The Spectral-Face Analysis for Face Recognition

Yan Zhu and Eric Sung
School of Electrical and Electronic Engineering, Nanyang Technological University
Nanyang Avenue, Singapore 639798
E-mail: {*zhuyan@pmail.ntu.edu.sg, eericsung@ntu.edu.sg*}

Abstract

Subspace methods have been widely used for face recognition possibly because of their robustness and simplicity. Due to high dimensionality of image space, these methods are likely to encounter computational problem when having to deal with very large number of face training samples. In this paper, a new subspace approach called the spectral-face analysis is developed to overcome this. It handles pixel information in matrix form rather than as vectors, and in so doing, keeps the basis computation invariant to the size of the training samples. Two types of statistics are implemented for the spectral-face analysis: the "covariance face" and the "error face". As they employ smaller vector space, the recognition rates are, as expected, not as good as conventional subspace methods such as PCA and LDA. To improve the performance, we extend the spectral-face methodology by some "stacking" technique, the sole purpose being to increase the vector space dimension. Extensive tests have been carried out on the ORL face database with good and interesting results. Application of LDA to the spectral-face analysis also showed marked improvements.

1. Introduction

The challenge to face recognition resides in the large appearance variation of human faces [3]. It may either come from intrinsic face changes in expression, makeup and aging effect, or result from extrinsic sources such as ambient lights and viewpoints. Appearance invariant face recognition can be tackled in two ways: 1) use of multiple 2D facial images and to derive class relevant features [13, 7]; and 2) using 3D face data, or 3D face synthesis, such that the information on face shape and texture can be extracted and applied for classification [4, 15].

Subspace methods, being major approaches in the first category, are popular because of their simplicity and good ability of feature generalization; successful examples of which include Principal Component Analysis (PCA) [12] and its class specific version of Linear Discriminant Analysis (LDA) [1, 16]. These methods deal with high dimensional image space, since they concatenate image pixel values into vector form to ensure fine statistical measure of pixel distribution. As a result, basis computation can be prohibitive when handling large training datasets. This is contradictory to the fact that the subspace system is appearance-based, and would need more face samples, covering wider appearance variations, to achieve robust recognition. A possible solution is to design a computationally effective subspace representation, as what we suggested in the Spectral-Face Analysis.

Instead of vectorizing image pixels, the Spectral-Face Analysis treats facial images as matrices, and use Singular Value Decomposition (SVD) to derive subspace basis. In this way, the basis computation can be based on a statistical measurement that is reduced from $R^{mn \times mn}$ to $R^{m \times n}$, assuming the image size is $m \times n$, and the number of training samples, $N > mn$. The spectral-face analysis simplifies the subspace construction by adopting a coarser statistics of pixel distribution, whose small size is invariant to the number of images that are used for training. The method suggests the feasibility of subspace methods to handle large capacity of face datasets, yet it may not preserve as much class/appearance relevant information as conventional subspace methods. In this paper, two different statistics are implemented for the proposed approach: the "covariance face" and the "error face". A newly developed LDA algorithm is also applied to the subspace on covariance face to improve its recognition performance. To evaluate the effect of reduced vector space dimension, some "stacking" technique is employed to enhance the vector space dimension. Experiments are carried out on the ORL database, which consists of 40 subjects and 10 pictures for each of them.

2. Subspace methods for face recognition

This section provides a brief background of the subspace methods, and introduces a direct LDA algorithm proposed by Yang *et al.* recently [14], which will later be used to explore the spectral-face space.

Subspace methods find their application in pattern recognition since the early days of computer vision, wherein pattern classes are not primarily defined as bounded regions or zones in the features space, but rather given in terms of linear subspaces that are obtained from statistical analysis [10]. To make a system able to identify query faces, a subspace (as specified by a specific set of basis vectors) that carries class-relevant information must be derived from a given set of face samples. Identification is done by projecting new face images onto those basis vectors, and comparing the projections with the existing prototype(s) for each person. Successful subspace methods include the PCA and the LDA. Recent development with Independent Component Analysis (ICA) [9] shows promise; however, in this paper, we will not deal with ICA.

