Enhanced Fisherfaces for Robust Face Recognition

Juneho Yi, Heesung Yang, and Yuho Kim

School of Electrical and Computer Engineering
Sungkyunkwan University
300, Chunchun-dong, Jangan-gu
Suwon 440-746, Korea

Abstract. This research features a new method for automatic face recognition robust to variations in lighting, facial expression and eyewear. The new algorithm named SKKUfaces (Sungkyunkwan University faces) employs PCA (Principal Component Analysis) and FLD (Fisher's Linear Discriminant) in series similarly to Fisherfaces. The fundamental difference is that SKKUfaces effectively eliminates, in the reduced PCA subspace, portions of the subspace that are responsible for variations in lighting and facial expression and then applies FLD to the resulting subspace. This results in superb discriminating power for pattern classification and excellent recognition accuracy. We also propose an efficient method to compute the between-class scatter and within-class scatter matrices for the FLD analysis. We have evaluated the performance of SKKUfaces using YALE and SKKU facial databases. Experimental results show that the SKKUface method is computationally efficient and achieves much better recognition accuracy than the Fisherface method [1] especially for facial images with variations in lighting and eyewear.

1 Introduction

In face recognition, a considerable amount of research has been devoted to the problem of feature extraction for face classification that represents the input data in a low-dimensional feature space. Among representative approaches are Eigenface and Fisherface methods. Eigenface methods [7] [9] are based on PCA and use no class specific information. They are efficient in dimensionality reduction of input image data, but only provides us with feature vectors that represent main directions along which face images differ the most. On the other hand, Fisherface methods [1] [5] are based both PCA and FLD [10]. They first use PCA to reduce the dimension of the feature space and then applies the standard FLD in order to exploit class specific information for face classification. It is reported that the performance of Fisherface methods is far better in recognition accuracy than that of Eigenface methods.

The analysis of our method is similar to the Fisherface method suggested in [1]. The fundamental difference is that we apply FLD to a reduced subspace that is more appropriate for classification purpose than the reduced PCA subspace

S.-W. Lee, H.H. Bülthoff, T. Poggio (Eds.): BMCV 2000, LNCS 1811, pp. 502–511, 2000.

that Fisherface methods use. It has been suggested in the PCA based methods such as Eigenfaces that by discarding the three most significant principal components, the variation due to lighting is reduced [1]. However, this idea in concert with FLD has not been employed. We apply FLD to the reduced subspace that is computed by ignoring the first few eigenvectors from PCA corresponding to the top principal eigenvalues as illustrated in Figure 2. The effect is that, in this reduced subspace, portions of the vector space that are responsible for variations in lighting and facial expression are effectively eliminated. The reduced subspace is more appropriate for the FLD analysis than the reduced PCA subspace that Fisherfaces employ. That is, class separability is improved, and applying FLD to this reduced subspace can improve the discriminating power for pattern classification. Another important contribution of SKKUfaces is an efficient method to compute the between-class scatter and within-class scatter matrices.

We have evaluated our method using YALE and SKKU (Sungkyunkwan University) facial databases and have compared the performance of SKKUfaces with that of Fisherfaces. Experimental results show that our method achieves much better recognition accuracy than the Fisherface method especially for facial images with variations in lighting. In addition, a class separability measure computed for SKKUfaces and Fisherfaces shows that SKKUfaces has more discriminating power for pattern classification than Fisherfaces.

This paper is organized as follows. The following section briefly reviews Eigenface and Fisherface approaches. In section 3, we present our approach to feature extraction for robust face recognition and also describe a computationally very efficient method to compute within-class scatter and between-class scatter matrices. Section 4 presents experimental results using YALE and SKKU (Sungkyunkwan University) facial databases.

2 Related Works

2.1 Eigenface Method

Eigenface methods are based on PCA (or Karhunen-loeve transformation) that generates a set of orthonormal basis vectors. These orthonormal basis vectors are known as principal components that capture the main directions which face images differ the most. A face image is represented as a coordinates in the orthonormal basis. Kirby and Sirovish [7] first employed PCA for representing face images and PCA was used for face recognition by Turk and Pentland [2]. Eigenface methods are briefly described as follows.

