Digital Signal Processing 55 (2016) 78–84

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Adaptive linear discriminant regression classification for face recognition

Pu Huang [a,e], Zhihui Lai [b,*], Guangwei Gao [c], Geng Yang [a], Zhangjing Yang [d]

[a] *School of Computer Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China*
[b] *College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, China*
[c] *Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China*
[d] *School of Technology, Nanjing Audit University, Nanjing, 211815, China*
[e] *Key Laboratory of Image and Video Understanding for Social Safety (Nanjing University of Science and Technology), Nanjing, 210094, China*

A R T I C L E I N F O

Article history:
Available online 6 May 2016

Keywords:
Feature extraction
Face recognition
LRC
Discriminant analysis
Adaptive

A B S T R A C T

Linear discriminant regression classification (LDRC) was presented recently in order to boost the effectiveness of linear regression classification (LRC). LDRC aims to find a subspace for LRC where LRC can achieve a high discrimination for classification. As a discriminant analysis algorithm, however, LDRC considers an equal importance of each training sample and ignores the different contributions of these samples to learn the discriminative feature subspace for classification. Motivated by the fact that some training samples are more effectual in learning the low-dimensional feature space than other samples, in this paper, we propose an adaptive linear discriminant regression classification (ALDRC) algorithm by taking special consideration of different contributions of the training samples. Specifically, ALDRC makes use of different weights to characterize the different contributions of the training samples and utilizes such weighting information to calculate the between-class and the within-class reconstruction errors, and then ALDRC seeks to find an optimal projection matrix that can maximize the ratio of the between-class reconstruction error over the within-class reconstruction error. Extensive experiments carried out on the AR, FERET and ORL face databases demonstrate the effectiveness of the proposed method.

1. Introduction

At present, face recognition has attracted considerable attention in the fields of computer vision and pattern recognition due to its wide range of applications such as access control, card identification, security monitoring, and so on. Feature extraction and classification are two key steps in a face recognition system. Over the past few decades, numerous algorithms have been proposed for feature extraction, among which, principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2] are the two most popular techniques. PCA aims at finding a set of projection vectors maximizing the data's variance. PCA is an unsupervised method, and it might not be a suitable feature extraction algorithm because maximizing the data's variance cannot guarantee the separability of different classes. In contrast, LDA is a supervised algorithm and it attempts to find a set of projection vectors to maximize the ratio of the between-class scatter to the within-class scatter. The projection of LDA makes the data from the same class as close as possible while the data from different classes as far away as possible. Compared with PCA, LDA is able to discover the essential discriminant structure of the data, and thus LDA is generally superior to PCA in terms of recognition rate.

Although LDA have been successfully employed to solve many practical problems, it cannot be directly applied for face recognition since face recognition is typically a small sample size (SSS) problem [3] which makes the within-class scatter matrix be usually singular. To overcome this limitation, many LDA variants [4–8] have been developed and demonstrate remarkable results for face recognition. For example, Belhumeur et al. [4] presented the famous Fisherface method, in which, a PCA plus LDA framework was used to extract features for face recognition. In [5] and [6], Chen and Jin proposed null-space LDA and uncorrelated discriminant analysis method respectively. In [7], Li et al. used the difference of between-class scatter and within-class scatter as the discriminant criterion called maximum margin criterion (MMC), in which the SSS problem in traditional LDA is naturally avoided since the inverse matrix of the within-class scatter matrix does not need to be computed.

Besides, motivated by manifold learning [9–11], a number of locality-based discriminant algorithms [12–20] have been developed, such as marginal fisher analysis (MFA) [12], multi-manifold

* Corresponding author.
E-mail addresses: huangpu3355@163.com (P. Huang), lai_zhi_hui@163.com (Z. Lai).

http://dx.doi.org/10.1016/j.dsp.2016.05.001
1051-2004/

discriminant analysis (MMDA) [13], discriminant similarity and variance preserving projection (DSVPP) [14], local structure preserving discriminant analysis (LSPDA) [15], and so on. Different from LDA, these methods combine both local information and discriminant information to learn a projection matrix that can reveal the local geometrical structure and the discriminant structure simultaneously. Despite different motivations of these algorithms, they can be nicely integrated in a graph embedding framework [12].

After extracting features, we need to choose a proper classifier to predict the labels of the features. Nearest neighbor classifier (NNC) [21] and minimum distance classifier (MDC) [22] are two of the most popular used classification methods. According to different decision rules, NNC assigns the label of a test sample $\mathbf{x}$ as that of a training sample which is closest to $\mathbf{x}$ and MMC classifies $\mathbf{x}$ into a class whose mean is closest to $\mathbf{x}$, respectively. However, both NNC and MMC are quite sensitive to noises (outliers), in particular for face recognition problems, to recognize face images with varying expression and illumination as well as occlusion and disguise, the two classifiers may not work well. In [23], Wright et al. proposed a robust face recognition framework called sparse representation based classification (SRC). This framework can handle errors due to occlusion and corruption, and extensive experimental results show its robustness and effectiveness for face recognition. However, SRC is much time-consuming due to solving the l_1-norm minimization problem. In [24], Naseem et al. proposed a linear regression classification (LRC) algorithm based on the assumption that face images from a specific class lie on a linear subspace. In LRC, the task of face recognition is defined as a problem of linear regression. The regression coefficients are estimated by using the least-squares estimation method, and then the decision is made in favor of the class which leads to the minimum reconstruction error. LRC has been successfully applied for face recognition and consumes less time than SRC on computation.

