Knowledge-Based Systems 89 (2015) 47–55

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Discriminant sparse local spline embedding with application to face recognition

Ying-Ke Lei [a,b,c,*], Hui Han [a], Xiaojun Hao [a]

[a] The State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System (CEMEE), Luoyang, Henan 471003, China
[b] Science and Technology on Communication Information Security Control Laboratory, Jiaxing, Zhejiang 314000, China
[c] Electronic Engineering Institute, Hefei, Anhui 230037, China

ARTICLE INFO

Article history:
Received 25 September 2014
Received in revised form 16 June 2015
Accepted 24 June 2015
Available online 29 June 2015

Keywords:
Sparse neighborhood graph
Local spline embedding
Sparse subspace learning
Maximum margin criterion
Discriminant sparse local spline embedding
Manifold learning
Face recognition

ABSTRACT

In this paper, an efficient feature extraction algorithm called discriminant sparse local spline embedding (D-SLSE) is proposed for face recognition. A sparse neighborhood graph of the input data is firstly constructed based on a sparse representation framework, and then the low-dimensional embedding of the data is obtained by faithfully preserving the intrinsic geometry of the data samples based on such sparse neighborhood graph and best holding the discriminant power based on the class information of the input data. Finally, an orthogonalization procedure is perfomred to improve discriminant power. The experimental results on the two face image databases demonstrate that D-SLSE is effective for face recognition.

1. Introduction

It is well known that there are large volumes of high-dimensional data in numerous real-world applications. Operating directly on such high-dimensional image space is ineffective and may lead to high computational and storage demands as well as poor performance. From the perspective of pattern recognition, dimensionality reduction is a typical way to circumvent the "curse of dimensionality" problem [1] and other undesired properties of high-dimensional spaces. The goal of dimensionality reduction is to construct a meaningful low-dimensional representation of high-dimensional data. Ideally the reduced representation in the low-dimensional space should have a dimensionality that corresponds to the intrinsic dimensionality of the data.

Researchers have developed many useful dimensionality reduction techniques. These techniques can be broadly categorized into two classes: linear and nonlinear. Classical linear dimensionality reduction approaches seek to find a meaningful low-dimensional subspace in a high-dimensional input space by linear transformation. This subspace can provide a compact representation of high-dimensional input data when the intrinsic structure of data embedded in the input space is linear. Among them, the most well known are principal component analysis (PCA) [2] and linear discriminant analysis (LDA) [3]. Linear models have been extensively used in pattern recognition and computer vision areas and have become the most popular techniques for face recognition [4–8].

Linear techniques, however, may fail to discover the intrinsic structures of complex nonlinear data. In order to address this problem, a number of nonlinear manifold learning techniques have been proposed under the assumption that the input data set lies on or near some low-dimensional manifold embedded in a high-dimensional unorganized Euclidean space [9]. The motivation of manifold learning techniques is straightforward as it seeks to directly find the intrinsic low-dimensional nonlinear data structures hidden in the observation space. Examples include isometric feature mapping (ISOMAP) [10], locally linear embedding (LLE) [11], Laplacian eigenmaps (LE) [12], Hessian-based locally linear embedding (HLLE) [13], maximum variance unfolding (MVU) [14], manifold charting [15], local tangent space alignment (LTSA) [16], Riemannian manifold learning (RML) [17], and local spline embedding (LSE) [18], elastic embedding (EE) [19], Cauchy graph embedding (CGE) [20], adaptive manifold learning [21], and neighborhood preserving polynomial embedding (NPPE) [22].

* Corresponding author at: The State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System (CEMEE), Luoyang, Henan 471003, China. Tel.: +86 0551 65591108.
E-mail address: leiyingke@163.com (Y.-K. Lei).

http://dx.doi.org/10.1016/j.knosys.2015.06.016
0950-7051/

Each manifold learning algorithm attempts to preserve a different geometrical property of the underlying manifold. Local approaches, such as LLE, HLLE, LE, LTSA, and LSE, aim to preserve the proximity relationship among the data, while global approaches like ISOMAP and LOGMAP aim to preserve the metrics at all scales. Some experiments have shown that these methods can find perceptually meaningful embeddings for face or digit images. They also do yield impressive results on other artificial and real-world data sets. However, these manifold learning methods have to confront with the out-of-sample problem when they are applied to pattern recognition. They can yield an embedding directly based on the training data set, but, because of the implicitness of the nonlinear map, when applied to a new sample, they cannot find the image of the sample in the embedding space. It limits the applications of these algorithms to pattern recognition problems. To overcome the drawback, Bengio et al. proposed a kernel method to embed the new data points by utilizing the generalization ability of Mercer kernel [23]. He et al. proposed a method named locality preserving projection (LPP) to approximate the eigen-functions of the Laplace–Beltrami operator on the manifold and the new testing points can be mapped to the learned subspace without trouble [24]. Yan et al. utilized the graph embedding framework for developing a novel algorithm called marginal Fisher analysis (MFA) to solve the out-of-sample problem [25].

