J. Vis. Commun. Image R. 25 (2014) 296–305

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Local maximal margin discriminant embedding for face recognition

Pu Huang [a,*], Zhenmin Tang [a], Caikou Chen [b], Zhangjing Yang [a]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[b] College of Information Engineering, Yangzhou University, Yangzhou 225009, China

ARTICLE INFO

Article history:
Received 10 November 2012
Accepted 24 November 2013
Available online 7 December 2013

Keywords:
Local maximal margin discriminant embedding
Locality preserving projection
Maximum margin criterion
Small sample size problem
Local structure
Appearance-based
Dimensionality reduction
Manifold learning
Face recognition

ABSTRACT

In this paper, a manifold learning based method named local maximal margin discriminant embedding (LMMDE) is developed for feature extraction. The proposed algorithm LMMDE and other manifold learning based approaches have a point in common that the locality is preserved. Moreover, LMMDE takes consideration of intra-class compactness and inter-class separability of samples lying in each manifold. More concretely, for each data point, it pulls its neighboring data points with the same class label towards it as near as possible, while simultaneously pushing its neighboring data points with different class labels away from it as far as possible under the constraint of locality preserving. Compared to most of the up-to-date manifold learning based methods, this trick makes contribution to pattern classification from two aspects. On the one hand, the local structure in each manifold is still kept in the embedding space; one the other hand, the discriminant information in each manifold can be explored. Experimental results on the ORL, Yale and FERET face databases show the effectiveness of the proposed method.

1. Introduction

Face recognition has attracted wide attention of the researchers in the fields of pattern recognition and computer vision because of its immense application potential. Many face recognition methods have been developed over the past few decades. One of the most successful and well-studied techniques to face recognition is the appearance-based method. In an appearance-based technique, a two-dimensional face image of size w by h pixels is represented by a vector in a $w \times h$-dimensional space. In practice, however, these $w \times h$-dimensional spaces are too large to allow robust and fast recognition. A common way to attempt to resolve this problem is to use dimensionality reduction techniques. Two of the most popular dimensionality reduction methods are principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2].

PCA is a classical dimensionality reduction and data representation technique widely used in pattern classification and visualization tasks. PCA is an unsupervised method, which aims to find a linear mapping that preserves the total variance by maximizing the trace of feature variance. The optimal mapping is the leading eigenvectors corresponding to the largest eigenvalues of the covariance matrix for data of all classes.

LDA produces an optimally discriminative projection for certain cases. LDA searches for the transformation that maximizes the between-class scatter and at the same time minimizes the within-class scatter. Different from PCA which is completely unsupervised with regard to the class information of the data, LDA takes full consideration of the class labels and it is generally believed that LDA is able to enhance class separability. Despite the success of the LDA algorithm in many applications, its effectiveness is still limited since, in theory, the number of available projection directions is lower than the class number. Furthermore, class discrimination in LDA is based upon within-class and between-class scatters, which is optimal only in cases where the data of each class is approximately Gaussian distributed, a property that cannot always be satisfied in real-world applications. At the same time, LDA cannot be applied directly to small sample size problem [3] because the within-class scatter matrix is singular [2]. To avoid the singularity problem of LDA, Li et al. [4] used the difference of both between-class scatter and within-class scatter as discriminant criterion, called maximum margin criterion (MMC). MMC has the advantages of effectiveness and simplicity.

Recent studies [5–7] have shown that the high-dimensional data possibly resides on a nonlinear sub-manifold. However, both PCA and LDA effectively see only the global Euclidean structure. When they are applied to face recognition, they fail to discover the underlying structure, if the face images lie on a nonlinear sub-manifold hidden in the image space. Some nonlinear

* Corresponding author.
E-mail address: huangpu3355@163.com (P. Huang).

1047-3203/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jvcir.2013.11.007

techniques have been proposed to discover the nonlinear structure of the manifold. The basic assumption of manifold learning is that the input data lie on a smooth low-dimensional manifold. Each manifold learning based method attempts to preserve a different geometrical property of the underlying manifold. The representative ones include Isomap [5], LLE [6], Laplacian Eigenmap [7] and local tangent space alignment (LSTA) [8]. These nonlinear methods do yield impressive results on some benchmark artificial data sets. However, they yield maps that defined only on the training data points and how to evaluate the maps on novel test data points remains unclear. To overcome this limitation, He et al. extended Laplacian Eigenmap to its linearized version, i.e. locality preserving projection (LPP) [9–13] for an explicit map. LPP attempts to construct a nearest neighbor graph and then evaluate the low-dimensional embedding to best preserve local structure of the data set.

