J. Vis. Commun. Image R. 38 (2016) 735–744

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Intrinsic structure based feature transform for image classification ☆

Zhengjuan Zhou [a,*], Jadoon Waqas [b]

[a] College of Computer Science, Sichuan University, Chengdu 610065, PR China
[b] Department of Computer Science, COMSATS Institute of Information Technology, Abbottabad 22060, Pakistan

A R T I C L E   I N F O

Article history:
Received 1 January 2016
Revised 11 March 2016
Accepted 20 April 2016
Available online 23 April 2016

Keywords:
Dimensionality reduction
Manifold learning
Subspace learning
Image classification

A B S T R A C T

Most dimensionality reduction works construct the nearest-neighbor graph by using Euclidean distance between images; this type of distance may not reflect the intrinsic structure. Different from existing methods, we propose to use sets as input rather than single images for accurate distance calculation. The set named as neighbor circle consists of the corresponding data point and its neighbors in the same class. Then a supervised dimensionality reduction method is developed, i.e., intrinsic structure feature transform (ISFT), it captures the local structure by constructing the nearest-neighbor graph using the Log-Euclidean distance as measurements of neighbor circles. Furthermore, ISFT finds representative images for each class; it captures the global structure by using the projected samples of these representatives to maximize the between-class scatter measure. The proposed method is compared with several state-of-the-art dimensionality reduction methods on various publicly available databases. Extensive experimental results have demonstrated the effectiveness of the proposed approach.

1. Introduction

Objects such as images are usually represented as high-dimensional vectors in applications. Direct computation by using high dimensional data not only bears huge computational burden, but also impacts the accuracy of final results. There is an increasing interest in dimensionality reduction methods among research community to address these issues [1,2]. These methods represent the data in a lower dimensional space without loss of information. Broadly the dimensionality reduction techniques can be grouped into two categories, i.e., linear and non-linear approaches.

The three classical linear dimensionality reduction techniques are: principal component analysis (PCA) [3,4], multidimensional scaling (MDS) [5] and Linear Discriminant Analysis (LDA) [4]. PCA performs dimensionality reduction by projecting the original d-dimensional data onto the $r \ll d$ dimensional linear subspace spanned by the leading eigenvectors of the data's covariance matrix. MDS finds an embedding that preserves pairwise distances between data points, and it is equivalent to PCA when those distances are Euclidean. LDA searches for the projective axes on which the data points of different classes are far from each other (maximize between class scatter), while constraining the data points of the same class to be as close to each other as possible (minimizing within class scatter). Although these linear methods can be applied to handle linear separable data, they might fail to discover the non-linear data structure [6]. Moreover, these techniques only preserve the global structure of the data, and it has been proved that local structure embedded in high-dimensional vector plays crucial role to characterize both the intrinsic and discriminative structure of data [7–11].

To explore the nonlinear structures, algorithms that can generate nonlinear mappings have also been developed, such as self-organizing map [12], autoencoder neural network [13], generative topographic map [14], principal curve and manifold [15–21] and kernel PCA [22,23]. However, most of these algorithms may only output a local minimum far away from the true nonlinear structure hidden in the data. Following this, manifold learning-based methods are developed under the assumption that the data points are sampled from an underlying manifold embedded in a high-dimensional Euclidean space.

The most representative manifold learning methods include Laplacian Eigenmap (LE) [7], Isometric feature mapping (Isomap) [10], and Local Linear Embedding (LLE) [8]. Because of the locality-preserving properties, these methods can straightforwardly be used to discover the nonlinear data structure hidden in the input space. However, a major issue arises in these methods is the out-of-sample problem [24]. In order to address this, numerous methods have been proposed, e.g., Locality Preserving Projection (LPP) [25], Neighborhood Preserving Embedding (NPE) [26], and Isometric Projection [27], which are linear approximation to

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail address: zhengjuanzhou@yeah.net (Z. Zhou).

http://dx.doi.org/10.1016/j.jvcir.2016.04.016
1047-3203/

these nonlinear algorithms. Furthermore, to effectively utilize the class label information, Zhang et al. presented Discriminant Neighborhood Embedding (DNE) [28]. It maps nearby points having the same class label in the high-dimensional space to nearby points with the low-dimensional representation. Moreover, it maps nearby points having different class labels to be as distant as possible in the reduced space.

Although these manifold based learning approaches have different motivations, the underlying mechanism is similar to graph embedding technique. The graph embedding directly preserves the local structure information by constructing the neighborhood graph, and then retains the graph structure in the projection for lower-dimensional space. A key step in these approaches is the distance measurements for the neighborhood calculations. All these methods construct the nearest-neighbor graph using, for example, Euclidean distance to calculate the distance between images to select neighbors. The smaller the distance value, the much closer these images are to each other. However in practice, the Euclidean distance cannot find correct neighbors since high-dimensional data such as images usually not lie on the Euclidean space.

