

Inexact Matching of Large and Sparse Graphs Using Laplacian Eigenvectors

David Knossow, Avinash Sharma, Diana Mateus, and Radu Horaud

Perception team, INRIA Grenoble Rhone-Alpes
655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier Cedex, France
firstname.lastname@inrialpes.fr

Abstract. In this paper we propose an inexact spectral matching algorithm that embeds large graphs on a low-dimensional isometric space spanned by a set of eigenvectors of the graph Laplacian. Given two sets of eigenvectors that correspond to the smallest non-null eigenvalues of the Laplacian matrices of two graphs, we project each graph onto its eigenvectors. We estimate the histograms of these one-dimensional graph projections (eigenvector histograms) and we show that these histograms are well suited for selecting a subset of significant eigenvectors, for ordering them, for solving the sign-ambiguity of eigenvector computation, and for aligning two embeddings. This results in an inexact graph matching solution that can be improved using a rigid point registration algorithm. We apply the proposed methodology to match surfaces represented by meshes.

1 Introduction

Many problems in computer vision, shape recognition, document and text analysis can be formulated as a graph matching problem. The nodes of a graph correspond to local features or, more generally, to objects and the edges of the graph correspond to relationships between these objects. Solution to graph matching consists of finding an isomorphism (exact matching) or an optimal sub-graph isomorphism (inexact matching) between the two graphs. Spectral graph matching methods are attractive because they provide a framework that allows to embed graphs into isometric spaces and hence to replace the initial NP-hard isomorphism problem with a more tractable point registration problem.

An undirected weighted graph with N nodes can be represented by an $N \times N$ real symmetric matrix, or the adjacency matrix of the graph. Provided that this matrix has N distinct eigenvalues, the graph can be embedded in the orthonormal basis formed by the corresponding eigenvectors. Hence, an N -node graph becomes a set of N points in an N -dimensional isometric space. In [1] it is proved that the eigendecomposition of the adjacency matrices provide an optimal solution for exact graph matching, i.e., matching graphs with the same number of nodes. The affinity matrix of a shape described by a set of points can be used as the adjacency matrix of a fully connected weighted graph [2–5]. Although these methods can only match shapes with the same number of points,

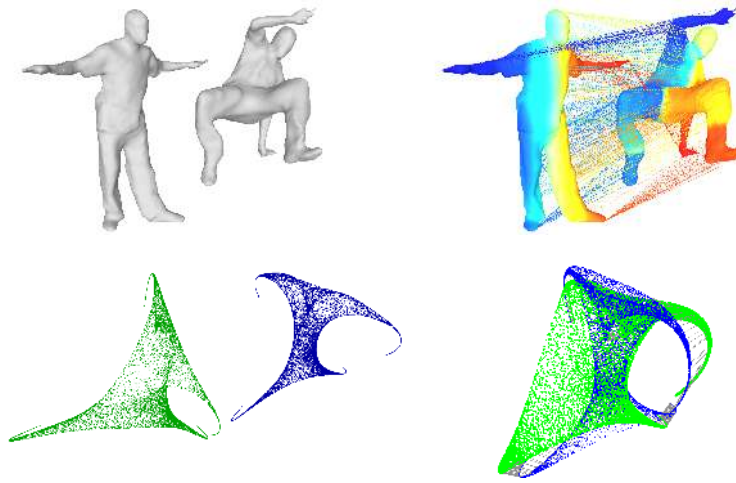


Fig. 1. Two graphs (meshes) of a hip-hop dancer (top-left) with respectively 31,600 and 34,916 nodes (vertices). The matching (top-right), subsampled for visualization purposes, was obtained by computing the embeddings of the two graphs into a 5-dimensional space (bottom-left) and by registering the two embeddings (bottom-right).

they introduce the heat kernel to describe the weights between points (nodes), which has a good theoretical justification [6].

Unfortunately, exact graph matching is not very practical, in particular when the two graphs have a different number of nodes, e.g., constructed from real data, such as visual sensors. Therefore one needs to combine spectral analysis with dimensionality reduction such that two graphs that are being matched are embedded in a common subspace with lower dimension than to the original graphs. This immediately calls for methods that allow many-to-many point correspondences. A clever idea is to combine matching with *agglomerative hierarchical clustering*, as done in [7]. We also note that spectral matching has strong links with spectral clustering [8] which uses the Laplacian matrix of a graph [9].

