Face Identification and Verification Using PCA and LDA

Lih-Heng Chan, Sh-Hussain Salleh, Chee-Ming Ting, A. K. Ariff.
Center for Biomedical Engineering, Faculty of Biomedical Engineering and Health Science
Universiti Teknologi Malaysia
81300 Skudai, Johor, Malaysia.

Abstract

Algorithms based on PCA (Principal Components Analysis) and LDA (Linear Discriminant Analysis) are among the most popular appearance-based approaches in face recognition. PCA is recognized as an optimal method to perform dimension reduction, yet being claimed as lacking discrimination ability. LDA once proposed to obtain better classification by using class information. Disputes over the comparison of PCA and LDA have motivated us to study their performance. In this paper, we describe both of these statistical subspace methods and evaluated them using The Database of Faces which comprises 40 subjects with 10 images each. Both identification and verification results have revealed the superiority of LDA over PCA for this medium-sized database.

1. Introduction

The mushrooming commercial and law-enforcement applications and availability of feasible technologies have triggered the evolution of computer vision research [1]. Face recognition is among the popular research subject which has grown into wide range of commercial products from small scale research system, especially biometric implementation. Besides, the arising machine visual learning technology has encouraged police and national defense of various countries to enhance surveillance system at migration hot spot and important functions etc.

Compared with other very reliable methods such as fingerprint and retinal scans etc, face recognition gains its own advantage in terms of collectable because it does not rely too much on users' cooperation. Biometric application has two main aspects to consider: identification and verification. Identification is a *one-to-many* concept, in which a user's identity is determined by comparing and matching its template with multiple biometric templates store in the system database. Verification on the contrary is a *one-to-one* concept which only measures the similarity of a user's template with the particular biometric template in database. Thus, it requires the user to claim his/her identify before recognition process. If the similarity scores obtained is within the acceptable range, user will be recognized as the client in the database and he/she will be allowed to access the biometric system. Generally, verification is more important than identification for most commercial applications.

Appearance-based approach is one of the genres of face recognition methodology, which employs a whole face region as an input to the recognition system. A technique proposed by [2] is found efficiently represent pictures of faces using PCA. This idea has inspired Turk and Pentland [3] to develop the renowned eigenface method which uses PCA to perform dimension reduction for face recognition. Eigenface method captures and utilizes maximum variance across the training images to find a basis vector which aims at obtaining most compact data representation. It is fast and relatively simple yet effective, thus spurring drastic growth of appearance-based approaches in the following years [4].

Unsatisfactory of PCA as mere feature extractor has prompted numerous researchers to seek better alternatives. Due to the drawback of PCA which retains unwanted variations during basis vector computation, Belhumeur et al [5] presented fisherface method (FLD) which utilizes class information by maximizing the ratio of between-class scatter to that of within-class scatter in order to improve discrimination ability. They reported positive results in which error rates, sensitivity to lighting and expression variation, and computation time are reduced. Contemporaneously, the hybrid classifier employing both PCA and LDA shown in [6] has surpassed original LDA method. Original LDA suffers from computational inefficiency due to the high dimension of original data. Besides, effectiveness of combining PCA and LDA has been demonstrated by another framework [7] with the so called Discriminant Karhunen-Loeve projection. These researches share the same basic idea where PCA is first utilized on raw input data for dimension reduction, subsequently applying LDA on the transformed data for classification. This

978-1-4244-2328-6/08/$25.00 © 2008 IEEE

approach not only improves computational efficiency but also alleviates the complication of singular within-class scatter matrix [5].

Since then, LDA is often referred as FLD, instead of the original LDA [8]. Based on the same literature, there is still no straightforward conclusion can be drawn to determine the best algorithms among PCA, LDA and ICA. Despite the tendency of preferring LDA to PCA, superiority of LDA over PCA is also challenged by [9] which claims that PCA can outperform LDA in the case when training dataset is small.

In this work, PCA and LDA approach are investigated for face identification and verification. Effect of using various numbers of training samples and size of database is studied. We show that in overall, face identification using more training samples perform better for both PCA and LDA. However, keep increasing training samples does not guarantee recognition improvement. Observation also reveals the effect of using different ways in selecting train dataset and test dataset, although there is no further analysis of the underlying distributed data. In conclusion, our results show that LDA outperform PCA in both identification and verification.

The paper is organized as follows. Section 2 describes the face recognition appearance-based approaches comprising PCA and LDA. Section 3 presents the experimental evaluation on PCA and LDA, and comparison of both in terms of identification and verification. Conclusion is given in the last section.

