SPATIAL-AWARENESS SPECTRAL EMBEDDING (SASE) FOR ROBUST SHAPE MATCHING

Huihui Xu, Jundong Liu

School of Electrical Engineering and Computer Science
Ohio University
Athens OH 45701, USA

ABSTRACT

Shape matching in the spectral domain has gained great popularity in recent years. Most algorithms, however, rely on invariant global spectral embeddings of the shapes to find correspondence, where spatial neighborhood information is not explicitly incorporated into the matching procedure. Misalignments of global as well as local structures are often resulted due to the lack of spatial guidance. In this paper, we identify a number of ambiguities existing in spectral embedding and matching, and subsequently propose a general framework to improve the matching coherence. At the center of the framework is a hybrid spatial-awareness spectral embedding (SASE), which allows various neighborhood and topological information, such as pair-wise distance, relative angles w.r.t. object centers, to be integrated into commute-time (CT) embeddings. A probabilistic expectation maximization (EM) algorithm with imposed regularity is employed to seek an optimal matching of the SASE embeddings. Experimental evaluations of the algorithm on 2D and 3D data demonstrate both the effectiveness and robustness of our approach.

Index Terms— point matching, spectral graph

1. INTRODUCTION

The analysis of 2D/3D shapes is a fundamental problem in many image understanding applications such as object recognition, tracking and brain mapping. Matching the input shapes, usually defined as the boundaries of the areas of interest, is often a prerequisite step for other analysis tasks to be carried out. Spatial domain approaches [1, 2, 3, 4] directly work in the 3D Euclidean space, and usually certain regularization constraint is imposed to ensure the smoothness and regularity of the estimated correspondence and deformation field. Spectral based methods [5, 6, 7, 8, 9], on the other hand, first map the input shapes or points into a low dimensional space and then seek the alignment in the spectral embedding space [10] through graph matching.

Spectral matching is pioneered by analytic methods [11, 12] that compute point correspondence through eigen-decomposition of the graph's adjacency matrices. Recently a variety of probabilistic approaches [13, 6, 8, 9] have been devised to enhance the robustness in handling graph variabilities and outliers. Mateus *et al.* [6] demonstrate the capability of Laplacian matrices in capturing the properties of locally-rigid structures, and subsequently propose an articulated shape matching algorithm that combines Laplacian embedding with probabilistic point matching. Escolano *et al.* [8] take a similar point-based approach, where each node of the graph is mapped to a point on a low dimensional manifold through the commute-time (CT) embeddings [14]. Sharma *et al.* [9] extend the solution in [6] through a normalized commute-time embedding. The eigenvalue-eigenvector ordering and sign issues are tackled using the *eigensignature* (histogram) which is invariant to the isometric shape deformations.

While being able to achieve greater matching flexibilities than spatial domain algorithms, spectral algorithms have a set of limitations. Without a Euclidean coordinate system as the reference, the information maintained in spectral embeddings is mainly composed of topological connectivity relationships rather than physical distances/orientations among neighbors. Since graph construction from scattered points is inherently a discrete procedure, flip and reflective ambiguities can be easily resulted — a rather small perturbation to the spatial locations of the vertex points may dramatically alter the underlying structure of the global and local graph components.

Our proposed framework: in this paper, we identify a number of ambiguities existing in spectral graph embeddings and matchings, and as a remedy to improve the coherence, a general framework is proposed. At the center of this framework is a hybrid spatial-awareness spectral embedding (SASE), which allows various geometric information to be integrated into spectral embeddings. To match SASEs, we adopt a probabilistic expectation maximization (EM) formulation with imposed regularity in the spectral domain. The combined framework proves to be effective in improving the matching accuracy and robustness, as well as removing spectral ambiguities.

