Pattern Recognition 47 (2014) 1433–1442

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Global plus local: A complete framework for feature extraction and recognition

Di Zhang [a,d,*], Jiazhong He [b], Yun Zhao [a], Zhongliang Luo [c], Minghui Du [d]

[a] School of Information Engineering, Guangdong Medical College, Dongguan, China
[b] Department of Physics, Shaoguan University, Shaoguan, China
[c] School of Computer Science, Shaoguan University, Shaoguan, China
[d] School of Electronics and Information, South China University of Technology, Guangzhou, China

ARTICLE INFO

Article history:
Received 4 February 2013
Received in revised form
16 May 2013
Accepted 2 October 2013
Available online 17 October 2013

Keywords:
Feature extraction
LDA
Global information
Local information
Pattern classification

ABSTRACT

Linear discriminant analysis (LDA) is one of the most popular supervised feature extraction techniques used in machine learning and pattern classification. However, LDA only captures global geometrical structure information of the data and ignores the geometrical structure information of local data points. Though many articles have been published to address this issue, most of them are incomplete in the sense that only part of the local information is used. We show here that there are total three kinds of local information, namely, local similarity information, local intra-class pattern variation, and local inter-class pattern variation. We first propose a new method called enhanced within-class LDA (EWLDA) algorithm to incorporate the local similarity information, and then propose a complete framework called complete global–local LDA (CGLDA) algorithm to incorporate all these three kinds of local information. Experimental results on two image databases demonstrate the effectiveness of our algorithms.

1. Introduction

In the past two decades, appearance-based image recognition has attracted considerable interest in computer vision, machine learning, and pattern classification [1–4]. It is well known that the dimension of an image is usually very high. For example, an image with a resolution of 120×120 can be viewed as a 14,400-dimensional vector. High dimensionality of feature vector has become a critical problem in practical applications. The data in the high-dimensional space is usually redundant and may degrade the performance of classifiers when the number of training samples is much smaller than the dimensionality of the image data. A common way to resolve this problem is to use either supervised or unsupervised feature extraction techniques. Principal component analysis (PCA) is a popular unsupervised feature extraction algorithm, which performs feature extraction by projecting the original d-dimensional data onto a l-dimensional ($l \ll d$) linear subspace spanned by the leading eigenvectors of the data's covariance matrix. LDA searches the projection axes on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. Since discriminating information is encoded, it is generally believed that LDA is superior to PCA [2]. However, when applying LDA to real-world applications, there are two problems needed to be carefully considered: (1) the singularity of within-class scatter matrix; and (2) the local geometrical structure information.

In the past, many LDA extensions have been developed to deal with the singularity of within-class scatter matrix, among which the most representative methods are Fisherface [3], Pseudo-inverse LDA (PLDA) [4], regular LDA (RLDA) [5], LDA/QR [6], LDA/GSVD [7], orthogonal LDA (OLDA) [8], null space LDA (NLDA) [28], direct-LDA (D-LDA) [29], SRDA [30], and CLDA [31]. Although these methods have been shown to be effective in experiments, their generalization capability on testing data cannot be guaranteed. The main reason is that they only capture global geometrical structure information of the data via equally minimizing the distance among data points from the same class and maximizing the distance among data points from different classes, and the local geometrical structure information is disregarded. It is just the local geometrical structure information that characterizes important modes of variability of data and helps to alleviate or even avoid the over-fitting problem, which will improve the generalization ability of the algorithms [9–11].

Recently, a number of graph-based feature extraction methods, which are also called manifold learning based discriminant approaches, have been successfully applied and became important methodologies in computer vision, machine learning and pattern classification. Some well-known graph-based algorithms are locally linear embedding (LLE) [12], Isomap [13], Laplacian eigenmap [14],

* Corresponding author at: School of Information Engineering, Guangdong Medical College, Dongguan, China. Tel./fax: +86 20 38769529.
E-mail address: changnuode@163.com (D. Zhang).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.10.005

graph embedding [15], linear discriminant projection (LDP) [32], Graph-based Fisher analysis (GbFA) [26], and locality preserving projection (LPP) [16]. All these algorithms were developed based on the assumption that the data lies on a manifold that can be modeled by a nearest-neighbor graph that preserves the local geometrical structure of the input space. Different from LLE, LDP, GbFA, Isomap and Laplacian eigenmap, LPP is a linear algorithm that is quite simple and easy to realize, thus has received much attention in the research community. In summary, two kinds of LPP-based local geometrical structure information have been proposed so far, they are unsupervised [16,17,19,21]/supervised [18,20,22–26] local similarity information and local intra-class pattern variation [27]. Literatures [16–27] show that they are all capable of providing useful discriminant information for classification. However, as we will show in Section 4, local geometrical structure information also includes the local inter-class pattern variation, which is also capable of providing useful discriminant information for classification.

