Proceedings of the 5th WSEAS Int. Conf. on SIGNAL, SPEECH and IMAGE PROCESSING, Corfu, Greece, August 17-19, 2005 (pp84-89)

A Novel Biometric Feature Extraction Algorithm using Two Dimensional Fisherface in 2DPCA subspace for Face Recognition

R. M. MUTELO, W.L. WOO, and S.S. DLAY
School of Electrical, Electronic and Computer Engineering
University of Newcastle
Newcastle upon Tyne, NE1 7RU
UNITED KINDOM

Abstract: - This paper describes a novel algorithm, 2D-FPCA, for face feature extraction and representation. The new algorithm fuses the two dimensional Fisherface method with the two dimensional principal component analysis (2DPCA). Our algorithm operates on the two dimensional image matrices. Therefore a total image covariance matrix can be constructed directly using the original image matrices and its eigenvectors are derived for feature extraction. Similarly, the between and the within image covariance matrices are constructed and transformed to a 2DPCA subspace. The result is that 2D-FPCA is faster and yields greater recognition accuracy. The ORL database is used as a benchmark. The new algorithm achieves a recognition rate of 95.50% compared to the recognition rate of 90.00% for the Fisherface method.

Key-Words: Biometrics, Recognition, Feature Extraction, Image Representation, two dimensional Fisherface, two dimensional principal component analysis.

1 Introduction

The early strategy for face identification was based on geometrical features such as nose width and length, mouth position and chin shape [1] almost all of the new approaches developed in recent years are based on holistic representations known as templates. In [2] a comparison of geometrical feature based matching with template (statistical features) matching is presented and the result favours the statistical feature based matching approach.

The most commonly used statistical representation for face recognition is the Eigenfaces [3], which uses principal component analysis (PCA) for dimensionality reduction. The produced templates are not always good for discrimination among classes. *Belhumeur et al.* [4] proposed Fisherfaces method, which is based on Fisher's Linear Discriminant and produces well separated classes in a low-dimensional subspace. Nonlinear PCA-related methods such as independent component analysis (ICA) and kernel principal component analysis (Kernel PCA) have been proposed for feature extraction [5-6]. These methods are superior to representations based on PCA. However, Kernel PCA and ICA are both computationally more expensive than PCA.

In all the PCA-based related techniques above, the 2D face image matrices must be transformed into 1D image vectors. The resulting image vectors of faces usually lead to a high dimensional image vector space, where it is difficult to evaluate the covariance matrix accurately due to its large size and the relatively small number of training samples. *Yang et al.* [7] proposed a straightforward image projection technique, called the image principal component analysis, which is directly based on analysis of original image matrices. Based on the idea of the image matrix *Li et al.* [8] proposed the two dimensional linear discriminant analysis where the Fisher linear projection criterion is used to find a good projection for better feature extraction.

In this paper, a straightforward image projection technique called two dimensional Fisherface Principal Component Analysis, 2D-FPCA, for feature extraction is presented. Our method uses Fisher linear projection criterion to find a good projection in the 2DPCA subspace. Performing 2DFLD in this new subspace has three important advantages over Fisherface. Firstly, it is easier to evaluate the covariance matrices accurately. Secondly, less time is required to determine the corresponding eigenvectors. Finaly, the face recognition accuracy increases significantly.

2 2D-FPCA

2.1 Idea and Algorithm

Suppose that there are C pattern classes in training set, $\{A_j\}_{j=1}^{N}$, where N denotes the size of training set. The *ith* class C_i has n_i samples, thus $N = \sum_{i=1}^{C} n_i$ .The mean image of all training set is

$\bar{A}$ and $\bar{A}_i, i = 1,....,C$ denotes the mean image of class C_i. The ith class and jth training image is denoted by A_j^i an $m \times n$ image matrix. X denotes an n-dimensional unitary column vector. Our novel idea is to project the jth image, A_j, onto X to the 2DPCA subspace by the following linear transformation [7], [9]:

$$Y_j = A_j X \quad (1)$$

We obtain a m dimensional projected vector which is the feature vector of the jth image A_j. We need to determine a good projection vector X. By definition, PCA finds a projection direction X, onto which all samples are projected, so that the total scatter of the resulting projected samples is maximized. The total covariance (scatter) matrix of the projected samples is characterized by the trace (tr) of the covariance matrix of the projected feature vectors as follows:

$$J(X) = tr(S_t) \quad (2)$$

Where $S_t = E(Y_j - EY_j)(Y_j - EY_j)^T \quad (3)$

$$= E[(A_j - EA_j)X][A_j - EA_j)X]^T$$

Hence $J(X) = tr(S_t) = X^T E[(A_j - EA_j)^T (A_j - EA_j)]X$.