2. METHODS

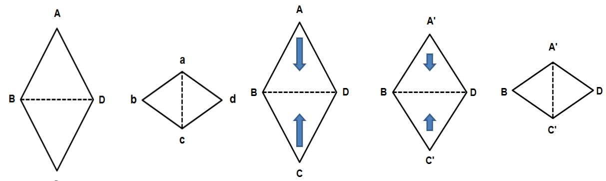
Let A and D be the adjacency and degree matrices of a graph $G = (V, E)$. Based on A and D, various graph Laplacian matrices can be generated. A common example is the combinatorial Laplacian matrix $L = D - A$. The set of all eigenvalues of L is called its spectrum, and they can be calculated through the generalized eigenvalue problem $Lv = \lambda Dv$, where v is an eigenvector. For each vertex X_i in the graph, the vector $(v_2(i), v_3(i), ...v_{K+1}(i))$ defines a K-dimensional spectral embedding.

Spectral embeddings contain a fair amount of topological and geometric information, therefore can be used to match point sets. For two given point sets in 3D Euclidean space, certain graph construction algorithm, e.g. Delaunay triangulation, needs to be applied to convert them to graphs, and generate the weight matrices accordingly. The graphs can then be matched through the optimization of certain objective function defined over their spectral embeddings. The estimated correspondence, when mapped back to the original Euclidean space, provides an alignment for the input point sets.

2.1. Instability and ambiguities in spectral embeddings and matching

Comparing with spatial-based methods, spectral methods have the advantage of better capturing the essential shape information, which leads to more power in matching and recognizing 'similar' shapes. Despite the versatility and flexibility, spectral methods possess a set of inherent limitations as well.

Flipping-induced matching instability and inconsistency After spatial points are mapped into spectral domain, the Euclidean coordinate system ceases to exist as a reference. As a result, the concepts of "above", "below", "left" and "right", which describes the relative position and orientation relationships among spatial points, are no longer applicable within spectral embeddings. Furthermore, matching of the embeddings captures topological similarity instead of spatial proximity, which reduces spatial orientation and scale to an unimportant role in this procedure. As a consequence of these two factors, alignment instability and inconsistency are prone to occur in spectral domain, due to the lack of spatial guidance. An example is given in Figure 1(a). In term of shape similarity, the flipped version of the left diamond shape is closer to the right one, therefore the estimated correspondence from spectral matching would be $ABCD \leftrightarrow dabc$.



(a) (b)

Fig. 1. Ambiguities existing in spectral embedding and matching.

For dynamically changing shapes, e.g., in human tracking, the flipping could lead to undesirable matching instability. Take Fig. 1(b) as an example, where the vertex A is moving down, and C moving up. The motion in the Euclidean coordinate system is continuous, while the structural changes are discrete under the graph embedding domain. Flipping in matching results occurs in parallel with structural changes. At the beginning, the matching is $ABCD \leftrightarrow A'BC'D$. At certain point, however, the alignment will be flipped to $ABCD \leftrightarrow DA'BC'$ due to the fact that the flipped diamond is more "similar" to the original shape. The tipping point is dependent on the graph neighborhood structure as well as graph weight setup, and Delaunay triangulation plays a deciding role if full graph is used.

Symmetric/Reflective ambiguity For a symmetric shape, reflecting it along the symmetry axis would yield an identical object. Without a spatial reference, matching ambiguity could be easily resulted, which is illustrated in Fig. 1(b). For the moving diamond shape, B could be mapped to match either B or D, and vice versa. In many computer vision applications, for example optical flow estimation, such a reflection as well as the estimated correspondence will generate illegitimate results.

