Progressive Principal Component Analysis

Jun Liu[1], Songcan Chen[1*] and Zhi-Hua Zhou[2]

[1] Department of Computer Science and Engineering,
Nanjing University of Aeronautics and Astronautics,
Nanjing 210016, P.R. China
maurice_nt@sohu.com, s.chen@nuaa.edu.cn
[2] National Laboratory for Novel Software Technology,
Nanjing University, Nanjing 210093, China
zhouzh@nju.edu.cn

Abstract. Principal Component Analysis (PCA) is a feature extraction approach directly based on a whole vector pattern and acquires a set of projections that can realize the best reconstruction for an original data in the mean squared error sense. In this paper, the progressive PCA (PrPCA) is proposed, which could progressively extract features from a set of given data with large dimensionality and the extracted features are subsequently applied to pattern recognition. Experiments on the FERET database show its face recognition performance is better than those based on both $E(PC)^2A$ and FLDA.

1 Introduction

The traditional PCA is an effective approach of extracting features and dimensionality reduction and has partially successfully been applied to pattern recognition such as face recognition with one training (face) pattern per person [1]. It operates directly on a whole vector pattern to extract so-needed global features for subsequent recognition by using a set of found projectors from a given training pattern set. The extracted features can maximally preserve or reconstruct original pattern information in the mean squared error sense. However, when the dimensionality of given pattern is very large, for example, 1000 or larger, extracting features directly from these large dimensional patterns *once* exists some processing difficulty such as computational complexity for large scale covariance matrix constructed by the training set. With the divide-and-conquer technique, the subpattern-based PCA (SpPCA) [2] eliminates such a difficulty via first dividing the whole pattern into K equally-sized subpatterns {Sp1, Sp2, …, SpK} as shown Fig. 1 and then separately performing PCA on each training subpattern set of sharing the same (feature) components so as to obtain a classification performance gain and robustness to partially missing subpatterns. However, SpPCA only performs independently PCAs in different divisions and thus disregards useful contextual information among different subpattern sets, which pro-

* Corresponding author: Tel: +86-25-84892805, Fax: +86-25-84893777
Email: s.chen@nuaa.edu.cn

duces unfavorable influence on classification performance. In this paper, a novel PCA (PrPCA) is developed to avoid the problem but still to preserve its robustness to partially missing information. It operates progressively or subpattern-by-subpattern on given pattern to extract features from a set of large dimensional patterns as shown Fig.1.

2 Proposed PrPCA

Except for extracting features in a *progressive* mode, the idea behind the proposed PrPCA is almost identical to the SpPCA so that PrPCA can keep both its simplicity and robustness but still possesses better recognition accuracy as shown in Table 1. The following is a formulation for the PrPCA.

PrPCA consists of two steps. The first step is a partition step, i.e., an original whole pattern is partitioned into a set of equally-sized (actually, not limited to this) subpatterns in non-overlapping ways and then all those subpatterns sharing the same original feature components are respectively collected from the training set to compose corresponding training subpattern sets. In the second step, PCA is performed progressively on these subpattern sets as shown Fig.1.

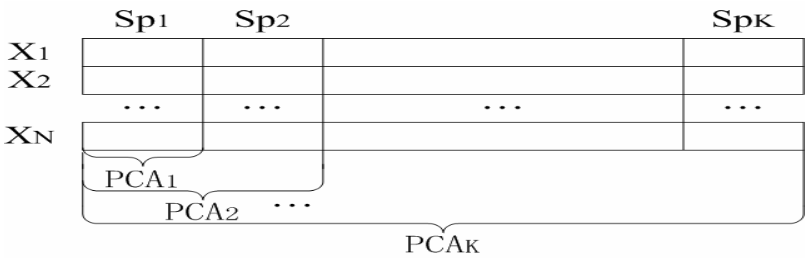


Fig. 1. Progressive feature extraction of PrPCA

More specifically, we are given a set of training patterns with large dimensionality:

$$X = \{X_1, X_2, \cdots, X_N\} \quad \textbf{(1)}$$

with each column vector X_i (i=1, 2, … , N) having m dimensions. Now according to the first step, an original whole pattern is first partitioned into K d-dimensional subpatterns in a nonoverlapping way and reshaped into a d-by-K matrix

$$X_i = (X_{i1}, X_{i2}, \cdots, X_{iK}) \quad \textbf{(2)}$$

with

$$X_{ij} = (x_{i((j-1)d+1)}, \cdots, x_{i(jd)})^T \quad \textbf{(3)}$$

being the jth subpattern of X_i and i=1,2,…, N and j=1,2,…,K. Then according to the second step, we perform PCA progressively from left to right as shown in Fig. 1, i.e.,

