Is Two-dimensional PCA a New Technique?[1)]

WANG Li-Wei WANG Xiao CHANG Ming FENG Ju-Fu

(*Center for Information Sciences, Peking University, Beijing* 100871)

(E-mail: {wanglw, wangxiao, changming, fjf}@cis.pku.edu.cn)

Abstract The principal component analysis (PCA), or the eigenfaces method, is a de facto standard in human face recognition. Numerous algorithms tried to generalize PCA in different aspects. More recently, a technique called two-dimensional PCA (2DPCA) was proposed to cut the computational cost of the standard PCA. Unlike PCA that treats images as vectors, 2DPCA views an image as a matrix. With a properly defined criterion, 2DPCA results in an eigenvalue problem which has a much lower dimensionality than that of PCA. In this paper, we show that 2DPCA is equivalent to a special case of an existing feature extraction method, *i.e.*, the block-based PCA. Using the FERET database, extensive experimental results demonstrate that block-based PCA outperforms PCA on datasets that consist of relatively simple images for recognition, while PCA is more robust than 2DPCA in harder situations.

Key words Face recognition, PCA, two-dimensional PCA, block-based PCA

1 Introduction

Within the past few decades, numerous algorithms have been proposed for human face recognition. Among these, PCA (Eigenfaces method)[1~3] has become a de facto standard and a common performance benchmark in the field. It uses the Karhunen-Loeve transform (KLT) to produce a most expressive subspace for face representation and recognition. The current state-of-the-art of face recognition is a class of subspace techniques. Examples of the subspace methods include Fisherfaces method (PCA+LDA)[4,5], Bayesian similarity measure[6], independent component analysis (ICA)[7,8], and their nonlinear generalizations using the kernel trick[9,10].

One disadvantage of PCA is the high computational complexity. Suppose that the images are of size $M \times N$, and the number of training images is greater than $M \times N$ (this is usually true for large database). Performing PCA needs $O((M \times N)^3)$ computations. This is quite expensive even for medium size images.

More recently, a technique called two-dimensional PCA (2DPCA)[11] (also called Image PCA in a previous paper by the same author[12]) was proposed to cut the computational cost of the standard PCA. Unlike PCA that treats images as vectors, 2DPCA views an image as a matrix. With a properly defined criterion, 2DPCA results in an eigenvalue problem which has a much lower dimensionality than that of PCA.

Another line of research in feature extraction and face recognition is the block-based methods. In these methods, an image is partitioned into several blocks. Often, all the blocks have the same size. Features are extracted from the block images. Block-based method first appeared in hidden Markov model (HMM) based face recognition systems[13]. In [14], the authors used block based principal components (eigenvectors) to train a hidden Markov model. Other works related to this idea include fragment-based feature extraction[15,16], which makes use of image fragments to represent objects.

The objective of this paper is twofold. First, we show that 2DPCA is equivalent to a special case of the block-based PCA. When each block is a line of the image, and taking all lines of the facial images as training samples, applying PCA to these "line" samples yields 2DPCA. Secondly, we investigate when block-based PCA (including 2DPCA) outperforms the standard PCA. Extensive experiments demonstrate that 2DPCA is superior to PCA on datasets that consist of relatively simple images for recognition. But PCA is more robust than 2DPCA in harder situations where large variations between the probes and the gallery images exist.

The remainder of this paper is organized as follows. Section 2 briefly reviews 2DPCA. Section 3 addresses the block-based PCA. We show an equivalence of 2DPCA to a special case of block-based PCA

1) Supported by National Key Basic Research Project of R. P. China (2004CB318000)
Received April 5, 2004; in revised form July 6, 2005

in Section 4. Section 5 provides experimental results and analysis on the comparison of block-based PCA to PCA.