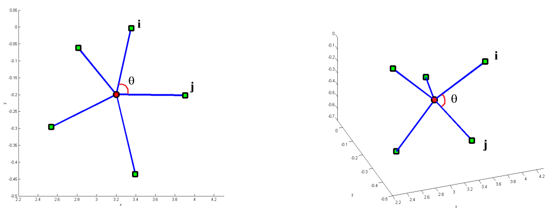
3. SPATIAL-AWARENESS SPECTRAL EMBEDDING (SASE) AND MATCHING

To overcome the aforementioned drawbacks in many spectral graph matching algorithms, we propose a *Spatial-Awareness Spectral Embedding (SASE) and Matching* framework in this paper to handle various embedding and matching ambiguities. The basic idea of Spatial-Awareness is to integrate as much relevant information as possible, such as discriminative local features or powerful geometric relationships beyond the simple weights or pair-wise distances on edges, into the spectral embedding and matching procedure, through which the ambiguities will be greatly reduced or even totally eliminated.

To test the concept, we choose *neighbors angles w.r.t. the point-set center* as the basis for the additional spatial information. Fig. 2(a) shows a group of two dimensional points (in green color). The red dot is the estimated center, through averaging the Euclidean coordinates of the point set. Integration of the additional spatial information can be carried out at different stages of the embedding and matching procedure. In this paper, we tackle the problem at the Laplacian matrix setup step. For each connected neighboring vertex pair i and j, other than the edge distance, we also calculate θ_{ij}, their angle w.r.t the center. The corresponding Laplacian weight W_{ij} is formulated as

$$W_{ij} = e^{\frac{\|x_i - xj\|^2}{\sigma_s^2} + \frac{\|\theta_{ij}\|^2}{\sigma_a^2}} \quad (1)$$

where σ_s and σ_a are the scale parameters (standard deviation) for edge distances and angles, respectively. If i and j are not connected, $W_{ij} = 0$. The same procedure can be conducted for 3D point sets, as shown in Fig. 2(b).



(a) (b)

Fig. 2. Neighbors' angle w.r.t. the point-set center as the basis for additional information in SASE.

Since all θ_{ij} are computed based on the same point-set center, integrating angle into the graph weights puts all neighboring structures into a unified framework. Flipping case in Fig. 1 can be avoided due to the imposed global control. On the other hand, as the angles represent the relative relations among neighbors, the resulted spectral embedding still allows shapes to be matched in a rotation and scaling invariant manner.

With the new "weights", we define our SASE based on a popular spectral embedding scheme: Commute Time (CT). Let $G = (P_X, E_X)$ be the undirected graphs generated from point-sets P_X, where E_X is specified by certain graph construction scheme and the number of nodes $N = |V_X|$. The weight matrix W_P is defined as in Eqn. 1, and corresponding degree matrix is D. Let $L_X = D_X W_X = \Phi_X \Lambda_X \Phi_T$ be the eigendecompositions of the Laplacian matrix of the graph, where $_X$ is the diagonal eigevalue matrices, and Φ_X the corresponding eigenvector matrix. The volumes of the graph is defined as $vol_X = trace(D_X)$. The SASE embedding is defined as:

$$\frac{\hat{\Theta}_X^{(i)}}{\sqrt{vol_X}} = (\frac{1}{\sqrt{\lambda_X^{(2)}}}\phi_X^{(2)}(i) ... \frac{1}{\sqrt{\lambda_X^{(d+1)}}}\phi_X^{(d+1)}(i))^T \qquad (2)$$

This justification of using CT lies in its strong robustness against modifications of graph structures, as well as its ability to maintain the Euclidean distance between nodes within the embeddings.

4. MATCHING SASES

Matching two SASEs can be reduced to finding a non-rigid alignment for two point sets in high-dimensional Euclidean space, where the dimension d is the number of eigenvectors chosen for the embedding. In all the experiments conducted in this paper, we chose $d = 5$ as the first six eigenvectors usually contain enough shape information for matching. To solve this problem, we adopted the probabilistic Expectation-maximization (EM) algorithm formulated in the Coherent Point Drift (CPD) algorithm [3] as well as in the Commute-time matching work by Escolano *et al.* [8]. CPD is appropriate in this regard because it is known of handling non-rigid alignment for an arbitrary number of dimensions.

