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, hence could deal with more complex cases than single affine transformations used in Procrustes analysis. (2) It could be easily extended to multi-manifolds alignment. In experiments, pose alignment on image sets (e.g., pose, face) may 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

\[ J(P_x, P_y) = \sum_{i \neq j \in C} ||P_x^T x_i - P_y^T y_j||^2 + \alpha_1 \sum_{i} ||P_x^T x_i - \sum_{k \neq i} w_{ik} P_y^T y_k||^2 \]

where \( C \) contains all index pairs \((i, j) \) of given correspondences. Ideally, on the underlying common manifold, the embedded labeled pairs \((P_x^T x_i, P_y^T y_j)\) 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:

\[ J(P_x, P_y) = \sum_{i \neq j \in C} ||P_x^T x_i - P_y^T y_j||^2 + \alpha_1 \sum_{i} ||P_x^T x_i - \sum_{k \neq i} w_{ik} P_y^T y_k||^2 \]

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

\[ J(P_x, P_y) = \sum_{i \neq j \in C} ||P_x^T x_i - P_y^T y_j||^2, \]

where set \( C \) contains all index pairs \((i, j) \) of given correspondences. Ideally, on the underlying common manifold, the embedded labeled pairs \((P_x^T x_i, P_y^T y_j)\) 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.

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:

\[ J(P_x, P_y) = J(P_x, P_y; X', Y') + \alpha_1 J(P_x, X) + \alpha_2 J(P_y, Y). \]

There are three terms in this energy function defined above. The first term \( J(P_x, P_y; X', Y') \) is called the correspondence preserving term. And the other two terms, \( J(P_x, X) \) and \( J(P_y, Y) \), are the manifold regularization terms which are used to preserve the intrinsic manifold structures of different datasets. The parameters \( \alpha_1 \) and \( \alpha_2 \) are used to balance the three terms and further influence the relative contribution of local structure preserving in each observation space.