Pattern Recognition 50 (2016) 74–87

Contents lists available at ScienceDirect

Pattern Recognition

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

Learning representations from multiple manifolds

Chan-Su Lee [a,*], Ahmed Elgammal [b], Marwan Torki [c]

[a] Department of Electronic Engineering, Yeungnam University, Gyeongsan, Republic of Korea
[b] Department of Computer Science, Rutgers, The State University of New Jersey, New Brunswick, USA
[c] Department of Computer and System Engineering, Alexandria University, Alexandria, Egypt

ARTICLE INFO

Article history:
Received 12 October 2014
Received in revised form
10 July 2015
Accepted 25 August 2015
Available online 5 September 2015

Keywords:
Manifold learning
Dimensionality reduction
Joint manifold representation
Correspondence

ABSTRACT

The problem we address in this paper is how to learn joint representation from data lying on multiple manifolds. We are given multiple data sets, and there is an underlying common manifold among the different data sets. Each data set is considered to be an instance of this common manifold. The goal is to achieve an embedding of all the points on all the manifolds in a way that preserves the local structure of each manifold and that, at the same time, collapses all the different manifolds into one manifold in the embedding space while preserving the implicit correspondences between the points across different data sets. We propose a framework to learn embedding of such data, which can preserve the intra-manifolds' local geometric structure and the inter-manifolds' correspondence structure. The proposed solution works as extensions to current state-of-the-art spectral-embedding approaches to handling multiple manifolds.

1. Introduction

Dimensionality reduction techniques have proven useful in many computer vision problems. In particular, nonlinear dimensionality reduction techniques [1–7] can achieve embedding of data lying on a nonlinear manifold by changing the metric from the original space to the embedding space, based on the manifold's local geometric structure. Many of the introduced nonlinear dimensionality reduction techniques are instances of graph spectral-embedding [8]. Spectral-embedding approaches, in general, construct an affinity matrix between data points that reflects the local manifold structure, i.e., constructing a graph with edge weights reflecting the local geometry of the manifold. Embedding is then achieved through solving an eigenvalue problem on the affinity matrix. Examples of nonlinear dimensionality reduction techniques in this category include: isometric feature mapping (Isomap) [1], locally linear embedding (LLE) [2], Laplacian eigenmaps [3], and manifold charting [4], among others. Bengio et al. [9] and Ham et al. [10] showed that these approaches are all instances of kernel-based learning, in particular, kernel principle component analysis (KPCA).

All these nonlinear embedding frameworks were shown to be able to embed data lying on a nonlinear manifold into a low-dimensional Euclidean space for toy examples, as well as for real images. Such approaches are able to embed image ensembles nonlinearly into low-dimensional spaces where various orthogonal perceptual aspects can be shown to correspond to certain directions or clusters in the embedding spaces. However, the application of such approaches is limited to embedding of a single manifold and, as we shall show, fails to embed data lying on multiple manifolds.

The problem we address in this paper is how to learn useful representation from data sets lying on multiple manifolds. We are given multiple data sets, and there is an underlying common manifold among the different data sets. Each data set is considered to be an instance of this manifold. For example, images of different objects with similar geometry are observed from different views, where the images of each object are one data set. The images of each object lie on the view manifold of that object, i.e., each data set is an instance of a view manifold of a different object. We can think of such manifolds as quasi-parallel in the space. Different objects' manifolds are distributed in the space. We can even think of such manifolds themselves as lying on a manifold (assume that we collapse each data set into a point). Another example comes from embedding motion manifolds for different people (e.g., consider data on different people walking). Each person's data set represents an instance of the "walking" manifold, while each person's walking manifold (as a whole) is lying on a different location in the input space. None of the existing manifold learning techniques can be used to learn such complex structures: both intra-manifold and inter-manifold structures.

* Corresponding author.
E-mail address: chansu@ynu.ac.kr (C.-S. Lee).

http://dx.doi.org/10.1016/j.patcog.2015.08.024

Our contribution: We propose a framework for learning embedded representation from different data sets, each assumed to lie on a nonlinear manifold. The embedding achieves representation of the common underlying manifold shared between the different data sets. The problem of learning a common embedded representation from multiple data sets becomes trivial if we assume that there are given correspondences between the different data sets, i.e., the data sets are aligned. In this paper, we do not assume that such correspondences are given. The results we achieve are superior to existing state-of-the-art embedding approaches when applied to such a setting. The proposed solution works as extensions for the current state-of-the-art spectral-embedding approaches, such as Isomap [11], LLE [2], and Laplacian eigenmaps [3], to handle multiple manifolds.

The organization of the paper is as follows. Section 2 discusses some motivating applications of the proposed approach. Section 3 presents related works. Section 4 presents the proposed joint manifold embedding with inter-manifold correspondence estimation. Section 5 presents some examples of experimental results on different data sets.

