

EIGENFACES AND FISHERFACES

Naotoshi Seo

University of Maryland
ENEE633 Pattern Recognition
Project 2-1

ABSTRACT

This project describes a study of two traditional face recognition methods, the Eigenface [10] and the Fisherface [7]. The Eigenface is the first method considered as a successful technique of face recognition. The Eigenface method uses Principal Component Analysis (PCA) to linearly project the image space to a low dimensional feature space. The Fisherface method is an enhancement of the Eigenface method that it uses Fisher's Linear Discriminant Analysis (FLDA or LDA) for the dimensionality reduction. The LDA maximizes the ratio of between-class scatter to that of within-class scatter, therefore, it works better than PCA for purpose of discrimination. The Fisherface is especially useful when facial images have large variations in illumination and facial expression. In this project, a comparison of the Eigenface and the Fisherface methods respect to facial images having large illumination variations is examined.

Index Terms— Eigenface, Fisherface, Face Recognition, PCA, LDA, Illumination Invariant

1. INTRODUCTION

Over the last couple of years, face recognition has become a popular area of research in computer vision and one of the most successful applications of image analysis and understanding. The face recognition problem can generally be formulated as: Given still or video images of a scene, identify or verify one or more persons in the scene using store database of faces.

The Eigenface [9], [11], [12] is the first method considered as a successful technique of face recognition. The Eigenface method uses Principal Component Analysis (PCA) to linearly project the image space to a low dimensional feature space.

The Fisherface [1] is an enhancement of the Eigenface method. The Eigenface method uses PCA for dimensionality reduction, thus, yields projection directions that maximize the total scatter across all classes, i.e., across all images of all faces. The PCA projections are optimal for representation in a low dimensional basis, but they may not be optimal from a discrimination standpoint. In stead, the Fisherface method

uses Fisher's Linear Discriminant Analysis (FLDA or LDA) which maximizes the ratio of between-class scatter to that of within-class scatter.

In section 2, the Eigenface method is investigated. In section 3, the Fisherface method is examined. In section 4, an empirical comparison of the Eigenface and the Fisherface methods subject to facial images having large illumination variations is performed. In section 5, conclusion of this study is provided.

2. EIGENFACE

The *Eigenface* method is based on linearly projecting the image space to a low dimensional feature space [9], [11], [12]. The Eigenface method, which uses principal components analysis (PCA) for dimensionality reduction, yields projection directions that maximize the total scatter across all classes, i.e., across all images of all faces.

Let us consider a set of N sample images $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ taking values in an n -dimensional image space, and assume that each image belongs to one of c classes $\{X_1, X_2, \dots, X_c\}$. Let us also consider a linear transformation mapping the original n -dimensional image space into an m -dimensional feature space, where $m < n$. The new feature vectors $\mathbf{y}_k \in \mathbb{R}^m$ are defined by the following linear transformation:

$$\mathbf{y}_k = W^T \mathbf{x}_k \quad k = 1, 2, \dots, N \quad (1)$$

where $W \in \mathbb{R}^{n \times m}$ is a matrix with orthonormal columns.

If the total scatter matrix S_T is defined as

$$S_T = \sum_{k=1}^N (\mathbf{x}_k - \mu)(\mathbf{x}_k - \mu)^T \quad (2)$$

where $\mu \in \mathbb{R}^n$ is the mean image of all samples, then after applying the linear transformation W^T , the scatter of the transformed feature vectors $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ is $W^T S_T W$. In PCA, the projection W_{opt} is chosen to maximize the determinant of the total scatter matrix of the projected samples, i.e.,

$$W_{\text{opt}} = \arg \max_W |W^T S_T W| \quad (3)$$

$$= [\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_m] \quad (4)$$

where $\{\mathbf{w}_i | i = 1, 2, \dots, m\}$ is the set of n -dimensional eigenvectors of S_T corresponding to the m largest eigenvalues $\{\lambda_i | i = 1, 2, \dots, m\}$ [4], i.e.,

$$S_T \mathbf{w}_i = \lambda_i \mathbf{w}_i, \quad i = 1, 2, \dots, m. \quad (5)$$

Since these eigenvectors have the same dimension as the original images, they are referred to as Eigenpictures in [9] and Eigenfaces in [11], [12]. Classification is performed using a nearest neighbor classifier in the reduced feature space.