Let a face image be a two-dimensional M by N array of intensity values. This image can be represented a vector X_i of dimension MN. Let $\mathbf{X} = [\mathrm{X}_1, \mathrm{X}_2, \cdots, \mathrm{X}_T]$ be the sample set of the face images. T is the total number of the face images. After subtracting the total mean denoted by Φ from each face image, we get a new vector set $\mathbf{\Phi} = [\mathrm{X}_1 - \Phi, \mathrm{X}_2 - \Phi, \cdots, \mathrm{X}_T - \Phi]$. Let Φ_i denote $\mathrm{X}_i - \Phi$. Then the covariance matrix is defined as:

$$\begin{aligned} S_T &= \textstyle\sum_{i=1}^{T} \Phi_i \Phi_i^T \\ &= \mathbf{\Phi}\mathbf{\Phi}^T . \end{aligned} \tag{1}$$

The eigenvector and eigenvalue matrices, Ψ, Λ are computed as:

$$S_T\Psi = \Psi\Lambda. \tag{2}$$

The size of the matrix, S_T is MNxMN and determining the MN eigenvectors and eigenvalues is an intractable task for typical image sizes. A computationally feasible method that employs the eigenanalysis of $\mathbf{\Phi}^T\mathbf{\Phi}$ instead of $\mathbf{\Phi}\mathbf{\Phi}^T$ is used [2]. The size of $\mathbf{\Phi}^T\mathbf{\Phi}$ is TxT.

$$(\mathbf{\Phi}^T\mathbf{\Phi})V = V\Lambda^{'} \tag{3}$$

$\mathbf{V} = [\mathrm{V}_1, \mathrm{V}_2, \cdots, \mathrm{V}_T]$ and $\Lambda^{'} = diag(\lambda_1, \lambda_2, \cdots, \lambda_T)$. Premultiplying $\mathbf{\Phi}$ on both sides, we have

$$\mathbf{\Phi}(\mathbf{\Phi}^T\mathbf{\Phi})V = (\mathbf{\Phi}\mathbf{\Phi}^T)(\mathbf{\Phi}V) = (\mathbf{\Phi}V)\Lambda^{'} \tag{4}$$

and $\mathbf{\Phi}V$ is the eigenvector matrix of $\mathbf{\Phi}\mathbf{\Phi}^T$. Assuming λ_i's are sorted as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_T$, we obtain eigenvectors of $\mathbf{\Phi}\mathbf{\Phi}^T$ corresponding to the first largest m eigenvalues as follows. These eigenvectors constitute the projection matrix W_{pca}

$$W_{pca} = [\mathbf{\Phi}\mathrm{V}_1, \mathbf{\Phi}\mathrm{V}_2, \cdots, \mathbf{\Phi}\mathrm{V}_m]. \tag{5}$$

$\mathbf{\Phi}\mathrm{V}_1, \mathbf{\Phi}\mathrm{V}_2, \cdots, \mathbf{\Phi}\mathrm{V}_m$ are refered to as eigenfaces. Refer to Figure 1 for an example of eigenfaces. A vector X_i that represents a face image is projected to a vector Y_i in a vector space of dimension, m using the following equation.

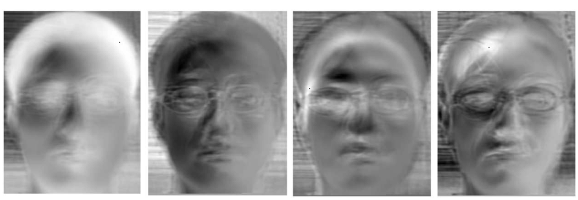

Fig. 1. The first four eigenfaces computed from SKKU facial images

$$\mathrm{Y}_i = W_{pca}^T(\mathrm{X}_i - \Phi) \tag{6}$$

A new face image X_i is recognized by comparison of Y_i with the projected vectors of the training face images that are computed off-line. Since PCA maximizes for all the scatter, it is more appropriate for signal representation rather than for recognition purpose.