2. Motivation and problem definition

Suppose we are given K sets of data lying on K different manifolds, e.g., different people performing the same set of gestures or facial expressions. These data represent two different factors: body configuration variability, which typically lies on a low-dimensional manifold, and different people's variability, which might be higher in its dimensionality. One objective is to learn an embedding of the body-configuration manifold invariant to the person performing the motion. Learning such joint person-invariant body-configuration manifold embedding is essential for estimation of the intrinsic configuration, for providing dynamic models for tracking, and for recognition of gestures and activities.

Obviously, we can achieve embedding of each person's data set individually, which yields person-specific body configuration manifold embedding, as can be seen in Fig. 1. Such embedding will be different from one person to another, and will not be useful in any general tracking or recognition task where the goal is to track and recognize the facial expression. On the other hand, if we put all the data together in one set and try to embed them using any nonlinear embedding technique, we will not be able to achieve meaningful embedding either, since the inter-manifold distance between data for different people will be much larger than the intra-manifold distance (within one specific person). So, we will end up with K separate clusters in the embedding space, as can be seen from the bottom-right embedding in Fig. 1. In this example, a Gaussian process latent variable model (GPLVM) [5] is used to achieve the embedding.

Another example is shown in Fig. 2, where we used shapes representing side-view walking sequences for multiple people

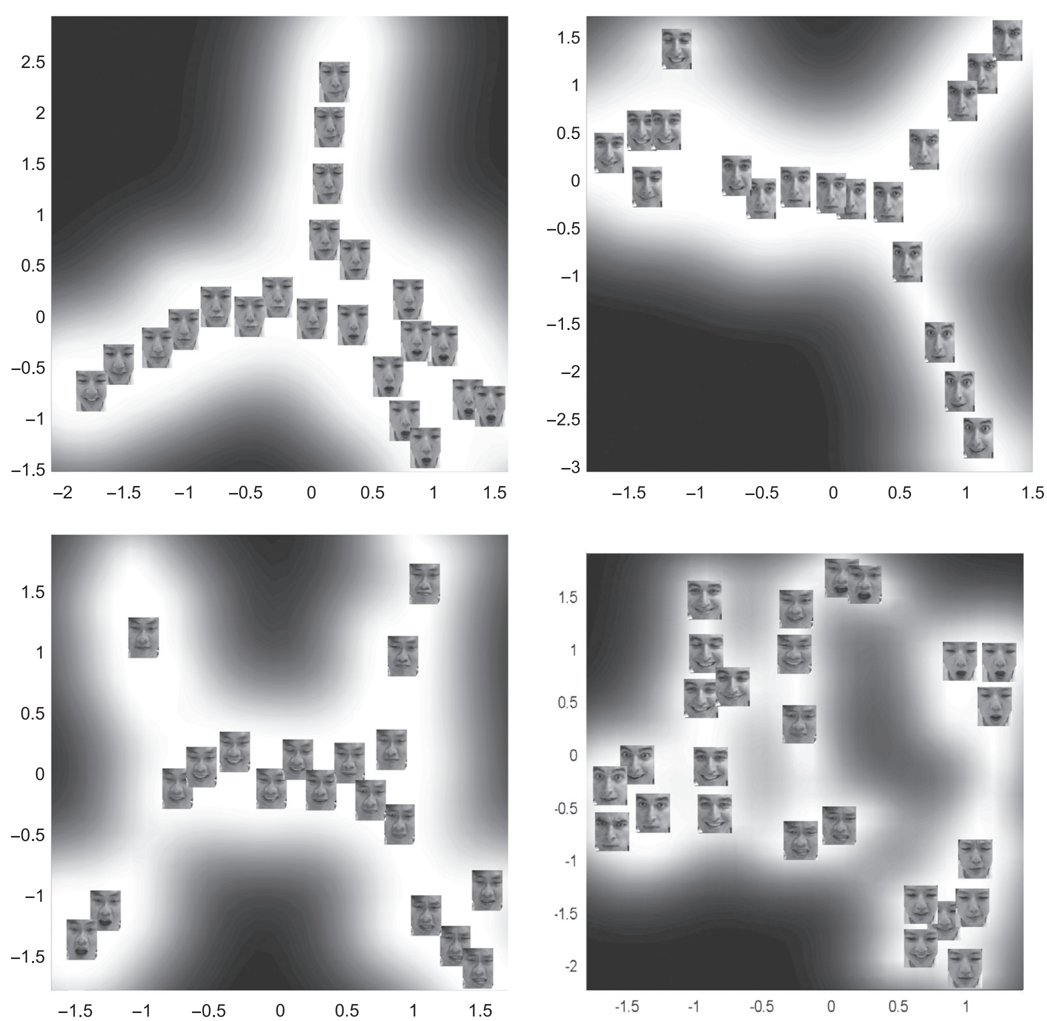

Fig. 1. Example embedding for facial expressions of three people: the first three plots show embedding for an individual person. The last plot shows embedding of three manifolds together, which is dominated by the inter-person manifold distance.