We define the total image covariance matrix as:

$$G_t = E[(A_j - EA_j)^T (A - EA_j)] \quad (4)$$

Therefore, (2) can be written as,

$$tr(S_t) = X^T G_t X \quad (5)$$

The unitary projection vector, X, that maximizes (5) is called the *optimal projection axis*,(X_{opt}) which is the eigenvector of G_t. It is not enough to have only one optimal projection axis. We need to select a set of projection axes, $X_{1,..........}X_d$, in such a way that not much discriminant information is lost while noise is reduced. The selected eigenvectors, $X_{1,..........}X_d$, contain principal components (PCs) which retain both the within variations (unwanted information) and between variations (important discriminant information).

Once the 2DPCA subspace has been created we use the *Fisher's linear projection criterion* in this new space to find a good projection that maximizes the ratio of the scatter among the face classes to the scatter within the face of the same class. The between image class covariance (scatter) matrix S_B and the within image class covariance (scatter) matrix S_W in the input space can be transformed to the 2DPCA subspace as S_{BB} and S_{WW} by,

$$S_{WW} = X^T S_W X \quad (6)$$

$$S_{BB} = X^T S_B X \quad (7)$$

We derive S_{BB} and S_{WW} as follows:

$$S_{BB} = \sum_{i=1}^{C} n_i (\bar{Y}_i - \bar{Y})^T (\bar{Y}_i - \bar{Y}) \quad (8)$$

$$= \sum_{i=1}^{C} n_i [(\bar{A}_i - \bar{A})X]^T [(\bar{A}_i - \bar{A})X]$$

$$= X^T \left(\sum_{i=1}^{C} n_i (\bar{A}_i - \bar{A})^T (\bar{A}_i - \bar{A}) \right) X$$

Where $S_B = \sum_{i=1}^{C} n_i (\bar{A}_i - \bar{A})^T (\bar{A}_i - \bar{A}) \quad (9)$

Therefore $S_{BB} = X^T S_B X \quad (10)$

Similarly $S_{ww} = \sum_{i=1}^{C} \sum_{j=1}^{n_i} (Y_j^{(i)} - \bar{Y}^{(i)})^T (Y_j^{(i)} - \bar{Y}^{(i)}) \quad (11)$

$$= \sum_{i=1}^{C} \sum_{j=1}^{n_i} [(A_j^{(i)} - \bar{A}^{(i)})X]^T [(A_j^{(i)} - \bar{A}^{(i)})X]$$

$$= X^T \left(\sum_{i=1}^{C} \sum_{j=1}^{n_i} (A_j^{(i)} - \bar{A}^{(i)})^T (A_j^{(i)} - \bar{A}^{(i)}) \right) X$$

Where $S_w = \left(\sum_{i=1}^{C} \sum_{j=1}^{n_i} (A_j^{(i)} - \bar{A}^{(i)})^T (A_j^{(i)} - \bar{A}^{(i)}) \right) \quad (12)$

