

Learning a Joint Manifold Representation from Multiple Data Sets

Marwan Torki, Ahmed Elgammal
Dept of Computer Science
Rutgers University
New Brunswick, NJ, USA
{mtorki,elgammal}@cs.rutgers.edu

Chan Su Lee
Department of Electronic Engineering
Yeungnam University
Gyeongsan, South Korea
chansu@ynu.ac.kr

Abstract—The problem we address in the paper is how to learn a 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 set. We propose a framework to learn an embedding of all the points on all the manifolds in a way that preserves the local structure on each manifold and, in 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. The proposed solution works as extensions to current state of the art spectral-embedding approaches to handle multiple manifolds.

I. INTRODUCTION

Dimensionality reduction techniques have proven useful in many computer vision problems. In particular, nonlinear dimensionality reduction (NLDR) techniques such as Isometric feature mapping (Isomap) [1], Local linear embedding (LLE) [2], Laplacian eigenmaps [3], and others [4], [5] can achieve embedding of data lying on a nonlinear manifold through changing the metric from the original space to the embedding space based on the manifold local geometric structure. 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. However, the application of such approaches is limited to embedding of a single manifold and they fail to embed data lying on multiple manifolds.

Our contribution: we propose a framework to learn an embedded manifold representation from multiple data sets. We are given multiple data sets and there is an underlying common manifold among the different data sets. An embedding can be achieved for each data set using any NLDR technique. However, if we put all the data sets together, no useful embedding can be achieved. This is because the inter-manifold distance between different data sets will be much larger than the intra-manifold distance within each data set. This can be seen in the example in Fig. 1 where GPLVM [5] is used to achieve the embeddings. The problem becomes trivial if we are given correspondences between the different data sets, i.e., the data sets are aligned, in such case

This research is partially funded by NSF CAREER award IIS-0546372

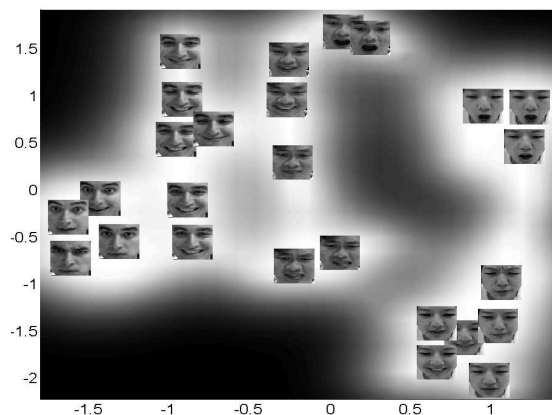


Figure 1. Embedding of three people’s facial expressions. As can be noticed, the embedding is dominated by the inter-person manifold distance.

corresponding data points can be stacked as a single point in a new space and any NLDR technique can be used. This is similar to the solution proposed by Ham et. al. in [6] for the problem of learning joint embedding across different spaces.

In this paper we do not assume that such correspondences are given. We solve the problem within the embedding framework where, from the data geometry, we can achieve the joint embedding. 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 [7], LLE [2], and Laplacian eigenmaps [3], to handle multiple manifolds.

II. JOINT MANIFOLD EMBEDDING

A. Inter- and Intra- Manifold Structure

The input is K different data sets in a D -dimensional space, denoted by $\mathbf{X}^k = \{\mathbf{x}_i^k \in \mathbb{R}^D, i = 1, \dots, N_k\}, k = 1, \dots, K$. The points on each data set are assumed to lie on a manifold and there are a common structure between the different manifolds. Each data set might have a different number of points. We denote by N_k the number of points in the data set k . Let N be the total number of points in all data

sets, i.e., $N = \sum_{k=1}^K N_k$.¹ The correspondences between the data sets are not known and, since the data sets are of different sizes, there is no one-to-one correspondences that can be assumed.

In such data, there are two different geometric structures: The intra-manifold structure and the inter-manifold structure. The intra-manifold structure is the local geometric structure within a given manifold. This is the geometric structure that is captured in any NLDR approach by constructing an affinity matrix \mathbf{W} between data points using a data dependent kernel $\mathbf{K}_D(\cdot, \cdot)$ depending on the technique used²

Given the k -th data set, we can construct an $N_k \times N_k$ symmetric weight matrix \mathbf{W}^k representing its local geometric structure as typically done in LLE, Isomap, etc. Therefore, we have K weight matrices where $\mathbf{W}_{ij}^k = \mathbf{K}_D(\mathbf{x}_i^k, \mathbf{x}_j^k)$.

