Bidirectional PCA With Assembled Matrix Distance Metric for Image Recognition

Wangmeng Zuo, David Zhang, *Senior Member, IEEE*, and Kuanquan Wang, *Member, IEEE*

***Abstract*—Principal component analysis (PCA) has been very successful in image recognition. Recent research on PCA-based methods has mainly concentrated on two issues, namely: 1) feature extraction and 2) classification. This paper proposes to deal with these two issues simultaneously by using bidirectional PCA (BD-PCA) supplemented with an assembled matrix distance (AMD) metric. For feature extraction, BD-PCA is proposed, which can be used for image feature extraction by reducing the dimensionality in both column and row directions. For classification, an AMD metric is presented to calculate the distance between two feature matrices and then the nearest neighbor and nearest feature line classifiers are used for image recognition. The results of the experiments show the efficiency of BD-PCA with AMD metric in image recognition.**

***Index Terms*—Face recognition, feature extraction, image recognition, nearest feature line, palm print recognition, principal component analysis (PCA).**

I. Introduction

PRINCIPAL component analysis (PCA) or PCA-based approaches have been very successful in image representation and recognition. In 1987, Sirovich and Kirby used PCA to represent human faces [1], [2]. Subsequently, Turk and Pentland proposed a PCA-based face recognition method, eigenfaces [3]. PCA has now been widely investigated and has been successfully applied to other image recognition tasks [4]–[6].

Despite the great success of PCA, some issues remain that deserve to be further investigated. First, because of the small sample size (SSS) problem, PCA is prone to be overfitted to the training set. Although no researchers have pointed out this problem directly, it can be addressed using some PCA-based approaches, such as $(\mathrm{PC})^2\mathrm{A}$ [7], [8], 2DPCA [9]–[11], and modular PCA [12]. The drawbacks of these approaches are that $(\mathrm{PC})^2\mathrm{A}$ alleviates the overfitting problem simply by blurring an image with an intrinsic low-dimensional image while 2DPCA and modular PCA introduce a much higher feature dimension than classical PCA [9], [10], [12]. Further work is thus required to solve the overfitting problem while avoiding the high feature dimension problem.

A second area requiring further investigation is the design of classifiers based on the PCA feature. One general classifier is the nearest neighbor (NN) classifier using the Euclidean distance measure. Other distance measures such as angle-based and Mahalanobis distance measures have been studied to further improve recognition performance [13]–[16]. Recently, the nearest feature line (NFL) classifier is introduced to eliminate the performance deterioration of NN caused by the reduction of prototypes [17]. Most recently, other variants or extensions of NFL have been investigated in [18]–[21]. Although previous studies of NN have shown that distance measures greatly affect recognition accuracy, with reference to NFL, distance measures have been little investigated.

This paper simultaneously investigates these two issues. First, to circumvent the overfitting problem, we propose a bidirectional PCA (BD-PCA) method that at the same time avoids the high feature dimension problem. Second, we present an assembled matrix distance (AMD) metric to calculate the distance between two feature matrices and then apply the AMD metric in the implementation of NN and NFL classifiers.

The organization of this paper is as follows. Section II investigates the classical PCA's overfitting problem and briefly reviews relevant previous work in addressing it. Section III presents a BD-PCA image feature extraction method and proposes an AMD metric. This section also introduces the implementation of NN and NFL classifiers. Section IV presents the results of experiments using the Olivetti Research Laboratory (ORL) face database and the PolyU palm print database. Section V offers our conclusion.

II. Overfitting Problem of PCA

A. Overfitting Problem

When applied to image recognition, due to the SSS problem, the classical PCA is apt to be overfitted to the training set. As a statistical method, the classical PCA's statistical meaning is problematic when the sample data dimension is much higher than the number of samples. To prove this, we carried out a series of experiments using the ORL database [22].

The ORL database contains 400 facial images of 40 persons with ten images per individual. All the images are taken against a dark homogeneous background but vary in sampling time, light conditions, facial expressions, facial details (glasses/no glasses), scale, and tilt. The size of these gray images is 112 × 92 [22]. Fig. 1 shows the ten images of one person.

We use the normalized mean square error (mse) to evaluate the overfitting problem. One statistical characteristic of PCA is that the mse between random vector $\mathbf{x}$ and its subspace projection is minimal [23]. Thus, the difference of mse on

Manuscript received December 28, 2004; revised July 18, 2005. This paper was recommended by Associate Editor V. Marino.

W. Zuo and K. Wang are with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: cswmzuo@gmail.com; wangkq@hit.edu.cn).

D. Zhang is with the Biometrics Research Centre, Department of Computing, Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: csdzhang@comp.polyu.edu.hk).

Digital Object Identifier 10.1109/TSMCB.2006.872274

Fig. 1. Ten images of an individual in the ORL face database.

the training set and the testing set indicates the degree of overfitting. If PCA is overfitted to the training set, the mse on the training set would be much lower than on the testing set.

Given the first L principal components (PCs), we define the PCA projector as $\mathbf{W}_L = [\varphi_1, \varphi_2, \ldots, \varphi_L]$. A vector $\mathbf{x}$ is then transformed into PCA subspace by

$$\mathbf{y} = \mathbf{W}_L^T(\mathbf{x} - \overline{\mathbf{x}}). \tag{1}$$

The reconstructed vector $\tilde{\mathbf{x}}$ can be represented as

$$\tilde{\mathbf{x}} = \overline{\mathbf{x}} + \mathbf{W}_L\mathbf{y} = \overline{\mathbf{x}} + \mathbf{W}_L\mathbf{W}_L^T(\mathbf{x} - \overline{\mathbf{x}}). \tag{2}$$

The normalized mse on the training set $\text{mse}_L^{\text{train}}$ is then defined as

$$\text{mse}_L^{\text{train}} = \frac{\sum_i \left\|\mathbf{x}_i^{\text{train}} - \tilde{\mathbf{x}}_i^{\text{train}}\right\|^2}{\sum_i \left\|\mathbf{x}_i^{\text{train}} - \overline{\mathbf{x}}^{\text{train}}\right\|^2} \tag{3}$$

where $\mathbf{x}_i^{\text{train}}$ is the ith training sample, $\tilde{\mathbf{x}}_i^{\text{train}}$ is the reconstructed vector of $\mathbf{x}_i^{\text{train}}$, and $\overline{\mathbf{x}}^{\text{train}}$ is the mean vector of all training samples. Similarly, we can calculate the normalized mse on the testing set $\text{mse}_L^{\text{test}}$.

We choose the first five images of each person for training, while the remaining five images are used for testing. Thus, we obtain a training set of 200 images and a testing set of 200 images. We then calculate $\text{mse}_L^{\text{train}}$ and $\text{mse}_L^{\text{test}}$ for a given $\mathbf{W}_L$, as shown in Fig. 2. We can observe that the difference between $\text{mse}_L^{\text{test}}$ and $\text{mse}_L^{\text{train}}$ would increase with the increase of L, and the difference could be up to 20% when $L = 100$. The great difference between $\text{mse}_L^{\text{train}}$ and $\text{mse}_L^{\text{test}}$ indicates that the classical PCA is overfitted to the training set.

B. Previous Work in Solving the PCA Overfitting Problem

To date, no researchers have directly pointed out the classical PCA overfitting problem, yet some PCA-based methods have been proposed to deal with it. In the following, we survey three representative approaches, namely, 1) (PC)^{2}A [7], [8], 2) 2DPCA [9]–[11], and 3) modular PCA [12], [24], [25].

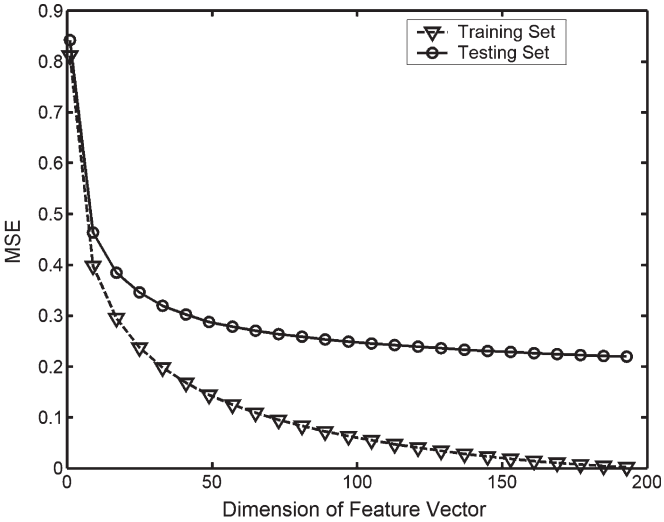


Fig. 2. PCA's normalized mse on the training set and the testing set as a function of feature dimension.