2 Two-dimensional PCA (2DPCA)

Consider an M by N image as $M \times N$ random matrix denoted by A. Let $\boldsymbol{x}$ be an N-dimensional unit column vector. Projecting A onto $\boldsymbol{x}$ yields an M-dimensional vector $\boldsymbol{y}$, *i.e.*,

$$\boldsymbol{y} = A\boldsymbol{x} \tag{1}$$

The purpose of 2DPCA is to select a good projection vector x. To evaluate the goodness of a projection vector, the use of the total scatter of the projected samples[11] is suggested, which can be characterized by the trace of the covariance matrix of the projected feature vectors. Thus, the criterion is to maximize the following:

$$J(x) = tr(S_x) \tag{2}$$

where S_x is the covariance matrix of the projected feature vectors, written by

$$S_x = E(\boldsymbol{y} - E\boldsymbol{y})(\boldsymbol{y} - E\boldsymbol{y})^{\mathrm{T}} = E[(A - EA)\boldsymbol{x}][(A - EA)\boldsymbol{x}]^{\mathrm{T}} \tag{3}$$

Hence,

$$J(x) = tr(S_x) = \boldsymbol{x}^{\mathrm{T}} E[(A - EA)^{\mathrm{T}}(A - EA)]\boldsymbol{x} \tag{4}$$

Given a set of training images $A(1), A(2), \cdots, A(n)$, criterion (4) becomes

$$J(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}} \left[\sum_{i=1}^{n} (A(i) - \bar{A})^{\mathrm{T}} (A(i) - \bar{A}) \right] \boldsymbol{x} \tag{5}$$

where $\bar{A}$ is the average of all training images. Let $G = \sum_{i=1}^{n} (A(i) - \bar{A})^{\mathrm{T}} (A(i) - \bar{A})$, and the optimal axis be the unit vector maximizing $J(x)$, *i.e.*, the eigenvector of G corresponding to the largest eigenvalue. Of course, one can compute m best projection axes, which are the m leading eigenvectors of G.

Without loss of generality, we will always assume that all the images have been shifted so that they have a zero mean, *i.e.*, $\bar{A} = \frac{1}{n} \sum_{i=1}^{n} A(i) = (0)_{M \times N}$. Thus, (5) becomes

$$J(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}} \left[\sum_{i=1}^{n} A(i)^{\mathrm{T}} A(i) \right] \boldsymbol{x} \tag{6}$$

3 Block-based PCA

The idea of extracting features from image blocks first appeared in hidden Markov model (HMM) based face recognition systems[13]. According to this model, a face image is divided into a number of blocks. Features are extracted from the blocks rather than from the whole image.

There are two types of features, the so-called block specific features[14] and the block universal features. Block universal features are extracted from all constitute blocks. On the other hand, for block specific features, one takes into account each block in the face image with different spatial properties depending on its position. Thus different features are extracted from different blocks. Features such as PCA, second-order PCA, and coefficients of two-dimensional discrete cosine transform (DCT) have been used as the block based features. In this paper, we focus on block-based PCA. We assume that an image is partitioned into K blocks having the same size.

Below, by block universal PCA (BUPCA) we mean to take all blocks of all images as training samples, and compute a unique set of leading eigenvectors of the sample covariance matrix. And by block specific PCA (BSPCA) we mean to compute K sets of eigenvectors, each for a block in a specific position of the face image.

More concretely, let $A(1), A(2), \cdots, A(n)$ be the training images, and $\boldsymbol{a}(i)_1, \boldsymbol{a}(i)_2, \cdots, \boldsymbol{a}(i)_K$ be the K blocks of $A(i)$, $i = 1, 2, \cdots, n$. BUPCA is to find the subspace spanned by the leading eigenvectors of the following sample covariance matrix:

$$S = \sum_{i=1}^{n} \sum_{k=1}^{K} \boldsymbol{a}(i)_k \boldsymbol{a}(i)_k^{\mathrm{T}}$$

And BSPCA solves K eigenvalue problems of the sample covariance matrices given below:

$$S_k = \sum_{i=1}^{n} \boldsymbol{a}(i)_k \boldsymbol{a}(i)_k^{\mathrm{T}}, \quad k = 1, 2, \cdots, K$$

where S_k is the sample covariance matrix of the kth block.

4 Equivalence of 2DPCA to line-based BUPCA

In this section, we show that 2DPCA is equivalent to a special case of the block based PCA. Specifically, if each block is a line of the image, then conducting PCA to all these "line" samples, *i.e.*, BUPCA with each block is a line, yields 2DPCA. We only outline the proof here, for details please refer to [15].