On the other hand, the inter-manifold structure is the geometric structure between the different data sets in the space. Given any two manifolds p and q , their pairwise inter-manifold structure is represented by a $N_p \times N_q$ kernel matrix \mathbf{U}^{pq} , such that $\mathbf{U}_{ij}^{pq} = G(\mathbf{x}_i^p, \mathbf{x}_j^q)$, where $G(\cdot, \cdot)$ is a global kernel between point \mathbf{x}_i^p and \mathbf{x}_j^q on manifolds p and q respectively. This collection of matrices captures the pairwise manifold relations. Notice that the matrices \mathbf{U}^{pq} are not symmetric. However, there exist a hyper-symmetry structure since $\mathbf{U}^{pq} = \mathbf{U}^{qpT}$.

B. Objective Function

Given the intra-manifold and the inter-manifold structures, the goal is to achieve an embedding of all the points on all the manifolds in a way that preserves the local structure on each manifold and, in 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. Since we want to find a common embedding among all the data sets, we do not need to preserve the inter-manifold structure. The inter-manifold structure is needed to solve for soft correspondences between the data sets.

If the correspondences between the different data sets are given, the problem become trivial. In our case, the correspondences are not given. Moreover, since the data sets are of different sizes, no optimal one-to-one correspondences is possible. Therefore, given the inter-manifold geometric structure, we aim to obtain *inter-manifold soft correspondence structure* by solving for soft correspondences between the different dataset. A soft correspondence matrix between manifold p and manifold q is denoted by an $N_p \times N_q$ matrix \mathbf{C}^{pq} . Generally speaking, a high \mathbf{C}_{ij}^{pq} value indicates a strong

¹Throughout this paper, we will use superscripts to indicate manifold index (equivalently, data set index) and subscripts to indicate point index, i.e., \mathbf{x}_i^k denotes point i on manifold k .

²For example, Isomap [7] finds shortest geodesic paths on the manifold. LLE [2] finds local linear weights to construct each point from its local neighbors. Laplacian eigenmaps [3] uses a heat kernel given nearest manifold neighbors.

correspondence between point \mathbf{x}_i^p and point \mathbf{x}_j^q . In section III we will show how can such soft correspondence matrices can be obtained.

Formally, we seek an embedding for the data, i.e., we seek the coordinates of N points $Y = \{\mathbf{y}_i^k \in \mathbb{R}^d, i = 1, \dots, N_k, k = 1, \dots, K\}$ in a d -dimensional embedding space where a point \mathbf{y}_i^k is the embedding of the data point \mathbf{x}_i^k . Given the above stated goals, we reach the following objective function on the embedded points Y , which need to be minimized

$$\phi(Y) = \sum_k \sum_{i,j} (\mathbf{y}_i^k - \mathbf{y}_j^k)^2 \mathbf{W}_{ij}^k + \sum_{\substack{p,q \\ p \neq q}} \sum_{i,j} (\mathbf{y}_i^p - \mathbf{y}_j^q)^2 \mathbf{C}_{ij}^{pq}. \quad (1)$$

The first term of the objective function preserves the intra-manifolds' local geometry since it tries to keep the embedding $(\mathbf{y}_i^k, \mathbf{y}_j^k)$ of any two points $(\mathbf{x}_i^k, \mathbf{x}_j^k)$ on a given manifold close to each other based on their intra-manifold weight \mathbf{W}_{ij}^k . The second term of the objective function tries to bring close the embedded points $(\mathbf{y}_i^p, \mathbf{y}_j^q)$ on manifolds p and q if their soft correspondence weight \mathbf{C}_{ij}^{pq} is high. This objective function can be rewritten using one set of weights defined on the whole set of input points as:

$$\phi(Y) = \sum_{p,q} \sum_{i,j} (\mathbf{y}_i^p - \mathbf{y}_j^q)^2 \mathbf{A}_{ij}^{pq}, \quad (2)$$

where the weight matrix \mathbf{A} is defined as:

$$\mathbf{A}_{ij}^{pq} = \begin{cases} \mathbf{W}_{ij}^k & p = q = k \\ \mathbf{C}_{ij}^{pq} & p \neq q \end{cases} \quad (3)$$

This construction defines an $N \times N$ weight matrix \mathbf{A} with $K \times K$ blocks where the $p - q$ block is of size $N_p \times N_q$. The p -th diagonal block is the intra-manifold weight matrix \mathbf{W}^p for the p -th manifold. The off-diagonal $p - q$ block is the soft correspondence matrix \mathbf{C}^{pq} . The matrix \mathbf{A} is symmetric by definition since diagonal blocks are symmetric and since $\mathbf{C}^{pq} = \mathbf{C}^{qpT}$.