Similarly, $S_{WW} = X^T S_W X \quad (13)$

We refer to S_{BB} and S_{WW} as the *between feature* and *within feature class covariance (scatter)* matrices with dimensions of $d \times d$, where d is the number of eigenvectors of G_t that we choose. The main idea at this stage is to project Y_j onto the unitary column vector Z as follows:

$$R_j = Y_j Z \quad (14)$$

To measure the class separability of the projected samples, $R_j, j = 1,2,......N$, we determine the trace of the between S_{BBB} and S_{WWW} within matrix of the projected samples as follows:

$$J(Z) = \frac{tr(S_{BBB})}{tr(S_{WWW})} \quad (15)$$

Using a similar idea as in (2), $tr(S_{BBB})$ and $tr(S_{WWW})$ is expressed as:

$$J(Z) = \frac{Z^T S_{BB} Z}{Z^T S_{WW} Z}$$

This is the Fisher linear projection criterion. The unitary vector Z that maximizes (15) is called the Fisher optimal projection axis (Z_{opt}). It means that the projected image samples on the direction Z has the minimal within image class scatter and the maximal between image class scatter in the subspace spanned by Z. If S_{WW} is nonsingular, the solution to above optimization problem is to solve the generalized eigenvalue problem [3-4]:

$$S_{BB}Z_{opt} = \lambda S_{WW} Z_{opt} \quad (16)$$

$$S_{WW}^{-1} S_{BB} Z_{opt} = \lambda Z_{opt}$$

In (16), λ is the eigenvalue of $S_{WW}^{-1}S_{BB}$. In the traditional LDA, we were faced with the singularity problem. However, 2D-FPCA overcomes this problem successfully. *Li et al. [8]* showed that the within image class covariance (scatter) matrix S_W is non-singular in real situation. In general, it is not enough to have only one Fisher optimal projection axis. We usually need to select a set of projection axis, $Z_{1,} .... Z_g$, subject to the orthonormal constraints.

2.2 Feature Extraction

The optimal projection vectors of 2DPCA, $X_{1,} .... X_d$, are used for the first feature extraction. For a given image sample A_j,

$$Y_j = A_j X_k, k = 1,........d. \quad (17)$$

The principal component vectors obtained are used to form an $m \times d$ matrix $Y_j = [B_{1,} .... B_d]$, which is called the *feature matrix* or *feature image* of the jth image sample A_j. We refer to B_k as the *principal component (vectors)* of the jth sample A_j. Here we can think of Y_j as a smaller image matrix with less noise from jth sample image A_j. The *Fisher feature vectors* $T_{1,} .... T_g$, form a $m \times g$ *Fisher feature matrix* $R_j = [T_1, ... T_g]$ for the jth image sample A_j as follows:

$$R_j = Y_j Z_p, p = 1,........g. \quad (18)$$

It should be noted that each *Fisher feature* of Fisherface is a scalar, whereas the *Fisher feature* of 2D-FPCA is a vector.

2.3 Classification Method

After a transformation by Z, a Fisher feature matrix is obtained for each image. Then, a nearest neighbour classifier is used for classification. Here, the distance between two feature matrices, $R_1 = [T_1^1, T_2^1, ... T_g^1]$ and $R_2 = [T_1^2, T_2^2, ... T_g^2]$ is defined as

$$D(R_1, R_2) = \sum_{p=1}^{g} \left\| T_p^1 - T_p^2 \right\| \quad (19)$$

Where $\left\| T_p^1 - T_p^2 \right\|$ denotes the Euclidean distance between the two principal component vectors T_p^1 and T_p^2. Given a test sample R, if $D(R, R_7) = \min D(R, R_j)$ and $R_7 \in C_k$, we decide that $R \in C_k$.

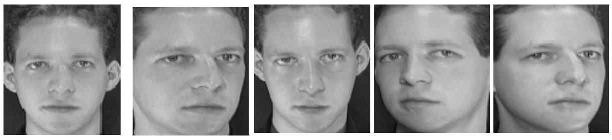

Fig. 1 Five sample images of one subject in the ORL database

3 Image Reconstruction

Image reconstruction for the 2D-FPCA can be performed in the following way. The original image can be reconstructed using (17) and (18).