1) (PC)^{2}A: (PC)^{2}A adopted an image preprocessing plus PCA mechanism [7]. Given an $m \times n$ image $I(x, y)$, Wu and Zhou define its vertical and horizontal integral projections by

$$V_p(x) = \frac{1}{n}\sum_{y=1}^{n} I(x, y) \tag{4}$$

$$H_p(y) = \frac{1}{m}\sum_{x=1}^{m} I(x, y). \tag{5}$$

Then, the projection map $M_p(x, y)$ is defined by

$$M_p(x, y) = \frac{V_p(x)H(y)}{\overline{I}} \tag{6}$$

where $\overline{I}$ is the average intensity of the image. We then obtain $I_\alpha(x, y)$, the projection-combined version of $I(x, y)$ with combination parameter α, i.e.,

$$I_\alpha(x, y) = (1 - \alpha)I(x, y) + \alpha M_p(x, y). \tag{7}$$

Finally, classical PCA is performed on the projection-combined version of $I(x, y)$.

Since the projection map $M_p(x, y)$ is generated by the vertical and horizontal integral projections $V_p(x)$ and $H_p(x)$, the intrinsic dimension of $M_p(x, y)$ should be less than $(m + n - 1)$ [7]. $I_\alpha(x, y)$ is a mixture of high-dimensional $I(x, y)$ and low-dimensional $M_p(x, y)$. The employment of $I_\alpha(x, y)$ in PCA can thus alleviate the overfitting problem. We considered that this is the reason for the better recognition performance of (PC)^{2}A over classical PCA.

2) 2DPCA: As we shall now show, 2DPCA is a row PCA that regards an $m \times n$ image matrix as an m-set of $1 \times n$ row vectors. Let $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N\}$ be a training set of N images. The image total scatter matrix $\mathbf{G}_t$ in [9] and [10] is defined by

$$\mathbf{G}_t = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{X}_i - \overline{\mathbf{X}})^T(\mathbf{X}_i - \overline{\mathbf{X}}) \tag{8}$$

where $\mathbf{X}_i$ denotes the ith image and $\overline{\mathbf{X}}$ denotes the global mean of all images. By representing $\mathbf{X}_i$ as an m-set of $1 \times n$ row vectors

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{x}_i^1 \\ \mathbf{x}_i^2 \\ \vdots \\ \mathbf{x}_i^m \end{bmatrix} \tag{9}$$

the image total scatter matrix $\mathbf{G}_t$ can be rewritten as

$$\mathbf{G}_t = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{m} \left(\mathbf{x}_i^j - \overline{\mathbf{x}}^j\right)^T \left(\mathbf{x}_i^j - \overline{\mathbf{x}}^j\right) \tag{10}$$

where $\mathbf{x}_i^j$ and $\overline{\mathbf{x}}_i^j$ denote the jth row of $\mathbf{X}_i$ and $\overline{\mathbf{X}}$, respectively. We can define the row total scatter matrix $\mathbf{S}_t^{\text{row}}$ as the scatter matrix of all the row vectors as

$$\mathbf{S}_t^{\text{row}} = \frac{1}{Nm} \sum_{i=1}^{N} \sum_{j=1}^{m} \left(\mathbf{x}_i^j - \overline{\mathbf{x}}\right)^T \left(\mathbf{x}_i^j - \overline{\mathbf{x}}\right) \tag{11}$$

where $\overline{\mathbf{x}}$ is the mean vector of all row vectors in $\overline{\mathbf{X}}$. Comparing (10) with (11), the computation of $\mathbf{G}_t$ and $\mathbf{S}_t^{\text{row}}$ is almost the same except for the substitution of $\overline{\mathbf{x}}$ for $\overline{\mathbf{x}}_i^j$ and the addition of the constant $1/m$. Consequently, we argue that 2DPCA is actually a variant of row PCA.

2DPCA treats an image as m $1 \times n$ row vectors and performs PCA on all row vectors in the training set. In 2DPCA, the actual vector dimension is n, the actual sample size is mN, and $n \ll mN$. Thus, the SSS problem is solved. Despite its advantages, 2DPCA still suffers from the high feature dimension problem (one typical dimension in [9] is 8 × 112). Yang proposed solving this problem with a 2DPCA plus PCA method, but the 2DPCA + PCA strategy would reintroduce the SSS problems.

3) Modular PCA: In modular PCA, an image is divided into n_1 subimages and PCA is performed on all these subimages [12]. Given an $m \times n$ image $I(x, y)$, these subimages can be represented as

$$I_{ij}(k, l) = I\left(\frac{m}{\sqrt{n_1}}(i-1) + k, \frac{n}{\sqrt{n_1}}(j-1) + l\right) \tag{12}$$

where I_{ij} denotes the vertical ith and horizontal jth subimages.

Since modular PCA divides an image into a number of subimages, the actual vector dimension in modular PCA will be much lower than in classical PCA. The number of training vectors used in modular PCA is much higher than the number used in classical PCA. Thus, modular PCA can be used to solve the overfitting problem.

Modular PCA, however, still has some problems. One is how to determine the number of subimages. For example, Toygar and Acan proposed dividing an image into five horizontal subimages [24], and Chen proposed a 5 × 5 and 5 × 3 partitions of an image [25]. Another problem is that the feature dimension will increase as we increase of the number of subimages.

In summary, none of these three methods provide an entirely satisfactory solution to the overfitting problem. (PC)^{2}A merely alleviates the overfitting problem by using a intrinsic low-dimensional image to blur the original image. 2DPCA and modular PCA both solve the overfitting problems by reducing the dimensionality and by increasing the training vectors yet introduce the high feature dimension problem. Finally, 2DPCA can be regarded as a special case of modular PCA where each row of the original image is regarded as a subimage.

III. BD-PCA WITH AMD METRIC

In this section, we first propose a BD-PCA method to solve the overfitting problem and then present an AMD metric and apply it to the NN or NFL classifiers.

A. BD-PCA

In classical PCA, an $m \times n$ image $\mathbf{X}$ should be mapped into a high-dimensional $mn \times 1$ vector $\mathbf{x}$ in advance, and then a $d_{\text{PCA}} \times 1$ feature vector $\mathbf{y}$ of $\mathbf{x}$ $(d_{\text{PCA}} \ll mn)$ is obtained by

$$\mathbf{y} = \mathbf{W}_{\text{pca}}^T \mathbf{x} \tag{13}$$

where $\mathbf{W}_{\text{pca}}$ is the PCA projector. BD-PCA, however, is a straightforward image projection technique where a $k_{\text{col}} \times k_{\text{row}}$ feature matrix $\mathbf{Y}$ of an $m \times n$ image $\mathbf{X}$ $(k_{\text{col}} \ll m,\ k_{\text{row}} \ll n)$ can be obtained by

$$\mathbf{Y} = \mathbf{W}_{\text{col}}^T \mathbf{X} \mathbf{W}_{\text{row}} \tag{14}$$

where $\mathbf{W}_{\text{col}}$ is the column projector and $\mathbf{W}_{\text{row}}$ is the row projector. BD-PCA has at least two advantages over PCA. First, being a straightforward image projection technique, BD-PCA does not require mapping an image $\mathbf{X}$ to an image vector $\mathbf{x}$. Second, BD-PCA-based feature extraction generally has a lower computational requirement than PCA-based feature extraction. PCA requires $m \times n \times d_{\text{PCA}}$ multiplications, while BD-PCA requires $m \times n \times k_{\text{row}} + m \times k_{\text{col}} \times k_{\text{row}}$ multiplications. Since, in general, $k_{\text{row}} < d_{\text{PCA}}$, BD-PCA-based feature extraction is computationally more efficient than PCA.

We next present our method for calculating $\mathbf{W}_{\text{col}}$ and $\mathbf{W}_{\text{row}}$. Let $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N\}$ be a training set of N images. By representing the ith image matrix $\mathbf{X}_i$ as an m-set of $1 \times n$ row vectors, the row total scatter matrix $\mathbf{S}_t^{\text{row}}$ is defined by

$$\mathbf{S}_t^{\text{row}} = \frac{1}{Nm} \sum_{i=1}^{N} (\mathbf{X}_2 - \overline{\mathbf{X}})^T (\mathbf{X}_2 - \overline{\mathbf{X}}) \tag{15}$$

where $\overline{\mathbf{X}}$ is the mean matrix of all training images. We choose the row eigenvectors corresponding to the first k_{row} largest eigenvalues of $\mathbf{S}_t^{\text{row}}$ to construct the row projector

$$\mathbf{W}_{\text{row}} = \left[\mathbf{w}_1^{\text{row}}, \mathbf{w}_2^{\text{row}}, \ldots, \mathbf{w}_{k_{\text{row}}}^{\text{row}}\right]. \tag{16}$$

By treating an image matrix $\mathbf{X}_i$ as an n-set of $m \times 1$ column vectors

$$\mathbf{X}_i = \left[\mathbf{x}_i^1,\ \mathbf{x}_i^2, \ldots, \mathbf{x}_i^n\right] \tag{17}$$

we define the column total scatter matrix

$$\mathbf{S}_t^{\mathrm{col}} = \frac{1}{Nn}\sum_{i=1}^{N}(\mathbf{X}_2 - \overline{\mathbf{X}})(\mathbf{X}_2 - \overline{\mathbf{X}})^T. \tag{18}$$

We then choose the first k_{col} column eigenvectors corresponding to the first k_{col} largest eigenvalues of $\mathbf{S}_t^{\mathrm{col}}$ to construct the column projector

$$\mathbf{W}_{\mathrm{col}} = \left[\mathbf{w}_1^{\mathrm{col}}, \mathbf{w}_2^{\mathrm{col}}, \ldots, \mathbf{w}_{k_{\mathrm{col}}}^{\mathrm{col}}\right]. \tag{19}$$

Note that BD-PCA is a generalization of Yang's 2DPCA. 2DPCA can be regarded as a special BD-PCA with $\mathbf{W}_{\mathrm{col}} = \mathbf{I}_m$, where $\mathbf{I}_m$ denotes an $m \times m$ identity matrix [10].