The analysis of the spectral methods cited above rely on the eigenvalues of the adjacency or Laplacian matrices. The strict ordering of the eigenvalues allows the alignment of the two eigenbases, while the existence of an eigengap allows the selection of a low-dimensional eigen space. In the case of inexact matching of large and sparse graphs, a number of issues remain open for the following reasons. The eigenvalues cannot be reliably ordered and one needs to use heuristics such as the ones proposed in [3, 10]. The graph matrices may have eigenvalues with geometric multiplicities and hence the corresponding eigenspaces are not uniquely defined. Dimensionality reduction relies on the existence of an eigengap. In the case of large graphs, e.g., ten thousands nodes, the eigengap analysis yields an eigen space whose dimension is not well suited for the task at hand. The sign ambiguity of eigenvectors is generally handled using simple heuristics [1, 7]. The link between spectral matching and spectral clustering has not yet been

thoroughly investigated. Existing spectral matching algorithms put the eigenvectors on an equal footing; the particular role played in clustering by the the Fiedler vector [9] has not been studied in the context of matching. We remark that the selection of strongly localized eigenvectors, which we define as the vector (eigenfunction) that spans over a small subset of the graph while being zero elsewhere, hence corresponding to *subgraph clusters*, has not been studied in depth. The only existing strategy for selecting such eigenvectors is based on eigenvalue ordering and the detection of an eigengap.

In this paper we propose an inexact spectral matching algorithm that embeds large graphs on an isometric space spanned by a *subset* of eigenvectors corresponding to the smallest eigenvalues of the Laplacian matrix. We claim that the tasks of (i) *selecting* a subset of eigenvectors, (ii) *ordering* them, (iii) finding a solution to the *sign-ambiguity* problem, as well as (iv) *aligning* two embeddings, can be carried out by computing and comparing the histograms of the projections of the graphs' nodes onto the eigenvectors. We postulate that the statistics of these histograms convey important geometric properties of the Laplacian eigenvectors [11]. In practice, we apply the proposed algorithm to match graphs corresponding to discrete surface representations of articulated shapes, i.e., *mesh registration*. Figure 1 shows a graph matching result obtained with our algorithm.

2 Laplacian embedding

We consider undirected weighted graphs. A graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ has a node set $\mathcal{V} = \{V_1, \dots, V_N\}$ and an edge set $\mathcal{E} = \{E_{ij}\}$. We use a Gaussian kernel to define the $N \times N$ weight and degree matrices:

$$W_{ij} = \exp(-d_{ij}^2)/\sigma^2 \quad (1)$$

$$D_{ii} = \sum_{j=1}^N W_{ij} \quad (2)$$

where d_{ij} is the geodesic distance between two nodes and σ is a smoothing parameter. In the case of meshes, a vertex is connected to its neighbors onto the surface. In practice we take the Euclidean distance between two connected vertices and $D_{ii} \approx 6$ which yields a very sparse graph. The Laplacian of a graph is defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$. We consider the *normalized graph Laplacian*: $\mathcal{L} = \mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-1/2}$. This is a semi-definite positive symmetric matrix with eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{N-1}$. The null space of this matrix is the constant eigenvector $\mathbf{U}_0 = (1 \dots 1)^\top$.

The eigenvectors of \mathcal{L} form an orthonormal basis, $\mathbf{U}_0^\top \mathbf{U}_i = 0, \forall i \in \{1, \dots, N-1\}$. Therefore we obtain the following property: $\sum_k U_{ik} = 0$. It is worthwhile to notice that $\mathcal{L} = \mathbf{I} - \mathcal{W}$ where $\mathcal{W} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ is the normalized adjacency matrix; matrices \mathcal{L} and \mathcal{W} share the same eigenvectors. Finally let $\mathcal{L}_t = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ be the *truncated* eigendecomposition where the null eigenvalue and constant eigenvector were omitted. Graph projections onto the eigenvectors corresponding to the smallest non-null eigenvalues of the Laplacian are displayed on the left side of figure 2. The right side of this

figure shows the density of these projections being considered as continuous equivalent of histograms. These densities were estimated using the MCLUST software¹ [12].