From (17), $Y_j = A_j X_k, k = 1,........d.$

Let $U = [X_{1,} .... X_d]$, then (17) can be written as

$$Y_j = A_j U \quad (20)$$

Since $X_{1,} .... X_d$ are orthonormal, it is easy to obtain the reconstructed image of the jth sample A_j:

$$\tilde{A}_j = Y_j U^T = \sum_{k=1}^{d} B_k X_k^T \quad (21)$$

But Y_j needs to be obtained from (18), let $V = [Z_1, ... Z_g]$. We can rewrite (18) as:

$$\tilde{Y}_j = R_j V^T = \sum_{p=1}^{g} T_p Z_p^T \quad (22)$$

Hence, $\tilde{A}_j = R_j V^T U^T$, which is the same size as the jth image sample A_j and represents the reconstructed sub images A_j. That is, the jth image A_j can be approximately reconstructed by adding up the first g and first d sub images. In particular, when the selected number of principal component vectors $g = d$ and $d = n$, we have $\tilde{A}_j = A_j$, i.e., the image is completely reconstructed by its principal component vectors without any loss of information. Otherwise, if either $g < d$ or $d < n$, the reconstructed image $\tilde{A}_j$ is an approximation for A_j.

4 Analysis

The new 2D-FPCA method was used for face recognition. The performance of our proposed algorithm was tested on a well-known face image database from the Olivetti Research Laboratory (ORL database). The ORL database [10] was used to evaluate the performance of the 2D-FPCA under conditions where the pose and number of 2DPCA eigenvectors varied.

4.1 Analysis using the ORL benchmark Database

This publicly available database comprises of 40 individuals, each providing 10 different images. For some subjects, the images were taken at different times. Furthermore, other within class variations such as facial expression e.g. eyes open, eyes closed, smiling, not smiling varied. Occlusion due to other accessories e.g. glasses was used for some subjects. All the images were taken against a dark background with the subjects in an upright, frontal position with tolerance for some side movement. The images were taken with a tolerance for tilting and rotation of the face of up to 20 degrees. The size of each image is 112×92 pixels, with 256 grey levels per pixel. Five samples from the ORL database are shown in Fig. 1

First, a face recognition experiment was performed using five image samples per class for training, and the remaining images for testing. The images were normalized to 56×46 pixels. A total number of training samples and testing samples were both 200. The 2D-FPCA algorithm was used for feature extraction. Here, the size of the total image covariance G_t , within image covariance S_W , and between image covariance S_B is $n \times n$ matrix, so it was relatively easy to evaluate the eigenvectors of G_t . We chose the eigenvectors corresponding to the nine (d = 9) largest eigenvalues, $X_{1,}.........X_d$, as projection axis. After the projection of the jth image sample A_j onto these axes using (17), we obtain nine principal component vectors $B_{1,}.........B_d$. We then transformed S_W and S_B with the same eigenvectors using (6) and (7) to obtain S_W and S_B in the 2DPCA subspace. These new $d \times d$ covariance matrices are called the *within feature and between feature class covariance matrices*, S_{WW} and S_{BB} . It is therefore relatively easier to calculate the eigenvectors of S_{WW} and S_{BB} compared to S_W and S_B . This results in faster and more accurate evaluation of S_{BB} and S_{WW} . The evaluation of G_t does not have to be as accurate as the evaluation of S_{BB} and S_{WW} , rather it is important to be able to keep a set of eigenvectors, $X_{1,}.........X_d$, that will optimize the evaluation of (16) without a loss of much information. The eigenvectors of (16) contain discriminant information which is essential in face classification. In this analysis we kept only five eigenvectors, $Z_{1,}......Z_g$, from (16) corresponding to the five (g = 5) largest eigenvalues. Finally, the Fisher features vectors are extracted using (18) and are used to form a $m \times g$ matrix R_j .

The magnitude of the eigenvalues obtained for the 2DPCA and for the 2DFLD in the subspace is plotted in decreasing order in Fig. 2 and Fig. 3.

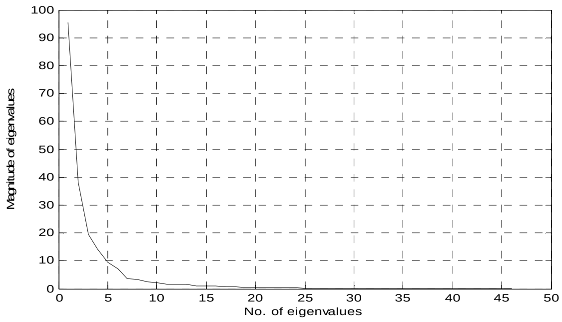

Fig. 2 The magnitude of the eigenvalues of 2DPCA in descending order

Fig. 2 shows that the magnitude of the eigenvalues quickly converges to zero, which is exactly consistent with the results in [3, 4, and 7]. Fig. 3 shows similar characteristics although it does not converge to zero in this case. Therefore, we can conclude that the energy of an image is concentrated on its first small number of component vectors. It is reasonable to use these component vectors to represent the image for identity recognition purposes.

