

Discriminative Common Images for Face Recognition

Vo Dinh Minh Nhat and Sungyoung Lee

Kyung Hee University – South of Korea
{vdmnhat, sylee}@oslab.khu.ac.kr

Abstract. Linear discrimination analysis (LDA) technique is an important and well-developed area of image recognition and to date many linear discrimination methods have been put forward. Basically, in LDA the image always needs to be transformed into 1D vector, however recently two-dimensional PCA (2DPCA) technique have been proposed. In 2DPCA, PCA technique is applied directly on the original images without transforming into 1D vector. In this paper, we propose a new LDA-based method that applies the idea of two-dimensional PCA. In addition to that, our approach proposes an method called Discriminative Common Images based on a variation of Fisher's LDA for face recognition. Experiment results show our method achieves better performance in comparison with the other traditional LDA methods.

Index Terms – Fisherfaces, Linear discrimination analysis (LDA), Discriminative Common Image, face recognition.

1. Introduction

The Fisherface method [4] combines PCA and the Fisher criterion [9] to extract the information that discriminates between the classes of a sample set. It is a most representative method of LDA. Nevertheless, Martinez *et al.* demonstrated that when the training data set is small, the Eigenface method outperforms the Fisherface method [7]. Should the latter be outperformed by the former? This provoked a variety of explanations. Liu *et al.* thought that it might have been because the Fisherface method uses all the principal components, but the components with the small eigenvalues correspond to high-frequency components and usually encode noise [11], leading to recognition results that are less than ideal. In [5], Yu *et al.* propose a direct LDA (DLDA) approach to solve this problem. It removes the null space of the between-class scatter matrix firstly by doing eigen-analysis. Then a simultaneous diagonalization procedure is used to seek the optimal discriminant vectors in the

(1) This research was supported by the MIC (Ministry of Information and Communication), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Assessment)
Corresponding Authors : Vo Dinh Minh Nhat (vo_dinhminhnhat@yahoo.com), and SungYoung Lee (sylee@oslab.khu.ac.kr)

subspace of the between-class scatter matrix. However, in this method, removing the null space of the between-class scatter matrix by dimensionality reduction would indirectly lead to the losing of the null space of the within-class scatter matrix which contains considerable discriminative information. Rui Huang [10] proposed the method in which the null space of total scatter matrix which has been proved to be the common null space of both between-class and within-class scatter matrix, and useless for discrimination, is firstly removed. Then in the lower-dimensional projected space, the null space of the resulting within-class scatter matrix is calculated. This lower-dimensional null space, combined with the previous projection, represents a subspace of the whole null space of within-class scatter matrix, and is really useful for discrimination. The optimal discriminant vectors of LDA are derived from it. In [14], a common vector for each individual class is obtained by removing all the features that are in the direction of the eigenvectors corresponding to the nonzero eigenvalues of the scatter matrix of its own class. The common vectors are then used for recognition. In their case, instead of using a given class's own scatter matrix, they use the within-class scatter matrix of all classes to obtain the common vectors.

In [15], a new PCA approach called Two-dimensional PCA (2DPCA), is developed for image feature extraction. As opposed to conventional PCA, 2DPCA is based on 2D matrices rather than 1D vectors. That is, the image matrix does not need to be transformed into vector. Instead, an image covariance matrix can be constructed directly using original image matrices. So, in this paper we improve the LDA algorithm based on the idea of two-dimensional PCA.

Generally, in this paper we improve the LDA-based algorithm by apply the 2D approach into the Discriminative Common Vectors [14] algorithm. Our new method takes the advantages of both 2DPCA method [15] for dealing with high dimensional data to avoid singularity and LDA-based algorithm [14] for dealing with small sample size problem. The remainder of this paper is organized as follows: In Section 2, the traditional LDA methods are reviewed. The idea of the proposed method and its algorithm are described in Section 3. In Section 4, experimental results are presented for the Yale face image databases to demonstrate the effectiveness of our method. Finally, conclusions are presented in Section 5.