B. AMD Metric

Distance measures seriously affect the recognition performance of the NN classifier. This motivates us to investigate distance measures subsequent to extracting the BD-PCA feature matrix. The distance of two feature matrices can be calculated using either a vector-based or matrix-based method. In a vector-based method, a feature matrix is first mapped to a vector and then a vector-based distance measure is used. In a matrix-based method, the distance between two feature matrices can be directly computed. In fact, given a vector-based distance measure, we could find an equivalent matrix-based distance measure to calculate the distance. For example, it is easy to prove the equivalence of the Euclidean distance (vector) and the Frobenius distance (matrix) [26]. Thus, we only investigate the matrix-based distance measure.

Definition 1: Given two feature matrices $\mathbf{A} = (a_{ij})_{k_{\mathrm{col}} \times k_{\mathrm{row}}}$ and $\mathbf{B} = (b_{ij})_{k_{\mathrm{col}} \times k_{\mathrm{row}}}$, the AMD distance $d_{\mathrm{AMD}}(\mathbf{A}, \mathbf{B})$ is defined as

$$d_{\mathrm{AMD}}(\mathbf{A}, \mathbf{B}) = \left(\sum_{j=1}^{k_{\mathrm{row}}}\left(\sum_{i=1}^{k_{\mathrm{col}}}|a_{ij} - b_{ij}|^{p_1}\right)^{\frac{p_2}{p_1}}\right)^{\frac{1}{p_2}}, \quad (p_1, p_2 > 0). \tag{20}$$

Definition 2: A matrix norm on $\mathbb{R}^{k_{\mathrm{col}} \times k_{\mathrm{row}}}$ is a function $f : \mathbb{R}^{k_{\mathrm{col}} \times k_{\mathrm{row}}} \to \mathbb{R}$ with the properties [26]

$$\begin{aligned} f(\mathbf{A}) &\geq 0, \quad \mathbf{A} \in \mathbb{R}^{k_{\mathrm{col}} \times k_{\mathrm{row}}}(f(\mathbf{A}) = 0 \Leftrightarrow \mathbf{A} = 0) \\ f(\mathbf{A} + \mathbf{B}) &\leq f(\mathbf{A}) + f(\mathbf{B}), \quad \mathbf{A}, \mathbf{B} \in \mathbb{R}^{k_{\mathrm{col}} \times k_{\mathrm{row}}} \\ f(\alpha\mathbf{A}) &\leq |\alpha| f(\mathbf{A}), \quad \alpha \in \mathbb{R}, \mathbf{A} \in \mathbb{R}^{k_{\mathrm{col}} \times k_{\mathrm{row}}}. \end{aligned} \tag{21}$$

Theorem 1: The function $\|\mathbf{x}\|_p = (\sum_i |x_i|^p)^{1/p}$ is a vector norm [26].

Theorem 2: The function $\|\mathbf{A}\|_{\mathrm{AMD}} = (\sum_{j=1}^{k_{\mathrm{row}}}(\sum_{i=1}^{k_{\mathrm{col}}} \times (|a_{ij}|)^{p_1})^{p_2/p_1})^{1/p_2}$ is a matrix norm.

Proof: It can be easily proved that

$$\begin{aligned} \|\mathbf{A}\|_{\mathrm{AMD}} &\geq 0 \\ \|\mathbf{A}\|_{\mathrm{AMD}} &= 0 \Leftrightarrow \mathbf{A} = 0 \\ \|\alpha\mathbf{A}\|_{\mathrm{AMD}} &= |\alpha|\|\mathbf{A}\|_{\mathrm{AMD}}. \end{aligned}$$

Next, we shall prove $\|\mathbf{A} + \mathbf{B}\|_{\mathrm{AMD}} \leq \|\mathbf{A}\|_{\mathrm{AMD}} + \|\mathbf{B}\|_{\mathrm{AMD}}$. From Theorem 1

$$\begin{aligned} \|\mathbf{A} + \mathbf{B}\|_{\mathrm{AMD}} &= \left(\sum_{j=1}^{k_{\mathrm{row}}}\left(\sum_{i=1}^{k_{\mathrm{col}}}|a_{ij} - b_{ij}|^{p_1}\right)^{\frac{p_2}{p_1}}\right)^{\frac{1}{p_2}} \\ &\leq \left(\sum_{j=1}^{k_{\mathrm{row}}}\left(\left\|\mathbf{a}^{(j)}\right\|_{p_1} + \left\|\mathbf{b}^{(j)}\right\|_{p_1}\right)^{p_2}\right)^{\frac{1}{p_2}}. \end{aligned}$$

From Theorem 1, the function $g(\mathbf{a}) = \sum_{j=1}^{k_{\mathrm{row}}}(\|\mathbf{a}^{(j)}\|_{p_1})^{p_2})^{1/p_2}$, $(\mathbf{a} = [\|\mathbf{a}^{(1)}\|_{p_1}, \ldots, \|\mathbf{a}^{(k_{\mathrm{row}})}\|_{p_1}]^T)$ is a vector norm. Let $\mathbf{b} = [\|\mathbf{b}^{(1)}\|_{p_1}, \ldots, \|\mathbf{b}^{(k_{\mathrm{row}})}\|_{p_1}]^T$

$$\begin{aligned} &\left(\sum_{j=1}^{k_{\mathrm{row}}}\left(\left\|\mathbf{a}^{(j)}\right\|_{p_1} + \left\|\mathbf{b}^{(j)}\right\|_{p_1}\right)^{p_2}\right)^{\frac{1}{p_2}} \\ &\quad = g(\mathbf{a} + \mathbf{b}) \leq g(\mathbf{a}) + g(\mathbf{b}) \\ &\quad = \left(\sum_{j=1}^{k_{\mathrm{row}}}\left(\left\|\mathbf{a}^{(j)}\right\|_{p_1}\right)^{p_2}\right)^{\frac{1}{p_2}} + \left(\sum_{j=1}^{k_{\mathrm{row}}}\left(\left\|\mathbf{b}^{(j)}\right\|_{p_1}\right)^{p_2}\right)^{\frac{1}{p_2}} \\ &\quad = \left(\sum_{j=1}^{k_{\mathrm{row}}}\left(\sum_{i=1}^{k_{\mathrm{col}}} a_{ij}^{p_1}\right)^{\frac{p_2}{p_1}}\right)^{\frac{1}{p_2}} + \left(\sum_{j=1}^{k_{\mathrm{row}}}\left(\sum_{i=1}^{k_{\mathrm{col}}} b_{ij}^{p_1}\right)^{\frac{p_2}{p_1}}\right)^{\frac{1}{p_2}} \\ &\quad = \|\mathbf{A}\|_{\mathrm{AMD}} + \|\mathbf{B}\|_{\mathrm{AMD}}. \end{aligned}$$

So $\|\mathbf{A} + \mathbf{B}\|_{\mathrm{AMD}} \leq \|\mathbf{A}\|_{\mathrm{AMD}} + \|\mathbf{B}\|_{\mathrm{AMD}}$, and $\|\mathbf{A}\|_{\mathrm{AMD}}$ is a matrix norm. ■

Definition 3: A metric in $\mathbb{R}^{k_{\mathrm{col}} \times k_{\mathrm{row}}}$ is a function $f : \mathbb{R}^{k_{\mathrm{col}} \times k_{\mathrm{row}}} \times \mathbb{R}^{k_{\mathrm{col}} \times k_{\mathrm{row}}} \to \mathbb{R}$ with the properties [27]

$$\begin{aligned} f(\mathbf{A}, \mathbf{B}) &\geq 0, \mathbf{A}, \mathbf{B} \in \mathbb{R}^{k_{\mathrm{col}} \times k_{\mathrm{row}}} \\ f(\mathbf{A}, \mathbf{B}) &= 0 \Leftrightarrow \mathbf{A} = \mathbf{B} \\ f(\mathbf{A}, \mathbf{B}) &= f(\mathbf{B}, \mathbf{A}) \\ f(\mathbf{A}, \mathbf{B}) &\leq f(\mathbf{A}, \mathbf{C}) + f(\mathbf{C}, \mathbf{B}). \end{aligned} \tag{22}$$

Theorem 3: $d_{\mathrm{AMD}}(\mathbf{A}, \mathbf{B})$ is a distance metric.