2.1. Further Observations

To obtain deeper understandings of the Eigenface methods, a few extra experiments are conducted in this section.

The Fig. 1 shows a facial image set taken from ORL face database [8]. The PCA analysis is applied to the image set, and the obtained meanface and eigenfaces are shown in the Fig. 2.



Fig. 1. The input facial image set

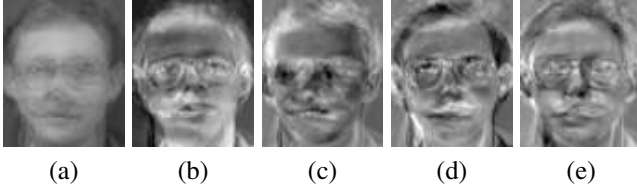


Fig. 2. (a) The meanface (b-e) The eigenfaces. The left-most eigenface is the most principal one.

In Fig. 3, the top row shows another image set and the bottom row shows the reconstructed images which are created by projecting images into the PCA subspace once and reconstructing them back into the full original space. The pictures illustrate the reason why the Eigenface method works well. The reconstructed images each other are more similar than the original images each other, thus, it possibly achieves better recognition by comparing the reconstructed images than comparing the original images like a typical template matching method.

Notice that we in fact do not need to compute the reconstructed images, but uses the projected points in the PCA subspace for the face recognition. A simple proof to show their equivalence is as follows: Without loss of generality, let $\mathbf{x} \in \mathbb{R}^n$ be a query face image vector to be classified and

$\{\mathbf{x}_i \in \mathbb{R}^n, i = 1, \dots, c\}$ ¹ be representatives from each face class i with zero empirical mean where n is the dimension of the vectors. Furthermore, let $\mathbf{y} \in \mathbb{R}^m$ and $\{\mathbf{y}_i \in \mathbb{R}^m, i = 1, \dots, c\}$ be the projected points into the PCA subspace of them, that is,

$$\mathbf{y} = \mathbf{W}^T \mathbf{x}, \quad (6)$$

$$\mathbf{y}_i = \mathbf{W}^T \mathbf{x}_i, \quad i = 1, \dots, c. \quad (7)$$

where $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_m]$ is a $n \times m$ matrix of a set of n dimensional orthonormal eigenvectors and $m \ll n$. Furthermore, let $\hat{\mathbf{x}} \in \mathbb{R}^n$ and $\{\hat{\mathbf{x}}_i \in \mathbb{R}^n, i = 1, \dots, c\}$ be the reconstructed images of them, that is,

$$\hat{\mathbf{x}} = \mathbf{W} \mathbf{y}, \quad (8)$$

$$\hat{\mathbf{x}}_i = \mathbf{W} \mathbf{y}_i, \quad i = 1, \dots, c. \quad (9)$$

Notice that

$$\mathbf{y} = \mathbf{W}^T \hat{\mathbf{x}}, \quad (10)$$

$$\mathbf{y}_i = \mathbf{W}^T \hat{\mathbf{x}}_i, \quad i = 1, \dots, c \quad (11)$$

are also satisfied and the orthonormal eigenvectors \mathbf{W} is the orthonormal “basis” for the reconstructed image space. The recognition task using the reconstructed images is to identify i^* by finding a representative $\hat{\mathbf{x}}_i$ with minimal distance to the query image $\hat{\mathbf{x}}$ in the reconstructed space, i.e.,

$$i^* = \arg \min_i \|\hat{\mathbf{x}} - \hat{\mathbf{x}}_i\|_2 \quad (12)$$

Using a property of the orthonormal basis:

$$\|\hat{\mathbf{x}}\|_2 = \|\mathbf{W}^T \hat{\mathbf{x}}\|_2, \quad (13)$$

the Eq. 12 is

$$i^* = \arg \min_i \|\mathbf{W}^T (\hat{\mathbf{x}} - \hat{\mathbf{x}}_i)\|_2 \quad (14)$$

$$= \arg \min_i \|\mathbf{y} - \mathbf{y}_i\|_2. \quad (15)$$

Therefore, the recognition task using the reconstructed images is equivalent with the one using the PCA projected points without reconstruction.