The Eigenface Method of Turk and Pentland [12], which is based on PCA, maximizes the subspace scattering of all projected samples to facilitate classification; while LDA projects images onto a subspace that purposely minimizes within-class differences, and at the same time, maximizes between-class differences [1, 16]. Many LDA approaches first use the PCA to discard the null space of within-class scattering, and then perform the LDA to maximize the discriminatory power. Recently, Yang *et al.* argued that this separate PCA step may lose discriminative information, and proposed a unified LDA/PCA algorithm for face recognition.

2.1. A direct LDA algorithm

For LDA, the between-class and within-class scattering of sample face images, S_B and S_W, are defined as:

$$S_B = E[(\mu_k - \mu)(\mu_k - \mu)^T], \quad (1)$$

$$S_W = E[(f_{k,i} - \mu_k)(f_{k,i} - \mu_k)^T], \quad (2)$$

where μ is the average face of all vectorized training faces, μ_k is the mean face of a class X_k, and $f_{k,i}$ is the i^{th} image sample of class X_k. The class separability is measured by Fisher's Criterion:

$$\mathcal{J}(W) = \arg\max_W \frac{|W^T S_B W|}{|W^T S_W W|}, \quad (3)$$

where W is the optimal projection matrix.

According to Yang *et al.*, the null space of S_W may contain useful information if the projection of S_B is not zero in that direction, but the null space of S_B can be safely discarded. They hence proposed a LDA algorithm that diagonalizes S_B and S_W in the way that

$$W^T S_W W = D_W, W^T S_B W = I, \quad (4)$$

where D_W is a diagonal matrix with diagonal elements sorted in a decreasing order. This is done by eigen-decomposing S_B first as

$$S_B = V D_B V^T, \text{i.e. } (D_B^{-1/2} V^T) S_B (V D_B^{-1/2}) = I. \quad (5)$$

Let $Z = V D_B^{-1/2}$, we have $Z^T S_B Z = I$. Now, eigen-decompose $Z^T S_W Z$ in a similar approach, so that $Z^T S_W Z = U D_W U^T$. It can be checked that

$$U^T Z^T S_W Z U = D_W, \text{and } U^T Z^T S_B Z U = I. \quad (6)$$

Thus, the LDA transformation, W, and the projection of a face f_i on the LDA subspace, a_i, are found by

$$W = (ZU), \text{and } a_i = D_W^{-1/2} W^T f_i, \quad (7)$$

with the projected values sphered. Since the objective is to maximize $\mathcal{J}(H)$, those eigenvectors corresponding to the smallest eigenvalues of D_W are the most discriminative dimensions. Yang's direct LDA approach unifies the PCA and LDA processes, while retaining equivalently good class separability.

3. The Spectral-Face Analysis

As mentioned in Section 1, the spectral-face analysis treats images as matrices, and use Singular Value Decomposition (SVD) to derive the subspace basis. It serves as an extension or rather a complement of PCA, yet chances are that the reduced subspace features are still representative across face objects and are still suitable for classification. The idea of subspace face recognition without image vectorization also appeared in early 1990's [6, 2]. However, they simply derive subspace features from SVD of the average face, which is proved to be lacking in discriminatory power according to our experiments. In this section, we perform spectral-face analysis based on the "covariance face" and the "error face". A "stacking" technique is also developed, to give freedom in feature space dimension.

3.1. Spectral-face analysis on the Covariance Face

Consider N face samples, with each face, M_i, of size $m \times n$. First to be computed is the covariance matrix:

$$\begin{aligned} C &= E[(M_i - M_{ave})(M_i - M_{ave})^T] \\ &\approx \frac{1}{N} \sum_{i=1}^{N} (M_i - M_{ave})(M_i - M_{ave})^T, \end{aligned} \quad (8)$$

where $M_{ave} = E[M_i]$ is the average face. Note that the $m \times m$ covariance C remains at a similar size as that of the images, irrespective of how many pictures are used for training. The idea now is to find the SVD of C to construct the subspace basis. As C is symmetrical, its singular values, s_k, and column eigenvectors, u_k, can be expressed as

