_{i=1}^C N_i$.

As determined in Section 3.1, each image $\mathbf{X}_j^{(i)}$ consists of n_B blocks, each block including r_B rows of the face image. Denoting the k th image block of $\mathbf{X}_j^{(i)}$ as $\mathbf{X}_j^{(i)}(k) \in \mathbb{R}^{r_B \times m}$, $k = 1, 2, \dots, n_B$, we have

$$\mathbf{X}_j^{(i)} = \begin{pmatrix} \mathbf{X}_j^{(i)}(1) \\ \mathbf{X}_j^{(i)}(2) \\ \vdots \\ \mathbf{X}_j^{(i)}(n_B) \end{pmatrix}. \quad (14)$$

Thus the k th block set $\mathbb{BS}_k$ can be formulated as

$$\mathbb{BS}_k = \{\mathbf{X}_1^{(1)}(k), \dots, \mathbf{X}_{N_1}^{(1)}(k), \mathbf{X}_1^{(2)}(k), \dots, \mathbf{X}_{N_2}^{(2)}(k), \dots, \mathbf{X}_1^{(C)}(k), \dots, \mathbf{X}_{N_C}^{(C)}(k)\}, \quad (15)$$

$$k = 1, 2, \dots, n_B.$$

Also let $\mathbf{X}_j^{(i)}(k, r) \in \mathbb{R}^{1 \times m}$ be the r th row of $\mathbf{X}_j^{(i)}(k)$, $r = 1, 2, \dots, r_B$. Then we can write

$$\mathbf{X}_j^{(i)}(k) = \begin{pmatrix} \mathbf{X}_j^{(i)}(k, 1) \\ \mathbf{X}_j^{(i)}(k, 2) \\ \vdots \\ \mathbf{X}_j^{(i)}(k, r_B) \end{pmatrix}. \quad (16)$$

For all training image matrices, the proposed B2D-MMC aims to find n_B orthogonal right-side projection matrices, one for each image block set; that is, given a desired dimensionality l , to find $\mathbf{V}(k) \in \mathbb{R}^{m \times l}$ for the k th block set $\mathbb{BS}_k$, mapping the k th image block $\mathbf{X}_j^{(i)}(k) \in \mathbb{R}^{r_B \times m}$ to $\mathbf{Y}_j^{(i)}(k) \in \mathbb{R}^{r_B \times l}$, such that

$$\mathbf{Y}_j^{(i)}(k) = \mathbf{X}_j^{(i)}(k) \mathbf{V}(k), \quad k = 1, 2, \dots, n_B. \quad (17)$$

And we use the following $\mathbf{Y}_j^{(i)} \in \mathbb{R}^{n \times l}$ as the feature of image $\mathbf{X}_j^{(i)}$ for training:

$$\mathbf{Y}_j^{(i)} = \begin{pmatrix} \mathbf{Y}_j^{(i)}(1) \\ \mathbf{Y}_j^{(i)}(2) \\ \vdots \\ \mathbf{Y}_j^{(i)}(n_B) \end{pmatrix} \quad i = 1, 2, \dots, C, \quad j = 1, 2, \dots, N_i. \quad (18)$$

For classification, features of testing images are stacked by subfeatures in the same form as above.

The following shows how to find the n_B projection matrices $\mathbf{V}(k)$, $k = 1, 2, \dots, n_B$. Let $\mathbf{M}_i(k) \in \mathbb{R}^{r_B \times m}$ and $\mathbf{M}(k) \in \mathbb{R}^{r_B \times m}$ denote the mean of the k th image blocks in the i th class and the mean of the k th block set $\mathbb{B}\mathbb{S}_k$, respectively, as follows

$$\begin{aligned}\mathbf{M}_i(k) &= \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{X}_j^{(i)}(k), \\ \mathbf{M}(k) &= \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^{N_i} \mathbf{X}_j^{(i)}(k).\end{aligned}\quad (19)$$