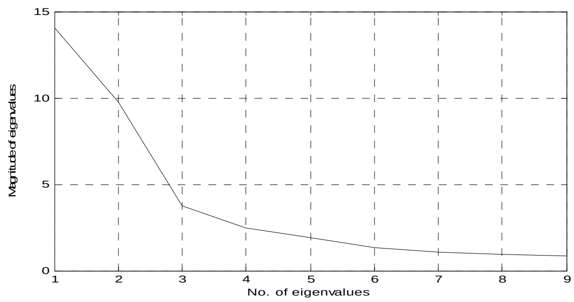

Fig. 3 The magnitude of the eigenvalues for 2DFLD in descending order

We evaluated the performance of 2D-FPCA under conditions where the number of eigenvectors for 2DPCA ($X_{1,}.........X_d$) is varied e.g. d = 5, 7, 9, 11, 13. The recognition results are shown in Fig. 4. The graphs indicates that the performance of 2D-FPCA is optimized when d = 9. The accuracy decreases for values of d > 9 and d < 9. 2D-FPCA achieves a top recognition of 95.50 % at it's optimum value (d =9) using five Fisher feature vectors. The lowest top recognition accuracy of 92.50 % was obtained when d = 5 using five Fisher feature vectors.

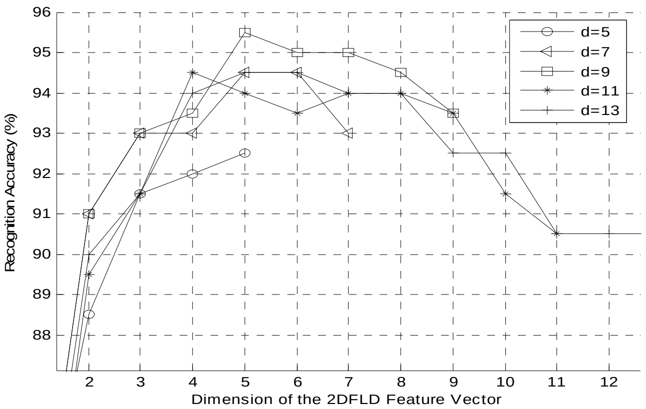

Fig. 4 The plot of the recognition accuracy (%) versus the dimension of the feature vector using in classification for varying d .

We further compared the performance of Fisherface and 2D-FPCA for d = 9. 2D-FPCA and the Fisherface method were used for feature extraction. Finally, a nearest neighbour classifier was employed for classification. For 2D-FPCA, (19) was used to calculate the distance between two feature matrices. In Fisherface, the common Euclidean distance measure is adopted. Fig. 5 presents the recognition accuracy plotted against the dimensions of the feature matrix used in classification for 2D-FPCA and for the Fisherface approach. 2D-FPCA performed better than the Fisherface method. A top recognition accuracy of 95.50 % for the 2D-FPCA was achieved and 90.00 % for the Fisherface. The 2D-FPCA method is also superior to Fisherface in terms of computational efficiency for feature extraction since in the Fisherface approach G_t is a $mn \times mn$ matrix compared to a $n \times n$ matrix for the 2D-FPCA. It is hard work to calculate the eigenvectors of a $mn \times mn$ matrix where $mn = 2576$ and $n = 46$ in this case.

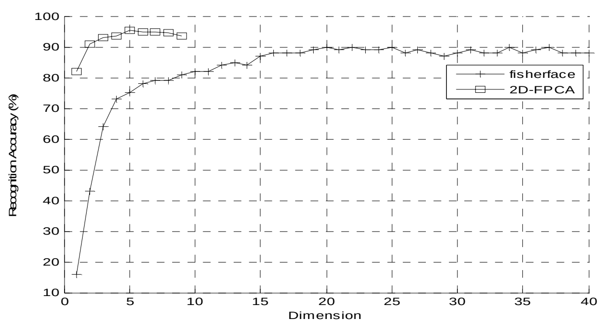

Fig. 5 Perfomance of 2D-FPCA and Fisherface.