Let $i \in V_X$ and $j \in V_Y$ be nodes of graphs X and Y and let T be a non-rigid transformation with aligns the SASE embeddings $\hat{\Theta}_X$ and $\hat{\Theta}_Y$. Then we can define $\widetilde{SASE}(i,u)$ = $\min_T ||\hat{\Theta}_X$ - $\hat{\Theta}_Y||^2$ = $\min_{T'}$ $\widetilde{SASE}(i,u)$ where

$$\widetilde{SASE}(i,u) = \sqrt{vol_{XY}} \sum_{z=2}^{d+1} \frac{1}{\sqrt{\lambda_{XY}^{(z)}}} (\phi_X^{(z)}(i) - T'(\phi_Y^{(z)}(u)))^2$$

where $vol_{XY} = vol_X \; vol_Y$, $\lambda_{XY}^{(z)} = \lambda_X^{(z)} \lambda_Y^{(z)}$, and T' aligns non-rigidly the SASE embedding ϕ_Y with those of ϕ_X. Let $\Phi_Y^{(u)}$ be the centers of d-dimensional Gaussian Mixtures (GMM). Aligning SASE is formulated as the minimization of

$$E(W, \sigma^2) = \frac{1}{2\sigma^2} \sum_{i=1,u=1}^{N,M} P_{iu} \widetilde{SASE}(i,u) + \frac{N^P}{2} log\sigma^2 + \gamma(W)$$

Given the $M \times N$ probability matrix P with entries P_{ui}, G a $M \times M$ Gaussian kernel matrix where $G(u,v) = exp - \frac{1}{2}\sigma^2 SASE(u,v)$, and matrices Θ_X^T, Θ_Y^T with dimensions $N \times d$ and $M \times d$ respectively, the parameters of W can be estimated by solving:

$$(G + \lambda\sigma^2 diag(P1)^{-1})W = diag(P1)^{-1} P \hat{\Theta}_X^T - \hat{\Theta}_Y^T \quad (3)$$

The following table summarizes the overall procedure of the SASE matching algorithm.

Algorithm 1 Pseudo-code of our SASE embedding and matching algorithm.

Input: A pair of 2D/3D point sets X and Y.

- Triangulate point-sets to meshes and construct a weighted graph.
  - Calculate pair-wise angles θ_{ij} w.r.t. the point-set center.
- Embedding:
  - Compute Laplacian matrices based on Eqn. (1).
  - Commute-time embeddings $\hat{\Theta}_X^{(t)}$ and $\hat{\Theta}_Y^{(u)}$ based on the Laplacian matrices.
- Match SASE embeddings based on probalistics EM:
  - Initialzation: $W = 0, \lambda > 0$,
  - $\sigma^2 = \frac{1}{DNM} \sum_{M,N}^{t,u=1} \left\| \hat{\Theta}_X^{(t)} - \hat{\Theta}_Y^{(u)} \right\|^2$
  - Construct G: $g_{ij} = exp^{-\frac{1}{2\beta^2}\left\| \hat{\Theta}_Y^{(i)} - \hat{\Theta}_Y^{(j)} \right\|^2}$
  - EM optimization, loop:
    - E-step:

$$p_{ui} = \frac{exp^{-\frac{1}{2\sigma^2}\left\| \hat{\Theta}_X^i - (\hat{\Theta}_Y^u + G(u,\cdot)W) \right\|^2}}{\sum_M^v exp^{-\frac{1}{2\sigma^2}\left\| \hat{\Theta}_X^i - (\hat{\Theta}_Y^v + G(v,\cdot)W) \right\|^2} + \frac{w}{1-w}\frac{(2\pi\sigma^2)^{D/2} M}{N}}$$

    - M-step: Solve Eqn. (3). The aligned embedding set is $T = (Y, W) = Y + GW$
  - The probability of correspondence is given by P

5. EXPERIMENTAL RESULTS

To demonstrate the effectiveness brought by our SASE embedding in reducing spectral ambiguities, we perform experiments on both 2D and 3D shapes. Comparisons are made with state-of-the-art spatial and spectral matching methods. The Coherent Point Drift (CPD) algorithm proposed by Myronenko *et al.* [3] has been used as the spatial category representative, due to its great popularity in recent years. The comparison with spectral solutions is with the matching algorithm proposed by Escolano *et al.* [8] in CVPR'2011.

