Available online at www.sciencedirect.com
SCIENCE DIRECT®

Pattern Recognition 38 (2005) 1125–1129

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Rapid and brief communication

Two-dimensional discriminant transform for face recognition

Jian Yang[a, b, *], David Zhang[a], Xu Yong[b], Jing-yu Yang[b]

[a]Department of Computing, Hong Kong Polytechnic University, Kowloon, Hong Kong
[b]Department of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, PR China

Received 13 October 2004; accepted 4 November 2004

Abstract

This paper develops a new image feature extraction and recognition method coined *two-dimensional linear discriminant analysis* (2DLDA). 2DLDA provides a sequentially optimal image compression mechanism, making the discriminant information compact into the up-left corner of the image. Also, 2DLDA suggests a feature selection strategy to select the most discriminative features from the corner. 2DLDA is tested and evaluated using the AT&T face database. The experimental results show 2DLDA is more effective and computationally more efficient than the current LDA algorithms for face feature extraction and recognition.

Keywords: Fisher linear discriminant analysis (FLD or LDA); Fisherfaces; Feature extraction; Face recognition; Two-dimensional data analysis

1. Introduction

Fisher linear discriminant analysis (LDA) has been successfully applied to face recognition area in the past few years. LDA is a 1D-data-based feature extraction technique, so, 2D image matrices must be converted into 1D image vectors before the application of LDA. Since the resulting image vectors are high dimensional, LDA usually encounters the small sample size (S3) problem in which the within-class scatter matrix becomes singular and thus the traditional LDA algorithm fails to use. To address this problem, a number of extended LDA algorithms have been suggested. Among them, the most popular one is to use PCA for dimension reduction prior to performing LDA [1,2]. This method has a computational complexity of $\mathcal{O}(M^3)$. When the training sample size M is large, the computation requirement of this method is still considerable.

To avoid the S3 problem LDA encounters, Liu [3] suggested a 2D image matrix-based linear discriminant technique. His idea is to perform LDA directly based on image matrices, while overleaping the process of turning image matrices into vectors. Thus, the difficulty resulting from high-dimensionality is artfully avoided. As a further development of Liu's method, the *uncorrelated image matrix-based linear discriminant analysis* (IMLDA) technique was proposed recently [4]. IMLDA has an advantage to eliminate the correlation between discriminant feature vectors so that it is more effective than Liu's method for face recognition [4].

A drawback of IMLDA is that it needs more coefficients than LDA for image representation. Thus, IMLDA needs more memory to store its features and costs more time to calculate distance (similarity) in classification phase. In this paper, we develop a new image feature extraction method coined *two-dimensional linear discriminant analysis* (2DLDA) to overcome the disadvantage of IMLDA. The initial idea of 2DLDA is to perform IMLDA twice: the first one is in horizontal direction and the second is in vertical

* Corresponding author. Tel.: +852 2766 7312; fax: +852 2774 0842.

E-mail addresses: csjyang@comp.polyu.edu.hk (J. Yang), csdzhang@comp.polyu.edu.hk (D. Zhang), laterfall@sina.com (X. Yong), yangjy@mail.njust.edu.cn (J.-y. Yang).

0031-3203/$30.00
doi:10.1016/j.patcog.2004.11.019

direction. After the two sequential IMLDA transforms, the discriminant information is compacted into the up-left corner of the image. A feature selection mechanism is followed to select the most discriminative features from the corner. The effectiveness of the proposed method is verified using AT&T database.

2. Outline of IMLDA

Suppose there are c known pattern classes. M is the total number of training samples, and M_i is the number of training samples in class i. In class i, the jth training image is denoted by an $m \times n$ matrix $\mathbf{A}_j^{(i)}$. The mean image of training samples in class i is denoted by $\bar{\mathbf{A}}^{(i)}$ and the mean image of all training sample is $\bar{\mathbf{A}}$.

Based on the given training image samples (image matrices), the *image between-class scatter matrix* and *image within-class scatter matrix* can be constructed by

$$\mathbf{G}_b = \frac{1}{M} \sum_{i=1}^{c} M_i (\bar{\mathbf{A}}_i - \bar{\mathbf{A}})^{\mathrm{T}} (\bar{\mathbf{A}}_i - \bar{\mathbf{A}}), \quad (1)$$

$$\mathbf{G}_w = \frac{1}{M} \sum_{i=1}^{c} \sum_{j=1}^{M_i} (\mathbf{A}_j^{(i)} - \bar{\mathbf{A}}^{(i)})^{\mathrm{T}} (\mathbf{A}_j^{(i)} - \bar{\mathbf{A}}^{(i)}). \quad (2)$$

By the definition, it is easy to verify that $\mathbf{G}_b$ and $\mathbf{G}_w$ are both $n \times n$ nonnegative definite matrices. It should be mentioned that in face recognition problems, $\mathbf{G}_w$ is usually invertible unless that there is only one training sample per class.