3 Matching using Laplacian eigenvectors

We consider two graphs \mathcal{G}_x and \mathcal{G}_y , with respectively N and M nodes. Exact graph matching, i.e., the special case $N = M$, can be written as the problem of minimizing the Frobenius norm:

$$\mathbf{P}^* = \arg \min_{\mathbf{P}} \|\mathcal{W}_x - \mathbf{P}\mathcal{W}_y\mathbf{P}^\top\|_F^2 \quad (3)$$

over the set of $N \times N$ permutation matrices \mathbf{P} . To solve this problem one can use the Laplacian embeddings previously introduced. Notice first that $\mathcal{W}_x - \mathbf{P}\mathcal{W}_y\mathbf{P}^\top = \mathcal{L}_x - \mathbf{P}\mathcal{L}_y\mathbf{P}^\top$. Let $\mathcal{L}_x = \mathbf{U}_x\mathbf{\Lambda}_x\mathbf{U}_x^\top$ and $\mathcal{L}_y = \mathbf{U}_y\mathbf{\Lambda}_y\mathbf{U}_y^\top$ be the truncated eigendecompositions of the two Laplacian matrices. The columns of $\mathbf{U} = [\mathbf{U}_1, \dots, \mathbf{U}_{N-1}]$ correspond to the eigenvectors which form an orthonormal basis, and $\mathbf{\Lambda} = \text{Diag}[\lambda_1, \dots, \lambda_{N-1}]$.

The spectral graph matching theorem [1] states that *if the eigenvalues of \mathcal{L}_x and \mathcal{L}_y are distinct and if they can be ordered, then the minimum of (3) is reached by:*

$$\mathbf{Q}^* = \mathbf{U}_x\mathbf{S}\mathbf{U}_y^\top, \quad (4)$$

where $\mathbf{S} = \text{Diag}[s_1, \dots, s_{N-1}]$, $s \in \{+1; -1\}$ accounts for the sign ambiguity in the eigendecomposition and where the domain of the objective function (3) has been extended to the group of orthogonal matrices. The entries of \mathbf{Q}^* are $Q_{ij}^* = \mathbf{x}_i^\top(\mathbf{s} \bullet \mathbf{y}_j)$, where $\mathbf{a} \bullet \mathbf{b}$ is the Hadamard product between two vectors. Since both \mathbf{U}_x and \mathbf{U}_y are orthonormal matrices, all entries Q_{ij}^* of \mathbf{Q}^* vary between -1 and 1 . Therefore, \mathbf{Q}^* can be interpreted as a *cosine node-similarity matrix*. Several heuristics were proposed in the past to solve for the sign ambiguity and hence to recover node-to-node assignments uniquely. In [1] the entries of \mathbf{U}_x and \mathbf{U}_y are replaced by their absolute values. The recovery of \mathbf{P}^* from \mathbf{Q}^* , i.e., exact matching, becomes a bipartite maximum weighted matching problem that is solved with the Hungarian algorithm.

In this paper we propose to perform the matching in a *reduced space* and let $K < (N, M)$ be the dimension of this space. We start with a solution provided by eigenvalue ordering followed by dimensionality reduction. This provides two sets of K ordered eigenvectors. However, ordering based on eigenvalues is not reliable simply because there may be geometric multiplicities that give rise, in practice, to numerical instabilities. To overcome this problem we seek a new eigenvector permutation which we denote by a $K \times K$ matrix $\bar{\mathbf{P}}$. Thus, equation (4) can be rewritten as:

$$\mathbf{Q} = \bar{\mathbf{U}}_x \bar{\mathbf{S}} \bar{\mathbf{P}} \bar{\mathbf{U}}_y^\top, \quad (5)$$

where $\bar{\mathbf{U}}_x$ and $\bar{\mathbf{U}}_y$ are $(N-1) \times K$ block-matrices. These matrices were obtained from \mathbf{U}_x and \mathbf{U}_y by retaining the first K columns and by re-normalizing each row such that

