An Implementation of Discriminative Common Vector Approach Using Matrices

Mehmet Koc, Atalay Barkana
Electrical and Electronics Engineering
Anadolu University
Eskisehir, Turkey
{mkoc6, atalaybarkan}@anadolu.edu.tr

***Abstract*— If one sample per class is available in a face recognition problem, vector-based methods which use within-class scatter will fail. The reason for that is the zero within-class matrix. In this paper a two dimensional extension of the discriminative common vector approach (2D-DCVA) is proposed. The performance of the proposed method is compared with discriminative common vector approach (1D-DCVA) and two dimensional Fisher linear discriminant analysis (2D-FLDA) in ORL, FERET, YALE, and UMIST face databases in one sample problem. Our proposed method outperforms 1D-DCVA and 2D-FLDA in all databases.**

Keywords- one sample problem; common vector; DCVA; two dimensional FLDA

I. INTRODUCTION

Face recognition has many application areas such as security, law enforcement, person identification [1,2]. If only one sample per person is available, then the problem gets difficult. This situation is called one sample problem [3]. Methods which use within-class scatter such as conventional Fisher discriminant analysis (1D-FLDA) will suffer from one sample problem because within-class matrix is a zero matrix. Many algorithms have been proposed to overcome this challenge [3,4,5,6,7]. General tendency at these methods is generating the virtual samples to increase the training set size. But this is not the solution of the singularity problem because in face recognition problems dimension of the feature space is high with respect to the number of feature vectors. This problem is called small sample size problem [8]. One solution to overcome the singularity problem is using the two dimensional variant of one dimensional methods. Two dimensional Fisher discriminant analysis (2D-FLDA) [9] is a solution of the singularity problem in 1D-FLDA. This method was used in [4] and [5] after generating virtual samples. Also discriminative common vector approach (1D-DCVA) which is a variation of FLDA comes up with a solution that overcomes the singularity problem of 1D-FLDA [10].

In this work we proposed a two dimensional extension of the discriminative common vector approach. In order to obtain unique common vector for each class, we use feature vectors instead of feature matrices in the first stage of this method. Then we convert the common vectors into matrices and calculate the discriminative common matrices. In [11], feature matrices are used to obtain common vectors. This method though cannot get unique common vectors. A brief review of the discriminative common vector approach (1D-DCVA) is given in Sec.II. Two dimensional extension of the discriminative common vector approach is given in Sec.III. We used QR decomposition with column pivoting (QRCP) method to generate the virtual samples. QRCP method is given in Sec.IV. We tested the performance of 2D-DCVA in four different databases. Database descriptions and the experiments are given in Sec.V, and finally the results are discussed in Sec.VI.

II. DISCRIMINATIVE COMMON VECTOR APPROACH

Discriminative common vector approach (1D-DCVA) is first introduced in [10]. The method gives a solution to the limitations of methods that use the null space of the within-class scatter matrix.

Let C be the number of classes, N be the number of feature vectors from each class, and let $\boldsymbol{x}_m^i$ be the m^{th} feature vector from i^{th} class. Then the within-class scatter matrix can be written as

$$\boldsymbol{S}_W = \sum_{i=1}^{C}\sum_{m=1}^{N}(\boldsymbol{x}_m^i - \boldsymbol{\mu}_i)(\boldsymbol{x}_m^i - \boldsymbol{\mu}_i)^T \quad (1)$$

where $\boldsymbol{\mu}_i = 1/N\sum_{m=1}^{N}\boldsymbol{x}_m^i$ is the mean of the i^{th} class. The method can be summarized as follows:

- Obtain the projection matrix $\boldsymbol{U} = [\boldsymbol{u}_1\boldsymbol{u}_2 \ldots \boldsymbol{u}_{NC-C}]$ where $\boldsymbol{u}_i, i = 1,2, \ldots NC - C$ are the eigenvectors corresponding to the nonzero eigenvalues of $\boldsymbol{S}_W$.
- Obtain the common vectors by projecting any feature vector from each class onto the null space of $\boldsymbol{S}_W$.

$$\boldsymbol{x}_{com}^i = \boldsymbol{x}_m^i - \boldsymbol{U}\boldsymbol{U}^T\boldsymbol{x}_m^i, m = 1, \ldots, N, i = 1, \ldots, C \quad (2)$$

- Compute the eigenvectors $\boldsymbol{w}_k$ of the scatter matrix of the common vectors $\boldsymbol{S}_{com}$, corresponding to the nonzero eigenvalues and obtain the projection matrix $\boldsymbol{W} = [\boldsymbol{w}_1\boldsymbol{w}_2 \cdots \boldsymbol{w}_{C-1}]$. In here $\boldsymbol{S}_{com}$ is defined as

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-202-8

$$\boldsymbol{S}_{com} = \sum_{i=1}^{C} (\boldsymbol{x}_{com}^{i} - \boldsymbol{x}_{ave})(\boldsymbol{x}_{com}^{i} - \boldsymbol{x}_{ave})^{T} . \quad (3)$$

where $\boldsymbol{x}_{ave}$ is the mean of the common vectors, i.e., $\boldsymbol{x}_{ave} = 1/C \sum_{i=1}^{C} \boldsymbol{x}_{com}^{i}$.