2. Appearance-Based Approaches

Algorithms designed have to cope with the challenge of identification and verification as stated in previous section. PCA and LDA are the two pattern recognition techniques that we are interested to study for face recognition.

a. *Principal Component Analysis*

Also known as Karhunen-Loeve method, PCA emphasizes the use of information "between the features". Therefore it is not the definitive solution for face recognition, as reminded in [4]. It should be noted that the "feature" obtained does not correspond to facial features such as eyes, noses etc. The following is the description of this algorithm in mathematic terms.

Given a total of M images with (N_x x N_y) pixels, we convert them into training set $\Gamma = [\Gamma_1\ \Gamma_2\ \ldots\ \Gamma_M]$ with lexicographic ordering the pixel elements. Difference matrix is the training data with their mean removed. Covariance matrix is computed step-by-step from the difference matrix as given by following equations:-

$$\Psi = \frac{1}{M}\sum_{i=1}^{M}\Gamma_i \quad (1)$$

$$\Phi_i = \Gamma_i - \Psi \quad (2)$$

$$A = [\Phi_1\ \ \Phi_2\ \ \cdots\ \ \Phi_M] \quad (3)$$

$$C = A \cdot A^T = \frac{1}{M}\sum_{i=1}^{M}\Phi_i\Phi_i^T \quad (4)$$

, where Ψ is mean of whole data in vector form, Φ is a mean subtracted image, and A is difference matrix. C is covariance matrix lying in a very high dimension which is (N_x x N_y)* (N_x x N_y). [2] has provided solution for this problem, by using a covariance matrix L with small dimension which is (M x M). Eigenvectors, v computed from covariance L is multiplied with A to yield another variable υ which is able to represent the actual eigenvectors of covariance C.

$$L = A^T \cdot A = \frac{1}{M}\sum_{i=1}^{M}\Gamma_i^T\Gamma_i \quad (5)$$

$$\upsilon_i = A \cdot v_i \quad (6)$$

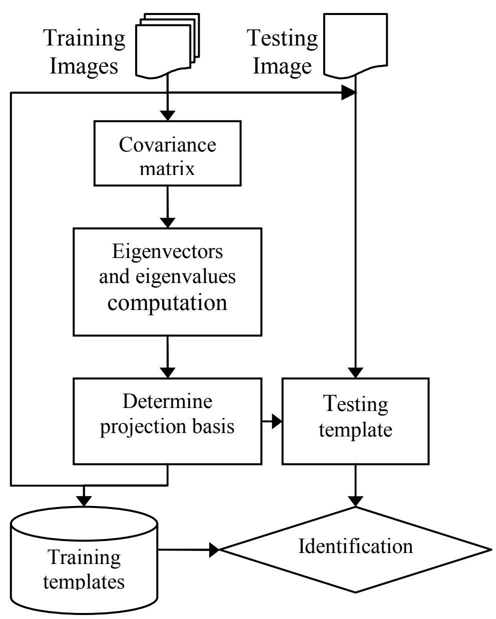

Figure 1. Simple flow chart of PCA algorithm.

Weight sets corresponding to respective subjects in training set can be obtained using projection basis is defined by following equations:-

978-1-4244-2328-6/08/$25.00 © 2008 IEEE

$$\omega_k = \upsilon_k^T \cdot \Phi = \upsilon_k^T \cdot (\Gamma - \Psi) \quad (7)$$

$$\Omega = [\omega_1 \;\; \omega_2 \;\; \cdots \;\; \omega_{M'}] \quad (8)$$

, where ω_k is weight, and Ω is weight set. Not all eigenvectors are needed, thus selecting M' eigenvectors will obtain projection basis, υ_k. In our experiment, selection of M' is based on 90% of the total generation power. For single test image identification, it has to undergo mean subtraction and being projected on υ_k to obtain its weight. Using Euclidean distance as classification method, similarity of test weight with every train weight is measured. The smallest distance yields identification result. Figure 1 shows a simple flow chart deriving the PCA algorithm.

b. Linear Discriminant Analysis

LDA utilizes face class information to represent face vector space efficiently. For such an attempt, LDA requires at least two training images per person. The keystone is to maximize Fisher Discriminant Criterion:

$$W_{opt} = \arg\max_{W} \frac{\left| W^T \cdot S_b \cdot W \right|}{\left| W^T \cdot S_w \cdot W \right|} \quad (9)$$

, where S_b is between-class scatter matrix and S_w is within-class scatter matrix. For c individuals having q_i training samples, S_b and S_w are obtained by using mean image per class, m_i and total mean, m_o, which are given by