The result of occluded face shows another interesting capability of the PCA as a interpolation or a noise removal method.

3. FISHERFACE

Since the learning set is labeled, it makes sense to use this information to build a more reliable method for reducing the dimensionality of the feature space. *Fisherface* method [1] uses a class specific linear method, Fisher’s Linear Discriminant (FLD) [5], for dimensionality reduction and simple classifiers

¹One representative from one class is the minimum requirement for the nearest neighbor classification.

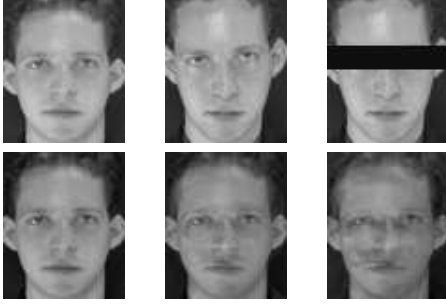


Fig. 3. Top row: the original images. Bottom row: the reconstructed images.

in the reduced feature space. This method selects W in [2] in such a way that the ratio of the between-class scatter and the within class scatter is maximized.

Again, let us consider a set of N sample images $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ taking values in an n -dimensional image space, and assume that each image belongs to one of c classes $\{X_1, X_2, \dots, X_c\}$. Let the between-class scatter matrix be defined as

$$S_B = \sum_{i=1}^c N_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (16)$$

and the within-class scatter matrix be defined as

$$S_W = \sum_{i=1}^c \sum_{\mathbf{x}_k \in X_i} (\mathbf{x}_k - \mu_i)(\mathbf{x}_k - \mu_i)^T \quad (17)$$

where μ_i is the mean image of class X_i , N_i is the number of samples in class X_i , and μ is the mean image of all samples. If S_W is nonsingular, the optimal projection W_{opt} is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples, i.e.,

$$\begin{aligned} W_{\text{opt}} &= \arg \max_W \frac{|W^T S_B W|}{|W^T S_W W|} \\ &= [\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_m] \end{aligned} \quad (18)$$

where $\{\mathbf{w}_i | i = 1, 2, \dots, m\}$ is the set of generalized eigenvectors of S_B and S_W corresponding to the m largest generalized eigenvalues $\{\lambda_i | i = 1, 2, \dots, m\}$, i.e.,

$$S_B \mathbf{w}_i = \lambda_i S_W \mathbf{w}_i, \quad i = 1, 2, \dots, m. \quad (19)$$

Note that there are at most $c - 1$ nonzero generalized eigenvalues, and so an upper bound on m is $c - 1$, where c is the number of classes. See [3].

To illustrate the benefits of class specific linear projection, a low dimensional analogue to the classification problem in which the samples from each class lie near a linear subspace

is shown. Fig. 4 is a comparison of PCA and FLD for a two-class problem in which the samples from each class are randomly perturbed in a direction perpendicular to a linear subspace. For this example, $N = 20$, $n = 2$, and $m = 1$. So, the samples from each class lie near a line passing through the origin in the 2D feature space. Both PCA and FLD have been used to project the points from 2D down to 1D. Comparing the two projections in the figure, *PCA actually smears the classes together* so that they are no longer linearly separable in the projected space. It is clear that, although PCA achieves larger total scatter, FLD achieves greater between-class scatter, and, consequently, classification is simplified.