- Obtain the discriminative common vectors by projecting any sample from each class onto the range space of $\boldsymbol{S}_{com}$.

$$\boldsymbol{\Omega}_{com}^{i} = \boldsymbol{W}^{T} \boldsymbol{x}_{m}^{i}, m = 1, \dots, N, i = 1, \dots, C \quad (4)$$

Let $\boldsymbol{x}_{test}$ be the test vector to be classified. Then classification can be done according to the following decision rule.

$$C^{*} = \underset{j}{\operatorname{argmin}} \left\{ \left\| \boldsymbol{\Omega}_{com}^{i} - \boldsymbol{W}^{T} \boldsymbol{x}_{test} \right\| \right\}, \ j = 1, \dots, C \quad (5)$$

III. Two Dimensional Extension Of DCVA

Let C be the number of image classes, N, be the number of feature vectors in each class and, $\boldsymbol{X}_{m}^{i}$ be the m^{th} two dimensional p by q pixel image of the i^{th} class. We convert the image matrix $\boldsymbol{X}_{m}^{i}$ to a vector $\boldsymbol{x}_{m}^{i}$ in the $n = p \times q$ dimensional space.

It is proved in [10] that the common vectors obtained from total within-class scatter matrix are unique for each class. In the first stage of the proposed method, we use $\boldsymbol{S}_{W}$, to take the advantage of the uniqueness of the common vectors. We apply the eigen decomposition to $\boldsymbol{S}_{W}$ and obtain the projection matrix $\boldsymbol{P}^{\perp}$ of its null space using the eigenvectors corresponding to the zero eigenvalues $\boldsymbol{v}_{i}$, $i = C(N-1)+1, \dots, n$. $\boldsymbol{P}^{\perp}$ can be calculated as follow,

$$\boldsymbol{P}^{\perp} = \sum_{i=C(m-1)+1}^{n} \boldsymbol{v}_{i} \boldsymbol{v}_{i}^{T} \quad (6)$$

Then the common vector of i^{th}class is calculated as

$$\boldsymbol{x}_{com}^{i} = \boldsymbol{P}^{\perp} \boldsymbol{x}_{m}^{i}, \ \ i = 1, \dots, C-1, m = 1, \dots, N \quad (7)$$

It should be noted that (2) and (7) give exactly the same results. We convert the common vectors $\boldsymbol{x}_{com}^{i}$ into p by q matrices, $\boldsymbol{X}_{com}^{i}$. The covariance matrix of the common matrices can be calculated as

$$\boldsymbol{S}_{com} = \sum_{i=1}^{C} (\boldsymbol{X}_{com}^{i} - \boldsymbol{X}_{ave})^{T} (\boldsymbol{X}_{com}^{i} - \boldsymbol{X}_{ave}) . \quad (8)$$

where $\boldsymbol{X}_{ave} = 1/C \sum_{i=1}^{C} \boldsymbol{X}_{com}^{i}$ is the mean of the common matrices. We are trying to find the optimal projection vectors $\boldsymbol{W} = [\boldsymbol{w}_{1} \vdots \boldsymbol{w}_{2} \vdots \cdots \vdots \boldsymbol{w}_{d}]$ which maximizes the criterion $J(\boldsymbol{W}) = \boldsymbol{W}^{T} \boldsymbol{S}_{com} \boldsymbol{W}$. Here d can be at most $\min(C-1, n)$.