2.2 Fisherface Method

The idea of the Fisherface method is that one can perform dimensionality reduction using W_{pca} and still preserve class separability. It applies FLD to the reduced PCA subspace to achieve more reliability for classification purpose. The Fisherface method is briefly described as follows. Let $\omega_1, \omega_2, \cdots, \omega_c$ and $N_1, N_2, \cdots, N_c$ denote the classes and the number of face images in each class, respectively. Let $\mathrm{M}_1, \mathrm{M}_2, \cdots, \mathrm{M}_c$ and M be the means of the classes and the total mean in the reduced PCA subspace. Since $\mathrm{Y}_{ij} = W_{pca}^T \mathrm{X}_{ij}$, we can then have $\mathrm{M}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathrm{Y}_{ij} = W_{pca}^T (\frac{1}{N_i} \sum_{j=1}^{N_i} \mathrm{X}_{ij})$. X_{ij} denotes the j^{th} face image vector belonging to the i^{th} class (i. e. subject). The between-class scatter and within-class scatter matrices $S_b^{'}$ and $S_w^{'}$ of Y_{ij}'s are expressed as follows.

$$S_b^{'} = \sum_{i=1}^{C} N_i (\mathrm{M}_i - \mathrm{M})(\mathrm{M}_i - \mathrm{M})^T = W_{pca}^T S_b W_{pca} \tag{7}$$

$$S_w^{'} = \sum_{i=1}^{C} \frac{1}{N_i} \sum_{j=1}^{N_i} (\mathrm{Y}_{ij} - \mathrm{M}_i)(\mathrm{Y}_{ij} - \mathrm{M}_i)^T = W_{pca}^T S_w W_{pca} \tag{8}$$

S_b and S_w denote the between-class scatter and within-class scatter matrices of X_{ij}'s, respectively. The projection matrix W that maximizes the ratio of the determinant, $\frac{|W^T S_b^{'} W|}{|W^T S_w^{'} W|}$ is chosen as the optimal projection, W_{fld}. The columns of W_{fld} are computed as the (C-1) leading eigenvectors of the matrix $(S_w^{'})^{-1} S_b^{'}$ [11] where C denotes the number of classes. For recognition, given an input face image X_k, it is projected to $\Omega_k = W_{fld}^T W_{pca}^T \mathrm{X}_k$ and classified by comparison with the vectors Ω_{ij}'s that were computed off-line from a set of training face images.

3 SKKUfaces

3.1 SKKUface Method

The SKKUface method proposed in this research is illustrated in Figure 2. It is similar to Fisherface methods in that it applies PCA and FLD in series. Our algorithm is different from Fisherface methods in that face variations due to lighting, facial expression and eyewear are effectively removed by discarding the first few eigenvectors from the results of PCA, and then apply FLD to the reduced subspace to get the most class separability for face classification. The result is an efficient feature extraction that carries only features inherent in each face, excluding other artifacts such as changes in lighting and facial expression. Classification of faces using the resulting feature vectors leads to a considerably improved recognition accuracy than Fisherface methods.

As illustrated in Figure 2, we apply FLD to the reduced subspace that is computed by ignoring the first few eigenvectors corresponding to the top principal eigenvalues. For the experimental results, we have only discarded the first eigenvector. Another important contribution of SKKUfaces is the efficient computation of the between-class scatter and within-class scatter matrices $S_b^{'}$ and $S_w^{'}$ of Y_{ij}. The following section describes the method.