Given this construction, the objective function in Eq. 1 reduces to the problem of Laplacian embedding of a graph defined by the weight matrix \mathbf{A} . Therefore the objective function reduces to

$$\mathbf{Y}^* = \arg \min_{\mathbf{Y}^T \mathbf{D} \mathbf{Y} = 0} \text{tr}(\mathbf{Y}^T \mathbf{L} \mathbf{Y}), \quad (4)$$

where \mathbf{L} is the Laplacian of the matrix \mathbf{A} , i.e., $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where \mathbf{D} is the diagonal matrix defined as $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$. The $N \times d$ matrix \mathbf{Y} is the stacking of the desired embedding coordinates such that,

$$\mathbf{Y} = [y_1^1, \dots, y_{N_1}^1, y_1^2, \dots, y_{N_2}^2, \dots, y_1^K, \dots, y_{N_K}^K]^T.$$

Minimizing this objective function is a straight forward generalized eigenvector problem: $\mathbf{L} \mathbf{y} = \lambda \mathbf{D} \mathbf{y}$. The desired embedding is obtained by the bottom d nonzero eigenvectors.

III. INTER-MANIFOLD CORRESPONDENCES

Given two data sets X^p and X^q with their inter-manifold geometric structure weight matrix \mathbf{U}^{pq} , a maximum weight matching can be achieved by solving for a permutation matrix \mathbf{P} that permutes the rows of \mathbf{U}^{pq} in order to maximize its trace, i.e.,

$$\psi(\mathbf{P}) = \text{tr}(\mathbf{P}^T \mathbf{U}^{pq})$$

The permutation matrix constraint can be relaxed into an orthonormal matrix constraint on the matrix \mathbf{P} . Therefore, the goal is to find optimal an orthonormal matrix \mathbf{P}^* such that

$$\mathbf{P}^* = \arg \max_{s.t. \mathbf{P}^T \mathbf{P} = \mathbf{I}} \text{tr}(\mathbf{P}^T \mathbf{U}^{pq}) \quad (5)$$

It was shown in [8] that the optimal solution for 5 is $\mathbf{P}^* = \mathbf{U} \mathbf{E} \mathbf{V}^T$ where the SVD decomposition of $\mathbf{U}^{pq} = \mathbf{U} \mathbf{S} \mathbf{V}^T$ and \mathbf{E} is obtained by replacing the singular values on the diagonal of \mathbf{S} by ones.

In our case we have $K * (K - 1)/2$ inter-manifold weight matrices \mathbf{U}^{pq} and we need to obtain $K * (K - 1)/2$ correspondence matrices \mathbf{C}^{pq} that simultaneously maximizes

$$\psi(\mathbf{C}) = \sum_{p=1:K, q=p+1:K} \text{tr}(\mathbf{C}^{pq^T} \mathbf{U}^{pq}) \quad (6)$$

Solving for simultaneous hard correspondences is a weighted multipartite graph matching problem, which is a much harder combinatorial problem. In our case the objective function in 6 can be directly maximized by finding each of the pairwise correspondences \mathbf{C}^{pq} by solving 5.

We use a Gaussian kernel to encode the inter-manifold geometry, i.e., $\mathbf{U}_{ij}^{pq} = \exp(-\|x_i^p - x_j^q\|)/2\sigma^2$ where σ is a global scale that is estimated as a percentile of the overall data scale. One obvious question that arises: can the inter-manifold kernels \mathbf{U}^{pq} directly (or after scaling) be used in the objective function 1 instead of \mathbf{U}^{pq} since it also provides a measure of affinity between the points from the different data sets. The answer is no. The soft correspondences obtained by solving 5 incorporates the principle of exclusion because of the orthogonality constraints.