Recently, sparse representation has attracted considerable interests in machine learning and pattern recognition. Some researchers proposed some new methods integrating the theory of sparse representation and subspace learning. They are considered as a special family of dimensionality reduction methods which consider "sparsity". It has either of the following two characteristics: (1) Finding a subspace spanned by sparse base vectors. The sparsity is enforced on the projection vectors and associated with the feature dimensionality. The representative techniques are sparse principal component analysis (SPCA) [26] and nonnegative sparse PCA [27]. (2) Aiming at the sparse reconstructive weight which is associated with the sample size. The representative methods include sparse neighborhood preserving embedding (SNPE) [28] and sparsity preserving projections (SPP) [29].

In this paper, inspired by the idea of LSE [18] and sparse representation, we propose a novel sparse subspace learning technique, called discriminant sparse local spline embedding (D-SLSE). Specifically, A sparse neighborhood graph of the input data is firstly constructed based on a sparse representation framework, and then the low-dimensional embedding of the data is obtained by faithfully preserving the intrinsic geometry of the data samples based on such sparse neighborhood graph and best holding the discriminant power based on the class information of the input data. Finally, an orthogonalization procedure is perfomred to improve discriminant power. We now enumerate several characteristics of our proposed algorithm as follows:

(1) D-SLSE does not have to encounter setting the neighborhood size in constructing a neighborhood graph incurred in LSE. An unsuitable neighborhood may result in "short-circuit" edges (see Fig. 1a) or a large number of disconnected regions (see Fig. 1b). In contrast, graph construction based on sparse representation makes our proposed method very simple to use in practice.
(2) D-SLSE computes an explicit linear mapping from the input space to the reduced space, which attempts to manage the trade-off between holding discriminant power and preserving local geometry structure.
(3) D-SLSE seeks to find a set of orthogonal basis functions and significantly improves its recognition accuracy.

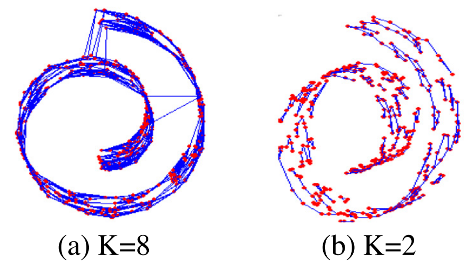

Fig. 1. The K-nearest neighborhood graph of Swiss roll data. (a) Short-circuit edge. (b) Disconnected regions.

The rest of this paper is organized as follows: The D-SLSE algorithm is developed in Section 2. Section 3 demonstrates the experimental results. Finally, conclusions are presented in Section 4.

2. Discriminant sparse local spline embedding

2.1. Local spline embedding

Xiang et al. [18] proposed a general dimensionality reduction framework called compatible mapping. They used the compatible mapping framework as a platform and developed a novel local spline embedding (LSE) manifold learning algorithm. This method includes two steps: part optimization and whole alignment. Each data point is represented in different local coordinate systems by part optimization. But its global coordinate should be maintained unique. Whole spline alignment is used to achieve this goal. The algorithmic procedure is listed as follows:

1. **Constructing the adjacency graph**: Let G denote a graph with n nodes. We use KNN criterion to construct the adjacency graph, i.e., putting an edge between nodes i and j *if* i is among k nearest neighbors of j or j is among k nearest neighbors of i.
2. **Obtaining tangent coordinates**: For each data point x_i, Let $X_i = [x_{i_1}, x_{i_2}, \ldots, x_{i_k}] \in R^{D\times k}$ denote its k nearest neighbors. Perform a singular decomposition of the centralized matrix of X_i, we have

$$X_i H_k = U_i \begin{bmatrix} \sum_i \\ 0_{(D-k)\times k} \end{bmatrix} V_i^T, \quad i = 1, \ldots, n, \tag{1}$$

where $H_k = I - e_k e_k^T/k$ is the centering operator, I is a $k \times k$ identity matrix, e_k is a k-dimensional vector with $e_k = [1, 1, \ldots, 1]^T \in R^k, \sum_i = diag(\sigma_1, \ldots, \sigma_k)$ contains the singular values in descending order. U_i is a $D \times D$ matrix whose column vectors are the left singular vectors, and V_i is a $k \times k$ matrix whose column vectors are the right singular vectors. The local tangent coordinates Θ_i of X_i can be obtained from the following formula:

$$\Theta_i = (U_i)^T X_i H_k = [\theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_k^{(i)}], \quad i = 1, \ldots, n, \tag{2}$$

where $\theta_j^{(i)}$ is the local tangent coordinate of the jth nearest neighbor of the data point x_i.
3. **Aligning global coordinates**: For the ith local tangent space projection Θ_i, let $Y_i = [y_{i_1}, y_{i_2}, \ldots, y_{i_k}] \in R^{d\times k}$ contain the corresponding global coordinates of the k data points. Further, denote the rth row of Y_i by $[y_{i_1}^{(r)}, y_{i_2}^{(r)}, \ldots, y_{i_k}^{(r)}]$. We determine the d spline functions $g_i^{(r)} : R^d \mapsto R, r = 1, 2, \ldots, d$, such that the coordinate components can be faithfully mapped:

$$y_{i_j}^{(r)} = g_i^{(r)}(\theta_j^{(i)}), \quad j = 1, 2, \ldots, k. \tag{3}$$