We use the nearest neighbor classifier for classification. The discriminant features of an image $\boldsymbol{X}_{r}$ is calculated as

$$\boldsymbol{Y}_{r} = \boldsymbol{X}_{r} \boldsymbol{W} = [\boldsymbol{y}_{1}^{r} \vdots \boldsymbol{y}_{2}^{r} \vdots \cdots \vdots \boldsymbol{y}_{d}^{r}]. \quad (9)$$

Let $\boldsymbol{X}_{test}$ be the test image to be classified. The optimal projection vectors of the test image can be given as $\boldsymbol{Y}_{test} = \boldsymbol{X}_{test} \boldsymbol{W} = [\boldsymbol{y}_{1}^{test} \vdots \boldsymbol{y}_{2}^{test} \vdots \cdots \vdots \boldsymbol{y}_{d}^{test}]$. Then the test image is classified according to the following decision rule.

$$C^{*} = \underset{i}{\operatorname{argmin}} \left\{ \sum_{k=1}^{d} \left\| \boldsymbol{y}_{k}^{test} - \boldsymbol{y}_{k}^{i} \right\| \right\} \quad (10)$$

IV. Image Decomposition With QR

QR decomposition is a well-known matrix factorization method [12]. If $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, then it can be decomposed as $\boldsymbol{A} = \boldsymbol{QR}$ where $\boldsymbol{Q} \in \mathbb{R}^{m \times n}$ with orthogonal columns which span the same subspace with the columns of $\boldsymbol{A}$, and $\boldsymbol{R}$ is an upper triangular matrix. QR-decomposition with column pivoting (QRCP) [13,14] is a modified version of QR. In this method the column of the matrix $\boldsymbol{A}$ are sorted such that the absolutes values of the diagonal elements of the matrix $\boldsymbol{R}$ are sorted in descending order. In this way, most of the energy of an image is concentrated into some basis images [5]. The basis images of $\boldsymbol{A}$ can be calculated as $\boldsymbol{q}_{i} \boldsymbol{\tau}_{i}$ where $\boldsymbol{q}_{i}$ is the i^{th} column of $\boldsymbol{Q}$ and $\boldsymbol{\tau}_{i}$ is the i^{th} row of $\boldsymbol{R}$. The orders of columns of $\boldsymbol{A}$ are stored in a permutation matrix $\boldsymbol{P}$ such that the equation $\boldsymbol{Q}^{T} \boldsymbol{AP} = \boldsymbol{R}$ holds. Let the approximation of an image matrix $\boldsymbol{A}$ be $\widehat{\boldsymbol{A}}$. Then it can be calculated as

$$\widehat{\boldsymbol{A}} = \sum_{i=1}^{k} \boldsymbol{q}_{i} \boldsymbol{\tau}_{i}. \quad (11)$$

Here k is selected according to the ratio E given below.

$$\frac{\sum_{i=1}^{k} d_{i}}{\sum_{i=1}^{m} d_{i}} \geq E \quad (12)$$

$d_{i}, i = 1,2, \dots, m$ are the absolute values of the diagonal elements of $\boldsymbol{R}$. In experiments we selected $E = 97\%$ as in [5]. In Figure 1 a sample image selected from YALE face database and its two approximations evaluated from image and its transpose are shown. The image and two reconstructed images evaluated from the image and its transpose are labeled as the training images of that subject.

Figure 1. Sample image and its virtual variants evaluated from the image and its transpose.

V. EXPERIMENTS

In the experimental stage, the performances of DCVA, our proposed method 2D-DCVA, and 2D-FLDA are compared in four face databases namely, ORL [15], FERET [16], YALE [17], and UMIST [18].

ORL face database contains 10 grayscale images from each 40 subjects which are taken in the lab. Images contain different lighting conditions and facial expressions (e.g., closed eyes, glasses, smile). Also images were taken at dark background and subjects are in the frontal position with tolerance to some side movement. The original size of the images is 112×92. In the experiments we used the original images of this database. FERET database contains 14,051 grayscale images from 1199 subjects. In the experiments a subset of the database that contains 200 subjects is used. Each subject has two images from f_a and f_b probes. YALE face database contains 11 images from each 15 subjects. Database includes different facial expressions and illumination conditions (i.e., with/without glasses, happy, sad, sleepy, surprised, wink, center-light, right-light, normal). UMIST database contains 20 individuals. The number of pictures per person varies from 19 to 36. Images were taken at various angles from left profile to right profile.

We preprocessed the images by cropping, scaling, resizing. In TABLE I. , the number of subjects, the number of images from each subject, and the size of the images taken from ORL, FERET, YALE, and UMIST databases after the preprocessing operations are summarized.

TABLE I. THE SUMMARY OF THE DATABASES AFTER THE PREPROCESSING STEP

Database	Number of classes	Number of images per class	Dimension
ORL	40	10	112x92
FERET	200	2	100x100
YALE	15	11	120x110
UMIST	20	19	112x92

In the experiments we randomly select an image from each class. Two virtual images are constructed using this image with the QRCP decomposition. The original image and the two virtual images are used to generate the training set images of the subject. The remaining images are used as test images. This procedure is repeated 5 times and the recognition rates are obtained by averaging each run. We implement this process to all databases. The top recognition rates of DCVA, 2D-DCVA, and 2D-FLDA and their standard deviations on the databases are shown in TABLE II.