Suppose again that $A(1), A(2), \cdots, A(n)$ are training images of size $M \times N$. Let $\boldsymbol{a}(i)_1, \boldsymbol{a}(i)_2, \cdots, \boldsymbol{a}(i)_M$ be the M lines of $A(i)$, $i = 1, 2, \cdots, n$. More concretely,

$$A(i) = \begin{bmatrix} a(i)_{11} & a(i)_{12} & \cdots & a(i)_{1N} \\ a(i)_{21} & a(i)_{22} & \cdots & a(i)_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ a(i)_{M1} & a(i)_{M2} & \cdots & a(i)_{MN} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}(i)_1^{\mathrm{T}} \\ \boldsymbol{a}(i)_2^{\mathrm{T}} \\ \cdots \\ \boldsymbol{a}(i)_M^{\mathrm{T}} \end{bmatrix}, \quad i = 1, 2, \cdots, n \tag{7}$$

and

$$\boldsymbol{a}(i)_m^{\mathrm{T}} = (a(i)_{m1}, a(i)_{m2}, \cdots, a(i)_{mN}), \quad i = 1, 2, \cdots, n, \ m = 1, 2, \cdots, M \tag{8}$$

Consider the criterion of 2DPCA (see (6)). Let $G = \sum_{i=1}^{n} A(i)^{\mathrm{T}} A(i)$. 2DPCA is to compute the leading eigenvectors of G. Taking into consideration of (7), (8), and rewriting G in terms of $a(i)_m$, $i = 1, 2, \cdots, n$, $m = 1, 2, \cdots, M$, we have

$$G = \sum_{i=1}^{n} A(i)^{\mathrm{T}} A(i) = \sum_{i=1}^{n} \sum_{m=1}^{M} \boldsymbol{a}(i)_m \boldsymbol{a}(i)_m^{\mathrm{T}} \tag{9}$$

Clearly, G is the sample covariance matrix of all the lines $a(i)_m$, $i = 1, 2, \cdots, n$, $m = 1, 2, \cdots, M$. That is, if we consider each line of each training image as a sample vector, G is the sample covariance matrix. 2DPCA is therefore equivalent to BUPCA with each block being a line.

5 Experiments and analysis

In this section we investigate when block-based PCA outperforms the standard PCA in human face recognition. It was reported that 2DPCA (a special case of block-based PCA) was superior to PCA. However, this result was obtained from the ORL, AR and Yale databases[16~18]. The three databases contain 40, 126 and 15 individuals respectively, which are relatively small.

In our experiments, we compare block-based PCA to PCA using the well-known FERET database. FERET dataset contains a gallery of size 1196 individuals, and four probe categories, named FB, Duplicate 1, Duplicate 2 and fc respectively. (A probe set consists of images of unknown faces, and the gallery contains images of known individuals. For details about the meaning of the terminologies and the FERET database, the reader is referred to [19].)

For human face recognition, it is also important to use a training set independent of the gallery and the probes[19]. We choose 1068 frontal images from 554 individuals to train models.

We conducted two sets of experiments. The primary goal of the first set was to compare block-based PCA with PCA. In the second set of experiments, we further compared BSPCA to BUPCA.

5.1 Comparison of block-based PCA to PCA

We compare BSPCA to PCA in this section. We test three partition methods:

Type I: each block is a line.

Type II: each block is a multi-line window.

Type III: each block is a rectangle window.

Experimental results for PCA and the three BSPCA′s on the four probe categories (FB, Duplicate 1, Duplicate 2, fc) are plotted in Fig. 1 to Fig. 4, where we show the cumulative match score for both top 1 candidate and top 50 candidates. In these figures, BSPCA I, BSPCA II and BSPCA III denote the three partition methods for BSPCA respectively.