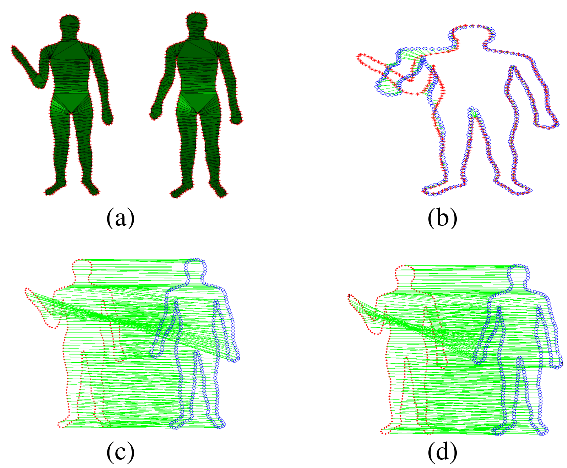


Fig. 3. Experiment on 2D human figures. (a) input point sets and the constructed triangulations. (b) matching result from CPD. (c) result from Escolano's method. (d) result from SASE matching.

The first experiment is conducted on a pair of 2D human figures with manually extracted boundary points. Delaunay triangulation has been used to define point neighborhoods and construct the connection graphs. The resulted triangulation is shown in Fig. 3(a). The two point sets mainly differ in the right arm area. Other parts of the bodies are relatively well corresponded, which implies that the deformation field would be vastly nonlinear around the elbow area, if a perfect matching is achieved. The matching result using the CPD is shown in Fig. 3(b). As a typical spatial domain matching algorithm, CPD assumes the estimated deformation field should be smooth with regularity. Not surprisingly, it fails to account for the huge position disparities around right arm. Matching result based on Escolano's method, as shown in Fig. 3(c), has both the flipping and reflective ambiguity issue. Mirror reflective mismatches occur below the shoulder — right hand in other figure is mapped to left hand in the other. Matching result from our SASE-based method is given in Fig. 3(d). As evidence, our method, with the additional spatial angle information integrated in SASE embeddings, achieved rather accurate matchings across the board.

The second experiment is carried out using a pair of 3D

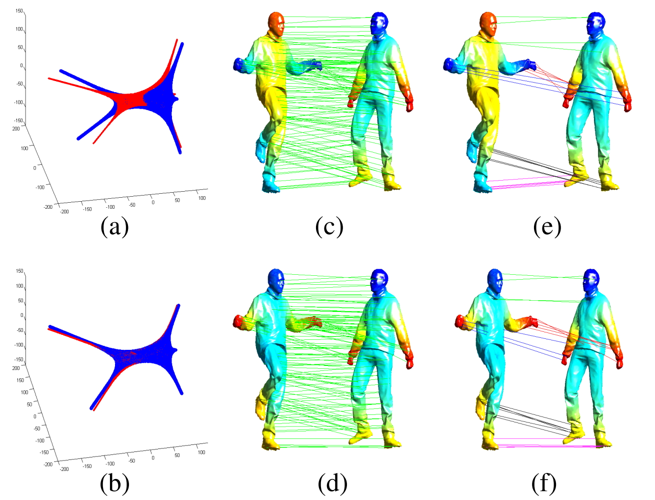


Fig. 4. Example of 3D Shape. (a) and (b) are the CT embedding and SASE embedding; (c) and (d) are the results of Escolano's method and SASE; (e) and (f) are zoom-in of (c) and (d).