IV. EXPERIMENTAL RESULTS

We run experiments on different data sets. Here we show experiments on three of them.

Example I: We used 5 objects with similar geometry from the COIL-20 data set [9] with different views for each object taken along a view circle, Fig. 2-a. We used 72 views for each objects. Each instance is 128×128 grayscale image. Putting all data together, traditional manifold embedding approaches fail to discover the structure of the data as can be seen in Fig. 2-b,c. The proposed approach succeeds in learning a joint view manifold of all the five objects, Fig. 2-d. This clearly shows that all the objects are arranged in this representation according to the view point and invariant of the object.

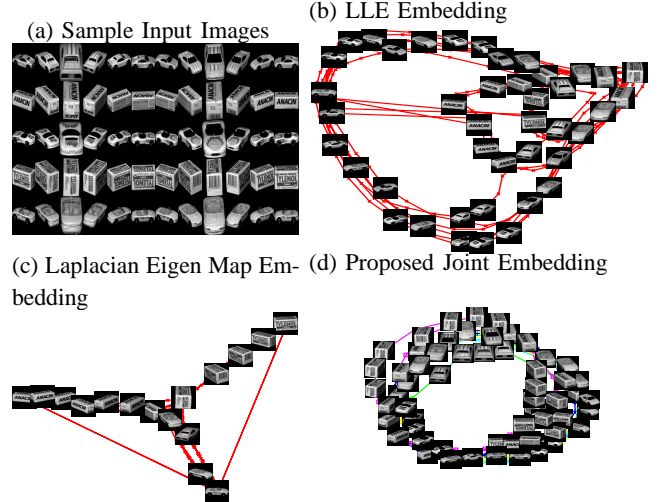


Figure 2. Coil Dataset Example: Learning a joint view manifold from multiple objects: (a) Input object appearance with view variations. (b,c) LLE and Laplacian Eigen map embedding for all the data. (d) The joint embedding of the view manifold obtained using the proposed algorithm.

Example II: The goal is to learn an embedding of the gait manifold from shape observations (silhouettes) from different people. The data are from CMU-Mobo gait data set [10]. For this experiment, we used 7 people data, six walking cycles each, side view. The number of frames in each data set vary from 189 to 234 (depending on the person’s walking speed) i.e., frames are not in correspondences. Putting all the data together, LLE and other embedding techniques fail to obtain a useful embedded representation for the propose of visualization or analysis. The proposed approach can successfully learn a joint embedding of the gait manifold across the different subjects, invariant to the different people shape variability as shown in Figure 3. The embedding shows a figure “8” embedding of the gait manifold.

Example III: The goal is to learn an embedding of a joint facial expression manifold across different subjects. We used four facial expression sets from CMU AMP facial expression database, each containing a person performing three expressions (smile, angry, surprise). Fig. 4-a shows examples of the input images for two of the subjects. The number of images in each set varies from 45 to 59 images. Each input instance is a 64×64 grayscale images.

Fig. 4-b shows the embedding obtained using Laplacian Eigenmaps with all data sets together. As can be noticed in this figure, the embedding is dominated by the inter-manifold structure and shows 4 separate clusters, one for each subject. LLE, regardless of the settings, also resulted in embeddings with 4 different clusters. Using the proposed approach, we can achieved a common facial expression manifold embedding, invariant of the subject, Fig. 4-c shows the embedding obtained. We can clearly see that the three different expressions are located at three different parts of

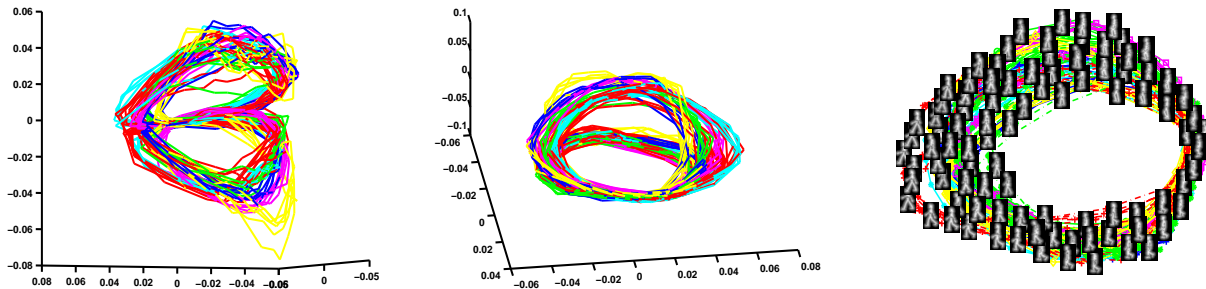


Figure 3. CMU Gait Example: Embedding the gait manifold from seven subjects. Different views of the obtained 3D embedding. We can see similar body postures across different people along the manifold.

