

Probabilistic two-dimensional principal component analysis and its mixture model for face recognition

Haixian Wang · Sibao Chen ·
Zilan Hu · Bin Luo

Received: 6 August 2006 / Accepted: 5 October 2007 / Published online: 25 October 2007
© Springer-Verlag London Limited 2007

Abstract Recently, two-dimensional principal component analysis (2DPCA) as a novel eigenvector-based method has proved to be an efficient technique for image feature extraction and representation. In this paper, by supposing a parametric Gaussian distribution over the image space (spanned by the row vectors of 2D image matrices) and a spherical Gaussian noise model for the image, we endow the 2DPCA with a probabilistic framework called probabilistic 2DPCA (P2DPCA), which is robust to noise. Further, by using the probabilistic perspective of P2DPCA, we extend the P2DPCA to a mixture of local P2DPCA models (MP2DPCA). The MP2DPCA offers us a method of being able to model faces in unconstrained (complex) environment. The model parameters could be fitted on the basis of maximum likelihood (ML) estimation via the expectation maximization (EM) algorithm. The experimental recognition results on UMIST, AR face database, and the face recognition (FR)

data collected at University of Essex confirm the effectivity of the proposed methods.

Keywords Two-dimensional principal component analysis · Probabilistic two-dimensional principal component analysis · Mixture model · EM algorithm · Face recognition

1 Introduction

The importance of research on face recognition (FR) is driven by both its wide range of potential applications and scientific challenges. And the appearance-based paradigm using the principal component analysis (PCA) [1], also known as Karhunen–Loève transformation, to extract features has exhibited great advantage, producing the well-known eigenfaces method [2, 3]. Now, the eigenfaces method has become one of the most successful approaches in FR (to see Ref. [4], for an early survey of FR). However, it could be noted that eigenfaces approach is vector-oriented; that is, in FR system using the eigenfaces method, the two-dimensional (2D) facial image matrices must be transformed into one-dimensional (1D) vectors prior to performing PCA. Typically, the resulting image vectors may be very high-dimensional, say 16,384-dimensional (a facial image with a resolution of 128×128), where it is difficult to handle the (possibly singular) covariance matrix because of the high dimension and relatively small training sample size even though the eigenvectors could be efficiently computed without the accurate calculation of the covariance matrix.

Recently, to attack the problem, Yang et al. [5] proposed the two-dimensional principal component analysis (2DPCA), which was directly based on 2D image matrices rather than 1D vectors and thus obviated the transformation

H. Wang (✉)
Key Laboratory of Child Development and Learning Science
of Ministry of Education, Southeast University,
Nanjing, Jiangsu 210096, China
e-mail: hxwang@seu.edu.cn

S. Chen
Department of Electronic Engineering and Information Science,
University of Science and Technology of China,
Hefei, Anhui 230027, China

Z. Hu
School of Mathematics and Physics,
Anhui University of Technology,
Maanshan, Anhui 243002, China

B. Luo
Key Lab of ICSP of Ministry of Education,
Anhui University, Hefei, Anhui 230039, China

from 2D matrices to 1D vectors by lexicographic ordering of the pixel elements of each image. Following the introduction of 2DPCA immediately, Wang et al. [6] showed the equivalence of 2DPCA to line-based PCA. And the generalization of 2DPCA to bilateral and kernel-based versions were also presented [7, 8]. The 2DPCA(-based) method is appealing, since the 2D spatial information of the image is well preserved and the computational complexity is significantly reduced compared with the PCA. Moreover, the 2DPCA could naturally and effectively avoid the small sample size (SSS) problem.

Like PCA, one flaw of 2DPCA model, however, is the absence of an associated probability density; that is, the process of implementing 2DPCA is distribution-free—a mathematical method with no underlying statistical model. The proposal of probabilistic PCA (PPCA) [9], which defines PCA from the probabilistic perspective, aims to remedy the disadvantage of PCA. In this paper, by supposing a parametric Gaussian distribution over the image space (spanned by the row vectors of 2D image matrices) and a spherical Gaussian noise model for the image, we endow the 2DPCA with a probabilistic framework, which we will refer to as probabilistic 2DPCA (P2DPCA). Such a probabilistic formulation enables the application of Bayesian method. And on the basis of maximum likelihood (ML) estimation, the estimates of model parameters could be implemented via the expectation maximization (EM) algorithm (which, under mild conditions, converges to the ML estimation) [10]. Besides, the probability model could fit the noise in the data whereas the 2DPCA is not robust to independent noise. Further, whereas the P2DPCA only defines a global projection of the sample vectors in the image space, we extend the P2DPCA to a mixture of local P2DPCA models (MP2DPCA). We adopt the P2DPCA and MP2DPCA as mechanisms for extracting facial information followed by discrimination. The experimental results demonstrate their superiority over the 2DPCA.