Although LPP is effective in many domains, it is unsupervised and its unsupervised nature restricts its discriminating capability. To consider class label information in LPP, several supervised LPP methods [14–21] have been developed. Local discriminant embedding (LDE) [15] and marginal fisher analysis (MFA) [16], whose objective functions are very similar, can also be viewed as supervised LPP methods. This is because their training phases both exploit the class label information of samples. They are derived by using a motivation partially similar to LPP and each of them is based on an eigen-equation formally similar to the eigen-equation of LPP. On the other hand, since LDE and MFA partially borrow the idea of discriminant analysis and try to produce satisfactory linear separability, their ideas are also somewhat different from the idea of preserving the local structure of LPP. LDE and MFA can be viewed as two combinations of the locality preserving technique and the linear discriminant analysis [22]. Compared with LDA, both LDE and MFA do not depend on the assumption that the data of each class is Gaussian distributed and can obtain more available projection directions and better characterize the separability of different classes.

The purpose of LPP is to preserve the proximity relationship of the input data. In LPP, by applying k nearest neighbor (k-NN) criterion, any point and its k nearest neighbors are viewed as located on a super-plane, where all the descriptions in linear space can be performed. A common problem with the classical LPP and several supervised LPP methods [14,17,18] is that they might not necessarily discover the most discriminative manifold for pattern classification tasks because the manifold learning is originally modeled based on a characterization of "locality", a model that has no direct connection to classification. This is unproblematic for existing LPP algorithms as they seek to model a simple manifold, for example, to recover an embedding of one person's face images. In face recognition each person forms his or her own manifold in the feature space [23]. If one person's face images do exist on a manifold, different persons' face images could lie on different manifolds. If the images needed to be classified reside on multi-manifolds and two or more models have a common axis, then the locality preserving algorithms of manifold learning may result in overlapped embedding belonging to different classes because to recognize faces it would be necessary to distinguish between images from different manifolds. This problem is referred to as "overlearning of locality" [24].

In order to solve the problem of "overlearning of locality", Yang et al. proposed an unsupervised discriminant projection (UDP) [25] method, which can be viewed as simplified LPP on the assumption that the local density is uniform [26]. In the proposed method, locality and non-locality are discussed in detail, where locality means the sum of the squared distance between the points in k nearest neighbors, and the non-locality denotes the sum of the squared distance between two points not belonging to any k nearest neighbors. In order to achieve a discriminative map, UDP aims to find a linear transformation that maximizes the ratio of the non-locality to the locality. In the literature [27], there is another algorithm named locally preserving and globally discriminant projection with prior information (LPGDP) introduced to address this problem. The LPGDP method utilizes prior misclassification rate of between-class in the training data for the global discriminant measure while using class labels for preserving locality. Besides, Li et al. proposed a linear multi-manifolds learning based approach called constrained maximum variance mapping (CMVM) [28]. CMVM aims at globally maximizing the distances between different manifolds. After the local scatters have been characterized, the CMVM algorithm focuses on developing a linear transformation that maximizes the dissimilarities between all the manifolds under the constraint of locality preserving.

As discussed above, when LPP is used to map the high-dimensional data into a low-dimensional feature space, it may produce high between-class overlaps because of the "overlearning of locality". To solve this problem, the methods including UDP, LPGDP and CMVM seek to find a transformation that separates different manifolds after the local structure has been characterized. It is unproblematic for these methods to effectively separate different classes when the data distributed on a manifold have the same label. However, in practice, the local scatter is usually constructed according to the k-NN criterion, which will bring another problem. It is that, when there is large variation within the same class, the within-class variation may be larger than the between-class variation, which means that the neighbor relationship measured by the k-NN criterion may be distorted. In other words, data samples residing on a manifold possibly have different labels. In this case, these methods may not work well because of their common assumption that the data distributed on a manifold have the same label.

In this paper, we propose an effective supervised manifold learning algorithm, called local maximal margin discriminant embedding (LMMDE) for feature extraction and recognition. The proposed algorithm LMMDE incorporates LPP and MMC for data analysis. Similar to MFA, LMMDE characterizes intra-class compactness and inter-class separability to maximize the margins between different classes. One difference between MFA and the proposed method lies that MFA neglects the local structure based on the overall samples which may be helpful for classification. In addition, both CMVM and LMMDE have the common purpose that is to take class label information into account based on the property of locality preserving, but they are essentially different because: (1) CMVM is originally designed to separate different manifolds based on the assumption that the data distributed on a manifold have the same label, while LMMDE is designed to reduce the between-class overlaps based on the assumption that the data distributed on a manifold may have different labels and (2) CMVM characterizes only the inter-class separability in a global way, while LMMDE measures both the inter-class separability and the intra-class compactness in a local way like MFA.

The rest of this paper is structured as follows: In Section 2, the PCA, LDA, LPP are briefly reviewed. Section 3 describes the proposed algorithm in detail. In Section 4 the proposed algorithm is examined on three data sets and the experimental results are offered. Section 5 finishes this paper with some conclusions.

2. Outline of PCA, LDA, LPP

Let us consider a set of n samples $\{x_1,\ldots,x_n\}$ takes values in an N-dimensional image space, and assume that each image belongs to one of C classes. Let us also consider a linear transformation that maps the original N-dimensional space into a d-dimensional feature space, where $N > d$. The new feature vectors in the d-dimensional space are defined by the following linear transformation:

$$y_k = A^{\mathrm{T}} x_k, \quad k = 1, \ldots, n \tag{1}$$