Proof: The function $\|\mathbf{A}\|_{\mathrm{AMD}}$ is a matrix norm, and it is easy to prove that $d_{\mathrm{AMD}}(\mathbf{A}, \mathbf{B}) = \|\mathbf{A} - \mathbf{B}\|_{\mathrm{AMD}}$ is a distance metric derived from the matrix norm $\|\mathbf{A}\|_{\mathrm{AMD}}$. ■

Corollary 1: The Frobenius distance measure [26] $d_F(\mathbf{A}, \mathbf{B}) = (\sum_{j=1}^{k_{\mathrm{row}}}\sum_{i=1}^{k_{\mathrm{col}}}(a_{ij} - b_{ij})^2)^{1/2}$ is a special case of the AMD metric with $p_1 = p_2 = 2$.

Corollary 2: The Yang distance measure [10] $d_Y(\mathbf{A}, \mathbf{B}) = \sum_{j=1}^{k_{\mathrm{row}}}(\sum_{i=1}^{k_{\mathrm{col}}}(a_{ij} - b_{ij})^2)^{1/2}$ is a special case of AMD metric with $p1 = 2$ and $p_2 = 1$.

Next, we shall prove that when using an NN classifier based on the AMD metric with $p_1 = 2$, 2DPCA is equivalent to BD-PCA with $k_{\mathrm{col}} = m$. It is easy to deduce the correctness of this statement from the proof of the following theorem.

Theorem 4: Given an $m \times n$ image $\mathbf{X}$, 2DPCA extracts the feature matrix $\mathbf{A}$ by $\mathbf{A} = \mathbf{X}\mathbf{W}_{\mathrm{row}}$ and BD-PCA extracts the feature matrix $\mathbf{B}$ by $\mathbf{B} = \mathbf{W}_{\text{col}}\mathbf{X}\mathbf{W}_{\text{row}} = \mathbf{W}_{\text{col}}\mathbf{A}$. If $p_1 = 2$ and $\mathbf{W}_{\text{col}} = [\mathbf{w}_1^{\text{col}}, \mathbf{w}_2^{\text{col}}, \ldots, \mathbf{w}_m^{\text{col}}]$, the matrices $\mathbf{A}$ and $\mathbf{B}$ have the same AMD norms, i.e., $\|\mathbf{A}\|_{\text{AMD}} = \|\mathbf{B}\|_{\text{AMD}}$.

Proof: From the definition of the AMD norm, we calculate $\|\mathbf{A}\|_{\text{AMD}} = \left(\sum_{j=1}^{k_{\text{row}}}\left(\sum_{i=1}^{m}(a_{ij})^2\right)^{p_2/2}\right)^{1/p_2}$ and $\|\mathbf{B}\|_{\text{AMD}} = \left(\sum_{j=1}^{k_{\text{row}}}\left(\sum_{i=1}^{m}(b_{ij})^2\right)^{p_2/2}\right)^{1/p_2}$. First, we shall prove $\sum_{i=1}^{m}(a_{ij})^2 = \sum_{i=1}^{m}(b_{ij})^2$, i.e.,

$$\begin{aligned}\sum_{i=1}^{m}(b_{ij})^2 &= \mathbf{b}_j^T\mathbf{b}_j, \quad \left(\mathbf{b}_j = [b_{1j}, \ldots, b_{ij}, \ldots, b_{mj}]^T\right)\\ &= (\mathbf{W}_{\text{col}}\mathbf{a}_j)^T(\mathbf{W}_{\text{col}}\mathbf{a}_j)\\ &\quad \times \left(\mathbf{a}_j = [a_{1j}, \ldots, a_{ij}, \ldots, a_{mj}]^T\right)\\ &= \mathbf{a}_j^T\left(\mathbf{W}_{\text{col}}^T\mathbf{W}_{\text{col}}\right)\mathbf{a}_j\\ &= \mathbf{a}_j^T\mathbf{a}_j \quad (\mathbf{W}_{\text{col}} \text{ is a complete orthogonal}\\ &\qquad\qquad \text{basis of } m\text{D space})\\ &= \sum_{i=1}^{m}(a_{ij})^2.\end{aligned}$$

Because $\sum_{i=1}^{m}(a_{ij})^2 = \sum_{i=1}^{m}(b_{ij})^2$, it is obvious that $\|\mathbf{A}\|_{\text{AMD}} = \|\mathbf{B}\|_{\text{AMD}}$. ■

C. Classifiers

We use two classifiers for image recognition, namely: 1) NN and 2) NFL classifiers. In the NN classifier, the feature matrix is classified as belonging to the class with the nearest template. Given all the templates $\{\mathbf{M}_{cl},\ 1 \le c \le C,\ 1 \le l \le n_c\}$ and the query feature $\mathbf{Y}$, the NN rule can be expressed as

$$d(\mathbf{Y}, \mathbf{M}_{\hat{c}\hat{l}}) = \min_{\{1\le c\le C, 1\le l\le n_c} d(\mathbf{Y}, \mathbf{M}_{cl}) \Rightarrow \mathbf{Y} \in w_{\hat{c}} \tag{23}$$

where C is the number of classes, n_c is the number of templates in class w_c, and $d(\mathbf{Y}, \mathbf{M}_{cl})$ denotes the distance between $\mathbf{Y}$ and $\mathbf{M}_{cl}$.

NFL, an extension of NN, can extend the representative capacity of templates by using linear interpolation and extrapolation [17]. Given two templates $\mathbf{M}_{cl}$ and $\mathbf{M}_{ck}$, the distance between the feature point $\mathbf{Y}$ and the feature line $\overline{\mathbf{M}_{cl}\mathbf{M}_{ck}}$ is defined as

$$d(\mathbf{Y}, \overline{\mathbf{M}_{cl}\mathbf{M}_{ck}}) = d(\mathbf{Y}, \mathbf{Y}_p) \tag{24}$$

where $\mathbf{Y}_p = \mathbf{M}_{cl} + \mu(\mathbf{M}_{ck} - \mathbf{M}_{cl})$ and $\mu = (\mathbf{Y} - \mathbf{M}_{cl}) \times (\mathbf{M}_{ck} - \mathbf{M}_{cl})/(\mathbf{M}_{ck} - \mathbf{M}_{cl}) \times (\mathbf{M}_{ck} - \mathbf{M}_{cl})$. Then, NFL determines the class $w_{\hat{c}}$ of the query feature $\mathbf{Y}$ according to

$$d(\mathbf{Y}, \overline{\mathbf{M}_{\hat{c}\hat{l}}\mathbf{M}_{\hat{c}\hat{k}}}) = \min_{\{1\le c\le C, 1\le l<k\le n_c\}} d(\mathbf{Y}, \overline{\mathbf{M}_{cl}\mathbf{M}_{ck}}) \Rightarrow \mathbf{Y} \in w_{\hat{c}}. \tag{25}$$

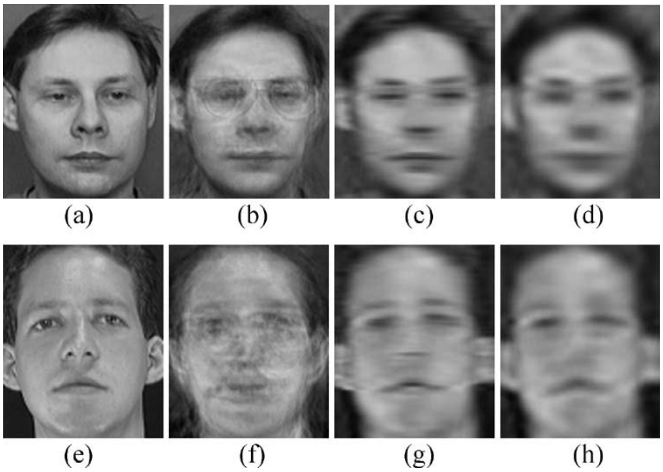


Fig. 3. Comparison of the reconstruction capability of PCA, 2DPCA, and BD-PCA. (a) and (e) Original images. (b) and (f) Reconstructed images by PCA. (c) and (g) Reconstructed images by 2DPCA. (d) and (h) Reconstructed images by BD-PCA.