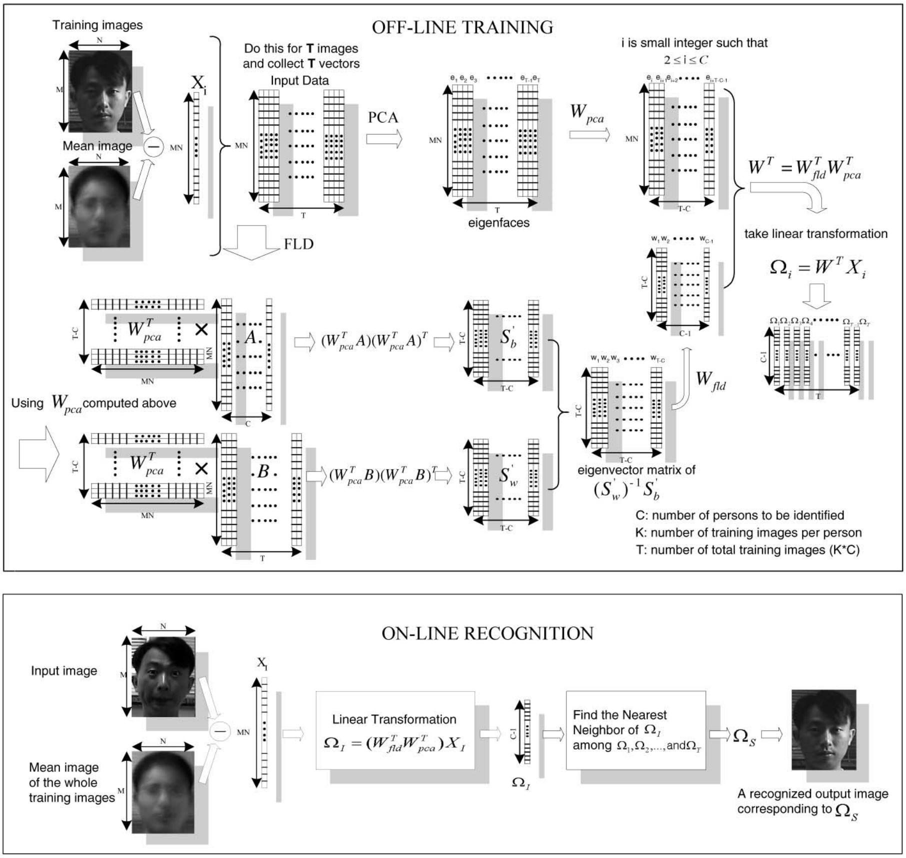

Fig. 2. The overview of the SKKUface method

3.2 Efficient Computation of Within-class Scatter and Between-class Scatter Matrices

After dimensionality reduction of the face vector space by the linear projection, W_{pca}, we need to compute the within-class scatter and between-class scatter matrices, $S_w^{'}$ and $S_b^{'}$ to apply the Fisher linear discriminant analysis to the reduced subspace. The resulting projection matrix, W_{fld} consists of columns of eigenvectors of $(S_w^{'})^{-1}S_b^{'}$ corresponding to the largest (C-1) leading eigenvalues. In computing $S_w^{'}$ and $S_b^{'}$ represented by $W_{pca}^T S_w W_{pca}$ and $W_{pca}^T S_b W_{pca}$, respectively, we do not explicitly evaluate S_w and S_b. The size of the matrices, S_w and S_b, is MNxMN and it is an intractable task to compute them for typical image sizes. On the other hand, S_b can be expressed using equation (9) assuming the same size of each class.

$$\begin{aligned} S_b &= \textstyle\sum_{i=1}^{C} \frac{1}{C}(\mathrm{M}_i - \mathrm{M})(\mathrm{M}_i - \mathrm{M})^T \\ &= \tfrac{1}{C}(\mathrm{M}_1 - \mathrm{M})(\mathrm{M}_1 - \mathrm{M})^T + \cdots + \tfrac{1}{C}(\mathrm{M}_c - \mathrm{M})(\mathrm{M}_c - \mathrm{M})^T \\ &= \tfrac{1}{C}\left[\mathrm{M}_1 - \mathrm{M}, \mathrm{M}_2 - \mathrm{M}, \cdots \mathrm{M}_C - \mathrm{M}\right] \begin{bmatrix} (\mathrm{M}_1 - \mathrm{M})^T \\ (\mathrm{M}_2 - \mathrm{M})^T \\ \vdots \\ (\mathrm{M}_C - \mathrm{M})^T \end{bmatrix} \\ &= \tfrac{1}{C} A A^T \end{aligned} \tag{9}$$

where $A = [\mathrm{M}_1 - \mathrm{M}, \mathrm{M}_2 - \mathrm{M}, \cdots \mathrm{M}_c - \mathrm{M}]$ and M_i, M denote the i^{th} class mean and the total mean, respectively. $\frac{1}{C}$ is *prior* probability that represents the size of each class.

