Face Recognition with One Training Image per Person

Jianxin Wu and Zhi-Hua Zhou*

National Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, P.R.China

Abstract

At present there are many methods that could deal well with frontal view face recognition. However, most of them cannot work well when there is only one training image per person. In this paper, an extension of the eigenface technique, i.e. $(PC)^2A$, is proposed. $(PC)^2A$ combines the original face image with its horizontal and vertical projections and then performs principal component analysis on the enriched version of the image. It requires less computational cost than the standard eigenface technique and experimental results show that on a gray-level frontal view face database where each person has only one training image, $(PC)^2A$ achieves 3%-5% higher accuracy than the standard eigenface technique through using 10%-15% fewer eigenfaces.

Keywords: Face recognition; Face identification; Principal component analysis; Eigenface; Pattern recognition

1. Introduction

In the past two decades, face recognition has been an active research area of computer vision and pattern recognition. According to Brunelli and Poggio [3], face recognition techniques could be categorized into two classes, i.e. geometric feature-based techniques and template-based techniques. Although many new techniques have been developed after the publication of this taxonomy, most of them still can be categorized into those two classes. In the first category, the most often used techniques include elastic bunch graph matching [20] and active appearance model [5]. In the second category, the most often used technique is eigenface [18]. Besides, neural networks are extensively used in this area [19, 21]. Good surveys on human and machine recognition of faces can be found in Chellappa *et al.* [4], Fromherz [7], and Zhao *et al.* [22].

Template-based techniques often follow the subspace method called eigenface originated by Turk and Pentland [18]. This technique is based on the Karhunen-Loève transformation, which is also referred to as PCA, i.e. Principal Component Analysis, and was introduced into face processing by Kirby and Sirovich [11]. It has gained great success and become a *de facto* standard and a common performance benchmark in face recognition [12]. Eigenface is a simple subspace method. At present there are many techniques follow the idea behind it and try to find more effective subspaces, such as LDA, i.e. Linear Discriminant Analysis [6], discriminant eigenfeatures [17], fisherface [1], view-based and modular eigenspace [14], probabilistic visual learning [13], and Bayesian face recognition [12].

Although those subspace techniques can achieve fast recognition speed and high recognition rate, they do have some defects [16]. In the standard eigenface technique [18], face images are represented as points in a high dimensional space. The PCA projection efficiently spreads the points but it does not consider how the points are assigned to different classes, which may cause serious consequences in some cases. Those techniques often deal with this problem by taking class labels of the points into consideration, therefore give better recognition results.

* Corresponding author. Tel.: +86-25-359-3163; fax: +86-25-330-0710.
E-mail addresses: wujx@ai.nju.edu.cn (J. Wu), zhouzh@nju.edu.cn (Z.-H. Zhou).

But in order to obtain good recognition performance, most of them require that there are at least two training images per person so that the intra-class variation could be considered against the inter-class variation. Unfortunately, in real-world tasks such a requirement can not always be satisfied. For example, in the law enforcement scenarios, only one image per person may be available for training. Therefore, difficulties may arise, e.g. parameter estimation may be very difficult (if not impossible) and unreliable, and the small sample size problem may occur [8].

In this paper, an extension of the standard eigenface technique called $(PC)^2A$, i.e. Projection-Combined Principal Component Analysis, is proposed. This technique requires less computational cost than the standard eigenface technique. Experiments on a gray-level frontal view face database show that, when only one training image per person is available, $(PC)^2A$ is superior to the standard eigenface technique.

The rest of this paper is organized as follows. In Section 2, we briefly introduce the eigenface technique. In Section 3, we propose $(PC)^2A$. In Section 4, we report on the experimental results. In Section 5, we explore the properties of $(PC)^2A$ with further experiments. Finally in Section 6, we summarize the contributions of this paper and raise several issues for future works.