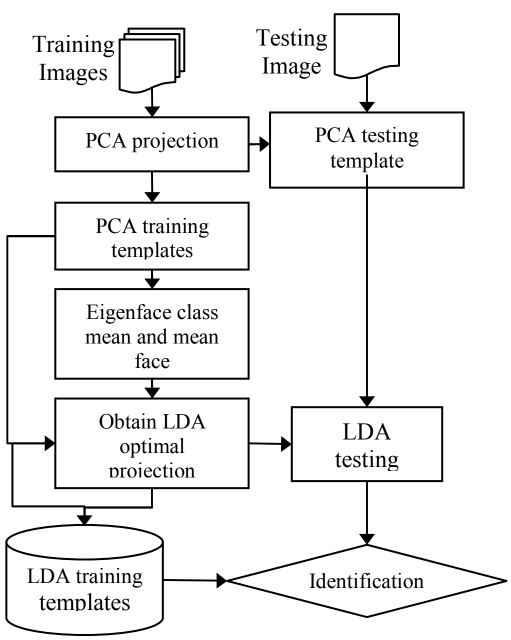

FIGURE 2. Simple flowchart of LDA algorithm.

$$m_i = \frac{1}{q_i} \sum_{k=1}^{q_i} \Omega_k \quad (10)$$

$$m_o = \frac{1}{M} \sum_{k=1}^{M} \Omega_k \quad (11)$$

$$S_W = \sum_{i=1}^{c} P(C_i)(\Omega - m_i) \cdot (\Omega - m_i)^T \quad (12)$$

$$S_b = \sum_{i=1}^{c} P(C_i)(m_i - m_o) \cdot (m_i - m_o)^T \quad (13)$$

, where $i = 1,2,\ldots,c$. Ω_k are the weight sets obtained in the PCA eigenface space. $P(C_i) = 1/c$ if each class has equal prior probability. Training image templates are obtained by the dot product of W_{opt} and weight sets. To perform face recognition, a test image with total mean subtracted is projected on υ and W_{opt}, subsequently followed by similarity measure with training image templates. Figure 2 illustrates the overview of LDA algorithm.

3. Experimental Evaluation

a. Database and Experiment Conditions

PCA and LDA were evaluated using The Database of Faces [10] provided by AT&T Laboratories Cambridge, which is much more renowned by its formal name: The ORL Database of Faces. This database contains 10 different images of 40 distinct subjects. The images vary in terms of lighting, facial expressions including open/closed eyes, facial details such as glass/without glasses, and different time of snapping pictures. In our experiment, these images were separated into training set and testing set and being processed in 56 x 46 pixels. Based on Euclidean distance as the classification measurement, identification was performed on every image in testing set using the templates of training images

b. Identification Performance of PCA and LDA.

Our experiment first investigate the performance of PCA and LDA from the effect of varying size of database, M and number of training samples, P. Size of database is the amount of subject involved in experiment. 5 testing images per subject were fixed at certain set for each different P to ensure fair comparison. Figure 3 is an example showing 10 images of a subject in database. The increasing P employs some/all images at the first row (1-5); while all images in second row (6-10) belong to testing set. Thus there were (M x 5) times of identification performed for each

M. Experiments were repeated with training set and testing set swapped. Be noted that LDA had $2 \leq P \leq 5$ due to the requirement that at least 2 training images needed.

Average recognition rate for PCA and LDA are recorded in Figure 4 and Figure 5 respectively. Both algorithms bear performance deterioration when they are to support more subjects which increased from M=10 to 20, 30, and 40. Thus large database again proved to inflict more recognition difficulty to algorithms. On the other hand, significant improvement is observed when more training samples are applied, especially from P=1 to 3 for PCA. However, it is surprised to find that, recognition rate of M=10 and M=20 achieve highest accuracy during P=3, deviating from our normal expectation. Thus an initial assumption could be drawn is that, more training samples does not guarantee higher recognition rate.

c. Identification Performance Comparison of PCA and LDA

It has been shown empirically by A.M. Martinez et al [9] that the performance of algorithms is regarding different ways of selecting images for training set and testing set. Thus in order to compare PCA and LDA in a fairer and reliable manner, we examined the recognition rates of 40 subjects by both algorithms in 6 different ways as shown in Table I. In the table, number 1-10 under the column of train set and test set refer to 10 images of single subject, which is depicted in Figure 3.