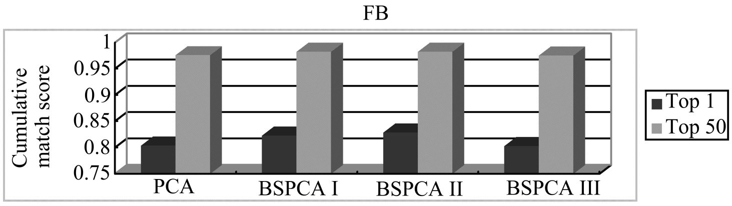

Fig. 1 Comparison of PCA and BSPCA on FB probe set

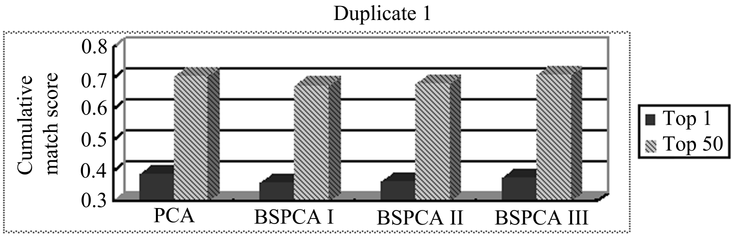

Fig. 2 Comparison of PCA and BSPCA on Duplicate 1 probe set

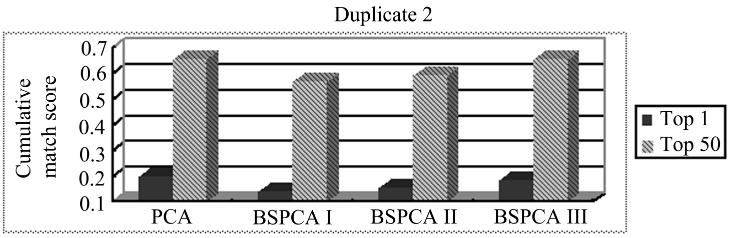

Fig. 3 Comparison of PCA and BSPCA on Duplicate 2 probe set

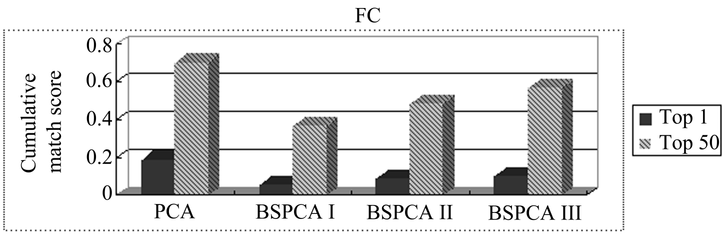

Fig. 4 Comparison of PCA and BSPCA on fc probe set

These four experiments reveal the following points:

1) BSPCA outperforms PCA when the probe set contains relatively simple images for recognition. Note that FB probes have just expression variation, and almost all algorithms behave well on this probe set[21].

2) PCA is more robust than BSPCA in harder situations. Images in Duplicate I, Duplicate II and fc are much more difficult to recognize due to the time and lighting variations. Please see[21] for comparison of the difficulties of the four probes.

5.2 Comparison of BSPCA to BUPCA

Some authors argued that block specific features were superior to block universal features[14]. To our knowledge, however, there is no paper that rigorously compares the performance of these two kinds of features for face recognition. In this section, we compare them with the afore-mentioned partition methods.

The cumulative match score up to the top 50 candidates are listed in Tables 1, 2, and 3, respectively.

Remember that BUPCA using type I blocks (each block is a line) is equivalent to 2DPCA.

Table 1 Cumulative match score (%) of type I block (single line)

	FB	Dup. I	Dup. II	Fc
BSPCA	98	68	56	37
BUPCA	98	67	55	29

Table 2 Cumulative match score (%) of type II block (multi- line window)

	FB	Dup. I	Dup. II	Fc
BSPCA	98	68	59	49
BUPCA	98	68	56	44

Table 3 Cumulative match score (%) type III block (rectangle window)

	FB	Dup. I	Dup. II	Fc
BSPCA	98	71	65	57
BUPCA	97	72	65	54

The experimental results suggest a number of conclusions:

1) BSPCA is slightly, but almost consistently better than BUPCA.

2) No partition method is superior to all the others on all probe sets.

6 Conclusion

In this paper, we show that 2DPCA is equivalent to BUPCA if each block is a line of the image. Experiments on FERET database demonstrated a performance improvement of block-based PCA over PCA on relatively simple probes. However, PCA is more robust than block-based PCA on harder situations.

References

1 Kirby M, Sirovich L. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, **12**(1): 103~108

2 Jolliffe I T. Principal Component Analysis. New York, Springer-Verlag, 1986

3 Turk M A, Pentland A P. Face recognition using eigenfaces. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Maui Hawaii: IEEE Press, 1991. 586~591.