¹ <http://www.stat.washington.edu/mclust/>

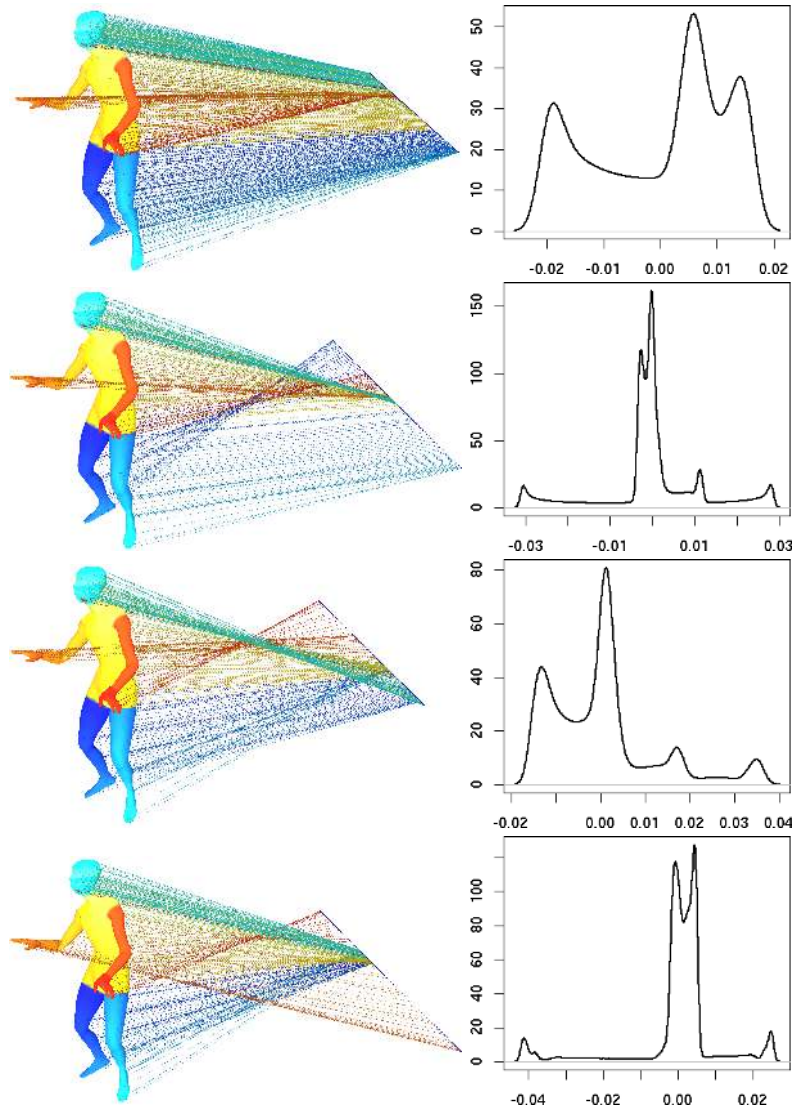


Fig. 2. A mesh with 7,063 vertices and with approximately six edges per vertex is shown here projected onto four Laplacian eigenvectors corresponding to the four smallest non-null eigenvalues. The curves onto the right correspond to histograms of these graph projections. The first one of this vector, the Fiedler vector is supposed to split the graph along a *cut* but in this case such a cut is difficult to interpret. Other vectors, such as the second and fourth ones are very well *localized* which makes them good candidates for clustering and for matching. These histograms also reveal that not all these eigenvectors are well localized. This suggests that some of the eigenvectors shown here are not well suited for spectral clustering.

the $N - 1$ row vectors lie on a hypersphere of dimension K , [8, 7]. As above, $\bar{\mathbf{S}}$ is a $K \times K$ sign-ambiguity diagonal form. Notice that, unlike \mathbf{Q}^* in (4), matrix \mathbf{Q} is not an orthonormal matrix anymore (since $\bar{\mathbf{U}}_x$ and $\bar{\mathbf{U}}_y$ are no longer orthonormal) and it has rank K . Consequently, it can only define an *inexact* (many-to-many) matching.