Also let $\mathbf{m}_i(k, r) \in \mathbb{R}^{1 \times m}$ and $\mathbf{m}(k, r) \in \mathbb{R}^{1 \times m}$ be the r th row of $\mathbf{M}_i(k)$ and $\mathbf{M}(k)$, respectively, $r = 1, 2, \dots, r_B$. Then we have

$$\begin{aligned}\mathbf{m}_i(k, r) &= \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{X}_j^{(i)}(k, r), \\ \mathbf{m}(k, r) &= \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^{N_i} \mathbf{X}_j^{(i)}(k, r).\end{aligned}\quad (20)$$

Let us define the between-class block scatter matrix $\mathbf{S}_b(k)$ and within-class block scatter matrix $\mathbf{S}_w(k)$ of the k th block set $\mathbb{B}\mathbb{S}_k$ respectively as follows

$$\begin{aligned}\mathbf{S}_b(k) &= \sum_{i=1}^C N_i (\mathbf{M}_i(k) - \mathbf{M}(k))^T (\mathbf{M}_i(k) - \mathbf{M}(k)), \\ \mathbf{S}_w(k) &= \sum_{i=1}^C \sum_{j=1}^{N_i} (\mathbf{X}_j^{(i)}(k) - \mathbf{M}_i(k))^T (\mathbf{X}_j^{(i)}(k) - \mathbf{M}_i(k)),\end{aligned}\quad (21)$$

$k = 1, 2, \dots, n_B$. It is easy to verify that $\mathbf{S}_b(k)$ and $\mathbf{S}_w(k)$ are two $m \times m$ nonnegative definite matrices from their definitions.

In the low dimensional space resulting from the k th linear transformation $\mathbf{V}(k)$, as in 2DMMC [8], we adopt the Frobenius norm $\|\cdot\|_F$ [10] as the metric of matrices, that is, $\|\mathbf{A}\|_F^2 = \text{tr}(\mathbf{A}\mathbf{A}^T) = \text{tr}(\mathbf{A}^T\mathbf{A})$ for any matrix $\mathbf{A}$. Under this metric, the projected between-class block scatter $\tilde{\mathbf{S}}_b(k)$ and projected within-class block scatter $\tilde{\mathbf{S}}_w(k)$ can be respectively defined as follows

$$\begin{aligned}\tilde{\mathbf{S}}_b(k) &= \sum_{i=1}^C N_i \|(\mathbf{M}_i(k) - \mathbf{M}(k)) \mathbf{V}(k)\|_F^2 \\ &= \text{tr} \left(\sum_{i=1}^C N_i \mathbf{V}(k)^T (\mathbf{M}_i(k) - \mathbf{M}(k))^T \right. \\ &\quad \left. \times (\mathbf{M}_i(k) - \mathbf{M}(k)) \mathbf{V}(k) \right),\end{aligned}$$

$$\begin{aligned}\tilde{\mathbf{S}}_w(k) &= \sum_{i=1}^C \sum_{j=1}^{N_i} \|(\mathbf{X}_j^{(i)}(k) - \mathbf{M}_i(k)) \mathbf{V}(k)\|_F^2 \\ &= \text{tr} \left(\sum_{i=1}^C \sum_{j=1}^{N_i} \mathbf{V}(k)^T (\mathbf{X}_j^{(i)}(k) - \mathbf{M}_i(k))^T \right. \\ &\quad \left. \times (\mathbf{X}_j^{(i)}(k) - \mathbf{M}_i(k)) \mathbf{V}(k) \right), \\ &\quad k = 1, 2, \dots, n_B.\end{aligned}\quad (22)$$

The proposed B2D-MMC finds the orthogonal projection matrix $\mathbf{V}(k)$ for the k th block set $\mathbb{B}\mathbb{S}_k$ by the following optimization:

$$\begin{aligned}\max_{\mathbf{V}(k)} \quad & \tilde{\mathbf{S}}_b(k) - \mu_k \tilde{\mathbf{S}}_w(k), \\ \text{s.t.} \quad & \mathbf{V}(k)^T \mathbf{V}(k) = \mathbf{I},\end{aligned}\quad (23)$$

where μ_k is a weighted parameter, $k = 1, 2, \dots, n_B$.