$$C = USU^T = \sum_{k=1}^{m} s_k u_k u_k^T, \quad (9)$$

which is the spectral decomposition of C, with $S = \text{diag}(s_1, s_2, \cdots, s_m)$, and $U = [u_1, u_2, \cdots, u_m]$. The singular values are sorted in descending order, and the same order applies to their corresponding eigenvectors. We then select $S_K = \text{diag}(s_1, s_2, \cdots, s_K)$ and $U = [u_1, u_2, \cdots, u_K]$

for recognition, with integer $K < m$ being the chosen dimension of the new subspace, also known as feature size. Hence, a covariance face can be linearly decomposed onto the face subspace as

$$(M_i - M_{ave})(M_i - M_{ave})^T = U_K A_i S_K U_K^T = \sum_{k=1}^{K} a_{ik} s_k u_k u_k^T, \quad (10)$$

where $A_i = \text{diag}(a_{i1}, a_{i2}, \cdots, a_{iK})$ is a diagonal matrix with the spectral-face projection values at diagonal, and $u_k u_k^T$ is the spectral-face basis. Applying orthogonality of U_K, the projection can be computed as

$$A_i = U_K^T (M_i - M_{ave})(M_i - M_{ave})^T U_K S_K^{-1}. \quad (11)$$

In this approach, the face decomposition is based on the components of covariance matrix, $(M_i - M_{ave})(M_i - M_{ave})^T$, instead of on the face image itself. This kind of subspace is said to describe the "covariance face". Fig. 1 and 2 demonstrate the first 18 basis and the face reconstruction with the subspace features based on Eq. 10. It can be seen that the first few spectral vectors carry global information, while the rest contain more picture detail.

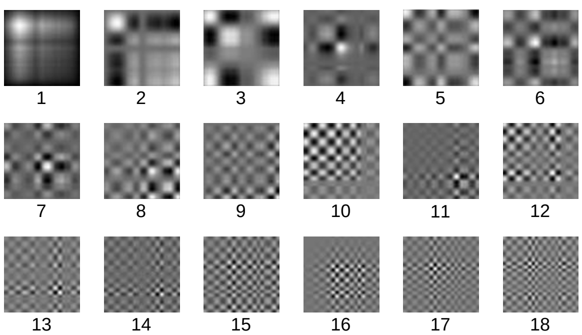


Figure 1. First 18 Spectral-Face Basis of Covariance Face.

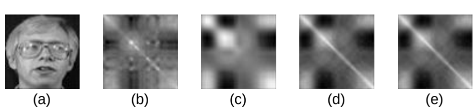


Figure 2. Face Reconstruction with Spectral-Face Basis on Covariance Face (a) Original image, M_i. (b) Covariance Face, $(M_i - M_{ave})(M_i - M_{ave})^T$. (c)(d)(e) Reconstructed covariance face using the first 5, 30, 64 features respectively.

3.2. Spectral-face analysis on the Error Face

Besides the covariance face method, an accidental spectral-face method is to be introduced here. It uses a pseudo random matrix C for basis derivation, and possesses good class separation. Consider

$$C = E[M_i - M_{ave}] = E[M_i] - M_{ave}. \quad (12)$$

The matrix should be mathematically zero, but it results in some small, trivial values, due to the round-up error by the digital computer. As shown in Fig. 3, the value of the pixels ranges from -7×10^{-14} for black and 5×10^{-14} for white.

The matrix obtained is in fact of small random numbers, and the subspace representation and projection can be computed using regular SVD approach:

$$C = USV^T = \sum_{k=1}^{K} s_k u_k v_k^T, \quad (13)$$

$$M_i - M_{ave} = UA_iV^T = \sum_{k=1}^{K} a_{ik} u_k v_k^T, \text{ and} \quad (14)$$

$$A_i = U^T (M_i - M_{ave}) V. \quad (15)$$

The basis derived from the Error Face, $u_k v_k^T$, is shown in Fig. 4. They bias on face pattern divergences, which should play a very important role for good feature separation. The phenomenon suggests that the semantics of the basis construction is not crucial to subspace separation, as long as the subspace resides in the face space that records face image variations.

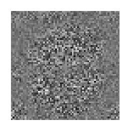


Figure 3. Error Face, $C = E[M_i - M_{ave}]$, from computer.