Suppose now that matrices $\bar{\mathbf{S}}$ and $\bar{\mathbf{P}}$ are correctly computed. By extension of the spectral graph matching theorem mentioned above we obtain the following interpretation. The entries of \mathbf{Q} in equation (5) can be written as $Q_{ij} = \bar{\mathbf{x}}_i^\top \bar{\mathbf{y}}_j$, where $\bar{\mathbf{x}}_i^\top$ and $\bar{\mathbf{y}}_j$ are K -dimensional vectors corresponding respectively to the rows of $\bar{\mathbf{U}}_x$ and to the columns of $\bar{\mathbf{S}} \bar{\mathbf{P}} \bar{\mathbf{U}}_y^\top$. Because of the normalization performed onto the rows of $\bar{\mathbf{U}}_x$ and of $\bar{\mathbf{U}}_y$, and because the two eigenvector bases are correctly aligned, the points $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{y}}_j$ lie *on the same hypersphere*. This suggests that \mathbf{Q} can be interpreted as a *loose* cosine node-similarity matrix, i.e., $-1 \leq Q_{ij} \leq 1$, and that good matches $\bar{\mathbf{x}}_i \leftrightarrow \bar{\mathbf{y}}_j$ may be chosen using a simple strategy such as $Q_{ij} > t > 0$. As a consequence, many-to-many matches may be established by seeking for each point in one set its nearest neighbors in the other set, and vice-versa.

This result is very important because it allows to bootstrap the matching of very large graphs through the alignment of two eigenbases of size K with $K \ll N$. We return now to the more general case when the two graphs have different cardinalities. The best one can hope in this case is to find the largest isomorphic subgraphs of the two graphs. In terms of spectral embeddings, this means that one has to find the largest subsets of points of the two sets of K -dimensional points. This problem can be stated as the following minimization problem:

$$\min_{\bar{\mathbf{R}}} \sum_{i,j} \|\bar{\mathbf{x}}_i - \alpha_{ij} \bar{\mathbf{R}} \bar{\mathbf{y}}_j\|^2 \quad (6)$$

This is an instance of the rigid point registration problem that can be solved by treating the assignments α_{ij} as missing variables in the framework of the expectation-maximization algorithm. A detailed solution is described in [13]. If matrices $\bar{\mathbf{S}}$ and $\bar{\mathbf{P}}$ are not correctly estimated, matrix $\bar{\mathbf{R}}$ belongs to the orthogonal group, i.e., rotations and reflections. However, if $\bar{\mathbf{S}}$ and $\bar{\mathbf{P}}$ are correctly estimated, $\bar{\mathbf{R}}$ belongs to the special orthogonal group, i.e., it is a rotation, which means that the two sets of points can be matched via an Euclidean transformation. The estimation of the latter is much more tractable than the more general case of orthogonal transformations.

Hence, the inexact graph matching problem at hand can be solved with the following three steps: (i) estimate matrices $\bar{\mathbf{S}}$ and $\bar{\mathbf{P}}$ using properties associated with the Laplacian eigenvectors, (ii) establish a match between the two sets of embedded nodes (K -D points) based on a nearest-neighbor strategy, and (iii) achieve point registration probabilistically by jointly estimating point-to-point assignments and a rotation matrix between the two sets of points.

4 Eigenvector alignment based on histograms

Each column U_{xi} of \bar{U}_x (as well as U_{yj} of \bar{U}_y), $1 \leq i, j \leq K$, corresponds to an eigenvector of dimension $N - 1$ (and of dimension $M - 1$). Hence, vector U_{xi} (and U_{yj}) can be interpreted as function projecting the nodes of \mathcal{G}_x (and of \mathcal{G}_y) onto the real line defined by $U_{xi} : \mathbb{R}^{N-1} \rightarrow \mathbb{R}$ (and by $U_{yj} : \mathbb{R}^{M-1} \rightarrow \mathbb{R}$).