Furthermore, the Fisherface approach selects $d = N - C$ eigenvectors (N is the number of training images and C is the number of classes) from G_t to make S_W non-singular whereas in 2D-FPCA we seek a value of d that optimizes the recognition accuracy while maintaining a faster speed. In the Fisherface approach S_{BB} and S_{WW} are of the order $d = N - C$ which can still be a large number depending on the size of the training sample compared to the optimal value of d for 2D-FPCA. In this analysis the former was a 160×160 matrix compared to a 9×9 matrix for 2D-FPCA. When compared to recently published results [7-13], 2D-FPCA has superior performance.

4 Conclusion

A novel technique for image feature extraction and representation was developed using the two dimensional Fisher's Linear Discriminant (2DFLD) combined with 2DPCA subspace. 2D-FPCA has many advantages over Fisherface. Firstly, since 2D-FPCA is based on the image matrix, it is simpler and more straightforward to use for image feature extraction. Secondly, 2D-FPCA outperforms Fisherface in terms of recognition accuracy as it achieves a high recognition accuracy of 95.50% compared to 90.00% for Fisherface. Thirdly, 2D-FPCA selects a smaller set of 2DPCA's eigenvectors that optimizes the accuracy and speed. Finally, 2D-FPCA is computationally more efficient than the Fisherface and it improves the speed of image feature generation significantly.

References:

[1] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face recognition: A literature survey," *CVL Technical Report, University of Maryland*, Oct.2000,

[2] R Brunelli and T Pogio, "Face Recognition through Geometrical Features," *Proc. European Conference on Computer Vision, pp 792-800, 1992*

[3] M. A. Turk and A. P. Pentland, "Eigenfaces for recognition," *J. Cognitive Neurosci.*, vol. 3, pp. 71–86, 1991.

[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces versus fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intell., vol. 19, pp. 711–720, 1997.*

[5] S. Bartlett, J. R. Movellan, T. J. sejnowski, "Face Recognition by Independent Component Analysis", IEEE Trans. On Neural Networks, Vol. 13, No. 6, Nov. 2002

[6] Ming-Hsuan Yang, "Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods," in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, Washington D. C., May 2002, pp. 215–220.

[7] J. Yang, D. Zhang, A. F. Frangi, J. Y. Yang, Two dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Trans. Pattern Anal. Machine Intell. Vol. 26, no. 1, pp. 131–137, 2004.*

[8] Ming Li and Baozong Yuan, 2D-LDA: A statistical linear discriminant analysis for image matrix, *Pattern Recognition Letters, Vol. 26, no. 5, pp. 527-532, 2005.*

[9] J. Yang, J.Y. Yang, "From Image Vector to Matrix: A Straightforward Image Projection Technique—IMPCA vs. PCA," *Pattern Recognition, Vol. 35, no. 9, pp. 1997-1999, 2002.*

[10] The Olivetti Research Laboratory (ORL), Olivetti database, 1994 <http://www.*cam-orl*.co.uk/facedatabase.html>

[11] Q. Liu, H. Lu, S.Ma, "Improving Kernel Fisherface Discriminant Analysis for Face Recognition," *IEEE Trans. Pattern Anal. Machine Intell. Vol. 14, no. 1, pp. 42–49, 2004.*

[12] J. Yang, A. F. Frangi, J. Y. Yang, D. Zhang, Z. Jin, "KPCA Plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature Extraction and Recognition," *IEEE Trans. Pattern Anal. Machine Intell. Vol. 27, no. 2, pp. 230–244, 2005.*

[13] J. Lu, K. N. Plataniotis, A.N. Venetsanopoulos, "Face Recognition Using Kernel Direct Discriminant Analysis Algorithms," *IEEE Trans. Neural Networks, Vol. 14, no. 1, pp. 117-126, 2003.*