human surface point sets obtained from the Multi-view Silhouettes [15] project. The graphs are constructed based on the surface meshes available in the data sets. The embedding and matching results are shown in Fig. 4. Like in the 2D experiment, matching using Escolano's method (Fig. 4(c)) still generates mirror reflective mismatches, where left hand/foot are incorrectly mapped to right hand/foot. Fig. 4.(e) provides a "zoom-in" version to demonstrate the details. Matching results and zoom-in from our SASE-based method is given in Fig. 4(d) and (f), where the flipping/reflective ambiguities have been removed and perfect matching has been achieved. To explore the underlying reason for matching ambiguities as well the remedy, we display the "CT alone" (as in Escolano's method) and our SASE embeddings in Fig. 4(a) and (b) respectively. Intuitively speaking, the ambiguities existing in CT alone come from the twisting structures below the chests, which makes left-to-left and left-to-right matchings equally plausible. The SASE embeddings, on the other hand, eliminate the twisting with the help from the spatial information, and the matching becomes unique and straightforward under the spectral embedding domain.

6. CONCLUSION

In this paper, we identify a set of inherent spectral embedding ambiguities, and propose a general framework to improve the matching coherence. Various spatial features can be added into the graph construction, Laplacian embeddings, or matching steps. Pair-wise angles w.r.t. point-set center have shown the potential to greatly reduce ambiguities and achieve robustness, and a direction of future investigation would be the exploration of other choices of geometric integration.

7. REFERENCES

[1] Paul J. Besl and Neil D. McKay, "A method for registration of 3-d shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, pp. 239–256, February 1992.

[2] Haili Chui and Anand Rangarajan, "A new point matching algorithm for non-rigid registration," *Computer Vision and Image Understanding*, vol. 89, no. 2-3, pp. 114–141, 2003.

[3] A. Myronenko and Xubo Song, "Point set registration: Coherent point drift," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 12, pp. 2262 –2275, dec. 2010.

[4] Bing Jian and Baba C. Vemuri, "Robust point set registration using gaussian mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1633–1645, 2011.

[5] Larry S Shapiro and J Michael Brady, "Feature-based correspondence: an eigenvector approach," *Image and Vision Computing*, vol. 10, no. 5, pp. 283 – 288, 1992, ¡ce:title¿BMVC 1991¡/ce:title¿.

[6] D. Mateus, R. Horaud, D. Knossow, F. Cuzzolin, and E. Boyer, "Articulated shape matching using laplacian eigenfunctions and unsupervised point registration," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[7] Rongjie Lai, Yonggang Shi, K. Scheibel, S. Fears, R. Woods, A.W. Toga, and T.F. Chan, "Metric-induced optimal embedding for intrinsic 3d shape analysis," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, june 2010, pp. 2871 –2878.

[8] F. Escolano, E. Hancock, and M. Lozano, "Graph matching through entropic manifold alignment," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, june 2011, pp. 2417 –2424.

[9] A. Sharma, R. Horaud, D. Mateus, et al., "3d shape registration using spectral graph embedding and probabilistic matching," 2011.

[10] R. Lai, *Computational Differential Geometry and Intrinsic Surface Processing*, Ph.D. thesis, University of California Los Angeles, 2010.

[11] S. Umeyama, "An eigendecomposition approach to weighted graph matching problems," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 10, no. 5, pp. 695–703, 1988.

[12] G.L. Scott and H.C. Longuet-Higgins, "An algorithm for associating the features of two images," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 244, no. 1309, pp. 21–26, 1991.

[13] B. Luo and ER Hancock, "A unified framework for alignment and correspondence," *Computer Vision and Image Understanding*, vol. 92, no. 1, pp. 26–55, 2003.

[14] H. Qiu and E.R. Hancock, "Clustering and embedding using commute times," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 11, pp. 1873–1890, 2007.

[15] Daniel Vlasic, Ilya Baran, Wojciech Matusik, and Jovan Popovi, "Articulated mesh animation from multi-view silhouettes," 2008.