In the case of two isomorphic graphs, there should be a one-to-one match between the eigenfunctions of the first graph and the eigenfunctions of the second graph, *provided that both \bar{S} and \bar{P} are known*. Indeed, in the isomorphic case, the node-to-node assignment between the two graphs is constrained by (4). Unfortunately, as already explained in the previous section, the two eigenbases cannot be reliably selected such that they span the same Euclidean space. Alternatively, we consider the histograms of the eigenfunctions just defined, namely $h([U_{xi}])$ and $h([U_{yj}])$, where the notation $h([U])$ corresponds to the histogram of the set of values returned by the eigenfunction U . The first observation is that these histograms are invariant to node permutation, i.e., invariant to the order in which the components of the eigenvectors are considered. Therefore, the histogram of an eigenfunction can be viewed as an invariant signature of an eigenvector. The second important observation is that $h([-U]) = h(B - [U])$, where B is the total number of bins used to build the histograms; Hence, the histograms can be used to detect sign flips. The third important observation is that the *shape* of the histogram is not too sensitive to the number of nodes in the graph and it is therefore possible to compare histograms arising from graphs with different cardinalities.

The problem of estimating matrices \bar{S} and \bar{P} can therefore be replaced by the problem of finding a set of assignments $\{U_{xi} \Leftrightarrow \pm U_{yj}, 1 \leq i, j \leq K\}$ based on the comparison of their histograms. This is an instance of the bipartite maximum matching algorithm already mentioned with complexity $O(K^3)$. Since the eigenvectors are defined up to a sign (modeled by \bar{S}), we must associate two histograms with each eigenfunction. Let $C(h_i, h_j)$ be a measure of similarity between two histograms. By computing the similarities between all pairs of histograms we can build a $K \times K$ matrix \bar{A} whose entries are defined by:

$$A_{ij} = \min\{C(h([U_{xi}]), h([U_{yj}])), C(h([U_{xi}]), h([-U_{yj}]))\}$$

as well as another matrix whose entries contain the signs of U_{yj} which are eventually retained. The Hungarian algorithm finds an optimal permutation matrix \bar{P} as well as a sign matrix \bar{S} .

5 Results

As a first example, consider a motion sequence of an articulated shape and their registration shown in figure 3. The articulated shape is described by a mesh/graph with 7,063 vertices and the degree of each vertex is approximately equal to six. The graphs were matched using the method described above, i.e., alignment of eigenvectors based on their histograms and *naive* point registration based on a nearest neighbor classifier.

On an average, the algorithm found 4,000 one-to-one matches and 3,000 many-to-many matches. Notice that in this case the two graphs are isomorphic.

Figure 4 shows two sets of eigenvector histograms (top and bottom) corresponding to the first pair of registered shapes of figure 3. The histograms shown on each row correspond to the five eigenvectors associated to the smallest non-null eigenvalues, shown in increasing order from left to right. There is a striking similarity between these histograms in spite of large discrepancies between the two shapes’ poses. This clearly suggests that these histograms are good candidates for both exact and inexact graph matching.



Fig. 3. The result of applying the graph matching algorithm to a dance sequence.

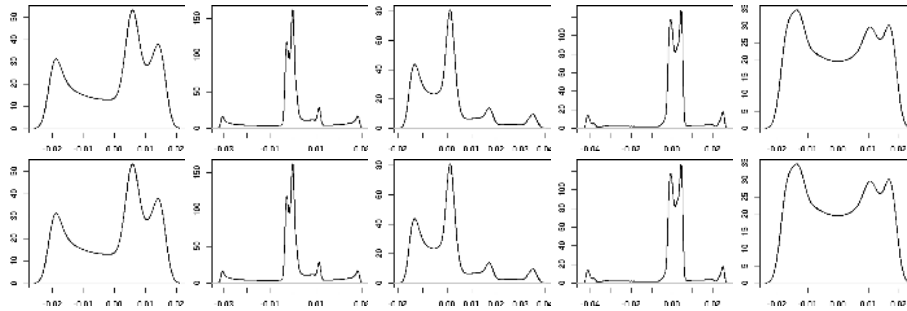


Fig. 4. Two sets of histograms associated with two sets of Laplacian eigenvectors. One may notice the striking similarity between these histograms that correspond to two isomorphic graphs.