IV. Experimental Results and Discussion

To evaluate BD-PCA with AMD metric (BD-PCA-AMD), we used two image databases, namely: 1) the ORL face database and 2) the PolyU palm print database. For each database, we investigate the effect of BD-PCA-AMD parameters and compare the error rates obtained using different distance measures.

A. Experimental Results on the ORL Database

In order to test BD-PCA-AMD, a series of experiments are carried out using the ORL database. First, we give an intuitive illustration of the reconstruction capability of BD-PCA. Then, we evaluate the capability of BD-PCA to solve the overfitting problem. We also evaluate the efficacy of the AMD metric and compare the recognition performance of BD-PCA-AMD with that of PCA and 2DPCA.

In the first set of experiments, we compare the reconstruction capability of PCA, 2DPCA, and BD-PCA. PCA-based reconstruction always involves the determination of feature dimensions. For PCA reconstruction, we evaluate the reconstructed training images based on 50, 100, 150, and 199 PCs, and find that the reconstruction quality is satisfactory when the number of PCs arrives at 100. For simplicity, we do not show these reconstructed images. For 2DPCA reconstruction, we evaluate the reconstructed images based on 5, 10, 15, and 20 row eigenvectors, and find that the reconstruction quality is satisfactory when the number of row eigenvectors k_{row} arrives at 10. For BD-PCA reconstruction, we first let $k_{\text{row}} = 10$ according to the reconstruction evaluation on 2DPCA. We then compare the reconstructed images based on 10, 15, 20, and 30 column eigenvectors, and find that the reconstruction quality is satisfactory when $k_{\text{col}} = 20$. Thus, we determine the feature dimensions of PCA, 2DPCA, and BD-PCA. The feature dimension of 2DPCA is $112 \times 10 = 1120$, much higher than that of BD-PCA ($20 \times 10 = 200$) and of PCA (100).

Fig. 3 shows two original images and their reconstructed images by PCA, 2DPCA, and BD-PCA. Fig. 3(a) is an original

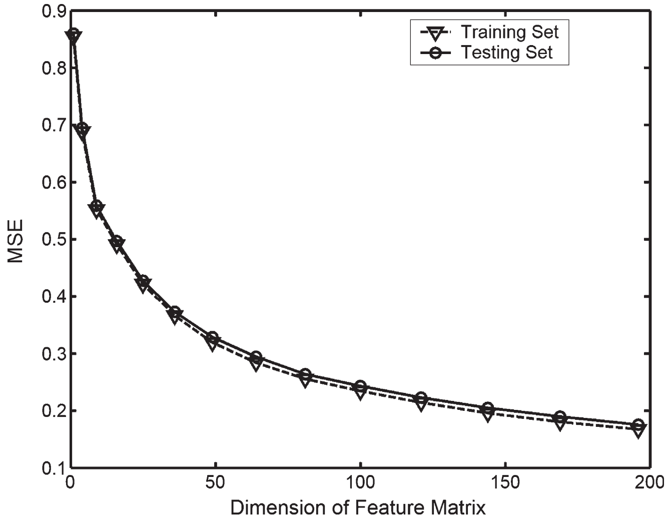


Fig. 4. BD-PCA's normalized mse on the training set and the testing set as a function of feature dimension.

image from the training set. Satisfactorily reconstructed images in Fig. 3(a) can be obtained using all of these three methods, as shown in Fig. 3(b)–(d). Fig. 3(e) is an image from the testing set and its reconstructed images are shown in Fig. 3(f)–(h). The quality of the reconstructed image by PCA deteriorates greatly, while both 2DPCA and BD-PCA can obtain a satisfactory reconstruction quality.

In the second set of experiments, we use the normalized mse to evaluate BD-PCA's capability to solve the overfitting problem. Given a column projector $\mathbf{W}_{\text{col}}$ and a row projector $\mathbf{W}_{\text{row}}$, an image $\mathbf{X}$ is mapped into its BD-PCA representation

$$\mathbf{Y} = \mathbf{W}_{\text{col}}^{T}(\mathbf{X} - \overline{\mathbf{X}})\mathbf{W}_{\text{row}}. \tag{26}$$

The reconstructed image $\tilde{\mathbf{X}}$ can then be represented as

$$\tilde{\mathbf{X}} = \overline{\mathbf{X}} + \mathbf{W}_{\text{col}}\mathbf{Y}\mathbf{W}_{\text{row}}^{T}. \tag{27}$$

Then, the normalized mse on the training set $\text{mse}^{\text{train}}$ can be defined as

$$\text{mse}^{\text{train}} = \frac{\sum_i \left\| \mathbf{X}_i^{\text{train}} - \tilde{\mathbf{X}}_i^{\text{train}} \right\|_F^2}{\sum_i \left\| \mathbf{X}_i^{\text{train}} - \overline{\mathbf{X}}^{\text{train}} \right\|_F^2} \tag{28}$$

where $\mathbf{X}_i^{\text{train}}$ is the ith training image matrix, $\tilde{\mathbf{X}}_i^{\text{train}}$ is the reconstructed image of $\mathbf{X}_i^{\text{train}}$, and $\overline{\mathbf{X}}^{\text{train}}$ is the mean matrix of all training images. Similarly, we can define the normalized mse on the testing set as mse^{test}.

Using the method discussed in Section II-A to construct the training set and the testing set, we calculate $\text{mse}^{\text{train}}$ and mse^{test}, as shown in Fig. 4. For simplicity, we let the number of the row eigenvectors k_{row} equal the number of column eigenvectors k_{col} and thus the dimension of the feature matrix $L = k_{\text{row}}^2$. From Fig. 4, the difference of $\text{mse}^{\text{train}}$ and mse^{test} is very small. Thus, BD-PCA can solve the overfitting problem.

In all the following experiments, we randomly choose n_p images of each person for training, resulting in a training set of $40 \times n_p$ images and a testing set of $40(10 - n_p)$ images with no overlap between the two sets. To reduce the variation of recognition results, an average error rate (AER) is adopted by calculating the mean of error rates over 20 runs.

In the third set of experiments, we study the effect of BD-PCA-AMD parameters. In BD-PCA-AMD, there are four parameters $(k_{\text{row}},\ k_{\text{col}},\ p_1,\ p_2)$. A cross-validation strategy is used to determine these parameters. In our experiments, we run BD-PCA-AMD 40 times. Each time we construct a training set and a testing set. The first 20 runs are used to determine the values of these parameters that correspond to the lowest AER. After parameter selection, we use the AER of the other 20 runs to evaluate recognition performance.

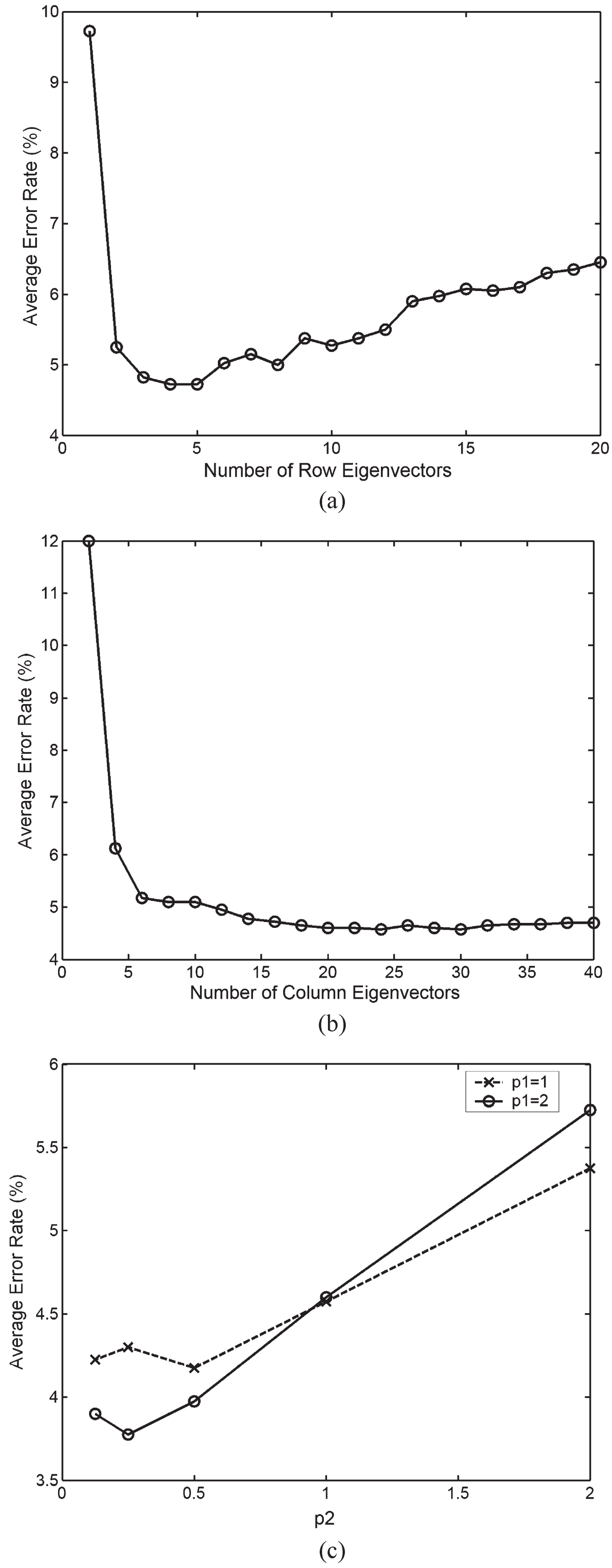
Since it is very difficult to determine these four parameters at the same time, we adopt a stepwise selection strategy. We prefix k_{col} to its maximum $(p_1 = 2, p_2 = 1)$ and try to find the optimal k_{row} value, as shown in Fig. 5(a). After the determination of $k_{\text{row}} = 4$, we pre-fix k_{row} to its optimal value $(p_1 = 2, p_2 = 1)$ and try to find the optimal k_{col} value, as shown in Fig. 5(b). After the determination of $k_{\text{col}} = 18$ and $k_{\text{row}} = 4$, we pre-fix k_{col} and k_{row} to its optimal values and study the effect of p_1 and p_2 on AER, as shown in Fig. 5(c). Stimulated by the study on PCA-based distance measure, we consider only two possible values for p_1, 1 and 2. Thus, we determine the values of these four parameters $(k_{\text{row}} = 4, k_{\text{col}} = 18, p_1 = 2, p_2 = 0.25)$.