The results obtained are plotted in Figure 6 where the performance of PCA and LDA are compared with increasing P. Be noted that P=1 is omitted because of LDA restriction. From the figure, LDA outperforms PCA even when there are only 2 training samples per subject. From Table I, LDA is found to be outperformed slightly by PCA at certain comparison especially when there are few training samples. Nevertheless as shown in Figure 6, LDA obviously performs better in overall. This finding is inconsistent with the conclusion drawn by [9] which claims the superiority of PCA for small training dataset.

d. Verification Performance Comparison of PCA and LDA

To evaluate PCA and LDA on verification task, this experiment employed 20 subjects in database as clients.

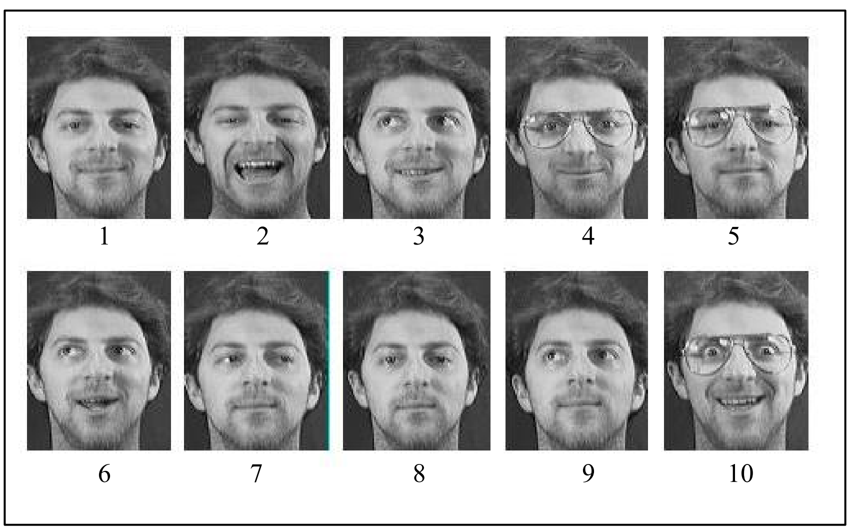

Figure 3. Example of a subject in the Database of Faces.

978-1-4244-2328-6/08/$25.00 © 2008 IEEE

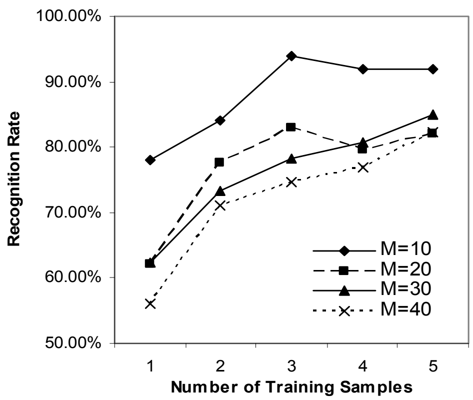

Figure 4. Recognition rate of PCA with varying size of database and number of training samples.

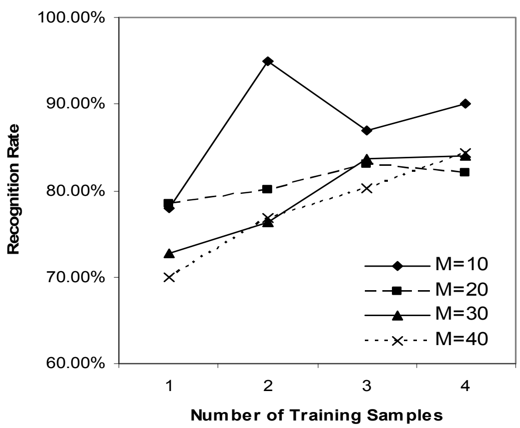

Figure 5. Recognition rate of LDA with varying size of database and number of training samples.

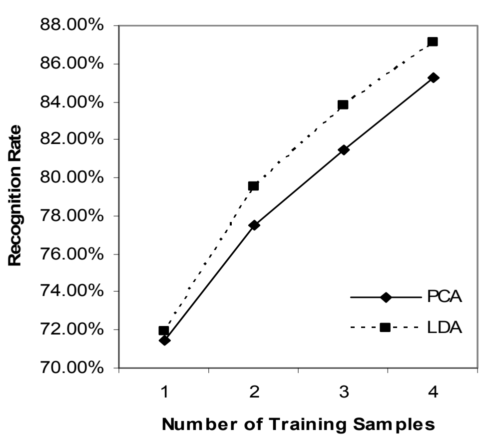

Figure 6. Comparison of PCA and LDA for 40 subjects with varying number of samples per person

Table 1. Codnition rate of PCA and LDA for 40 Subjects under permutation of trainings set ans testing set.