In order to compute $\mathbf{V}(k)$, $k = 1, 2, \dots, n_B$, comparing (21) with (22), the following relation is held:

$$\tilde{\mathbf{S}}_b(k) - \mu_k \tilde{\mathbf{S}}_w(k) = \text{tr} \left(\mathbf{V}(k)^T (\mathbf{S}_b(k) - \mu_k \mathbf{S}_w(k)) \mathbf{V}(k) \right). \quad (24)$$

Thus the optimal $\mathbf{V}(k)$ can be computed by solving a eigen-decomposition on $m \times m$ matrix $\mathbf{S}_b(k) - \mu_k \mathbf{S}_w(k)$; that is,

$$\mathbf{V}(k) = \{\mathbf{v}_1(k), \mathbf{v}_2(k), \dots, \mathbf{v}_l(k)\}, \quad (25)$$

where $m \times 1$ vector $\mathbf{v}_i(k)$ is the eigenvector corresponding to the i th largest eigenvalue of $\mathbf{S}_b(k) - \mu_k \mathbf{S}_w(k)$, $i = 1, 2, \dots, l$, $k = 1, 2, \dots, n_B$.

From the description above, it is easy to see that our B2D-MMC has the following two advantages compared with 2DMMC [8].

(a) *Computational Complexity.* B2D-MMC seeks a closed form solution of unilateral projection matrix for each block set instead of finding iterative solutions of two projection matrices for the entire image matrix, avoiding iterations and alternations as in 2DMMC, which saves the computational effort.

(b) *Locality.* Based on the block-wise model, B2D-MMC learns local characteristics of input image by dividing the face image into non-overlapping image blocks. Expectedly, distribution of data is much less complex inside these block manifolds.

3.3. Algorithm Design. Based on the analysis above, our B2D-MMC algorithm is designed as in Algorithm 1.

In our experiments reported in Section 4, the parameter μ_k is set as $\text{tr} \mathbf{S}_b(k) / \text{tr} \mathbf{S}_w(k)$ according to [11], $k = 1, 2, \dots, n_B$.

Input: Training set of N matrices $\{\mathbf{X}_j^{(i)} \in \mathbb{R}^{n \times m} \mid j = 1, 2, \dots, N, i = 1, 2, \dots, C\}$, desired dimensionality l , number of image blocks n_B , and number of rows in each image block $r_B = n/n_B$.

Output: Projection matrices $\mathbf{V}(k) \in \mathbb{R}^{m \times l}, k = 1, 2, \dots, n_B$.

Step 1. Initialize $k = 1$.

Step 2. Assign rows $(k-1)r_B + 1 \sim kr_B$ of $\mathbf{X}_j^{(i)}$ to $\mathbf{X}_j^{(i)}(k) \in \mathbb{R}^{r_B \times m}$; then compute $\mathbf{M}_i(k)$ and $\mathbf{M}(k)$ using (19), $j = 1, 2, \dots, N, i = 1, 2, \dots, C$.

Step 3. Compute $\mathbf{S}_b(k)$ and $\mathbf{S}_w(k)$ using (21).

Step 4. Compute the l eigenvectors corresponding to the largest l eigenvalues of $\mathbf{S}_b(k) - \mu_k \mathbf{S}_w(k)$ to form $\mathbf{V}(k)$ as in (25).

Step 5. If $k < n_B$, then $k = k + 1$, go to **Step 2**; else stop.

ALGORITHM 1

FIGURE 2: Images of one person from the ORL face database.

3.4. Computational Complexity Analysis. Most of the algorithms involve computations scale to $O(h^3)$ for eigen-decomposition of an $h \times h$ matrix [10]. The eigen-decomposition of the scatter matrices in B2D-MMC amounts to a complexity of $O(m^3)$. However, as reviewed in Section 2, in 2DLDA and 2DMMC, the scatter matrices in each iteration are of size $n \times n$, so the overall computation complexity of 2DMMC is $O(tn^3)$, where t is the number of iterations. Obviously we can expect that $O(m^3)$ is smaller than $O(tn^3)$ when t is considerable.

4. Experiments

In this section, to investigate the performance of the proposed B2D-MMC for face recognition, we compare our method with PCA [1], LDA [2], MMC [3], GLRAM [6], 2DLDA [7], and 2DMMC [8], in both accuracy and efficiency. Furthermore, the effect of image block size on recognition results is investigated.