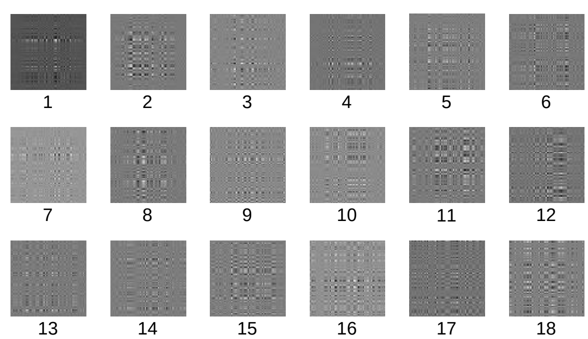
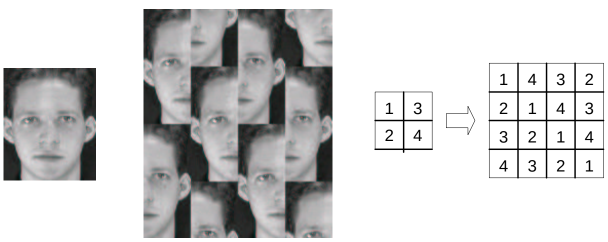


Figure 4. First 18 Spectral-Face Basis of Error Face.

3.3. Feature space expansion with Image Stacking

The basis obtained from the methods above is compact, yet suffers from low vector space dimension, which would be hard to preserve optimal class separation based on the premise of Cover's theorem [5]. In the spectral-face analysis, the dimension of the feature space solely depends on the row size of face images. A stacking technique, called "circulant stacking", is hence introduced to expand the space dimension.

For circulant stacking, a face image is cut and numbered into $L \times L$ smaller parts, and then piled up with the part numbers arranged in a circulant matrix. Fig. 5 describes a case where $L = 2$. In general, this stacking method enables the face images to be reshaped from $m \times n$ to $Lm \times Ln$ matrices. Thus, a feature space dimension is expanded by L times, at the expense of increased computational load.



(a) (b) (c)

Figure 5. Circulant Stacking of a face image
(a) original image, $m \times n$. (b) stacking by circulant expansion, $2m \times 2n$. (c) schematic description of (b).

4. Experiments

4.1. Experimental setup

The proposed spectral-face methods are evaluated with the ORL Database from AT&T laboratories [11], which consists of 40 distinct subjects, each having ten different images. For all the accuracy tests below, 5 images for each of the 40 persons are randomly selected to train the system, with the rest projected onto the subspace obtained to see if they can find the correct match. Unless otherwise specified, the simple euclidian distance matching method, Nearest Neighbor (NN), is used for classification. A new linear classifier called the Nearest Feature Line method (NFL) [8] is also employed to evaluate the importance of more training samples for the spectral-face methods.

4.2. Recognition performance of various methods

Table 1 lists the average recognition performance of the spectral-face methods as compared with the PCA and the algebraic method of Hong and Cheng [6, 2]. The results are averaged over 100 times random tests, for 64×64 images using NN classifier and 64 features. The table shows that the PCA still has the highest accuracy. However, the covariance face and the error face methods also give reasonable performance, considering that they are in lower dimensional spaces. The algebraic method that is based on the average face does not work well, which suggests that a subspace basis without knowledge of face differences is not suitable for classification.

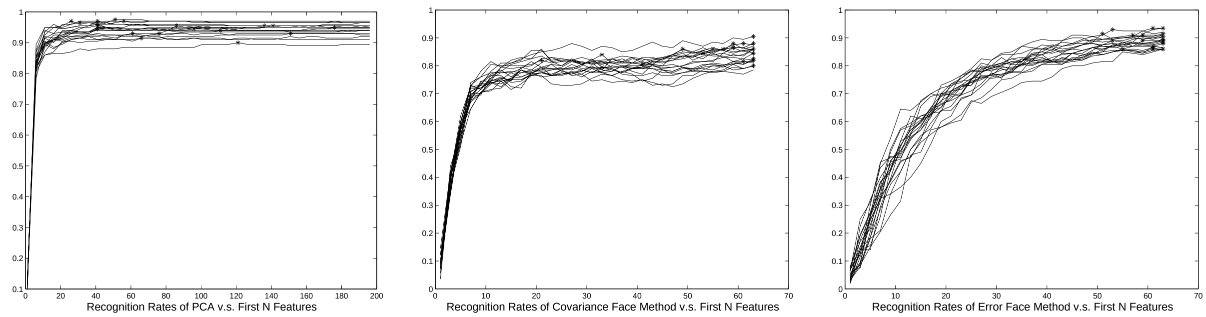
Table 1. Average accuracy of various methods on ORL face database.

