Digital Signal Processing 76 (2018) 84–93

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Collaborative representation based local discriminant projection for feature extraction

Pu Huang [a,b,*], Tao Li [a,b], Guangwei Gao [c,e], Yazhou Yao [d], Geng Yang [a,b]

[a] School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China
[b] Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
[c] Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China
[d] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China
[e] Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Nanjing University of Science and Technology, Nanjing 210094, China

A R T I C L E I N F O

Article history:
Available online xxxx

Keywords:
Feature extraction
Graph construction
Manifold learning
Collaborative representation
Face recognition

A B S T R A C T

This paper introduces a novel dimensionality reduction algorithm, called collaborative representation based local discriminant projection (CRLDP), for feature extraction. CRLDP utilizes collaborative representation relationships among samples to construct adjacency graphs. Different from most graph-based algorithms which manually construct the adjacency graphs, CRLDP is able to automatically construct the graphs and avoid manually choosing nearest neighbors. In CRLDP, two graphs (the within-class graph and the between-class graph) are constructed. Based on the two constructed graphs, the within-class scatter and the between-class scatter are computed to characterize the compactness and separability of samples, respectively. Then CRLDP seeks to find an optimal projection matrix to maximize the ratio of the between-class scatter to the within-class scatter. Experimental results on ORL, AR and CMU PIE face databases validate the superiority of CRLDP over other state-of-the-art algorithms.

1. Introduction

Feature extraction (or dimensionality reduction) [1–3] has become an essential step in computer vision and pattern recognition areas, which seeks to extract the distinctive features of the data by mapping the original data into a low-dimensional subspace. Principle component analysis (PCA) [4] and linear discriminant analysis (LDA) [5] are the two most well-known dimensionality reduction (DR) methods. However, they failed to discover the underlying manifold structure of the data.

In recent years, graph-based DR methods [6–12], represented by locality preserving projection (LPP) [6], marginal discriminant analysis (MFA) [7], and local discriminant embedding (LDE) [8], have attracted much attention for feature extraction. LPP seeks to find a set of projection axes such that the neighborhood structure of the data can be preserved after projection. However, LPP is an unsupervised method which neglects the discriminant structure of the data. Different from LPP, both MFA and LDE are supervised methods, and they seek to find a subspace where the neighboring data points from the same class are close to each other and the neighboring data points from different classes are separated from each other. Compared with LPP, MFA and LDE not only consider the neighborhood structure of the data but also consider the discriminant structure of the data. From the viewpoint of classification, both MFA and LDE could receive better performance than LPP for pattern recognition tasks.

In graph-based DR methods, the key problem is how to construct adjacency graphs to discover the intrinsic structure of the data. Graph construction usually involves two steps: the first step is to determine the neighborhood relationships between samples, and the second step is to set edge weights between sample pairs. There are two popular ways to determine neighborhood of samples, one is the k nearest neighbor method, and the other is the ε-ball method, where, for each datum, the samples within its surrounding ε-ball are connected. Once the neighborhood is determined, then several approaches, e.g., binary, Gaussian-kernel and L2-reconstruction [13], can be used to set the graph edge weights. Unfortunately, it is quite difficult to manually construct graph due to the difficulty of parameter selection. Accordingly, how to automatically construct the adjacency graphs becomes particularly important.

More recently, sparse representation based techniques have received much attention in the fields of computer vision and pattern recognition. In [14], Wright et al. proposed a sparse representation based classifier (SRC) for face recognition and they showed

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

https://doi.org/10.1016/j.dsp.2018.02.009
1051-2004/

that SRC could effectively handle facial images with occlusion and illumination variations. The success of SRC greatly facilitated the research of sparse representation based theories and applications [15–19]. Zhang et al. [20] analyzed the mechanism of SRC and they pointed out that it is the collaborative representation, not the L1-norm sparsity constraint, that makes SRC effective. To this end, they proposed a collaborative representation based classification (CRC) method and demonstrated that CRC could achieve competitive results but with significantly less time than SRC.

Due to the promising property of sparse representation, many researchers focus on learning DR algorithms [21–27] based on sparse representation. For example, Qiao et al. [21] proposed a sparsity preserving projection (SPP) method for feature extraction. SPP aims to preserve the sparse reconstruction relationship of the data, which is achieved by solving a L1 regularization related objective function like SRC. SPP inherits some merits of SRC, e.g., containing natural discriminant information, choosing neighborhood automatically etc. Besides, Chen et al. [13] gave a L1-graph construction method, and Zhang et al. [22] presented graph optimization for dimensionality reduction with sparsity constraints (GODRSC) which aims to simultaneously seek the sparse representation coefficients and the projection matrix. In addition, some other sparse representation based algorithms, such as discriminant sparse neighborhood preserving embedding (DSNPE) [23], weighted discriminative sparsity preserving embedding (WDSPE) [24], and multilinear sparse principal component analysis (MSPCA) [25], have also been proposed for feature extraction.

