Neurocomputing 173 (2016) 196–211

Contents lists available at ScienceDirect

Neurocomputing

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

On some variants of locality preserving projection

Gitam Shikkenawis, Suman K. Mitra *

Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, India

ARTICLE INFO

Article history:
Received 1 July 2014
Received in revised form
10 November 2014
Accepted 15 January 2015
Available online 6 September 2015

Keywords:
Locality preserving projection
Subspace analysis
Dimensionality reduction
Data discrimination

ABSTRACT

High dimensional data is hard to interpret and work with in its raw form; hence dimensionality reduction is applied beforehand to discover underlying low dimensional manifold. Locality Preserving Projection (LPP) was introduced using the concept that neighboring data points in the high dimensional space should remain neighbors in the low dimensional space as well. In a typical pattern recognition problem, true neighbors are defined as the patterns belonging to same class. Ambiguities in regions having data points from different classes close by, less reducibility capacity and data dependent parameters are some of the issues with conventional LPP. In this article, some of the variants of LPP have been introduced that try to resolve these problems. A weighing function that tunes the parameters depending on data and takes care of the other issues is used in Extended version of LPP (ELPP). Better class discrimination is obtained using the concept of intra and inter-class distance in a supervised variant (ESLPP-MD). To capture the non-linearity of the data, Kernel based variants are used, that first map the data to feature space. Data representation, clustering, face and facial expression recognition performances are reported on a large set of databases.

1. Introduction

Number of pixels used to represent the image constitute dimensions of the image space in which each image of that space is considered as a data point. Out of many possible combinations in this space, only a few turn out to be meaningful images. Generally, the distribution of such images i.e. data points in the original space is not uniform and may seek representation in a lower dimensional subspace. It becomes difficult to deal with a very high dimensional data, be it classification, recognition or analysis task. Hence, image representation in subspace is advantageous. This reduces the storage space as well as the computational complexity.

One simple experiment can easily convince us about the image data residing compactly in the image space. The experiment exploits redundancy present in the image as discussed in Chen et al. [26]. The plots of log probabilities of horizontal and vertical derivatives of various kinds of images are shown in Fig. 1. It can be observed that the peak occurs at the values where derivatives are nearly zero which signifies high correlation among the neighboring pixels of the images. Hence, it can be inferred that the data, visually varying, in the high dimensional space actually resides in a very compact lower dimensional space.

To represent the images in a subspace having much lower dimensions, liner and non-linear dimensionality reduction techniques are present in literature. These techniques try to represent the original image in a much compact way keeping the information content intact. Principle Component Analysis [20,31] is one of the most popular dimensionality reduction methods. This linear approach finds the subspace from the data covariance information and preserves the directions of maximum variance. Linear Discriminant Analysis (LDA) [3], a supervised linear dimensionality reduction approach, maximizes inter class variability whereas minimizes intra-class variability in order to have better separation between different classes. Another popular liner approach is Independent Component Analysis (ICA) [18] that aims at making the components as independent as possible.

Data not always lies on a liner manifold, many a times, the manifold on which data lies happens to be non-linear. In such cases, linear dimensionality reduction methods fail to discover the non-linearity present in the data. Isomap [30] is a non-linear dimensionality reduction technique that preserves the intrinsic geometry of the data using geodesic distances between the data points. Like Isomap, Locally Linear Embedding (LLE) [25] finds non-linear manifold by stitching small linear neighborhoods. LLE finds a set of weights that perform local linear interpolations to closely approximate the data. But in these approaches, it is not

* Corresponding author at: DA-IICT Post Bag 4, Near Indroda Circle, Gandhinagar 382007, Gujarat, India. Tel.: +91 79 30510584; fax: +91 79 30520010.

E-mail addresses: 201221004@daiict.ac.in (G. Shikkenawis), suman_mitra@daiict.ac.in (S.K. Mitra).

URL: http://intranet.daiict.ac.in/~suman_mitra/newSite/ (S.K. Mitra).