Since $MN \gg C$, we can save a huge amount of computation by using the matrix A of size MNxC matrix rather than directly dealing with S_b of size MNxMN. Finally, $S_b^{'}$ is obtained using the following equation.

$$S_b^{'} = W_{pca}^T S_b W_{pca} = W_{pca}^T A A^T W_{pca} = (W_{pca}^T A)(A^T W_{pca}) \tag{10}$$

Notice that $S_b^{'}$ is simply computed by multiplication of $W_{pca}^T A$ and its transpose. Similarly, $S_w^{'}$ can be written as follows.

$$\begin{aligned} S_w &= \textstyle\sum_{i=1}^{C} \sum_{j=1}^{N_i} (\mathrm{X}_{ij} - \mathrm{M}_i)(\mathrm{X}_{ij} - \mathrm{M}_i)^T \\ &= [\mathrm{K}_{11}, \cdots, \mathrm{K}_{21}, \cdots, \mathrm{K}_{CN_C}] \begin{bmatrix} \mathrm{K}_{11}^T \\ \vdots \\ \mathrm{K}_{21}^T \\ \vdots \\ \mathrm{K}_{CN_C}^T \end{bmatrix} \\ &= BB^T \end{aligned} \tag{11}$$

$\mathrm{K}_{ij} = \mathrm{X}_{ij} - \mathrm{M}_{ij}$ and $B = [\mathrm{K}_{11}, \cdots, \mathrm{K}_{21}, \cdots, \mathrm{K}_{CN_C}]$. $S_w^{'}$ is computed as:

$$S_w^{'} = W_{pca}^T S_w W_{pca} = W_{pca}^T B B^T W_{pca} = (W_{pca}^T B)(B^T W_{pca}) \tag{12}$$

The size of matrix B is MNxT and $MN \gg T$. We could save a lot of computational effort using the matrix, B. Similarly to $S_w^{'}$, $S_b^{'}$ is simply computed by multiplication of $W_{pca}^T B$ and its transpose.

Suppose $M = N = 256$, $C = 10$, $K = 15$. The explicit computation of S_b and S_w involves matrices of size 65536 x 65536. Employing the proposed methods involves computation using a 65536 x 10 matrix for $S_b^{'}$ and a 65536 x 150 matrix for $S_w^{'}$. This achieves about 6,500 times and 43 times less computation for $S_b^{'}$ and $S_w^{'}$, respectively.

4 Experimental Results

To assess the performance of SKKUfaces, the recognition rate of SKKUfaces is compared with that of Fisherfaces [1] using Yale facial database and SKKU facial database. The recognition rates were determined by the "leaving-one-out"method [11]. A face image is taken from the database for classification and all the images except this image are used for training the classifier. Classification was performed using a nearest neighbor classifier.

SKKU facial database contains ten different images of each of ten different subjects. The size of an image is 50 x 40. For a subject, five images out of ten images were taken first and the rest five images at a different time. All the images are frontal views of upright faces with changes in illumination, facial expression (open/closed eyes, smiling/nonsmiling/surprised), facial details (glasses/no glasses) and hair style. Refer to Figure 3 for the whole set of SKKU face images. In Yale facial database, each of sixteen different subjects have ten images which consist of three images under illumination changes, six with changes in facial expression and one with glasses worn. Figure 4 shows a set of images of a subject in Yale facial database.

Figures 5 and 6 show the relative performance of the algorithms when applied to SKKU facial database and Yale facial database, respectively. As can be seen in Figures 5 and 6, the performance of SKKUfaces is far better than that of Fisherfaces in the cases of variations in illumination and eyewear. This experimentally proves our claim that we apply FLD to a reduced subspace that is more appropriate for classification purpose than the reduced PCA subspace that Fisherface methods use. Application of FLD to this reduced subspace yields the better discriminating power for pattern classification and the recognition accuracy is far improved. The amount of computational saving we could benefit in computing $S_w^{'}$ and $S_b^{'}$ from the method proposed in section 3.2 is as follows.