4.1. Performance Comparison

4.1.1. Face Datasets. In our experiment, we use two standard face recognition databases which are widely used as benchmark datasets in feature extraction literature.

The ORL Face Database. There are ten images for each of the 40 human subjects, which were taken at different times, varying the lighting, facial expressions and facial details. Images from one subject are shown in Figure 2. The original images (with 256 gray levels) have size 92×112 , which are resized to 32×32 for efficiency.

The Yale Face Database. It contains 11 gray scale images for each of the 15 individuals. The images demonstrate variations in lighting condition, facial expression, and with/without glasses. Images from one subject are shown in Figure 3. In our experiment, the images were also resized to 32×32 .

4.1.2. Parameter Settings for B2D-MMC. For each individual, $TN = 2, 3, 4$ images were randomly selected as training

FIGURE 3: Images of one person from the Yale face database.

samples, and the rest were used for testing. The training set was used to learn $n_B = 4$ subspaces, each for one block set. Thus the size of the block set is 8×32 . Features of images for classification were stacked by sub-features in the form of (18), and the recognition was performed by Nearest Neighbor Classifier, with the Frobenius norm as the similarity metric. Since the training set was randomly chosen, we repeated each experiment 20 times and calculated the average recognition accuracy. In general, the recognition rate varies with l , that is, the number of columns of the feature (projected image). We set l to the corresponding dimensionality when the best performance was obtained by 2DMMC [8].

4.1.3. Comparison on Classification Accuracy. Tables 1 and 2 show the experimental results of the proposed B2D-MMC on the two databases, respectively, with the *best* results of PCA, LDA, MMC, GLRAM, 2DLDA, and 2DMMC referred from [8] for comparison. For all the methods, the value in each entry represents the average recognition accuracy of 20 independent trials, and the number in brackets is the corresponding projection dimensionality.

Since the value of dimensionality l , which corresponds to the best performance obtained by 2DMMC, is not necessarily the best choice for our B2D-MMC, it is clear that B2D-MMC outperforms 2DMMC and the other feature extraction methods on both of the two data sets.

4.1.4. Comparison on Efficiency. In this subsection, B2D-MMC is compared with 2DMMC in computational efficiency. We take the ORL and the Yale datasets where $TN = 2$ for example; that is, two training samples are randomly selected for each subject.

For 2DMMC, we record the training time in the following way: taking the entries in Tables 1 and 2 as the best classification accuracies, that is, 78.75% as the best on the ORL and 54.37% as the best on the Yale dataset, the iteration of training process stops if the difference between the obtained classification accuracy and the best classification accuracy is smaller than 0.1%. And the projection dimensionality of the training process is set to the corresponding value of the best classification, that is, 12×12 for the ORL and 6×6 for the Yale dataset.

FIGURE 4: Training time of two methods on ORL and Yale datasets with $TN = 2$.

The average training time of B2D-MMC and 2DMMC, over 20 independent runs on a typical laptop using MATLAB, is shown in Figure 4. It can be seen that B2D-MMC is more efficient than 2DMMC, both on the ORL and on the Yale dataset. This conforms to the complexity analysis in Section 3.4. This must be because that, unlike 2DMMC which is to find iterative solutions of two projection matrices for the entire image matrix, B2D-MMC seeks a closed form solution of one-side projection matrix for each block set, avoiding iterations and alternations as in 2DMMC, which decreases the computational load.

4.2. Effect of Number of Blocks on Recognition Results. The proposed B2D-MMC has been applied on the ORL and the Yale datasets with the same settings as in Section 4.1 but for three different values of n_B , namely 3, 4, and 5. Results, shown in Figures 5 and 6, reveal that the performance of B2D-MMC achieves optimum when n_B takes an appropriate value, for example, $n_B^* = 4$, and neither raising nor reducing the value of n_B^* degrades the performance of B2D-MMC. It can be interpreted as follows. An increase in the number of blocks per image helps to learn more local characteristics; however, a decrease in the number helps to utilize more global

TABLE 1: Face recognition accuracies of different methods on the ORL database. TN means number of training samples per subject, and the number in brackets is the corresponding projection dimensionality. The bold value means the highest accuracy among all the methods.