As to the problem of incorporating local geometrical structure information when applying LDA, however, there are not so many articles have been published so far, and the most representative methods are local LDA (LocLDA) [19], local Fisher discriminant analysis (LFDA) [25], enhanced Fisher discriminant criterion (EFDC) [27], unsupervised discriminant projection (UDP) [33], and heteroscedastic LDA (HLDA) [38,39]. Though LocLDA integrates LDA and LPP in a unified framework, it only uses unsupervised local geometrical structure information and disregards label information in the LPP formulation, which is in contradiction to the supervised nature of LDA. EFDC redesigns a new between-class scatter matrix by combining the generic LDA between-class scatter matrix and a local intra-class pattern variation based scatter matrix. LFDA is still a LDA technique with the redesigned LPP-based local within-class and local between-class scatter matrices. Different from generic LDA, LFDA focuses only on the local structure and neglects the global structure of the data. UDP applies Fisher criteria to the nonlocal scatter and local scatter, i.e., finds projection axes on which the local scatter is minimized while the nonlocal scatter is maximized. Besides the unsupervised nature, UDP's nonlocal scatter and local scatter are complementary, the change of local scatter will ultimately lead to the change of nonlocal scatter, which is inconsistent with the nature of LDA. Moreover, LocLDA, LFDA, EFDC, and UDP only use part of the local geometrical structure information of the data. Unlike LocLDA, LFDA, EFDC or UDP, HLDA does not explicitly model local geometrical structure information, but it does so implicitly by not assuming homoscedasticity in the data's covariance structure. However, the local information incorporated in HLDA is modeled by the within-class covariance matrices of different classes, which are in fact the global statistics of different classes. We call such local information pseudolocal information to distinguish it from the real local information defined by LPP.

So, to a large degree, the complete use of all kinds of local geometrical structure information in LDA, has not been addressed adequately so far. In this paper, we will first develop a new supervised feature extraction algorithm, called enhanced within-class LDA (EWLDA) algorithm to incorporate the supervised local similarity information, and then based on the proposed EWLDA algorithm, we propose a complete framework called complete global–local LDA (CGLDA) algorithm to use all kinds of local information of the data.

The rest of the paper is organized as follows. In Section 2, we give a brief review of LDA. Section 3 introduces the proposed EWLDA algorithm and Section 4 presents the CGLDA algorithm. The connection between our proposed algorithms and other state-of-the-art feature extraction algorithms is summarized in Section 5. Extensive experiments for object recognition are conducted in Section 6 to verify the efficiency of our methods. Finally, we provide the concluding remarks and suggestions for future work in Section 7.

2. A brief review of LDA

In classification problems, given a set of n d-dimensional samples $\mathbf{x}_1$, $\mathbf{x}_2$,..$\mathbf{x}_n$, belonging to C known pattern classes, LDA seeks direction $\mathbf{v}$ on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other, i.e., LDA maximizes the objective function $J(\mathbf{v})$ (also known as the Fisher's criterion) as follows:

$$J(\mathbf{v}) = \frac{\mathbf{v}^T\mathbf{S}_B\mathbf{v}}{\mathbf{v}^T\mathbf{S}_W\mathbf{v}} \tag{1}$$

$$\mathbf{S}_B = \sum_{k=1}^{C} m_k(\boldsymbol{\mu}^k-\boldsymbol{\mu})(\boldsymbol{\mu}^k-\boldsymbol{\mu})^T$$

$$\mathbf{S}_W = \sum_{k=1}^{C}\left(\sum_{i=1}^{m_k}(\mathbf{x}_i^k-\boldsymbol{\mu}^k)(\mathbf{x}_i^k-\boldsymbol{\mu}^k)^T\right)$$

where $\boldsymbol{\mu}$ is the total sample mean vector, $\boldsymbol{\mu}^k$ is the centroid of the k-th class, m_k is the number of samples in k-th class, and $\mathbf{x}_i^k$ is the i-th sample in the k-th class. The matrices $\mathbf{S}_B$ and $\mathbf{S}_W$ are often called the between-class scatter matrix and within-class scatter matrix, respectively.

Actually, maximizing the objective function $J(\mathbf{v})$ is equivalent to solving the following two functions simultaneously:

$$\arg\max_{v}(\mathbf{v}^T\mathbf{S}_B\mathbf{v}) \tag{2}$$

$$\arg\min_{v}(\mathbf{v}^T\mathbf{S}_W\mathbf{v}) \tag{3}$$

The objective function (2) aims to push the points of different classes as distant as possible while the objective function (3) aims to make the points of the same class as close as possible. Since Eqs. (2) and (3) are based on the assumption that the data points approximately obey Gaussian distributions, they only capture the global geometrical structure information and ignore the geometrical structure information of local data points. In many real-world applications, however, data points are always in complicated distributions, the geometrical structure information of local data points can also provide useful discriminant information for classification, applying the objective functions Eqs. (2) and (3) may result in the following problems in real-world applications:

1. The objective function (3) tries to get a compact intra-class representation of the training sample by making data points from the same class as close as possible in the reduced feature space. However, it is easy to project the samples, which are not very close to each other and describe the local geometrical structure of a particular class, to a subspace in which they are stick to each other. It indicates that the local similarity information of data points from the same class is damaged in the reduced feature space. Unfortunately, such information is demonstrated to be effective in classification and many algorithms [16–27] have been proposed to try to maintain the local similarity information. Unlike the previously published works, in this paper, a novel algorithm is proposed to maintain the local similarity information.
2. Based on the assumption that all the data points from the same class are mapped onto one single point (the class center) in the feature space, the objective function (2) tries to maximize the distances among different classes by pushing the center of each class away from the center of the whole training samples. In such case, the objective function (2) disregards the pattern