After the determination of BD-PCA-AMD parameters, Table I compares the AER obtained using the Frobenius, the Yang, and the AMD distance measures on another 20 runs. The AMD metric achieves the lowest AER for NN. NFL with AMD measure also achieves a lower AER than NFL using the other two distance measures. In the following experiments, BD-PCA-NN denotes BD-PCA with AMD using the NN classifier and BD-PCA-NFL denotes BD-PCA with AMD using NFL.

Fig. 6 depicts the AER with different n_p values. An interesting point to note is that the improvement of BD-PCA-NFL over BD-PCA-NN is very small when the number of training samples $n_p \geq 7$. This observation indicates that the recognition performance of NN would be comparable to that of NFL when there are sufficient templates for each class.

In the fourth set of experiments, we carry out a comparative study of PCA and BD-PCA with $n_p = 5$. Fig. 7 plots the error rates of 20 runs obtained using PCA and BD-PCA, while Table II lists the AER and standard deviation (std.) of each method. For both NN and NFL, BD-PCA has a lower AER and standard deviation than PCA. For NN, the AER of BD-PCA is about 0.597 of that of PCA, and for NFL the AER of BD-PCA is about 0.577 of that of PCA.

In the fifth set of experiments, we compare the recognition performance of 2DPCA and BD-PCA with $n_p = 5$. Fig. 8 plots the error rates of 20 runs obtained using 2DPCA and BD-PCA, while Table II shows the AER and standard deviation of each method. We can observe that for the NN classifier the AER obtained using BD-PCA is about 0.813 of that obtained using 2DPCA, and for NFL is about 0.818. The standard deviation of BD-PCA is very close to that of 2DPCA for both NN and NFL. In addition, the feature dimension of BD-PCA is $18 \times 4 = 72$, much lower than the feature dimension of 2DPCA $(112 \times 4 = 448)$.



(a)

(b)

(c)

Fig. 5. Plots of the AERs over the variation of BD-PCA-AMD parameters. (a) AERs versus the number of row eigenvectors. (b) AERs versus the number of column eigenvectors. (c) AERs versus the value of p_2 with $p_1 = 1$ and $p_1 = 2$.

TABLE I
COMPARISON OF AERs OBTAINED USING DIFFERENT DISTANCE MEASURES AND CLASSIFIERS ON THE ORL DATABASE

Classifier	Frobenius	Yang	AMD
NN	5.05±1.69	4.48±1.27	3.55±1.01
NFL	4.10±1.54	3.48±1.36	2.70±0.86

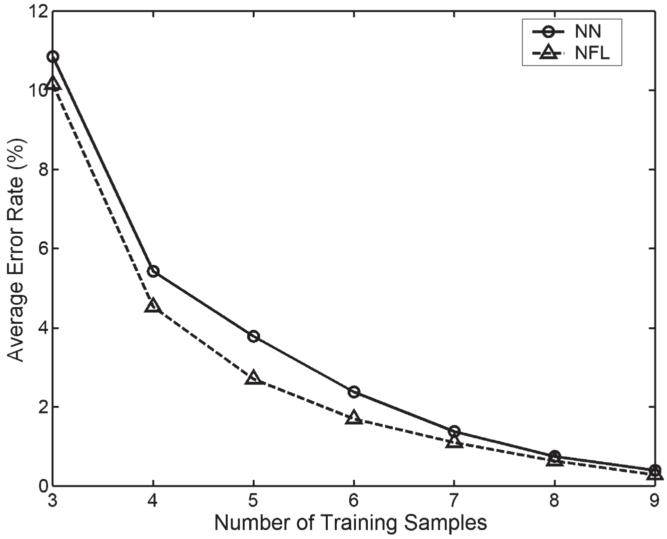


Fig. 6. Comparison of AER obtained using NN and NFL classifiers obtained using different numbers of training samples.

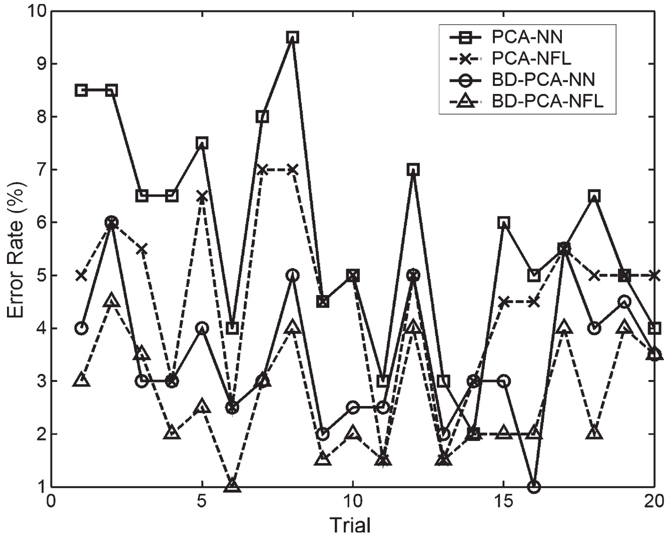


Fig. 7. Plots of error rates for each run.

TABLE II
COMPARISON OF AERs OBTAINED USING PCA, 2DPCA, BD-PCA, FISHERFACES, AND D-LDA ON THE ORL DATABASE

Classifier	PCA	2DPCA	BD-PCA	Fisherfaces	D-LDA
NN	5.78±1.9	4.28±1.1	3.48±1.0	7.93±2.1	5.40±1.7
NFL	4.63±1.5	3.30±0.9	2.70±0.9	7.65±2.0	5.10±1.6

In the last set of experiments, the performance of BD-PCA is compared with that of other appearance-based methods with $n_p = 5$. We implement two LDA-based methods, namely: 1) Fisherfaces [28] and 2) D-LDA [29]. Table II compares the AERs obtained using Fisherfaces, D-LDA, and BD-PCA. As

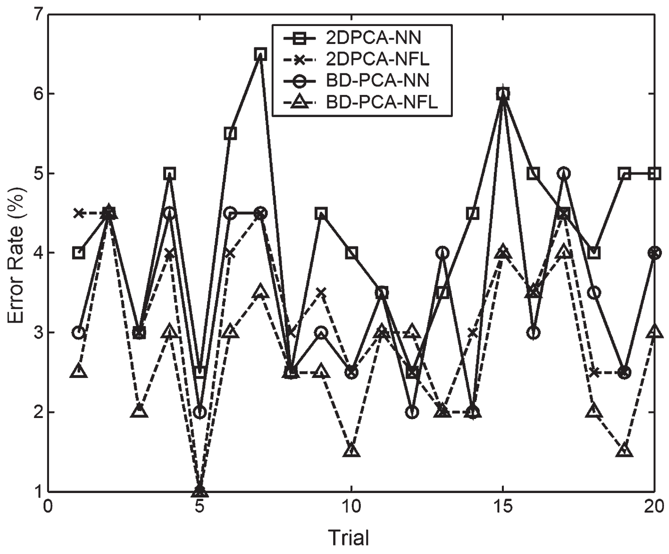


Fig. 8. Plots of error rates for each run.

