Neurocomputing 164 (2015) 300–306

Contents lists available at ScienceDirect

Neurocomputing

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

Bilateral filtering inspired locality preserving projections for hyperspectral images

Xinrong Li [a], Jing Pan [b,c,*], Yuqing He [b], Changshu Liu [b]

[a] Institute of Plant Nutrition and Natural Resources, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China
[b] School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China
[c] School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin 300222, China

ARTICLE INFO

Article history:
Received 29 July 2014
Received in revised form
3 January 2015
Accepted 10 January 2015
Communicated by D. Tao
Available online 20 April 2015

Keywords:
Locality preserving projections
Bilateral filtering
Dimensionality reduction
Subspace learning
Hyperspectral image classification

ABSTRACT

As a high-dimensional data, hyperspectral image contains rich information for agricultural remote sensing classification. *Locality preserving projections* (LPPs) have been widely used for extracting compact and discriminative information from such high-dimensional data. The objective function of LPP is formulated as a sum of the difference between transformed low dimensional vectors weighed by a function of the difference between images. The weights are crucial for LPP which enforce reduced feature vectors preserving the locality property in the original high dimensional space. In this paper, we borrow the idea of weight design of bilateral filtering to re-design the weights in LPP. The weights in bilateral filtering depend not only on the Euclidean distance of pixels (i.e., spatial weight) but also on the intensity differences (i.e., range weight). Analogously, we design the weights in our improved LPP (called bilateral LPP and abbreviated to BLPP) as a multiplication of a function of Euclidean distance $\|\mathbf{x}_i - \mathbf{x}_j\|$ of the original images (i.e., spatial weight) and a function of the Euclidean distance $\|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|$ of the features extracted from the images (i.e., range weight, a.k.a., feature weight). The spatial weight measures the similarity in spatial space whereas the feature weight measures the similarity in feature space which reveals the content of the images. Thus, the proposed BLPP utilizes both the spatial information and the image content information, which results in higher recognition rate. Experimental results on the Salinas and Indian Pine hyperspectral databases demonstrate the effectiveness of BLPP.

1. Introduction

Hyperspectral image, when concatenated as a long vector, is of high dimension which causes the curse of the dimensionality and limits the generalization ability of a agricultural remote sensing classification (e.g., crop recognition) system [1,49,50]. Therefore, developing proper dimensionality reduction techniques is a key for improving the performance. In this paper, we focus on linear dimensionality reduction methods.

Many linear dimensionality reduction methods were developed in the community of computer vision and pattern recognition where Eigenface, Fisherface, and Laplacianface are three classical dimensionality reduction methods. The cores of Eigenface, Fisherface, and Laplacianface are *principal component analysis* (PCA), *linear discriminant analysis* (LDA), and *locality preserving projections* (LPP), respectively [9–11]. PCA aims at minimizing reconstruction error whereas LDA targets at maximizing separability. Regularized LDA is able to deal with the singularity problem [51,42] and improve the generalization ability. As an unsupervised dimensionality reduction method, LPP is quite different from PCA [11,12,45]. The goal of LPP is to produce a subspace where the locality (i.e., neighborhood structure) of the original high dimensional data set is preserved in the subspace. One important contribution of LPP is that it reveals the connections to PCA and LDA from the point of view of graph embedding framework. Based on this framework, one can develop different methods by changing some of the connectiveness of the graph. *Marginal Fisher analysis* (MFA) proposed by Yan et al. [7] and *local Fisher discriminant analysis* (LFDA) proposed by Sugiyam [8] are two important instances of the graph embedding framework and can be regarded as supervised versions of LPP. More recent works include *supervised optimal locality preserving projection* (SOLPP) [41], *normalized Laplacian-based supervised optimal locality preserving projection* (NL-SOLPP) [41], constrained concept factorization [52], and non-negative local coordinate factorization [53]. Nonlinear methods are more powerful in capturing the intrinsic structure of the data at the cost of large complexity [54].

MFA and LFDA place emphasis on which pairs of graph vertexes should be connected or disconnected. In this paper, we concentrate on how to design the edge weights without changing the connectiveness status of the vertexes. In traditional LPP, the weight w_{ij} for the edge

* Corresponding author at: School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China.
E-mail address: jingpan23@gmail.com (J. Pan).

http://dx.doi.org/10.1016/j.neucom.2015.01.021
0925-2312/

determined by two neighboring vertexes (images) $\mathbf{x}_i \in \Re^D$ and $\mathbf{x}_j \in \Re^D$ is a function of their Euclidean distance $\|\mathbf{x}_i - \mathbf{x}_j\|$. The vector $\mathbf{x}_i$ is a point in the D-dimensional space. In this space, the contribution of $\mathbf{x}_j$ to $\mathbf{x}_i$ and hence w_{ij} is determined by the spatial distance between $\mathbf{x}_j$ and $\mathbf{x}_i$. This is similar to the Gaussian smoothing filter used in image denosing in the sense that only spatial weight (i.e., geometric closeness) is employed. It is known that bilateral filter proposed by Tomasi and Manduchi [26] is more effective than Gaussian smoothing filter in simultaneously smoothing noise and preserving edge information. The merit of bilateral filter stems from the fact that it makes use of both geometric closeness (i.e., spatial weight) and photometric similarity (i.e., range weight). Inspired by the success of bilateral filtering, we, in this paper, propose to redesign the edge weight by combining the geometric closeness $\|\mathbf{x}_i - \mathbf{x}_j\|$ (i.e., spatial weight) and image content similarity $\|f(\mathbf{x}_i) - \mathbf{f}(\mathbf{x}_j)\|$ (i.e., range weight, a.k.a. feature weight) where the function f extracts the content information of the image $\mathbf{x}$. Properly designing of f is able to make this bilateral filtering inspired LPP (called BLPP) much more robust to noise and geometric transformation (caused by alignment error). Importantly, because the discriminative information contained in $f(\mathbf{x}_i)$ and feature weight is used, BLPP also has higher recognition performance than LPP.