4 Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces *vs*. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 711~720

5 Etemad K, Chellappa R. Discriminant analysis for recognition of human face images. *Journal of Optical Society of Americia A*, 1997, **14**(8): 1724~1733

6 Moghaddam B. Bayesian face recognition. *Pattern Recognition*, 2000, **13**(11): 1771~1782

7 Bell A J, Sejnowski T J. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 1995, **7**(6): 1129~1159

8 Comon P. Independent Component analysis—a new concept? *Signal Processing*, 1994, **36**: 287~314

9 Scholkopf B, Smola A, Muller K R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 1998, **10**(5): 1299~1319

10 Mika S, Ratsch G, Weston J, Scholkopf B, Muller K. Fisher discriminant analysis with kernels. Neural Networks for Signal Processing IX, Hu Y H, Larsen J, Wilson E, Douglas D, eds., 1999. 41~48

11 Yang J, Zhang D. Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(1): 131~137

12 Yang J, Yang J Y. From image vector to matrix: a straightforward image projection technique–IMPCA *vs*. PCA. *Pattern Recognition*, 2002, **35**(9): 1997~1999

13 Samaria F, Young S. HMM-based architecture for face identification. *Image and Vision Computing*, 1994, **12**(8): 537~543

14 Kim M, Kim D, Lee S. Face recognition using the embedded HMM with second-order block-specific observations. *Pattern Recognition*, 2003, **36**(11): 2723~2735

15 Wang L, Wang X, Zhang X, Feng J. The equivalence of two-dimensional PCA to line-based PCA. *Pattern Recognition Letters*. 2005, **26**(1): 57~60

16 AT&T Laboratories Cambridge. http://www.uk.research.att.com/facedatabase.html, 2002

17 Martinez A M, Benavente R. The AR face database. CVC Technical Report #24, 1998

18 http://cvc.yale.edu/projects/yalefaces/yalefaces.html.

19 Phillips P.J, Moon H, Rizvi S.A, Rauss P.J. The FERET evaluation methodology for face-recognition algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(10): 1090~1104

WANG Li-Wei Ph. D. His research interests include pattern recognition, computer vision, and machine learning.

WANG Xiao Master. Her research interests include computer vision and pattern recognition.

CHANG Ming Graduate student. Her research interests include computer vision and pattern recognition.

FENG Ju-Fu Professor. His research interests include pattern recognition, computer vision, and machine learning.

(From Page 674)

Major topics of the congress include, but are not limited to the following areas.

A. Theory and methods

A1. Control theory

1. System and control theory 2. Nonlinear systems 3. Large-scale systems
4. Hybrid systems and DEDS 5. Control system with distributed parameters
6. Modeling, identification, estimation and optimization
7. Advanced control (adaptive control, variable structure control, robust control, H_∞ control)

A2. Intelligent control

1. Artificial intelligence and expert systems 2. Neural networks
3. Fuzzy algorithms, genetic algorithms, evolutionary algorithms 4. Fuzzy control, learning control
5. Intelligent information processing 6. Networked control

B. Industrial systems and control

B1. Modeling, sensing and fault diagnosis

1. Process modeling techniques 2. Soft measurement techniques
3. Sensors, measurement and intelligent instruments 4. Fault diagnosis 5. Data mining
6. Simulation and CAD of control systems

B2. Control techniques and integrated automation systems

1. Advanced control techniques 2. Optimized control techniques
3. Integrated automation systems of pro cess industry 4. Computer integrated manufacturing systems
5. Decision supporting systems 6. Enterprise resource planning and manufacturing execution systems
7. Production planning and intelligent scheduling

C. Applications

C1. Automation and intelligence of process industry

1. Power systems 2. Petrochemical processes 3. Metallurgical processes
4. Paper making processes 5. Others

C2. Automation and intelligence of manufacturing industry

1. Intelligent manufacturing systems 2. Advanced digital control systems 3. Motion control
4. Propulsion system control 5. Micro and nano scale sensors, actuators and robots 6. Others

C3. Other application systems

1. Intelligent transportation systems 2. Intelligent building systems 3. Intelligent Robotics
4. Environmental and biomedical systems 5. Human-machine systems
6. Pattern recognition and image processing

Paper Submission

Prospective authors are invited to submit pdf files of their full papers in Chinese or English (within 5 pages) at the congress website http : //wcica06.dlut.edu.cn. The instructions on the manuscript format and submission details can be seen from IEEE Xplore guidelines, and also are available at the congress website. The cover page should contain paper title, author names and affiliations, the address, telephone number, and the email address of the corresponding author, a paper abstract, 3~5 keywords. Proposals for organized special sessions are invited and encouraged. The proposal should contain the title of the proposed session, a list of 6 contributed paper titles, full papers in pdf files, and the contact information of the author of each paper.

Important Dates

Paper submission deadline	Nov. 1, 2005
Notification of paper acceptance	Feb. 1, 2006
Final version of paper submission deadline	March 1, 2006