http://dx.doi.org/10.1016/j.neucom.2015.01.100
0925-2312/

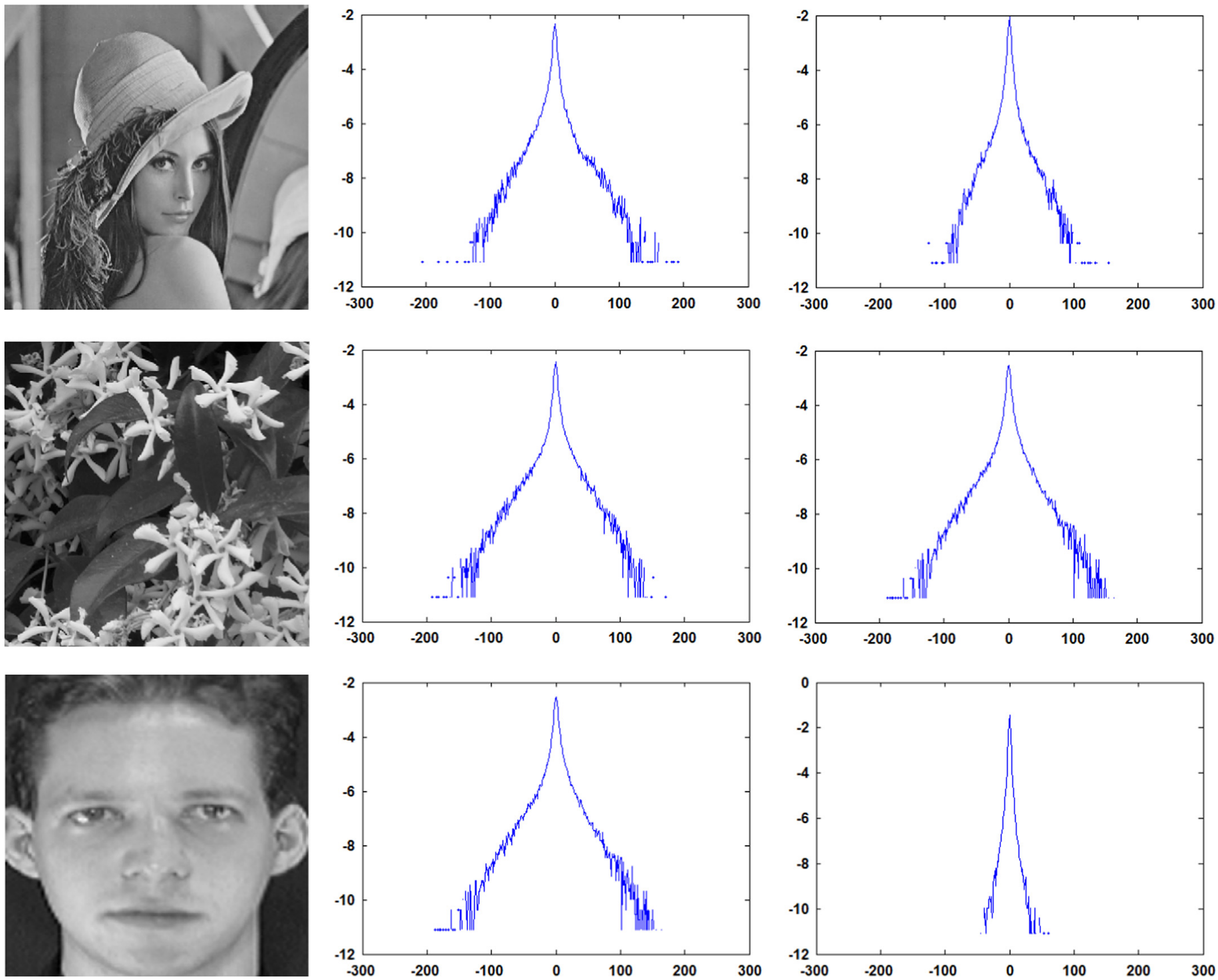

Fig. 1. Log probabilities of horizontal (column 2) and vertical (column 3) derivatives of different kind of images.

clear how to project a new data point in the lower dimensional projection space. Neighborhood Preserving Embedding (NPE) [15], is a linear approximation of LLE that aims to discover the local structure of the data manifold.

Locality Preserving Projection (LPP) [16,14] a linear dimensionality reduction approach, tries to capture the non-linearity present in the data using neighborhood information. Non-orthonormal basis vectors are obtained using notion of Laplacian of graph constructed by considering the data points i.e. images as nodes. Weight of the edges in the graph is assigned using the Euclidean distance between the data points in the original space. Orthogonal version of LPP, Orthogonal Locality Preserving Projection (OLPP) [7] produces the orthogonal basis functions with the aim of having more locality preserving power. Conventional LPP is sensitive to noise and outliers and depends highly on the parameters for constructing the neighborhood graph and the weight matrix. A few extensions [11,12,32,28,22,6] have been proposed to overcome these issues. Robust path based similarity is used in Enhanced Locality Preserving Projection [12] to obtain robustness against noise and outliers. Parameter free LPP [11] is also developed using Pearson correlation and adaptive neighborhood information. Ambiguities can take place in the regions having data points from two or more classes in case of LPP as it considers only a few nearest neighbors in order to preserve the local structure [29]. Supervised LPP [34] uses class labels to have better separation between different classes. The data points having same class labels are only considered as neighbors, thereby resolving the ambiguous situations.

To find the non-linear manifold of the data more precisely, various kernels are used to map the data non-linearity in the feature space before applying the conventional dimensionality reduction approaches such as Kernel PCA [1] and Kernel Discriminant Analysis [2]. LPP well preserves the local structure of the data, still to find the non-linear manifold of the data in a much better way, Supervised Kernel LPP [19] uses class label information as well as non-linear kernel mapping for better data discrimination.

The paper covers locality preserving projection (LPP) and some of its variants. The regions having data points from two or more different classes nearby may have ambiguous mappings in the LPP projection space. Also, values of parameters of LPP play important role in finding the data embedding and are data dependent. An extension of LPP (ELPP) taking care of these issues is explained in detail. A z-shaped weighing function, that automatically tunes the parameters depending on data is used. In face and facial expression recognition tasks, many a times the class labels of the training