Enhanced $(PC)^2A$ for Face Recognition with One Training Image per Person

Songcan Chen[1, 2*], Daoqiang Zhang[1] and Zhi-Hua Zhou[2]

[1] *Department of Computer Science and Engineering*
Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

[2] *State Key Laboratory for Novel Software Technology*
Nanjing University, Nanjing 210093, China

Abstract: Recently, a method called $(PC)^2A$ was proposed to deal with face recognition with one training image per person. As an extension of the standard eigenface technique, $(PC)^2A$ combines the original face image with its first-order projection and then performs principal component analysis (PCA) on the enriched version of the image. It was reported that $(PC)^2A$ could achieve higher accuracy than the eigenface technique through using 10%-15% fewer eigenfaces. In this paper, we generalize and further enhance $(PC)^2A$ along two directions. In the first direction, we combine the original image with its second-order projection as well as its first-order projection in order to acquire more information, and then apply PCA on the derived images. In the second direction, instead of combining the original image with its projections, we regard the projections of the original images as derived images that could augment training information, and then apply PCA on all the training images available, including the original ones and the derived ones. Experiments on the well-known FERET database show that the enhanced versions of $(PC)^2A$ are about 1.6% to 3.5% more accurate and use about 47.5% to 64.8% fewer eigenfaces than $(PC)^2A$.

Keywords: Face recognition; Principal component analysis; Eigenface; Extended PCA

[*] Corresponding author. Tel.:+86-25-8489-2805, fax: +86-25-8489-3777.
E-mail address: s.chen@nuaa.edu.cn (S. Chen), daoqz@mail.com (D. Zhang), zhouzh@nju.edu.cn (Z.-H. Zhou)

1. Introduction

Face Recognition has been an active research area of computer vision and pattern recognition for decades (Turk and Pentland, 1991; Brunelli and Poggio, 1993; Chellappa et al., 1995; Moghaddam and Pentland, 1997; Moghaddam et al., 2000; Sukthankar, 2000; Wiskott et al., 1997; Zhao et al., 2000; Chen and Huang, 2003). Many face recognition methods have been proposed to date and according to (Brunelli and Poggio, 1993), these methods can be roughly classified into two categories, i.e., geometric feature-based and template-based. In the first category, the most often used method is the elastic bunch graph matching (Wiskott et al., 1997), while in the second category, the most widely used algorithm is the eigenface (Turk and Pentland, 1991). Recently, neural networks (Valentin et al., 1994; Zhang et al., 1997; Raytchev and Murase, 2003), support vector machines (Pang et al., 2003), kernel methods (Lu et al., 2003), and ensemble techniques (Pang et al., 2003) also find great applications in this area.

In some specific scenarios such as law enforcement, only one image per person can be used for training the face recognition system. It is unfortunate that most face recognition algorithms may have problems in such scenarios. For example, most subspace methods such as Linear Discriminant Analysis (LDA) (Etemad and Chellappa, 1997; Lu et al., 2003), discriminant eigenfeatures (Swets and Weng, 1996) and fisherface (Belhumeur et al., 1997) can hardly be used because in order to obtain good recognition performance, they require there exist at least two training images per person so that the intra-class variation could be considered against the inter-class variation. Recently, a few researchers begin to address this issue (Wu and Zhou, 2002; Martinez, 2002;). In (Wu and Zhou, 2002), a method called $(PC)^2A$ was proposed as an extension of the standard eigenface technique, which combines the original face image with its first-order projected image and then performs principal component analysis (PCA) on the enriched version of the image. It was reported that $(PC)^2A$ outperformed the standard eigenface technique when only one training image per person is available (Wu and Zhou, 2002). In (Martinez, 2002), a probabilistic approach was described, in which the model parameters were estimated by using a set of images generated around a so-called representative sample image, each with small localized errors within the eigenspace, or partially occluded and expression-variant faces corresponding to the sample image.

In this paper, we follow the line of Wu and Zhou (2002) but generalize and enhance $(PC)^2A$ in two ways. In the first way, besides the first-order projected image, we construct second-order projected images, combine these first and second-order projected images with the original image, and then perform PCA on the combined image. In the second way, instead of combining the original image with the projected images, we enlarge the training image database using a series of *n*-order projected images. That is to say, if there are *M* face images in the image database corresponding to *M* different persons, we can generate *n* additional images for each person and therefore obtaining an enlarged training database comprising $(n+1)M$ face images. Then we perform PCA on the enlarged image database. The idea behind these two ways is to squeeze as much information as possible from the single face images. These information can derive some salient features that are important in face recognition with one training image per person, therefore we get the first extended version. These information can also be used to provide each person with several imitated face images so that the problem of face recognition with one training image per person becomes a common face recognition problem, therefore we get the second extended version. Experiments have been performed on a subset of the well-known FERET database, and the experimental results show that both the enhanced versions of $(PC)^2A$ get improved recognition accuracy while the number of eigenfaces used is only about half of that used by $(PC)^2A$.