2. Linear Discriminant Analysis

Suppose that we have N sample images $\{x_1, x_2, \dots, x_N\}$ taking values in an n -dimensional image space. 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 $y_k \in \mathbb{R}^m$ are defined by the following linear transformation :

$$y_k = W^T x_k \quad (1)$$

where $k = 1, 2, \dots, N$ and $W \in \mathbb{R}^{n \times m}$ is a matrix with orthonormal columns.

Different objective functions will yield different algorithms with different properties. While PCA seeks directions that are efficient for representation, Linear Discriminant Analysis seeks directions that are efficient for discrimination. Assume that each image belongs to one of C classes $\{C_1, C_2, \dots, C_C\}$. Let N_i be the number of the samples in class $C_i (i = 1, 2, \dots, C)$, $\mu_i = \frac{1}{N_i} \sum_{x \in C_i} x$ be the mean of the samples in class X_i , $\mu = \frac{1}{N} \sum_{i=1}^N x_i$ be the mean of all samples. Then the between-class scatter matrix S_b is defined as

$$S_b = \frac{1}{N} \sum_{i=1}^C N_i (\mu_i - \mu)(\mu_i - \mu)^T = \frac{1}{N} \Phi_b \Phi_b^T \quad (2)$$

and the within-class scatter matrix S_w is defined as

$$S_w = \frac{1}{N} \sum_{i=1}^C \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T = \frac{1}{N} \Phi_w \Phi_w^T \quad (3)$$

In LDA, the projection W_{opt} is chosen to maximize 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.,

$$W_{opt} = \arg \max_w \frac{|W^T S_b W|}{|W^T S_w W|} = [w_1 w_2 \dots w_m] \quad (4)$$

where $\{w_i | i = 1, 2, \dots, m\}$ 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.,

$$S_b w_i = \lambda_i S_w w_i \quad i = 1, 2, \dots, m \quad (5)$$

3. Discriminative Common Vectors and Discriminative Common Images

The Discriminative Common Vectors approach can be summarized as follows, details can be referenced at [14] :

- Step 1: Compute the nonzero eigenvalues and corresponding eigenvectors of S_w . Set $Q = [\alpha_1 \dots \alpha_r]$, where r is the rank of S_w .

4 **Vo Dinh Minh Nhat and Sungyoung Lee**

- Step 2: Choose any sample from each class and project it onto the null space of S_w to obtain the common vectors.

$$x_{c_com} = x_i - QQ^T x_i, \quad x_i \in C_c, \quad c = 1, \dots, C \quad (6)$$

- Step 3: Compute the eigenvectors w_k of S_{com} , corresponding to the nonzero eigenvalues. With S_{com} is defined below. There are at most $C-1$ eigenvectors that correspond to the nonzero eigenvalues. Use these eigenvectors to form the projection matrix $W = [w_1 w_2 \dots w_{C-1}]$, which will be used to obtain feature vectors.

$$S_{com} = \sum_{c=1}^C (x_{c_com} - \mu_{com})(x_{c_com} - \mu_{com})^T \quad (7)$$

$$\mu_{com} = \frac{1}{C} \sum_{c=1}^C x_{c_com}$$

In Discriminative Common Images approach, the image matrix does not need to be previously transformed into a vector, so a set of N sample images is represented as $\{X_1, X_2, \dots, X_N\}$ with $X_i \in \mathbb{R}^{k \times s}$. Then the between-class scatter matrix S_b is re-defined as

$$S_b = \frac{1}{N} \sum_{i=1}^c N_i (\mu_{C_i} - \mu_X)(\mu_{C_i} - \mu_X)^T \quad (8)$$

and the within-class scatter matrix S_w is re-defined as

$$S_w = \frac{1}{N} \sum_{i=1}^c \sum_{X_k \in C_i} (X_k - \mu_{C_i})(X_k - \mu_{C_i})^T \quad (9)$$

with $\mu_X = \sum_{i=1}^N X_i \in \mathbb{R}^{k \times s}$ is the mean image of all samples and

$\mu_{C_i} = \frac{1}{N_i} \sum_{X \in C_i} X$ be the mean of the samples in class C_i . The common vectors in (8) now become common images, and defined as

$$X_{c_com} = X_i - QQ^T X_i, \quad X_i \in C_c, \quad c = 1, \dots, C \quad (10)$$