PCA_1 performs on the first subpattern set $\{X_{11}, X_{21}, \ldots, X_{N1}\}$ and then obtains corresponding reduced subpatterns $\{P_{11}, P_{21}, \ldots, P_{N1}\}$ which are, in turn, used to augment the subpattern dimensionality of the next set to construct new progressive training set

$$\left\{\begin{bmatrix} P_{11} \\ X_{12} \end{bmatrix}, \begin{bmatrix} P_{21} \\ X_{22} \end{bmatrix}, \cdots, \begin{bmatrix} P_{N1} \\ X_{N2} \end{bmatrix}\right\} \quad \textbf{(4)}$$

on which PCA_2 is performed, …, and so on. Finally we will terminate to some PCA_k ($k \leqslant K$) depending on the predefined recognition accuracy. Here for completeness of description, we give a concise introduction to the PCA as follows:

Without loss of generality, let us abuse a notation of the training set

$$X = \{X_1, X_2, \cdots, X_N\} \quad \textbf{(5)}$$

and perform PCA on it. Define a covariance matrix for the X,

$$C = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})(X_i - \overline{X})^T \quad \textbf{(6)}$$

where

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \quad \textbf{(7)}$$

is the sample mean. PCA can find a set of optimal projection vectors

$$\Phi = (\varphi_1, \varphi_2, \cdots, \varphi_l) \quad \textbf{(8)}$$

to ensure the minimal reconstruction error by solving the eigenvalue-eigenvector system

$$C\Phi = \Phi\Lambda \quad \textbf{(9)}$$

under the constraints that

$$\Phi^T \Phi = I \quad \textbf{(10)}$$

where I is an identity matrix and

$$\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_l) \quad \textbf{(11)}$$

a diagonal matrix composed by the first l largest non-negative eigenvalues of C in a descending order and thus their corresponding first l eigenvectors compose the Φ. l is the smallest value that is determined by the criterion

$$\sum_{i=1}^{l}\lambda_i \Big/ \sum_{i=1}^{N}\lambda_i \geq \theta \text{ (Here } \theta \text{ is a user prespecified parameter)}. \quad \textbf{(12)}$$

In this way, the final obtained PCA_k ($k \leqslant K$) projections can be used to extract features for any pattern and subsequently those extracted features are in turn used to pattern recognition with 1-nearest neighbor (1NN). To verify the feasibility of the PrPCA, we use a real FERET face database [3] to carry out the following experiments in the next section.

3 Experimental Results

3.1 Dataset

The FERET database comprises 400 gray-level frontal view face images from 200 persons, with the size of 256x384. There are 71 females and 129 males, each of whom has two images (*fa* and *fb*) with different race, different gender, different age, different expression, different illumination, different occlusion, different scale, etc. The *fa* images are used for training while the *fb* images for testing. In our experiments, all faces are normalized to satisfy some constraints so that each face could be appropriately cropped. Those constraints include that the line between the two eyes is parallel to the horizontal axis, the inter-ocular distance (distance between the two eyes) is set to a fixed value, and the size of the image is fixed. Here the eyes are manually located and after a series of rotating and resizing, the cropped image size is 60×60 pixels and the inter-ocular distance is 28 pixels. In the experiments, we use two partition ways for the whole pattern: sequent (seq) and random (rand) and investigate their influence on recognition results.

3.2 Results and conclusions

We compare the obtained classification results with those of $E(PC)^2A$ (best 85.5%)[1][Appendix] and FLDA(best 86.5%) [4].

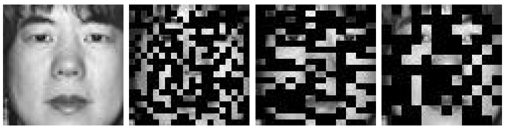
Table 1 shows the recognition accuracies (RAs) of both PrPCAs and SpPCAs for different partitions with changeable sizes of blocks and the testing patterns including differently missing blocks.