The rest of this paper is organized as follows. In Section 2, we present the ways to generalize and enhance $(PC)^2A$. In Section 3, we report our experiments. Finally in Section 4, we conclude.

2. Enhanced $(PC)^2A$

2.1 $E(PC)^2A1$ and $E(PC)^2A2$

In $(PC)^2A$, the original image $I(x, y)$ is combined with its first-order projection to derive a new version of the original image. It was demonstrated that such a combination is helpful to subsequent recognition process. Therefore, a natural extension of $(PC)^2A$ is to exploit some higher-order projection to enhance the recognition process .

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

$$P_2(x,y) = \frac{V_2(x)H_2(y)}{\bar{J}} \qquad (1)$$

where $\bar{J}$ is the mean value of $J(x,y)$ which is defined as the square of $I(x,y)$, i.e., $J(x,y) = I(x,y)^2$, and V_2 and H_2 are defined respectively as

$$V_2(x) = \frac{1}{N_2}\sum_{y=1}^{N_2} J(x,y) \qquad (2)$$

$$H_2(y) = \frac{1}{N_1}\sum_{x=1}^{N_1} J(x,y) \qquad (3)$$

It is also possible to derive another version of second-order projection of $I(x, y)$ as

$$P_2'(x,y) = \frac{V_2'(x)H_2'(y)}{\bar{J}} \qquad (4)$$

where V_2' and H_2' are defined as Eqs.(5) and (6) respectively.

$$V_2'(x) = \frac{1}{C_{N_2}^2}\left(\sum_{y=1}^{N_2} I(x,y)\right)^2 \qquad (5)$$

$$H_2'(y) = \frac{1}{C_{N_1}^2}\left(\sum_{x=1}^{N_1} I(x,y)\right)^2 \qquad (6)$$

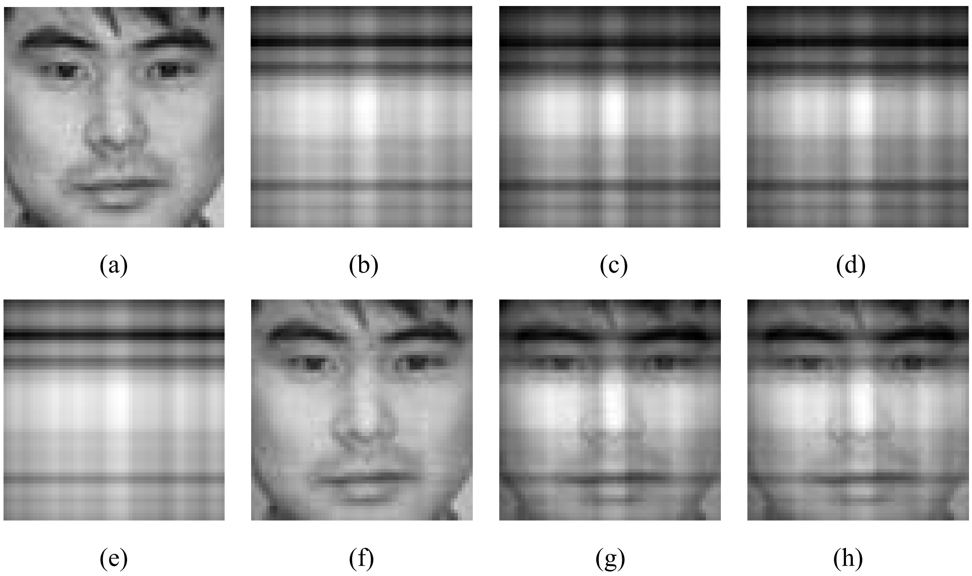
where C_N^2 is a combinatorial number, whose value is *N(N-1)/2*, which denotes the number of terms in the brackets in (5) or (6). The difference between $P_2(x,y)$ and $P_2'(x,y)$ is that the latter introduces mutual correlations between pixels as shown in Eq. (5) and (6).

In (Wu and Zhou, 2002), the first-order projected image P_1 is defined similarly as in Eq. (1), which was obtained by using the original image $I(x,y)$ to replace $J(x,y)$ in Eqs. (1)-(3). Then, through combining the original image with its first and second-order projection, we obtain two new projection-combined images, i.e. $I_2(x, y)$ and $I'_2(x, y)$, as shown in Eqs.(7) and (8), respectively, where α and β are parameters to control the bias of the projections.

$$I_2(x,y) = \frac{I(x,y) + \alpha P_1(x,y) + \beta P_2(x,y)}{1+\alpha+\beta} \qquad (7)$$

$$I_2'(x,y) = \frac{I(x,y) + \alpha P_1(x,y) + \beta P_2'(x,y)}{1 + \alpha + \beta} \qquad (8)$$

Fig. 1 shows an example of the original image, its projections and the derived projection-combined images, where α and β are set to 0.25 and 1.5 respectively. Since the pixels of the projections and the projection-combined images may fall out of [0 1], these images are normalized into [0 1] for better display.



a) original face image b) 1-order projection, c) 2-order projection generated by Eq.1, d) 2-order projection generated by Eq.4, e) 1/2 order projection, f) 1-order projection-combined image, g) 2-order projection-combined image generated by Eq.7, h) 2-order projection-combined image generated by Eq.8

Fig. 1 Example of an original image, its projections, and the projection-combined images

It is not difficult to derive that the projections P_1, P_2 and P_2', and the projection-combined images I_2 and I_2' have the following properties:

1) As α and β approach 0, I_2 and I_2' turn out to be exactly as the original image I .

2) As α approaches infinity and β approaches 0, I_2 and I_2' approach P_1 .

3) As α approaches 0 and β approaches infinity, I_2 and I_2' approach P_2 and P_2', respectively.

4) P_1, P_2 and P_2' have the intrinsic dimensionality (Bishop, 1995), or rank, not more than 1.

Properties 1)-3) are apparent. To prove Property 4), it is helpful to represent projections P_1, P_2 and P_2' in matrix form. For example, P_2 can be rewritten as $P_2 = \frac{V_2 H_2}{\bar{J}}$, where V_2 is a N_1-dimensional column vector and H_2 a N_2-dimensional row vector. Thus, P_1 is a $N_1 \times N_2$ matrix and is exactly an out-product of vectors V_2 and H_2^T, where H_2^T is the transpose of H_2. From the matrix theory, the rank of a matrix generated by two vector outer-product is at most 1. Thus, we get

$$rank(P_2) = rank(V_2 H_2) = rank(H_2 V_2) \leq 1 . \qquad (9)$$

Similarly, we can prove that both the ranks of P_1 and P_2' are at most 1.

Such a property shows that the main information of the original image is still kept after combining it with its projections. Moreover, since the projections blur the original image to some extent, it can be anticipated that some important features for recognition may become more salient.

In principle, the third, the fourth and even more higher-order projections can be combined with the original image. However, such a process will introduce too many control parameters to be adjusted, which will significantly increase the complexity of the method. Therefore, we choose to use only the first- and second-order projections.

After deriving the new image $I_2(x, y)$ or $I_2'(x, y)$ via Eq.(7) or Eq.(8), PCA can be performed to build the corresponding eigenspaces for subsequent recognition process. Here we call the method applying PCA on $I_2(x, y)$ as E(PC)^{2}A1 and that on $I_2'(x, y)$ as E(PC)^{2}A2. From the above discussions, both E(PC)^{2}A1 and E(PC)^{2}A2 turn into PCA as α and β both approach 0, and (PC)^{2}A as only β approaches 0.

2.2 (PC)^{2}A+, E(PC)^{2}A1+ and E(PC)^{2}A2+

In fact, the main difficulty of face recognition with one training image per person lies in that since only one training image is available for each person, the intra-class variation can hardly be considered against the inter-class variation. Since the projections derived for each image can be regarded as a new image for a specific person, it is interesting to see that whether these projected images can act as additional training images for the person. Therefore, the problem of face recognition with one training image becomes a common face recognition problem where each person has several training images.

Let's generalize Eq. (1) to a universal form

$$P_n(x,y) = \frac{V_n(x)H_n(y)}{\bar{J}} \qquad (10)$$

where $\bar{J}$ is the mean value of $J(x,y)$ which is defined as $J(x,y) = I(x,y)^n$, and V_n and H_n are defined respectively as

$$V_n(x) = \frac{1}{N_2}\sum_{y=1}^{N_2} J(x,y) \qquad (11)$$

$$H_n(y) = \frac{1}{N_1}\sum_{x=1}^{N_1} J(x,y) \qquad (12)$$

Note that *n* can be integers as well as fractions such as 1/2, 1/3, etc. Then, assume that we have obtained a series of projected images for each original face image. Without loss of generality, assume that there are *n* projected images for each person. Therefore, together with the original image, each person has (*n*+1) images for training, and PCA can be performed on (*n*+1)*M* training images, where *M* is the number of persons to be recognized.

Thus, we develop another three extensions to $(PC)^2A$. The first one, i.e. $(PC)^2A+$, applies PCA on the original face images and their first-order projected images. The second one, i.e. $E(PC)^2A1+$, applies PCA on the original face images, their 1/2-order and first-order projected images, and their second-order projected images derived according to Eq. (1). Finally the third one, i.e. $E(PC)^2A2+$, applies PCA on the original face images, their 1/2 order and first-order projected images, and second-order projected images derived according to Eq. (4). Note that for simplifying our following discussion, we only use 1/2-, first- and second-order projected images, but projected images with other orders can also be used in principle.

3. Experiments

3.1 Data Set

In our experiments, the new methods presented in Section 2 are compared with both $(PC)^2A$ and the standard eigenface technique. The experimental configuration is similar as that was described in (Wu and Zhou, 2002). The experimental face 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 person has two images (**fa** and **fb**) with different facial expressions. The **fa** images are used as gallery for training while the **fb** images as probes for testing. All the images are randomly selected from the FERET face database (Phillips et al., 1998). No special criterion is set forth for the selection. So, the face images used in the experiments are very diversified, e.g. there are faces with different race, different gender, different age, different expression, different illumination, different occlusion, different scale, etc., which greatly increases the difficulty of the recognition task. See (Wu and Zhou, 2002) for some concrete face samples.

Before the recognition process, the raw images are normalized according to some constraints so that the face area 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 in our experiments, the eyes are manually located, the cropped image size is 60x60 pixels and the inter-ocular distance is 28 pixels.

3.2 Results on $E(PC)^2A1$ and $E(PC)^2A2$

At first, we compare the recognition performance of the methods proposed in Section 2.1 with that of $(PC)^2A$ and the standard eigenface technique when the size of the face database increases gradually form 20 to 200 with 20 as the interval. When a probe, i.e., an unknown face image, is presented, its corresponding feature vector is constructed from the eigenfaces. Then the distance between the probe's feature vector and that of the gallery images are computed, and the k best-matched image (with the minimum distance) in the gallery is considered as the *top k match* result.

The *top 1 match* result is depicted in the left part of table 1. Here the number of eigenvectors or eigenfaces, i.e. d, is controlled by setting a threshold as follows

$$\frac{\sum_{i=1}^{d} \lambda_i}{\sum_{i=1}^{m} \lambda_i} \geq \theta \qquad (13)$$

where $\lambda_1, \lambda_2, \ldots, \lambda_m$ are the *m* biggest eigenvalues and θ is set to 0.95. α is set to 0.25, and β is set to 1.5. These values will be used in the rest of this paper if no specific value is explicitly stated.

Table 1 Comparison of recognition accuracies at different size of database

size of database	Eigenface	$(PC)^2A$	$E(PC)^2A1$	$E(PC)^2A2$	$(PC)^2A+$	$E(PC)^2A1+$	$E(PC)^2A2+$
20	0.9500	0.9500	0.9500	0.9000	0.9500	0.9500	0.9000
40	0.8500	0.8750	0.8750	0.9250	0.8500	0.8750	0.9250
60	0.8500	0.8667	0.8833	0.9333	0.8667	0.8833	0.9000
80	0.8250	0.8250	0.8250	0.8875	0.8250	0.8500	0.8500
100	0.7600	0.7500	0.7700	0.8400	0.7600	0.8000	0.7900
120	0.7750	0.7750	0.8167	0.8250	0.7833	0.8000	0.8167
140	0.7929	0.8000	0.8214	0.8143	0.8071	0.8214	0.8143
160	0.8187	0.8187	0.8313	0.8250	0.8250	0.8438	0.8187
180	0.8167	0.8278	0.8389	0.8222	0.8278	0.8556	0.8500
200	0.8300	0.8350	0.8450	0.8400	0.8400	0.8550	0.8400

Table 1 reveals that for nearly all cases, both $E(PC)^2A1$ and $E(PC)^2A2$ achieve higher recognition accuracy than $(PC)^2A$ and the standard eigenface technique. It also shows that on databases of relatively small size (about 30 to 130), $E(PC)^2A2$ obtains the best performance among all the compared methods, while for databases of relatively large size (about bigger than 130), $E(PC)^2A1$ is the best one. In average, the recognition accuracy of the standard eigenface technique, $(PC)^2A$, $E(PC)^2A1$ and $E(PC)^2A2$ under different size of databases is 82.68%, 83.23%, 84.57% and 86.12%, respectively. In other words, the recognition accuracy of $E(PC)^2A1$ and $E(PC)^2A2$ is about 1.6% and 3.5% higher than that of $(PC)^2A$, respectively.

Although $(PC)^2A$ can achieve better recognition accuracy than the standard eigenface technique, its biggest strength is that it can use significantly fewer (about 10-15% (Wu and Zhou, 2002)) eigenfaces to achieve similar performance of the standard eigenface technique. Therefore, it is interesting to compare the number of eigenfaces used by $E(PC)^2A1$ and $E(PC)^2A2$. Part of Fig. 2 shows the comparison results. It is impressive that the number of eigenfaces used by $E(PC)^2A1$ and $E(PC)^2A2$ is even far fewer than that used by $(PC)^2A$, and the difference is more and more distinct as the size of database increases. In average, the number of eigenfaces used under different size of database by the standard eigenface technique, $(PC)^2A$, $E(PC)^2A1$ and $E(PC)^2A2$ are 47.8, 41.5, 21.8, and 14.6, respectively. In other words, $E(PC)^2A1$ and $E(PC)^2A2$ use about 47.5% and 64.8% fewer eigenfaces than $(PC)^2A$. Recall that the number of eigenfaces used determines the dimensionality of the feature vectors that are extracted for representing the face images. So, it is obvious that using fewer eigenfaces means that less computational cost, less storage cost, and less matching time are required, which is of great benefit for large-size face databases in real-world tasks.

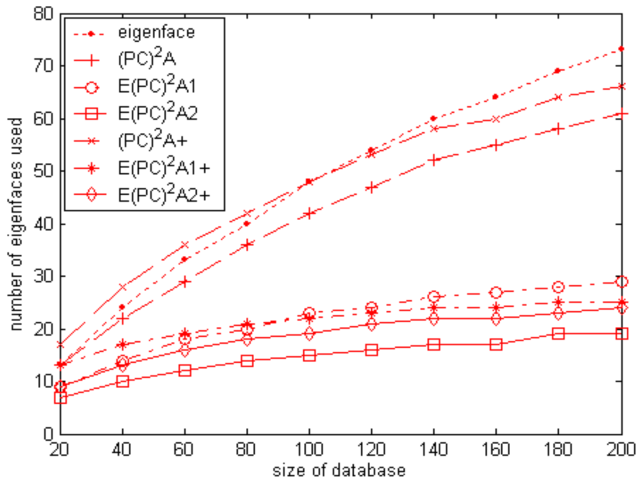


Fig. 2 Comparison of number of eigenfaces used corresponding to recognition accuracies in table 1

In the above experiments, we fix the number of eigenvalues by setting θ in Eq. (13) to a constant, i.e. 0.95. Fig. 3 depicts the recognition accuracies of the standard eigenface technique, $(PC)^2A$, $E(PC)^2A1$ and $E(PC)^2A2$ respectively under different number of eigenfaces. Clearly, both $E(PC)^2A1$ and $E(PC)^2A2$ achieve higher recognition accuracy than $(PC)^2A$ and the standard eigenface technique again.

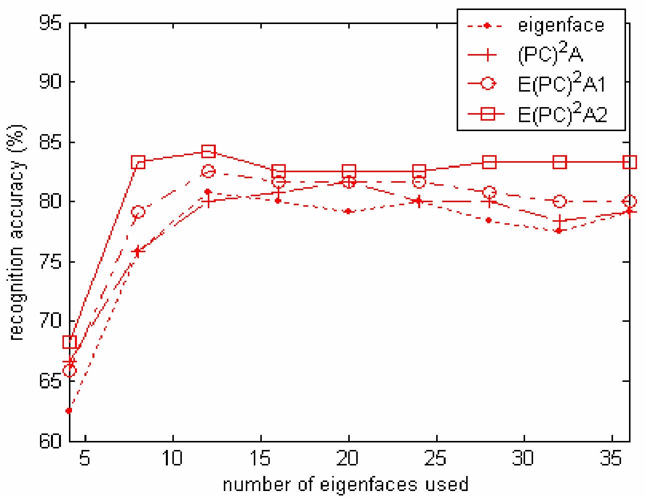


Fig. 3 Comparison of recognition accuracies under different number of eigenfaces used

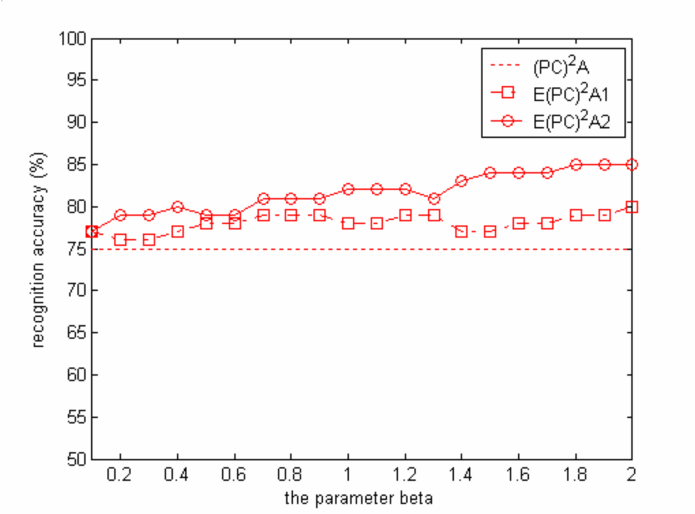


Fig. 4 Recognition accuracy under different values of β

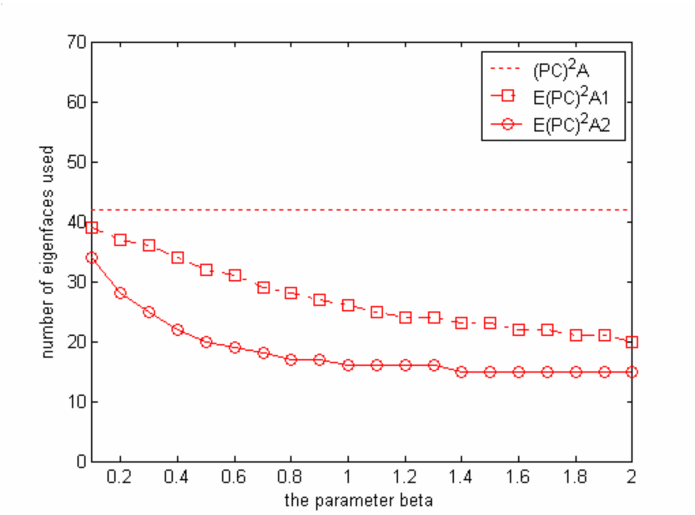


Fig. 5 Number of eigenfaces used under different values of β corresponding to recognition accuracy in Fig. 4

In $(PC)^2A$, there is a combination parameter α. In (Wu and Zhou, 2002), α is set to 0.25 and it has been shown that this value helps reach a good trade-off between the recognition accuracy and the number of eigenfaces used. In $E(PC)^2A1$ and $E(PC)^2A2$, however, besides α, there is another parameter β. In order to know the influence of β, more experiments are performed on $E(PC)^2A1$ and $E(PC)^2A2$ with different values of β. Fig. 4 shows the top 1 match recognition accuracy, and Fig. 5 shows the corresponding number of eigenfaces used by the methods. The size of database used in the experiments is 100. For comparison, the performance of $(PC)^2A$ is also depicted as a baseline. Fig. 4 shows that as β gradually increases, there are a series of fluctuations of recognition accuracy for both $E(PC)^2A1$ and $E(PC)^2A2$, but no matter how to choose β, the recognition of both $E(PC)^2A1$ and $E(PC)^2A2$ always outperform $(PC)^2A$. On the other hand, as β gradually increases, the number of eigenfaces used in $E(PC)^2A1$ and

$E(PC)^2A2$ first decreased greatly and approaches to be steady after 1.5. Recall that fewer eigenfaces means less computational cost, we suggest setting the value of β as 1.5.

3.3 Results on $(PC)^2A+$, $E(PC)^2A1+$ and $E(PC)^2A2+$

Then, we compare the recognition performance of the methods proposed in Section 2.2 with that of $(PC)^2A$ and the standard eigenface technique. The experimental methodology is the same as that described in Section 3.2.

The right part of table 1 and Fig. 2 depict the recognition accuracy and the corresponding number of eigenfaces used, which shows that the performance of $(PC)^2A+$ is very comparable to that of $(PC)^2A$. Moreover, from table 1 and Fig. 2, it can be found that the performance of $E(PC)^2A1+$ and $E(PC)^2A2+$ are comparable to that of $E(PC)^2A1$ and $E(PC)^2A2$. In fact, the average recognition accuracy of $(PC)^2A+$, $E(PC)^2A1+$ and $E(PC)^2A2+$ under different sizes of database are 83.35%, 85.34% and 85.05%, while the averaged numbers of eigenfaces used are 47.2, 21.3 and 18.7, respectively. The results are comparable to those in Section 3.2, while no extra parameters are needed in the algorithms.

These results demonstrate that the methods proposed in Section 2.2, especially $E(PC)^2A1+$ and $E(PC)^2A2+$, can also improve the performance of $(PC)^2A$. This supports our claim that through employing the projections of the original images to enlarge the training face database, the problem of face recognition with one training image per person can be transformed to be a common face recognition problem to solve. In the above experiments, we just use 1/2, first- and second- order projections. It is natural to think about whether the recognition performance can be further improved through employing more projected images. Intuitively, we guess there would be a *critical point* for the number of projected images, that is, if the number of projected images used beyond the critical point, the recognition performance might gradually decrease. To verify this idea, more experiments are performed.

Fig. 6 shows the average recognition accuracy as the number of projected images per person increases gradually. Fig. 7 shows the average number of eigenfaces used as the number of projected images per person increases gradually. Here we use two methods to choose the projected images. In method 1, as the number of projected images per person increases, the images selected are like I, P_1, P_2, $P_{1/2}$, P_3, $P_{1/3}$… While in method 2, the images are selected like I, P_1, $P_{1/2}$, P_2, $P_{1/3}$, P_3… Here I is the original face image, and P_n is the *n*-order projected image. According to Fig. 6 and Fig. 7, we find as the numbers of faces per person exceed 4, although there may be some fluctuations, the recognition accuracy tends to decrease and the number of the extracted eigenfaces tends to increase. So, we guess the critical point may locate at somewhere near the point 4. However, the accurate critical point can only be determined by more experiments and rigorous theoretical justifications. We believe that such a phenomenon can be explained from the fact that as the order increases, $(I(x,y))^n$ will be more and more close to zero, while $(I(x,y))^{1/n}$ to 1, which means their uselessness gradually increases. In addition, for high order projections, the numerical computations get noisier and less reliable. It is obvious that such a statement should also undergo rigorous theoretical justification in future work.

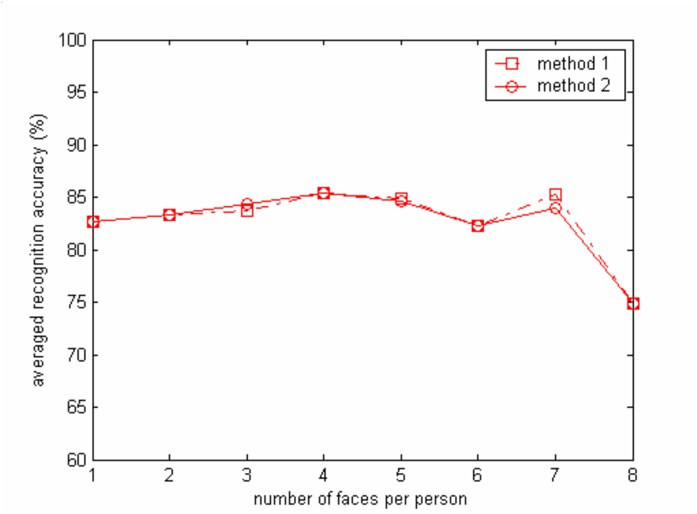


Fig. 6 Averaged recognition accuracy using different numbers of faces per person

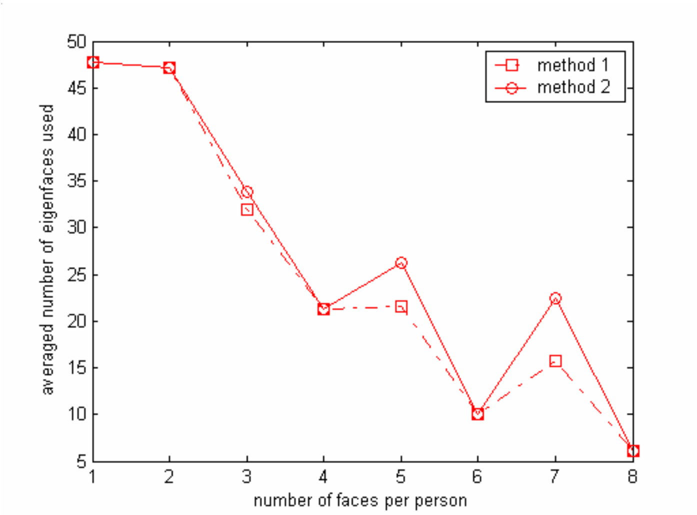


Fig. 7 Averaged number of eigenfaces with different numbers of faces per person corresponding to averaged recognition accuracies in Fig. 6.

4. Conclusions

Most face recognition techniques require that there exist at least two training images per person. Recently, a method called (PC)^{2}A was proposed to address the issue of face recognition with one training image per person. In this paper, two directions for generalizing and enhancing (PC)^{2}A are identified and several new algorithms are proposed. These algorithms utilize second-order projections besides the first-order projection used by (PC)^{2}A. Experiments show that algorithms proposed from the aspect of both directions can significantly improve the performance of (PC)^{2}A, that is, achieving higher recognition accuracy with far fewer eigenfaces.

Moreover, the second direction we proposed to improve (PC)^{2}A is a general paradigm for dealing with 'small sample problem', in which we enlarge the original image database by appending the *n*-order projected images. This paper shows that this paradigm works well in the scenario of face recognition with one training image per person. However, we believe that this method is also effective in scenarios where each person has two (or more, but still 'small sample') training images, which is another interesting issue for future work, besides the issues raised at the end of Section 2.2.

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, and the *Qinglan* project foundation of Jiangsu province. Portions of the research in this paper use the FERET database of facial images collected under the FERET program.

References

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

Bishop, C. M. Neural Networks for Pattern Recognition, 1995.Oxford University Press, New York.

Brunelli, R. and Poggio, T., 1993. Face recognition: Features versus templates. IEEE Trans. on Pattern Analysis and Machine Intelligence 15(10), 1042-1062.

Chellappa, R., Wilson, C.L.,and Sirohey, S., 1995. Human and machine recognition of faces: a survey. Proceedings of the IEEE 83(5), 705-740.

Chen, X., Huang, T., 2003. Facial expression recognition: a clustering-based approach. Pattern Recognition Letters 24(9-10), 1295-1302.

Etemad, K., Chellappa, R., 1997. Discriminant analysis for recognition of human face images. Journal of the Optical Society of America A: Optics Image Science and Vision 14(8), 1724-1733.

Lawrence, S.,Giles, C.L., Tsoi, A. and Back, A., 1997. Face recognition: A convolutional neural-network approach. IEEE Trans. on Neural Networks, 8(1), 98-113.

Lu, J., Plataniotis, K.N.and Venetsanopoulos, A.N., 2003. Face Recognition using kernel direct discriminant analysis algorithms. IEEE Trans. on Neural Networks 14(1), 117-126.

Martinez, A.M., 2002. Recognition imprecisely localized, partially occluded, and expression variant faces from a single sample per class, IEEE Trans. on Pattern Analysis and Machine Intelligence 25(6), 748-763.

Moghaddam, B., Pentland, A., 1997. Probabilistic visual learning for object representation. IEEE Trans. on Pattern Analysis and Machine Intelligence 19(7), 696-710.

Moghaddam, B., Jebara, T. and Pentland, A., 2000. Bayesian face recognition. Pattern Recognition 33(11), 1771-1782.

Pang, S., Kim, D., Bang, S.Y., 2003, Membership authentication in the dynamic group by face classification using SVM ensemble. Pattern Recognition Letters 24(1-3), 215-225.

Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.J., 1998. The FERET database and evaluation procedure for face-recognition algorithms. Image and Vision Computing, 16(5), 295-306.

Raytchev, B., Murase, H., 2003. Unsupervised face recognition by associative chaining. Pattern Recognition 36(1), 245-257.

Sukthankar, G., 2000. Face recognition: a critical look at biologically-inspired approaches. Technical Report: CMU-RI-TR-00-04, Carnegie Mellon University, Pittsburgh, PA.

Swets, D.L., Weng., J., 1996. Using discriminant eigenfeatures for image retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence 18(8), 831-836.

Turk, M. A. and Pentland, A. P., 1991. Face Recognition Using Eigenfaces. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 586-591.

Valentin, D., Abdi, H., O'toole, A.J. and Cottrell, G.W., 1994. Connectionist models of face processing: a survey. Pattern Recognition 27(9), 1209-1230.

Wiskott, L., Fellous, J.-M., Kruger, N. and Malsburg, C. von der, 1997. Face recognition by elastic bunch graph matching. IEEE Trans. on Pattern Analysis and Machine Intelligence 19 (7), 775-779.

Wu, J. and Zhou, Z.-H.,2002. Face Recognition with one training image per person. Pattern Recognition Letters 23(14), 1711-1719.

Zhao, W., Chellappa, R.,Rosenfeld, A. and Phillips, P.J., 2000. Face recognition: a literature survey. Technical Report: UMD CfAR-TR-948, University of Maryland, College Park, MD, 2000.

Zhang, J., Yan, Y. and Lades, M., 1997. Face recognition: Eigenfaces, elastic matching, and neural nets. Proceedings of the IEEE 85(9), 1422-1435.