To boost the robustness of LRC for pattern or face recognition, Huang et al. [25] incorporated the discriminant analysis concept into LRC and proposed linear discriminant regression classification (LDRC) algorithm. By applying discriminant analysis, they aim to learn a subspace where samples from different classes are far away from each other while samples from the same class are close to each other. Therefore, LDRC estimates an optimal projection in such a way that the ratio of the between-class reconstruction error over the within-class reconstruction error achieved by the LRC is maximized. However, LDRC considers an equal importance of each training sample and neglects the different contributions of these samples to learn the discriminative feature subspace for classification.

Motivated by the fact that some training samples are more effectual in learning the low-dimensional feature space than other samples, in this paper, we propose an adaptive linear discriminant regression classification (ALDRC) algorithm by taking into account of different contributions of the training samples. To be specific, we use different weights to characterize different contributions of training samples and integrate such weighting information into the estimation of the between-class and within-class reconstruction errors, and then a concise discriminant criterion is derived to learn a projection via maximizing the ratio of the between-class reconstruction error over the within-class reconstruction error.

The remainder of this paper is organized as follows. In Section 2, we briefly review LRC and LDRC. In Section 3, the proposed algorithm ALDRC is introduced in detail. Sections 4 provides the justification of penalty functions. To evaluate the effectiveness of the proposed algorithm, experiments on the AR, FERET and ORL face databases are presented in Section 5. Finally, some conclusions are made in Section 6.

2. LRC and LDRC

Assume we have n training samples denoted by $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$ from c classes, where $\mathbf{x}_i \in \mathrm{R}^{N\times 1}$ $(i = 1, 2, \ldots, n)$ is a column vector which represents an image sample. In this section, we briefly review LRC and LDRC.

2.1. LRC

LRC is developed based on the concept that patterns from the same class lie on a linear subspace.

Let $\mathbf{y} \in \mathrm{R}^{N\times 1}$ be a test image vector and the problem is to classify $\mathbf{y}$ as one of the classes $i = 1, 2, \ldots, c$. If $\mathbf{y}$ belongs to the ith class, it should be represented as a linear combination of the training images from the same class (lying in the same subspace), i.e.,

$$\mathbf{y} = \mathbf{X}_i \boldsymbol{\beta}_i, \quad i = 1, 2, \ldots, c \tag{1}$$

where $\mathbf{X}_i$ denote the set of samples from the ith class, $\boldsymbol{\beta}_i \in \mathrm{R}^{n_i\times 1}$ is the vector of reconstruction parameters, and n_i is the number of samples from the ith class. Given that $N \geq n_i$, the system of equations in (1) is well conditioned and $\boldsymbol{\beta}_i$ can be estimated using least-squares estimation [26,27]:

$$\boldsymbol{\beta}_i = \left(\mathbf{X}_i^{\mathrm{T}}\mathbf{X}_i\right)^{-1}\mathbf{X}_i^{\mathrm{T}}\mathbf{y} \tag{2}$$

The estimated vector of parameters $\boldsymbol{\beta}_i$ along with the predictor $\mathbf{X}_i$ is used to predict the response vector for each class i. By substituting (2) into (1), we have:

$$\begin{aligned} \hat{\mathbf{y}}_i &= \mathbf{X}_i(\mathbf{X}_i^{\mathrm{T}}\mathbf{X}_i)^{-1}\mathbf{X}_i^{\mathrm{T}}\mathbf{y} \\ &= \mathbf{H}_i\mathbf{y} \end{aligned}, \quad i = 1, 2, \ldots, c \tag{3}$$

where the predicted vector $\hat{\mathbf{y}} \in \mathrm{R}^{N\times 1}$ is the projection of $\mathbf{y}$ onto the ith subspace by the class projection matrix:

$$\mathbf{H}_i = \mathbf{X}_i\left(\mathbf{X}_i^{\mathrm{T}}\mathbf{X}_i\right)^{-1}\mathbf{X}_i^{\mathrm{T}}, \quad i = 1, 2, \ldots, c \tag{4}$$

The decision of LRC is ruled based on the minimum distance between the original vector and its predicted vector. Thus the predicted identity (label) i^* of the test image vector $\mathbf{y}$ is determined as:

$$i^* = \min_i \|\mathbf{y} - \hat{\mathbf{y}}_i\|, \qquad i = 1, 2, \ldots, c \tag{5}$$

2.2. LDRC

LDRC is a feature extraction method which employs discriminant analysis in the LRC to guarantee efficient for discrimination. It uses the labeled training data to construct a more reliable subspace on which an effective discriminant regression classification could be conducted. The outline of LDRC is given as follows.

To obtain a more effective discriminant subspace for LRC, LDRC seeks to find a projection by maximizing the ratio of the between-class reconstruction error over the within-class reconstruction error. Let $\mathbf{A} \in \mathrm{R}^{N\times d}(d < N)$ denote the optimal projection matrix obtained by LDRC and $\mathrm{tr}(\cdot)$ denote the trace operator, and then the objective function of LDRC is defined as [25]:

$$\mathbf{A}^* = \max_{\mathbf{A}} \frac{\mathrm{tr}(\mathbf{A}^{\mathrm{T}}\mathbf{E}_b\mathbf{A})}{\mathrm{tr}[\mathbf{A}^{\mathrm{T}}(\mathbf{E}_w + \varepsilon\mathbf{I})\mathbf{A}]} \tag{6}$$

where ε is a small positive number and $\mathbf{I}$ is an identity matrix, $\mathbf{E}_b$ and $\mathbf{E}_w$ are the between-class and the within-class reconstruction error matrices respectively, and they are defined as [25]:

$$\mathbf{E}_b = \frac{1}{n(c-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq l(x_i)}^{c} \left(\mathbf{x}_i - \mathbf{x}_{ij}^{\mathrm{inter}}\right)\left(\mathbf{x}_i - \mathbf{x}_{ij}^{\mathrm{inter}}\right)^{\mathrm{T}} \tag{7}$$