Table 1. Recognition accuracies (RA) (%) and percentages of kept blocks (POKB)

block size	3×3		3×5		5×5	
	RA	POKB	RA	POKB	RA	POKB
PrPCA(seq)	86.5	93.75	86.5	92.5	86.5	92.36
PrPCA(rand)	89	41.3	88	37.5	88	38.9
SpPCA[a]	85	100	85	100	85	100
SpPCA	87(0.9)[b]	100	87.5(0.8)[b]	100	87(0.8)[b]	100

Notice: *a*: θ are all set to 0.99999 in this row to obtain the RAs and POKBs;
b: θ values corresponding to achieve the best RAs.

Fig.2 shows three randomly missing block cases able to achieve corresponding best RAs with different block-size partitions.



(a) (b) (c) (d)

Fig. 2. Randomly kept block percentages and corresponding best recognition accuracies (%). (a) Original face; (b) 41.3 (3×3) and 89; (c) 37.5 (3×5) and 88; (d) 38.9 (5×5) and 88. Note: the dark blocks represent discarded or missing parts.

From the table 1, we can observe that the RAs (88-89%) of PrPCA with random partitions are higher than the RAs (86.5% and 87.5%) of both the counterparts with sequential partitions, the SpPCAs and FLDA[4]. At the same time, the tests on the SpPCA indicate the RAs (85%s) with preserving all original information are not necessarily better than the ones (87-87.5%) with partial original information preserved (θ values are respectively set to 0.9, 0.8 and 0.8 as listed in Table 1). Furthermore, in order to test the robustness of the proposed PrPCA, we adopt a randomly discarding image block way for the faces to be recognized (as shown in Fig. 2) to examine its RAs and find that it can still achieve the best RA of 88% in the random partition even when blocks of 61.1% in faces are missed (equivalently, 38.9% blocks are kept.). These results confirm feasibility and effectiveness of the PrPCA and finally the idea used here can be applied to other similar problems.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under the Grant No. 60271017, the National Outstanding Youth Foundation of China under the Grant No. 60325207, the Natural Science Foundation of Jiangsu Province under the Grant No. BK2002092, the *QingLan* Project Foundation of Jiangsu Province, and the Returnee Foundation of China Scholarship Council. Portions of the research in this paper use the FERET database of facial images collected under the FERET program.

References

1. Chen, S.C., Zhang, D.Q., Zhou, Z.H.: Enhanced $(PC)^2A$ for face recognition with one training image per person. Pattern Recognition Letters, 2004, in press.
2. Chen, S.C., Zhu, Y. L.: Subpattern-based principle component analysis. Pattern Recognition, 2004, 37: 1081-1083.
3. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.J.: The FERET database and evaluation procedure for face-recognition algorithms. Image and Vision Computing, 1998, 16(5): 295-306.

4. Chen, S.C., Liu, J., Zhou, Z.H.: Making FLDA applicable to face recognition with one sample per person. Pattern Recognition, 2004, in press.

Appendix: E(PC)^{2}A

The aim of E(PC)^{2}A is to augment the sample size by constructing the original face's n-order projections. Let $I(x,y)$ be an intensity image of size $N_1 \times N_2$, where $x \in [1,N_1]$, $y \in [1,N_2]$, and $I(x,y) \in [0,1]$, we can define the second-order projection of the original image as

$$P_2(x,y) = \frac{V_2(x) H_2(y)}{J_{mean}} \quad \textbf{(13)}$$

where J_{mean} is the mean value of $J(x,y)$ which is defined as the square of $I(x,y)$, that is, $J(x,y) = I(x,y)^2$, and V_2 and H_2 are defined respectively as:

$$V_2(x) = \frac{1}{N_2} \sum_{y=1}^{N_2} J(x,y) \quad \textbf{(14)}$$

$$H_2(y) = \frac{1}{N_1} \sum_{x=1}^{N_1} J(x,y) \quad \textbf{(15)}$$

Then, through combining the original image with its first and second-order projections, a new projection-combined image, i.e. $I_2(x, y)$ as shown in Eq.A.4, can be obtained, where α and β are parameters used to control the bias of the projections.

$$I_2(x,y) = \frac{I(x,y) + \alpha P_1(x,y) + \beta P_2(x,y)}{1 + \alpha + \beta} \quad \textbf{(16)}$$