The generalized Fisher criterion can be defined by

$$J(\boldsymbol{\varphi}) = \frac{\boldsymbol{\varphi}^{\mathrm{T}} \mathbf{G}_b \boldsymbol{\varphi}}{\boldsymbol{\varphi}^{\mathrm{T}} \mathbf{G}_w \boldsymbol{\varphi}}. \quad (3)$$

It is easy to find a vector $\mathbf{u}^*$ to maximize the Rayleigh quotient function $J(\boldsymbol{\varphi})$. After the projection of samples onto $\mathbf{u}^*$, the ratio of the between-class scatter to the within-class scatter is maximized. So, the vector $\mathbf{u}^*$ is called the optimal image projection direction. Generally, a single projection direction is not enough for the discrimination of multi-class problems so that we need a set of discriminant vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_q$, which maximize the generalized Fisher criterion and satisfy $\mathbf{G}_t$-orthogonal constraints, i.e.,

$$\mathbf{u}_i^{\mathrm{T}} \mathbf{G}_t \mathbf{u}_j = 0, \quad \text{where } \mathbf{G}_t = \mathbf{G}_b + \mathbf{G}_w,\ i \neq j,\ i, j = 1, \ldots, q. \quad (4)$$

The role of these constraints is to make the resulting discriminant feature vectors uncorrelated and thereby more discriminative for classification [4].

Actually, the discriminant feature vectors subject to the above constraints can be selected as the $\mathbf{G}_t$-orthogonal generalized eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_q$ of $\mathbf{G}_b$ and $\mathbf{G}_w$ corresponding to q largest generalized eigenvalues, i.e., $\mathbf{G}_b \mathbf{u}_j = \lambda_j \mathbf{G}_w \mathbf{u}_j$, where $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_q$. These eigenvectors can be calculated using the algorithm suggested in Ref. [4]. The obtained eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_q$ are used for image feature extraction. Let

$$\mathbf{B} = \mathbf{A}\mathbf{U}, \quad \text{where } \mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_q). \quad (5)$$

The resulting feature matrix $\mathbf{B}$ is used to represent image $\mathbf{A}$ for classification.

3. 2DLDA

3.1. Idea

IMLDA can eliminate the correlations between image columns and compress the discriminant information optimally into a few of columns in horizontal direction. However, it disregards the correlations between image rows and the data compression in vertical direction. So, its compression rate is far lower than LDA and more coefficients are needed for the representation of images. This must lead to a slow classification speed and large storage requirements for large-scaled databases.

In this section, we will suggest a way to overcome the weakness of IMLDA. Our idea is simple, just to perform IMLDA twice: the first one is in horizontal direction and the second is in vertical direction (*note that any operation in vertical direction can be equivalently implemented by an operation in horizontal direction by virtue of the transpose operation of matrix*). Specifically, given image $\mathbf{A}$, we obtain its feature matrix $\mathbf{B}$ after the first IMLDA transform. Then, we transpose $\mathbf{B}$ and input $\mathbf{B}^{\mathrm{T}}$ into IMLDA, and determine the transform matrix $\mathbf{V}$. Projecting $\mathbf{B}^{\mathrm{T}}$ onto $\mathbf{V}$, we obtain $\mathbf{C}^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{V}$. The resulting feature matrix is $\mathbf{C} = \mathbf{V}^{\mathrm{T}}\mathbf{B}$. This process is illustrated in Fig. 1.

In the whole process, the first IMLDA transform $\mathbf{B} = \mathbf{A}\mathbf{U}$ performs the compression of 2D-data in horizontal direction, making the discriminant information pack into a small number of columns. While the second IMLDA transform $\mathbf{C} = \mathbf{V}^{\mathrm{T}}\mathbf{B}$ performs the compression of 2D-data in vertical direction, eliminating the correlations between columns of image $\mathbf{B}$ and making its discriminant information further compact into a small number of rows. Ultimately, the discriminant information of the whole image is packed into the up-left corner of the image matrix.

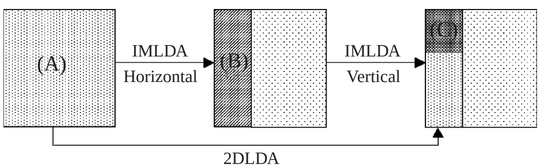

Fig. 1. Illustration of 2DLDA transform.