Recently, many studies have demonstrated that sparse representation can well characterizes the local relationship among data [29–32]. Inspired by these works, various methods have been proposed to adaptively construct the adjacency graph by minimizing l_1-optimization problem and combined subspace learning for dimensionality reduction. Qiao et al. [33] proposed sparsity preserving projection (SPP), which has higher recognition accuracy than PCA and NPE. Zang and Zhang [34] introduced the label information in minimizing l_1-regularization objective function and then combined the sparse reconstruction error within class scatter of data to obtain the projection matrix for dimensionality reduction. Gui et al. [35] proposed a discriminant sparse neighborhood preserving embedding (DSNPE) algorithm by combining SPP and maximum margin criterion (MMC) for face recognition. However, these algorithms ignore or impair the local discriminant information embedded in data, which is very important for image classification. To address this issue, Gao et al. [36] proposed the discriminative sparsity preserving projections (DSPP), which employs sparse representation model to adaptively build both intrinsic adjacency graph and penalty graph with weight matrix. However, this work still needs other distance measure such as Euclidean distance to decide the ε-neighborhoods. The choice of parameter for accurate neighborhood selection will affect the classification results.

In this work, we propose accurate distance calculation for neighborhood selection based on image sets as input rather than single images. The image set that is named as neighbor circle consists of the data point and its corresponding neighbors in the same class. By considering that, an image may looks more similar with images of different classes than its own class if large variations exist, for instance, viewpoint and illumination in one class. Therefore, if variations are not considered in the calculation, the calculated distance between images from the same class may be larger than that from different classes. However, if the local neighbors of each data point in same class were treated as the background, then the neighboring points in same class have similar background, and the neighboring points in different classes have different background. In this way, if background is considered in the distance calculation, the distance between images from the same class will be smaller than that from different classes. In general, the distance between sets can be viewed as extending the distance between images to account for more general and complex data variations. Furthermore, local structure can be included in this kind of distance calculation.

Motivated by the above analysis, we propose a novel supervised subspace learning method called intrinsic structure feature transform (ISFT), which considers both local and global structure in an integrated modeling environment. We use LSR [37] to construct neighbor circles that can preserves local structure for every data point, and then formulated the problem for calculating the distance of neighbor circle as distances of data points lying on a Riemannian manifold spanned by SPD matrices, i.e., nonsingular covariance matrices. Finally, an embedding is learnt that not only preserves the local intrinsic structure of data points from same class, but also enhances the discrimination among data points from different classes. Moreover, global between-class structure is integrated into the discriminant manifold learning objective function for dimensionality reduction. Experiments on four popular image databases have demonstrated the effectiveness of the proposed method for classification.

The remainder of this work is organized as follows. Section 2 briefly summarizes LDA and DNE. Intrinsic structure based feature transform is presented in Section 3. The experimental results in four image databases are presented in Section 4. Conclusion and future works are presented in Section 5.

2. LDA and DNE

This section briefly reviews the LDA and DNE, serving as a ground for further developments. LDA focuses on preserving global structure of data, while DNE focuses on preserving local structure of data. Given an image set $X = [X_1, X_2, \ldots, X_c] = [x_i \in \mathbb{R}^d]_{i=1}^N$ drawn from a union of c subspaces, both methods try to find a transformation matrix $P \in \mathbb{R}^{d \times r}$ that can maps the original d-dimensional space to a new r-dimensional space through linear transform as follows,

$$y_i = P^T x_i, \ i = 1 \ldots N, \tag{1}$$

where $y_i \in \mathbb{R}^r$ and $r \ll d$.

2.1. LDA

The goal of LDA is to maximize the between-class measure while minimizing the within-class measure. The procedure can be summarized in three steps:

(1) Constructing within-class scatter matrix

$$S_w = \sum_{l=1}^{c} \sum_{i=1}^{N_l} (x_i^l - \mu_l)(x_i^l - \mu_l)^T, \tag{2}$$

where x_i^l is the ith sample of class l, c is the number of classes, N_l is the number of samples in class l, and $\mu_l = \frac{1}{N_l} X_l, X_l = \sum_{i=1}^{N_l} x_i^l$.

(2) Constructing between-class scatter matrix

$$S_b = \sum_{l=1}^{c} N_l (\mu_l - \mu)(\mu_l - \mu)^T, \tag{3}$$

where $\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$.

(3) Finding eigenmaps. Compute the eigenvectors and eigenvalues for the problem

$$S_b P = \lambda S_w P, \tag{4}$$

where P is formed from the optimal $r \leqslant c - 1$ eigenvectors corresponding to the r largest eigenvalues, i.e. $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_r$.