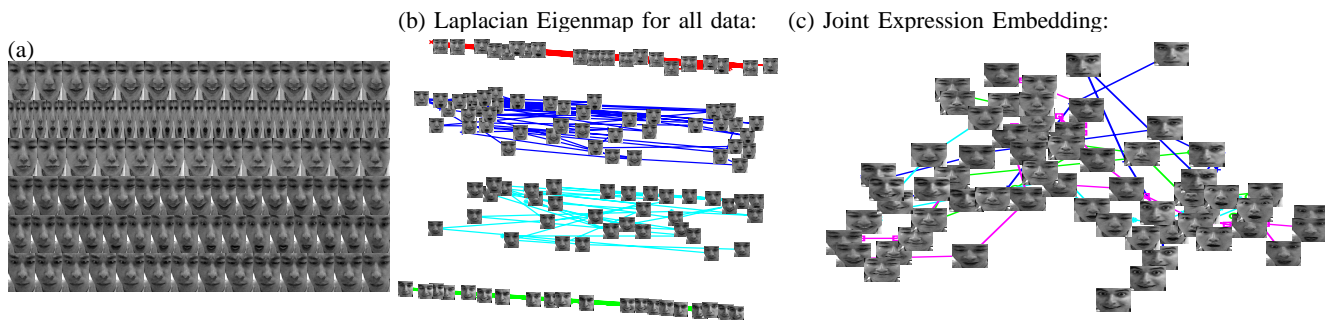


Figure 4. Embedding of a joint facial expression across different people: a) sample input data for two subjects. b) Embedding of all the data with laplacian eigenmaps. Notice the separation between the different people’s manifolds c) The joint expression embedding obtained by the proposed approach. Notice how expressions are embedded invariant of people.

the embedding space where the corresponding frames across the different subjects are close to each other.

V. CONCLUSION

We introduced an approach for learning joint embedded representations from multiple manifold. This is an important problem in data analysis where the goal is to learn an embedded representation of data from multiple sources (e.g., different people, objects, different spaces) regardless of the variability of the source. We successfully applied the approach to several real data sets including, visual gait data, facial expression data, and kinematic locomotion data. In all cases we can reach a meaningful embedding of the underlying manifold across different input data sets. The approach we introduce provide an important extension to the state-of-the-art spectral embedding techniques, such as LLE, Isomap, and Laplacian eigenmaps, to handle the case of multiple manifolds.

REFERENCES

- [1] J. Tenenbaum, “Mapping a manifold of perceptual observations.” in *Advances in Neural Information Processing*, vol. 10, 1998, pp. 682–688.
- [2] S. Roweis and L. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [3] M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [4] M. Brand and K. Huang, “A unifying theorem for spectral embedding and clustering,” in *Proc. of the Ninth International Workshop on AI and Statistics*, 2003.
- [5] N. Lawrence, “Gaussian process latent variable models for visualization of high dimensional data,” in *NIPS*, 2003. [Online]. Available: citeseer.ist.psu.edu/lawrence03gaussian.html
- [6] J. H. Ham, D. D. Lee, and L. Saul, “Learning high dimensional correspondences with low dimensional manifolds,” in *The ICML 2003 Workshop on the Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining*, 2003.
- [7] J. Tenenbaum, V. de Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, no. 5500, pp. 2319 – 2323, 2000.
- [8] G. L. Scott and H. C. Longuet-Higgins, “An algorithm for associating the features of two images,” *Biological Sciences*, vol. 244, no. 1309, 1991.
- [9] S. A. Nene, S. K. Nayar, and H. Murase, “Columbia object image library(coil-20),” Columbia University, Tech. Rep. Technical Report CUCS-005-96, 1996.
- [10] R. Gross and J. Shi, “The cmu motion of body (mobo) database,” Carnegie Mellon University, Tech. Rep. TR-01-18, 2001.