2. Eigenface

Eigenface technique regards each face image as a feature vector in a high dimensional space by concatenating the rows of the image and using the intensity of each pixel as a single feature. Thus each image can be represented as an n-dimensional random vector $\mathbf{x}$. The dimensionality n may be very large, on the order of several thousands [17]. The main function of PCA is dimensionality reduction, that is, mapping the n-dimensional vector $\mathbf{x}$ into an m-dimensional space, where $m << n$.

The vector $\mathbf{x}$ can be approximated as a linear combination of a set of orthonormal vectors, i.e. $\mathbf{u}_i$ (i = 1, 2, …, m), as expressed in Eq. (1).

$$\hat{\mathbf{x}} = \sum_{i=1}^{m} \mathbf{z}_i \mathbf{u}_i \tag{1}$$

The best approximation is defined as the one that minimizes the difference between $\hat{\mathbf{x}}$ and $\mathbf{x}$, which is often measured by the mean squared error. Let $\Sigma_{\mathbf{x}}$ be the covariance matrix of $\mathbf{x}$, i.e.

$$\Sigma_{\mathbf{x}} = E\left[(\mathbf{x}-\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})^{\mathbf{T}}\right] \tag{2}$$

where $\overline{\mathbf{x}}$ is the mean vector of $\mathbf{x}$. It has been proved [2] that the best $\mathbf{u}_i$ (i = 1, 2, …, m) are the unit eigenvectors associated with the m largest eigenvalues of $\Sigma_{\mathbf{x}}$. Those unit eigenvectors are called *eigenfaces* [18] in face recognition, and the corresponding m-dimensional vector $\mathbf{z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m]$ is treated as features extracted by PCA, which is used in later recognition process. Since m is usually far smaller than n, the dimensionality is greatly reduced so that the difficulty of the problem usually decreases. Note that in real-world applications, $\Sigma_{\mathbf{x}}$ is often substituted by the sample scatter matrix.

In order to determine the proper value for m, the eigenvalues of $\Sigma_{\mathbf{x}}$ is ranked in descending order. Suppose there are n eigenvalues, i.e. $\lambda_1, \lambda_2, \ldots, \lambda_n$, where λ_1 is the largest eigenvalue while λ_n is the smallest one. Then m is determined as the first integer that satisfies Eq. (3), where θ is a pre-set threshold [2].

$$\frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \geq \theta \tag{3}$$

It has been proved [2] that following the selection criterion corresponding to Eq. (3), the m largest eigenvectors can capture $\theta \times 100\%$ of the whole variation of $\mathbf{x}$.

The value of θ is often set to be larger than 0.9 so that most of the variation of the input data can be captured. In this sense, features extracted by PCA are '*the most expressive features*'. However, this does not necessarily mean that they are '*the most discriminating features*' [16, 17]. As mentioned in Section 1, although the PCA projection efficiently spreads the points, it does not take into account how the points are assigned to different classes. A possible solution to this problem is to utilize information in the class labels. However, such a supervised learning technique may not work well when there is only one training image per person.

3. (PC)^{2}A

As stated by Jain *et al.* [9], the training phase in a statistical pattern recognition system can be divided into three successive stages, i.e. the preprocessing stage, the feature extraction/selection stage, and the learning stage. At present, most of the extensions of the eigenface technique focus on the feature extraction/selection stage [1, 6, 17] or the learning stage [12, 13, 14]. There is no specific preprocessing except for a few standard image processing techniques such as histogram equalization. Here we propose an extension of eigenface technique, i.e. (PC)^{2}A, whose key is a novel preprocessing mechanism specially designed for face recognition purposes.