Also, S_{com} in (9) is re-defined as

$$S_{com} = \sum_{c=1}^C (X_{c_com} - \mu_{com})(X_{c_com} - \mu_{com})^T \quad (11)$$

$$\mu_{com} = \frac{1}{C} \sum_{c=1}^C X_{c_com}$$

4. Experimental results

This section evaluates the performance of our proposed algorithm Discriminative Common Images (DCI) compared with that of the original Fisherface algorithm, Direct LDA algorithm, and Discriminative Common Vectors (DCV) based on using Yale face database. The database contains 5760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses x 64 illumination conditions). For every subject in a particular pose, an image with ambient (background) illumination was also captured.

Table 1. The recognition rates on Yale databases

k	Fisherface	Direct LDA	DCV	DCI
2	77.95	79.97	83.01	85.73
3	86.29	86.59	89.11	93.78
4	91.62	92.32	92.97	95.80
5	93.31	93.98	94.42	97.51
6	95.36	95.80	96.87	98.67

Firstly, we tested the recognition rates with different number of training samples. $k(k = 2, 3, 4, 5, 6)$ images of each subject are randomly selected from the database for training and the remaining images of each subject for testing. For each value of k , 50 runs are performed with different random partition between training set and testing set, and *Table 1* shows the average recognition rates (%) with Yale database.

5. Conclusions

A new LDA-based method for face recognition has been proposed in this paper. In this paper, we propose a method that applies the idea of two-dimensional PCA. In addition to that, we propose a method called Discriminative Common Images based on a variation of Fisher's LDA for face recognition. By solving the small sample size problem and high dimensionality, this paper proposes a practical algorithm for applying LDA on image recognition applications, and shows the efficiency in face recognition application. It has the advantage of easy training, efficient testing, and good performance compared to other linear classifiers.

Reference

1. M. Turk, A Pentland: Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, Vol. 3 (1991) 71-86.
2. W. Zhao, R. Chellappa, P.J. Phillips: Subspace Linear Discriminant Analysis for Face Recognition. Technical Report CAR-TR-914, 1999.
3. D. L. Swets, J. J. Weng: Using discrimination eigenfeatures for image retrieval. *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18 (1996) 831–836.
4. P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman: Eigenfaces vs. fisherface: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 19 (1997) 711–720.
5. H. Yu, J. Yang: A direct LDA algorithm for high-dimensional data with application to face recognition. *Pattern Recognit.*, Vol. 34 (2001) 2067–2070.
6. M. Loog, R. P. W. Duin, R. Haeb-Umbach: Multiclass linear dimension reduction by weighted pairwise fisher criteria. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 23 (2001) 762–766.
7. A. M. Martinez, A. C. Kak: PCA versus LDA. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 23 (2001) 228–233.
8. D. H. Foley, J. W. Sammon: An optimal set of discrimination vectors. *IEEE Trans. Comput.*, Vol. C-24 (1975) 281–289.
9. R. A. Fisher: The use of multiple measurements in taxonomic problems. *Ann. Eugenics*, Vol. 7 (1936) 178–188.
10. Rui Huang, Qingshan Liu, Hanqing Lu, Songde Ma: Solving the small sample size problem of LDA. *Pattern Recognition*, 2002. Proceedings. 16th International Conference on , Vol 3 (2002)
11. C. Liu, H. Wechsler: Robust coding scheme for indexing and retrieval from large face databases. *IEEE Trans. Image Processing*, Vol. 9 (2000) 132–137.
12. Chengjun Liu, Wechsler H.: A shape- and texture-based enhanced Fisher classifier for face recognition. *IEEE Trans. Image Processing*, Vol. 10(2001) 598–608.
13. L. Chen, H. M. Liao, M. Ko, J. Lin, G. Yu: A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognit.*, Vol. 33 (2000) 1713–1726.
14. Cevikalp H., Neamtu, M., Wilkes, M.; Barkana, A.: Discriminative common vectors for face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , Vol. 27 (2005) 4 – 13
15. Jian Yang, Zhang, D., Frangi, A.F., Jing-yu Yang : Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , Vol. 26 (2004) 131 – 137
16. “The Yale face database” <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.