Since $M = 50$, $N = 40$, $C = 10$, $K = 10$ in the case of SKKU facial database, directly evaluating with S_b and S_w should involve matrices of size 2000 x 2000. However, employing the proposed method only deals with a 2000 x 10 matrix for $S_b^{'}$ and a 2000 x 100 matrix for $S_w^{'}$, respectively. The saving amounts to about 200 times and 20 times less computation for $S_b^{'}$ and $S_w^{'}$, respectively.

5 Conclusion

We have proposed SKKUfaces for automatic face recognition robust to variations in lighting, facial expression and eyewear. In the reduced PCA subspace, SKKUfaces effectively removes portions of the vector space that are responsible for variations in lighting and facial expression, and applies FLD to this reduced subspace. The experimental results show that the discriminating power for pattern classification is considerably improved and excellent recognition accuracy is achieved. A study on the relationship between the number of eigenvectors to be discarded in the reduced PCA subspace and the degree of variations in lighting or facial expression will enable us to achieve the optimum performance of SKKUfaces.

Fig. 3. The whole set of SKKU facial images [13]

Fig. 4. Example images from Yale facial database [12]

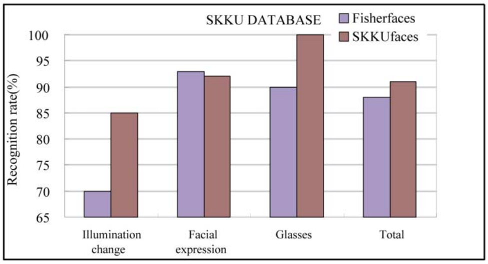

Fig. 5. The relative performance of the SKKUface and the Fisherface methods for SKKU facial images

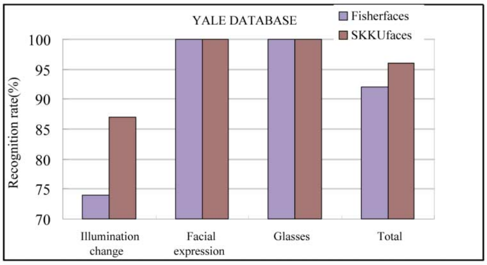

Fig. 6. The relative performance of the SKKUface and the Fisherface methods for Yale facial images

Acknowledgement

This work was supported by grant number 1999-2-515-001-5 from interdisciplinary research program of the KOSEF.

References

1. P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. on PAMI*, vol. 19, no. 7, pp. 711-720, 1997.
2. M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.
3. R. Brunelli and T. Poggio, "Face Recognition: Features vs. Templates," *IEEE Trans. on PAMI*, vol. 15, no. 15, pp. 1042-1052, 1993.
4. Shang-Hung Lin et al., "Face Recognition and Detection by Probabilistic Decision Based Neural Network," *IEEE Trans. on Neural Network*, vol. 8, no. 1, pp. 114-132, 1997.
5. Chengjun Liu and Harry Wechsler, "Enhanced Fisher Linear Discriminant Models for Face Recognition," Proceedings of the 14th International Conference on Pattern Recognition, vol. 2, pp. 1368-1372, 1998.
6. Rama Chellappa, Charles L. Wilson, and Saad Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proceedings of IEEE*, vol. 83, no. 5, 1995.
7. M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces," *IEEE Trans. on PAMI*, vol. 12, no. 1, pp. 103-108, 1990.
8. K. Etemad and R. Chellappa, "Discriminant Analysis for Recognition of Human faces image," *Journal of Optical Society of America*, vol. 14, no. 8, pp. 1724-1733, 1997.
9. A. Pentland, B. Moghaddam, T. Starner, and M. Turk, "View-Based and Modular Eigenspaces for Face Recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 84-91, 1994.
10. R. A. Fisher, "The Use of Multiple Measures in Taxonomic Problems," *Ann. Eugenics*, vol. 7, pp. 179-188, 1936.
11. K. Fukunaga, Introduction to Statistical Pattern Recognition. Academic Press, second edition, 1991.
12. http://cvc.yale.edu/projects/yalefaces/yalefaces.html
13. http://vulcan.skku.ac.kr/research/skkufaces.html