Let $P(x, y)$ be an intensity image of size $N_1 \times N_2$, where $x \in [1, N_1]$, $y \in [1, N_2]$, and $P(x, y) \in [0, 1]$. The vertical and horizontal integral projections are defined respectively as:

$$V_P(x) = \sum_{y=1}^{N_2} P(x, y) \tag{4}$$

$$H_P(y) = \sum_{x=1}^{N_1} P(x, y) \tag{5}$$

Now we define the *projection map* $M_P(x, y)$ of $P(x, y)$ as:

$$M_P(x, y) = \frac{V_P(x) H_P(y)}{N_1 N_2 \overline{P}} \tag{6}$$

where $\overline{P}$ is the average intensity of the image, i.e.

$$\overline{P} = \frac{\sum_{x=1}^{N_1} \sum_{y=1}^{N_2} P(x, y)}{N_1 N_2} \tag{7}$$

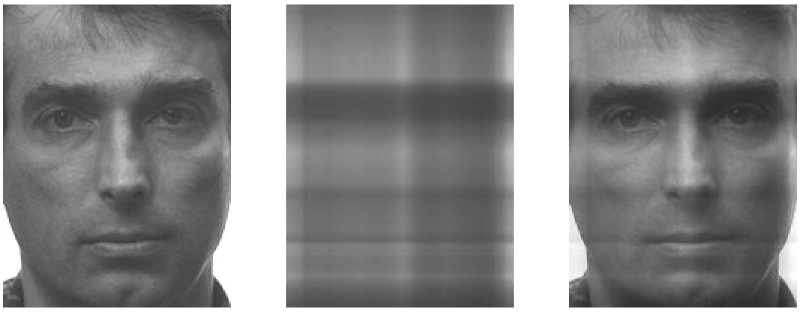
Then, we define the projection-combined version of $P(x, y)$ as Eq. (8), where α is called *combination parameter*.

$$P_\alpha(x, y) = \frac{P(x, y) + \alpha M_P(x, y)}{1 + \alpha} \tag{8}$$

See Fig.1 for an example of the projection map and the projection-combined version of an image, where α is set to 0.25. Note that since $P_\alpha(x, y)$ may fall out of [0, 1], the display of the image may be distorted although the recognition result will not be affected. For better display, the rightest image in Fig.1 has been adjusted according to:

$$P_\alpha'(x, y) = \frac{P_\alpha(x, y) - \min\left(P_\alpha(x, y)\right)}{\max\left(P_\alpha(x, y)\right) - \min\left(P_\alpha(x, y)\right)} \tag{9}$$

In (PC)^{2}A, PCA is performed on the projection-combined version of the image, i.e. $P_\alpha(x, y)$, instead of on the original face image, i.e. $P(x, y)$. This is the exact reason that we call the proposed method as projection-



a) original face image b) projection map c) projection-combined image

Fig.1 Example of the projection map and the projection-combined image

combined principal component analysis.

It is easy to derive that $P(x, y)$, $M_P(x, y)$, and $P_\alpha(x, y)$ have the following properties:

1) The average intensity of $P(x, y)$ is equal to the average intensity of $M_P(x, y)$.
2) So long as $\alpha \neq -1$, the average intensity of $P(x, y)$ is equal to the average intensity of $P_\alpha(x, y)$.
3) As α approaches 0, $P_\alpha(x, y)$ turns out to be exactly $P(x, y)$.
4) As α approaches infinity, $P_\alpha(x, y)$ approaches $M_P(x, y)$.
5) The *intrinsic dimensionality* [2] of $P(x, y)$ is less than $(N_1 + N_2 - 1)$.
6) The intrinsic dimensionality of $M_P(x, y)$ is much smaller than that of $P(x, y)$.

Note that $M_P(x, y)$ is determined by the average intensity of the column and row that include the point (x, y). So, as long as α is not too large, integrating $M_P(x, y)$ into $P(x, y)$ blurs the original image while keeping its main information. Such a blurred version, i.e. $P_\alpha(x, y)$, could be expected to work better than the standard eigenface technique against minor changes of expression, illumination, occlusion, *etc*.

4. Experiments

4.1. Configuration

In our experiments, $(PC)^2A$ is compared against the standard eigenface technique on a gray-level frontal view face database that comprises 400 images from 200 persons. There are 71 females and 129 males, each of whom 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 [15]. No special criterion is set forth for the selection. So, the face images used in our 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. Fig.2 shows some raw images in the database.