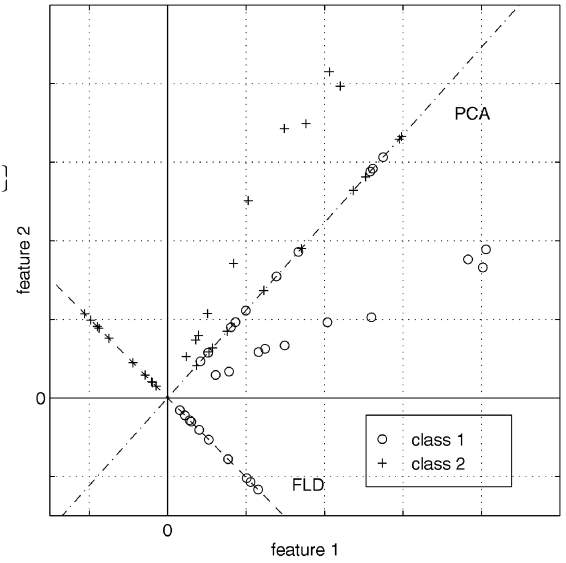


Fig. 4. A comparison of principal component analysis (PCA) and Fisher's linear discriminant (FLD) for a two class problem where data for each class lies near a linear subspace.

In the face recognition problem, one is confronted with the difficulty that the within-class scatter matrix $S_W \in \mathbb{R}^{n \times n}$ is always singular. This stems from the fact that the rank of S_W is at most $N - c$, and, in general, the number of images in the learning set N is much smaller than the number of pixels in each image n . This means that it is possible to choose the matrix W such that the within-class scatter of the projected samples can be made exactly zero.

In order to overcome the complication of a singular S_W , an alternative to the criterion in (18) is proposed. This method, named *Fisherfaces*, avoids this problem by projecting the image set to a lower dimensional space so that the resulting within-class scatter matrix S_W is nonsingular. This is achieved by using PCA to reduce the dimension of the feature space to $N - c$, and then applying the standard FLD defined by (18) to reduce the dimension to $c - 1$. More formally, W_{opt} is given by

$$W_{\text{opt}}^T = W_{\text{fld}}^T W_{\text{pca}}^T \quad (20)$$

where

$$W_{\text{pca}} = \arg \max_W |W^T S_T W| \quad (21)$$

$$W_{\text{fld}} = \arg \max_W \frac{|W^T W_{\text{pca}}^T S_B W_{\text{pca}} W|}{|W^T W_{\text{pca}}^T S_W W_{\text{pca}} W|} \quad (22)$$

Note that the optimization for W_{pca} is performed over $n \times (N - c)$ matrices with orthonormal columns, while the optimization for W_{fld} is performed over $(N - c) \times m$ matrices with orthonormal columns. In computing W_{pca} , we have thrown away only the smallest c -1 principal components.

4. EXPERIMENTAL RESULTS

The PIE database is used in the experiments. The PIE database contains 21 face images of 68 people having a large illumination variations. A few sample images from PIE database are shown in Fig. 5. The former 16 images of each person are used as a training set and the latter 5 images are used as a testing set.



Fig. 5. Examples from PIE database. Original image size 48x40.

The Fig. 6 shows a comparison of the Eigenface and Fisherface methods with respect to the recognition rate versus number of feature dimensions used. The Fig. 7 shows an evaluation based on cumulative match score (CMS) [6] using 40 dimensional feature vectors.

5. CONCLUSION

The Eigenface and Fisherface method were investigated and compared. The comparative experiment showed that the Fisherface method outperformed the Eigenface method. The usefulness of the Fisherface method under varying illumination was verified.