Method	TN = 2	TN = 3	TN = 4
PCA	70.67% (79)	78.88% (118)	84.21% (152)
LDA	72.80% (25)	83.79% (39)	90.13% (39)
MMC	77.97% (39)	86.32% (39)	91.63% (39)
GLRAM	71.30% (17×17)	79.84% (11×11)	84.73% (16×16)
2DLDA	78.13% (11×11)	86.79% (16×16)	92.08% (15×15)
2DMMC	78.75% (12×12)	87.50% (10×10)	92.92% (8×8)
B2D-MMC	79.14% ($l = 12$)	87.63% ($l = 10$)	92.88% ($l = 8$)

TABLE 2: Face recognition accuracies of different methods on the Yale database. TN means number of training samples per subject, and the number in brackets is the corresponding projection dimensionality. The bold value means the highest accuracy among all the methods.

Method	TN = 2	TN = 3	TN = 4
PCA	46.04% (29)	49.96% (44)	55.67% (58)
LDA	42.81% (11)	60.33% (14)	68.10% (13)
MMC	52.37% (14)	61.83% (14)	67.95% (15)
GLRAM	49.33% (6×6)	54.17% (6×6)	57.76% (5×5)
2DLDA	44.37% (7×7)	59.71% (5×5)	68.71% (5×5)
2DMMC	54.37% (6×6)	63.50% (9×9)	68.86% (15×15)
B2D-MMC	56.11% ($l = 6$)	64.35% ($l = 9$)	68.92% ($l = 15$)

FIGURE 5: Recognition performance on the ORL dataset for different number of blocks per image.

characteristics. The optimal recognition performance results from the tradeoff between local and global information.

5. Conclusions

This paper proposed a novel framework to extract discriminating features directly from 2D face images. The proposed B2D-MMC introduces a block-wise model for face

FIGURE 6: Recognition performance on the Yale dataset for different number of blocks per image.

recognition, performing one-side subspace projection inside each block manifold, in which a block is close to those belonging to the same class but far from those belonging to different classes. The unilateral projection and the block-wise learning avoid iterations and alternations as in current bilateral projection based two-dimensional feature extraction approaches, and have advantages in complexity and locality.

Computational complexity analysis shows that B2D-MMC consumes less time than 2DLDA and 2DMMC when

the number of iterations for the latter is considerable. Performance comparison experiments on the ORL and the Yale datasets illustrate that B2D-MMC is more effective and efficient than current bilateral projection based two-dimensional feature extraction techniques.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is partially supported by the Basic and Frontier Technology Research Project of Henan Province in China under Grant no. 122300410321 and Science and Technology Development Project of Henan Province under Grant no. 132102210186.

References

- [1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [3] H. Li, T. Jiang, and K. Zhang, "Efficient and robust feature extraction by maximum margin criterion," *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 157–165, 2006.
- [4] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Computer Science and Scientific Computing, Academic Press, Boston, Mass, USA, 2nd edition, 1990.
- [5] W.-S. Zheng, J. H. Lai, and S. Z. Li, "1D-LDA vs. 2D-LDA: when is vector-based linear discriminant analysis better than matrix-based?" *Pattern Recognition*, vol. 41, no. 7, pp. 2156–2172, 2008.
- [6] J. Ye, "Generalized low rank approximations of matrices," in *Proceedings of the 21th International Conference on Machine Learning*, Banff, Canada, July 2004.
- [7] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," in *Advances in Neural Information Processing Systems 17 (NIPS '04)*, Vancouver, Canada, December 2004.
- [8] Q. Gu and J. Zhou, "Two dimensional maximum margin criterion," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 1621–1624, Taipei, Taiwan, April 2009.
- [9] A. Eftekhari, M. Forouzanfar, H. Abrishami Moghaddam, and J. Alirezaie, "Block-wise 2D kernel PCA/LDA for face recognition," *Information Processing Letters*, vol. 110, no. 17, pp. 761–766, 2010.
- [10] J. W. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, Pa, USA, 1997.
- [11] F. Nie, S. Xiang, Y. Song, and C. Zhang, "Extracting the optimal dimensionality for discriminant analysis," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, vol. 2, pp. 617–620, Honolulu, Hawaii, USA, April 2007.