The sparse representation (SR) based DR algorithms were designed via the L1-graph which owns several excellent merits such as sparsity, robustness to noise, and automatically neighborhood determination. However, it is much expensive due to solving the L1-norm minimization problem. To alleviate this problem, Yang et al. [28] used the idea of CRC to construct graph, i.e., the L2-graph. L2-graph calculates the edge weights using the overall samples, and could avoid manually choosing the nearest neighbors. Based on L2-graph, Yang et al. proposed collaborative representation based projections (CRP) like SPP. CRP shows competitive results with SPP, and it is much faster than SPP because the weights are obtained by solving a simple L2-norm minimization problem.

Despite the success of CRP, there are still some limitations. First, CRP is an unsupervised method in nature. In supervised scenario, the prior label information is available, and encoding the label information could significantly improve the performance. Second, CRP seeks to preserve the local compactness of samples. However, the local compactness information is obtained by solving a L2-norm regularization problem which is much weaker than the L1-norm sparsity constraint. This implies that many samples from different classes are probably clustered together after projection. Third, CRP utilizes the global scatter to characterize the separability of all the samples, which is unreasonable since it is always expected that samples from the same class are close to each other while samples from different classes are far away from each other.

Motivated by the above discussion, in this paper, we propose a novel DR algorithm, called collaborative representation based local discriminant projection (CRLDP), for feature extraction. Similar to CRP, CRLDP also utilizes collaborative representation relationships among samples to construct adjacency graphs. In CRLDP, two graphs called the within-class graph and the between-class graph are constructed. Based on the two constructed graphs, the within-class scatter and the between-class scatter are computed to characterize the compactness and separability of samples, respectively. Then CRLDP seeks to find an optimal projection matrix to maximize the ratio of the between-class scatter to the within-class scatter. It is worthwhile to highlight the novelties of our method as follows:

(1) Different from CRP, CRLDP is a supervised method. CRLDP takes full advantage of the label information which is beneficial for classification tasks.

(2) Like CRP, CRLDP constructs graphs using the L2-graph. CRLDP can automatically set the edge weights and avoid predetermining the neighborhood of each sample, and find a stable solution. Moreover, it is much faster than the sparse representation based techniques (e.g. SPP).

(3) CRLDP explicitly considers the local compactness of samples sharing the same label and local separability of samples with different labels. Thus CRLDP could effectively detect the discriminant structure of the data which is favorable for classification.

The remainder of this paper is organized as follows. In Section 2, we briefly review some related methods (LDE, CRC and CRP). In Section 3, we introduce CRLDP in detail. In Section 4, we compare CRLDP and other DR methods. In Section 5, we carry out experiments on the benchmark datasets such as ORL, AR and CMU PIE face databases to demonstrate the superiority of our method. In Section 6, conclusions and future work are made to summarize this paper.

2. Related work

Given a set of n training samples $X = [x_1, x_2, \ldots, x_n] \in \mathrm{R}^{m\times n}$, where each sample x_i $(i = 1, 2, \ldots, n)$ denotes an n-dimensional column vector. In this section, we briefly review some related methods including LDE [8], CRC [20] and CRP [28].

2.1. LDE

In LDE, two adjacency graphs, denoted by $G = \{X, W\}$ and $G' = \{X, W'\}$, are constructed, where the data set X corresponds to the vertices, W and W' are the weight matrices used to characterize the similarity between data pairs.

The graph G connects each sample with its homogeneous neighboring samples, and W is defined as follows:

$$W_{ij} = \begin{cases} \exp[-\|x_i - x_j\|^2/t], \\ \quad \text{if } l_i = l_j, x_i \in N_w(x_j) \text{ or } x_j \in N_w(x_i) \\ 0, \quad \text{otherwise} \end{cases} \tag{1}$$

where $\|x_i - x_j\|$ denotes Euclidean distance between x_i and x_j, $t > 0$ is a parameter, l_i is the class label of x_i, and $N_w(x_i)$ represents the set of k homogeneous nearest neighbors of x_i.

The other graph G' connects each sample with its heterogeneous neighboring samples, and W' is defined by:

$$W'_{ij} = \begin{cases} \exp[-\|x_i - x_j\|^2/t], \\ \quad \text{if } l_i \neq l_j, x_i \in N_b(x_j) \text{ or } x_j \in N_b(x_i) \\ 0, \quad \text{otherwise} \end{cases} \tag{2}$$

where $N_b(x_i)$ is the set of k' heterogeneous nearest neighbors of x_i.

Let $P \in \mathrm{R}^{m\times d} (d < m)$ denote the projection matrix. Using the graph embedding theory [7], the within-class scatter S_w and between-class scatter S_b can be defined as:

$$S_w = \sum_{ij} \left\| P^{\mathrm{T}} x_i - P^{\mathrm{T}} x_j \right\|^2 W_{ij} = 2\,\mathrm{tr}\left\{ P^{\mathrm{T}} X (D - W) X^{\mathrm{T}} P \right\} \tag{3}$$

and

$$S_b = \sum_{ij} \left\| P^{\mathrm{T}} x_i - P^{\mathrm{T}} x_j \right\|^2 W'_{ij} = 2\,\mathrm{tr}\left\{ P^{\mathrm{T}} X (D' - W') X^{\mathrm{T}} P \right\} \tag{4}$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator of a matrix, D and D' are diagonal matrices, and their elements are computed as: $D_{ii} = \sum_j W_{ij}$ and $D'_{ii} = \sum_j W'_{ij}$.