Both the standard eigenface technique and $(PC)^2A$ represent an image as an m-dimensional feature vector, where the value of m depends on the value of θ in Eq. (3). In our experiments θ is set to 0.9 if it is not explicitly stated. Note that $(PC)^2A$ has another parameter, i.e. the combination parameter α. In our experiments α is set to 0.25.

It is worth mention that the face images are normalized before they are presented to subspace techniques. The normalized version of a face image satisfies some constraints so that the 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. Note that accurately locating eyes is critical to the normalization. Here in our experiments, the eyes are manually located

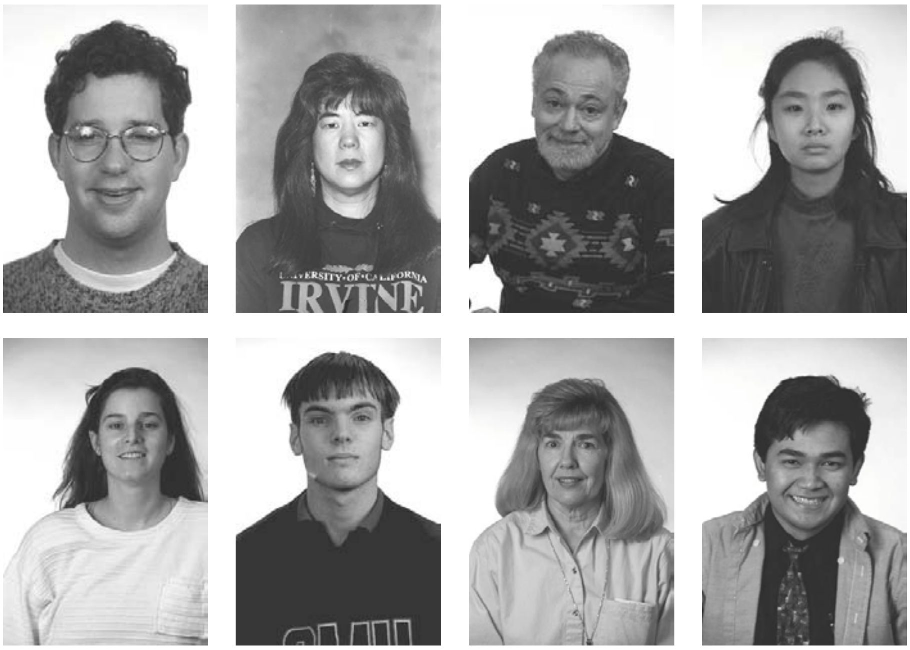


Fig.2 Some raw images in the experimental database

and then the raw images, whose size is 256 × 384 pixels and the inter-ocular distances range from 40 to 60 pixels, are rotated and resized using the bilinear method. The cropped face images are in the size of 60 × 60 pixels and their inter-ocular distance is 28 pixels.

When a probe is presented, its corresponding feature vector is constructed from the eigenfaces. Then the angles between the probe's feature vector and that of every image in the gallery are computed. After that, the images in the gallery are ranked according to the ascending order of the angles, and the identity of the top image in the list is considered as the recognition result. Note that conforming to the FERET testing protocol [15], we consider not only '*is the top match correct?*' but also '*is the correct answer in the top k matches?*'.

4.2. Results

At first, we compare the recognition performance of the standard eigenface technique and that of $(PC)^2A$ when the size of the face database increases gradually from 20 to 200 with 20 as the interval. The results are shown in Fig.3 where the standard eigenface technique and $(PC)^2A$ are labeled as "*standard*" and "*extension*" respectively, and the number of eigenfaces they used is determined by Eq.(3). Note that Fig.3 shows not only the *top 1 match* rate but also the *top 3 match* rate.