Specifically, the contributions of the paper are as follows:

1. A new way to design the edge weights of LPP is proposed. In classical LPP, an edge weight between two images reflects merely the distance in spatial domain. In contrast, the proposed method combines both the spatial weight and the range weight. The range weight contains the information of image content.
2. We formulate the problem of dimensionality reduction as a process of bilateral filtering and the idea of combining spatial and range weights are inspired by bilateral filtering. The method is effective for classification of hyperspectral images.

The rest of the paper is organized as follows. We first discuss related work in Section 2. Then, we describe LPP in Section 3. Subsequently, we introduce the proposed BLPP in Section 4. Experimental results are presented in Section 5. Finally, Section 6 draws conclusions based on these experimental results.

2. Related works

In this section, we briefly overview both dimensionality reduction methods and bilateral filtering methods because they are closely related to the proposed method.

2.1. Some dimensionality reduction methods

Based on whether and how the class label information is used, dimensionality reduction methods can be categorized into supervised, unsupervised, and semi-supervised classes. The original LPP is an unsupervised method. Two unsupervised methods, *neighborhood preserving projection* (NPP) [2,3] and *neighborhood preserving embedding* (NPE) [4], are inspired by LPP. NPP and NPE are similar to LPP in the sense of locality preserving. But in NPP and NPE, the weights between neighboring points are computed so that they can optimally construct the current point. The idea of locality preserving was also applied to *nonnegative matrix factorization* (NMF) by using KL-divergence for evaluating the similarity on the hidden topics [5,6] or representing each data point as a linear combination of only few nearby anchor points [14].

Many algorithms were also developed to include the information of class labels in order to improve LPP for higher recognition performance. MFA [7] and LFDA [8] are two classical supervised algorithms inspired by LPP. *Linear discriminant projections* (LDP), used for reducing the dimensionality of SIFT descriptor, can be viewed as a global version of MFA and it is insensitive to noise because it does not rely merely on the nearest neighbors [13].

In addition to locality preserving, sparsity preserving can also be imposed on dimensionality reduction [15,16]. Orthogonality is also useful to improve NPP [17,18]. Both LPP and NPP(NPE) can be extended to their nonlinear versions by kernel trick [17,19–21]. In [19], LDA and implicitly nonlinear mapping are combined in the framework of local linear embedding.

Under the framework of LPP and graph embedding, Lin et al. [20] proposed to use multiple kernel learning for learning a unified space of low dimension for data in multiple feature representation. But they employed the same edge weights as LDA [51] and LDE (*local discriminant embedding*) [22]. *Graph-optimized LPP* (GoLPP) [23] and *sparse neighborhood embedding* (SNE) [25] are flexible and optimal for constructing neighborhood graph. In the application of image re-ranking, relevance information can also be introduced into the graph embedding [44].

Dimensionality reduction (DR) methods can also be classified as vector based or tensor based methods according to whether an image is vectorized or not [46–48]. The above-mentioned methods are vector-based ones. *Orthogonal tensor neighborhood preserving* (OTNPE) regards two-dimensional images as points in the second-order tensor space and construct neighborhood graph with sparse construction [24].

2.2. Bilateral filtering

Smooth filtering is important for denoising in images. Before emergence of bilateral filtering, the most widely used method is Gaussian low-pass filtering which computes a weighted average of pixels' values in a small neighborhood. Because the weights decrease with distances from the neighborhood center, the edges nearby the center are blurred by the Gaussian low-pass filtering [26,29,30]. Bilateral filtering aims at averaging within smooth regions and not averaging across edges. Bilateral filtering measures the relationship of two pixels in terms of both closeness in the domain and similarity in the range. The output of bilateral filter at each pixel is also a weighted average of its neighbors. But different to the Gaussian filter, in bilateral filter the weight assigned to each neighbor decreases with both the distance in the image plane (i.e., the spatial domain) and the distance on the intensity axis (i.e., the range domain). Barash [27] discovered the fundamental relationship between the bilateral filtering adaptive smoothing and the nonlinear diffusion equation.

There has been rapid progress in fast algorithms of bilateral filtering. Durand and Dorsey [31] proposed to linearize the bilateral filter and use fast Fourier transforms for acceleration. Pairs and Durand [28] considered the bilateral as a higher-dimensional convolution followed by two nonlinearities.

The original and fast bilateral filters have been adopted for image denosing, relighting and texture manipulation, dynamic range compression, image enhancement, mesh fairing, volumetric denoising, optical flow and motion estimation, video processing [28], and image dehazing. To the best of our knowledge, it is never directly used for dimensionality reduction.

3. Locality preserving projections

The original *locality preserving projection* (LPP) is an unsupervised linear dimensionality reduction method [11,12]. As a classical unsupervised dimensionality reduction method, *principal component analysis* (PCA) aims to preserve the global structure of the data. In contrast, LPP seeks to preserve the intrinsic geometry and local structure of the data [11,12].

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \in \mathrm{R}^{D \times N}$ be the D-dimensional training data where N is the number of total training samples. The task of the