Manifold Alignment via Corresponding Projections

Deming Zhai¹² dmzhai@jdl.ac.cn
Bo Li¹² bli@jdl.ac.cn
Hong Chang²³ hchang@jdl.ac.cn
Shiguang Shan²³ sgshan@jdl.ac.cn
Xilin Chen²³ xlchen@jdl.ac.cn

Wen Gao¹⁴ wgao@jdl.ac.cn ¹ School of Computer Science and Technology, Harbin Institute of Technology, China

² Digital Media Research Center, Institute of Computing Technology, CAS, China

³ Key Laboratory of Intelligent Information Processing, Chinese Academy of Sciences, China

⁴ Institute of Digital Media, Peking University, China

In many real-world applications, the same object (e.g., pose, face) may have different observations (or descriptions) which are highly related but sometimes look different from each other, such as the videos of the same scene from different viewpoints, the image sequences of the same action for different objects, the video and audio segments that come from the same circumstances, and so on. How to find the correspondence between the data points of different observation sets is a hot topic. Due to the fact that different datasets might be located in different high-dimensional spaces and represented by different features, it is difficult to match the data in their original observation spaces. From the geometric perspective, each observation set (e.g., a sequence of face images under various poses) forms a manifold. Given that these observations are from the same object, it is reasonable to assume that some common features across different observation spaces can be represented in an underlying common manifold. The shared low-dimensional embeddings, as better descriptions of the intrinsic geometry and relationship between different manifolds, are expected to benefit the subsequent task of datasets alignment.

In this paper, we propose a novel manifold alignment method via corresponding projections under the semi-supervised learning setting. In our algorithm, manifold alignment is formulated as a minimization problem with proper constraints. Our optimization could be solved in an analytical manner with closed-form solution. The proposed method is more general in the following three senses: (1) It learns explicit corresponding mappings from different manifolds to the underlying common embeddings, thus overcome the limitation of most previous methods on out-of-sample extension [1] ability. (2) It defines the corresponding mappings for different datasets at the same time and doesn't have any assumption on the data structures, hence could deal with more complex cases than single affine transformation used in Procrustes analysis. (3) It could be easily extended to multi-manifolds alignment. In experiments, pose alignment on image sequences of different objects and face images, are used to verify the efficiency of our algorithm.

The target of our manifold alignment method is to learn the corresponding mappings which could project data points from different datasets to the intrinsic common embeddings. By doing this, we can compare the embeddings of the two datasets instead of their original high-dimensional representations.

More specifically, our algorithm learns mapping matrices P_x and P_y for data sets X and Y, respectively. In mapping matrices learning, two important issues should be taken into consideration: (1) The common embeddings should be consistent with the given labeled correspondence pairs; (2) The common embeddings should preserve the local geometric structures in all original input spaces. We achieve the overall objective by minimizing the following energy function:

$$J(\mathbf{P}_{r}, \mathbf{P}_{v}) = J(\mathbf{P}_{r}, \mathbf{P}_{v}, \mathbf{X}^{l}, \mathbf{Y}^{l}) + \alpha_{r}J(\mathbf{P}_{r}, \mathbf{X}) + \alpha_{v}J(\mathbf{P}_{v}, \mathbf{Y}). \tag{1}$$

There are three terms in this energy function defined above. The first term $J(\mathbf{P}_x, \mathbf{P}_y, \mathbf{X}^l, \mathbf{Y}^l)$ is called the *correspondence preserving term*. And the other two terms, $J(\mathbf{P}_x, \mathbf{X})$ and $J(\mathbf{P}_y, \mathbf{Y})$, are the *manifold regularization terms* which are used to preserve the intrinsic manifold structures of different datasets. The parameters α_x and α_y are used to balance the three terms and further influence the relative contribution of local structure preserving in each observation space.

The term of correspondence preserving cost is defined according to the given correspondence pairs as

$$J(\mathbf{P}_x, \mathbf{P}_y, \mathbf{X}^l, \mathbf{Y}^l) = \sum_{(i,j) \in C} ||\mathbf{P}_x^T \mathbf{x}_i - \mathbf{P}_y^T \mathbf{y}_j||^2,$$
(2)

where set C contains all index pairs (i, j) of given correspondences. Ideally, on the underlying common manifold, the embedded labeled pairs $(\mathbf{P}_x^T \mathbf{x}_i, \mathbf{P}_y^T \mathbf{y}_j)$ should be as close as possible. Therefore, we formulate the correspondence preserving term as a sum of squared differences (SSD) between the embeddings of all labeled correspondence data pairs.