Method	Accuracy
Turk and Pentland's PCA	94%
Hong and Cheng's Algebraic Method	68.5%
The Covariance Face Method	84.44%
The Error Face Method	90%

Fig. 6 summaries the recognition performance of the PCA, the covariance face method and the error face method with respect to varying feature length, where all the features are ranked with the corresponding eigen or singular values in descending order. The plots trace the performance of the three methods for 20 randomly selected datasets by NN classification. The full feature size is 199 for PCA, and 64 for the spectral-face methods, as there are 200 64×64 facial images in each of the datasets. For covariance face method, it can be seen that the features with smaller singular values are less significant for classification, but including them in the process would ensure a relatively higher accuracy. For error face method, however, the performance saturates much slower or never, which suggests that the most discriminatory features of this method are randomly distributed among the basis. It shall be noticed that the error face method can achieve very high accuracy with proper feature selection, which is, in some cases, even comparable to the PCA. However, the new method is less stable as demonstrated by the larger performance deviation among different datasets. Thus, it shows a lower accuracy in Table 1, on average.

For the spectral-face methods, larger images proved to have better class separation, as they enhance the vector space dimension. The average accuracy from 100 random tests reveals $1 \sim 2\%$ improvement for 112×92 images over 64×64 ones, which is much more significant than that of the PCA. It is also observed that the performance perturbation for both the spectral-face methods are quite large. This confirms that the lower resolution measurements of pixel distribution make the system more fluctuated. A feature optimization scheme needs to be embedded for better recognition performance.

Experiments also show that the spectral-face methods



(a) (b) (c)

Figure 6. Recognition performance of (a) the PCA, (b) the Covariance Face Method and (c) the Error Face Method under varying feature length.

prefer classifiers with good sample generalization ability. Comparison tests of NFL over NN show improvement of $2 \sim 4\%$ on average. The use of NFL is somehow equivalent to adding more samples to the training datasets. the results suggest that more knowledge on possible appearance variation tends to enhance class separation, especially for methods like spectral-face that operates in lower dimensions.

4.3. LDA enhancement of the covariance face method

The direct LDA algorithm introduced in Section 2.1 is adapted into the covariance face method by substituting face matrices into the scattering formulae, and finding transformation matrices that simultaneously diagonalize S_B and S_W using SVD instead of eigen-decomposition. The LDA method is not suitable for the error face method, as the scattering matrices it used would again be random with very small, trivial values, which is hard to retain discriminatory power across face classes. Table 2 demonstrates the effects of LDA on the PCA and the covariance face features. 64×64 images were used for testing, and recognition utilizes the full span of features.

The most significant enhancement is found in the covariance face method, which proves that its features, though only based on the left-eigenvectors, can actually preserve class relevant information. Nevertheless, the performances of the spectral-face methods are still worse than the PCA, even with the LDA enhancement. This is mainly caused by the lower space dimension, therefore, we later tried to scale it up using circulant stacking.

4.4. Effect of image stacking

In this section, 64×64 images were reshaped by the stacking technique mentioned in Section 3.3 to expand the

Table 2. Recognition performance with LDA enhancement.

Method	PCA	Covariance Face
Without LDA	94%	84.4%
With LDA	95.69%	90.5%

feature space dimension for the covariance face method. Fig. 7 examines the effect of different feature length using $4m \times 4n$ circulant expansion with NFL classifier. The result confirms that the most discriminant features are related with larger singular values, and the best feature size is about 120, while including more features only introduce noises that eventually confuses decision.

In general, using higher dimensional vector space will enhance the class separation of the spectral-face analysis. However, careful feature selection shall be incorporated for optimization. Moreover, only the moderate level of stacking has practical meaning. Expanding with a factor greater than 4 would involve too much computation, which may defeat the original purpose that the spectral-face was designed for.