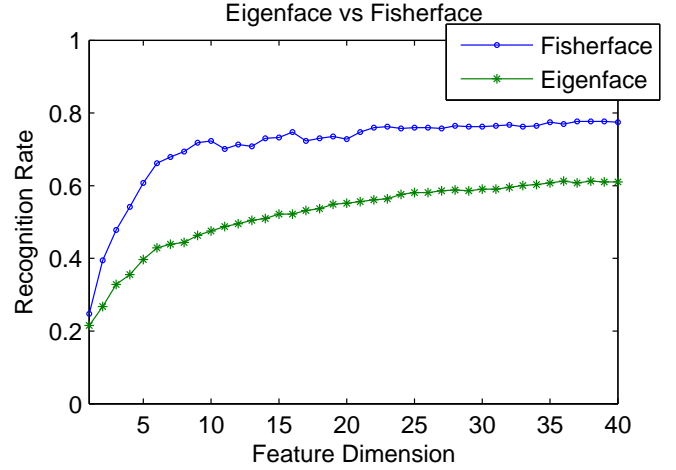


Fig. 6. Face recognition experiment: Eigenface v.s. Fisherface

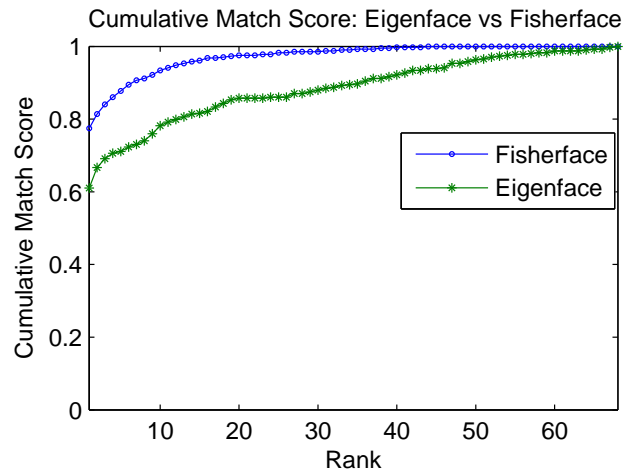


Fig. 7. Cumulative Match Score: Eigenface v.s. Fisherface. 40 dimensional features are used.

6. REFERENCES

- [1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, *Eigenfaces vs Fisherfaces Recognition Using Class Specific Linear Projection*, European Conf. Computer Vision (1996), 45–58.
- [2] R. Chellappa, C. Wilson, and S. Sirohey, *Human and Machine Recognition of Faces: A Survey*, Proc. IEEE **83** (1995), no. 5, 705–740.
- [3] R. Duda and P. Hart, *Pattern classification and scene analysis*, Wiley, 1973.
- [4] Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern classification*, Wiley, 2001.
- [5] R. A. Fisher, *The Use of Multiple Measures in Taxonomic Problems*, Ann. Eugenics **7** (1936).
- [6] P. J. Phillips, P. Rauss H. Moon, and S. A. Rizvi, *The feret evaluation methodology for face-recognition algorithms*, CVPR (1997), 137–143.
- [7] D.J. Kriegman P.N. Belhumeur, J.P. Hespanha, *Eigenfaces vs. fisherfaces: Recognition using class specific linear projection*, IEEE Transactions on Pattern Analysis and Machine Intelligence **19** (1997), no. 7, 711–720.
- [8] F. Samaria and A. Harter, *Parameterisation of a stochastic model for human face identification*, 2nd IEEE Workshop on Applications of Computer Vision (1994), <http://www.cl.cam.ac.uk/Research/DIG/attarchive/facedatabase.html>.
- [9] L. Sirovitch and M. Kirby, *Low-dimensional procedure for the characterization of human faces*, J Optical Soc of Am. A **2** (1987).
- [10] M. Turk and A. Pentland., *Eigenfaces for recognition*, Journal of Cognitive Neuroscience **3** (1991), no. 1, 71–86.
- [11] M. Turk and A. Pentland, *Eigenfaces for recognition*, J. Cognitive Neuroscience **3** (1991), no. 1.
- [12] ———, *Face recognition using eigenfaces*, Proc. IEEE Conf. on Computer Vision and Pattern Recognition (1991), 586–591.