Inspired by LLE [2] method, the two manifold regularization terms can be properly defined. Ultimately, we formulate the energy function as the following form:

$$J(\mathbf{P}_{x}, \mathbf{P}_{y})$$

$$= \sum_{(i,j)\in C} ||\mathbf{P}_{x}^{T}\mathbf{x}_{i} - \mathbf{P}_{y}^{T}\mathbf{y}_{j}||^{2} + \alpha_{x} \sum_{i=1}^{M} ||\mathbf{P}_{x}^{T}\mathbf{x}_{i} - \sum_{k \in N^{x}(i)} w_{ik}^{x} \mathbf{P}_{x}^{T}\mathbf{x}_{k}||^{2}$$

$$+ \alpha_{y} \sum_{j=1}^{N} ||\mathbf{P}_{y}^{T}\mathbf{y}_{j} - \sum_{k \in N^{y}(j)} w_{jk}^{y} \mathbf{P}_{y}^{T}\mathbf{y}_{k}||^{2}.$$
(3)

Here, w_{ik}^x is calculated through reconstructing the data point x_i using its k-nearest neighbors x_k ($k \in N^x(i)$) by minimizing $||\mathbf{x}_i - \sum_{k \in N^x(i)} w_{ik}^x \mathbf{x}_k||^2$ s.t.

 $\sum_{k \in N^{x}(i)} w_{ik}^{x} = 1, \text{ and } w_{jk}^{y} \text{ could be got in a similar manner.}$

Let us express the objective function Eq.(3) using a complete matrix form, the energy function is rewritten as

$$J(\mathbf{P}_{x}, \mathbf{P}_{y}) = J(\mathbf{P}_{x}, \mathbf{P}_{y}, \mathbf{X}^{l}, \mathbf{Y}^{l}) + \alpha_{x}J(\mathbf{P}_{x}, \mathbf{X}) + \alpha_{y}J(\mathbf{P}_{y}, \mathbf{Y})$$

$$= \operatorname{Tr}\left(\left[\begin{array}{cc} \mathbf{P}_{x}^{T}\mathbf{X} & \mathbf{P}_{y}^{T}\mathbf{Y} \end{array}\right] \left[\begin{array}{cc} \mathbf{L}^{x} + \alpha_{x}\mathbf{M}^{x} & -\mathbf{L}^{xy} \\ -(\mathbf{L}^{xy})^{T} & \mathbf{L}^{y} + \alpha_{y}\mathbf{M}^{y} \end{array}\right] \left[\begin{array}{cc} \mathbf{X}^{T}\mathbf{P}_{x} \\ \mathbf{Y}^{T}\mathbf{P}_{y} \end{array}\right]\right).$$

$$\operatorname{Let} \mathbf{P} = \left[\begin{array}{cc} \mathbf{P}_{x} \\ \mathbf{P}_{y} \end{array}\right], \mathbf{Z} = \left[\begin{array}{cc} \mathbf{X} \\ \mathbf{Y} \end{array}\right] \text{ and } \mathbf{A} = \left[\begin{array}{cc} \mathbf{L}^{x} + \alpha_{x}\mathbf{M}^{x} & -\mathbf{L}^{xy} \\ -(\mathbf{L}^{xy})^{T} & \mathbf{L}^{y} + \alpha_{y}\mathbf{M}^{y} \end{array}\right].$$

$$\operatorname{Consequently, we obtain a concise form as}$$

$$J(\mathbf{P}_x, \mathbf{P}_y) = \operatorname{Tr}\left(\mathbf{P}^T \mathbf{Z} \mathbf{A} \mathbf{Z}^T \mathbf{P}\right). \tag{5}$$

Finally, we add the constraints to achieve scaling and translation invariance, and solve the optimization problem of manifold alignment by minimizing

$$J(\mathbf{P}_x, \mathbf{P}_y)$$
 s.t. $\frac{1}{M+N} \mathbf{P}^T \mathbf{Z} \mathbf{Z}^T \mathbf{P} = \mathbf{I}$ and $\mathbf{P}^T \mathbf{Z} \mathbf{e} = 0$, (6)

where **I** is the identity matrix with size of $d \times d$, and $\mathbf{e} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$ is the vector of ones with M + N entries.

Let $\mathbf{E} = \mathbf{Z}\mathbf{A}\mathbf{Z}^T$ and $\mathbf{F} = \mathbf{Z}\mathbf{Z}^T$. The solution to the optimization problem with respect to \mathbf{P} is given by the second to (d+1)-th smallest generalized eigenvectors \mathbf{p} with $\mathbf{E}\mathbf{p} = \lambda\mathbf{F}\mathbf{p}$. Note that $\mathbf{p} = \begin{bmatrix} \mathbf{p}_x \\ \mathbf{p}_y \end{bmatrix}$ and d represents the dimensionality of the common embeddings.

- Y. Bengio, J. Paiement and P. Vincent. Out-of-Sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering. In Advances in Neural Information Processing Systems, pp 177-184, 2004.
- [2] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. Science, 290:2323-2326, 2000.