TABLE II. THE RECOGNITION RATES ON THE DATABASES

Methods	Databases			
	ORL (%)	FERET (%)	YALE (%)	UMIST(%)
1D-DCVA	69.8± 3.7	88.8± 0.9	58.3± 5.6	55.9± 3.6
2D-DCVA	**76.4**± 2.4	**90.3**± 0.3	**61.6**± 5.2	**64.4**± 4.1
2D-FLDA	76.0± 2.5	90.1± 0.2	59.5± 5.4	61.3± 3.7

VI. RESULTS AND CONCLUSION

One sample problem is an important challenge in face recognition. Methods which use within-class scatter matrix fail. In this work we proposed a two dimensional extension of the discriminative common vector approach. The performance of the proposed method is tested on four different databases namely, ORL, FERET, YALE, and UMIST. 2D-DCVA gave the best recognition results in all databases. 2D-FLDA outperformed 1D-DCVA in all databases. This may be due to fact that the matrix-based methods generally outperform vector based methods [19].

REFERENCES

[1] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," ACM Computing Surveys, vol. 35, no. 4, pp. 399-458, 2003.

[2] Z. Daugman, "Face and gesture recognition: Overview," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 675-676, 1997.

[3] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: a survey," Pattern Recognition, vol. 39, no. 9, pp. 1725-1745, 2006.

[4] Q.-x. Gao, L. Zhang, and D. Zhang, "Face recognition using FLDA with single training image per person," Applied Mathematics and Computation, vol. 205, no. 2, pp. 726-734, 2008.

[5] M. Koç and A. Barkana, "A new solution to one sample problem in face recognition using FLDA," Applied Mathematics and Computation, vol. 217, no. 24, pp. 10368-10376, 2011.

[6] H. Yin, P. Fu, and S. Meng, "Sampled FLDA for face recognitionwith single training image per person," Neurocomputing, vol. 69, no. 16-18, pp. 2443-2445, 2006.

[7] M. Apaydın, Ü.Ç. Turhal, and A. Duysak, "An SVD based common matrix method for face recognition: Single image per person," 25th International Symposium on Computer and Information Sciences, 2010, pp. 289-292.

[8] S. Theodoridis, and K. Koutroumbas, Pattern Recognition, Academic Press, USA, 1999.

[9] H. Kong, E.K. Teoh, J.G. Wang, and R. Venkateswarlu, "Two dimensional Fisher discriminant analysis: Forget about small sample problem," Proc. IEEE Intern. Conf. on Acoustics, Speech, and Signal Processing, 2005, pp. 761-764.

[10] H. Çevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative Common Vectors for Face Recognition," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 27, no. 1, pp. 4-13, 2005.

[11] V.D.M. Nhat and S. Lee, "Discriminative common images for face recognition," In proceedings of ICANN - Part I, vol. 3696, pp. 563-568, 2005.

[12] T. Kailath, A.H. Sayed, and B. Hassibi, Linear Estimation, Prentice Hall, 1999.

[13] S. Chakroborty and G. Saha, "Feature Selection Using Singular Value Decomposition and QR Factorization with Column Pivoting for Text-Independent Speaker Identification," Speech Communication, vol. 52, no. 9, pp. 693-709, 2010.

[14] S. Ari and G. Saha, "In Search of an SVD and QRcp Based Optimization Technique of ANN for Automatic Classification of Abnormal Heart Sounds," International Journal of Biological and Life Sciences, vol. 2, no. 1, pp. 1-9, 2007.

[15] ORL Face Database, AT&T Laboratories Cambridge1992-1994.

[16] P.J. Phillips, H. Moon, S.A. Rizvi, and P. Rauss, "The FERET Evaluation Methodology for Face Recognition Algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pp. 1090-1104, 2000.

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

[18] D.B. Graham and G.N.M. Allinson, "Characterizing Virtual Eigensignatures for General Purpose Face Recognition," Face Recognition: From Theory to Applications, (Ed: Wechsler, H. ve ark.), NATO ASI Series / Computer and Systems Sciences, 1998, ch. 163, pp. 446-456.

[19] W.-S. Zheng, J.H. Lai, and S.Z. Li, "1D-LDA vs. 2D-LDA: When is vector-based linear discriminant analysis better than matrix-based?," Pattern Recognition, vol. 41, no. 7, pp. 2156-2172, 2008.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-202-8