A preliminary work of this paper was appeared in Ref. [11]. The rest of the paper is organized as follows: In Sect. 2, we briefly review the formulation of 2DPCA. In Sect. 3, we propose the P2DPCA model, while the MP2DPCA model is presented in Sect. 4. Section 5 reports experimental results and Sect. 6 ends the paper by addressing some concluding remarks.

2 Two-dimensional principal component analysis

In the case of eigenfaces-based FR, the transformation from image matrices into vectors before performing PCA is a disadvantage and may lead to a singular covariance matrix for relatively small number of training samples. However, the 2DPCA seeks to carry out PCA immediately based on 2D

image matrices. The idea behind 2DPCA is to project image \mathbf{X} , an $s \times t$ random matrix, onto μ by the linear transformation $\mathbf{z} = \mathbf{X}\mu$. The projection direction μ is determined by maximizing the total scatter of the resulting projected samples which is characterized by the trace of the covariance matrix of the projected feature vectors. Namely, the 2DPCA seeks to maximize the criterion $J(\mu) = \text{tr}(\text{COV})$, where COV denotes the covariance matrix of the projected features vectors of the training samples and tr is the trace operator.

Mathematically, for a set of observed $s \times t$ image samples $\mathbf{X}_1, \dots, \mathbf{X}_n$, the q ($q \leq t$, usually $q \ll t$) principal axes μ_j are given by the q dominant eigenvectors of the image covariance matrix $\mathbf{S}_I = \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^T (\mathbf{X}_i - \bar{\mathbf{X}})$, where $\bar{\mathbf{X}}$ is the mean image $\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$, such that $\mathbf{S}_I \mu_j = \tau_j \mu_j$. For a given image sample \mathbf{X} , the principal components (vectors) are given by $\mathbf{z}_j = \mathbf{X}\mu_j$, $j = 1, \dots, q$. The principal component vectors obtained are to form an $s \times q$ feature matrix $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_q)$. The linear reconstruction of \mathbf{X} is given by $\hat{\mathbf{X}} = \mathbf{Z}\mathbf{U}^T$, where $\mathbf{U} = (\mu_1, \dots, \mu_q)$.

The essential of 2DPCA is to view each row vector of 2D image matrix as observed sample, and the q principal axes μ_j , $j = 1, \dots, q$, are those orthonormal axes onto which the sum of variances of row vectors under projection is maximal. Particularly, when each image matrix comprises only one row vector, the 2DPCA becomes PCA. Therefore, the 2DPCA is also an orthogonal transformation of the coordinate axes in which we describe the row vectors of observed 2D images. By statistical manipulation and matrix representation, the 2DPCA is reduced to solving an eigenvalue problem of image covariance matrix constructed directly via image matrices.

3 Probabilistic model for 2DPCA

Equations $\mathbf{z}_j = \mathbf{X}\mu_j$ above relate the principal component vectors \mathbf{z}_j to the observed random image \mathbf{X} . In general \mathbf{z}_j will not have zero mean. In order for the principal component vectors to have zero mean, they should be defined as $\mathbf{z}_j = (\mathbf{X} - \Xi)\mu_j$, $j = 1, \dots, q$, for mean matrix Ξ . In practice Ξ is the sample mean $\bar{\mathbf{X}}$. Below, for the sake of simplicity, we made the assumption that the image samples are already mean centered (which is easy to achieve). And we denote the k th row vector of the i th (mean-centered) image sample \mathbf{X}_i as \mathbf{X}_i^k for $k = 1, \dots, s$, $i = 1, \dots, n$; namely, $\mathbf{X}_i^T = (\mathbf{X}_i^1, \dots, \mathbf{X}_i^s)$. Then, we have the following:

Theorem Regarding \mathbf{X}_i^k , $k = 1, \dots, s$, $i = 1, \dots, n$, as t -dimensional observations of size sn , and then applying the model of PPCA to these samples:

$$\mathbf{X}_i^k = \Lambda F + \varepsilon, \text{ for } k = 1, \dots, s, i = 1, \dots, n, \quad (1)$$

in which Λ is a $t \times q$ factor loading matrix, F is q -dimensional latent (unobservable) variables (also known as

common factors), and ε is t -dimensional specific factors. This generative model assumes that the latent variables are independently and identically distributed (i.i.d.) as Gaussian with zero-mean and identity covariance, the zero mean specific factors are also i.i.d. as Gaussian with covariance σ^2 , and F is independent with ε and their joint distribution is Gaussian. Then one has that the columns of the ML estimator of Λ span the principal subspace as in the 2DPCA.

Proof Since the image samples are already mean centered, that is $\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i = 0$, we obtain $\frac{1}{sn} \sum_{i=1}^n \sum_{k=1}^s \mathbf{X}_i^k = 0$. Then we calculate the “sample” covariance matrix for \mathbf{X}_i^k , $k = 1, \dots, s, i = 1, \dots, n$ as

$$\mathbf{S} = \frac{1}{sn} \sum_{i=1}^n \sum_{k=1}^s \mathbf{X}_i^k \mathbf{X}_i^{kT} = \frac{1}{sn} \sum_{i=1}^n \mathbf{X}_i^T \mathbf{X}_i = \frac{1}{s} \mathbf{S}_I. \tag{2}$$

And by the properties of the ML estimators of PPCA [9], the ML estimator for Λ could be written as

$$\hat{\Lambda} = W_q (\Gamma_q - \hat{\sigma}^2 I_q)^{\frac{1}{2}} R, \tag{3}$$

where the q column vectors in the $t \times q$ matrix W_q are the principal eigenvectors of \mathbf{S} , with corresponding eigenvalues in the $q \times q$ diagonal matrix Γ_q , $\hat{\sigma}^2$ is the ML estimate of the global noise level. On account of that \mathbf{S} is just the image covariance matrix \mathbf{S}_I (up to a scalar), W_q comprises the principal axes as in the 2DPCA. \square

Such formulation may be termed generative model, since data vectors \mathbf{X}_i^k could be generated by sampling from the F and ε distributions and applying (1). Correspondingly, In FR the generative model (1) could be interpreted as follows. An observed facial image is assumed to be generated via three steps: firstly choose s points for the rows of image from the principal subspace spanned by $\hat{\Lambda}$, secondly add a zero-mean Gaussian “noise” ε (including real noise and conceptual noise such as whisker detail, eyelash detail and hair detail, etc.) to each point, and finally pile the chosen points (which are all t -dimensional) with added noise and mean image to form the observed facial image. Figure 1 schematically shows the procedures.

3.1 Feature extraction and reconstruction

The general objective of 2DPCA is to seek some reduced-dimensionality representation of the observed images. However, from the probabilistic perspective, it is natural to adopt the posterior mean $E[F|\mathbf{X}^k]$ (an estimation of factor scores) as the reduced-dimensionality transformation for \mathbf{X}^k (which is the k th row vector of an image \mathbf{X}). Specifically, in the model of P2DPCA, the posterior mean $E[F|\mathbf{X}^k]$ becomes

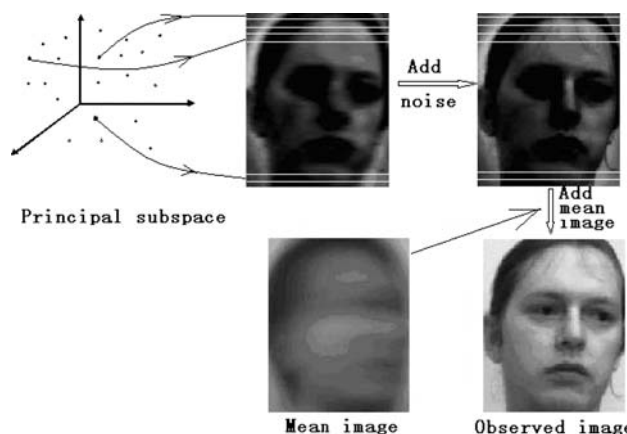


Fig. 1 The procedures of generating an observed facial image

$$\beta(\mathbf{X}^k) = (\hat{\Lambda}^T \hat{\Lambda} + \hat{\sigma}^2 I_q)^{-1} \hat{\Lambda}^T \mathbf{X}^k, \quad k = 1, \dots, s. \tag{4}$$

So, for an observed image \mathbf{X} , the reduced-dimensionality representation, that is an $s \times q$ feature matrix, is given by

$$\mathcal{Z} = (\beta(\mathbf{X}^1), \dots, \beta(\mathbf{X}^s))^T = \mathbf{X} \hat{\Lambda} (\hat{\Lambda}^T \hat{\Lambda} + \hat{\sigma}^2 I_q)^{-1}. \tag{5}$$

As a result, the reconstructed image of \mathbf{X} could be

$$\hat{\mathbf{X}} = (\hat{\Lambda} \beta(\mathbf{X}^1), \dots, \hat{\Lambda} \beta(\mathbf{X}^s))^T + \bar{\mathbf{X}} = \mathcal{Z} \hat{\Lambda}^T + \bar{\mathbf{X}}. \tag{6}$$

It can be seen that when $\hat{\sigma}^2 \rightarrow 0$, the rows of \mathcal{Z} represent projections onto the latent space and the conventional 2DPCA is recovered.

3.2 An EM algorithm for P2DPCA

We can obtain an EM algorithm for finding the model parameters by maximizing the corresponding likelihood function. There may be computational advantage in the EM approach for large t .

In the EM framework, the estimates for $\hat{\Lambda}$ and $\hat{\sigma}^2$ could be updated as follows:

$$\hat{\Lambda}^{(k+1)} = \mathbf{S}_I \hat{\Lambda}^{(k)} \left(s \hat{\sigma}^{2(k)} I_q + \Delta^{(k)-1} \hat{\Lambda}^{(k)T} \mathbf{S}_I \hat{\Lambda}^{(k)} \right)^{-1}, \tag{7}$$

$$\hat{\sigma}^{2(k+1)} = \frac{1}{st} \text{tr} \left(\mathbf{S}_I - \mathbf{S}_I \hat{\Lambda}^{(k)} \Delta^{(k)-1} \hat{\Lambda}^{(k+1)T} \right), \tag{8}$$

where superscript (k) denotes the k th estimate, and $\Delta^{(k)} = \hat{\sigma}^{2(k)} I_q + \hat{\Lambda}^{(k)T} \hat{\Lambda}^{(k)}$. The complexity of the EM algorithm is limited by $O(sntq)$ per iteration.

4 Mixtures of P2DPCA models

The 2DPCA (or P2DPCA) is, essentially, a linear model for data representation in a low dimensional subspace. It may

be insufficient to model data with large variation caused by, for example, pose, expression and lighting. An alternative choice is to model the complex manifold with a mixture of local linear sub-models from the probabilistic formulation of P2DPCA. For $j = 1, \dots, s; i = 1, \dots, n$, we suppose that the t -dimensional samples \mathbf{X}_i^j are generated independently from a mixture of g underlying populations with unknown proportion π_1, \dots, π_g

$$\mathbf{X}_i^j = M_m + A_m F_m + \varepsilon_m, \quad \text{with probability } \pi_m (m = 1, \dots, g), \quad (9)$$

where M_m are t -dimensional mean vectors, and π_m are mixing proportions with $\pi_m > 0$ and $\sum_{m=1}^g \pi_m = 1$. Note that a separate M_m is associated with each component of the mixture model, therefore allowing each to model the data covariance structure in a different region of the data space. While the P2DPCA method uses one set of features for the data points, the mixture of P2DPCAs (MP2DPCA) uses more than one set of features. Therefore, the MP2DPCA is expected to represent data more effectively and has better recognition performance than the P2DPCA method.

Also, the model parameters Λ_m , σ_m^2 and M_m could be estimated by using the EM algorithm. Specifically, in the E -step, the posterior probability that \mathbf{x}_i^j belongs to the m th component of the mixture, using the current estimate $\hat{\Theta}^{(k)}$ for Θ , is

$$\tau_m^{(k)}(i, j) = \frac{\pi_m^{(k)} \varphi(\mathbf{x}_i^j; M_m^{(k)}, \Sigma_m^{(k)})}{\sum_{l=1}^g \pi_l^{(k)} \varphi(\mathbf{x}_i^j; M_l^{(k)}, \Sigma_l^{(k)})}, \quad (10)$$

$$(m = 1, \dots, g; j = 1, \dots, s; i = 1, \dots, N),$$

where φ denotes the probability density function of normal distribution, and

$$\Sigma_m = \Lambda_m \Lambda_m^T + \sigma_m^2 I_t, \quad (m = 1, \dots, g).$$

In the M -step, for $m = 1, \dots, g$, we have the updates as follows

$$\pi_m^{(k+1)} = \frac{1}{sN} \sum_{i=1}^N \sum_{j=1}^s \tau_m^{(k)}(i, j), \quad (11)$$

$$M_m^{(k+1)} = \frac{\sum_{i=1}^N \sum_{j=1}^s \tau_m^{(k)}(i, j) \mathbf{x}_i^j}{\sum_{i=1}^N \sum_{j=1}^s \tau_m^{(k)}(i, j)}. \quad (12)$$

If let

$$S_m^{(k+1)} = \frac{1}{sN \pi_m^{(k+1)}} \sum_{i=1}^N \sum_{j=1}^s \tau_m^{(k)}(i, j) (\mathbf{x}_i^j - M_m^{(k+1)}) (\mathbf{x}_i^j - M_m^{(k+1)})^T,$$

then the updates for Λ_m , σ_m^2 could be obtained by using eigen-decomposition [9]. However, they still could be iteratively computed as

$$\Lambda_m^{(k+1)} = S_m^{(k+1)} \Lambda_m^{(k)} \left(\sigma_m^{2(k)} I_q + \Lambda_m^{(k)-1} \Lambda_m^{(k)T} S_m^{(k+1)} \Lambda_m^{(k)} \right)^{-1}, \quad (13)$$

$$\sigma_m^{2(k+1)} = \frac{1}{t} \text{tr} \left(S_m^{(k+1)} - S_m^{(k+1)} \Lambda_m^{(k)} \Lambda_m^{(k)-1} \Lambda_m^{(k+1)T} \right), \quad (14)$$

where

$$\Lambda_m^{(k)} = \sigma_m^{2(k)} I_q + \Lambda_m^{(k)T} \Lambda_m^{(k)}. \quad (15)$$

The derivations of these formula are similar with Ref. [9]. By applying the MP2DPCA model, all the samples are softly divided into g clusters each modelled by a local P2DPCA. We use the most appropriate local P2DPCA for a given sample in terms of the fitted posterior probability. Based on the probabilistic framework, a natural choice is to assign the sample to a cluster belong to which its posterior probability is the largest.

5 Experiments

In this section, we compare the recognition performances of 2DPCA, P2DPCA and MP2DPCA on three benchmark databases: the UMIST face database, AR face database, and FR data named “faces94” collected at University of Essex, UK. The UMIST database (available from: <http://images.ee.umist.ac.uk/danny/database.html>) consists of 564 images of 20 subjects [12]. The face images of each person cover a range of poses from profile to frontal views. Subjects vary with respect to race, sex and appearance. The files are all in PGM format, approximately 220×220 pixels with 256 grey levels. Pre-cropped versions (with a size of 112×92) of the images may be also made available from the same web site and they are used in our experiments. Figure 2 shows the 20 images of one subject. The AR face database (available from: http://rv11.ecn.purdue.edu/~aleix/aleix_face_DB.html) contains over 4,000 color images corresponding to 126 subjects (70 males and 56 females) [13]. There are variations of facial expressions, illumination conditions, and occlusions (sun glasses and scarf) with each person. Each individual consists of 26 frontal view images taken in two sessions (separated by 2 weeks), where each session has 13 images. The original images are of size 768×576 pixels and of 24 bits of RGB color values. Figure 3 shows the 26 images of one subject. In the experiment, we select 70 subjects (50 man and 20 women), and only use the non-occluded 14 images [i.e., those numbered (a) through (g) and (n) through (t)] of each selected subject for evaluation. These images are then cropped and resized to 100×100 pixels. The FR data named “faces94” collected at University of Essex (available from: <http://cswww.essex.ac.uk/mv/allfaces/index.html>) were taken when the participants sat



Fig. 2 Twenty face examples of one subject from the UMIST face database

at fixed distance from the camera and were asked to speak. This database contains 152 individuals, and each person has 20 color images. There are considerable expression changes with each person. The original image size is 200×180 pixels. In the experiment, we select 70 subjects (51 man and 19 women) for evaluation. The selected images are further down-sampled into 100×90 pixels.

The 2DPCA, P2DPCA and MP2DPCA methods are used for extracting features of facial images from the training samples, respectively, and then a nearest-neighbor classifier is used to find the most-similar face from the

training samples for a queried face. In our experiments, the measure of distance between two feature matrices $Z_{i_1} = (A_1^{(i_1)}, \dots, A_s^{(i_1)})^T$ and $Z_{i_2} = (A_1^{(i_2)}, \dots, A_s^{(i_2)})^T$, is defined as the sum of s Euclidean distances between each corresponding row of the two matrices.

5.1 UMIST database

The difference in representation becomes evident when considering the quality of reconstructed images using the

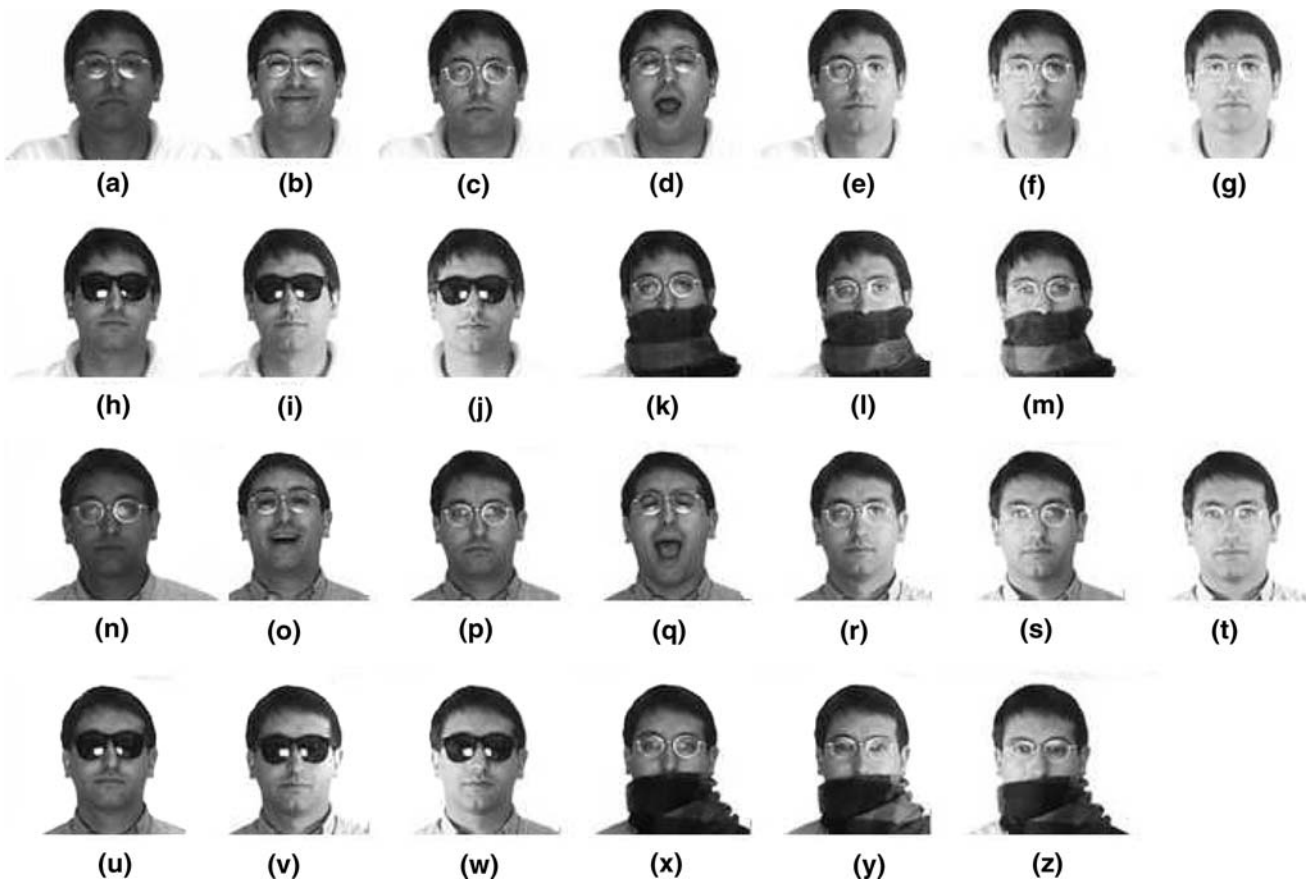
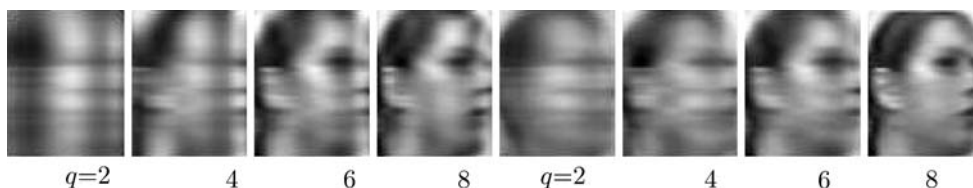


Fig. 3 Twenty-six face examples of one subject from the AR face database. The images (a) through (m) are from the first session, and the rest 13 images are from the second session

Fig. 4 2DPCA versus P2DPCA reconstructions for a novel testing view. The left and right four images correspond to 2DPCA and P2DPCA reconstructions with varying q , respectively



2DPCA and P2DPCA methods. Figure 4 compares the reconstructed images obtained with the two methods when trained on ten image samples per person. It could be seen that the P2DPCA provides a better characterization of the object.

We now adopt two strategies to evaluate the performances of 2DPCA, P2DPCA and MP2DPCA. In the first series of experiment the interpolation performance is tested by training on a subset of the available views and testing on the intermediate views. We select ten images per person for training and another ten images per person for testing. So, the total number of training samples and testing samples are both 200. The experiment is run ten times, and the average recognition rates obtained are recorded, as shown in Table 1. The second series of experiments test on the extrapolation performance by training on a range of views and testing on novel views outside the training range. The average recognition of ten times are listed in Table 2. From Table 1, it could be seen that the three methods achieve high recognition rate in the case of interpolation. While in the case of extrapolation, as illustrated in Table 2, the MP2DPCA method obtains the best recognition rate among the three systems. The reason may be that, when there is large pose variation, different pose may appear more separated by using the

local subspace. And the MP2DPCA is well suited to model such complex facial images. In this experiment, the 2DPCA is a benchmark for evaluation. As a matter of fact, for FR domain the 2DPCA works better than many other methods, including eigenfaces, fisherfaces, ICA, and Kernel eigenfaces, in terms of recognition rate [5]. From Table 2, we notice a paradoxical thing is that the performance declines as q increases. The reason could possibly be that in the extrapolation situation only the most principal components are effective and the samples are “over”-learned if q is large.

5.2 AR database

In the experiment, we randomly select seven images of each person as the training samples, and use the remaining images to form the testing sample set. The cross-validation strategy is applied, and the experiment is run ten times. The average recognition rates across ten rounds of experiments are shown in Table 3. From Table 3, we see that MP2DPCA achieves the overall best recognition performance, and P2DPCA slightly outperform 2DPCA. This two points are consistent with the experimental results in above subsection.

Table 1 Interpolation performances of 2DPCA, P2DPCA and MP2DPCA with varying q on the UMIST database (%)

q	2DPCA	P2DPCA	MP2DPCA			
			$g = 2$	3	4	5
2	98.5	98.5	99.0	98.0	99.0	96.0
4	100	100	100	100	99.0	100
6	99.5	99.5	100	100	100	99.5
8	99.5	100	100	100	100	100

Table 3 The comparison of average recognition rates of 2DPCA, P2DPCA and MP2DPCA with varying q on the AR database (%)

q	2DPCA	P2DPCA	MP2DPCA			
			$g = 2$	3	4	5
2	80.2	80.4	81.9	81.5	81.5	79.4
4	88.1	89.5	91.2	90.7	88.5	89.7
6	91.4	92.1	94.6	94.9	94.3	94.6
8	91.3	91.8	92.8	92.8	92.7	91.6

Table 2 Extrapolation performances of 2DPCA, P2DPCA and MP2DPCA with varying q on the UMIST database (%)

q	2DPCA	P2DPCA	MP2DPCA			
			$g = 2$	3	4	5
2	73.0	73.0	76.0	76.0	79.0	79.0
4	68.0	68.0	76.0	70.0	73.0	70.5
6	60.5	62.5	66.5	65.5	61.5	70.0
8	55.5	59.0	62.0	61.5	64.5	65.0

Table 4 The comparison of average recognition rates of 2DPCA, P2DPCA and MP2DPCA with varying q on the database faces94 of University of Essex (%)

q	2DPCA	P2DPCA	MP2DPCA			
			$g = 2$	3	4	5
2	81.6	83.7	88.9	86.2	83.7	86.5
4	88.9	89.6	93.8	90.3	92.4	92.9
6	83.2	85.8	86.3	89.5	87.9	87.3
8	82.5	82.3	87.4	87.6	86.2	84.3

5.3 Database faces94 of University of Essex

In the experiment, we randomly select ten images of each person to form the training sample set, and use the remaining images for testing. We use the same experimental procedure as above subsection. The average recognition rates across 10 rounds of experiments are shown in Table 4, from which we again see that MP2DPCA and P2DPCA overall outperform conventional 2DPCA.

6 Conclusion and future work

In this paper, a probabilistic model for the 2DPCA, termed P2DPCA, is developed. And the proposal of MP2DPCA offers a tempting prospect of being able to model faces in unconstrained (complex) environment, for example, with possible varying poses, facial expressions and illumination conditions. All the model parameters could be learned via the EM algorithm on the basis of ML estimation. The proposed methods allow a convenient way for image feature extraction and representation. The probabilistic models could deal properly with missing data problem, which cope well with the partially occluded images when regarding the occluded parts as missing values. This is our future work.

Acknowledgments The authors would like to express their gratitude to the anonymous referees for help recommendations. The authors also gratefully acknowledge the financial support of National Natural Science Foundation of China (Grant Nos. 10571001, 60603083, and 60375010).

References

- Jolliffe IT (1986) Principal component analysis. Springer, New York
- Turk M, Pentland A (1991) Eigenfaces for recognition. *J Cogn Neurosci* 3:71–86
- Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans Pattern Anal Mach Intell* 19:711–720
- Zhao W, Chellappa R, Rosenfeld A, Phillips PJ (2000) Face recognition: a literature survey. Technical Report CAR-TR-948, Vision Technologies Lab, Sarnoff Corporation, Princeton
- Yang J, Zhang D, Frangi AF, Yang Jy (2004) Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Trans Pattern Anal Mach Intell* 26:1–7
- Wang L, Wang X, Zhang X, Feng J (2005) The equivalence of two-dimensional PCA to line-based PCA. *Pattern Recognit Lett* 26:57–60
- Kong H, Wang L, Teoh EK, Li X, Wang JG, Venkateswarlu R (2005) Generalized 2D principal component analysis for face image representation and recognition. *Neural Netw* 18:585–594
- Zhang D, Zhou Z-H (2005) (2D)2PCA: two-directional two-dimensional PCA for efficient face representation. *Neurocomputing* 69:224–231
- Tipping ME, Bishop CM (1999) Mixtures of probabilistic principal component analysers. *Neural Comput* 11:443–482
- McLachlan GJ, Krishnan T (1997) The EM algorithm and extensions. Wiley, New York
- Wang H, Hu Z (2006) Face recognition using probabilistic two-dimensional principal component analysis and its mixture model. In: Jiao L et al (eds) The second international conference on natural computation, Xi'an, China, Lecture Notes in Computer Science 4221, pp. 337–340
- Graham DB, Allinson NM (1998) Characterizing virtual eigen-signatures for general purpose face recognition. In: Wechsler H, Phillips PJ, Bruce V, Fogelman-Soulie F, Huang TS (eds) Face recognition: from theory to applications, NATO ASI series F, Computer and systems sciences, vol 163, pp. 446–456
- Martinez AM, Benavente R (1998) The AR face database. CVC technical report 24, Barcelona, Spain