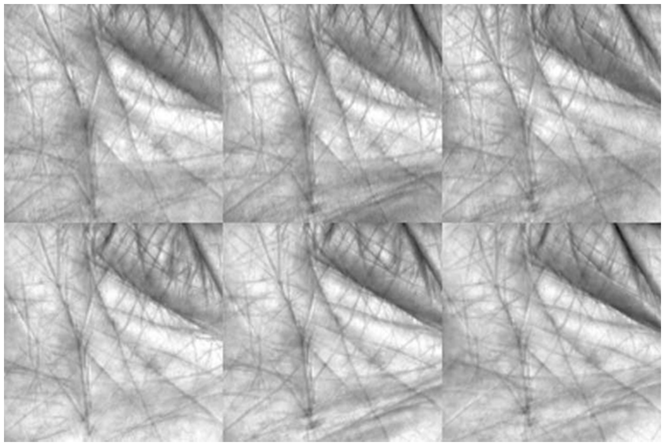
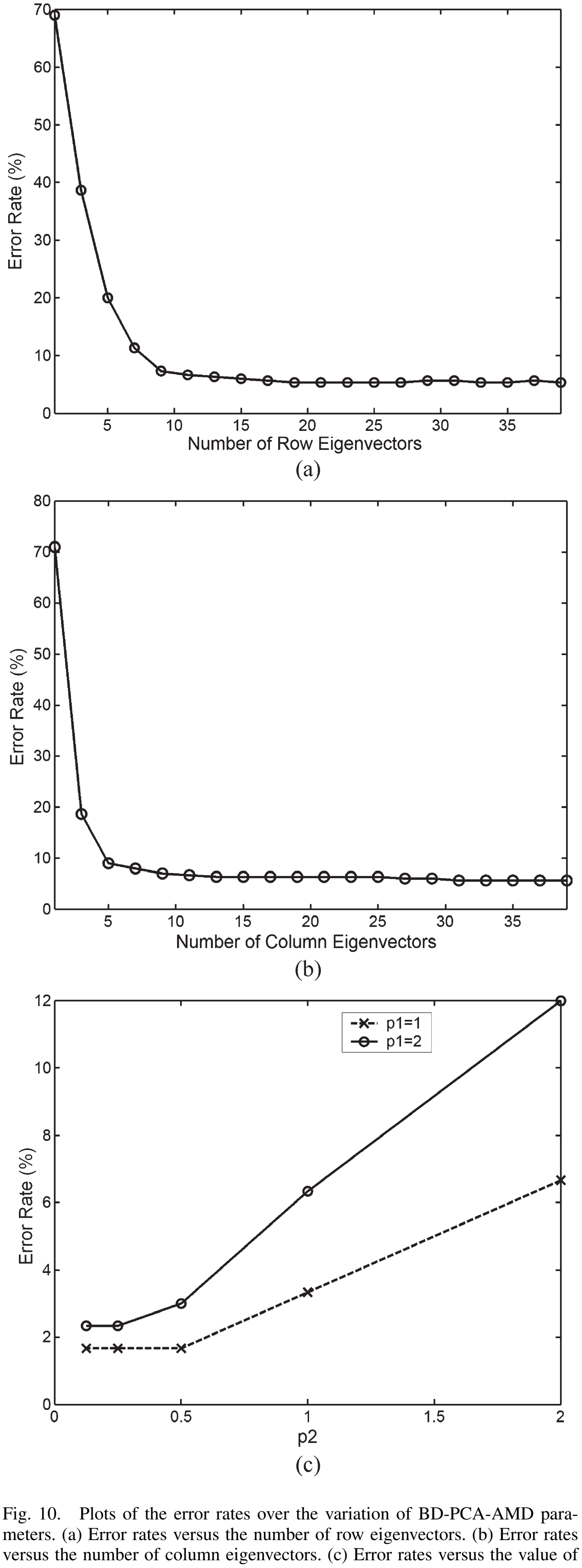


Fig. 9. Six palm print images of a palm in the PolyU palm print database.

references, we also compare the AER of BD-PCA with some reported results of other appearance-based methods on the ORL database. The error rate is 4.05 for Ryu's SHC method [19], 3.85 for Wang's CLSRD [20], 3.0 for Yang's complete PCA + LDA [30], 4.2 for Lu's DF-LDA [31], 4.9 for Liu's NKFDA [32], 4.2 for Song's LMLP [33], and 4.15 for Zheng's ELDA method [34]. The reported results are obtained on the average of different number of runs. Compared with these results, BD-PCA is very competitive.

B. Experimental Results on the PolyU Palm Print Database

Palm print sampling is low cost, nonintrusive, and the palm print contains stable structural features, making palm print recognition the object of considerable recent research interest [35]. In this section, we use the PolyU palm print database [36] to further evaluate BD-PCA-AMD. The PolyU palm print database contains 600 grayscale images of 100 different palms with six samples for each palm. Six samples from each of these palms were collected in two sessions, where the first three samples were captured in the first session and the other three in



(a)

(b)

(c)

Fig. 10. Plots of the error rates over the variation of BD-PCA-AMD parameters. (a) Error rates versus the number of row eigenvectors. (b) Error rates versus the number of column eigenvectors. (c) Error rates versus the value of p_2 with $p_1 = 1$ and $p_1 = 2$.

TABLE III
COMPARISON OF ERROR RATES OBTAINED USING DIFFERENT DISTANCE MEASURES ON THE PolyU PALM PRINT DATABASE

Classifiers	Frobenius	Yang	AMD
NN	11.00	4.67	1.67
NFL	11.00	4.33	1.33

TABLE IV
COMPARISON OF ERROR RATES OBTAINED USING DIFFERENT METHODS ON THE PolyU PALM PRINT DATABASE

Methods	PCA	Fisherfaces	D-LDA	2DPCA	BD-PCA
Error Rate (%)	11.33	6.67	6.00	4.40	1.33

the second session. The average interval between the first and the second session was two months.

In our experiments, a subimage of each original palm print was cropped to the size of 128 × 128 and preprocessed with histogram equalization. Fig. 9 shows six palm print images of one palm. For the PolyU palm print database, we choose the first three samples of each palm for training, thus using the 300 images captured in the first session as the training set and the 300 images captured in the second session as the testing set.

We adopt a stepwise strategy to determine the BD-PCA-AMD parameters, as shown in Fig. 10. The optimal parameters are then determined as $k_{\text{row}} = 19, k_{\text{col}} = 17, p_1 = 1, p_2 = 0.5$. Table III compares the error rates obtained using different distance measures. The AMD metric achieves the lowest error rates for both NN and NFL classifiers.

Table IV compares the error rates obtained using Eigenfaces, Fisherfaces, D-LDA, 2DPCA, and BD-PCA. The error rate of BD-PCA-NFL is 1.33, much lower than that obtained using the other four methods.

V. CONCLUSION

In this paper, we propose a BD-PCA with the AMD measure method (BD-PCA-AMD) for image recognition. The proposed method has some significant advantages. First, BD-PCA is directly performed on the image matrix, while classical PCA requires mapping an image matrix to a high-dimensional vector in advance. Second, BD-PCA can circumvent the overfitting problem associated with classical PCA. Third, the feature dimension of BD-PCA is much less than that of 2DPCA. We additionally present an AMD metric to reflect the fact that the BD-PCA feature is a matrix and apply the AMD metric to further improve the recognition performance of NN and NFL classifiers. BD-PCA can achieve an AER of 3.55 using the ORL with five training images of each person for the NN classifier, and 2.70 for the NFL classifier. On the PolyU palm print database, BD-PCA-NN achieved an error rate of 1.67 and BD-PCA-NFL achieved an error rate of 1.33.

ACKNOWLEDGMENT

The authors would like to thank J. Yang for many useful discussions on this paper, AT&T (Olivetti) Research Laboratories Cambridge for offering the ORL database, and the anonymous reviewers for their constructive suggestions.

REFERENCES

[1] L. Sirovich and M. Kirby, “Low-dimensional procedure for characterization of human faces,” *J. Opt. Soc. Amer.*, vol. 4, no. 3, pp. 519–524, Mar. 1987.

[2] M. Kirby and L. Sirovich, “Application of the Karhunen–Loeve procedure for the characterization of human faces,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 103–108, Jan. 1990.

[3] M. Turk and A. Pentland, “Eigenfaces for recognition,” *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[4] G. Lu, D. Zhang, and K. Wang, “Palmprint recognition using eigenpalms features,” *Pattern Recognit. Lett.*, vol. 24, no. 9/10, pp. 1463–1467, Jun. 2003.

[5] X. Wu, D. Zhang, and K. Wang, “Fisherpalms based palmprint recognition,” *Pattern Recognit. Lett.*, vol. 24, no. 15, pp. 2829–2838, Nov. 2003.

[6] R. Huber, H. Ramoser, K. Mayer, H. Penz, and M. Rubik, “Classification of coins using an eigenspace approach,” *Pattern Recognit. Lett.*, vol. 26, no. 1, pp. 61–75, Jan. 2005.