Train Set	Test Set	Recognition Rate	
		PCA	LDA
1,2	6,7,8,9,10	71.00%	70.00%
1,2,3		75.00%	76.00%
1,2,3,4		78.00%	78.00%
1,2,3,4,5		82.00%	82.00%
2,3	1,7,8,9,10	71.00%	71.50%
2,3,4		75.50%	77.50%
2,3,4,5		80.50%	81.00%
2,3,4,5,6		89.00%	87.50%
3,4	1,2,8,9,10	74.00%	77.00%
3,4,5		81.00%	84.00%
3,4,5,6		91.50%	91.50%
3,4,5,6,7		90.00%	91.50%
4,5	1,2,3,9,10	66.00%	65.00%
4,5,6		78.50%	82.50%
4,5,6,7		81.00%	85.50%
4,5,6,7,8		83.00%	87.00%
5,6	1,2,3,4,10	75.50%	78.00%
5,6,7		80.50%	79.50%
5,6,7,8		82.00%	84.50%
5,6,7,8,9		85.00%	88.00%
6,7	1,2,3,4,5	71.00%	70.00%
6,7,8		74.50%	77.50%
6,7,8,9		76.00%	82.50%
6,7,8,9,10		82.50%	86.50%

For each client, we have 5 images for training and another 5 images for testing. The 5 testing images were iterated for each training image to obtain a total of 25 scores to plot client distribution. On the other hand, the rest 20 subjects in database belong to the group of imposters. Since each imposter has 10 images, iteration of 5 images of a client with 200 images of 20 imposters produced a total of 1000 scores for each client to plot Gaussian graph.

Overlap of Gaussian graph determines FAR and FRR. For each client, speaker specific threshold was set *a posteriori* and adjusted to let FAR and FRR getting a similar value. This similar value becomes equal error rate (EER) which is an important evaluation for biometric system. Table II has shown the average EER of 20 clients using The Database of Faces, in which LDA is found to outperform PCA obviously, giving a percentage of improvement of 16%.

978-1-4244-2328-6/08/$25.00 © 2008 IEEE

Table 2. Average EER of 20 clients using PCA and LDA.

Algorithm	Equal Error Rate (%)
PCA	14.87
LDA	12.49

4. Conclusion

In this paper, face identification and verification performance using PCA and LDA are investigated using The Database of Faces. To enhance original PCA method's discrimination ability, LDA is applied on PCA face subspace for classification. Results have shown the superiority of LDA over PCA, even when the training dataset is small. Besides, significant improvement is observed when there are more training samples employed for both algorithms. However, amount of training samples should consider other factor such as size of database because it consumes more computational time to increase training samples. For 40 subjects and 5 training samples per subject, PCA and LDA achieves recognition rate of 85.25% and 87.08% respectively. 16% percentage of improvement is gained by LDA over PCA for verification task. As future work we suggest to extend the study on standard LDA improvement approaches, such as F-LDA (Fractional-step Linear Discriminant Analysis) and FD-LDA (Direct F-LDA)

5. Acknowledgment

This research project is supported by CBE (Center of Biomedical Engineering) at Universiti Teknologi Malaysia and funded by Minister of Higher Education (MOHE), Malaysia under grant "Support Vector Machines (SVM) As a Large Margin Classifier For Machine Learning Algorithm " Vot 78029.

6. References

[1] W. Zhao and R. Chellappa. Face processing Advanced Modeling and Method. Elsevier Inc, 2006

[2] M. Kirby and L. Sirovich. "Application of the karhunen-loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 12(1), January 1990.

[3] Turk, M. and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, 1991. 3(1): p. 71-86.

[4] M. Turk, "A random walk through eigenspace," *IEICE Trans. Inform. Syst.*, vol. E84-D, pp. 1586–1695, Dec. 2001.

[5] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, July 1997.

[6] W. Zhao, R. Chellapa, A. Krishnaswamy. "Discriminant Analysis of Principal Components for Face Recognition," *In Proceedings of the 3rd International Conference on Automatic Face and Gesture Recognition,* pages 336-341, 1998.

[7] D. Swets and J. Weng. "Using discriminant eigenfeatures for image retrieval," *IEEE trans. on Pattern Analysis and Machine Intelligence*, 18:831-836, 1996.

[8] K. Delac, M. Grgic, P. Liatsis. "Appearance-based Statistical Methods for Face Recognition," *47th, International Symposium ELMAR-2005,* 08-10 June 2005, Zadar, Croatia.

[9] A.M. Martinez and A.C. Kak. "PCA versus LDA," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 23, No.2, pp. 228-233, 2001.

[10] The Database of Faces, AT&T Laboratories Cambridge, http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

978-1-4244-2328-6/08/$25.00 © 2008 IEEE