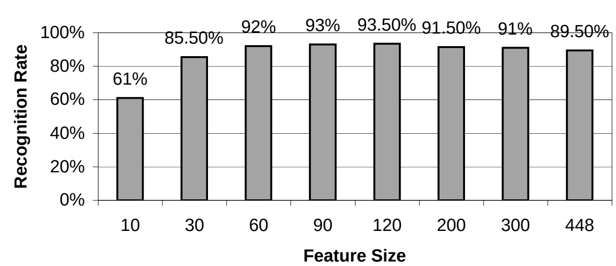


Figure 7. Performance of the covariance face method with $4m \times 4n$ circulant stacking.

5. Conclusions

In this paper, a new subspace method called the spectral-face analysis is developed, to facilitate face recognition in large image datasets. They use compact measurement of pixel distribution, and are efficient at subspace computation for this reason. Two types of implementations were developed based on the spectral-face scenario, i.e. the covariance method and the error face method. Feature space expansion and LDA algorithm are applied to the covariance method, in order to explore its subspace properties for class separation.

Intensive experiments carried on the ORL database show that the new methods are less stable as compared to the PCA method, because of the small vector space they used for feature derivation. However, with careful feature selection, some of the spectral-face method can achieve equivalent performance as the PCA method. In the future, the research will focus on the subspace property of the new methods as compared to the face space and the scattering matrices. With more knowledge on the actual subspace distribution, it may facilitate feature selection and hence, enhance the system performance. To generalize the experiment results, face database other than the ORL faces shall be tested on the new methods. Furthermore, it is worthwhile to study on the random basis as in the error face method, in order to see how generally it can be applied to face recognition problems.

References

[1] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19, pp.771-720, 1997.

[2] Y.Q. Cheng, K. Liu, J.Y. Yang and H.F. Wang, “A Robust Algebraic Method for Human Face Recognition,” in *Proceedings, International Conference on Pattern Recognition*, vol.2, pp.221-224, 1992.

[3] S. Gong, S.J. McKenna and A. Psarrou, *Dynamic Vision: From Images to Face Recognition*, Imperial College Press, 2000.

[4] G.G. Gordon, “Face Recognition Based on Depth and Curvature Features,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.808-810, 1992.

[5] S. Haykin, *Neural Networks, a Comprehensive Foundation*, 2nd Ed., Prentice-Hall, 1999.

[6] Z.Q. Hong, “Algebraic Feature Extraction of Image for Recognition,” *Pattern Recognition*, vol.24, no.3, pp.211-219, 1991.

[7] A.J. Howell and H. Buxton, “Towards Unconstrained Face Recognition from Image Sequences,” in *Proceedings, International Conference on Automatic Face and Gesture Recognition*, pp.224-229, 1996.

[8] S.Z. Li and J. Lu, “Face Recognition Using the Nearest Feature Line Method,” *IEEE Transactions on Neural Networks*, vol.10, no.2, pp.439-443, 1999.

[9] C. Liu and H. Wechsler, “Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition,” *AVBPA’99*, 1999.

[10] E. Oja, *Subspace Methods of Pattern Recognition*, Research Studies Press, 1983.

[11] ORL Face Database, *http://www.uk.research.att.com/facedatabase.html*, 1994.

[12] M. Turk and A. Pentland, “Eigenfaces for Recognition,” *Journal of Cognitive Neuroscience*, Vol.3, pp.72-86, 1991.

[13] Y. Yacoob, H. Lam and L.S. Davis, “Recognizing Faces Showing Expressions,” *International Workshop on Automatic Face- and Gesture-Recognition*, 1995.

[14] J. Yang, H. Yu and W. Kunz, “An Efficient LDA Algorithm for Face Recognition,” *ICARCV’2000*, Singapore, 2000.

[15] W. Zhao and R. Chellappa, “SFS Based View Synthesis for Robust Face Recognition,” in *Proceedings, International Conference on Automatic Face and Gesture Recognition*, 2000.

[16] W. Zhao, R. Chellappa, and A. Krishnaswamy, “Discriminant Analysis of Principal Components for Face Recogition,” in *Proceedings of International Conference on Automatic Face and Gesture Recognition*, pp.336-341, 1998.