Fig.3 reveals that $(PC)^2A$ is superior to the standard eigenface technique when the top 1 match rate is concerned. Its recognition accuracy, i.e. the correct recognition rate, is about 3% to 5% higher than that of the standard eigenface technique. Moreover, Fig.3 reveals that $(PC)^2A$ also outperforms the standard eigenface technique when the top 3 match rate is concerned. In fact, when observing the performance of the two compared methods in top k match rates, we find that for almost all values of k, the performance of $(PC)^2A$ is better than or at least equally as well as that of the standard eigenface technique. The reason for why Fig.3 only displays the results of $k = 3$ is that as k grows larger, especially when $k > 10$, the difference between the performance of those two methods shrinks to undistinguishable level. Moreover, in real-world tasks, k is usually a small number.

In order to further analyze the results depicted in Fig.3, we compare the number of eigenfaces used by the standard eigenface technique and $(PC)^2A$ when the size of the face database increases gradually from 20 to 200 with 20 as the interval. The results are shown in Fig.4.

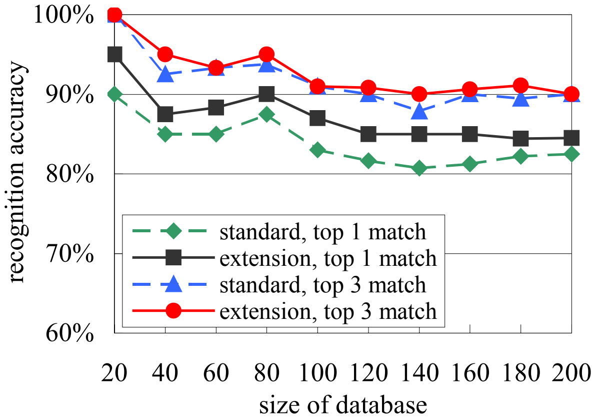


Fig.3 Comparison of the recognition performance

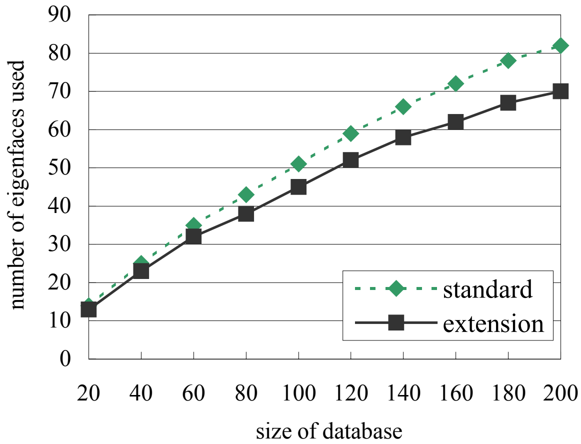


Fig.4 Comparison of the number of eigenfaces used

Fig.4 reveals that for almost all sizes of the database, $(PC)^2A$ uses 10% to 15% fewer eigenfaces than the standard eigenface technique. 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 processing time are required. In real-world tasks, the face database may contains tens of thousands or more images. Therefore the saving of the computational cost, storage cost, and processing time may be very profitable.

Then, we compare the recognition performance of those two techniques using the same number of eigenfaces on the database containing 200 images. Now the number of eigenfaces used is not determined by Eq.(3). Instead, it increases gradually from 10 to 100 with 10 as the interval. The results are plotted in Fig.5.

Fig.5 clearly demonstrates that when the same number of eigenfaces are used, $(PC)^2A$ always performs better than the standard eigenface technique. Pairwise two-tailed *t*-tests indicate that the difference between those two techniques is significant at 0.95 level of significance.

5. Further exploration of $(PC)^2A$

The vertical and horizontal integral projections themselves are very useful for face recognition tasks. For instance, Kanade [10] has employed them to determine the locations of facial features. Those projections could

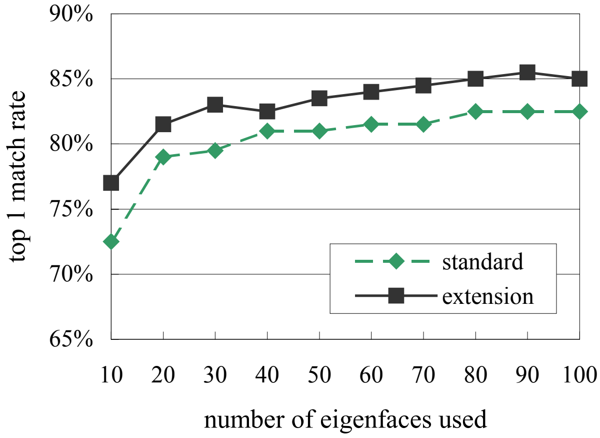


Fig.5 Comparison of the top 1 match rate when the same number of eigenfaces are used

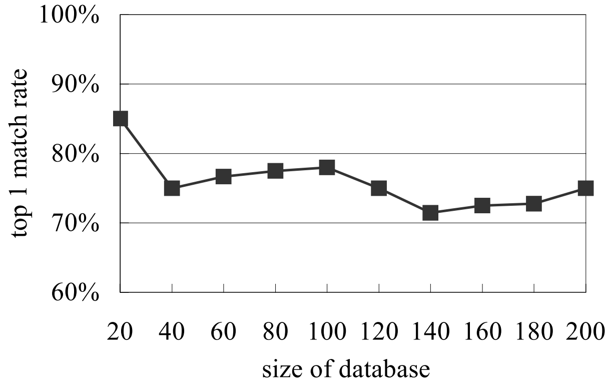


Fig.6 The top 1 match rate of the standard eigenface technique on the projection map

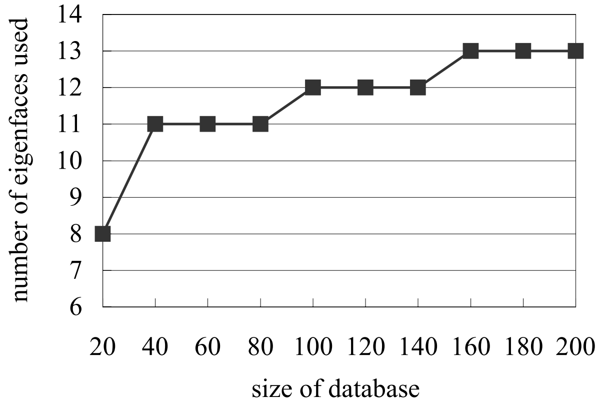


Fig.7 Number of eigenfaces used by the standard eigenface technique on the projection map

be more useful when they are combined to form the projection map. To exhibit this point, we use $M_P(x, y)$ instead of $P(x, y)$ as the input to the standard eigenface technique on databases with different sizes. The top 1 match rate is shown in Fig.6, and the number of eigenfaces used is shown in Fig.7, where θ is set as 0.9.

Fig.6 reveals that when presenting the projection map instead of the face image itself to the standard eigenface technique, the top 1 match rate is always better than 70%. Moreover, Fig.7 shows that in order to capture 90% of the whole variation, very few eigenfaces are required. Even when the size of the database is 200, only 13 eigenfaces are needed. This result illustrates the power of the projection map in face recognition. We believe that the projection map is so useful because it smoothes out the relatively unimportant features while emphasizes the relatively important ones for recognition. In other words, through fading out the unimportant features, it helps the eigenface technique focus on the important features that have become more salient after the preprocessing. We also believe that this property of the projection map contributes much to the success of $(PC)^2A$.

$(PC)^2A$ introduces the combination parameter, i.e. α. In order to know the influence of α on the performance of $(PC)^2A$, we perform some experiments on $(PC)^2A$ with different values of α. The number of eigenfaces used by $(PC)^2A$ is shown in Fig.8, the top 1 match rate is shown in Fig.9. Note that here the database used contains the images of 96 persons, which is a proper subset of the one used in Section 4.