[7] J. Wu and Z. H. Zhou, “Face recognition with one training image per person,” *Pattern Recognit. Lett.*, vol. 23, no. 14, pp. 1711–1719, Dec. 2002.

[8] S. Chen, D. Zhang, and Z. H. Zhou, “Enhanced (PC)2A for face recognition with one training image per person,” *Pattern Recognit. Lett.*, vol. 25, no. 10, pp. 1173–1181, Jul. 2004.

[9] J. Yang and J. Y. Yang, “From image vector to matrix: A straightforward image projection technique—IMPCA versus PCA,” *Pattern Recognit.*, vol. 35, no. 9, pp. 1997–1999, Sep. 2002.

[10] J. Yang, D. Zhang, A. F. Frangi, and J. Y. Yang, “Two-dimensional PCA: A new approach to appearance-based face representation and recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.

[11] S. Chen and Y. Zhu, “Subpattern-based principal component analysis,” *Pattern Recognit.*, vol. 37, no. 5, pp. 1081–1083, May 2004.

[12] R. Gottumukkal and V. K. Asari, “An improved face recognition technique based on modular PCA approach,” *Pattern Recognit. Lett.*, vol. 25, no. 4, pp. 429–436, Mar. 2004.

[13] P. Navarrete and J. Ruiz-del-Solar, “Eigenspace-based recognition of faces: Comparisons and a new approach,” in *Proc. ICIAP*, 2001, pp. 42–47.

[14] H. Moon and P. J. Phillips, “Analysis of PCA-based face recognition algorithms,” in *Empirical Evaluation Techniques in Computer Vision*, K. Bowyer and P. J. Phillips, Eds. Los Alamitos, CA: IEEE Computer Society Press, 1998.

[15] W. S. Yambor, B. A. Draper, and J. R. Beveridge, “Analyzing PCA-based face recognition algorithm: Eigenvector selection and distance measures,” in *Empirical Evaluation Methods in Computer Vision*, H. Christensen and J. Phillips, Eds. Singapore: World Scientific, 2002.

[16] V. Perlibakas, “Distance measures for PCA-based face recognition,” *Pattern Recognit. Lett.*, vol. 25, no. 6, pp. 711-724, Apr. 2004.

[17] S. Li and J. Lu, “Face recognition using the nearest feature line method,” *IEEE Trans. Neural Netw.*, vol. 10, no. 2, pp. 439–443, Mar. 1999.

[18] J. T. Chien and C. C. Wu, “Discriminant waveletfaces and nearest feature classifiers for face recognition,” *IEEE Trans. Pattern Anal. Mach Intell.*, vol. 24, no. 12, pp. 1644–1649, Dec. 2002.

[19] Y. S. Ryu and S. Y. Oh, “Simple hybrid classifier for face recognition with adaptively generated virtual data,” *Pattern Recognit. Lett.*, vol. 23, no. 7, pp. 833–841, May 2002.

[20] H. Wang and L. Zhang, “Linear generalization probe samples for face recognition,” *Pattern Recognit. Lett.*, vol. 25, no. 8, pp. 829–840, Jun. 2004.

[21] W. Zheng, L. Zhao, and C. Zou, “Locally nearest neighbor classifiers for pattern classification,” *Pattern Recognit.*, vol. 37, no. 6, pp. 1307–1309, Jun. 2004.

[22] AT&T (Olivetti) Research Laboratories Cambridge, *The ORL Face Database*, Cambridge, U.K.. [Online]. Available: http://www.cl.cam.ac.uk/Research/DTG/attarchive/facedatabase.html

[23] J. Karhunan and J. Joutsensalo, “Generalization of principal component analysis, optimization problems, and neural networks,” *Neural Netw.*, vol. 8, no. 4, pp. 549–562, 1995.

[24] O. Toygar and A. Acan, “Multiple classifier implementation of a divide-and-conquer approach using appearance-based statistical methods for face recognition,” *Pattern Recognit. Lett.*, vol. 25, no. 12, pp. 1421–1430, Sep. 2004.

[25] S. Chen, J. Liu, and Z. H. Zhou, “Making FLDA applicable to face recognition with one sample per person,” *Pattern Recognit.*, vol. 37, no. 7, pp. 1553–1555, 2004.

[26] G. H. Golub and C. F. Van Loan, *Matrix Computation*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.

[27] D. Chaudhuri, C. A. Murthy, and B. B. Chaudhuri, "A modified metric to compute distance," *Pattern Recognit.*, vol. 25, no. 7, pp. 667–677, 1992.

[28] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces versus. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[29] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data with application to face recognition," *Pattern Recognit.*, vol. 34, no. 10, pp. 2067–2070, 2001.

[30] J. Yang and J. Y. Yang, "Why can LDA be performed in PCA transformed space?," *Pattern Recognit.*, vol. 36, no. 2, pp. 563–566, 2003.

[31] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using LDA-based algorithms," *IEEE Trans. Neural Netw.*, vol. 14, no. 1, pp. 195–200, Jan. 2003.

[32] W. Liu, Y. Wang, S. Z. Li, and T. Tan, "Null space-based kernel fisher discriminant analysis for face recognition," in *Proc. 6th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2004, pp. 369–375.

[33] F. Song, J. Y. Yang, and S. Liu, "Large margin linear projection and face recognition," *Pattern Recognit.*, vol. 37, no. 9, pp. 1953–1955, Sep. 2004.

[34] W. Zheng, L. Zhao, and C. Zou, "An efficient algorithm to solve the small sample size problem for LDA," *Pattern Recognit.*, vol. 37, no. 5, pp. 1077–1079, May 2004.

[35] D. Zhang, W. Kong, J. You, and M. Wong, "On-line palmprint identification," *IEEE Trans. Pattern Anal. Mach Intell.*, vol. 25, no. 9, pp. 1041–1050, Sep. 2003.

[36] D. Zhang, *PolyU Palmprint Palmprint Database*. Biometric Research Centre, Hong Kong Polytechnic University. [Online]. Available: http://www.comp.polyu.edu.hk/~biometrics/

Wangmeng Zuo received the B.S. and M.S. degrees in material science from the Harbin Institute of Technology, Harbin, China, in 1999 and 2001, respectively, and is currently working toward the Ph.D. degree at the School of Computer Science and Technology, Harbin Institute of Technology.

From July to December 2004, he was a Research Assistant at the Department of Computing, Hong Kong Polytechnic University, Kowloon. He is the author of five scientific papers in pattern recognition and computer vision. His current research interests include biometrics, pattern recognition, computer vision, and medical image processing.

David Zhang (SM'95) graduated in computer science from Peking University, Beijing, China, in 1974, and received the M.Sc. and Ph.D. degrees in computer science from the Harbin Institute of Technology (HIT), Harbin, China, in 1982 and 1985, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada.

From 1986 to 1988, he was a Postdoctoral Fellow at Tsinghua University, Beijing, China, and then an Associate Professor at Academia Sinica, Beijing. He is currently a Chair Professor at the Hong Kong Polytechnic University, Kowloon, Hong Kong, where he is the Founding Director of the Biometrics Technology Centre (UGC/CRC) supported by the Hong Kong SAR Government. He also serves as an Adjunct Professor at Tsinghua University, Shanghai Jiao Tong University, Beihang University, HIT, and the University of Waterloo. He is the Founder and Editor-in-Chief of the *International Journal of Image and Graphics*, Book Editor of the Kluwer International Series on Biometrics, and an Associate Editor of more than ten international journals including *Pattern Recognition* and is the author of more than ten books.

Dr. Zhang is Program Chair of the International Conference on Biometrics Authentication (ICBA) and an Associate Editor for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A and PART C. He is a current Croucher Senior Research Fellow and Distinguished Speaker of the IEEE Computer Society.

Kuanquan Wang (M'01) received the B.E. and M.E. degrees in computer science from the Harbin Institute of Technology (HIT), Harbin, China, in 1985 and 1988, respectively, and the Ph.D. degree in computer science from Chongqing University, Chongqing, China, in 2001.

From 1988 to 1998, he was with the Department of Computer Science, Southwest Normal University, Chongqing, as a Tutor, Lecturer, and Associate Professor. Since 1998, he has been with the Biocomputing Research Center, HIT, as a Professor, Supervisor of Ph.D. candidates, and an Associate Director. From 2000 to 2001, he was a Visiting Scholar at the Hong Kong Polytechnic University, Kowloon, Hong Kong, supported by the Hong Kong Croucher Funding, and from 2003 to 2004 was a Research Fellow. He has published over 70 papers. His research interests include biometrics, image processing, and pattern recognition. He is an Editor for the *International Journal of Image and Graphics* and a reviewer of *Pattern Recognition*.

Dr. Wang is a reviewer for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS.