Bi-Dierectional PCA with Assembled Matrix Distance Metric

Wangmeng Zuo, Kuanquan Wang
School of Computer Science and Technology
Harbin Institute of Technology
Harbin, China

David Zhang
Department of Computing
The Hong Kong Polytechnic University
Kowloon, Hong Kong

Abstract—**Principal Component Analysis (PCA) has been very successful in image recognition. Recent researches on PCA-based methods are mainly concentrated on two issues, feature extraction and classification. In this paper we propose Bi-Directional PCA (BDPCA) with assembled matrix distance (AMD) metric to simultaneously deal with these two issues. For feature extraction, we propose a BDPCA approach which can reduce the dimension of the original image matrix in both column and row directions. For classification, we present an AMD metric to calculate the distance between two feature matrices. The results of our experiments show that, BDPCA with AMD metric is very effective in image recognition.**

Keywords-PCA, 2DPCA, image recognition, face recognition, palmprint recognition

I. INTRODUCTION

Principal component analysis or PCA-based approaches have been very successful in image recognition. In 1987, Sirovich and Kirby used PCA to represent human faces [1]. Subsequently, PCA was applied to face recognition. PCA has now been widely investigated and has been successfully applied to other image recognition tasks.

Despite the great success of PCA, some issues remain that deserve to be further investigated. First, as we have showed, PCA is prone to be over-fitted to the training set because of the small sample size (SSS) problem. Although no researchers directly point out the over-fitting problem, some PCA-based methodes, such as $(PC)^2A$ [2] and 2DPCA [3], can be used to address it. But $(PC)^2A$ just alleviate the over-fitting by blurring the training image with intrinsic low-dimensional image, and 2DPCA has higher feature dimension than that of classical PCA. Thus further work is needed to solve the over-fitting problem and avoid the high feature dimension problem.

Second, there still has some work to be investigated in the design of classifiers based on the PCA feature. One general classifier is nearest neighbor (NN) classifier using the Euclidean distance measure. Other measures had also been studied to improve the recognition accuracy of NN [4]. Recently, nearest feature line (NFL) classifier is introduced to eliminate the performance deterioration caused by the reduction of prototypes [5]. Yet, even though previous studies have shown distance measures greatly affect the recognition accuracy, with reference to NFL, distance measures have been little studied.

In this paper, we tried to simultaneously address these two issues. First we propose a Bi-Directional PCA (BDPCA) to circumvent the over-fitting problem. BDPCA also avoids the high feature dimension of 2DPCA. Second, we present an assembled matrix distance (AMD) metric to calculate the distance between two feature matrices. To test the efficiency of BDPCA with AMD metric, experiments were carried out using two image databases. Experimental results show that, the proposed method is very effective in image recognition.

The organization of this paper is as follows: Section 2 investigated the over-fitting problem. Section 3 proposes BDPCA with AMD metric. Section 4 presents the results of experiments, Section 5 offers our conclusion.

II. THE OVER-FITTING PROBLEM OF CLASSICAL PCA

When applied to image recognition, classical PCA is apt to be over-fitted to the training set due to the SSS problem. To verify this perspective, we carried out a series of experiments using the ORL database.

The ORL database (http://www.uk.research.att.com/facedatabase.html) contains 400 facial images, with 10 images per individual. The images vary in age, light conditions, facial expressions, facial details (glasses/no glasses), scale and tilt. The size of these images is 112×92.

Now we use the normalized mean-square error (MSE) to evaluate the over-fitting problem. One statistical characteristic of PCA is that the MSE between random vector x and its subspace projection is minimal. Thus the difference of MSE on the training set and the testing set can be used to investigate the over-fitting problem.

Given the first L principal components, we can obtain the corresponding projector $\mathbf{W}_L$. Then a vector $\mathbf{x}$ can be transformed into the PCA subspace by

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

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

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

0-7803-9134-9/05/$20.00 ©2005 IEEE

where $\bar{\mathbf{x}}$ is the mean vector. The MSE on the training set MSE_L^{train} is defined as

$$MSE_L^{train} = \sum_{i=1}^{N_1} \left\| \mathbf{x}_i^{train} - \tilde{\mathbf{x}}_i^{train} \right\|^2 \Bigg/ \sum_{i=1}^{N_1} \left\| \mathbf{x}_i^{train} - \overline{\mathbf{x}^{train}} \right\|^2 , \tag{3}$$

where N_1 is the size of training set, $\mathbf{x}_i^{train}$ is the ith training samples, $\tilde{\mathbf{x}}_i^{train}$ is the reconstructed vector of $\mathbf{x}_i^{train}$, and $\overline{\mathbf{x}^{train}}$ is the mean vector of all training samples. Similarly, we can calculate the MSE on the testing set MSE_L^{test}.

Next we select the first 5 images per individual for training and calculate MSE_L^{train} and MSE_L^{test} for given $\mathbf{W}_L$, as shown in Fig. 1. When the value of L is small, the difference between MSE_L^{train} and MSE_L^{test} is small. But the difference is becoming great rapidly with the increase of L. Thus classical PCA is over-fitted to the training set.

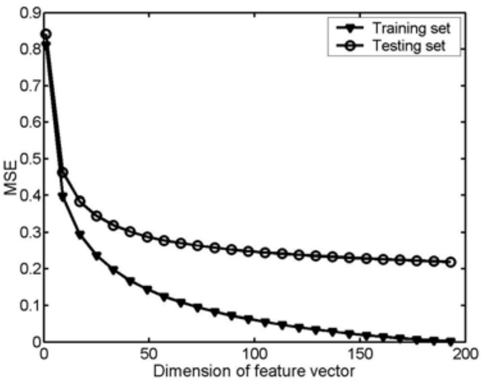

Fig. 1 The PCA's MSE on the training set and the testing set as the function of feature dimension.

III. BDPCA WITH ASSEMBLED MATRIX DISTANCE METRIC

In this Section, we first propose a novel PCA-based method, Bi-Directional PCA for image feature extraction, and then present an assembled matrix distance metric to calculate the distance of two feature matrices.

A. Bi-Directional PCA

Unlike classical PCA, Bi-Directional PCA directly extracts feature matrix $\mathbf{Y}$ from image matrix $\mathbf{X}$ by,

$$\mathbf{Y} = \mathbf{W}_{col}^T \mathbf{X} \mathbf{W}_{row} , \tag{4}$$

where $\mathbf{W}_{\mathrm{col}}$ and $\mathbf{W}_{\mathrm{row}}$ are the column and row projectors.

Next we present our method to calculate $\mathbf{W}_{\mathrm{col}}$ and $\mathbf{W}_{\mathrm{row}}$. Given a training set $\{\mathbf{X}_1, \cdots, \mathbf{X}_N\}$, N is the number of the samples, and the size of each image is $m \times n$. First we defines the row total scatter matrix $\mathbf{S}_t^{row}$ as

$$\mathbf{S}_t^{row} = \frac{1}{Nm} \sum_{i=1}^{N} (\mathbf{X}_i - \overline{\mathbf{X}})^T (\mathbf{X}_i - \overline{\mathbf{X}}) , \tag{5}$$

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 $\mathbf{S}_t^{row}$ to construct the row projector $\mathbf{W}_{row}$. Then we define the column total scatter matrix $\mathbf{S}_t^{col}$

$$\mathbf{S}_t^{col} = \frac{1}{Nn} \sum_{i=1}^{N} (\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})^T , \tag{6}$$

and choose the column eigenvectors corresponding to the first k_{col} largest eigenvalues of $\mathbf{S}_t^{col}$ to construct the column projector $\mathbf{W}_{col}$. Finally we use Eq. (4) to extract feature matrix $\mathbf{Y}$ from image $\mathbf{X}$.

Actually, BDPCA is a generalization of 2DPCA and 2DPCA can be regarded as a special BDPCA with $\mathbf{W}_{col} = \mathbf{I}_m$, where $\mathbf{I}_m$ is an $m \times m$ identity matrix.

B. Assemblem Matrix Distance Metric

BDPCA produces a feature matrix. So we propose an assembled matrix distance (AMD) metric to measure the distance between two feature matrices.

First we briefly reviewed some other matrix distance measures. Give two feature matrices $\mathbf{A} = (a_{ij})_{k_{col} \times k_{row}}$ and $\mathbf{B} = (b_{ij})_{k_{col} \times k_{row}}$, the Frobenius distance [6] is

$$d_F(\mathbf{A}, \mathbf{B}) = \left(\sum_{i=1}^{k_{col}} \sum_{j=1}^{k_{row}} (a_{ij} - b_{ij})^2 \right)^{1/2} , \tag{7}$$

Yang [3] proposed another distance (Yang distance)

$$d_Y(\mathbf{A}, \mathbf{B}) = \sum_{j=1}^{k_{row}} \left(\sum_{i=1}^{k_{col}} (a_{ij} - b_{ij})^2 \right)^{1/2} , \tag{8}$$

Here we define the AMD distance as

$$d_{AMD}(\mathbf{A}, \mathbf{B}) = \left(\sum_{j=1}^{k_{row}} \left(\sum_{i=1}^{k_{col}} (a_{ij} - b_{ij})^{p_1} \right)^{p_2/p_1} \right)^{1/p_2} , \tag{9}$$

Definition 1.[6] A *matrix norm* on $\mathbb{R}^{k_{col} \times k_{row}}$ is a function $f : \mathbb{R}^{k_{col} \times k_{row}} \to \mathbb{R}$ with the following properties:

$$f(\mathbf{A}) \geq 0, \quad \mathbf{A} \in \mathbb{R}^{k_{col} \times k_{row}} \, (f(\mathbf{A}) = 0 \Leftrightarrow \mathbf{A} = 0)$$

$$f(\mathbf{A} + \mathbf{B}) \leq f(\mathbf{A}) + f(\mathbf{B}), \quad \mathbf{A}, \mathbf{B} \in \mathbb{R}^{k_{col} \times k_{row}}$$

$$f(\alpha \mathbf{A}) \leq |\alpha| f(\mathbf{A}), \quad \alpha \in \mathbb{R}, \mathbf{A} \in \mathbb{R}^{k_{col} \times k_{row}}$$

Actually, both the Frobenius, Yang and AMD measures are matrix metrics, and next we will prove it.

Theorem 1.[6] $\|\mathbf{x}\|_p = (\sum_i |x_i|^p)^{1/p}$ is a vector norm.

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

Proof. It can be easily shown that

$$\|\mathbf{A}\|_{AMD} \geq 0 ,$$

$$\|\mathbf{A}\|_{AMD} = 0 \Leftrightarrow \mathbf{A} = 0 ,$$

$$\|\alpha\mathbf{A}\|_{AMD} = |\alpha|\|\mathbf{A}\|_{AMD} .$$

Now we prove $\|\mathbf{A}+\mathbf{B}\|_{AMD} \leq \|\mathbf{A}\|_{AMD} + \|\mathbf{B}\|_{AMD}$

$$\|\mathbf{A}+\mathbf{B}\|_{AMD} = \left(\sum_{j=1}^{k_{row}} \left(\sum_{i=1}^{k_{col}} (a_{ij}+b_{ij})^{p_1} \right)^{p_2/p_1} \right)^{1/p_2}$$
$$\leq (\sum_{j=1}^{k_{row}} (\|\mathbf{a}^{(j)}\|_{p_1} + \|\mathbf{b}^{(j)}\|_{p_1})^{p_2})^{1/p_2}$$

where $\mathbf{a}^{(j)}$ denotes the jth column vector of $\mathbf{A}$, and $\mathbf{b}^{(j)}$ denotes the jth column vector of $\mathbf{B}$. From **Theorem 1**,

$$g(\mathbf{a}) = \sum_{j=1}^{k_{row}} (\|\mathbf{a}^{(j)}\|_{p_1})^{p_2})^{1/p_2}, \ (\mathbf{a} = [\|\mathbf{a}^{(1)}\|_{p_1}, \cdots, \|\mathbf{a}^{(k_{row})}\|_{p_1}]^T)$$ is

a vector norm. Let $\mathbf{b} = [\|\mathbf{b}^{(1)}\|_{p_1}, \cdots, \|\mathbf{b}^{(k_{row})}\|_{p_1}]^T$,

$$(\sum_{j=1}^{k_{row}} (\|\mathbf{a}^{(j)}\|_{p_1} + \|\mathbf{b}^{(j)}\|_{p_1})^{p_2})^{1/p_2} = g(\mathbf{a}+\mathbf{b})$$
$$\leq g(\mathbf{a}) + g(\mathbf{b})$$
$$= (\sum_{j=1}^{k_{row}} (\|\mathbf{a}^{(j)}\|_{p_1})^{p_2})^{1/p_2} + (\sum_{j=1}^{k_{row}} (\|\mathbf{b}^{(j)}\|_{p_1})^{p_2})^{1/p_2}$$
$$= \left(\sum_{j=1}^{k_{row}} \left(\sum_{i=1}^{k_{col}} \mathbf{a}_{ij}^{\ p_1} \right)^{p_2/p_1} \right)^{1/p_2} + \left(\sum_{j=1}^{k_{row}} \left(\sum_{i=1}^{k_{col}} \mathbf{b}_{ij}^{\ p_1} \right)^{p_2/p_1} \right)^{1/p_2}$$
$$= \|\mathbf{A}\|_{AMD} + \|\mathbf{B}\|_{AMD}$$

So $\|\mathbf{A}\|_{AMD}$ is a matrix norm.

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

Proof. Since $\|A\|_{AMD}$ is a matrix norm, it is easy to prove that $d_{AMD}(\mathbf{A}, \mathbf{B}) = \|\mathbf{A}-\mathbf{B}\|_{AMD}$ is a distance metric.

Corollary 1. The Frobenius distance measure is a special case of AMD metric with $p_1 = p_2 = 2$.

Corollary 2. The Yang distance measure is a special case of AMD metric with $p_1 = 2$ and $p_2 = 1$.

IV. EXPERIMENTAL RESULTS AND DISCCUSSIONS

To evaluate the efficiency of BDPCA with AMD metric (BDPCA-AMD), we used two image databases, the ORL database and the PolyU palmprint database.

A. Experiments on the ORL Database

Using the ORL database, we compare the reconstruction capability of PCA, 2DPCA, and BDPCA. Fig. 2(a) is a training image. Satisfied reconstructed images are obtained using all these three methods, as shown in Fig. 2(b)-(d). The best reconstructed image is that reconstructed by PCA. Fig. 2(e) is a testing image and its reconstructed images are shown in Fig. 2(f)-(h). The quality of the reconstructed image by PCA deteriorates greatly, while 2DPCA and BDPCA still obtain satisfied reconstruction quality. Besides, the feature dimension of BDPCA is 8×30=240, lower than that of 2DPCA (896).

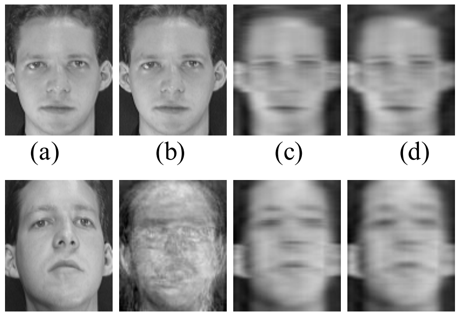

Fig. 2 Comparison of the reconstruction capability of PCA, 2DPCA and BDPCA, (a)(e) original images, reconstructed images by (b)(f) PCA, (c)(g) 2DPCA, and (d)(h) BDPCA.

Then we use MSE to evaluate BDPCA's capability in solving the over-fitting problem. By selecting the first 5 images per individual for training, we calculate MSE^{train} and MSE^{test} for given $\mathbf{W}_{col}$ and $\mathbf{W}_{row}$, as shown in Fig. 3. The difference of MSE^{train} and MSE^{test} is very small. Thus BDPCA can solve the over-fitting problem.

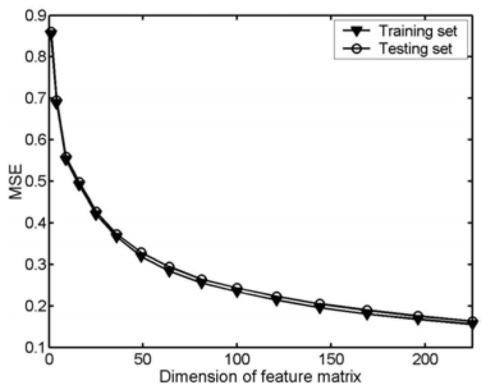

Fig. 3 The BDPCA's MSE on the training set and the testing set as the function of feature dimension

In all the following experiments, we randomly choose 5 images per individual for training, and use average error rate (the mean of error rates over 20 runs). For BDPCA, we set k_{row}=4, k_{col}=18, and p_1=2.

Fig. 4 shows the effect of the AMD parameter p_2. From Fig. 4, we set p_2=0.25. Table 1 compares the AER obtained using the Frobenius, the Yang and the AMD metrics. The AMD metric achieved the lowest AER for both NN (BDPCA-NN) and NFL (BDPCA-NFL).

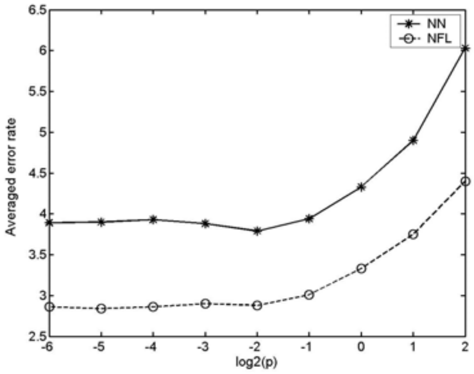

Fig. 4 AER of BDPCA-AMD with different p_2 values

Classifiers	Frobenius	Yang	AMD
NN	4.90	4.33	3.78
NFL	3.75	3.33	2.88

Table 1 Comparisons of the AER obtained using different distance metrics and classifiers

Table 2 lists the AER of PCA, 2DPCA and BDPCA. BDPCA outperforms PCA and 2DPCA for both NN and NFL classifiers. The feature dimension of BDPCA is 18 × 4 = 72, much lower than that of 2DPCA (112 × 4 = 448).

Methods	PCA	2DPCA	BDPCA
NN	5.78	4.28	3.45
NFL	4.63	3.30	2.70

Table 2 Comparisons of the AER obtained using PCA, 2DPCA, and BDPCA

B. Experiments on the PolyU Palmprint Database

The PolyU palmprint database (http://www.comp.polyu.edu.hk/~biometrics) [7] is also used to evaluate BDPCA-AMD. The database contains 600 images of 100 palms with six samples for each palm. In our experiments, sub-image of each palmprint was cropped to the size of 128×128 and pre-processed by histogram equalization. We choose the first 3 samples per palm for training.

With k_{row}=13, k_{col}=15 and p_1=1, we studied the effect of parameter p_2, and set p_2=0.25 experimentally. Table 3 compares the error rates obtained using different distance metrics, and the AMD metric achieved the minimum error rate for both NN and NFL classifiers. Table 4 lists the error rates obtained using PCA, 2DPCA, and BDPCA. The lowest error rate of BDPCA is 1.33, much lower than that obtained using PCA or 2DPCA.

Classifiers	Frobenius	Yang	AMD
NN	11.00	4.67	1.33
NFL	11.00	4.33	1.00

Table 3 Comparisons of the error rates obtained using different distance metrics and classifiers

Methods	PCA	2DPCA	BDPCA
Error Rate(%)	11.33	4.40	1.33

Table 4 Comparisons of the error rate obtained using PCA, 2DPCA, and BDPCA

CONCLUSIONS

In this paper, we propose a novel image recognition method, BDPCA with AMD metric, which has some significant advantages. First, BDPCA is directly performed on image matrix, while classical PCA is required to map an image matrix to a 1D vector in advance. Second, BDPCA can circumvent classical PCA's over-fitting problem. Third, the feature dimension of BDPCA is much less than 2DPCA. Fourth, an AMD metric is proposed to meet the fact that BDPCA feature is a matrix and the AMD metric can be used improve the recognition accuracy for either NN or NFL classifiers.

ACKNOWLEDGMENT

The work is partially supported by the UGC/CRC fund from the HKSAR Government, the central fund from the Hong Kong Polytechnic University and the NSFC fund under the contract No. 60332010 and No. 90209020. The authors thank Jian Yang for many useful discussions on this paper.

REFERENCES

[1] L. Sirovich, and M. Kirby, "Low-dimensional procedure for characterization of human faces," *J. Optical Soc. of Am.*, vol. 4, pp. 519-524, 1987.

[2] J. Wu, Z.-H. Zhou, "Face recognition with one training image per person, *Pattern Recognit. Lett.*, vol. 23, pp. 1711-1719, 2002.

[3] J. Yang, D. Zhang, A. F.Frangi, J. Yang, "Two-Dimensional PCA: a new approach to appearance-based face representation and recognition", *IEEE Trans. on PAMI*, vol. 26, no. 1, pp. 131-137, 2004.

[4] V. Perlibakas, "Distance measures for PCA-based face recognition," *Pattern Recognit. Lett.*, vol. 25, pp. 711-724, 2004.

[5] S. Li, and J. Lu, "Face recognition using the nearest feature line method," *IEEE Trans. Neural Netw.*, vol. 10, no. 2, pp. 439-443, 1999.

[6] G.H. Golub, and C. F. Van Loan, *Matrix Computation* (3rd edition), Baltimore: The Johns Hopkins University Press, 1996.

[7] D. Zhang, W.K. Kong, J. You, And M. Wong, "Online palmprint identification," *IEEE Trans. on PAMI*, vol. 25, no. 9, pp. 1041-1050, 2003.