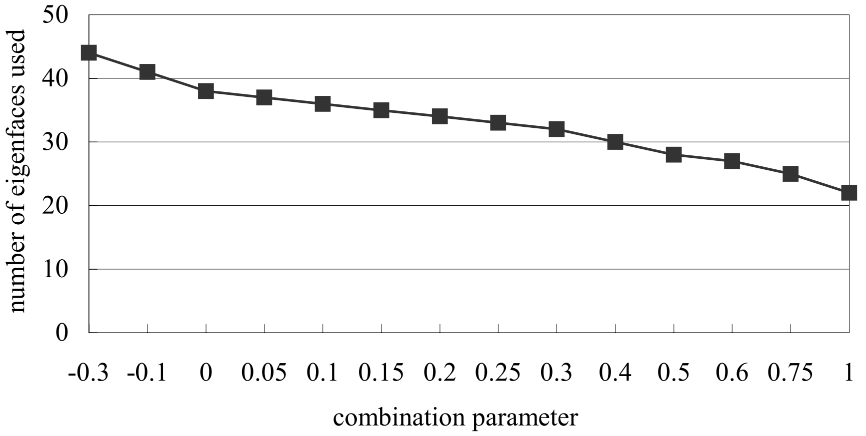


Fig.8 Number of eigenfaces used by $(PC)^2A$ with different values of the combination parameter, i.e. α

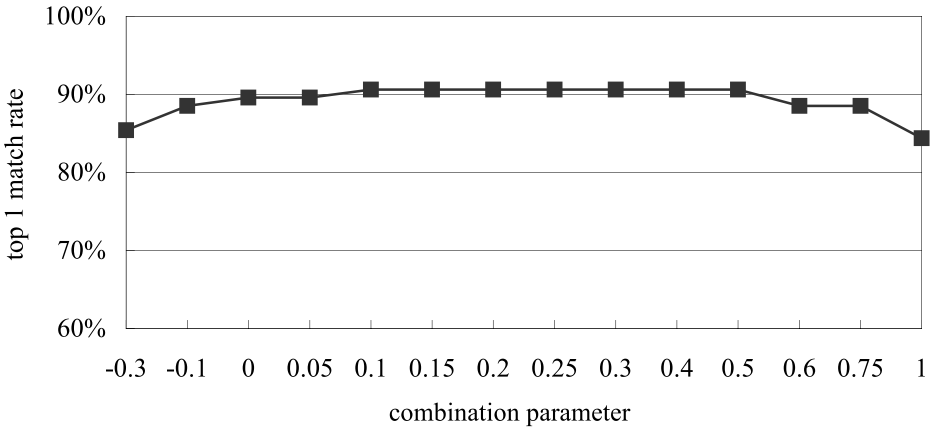


Fig.9 The top 1 match rate of $(PC)^2A$ with different values of the combination parameter, i.e. α

Fig.8 reveals that α greatly affects the dimensionality reduction of $(PC)^2A$ in the way that as α increases, the number of eigenfaces used significantly decreases. Recall that using fewer eigenfaces means that less computational cost, less storage cost, and less processing time are required. So, with large value of α, the computational cost, storage cost, and processing time can be reduced.

Fig.9 shows that as α gradually increases, the recognition accuracy increases at the beginning and decreases at the end. This may suggest that choosing an appropriate value for α is important to the recognition performance of $(PC)^2A$. Fortunately, Fig.9 also shows that there is a long flat area where the recognition accuracy of $(PC)^2A$ does not significantly change. So, in order to obtain good recognition performance in real-world tasks, a simple rule can work well in determining the value of α., that is, set α as a value which is neither too large nor too small.

6. Conclusion

At present there are many subspace techniques that work well in gray-level frontal view face recognition. However, in order to obtain good recognition performance, most of those techniques require that there are at least two training images per person so that the intra-class variation could be considered against the inter-class variation. Unfortunately, in real-world tasks such a requirement can not always be met. In this paper, a novel extension of the standard eigenface technique, i.e. $(PC)^2A$, is proposed. Instead of focusing on the feature extraction/selection or the learning stage of pattern recognition systems, $(PC)^2A$ focuses on the preprocessing stage where it smoothes the original face images through combining them with their vertical and horizontal projections. This technique requires less computational cost than the standard eigenface technique, and experiments show that when only one training image per person is available, $(PC)^2A$ is superior to the standard eigenface technique in that it achieves 3%-5% higher accuracy than the standard eigenface technique through using 10%-15% fewer eigenfaces.

We guess that the success of $(PC)^2A$ may lie in that it utilizes the projection map which fades out the unimportant features for face recognition and helps the eigenface technique focus on the important features that have become more salient after the preprocessing. However, such a guess can only be justified with more experiments on larger face databases, which is an interesting issue for future works.

If our guess on the reason of the success of $(PC)^2A$ is right, then it will be very interesting to explore that whether other kinds of smoothing, e.g. Gausian filters, can work equally well or even better than the projection map, which is another issue for future works.

Moreover, we hope to develop some methods along the way that $(PC)^2A$ goes, i.e. focusing on preprocessing, to deal with the situations where the test images contain *duplicate* probes [15], i.e. an image of a person whose corresponding gallery image is taken on a different date.

Furthermore, in our experiments the eyes in the face images are manually located. Since the locations of the eyes are very important in normalizing and cropping the face, we hope to design some mechanisms that could automatically and accurately extract such information from the images. This work is not required by $(PC)^2A$, but it is very important for the development of ideal face recognition systems.

Acknowledgements

The comments and suggestions from the anonymous reviewers greatly improve this paper. The National Natural Science Foundation of P.R.China, the Natural Science Foundation of Jiangsu Province, P.R.China, and the 10th Five Years Plan of Jiangsu Province, P.R.China, supported this research.

References

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

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

[3] Brunelli R, Poggio T. Face recognition: features versus templates. *IEEE Transactions on Pattern Analysis*

and Machine Intelligence 1993; 15(10): 1042-1052.

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

[5] Edwards G J, Cootes T F, Taylor C J. Face recognition using active appearance models. In: *Proceedings of the 5th European Conference on Computer Vision*, vol. 2, Freiburg, Germany, 1998, pp.581-595.

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

[7] Fromherz T. Face recognition: a summary of 1995-1997. Technical Report: TR-98-027, International Computer Science Institute, Berkeley, CA, 1998.

[8] Fukunaga K. *Introduction to Statistical Pattern Recognition*, 2nd edition, Academic Press, London, 1990.

[9] Jain A K, Duin R P W, Mao J. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2000; 22(1): 4-37.

[10] Kanade T. Picture processing by computer complex and recognition of human faces. Technical Report, Kyoto University, Department of Information Science, Kyoto, Japan, 1973.

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

[12] Moghaddam B, Jebara T, Pentland A. Bayesian face recognition. *Pattern Recognition* 2000; 33(11): 1771-1782.

[13] Moghaddam B, Pentland A. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1997; 19(7): 696-710.

[14] Pentland A, Moghaddam B, Starner T. View-based and modular eigenspaces for face recognition. In: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, Seattle, WA, 1994, pp.84-91.

[15] 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.

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

[17] Swets D L, Weng J. Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1996; 18(8): 831-836.

[18] Turk M, Pentland A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 1991; 3(1): 71-86.

[19] Valentin D, Abdi H, O'Toole A J, Cottrell G W. Connectionist models of face processing: a survey. *Pattern Recognition* 1994; 27(9): 1209-1230.

[20] Wiskott L, Fellous J M, Krüger N, Malsburg C. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1997; 19(7): 775-779.

[21] Zhang J, Yan Y, Lades M. Face recognition: eigenface, elastic matching, and neural nets. *Proceedings of the IEEE* 1997; 85(9): 1423-1435

[22] Zhao W, Chellappa R, Rosenfeld A, Phillips P J. Face recognition: a literature survey. http://citeseer.nj.nec.com/374297.html, 2000.