Figure 5 shows three more examples of inexact graph matching corresponding to different shape pairs: dog-horse, dancer-gorilla, and horse-seahorse. Each mesh in this figure is described by a sparse graph and there are notable differences in the number of nodes. For example, the horse graph has 3,400 nodes, the gorilla graph has 2,046 nodes,

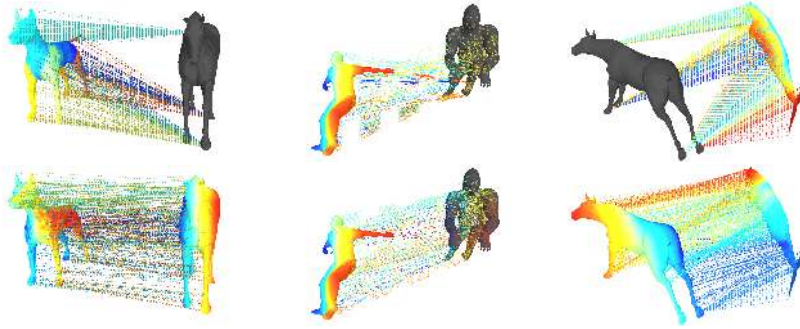


Fig. 5. Examples of matching different shapes (dog-horse, dancer-gorilla, and horse-seahorse) corresponding to graphs of different size. The top row shows the results obtained with the algorithm described in this paper while the second row shows the results after performing point registration with a variant of the EM algorithm. While in the first two examples (dog-horse and dancer-gorilla) the graphs have the same global structure, the third example shows the result obtained with two very different graphs.

the dancer graph has 34,000 nodes, and the seahorse graph has 2,190 nodes. The top row of figure 5 shows the result of many-to-many inexact matching obtained with the method described in this paper. The bottom row shows the result of one-to-one rigid point registration obtained with a variant of the EM algorithm [13] initialized from the matches shown on the top row.

6 Conclusion

In this paper, we proposed a framework for inexact matching of large and sparse graphs. The method is completely *unsupervised*, it does not need any prior set of matches between the two graphs. The main difficulty of the problem is twofold: (1) to extend the known spectral graph matching methods such that they can deal with graphs of very large size, i.e., of the order of 10,000 nodes and (2) to carry out the matching in a robust manner, i.e., in the presence of large discrepancies between the two graphs.

We showed that it is possible to relax the graph isomorphism problem such that inexact graph matching can be carried out when the dimension of the embedding space is much smaller than the number of vertices in the two graphs. We also showed that the alignment of the eigenbases associated with the two embedded shapes can be robustly estimated using eigenvector's density instead of eigenvalue ordering. The method starts with an initial solution based on ordering the eigenvalues and then it finds the optimal subset of eigenvectors that are aligned based on comparing their density distribution. This selects both a one-to-one eigenvector alignment and the dimension of the embedding. We also pointed out localization as an important property of eigenvectors and presented initial results to support our observations.

In future, we plan to investigate more thoroughly the link between graph matching and graph clustering. We believe localization property to be a promising direction to move forward.

References

1. Umeyama, S.: An eigen decomposition approach to weighted graph matching problems. *IEEE PAMI* **10** (1988) 695–703
2. Scott, G., Longuet-Higgins, C.: An algorithm for associating the features of two images. *Proceedings Biological Sciences* **244** (1991) 21–26
3. Shapiro, L., Brady, J.: Feature-based correspondence: an eigenvector approach. *Image and Vision Computing* **10** (1992) 283–288
4. Carcassoni, M., Hancock, E.R.: Correspondence matching with modal clusters. *IEEE PAMI* **25** (2003) 1609–1615
5. Carcassoni, M., Hancock, E.R.: Spectral correspondence for point pattern matching. *Pattern Recognition* **36** (2003) 193–204
6. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* **15** (2003) 1373–1396
7. Caelli, T., Kosinov, S.: An eigenspace projection clustering method for inexact graph matching. *IEEE PAMI* **26** (2004) 515–519
8. Ng, A., Jordan, M., Weiss, Y.: On spectral clustering: analysis and an algorithm. In: *NIPS*. (2002)
9. Chung, F.: *Spectral Graph Theory*. American Mathematical Society (1997)
10. Zhang, H., van Kaick, O., Dyer, R.: Spectral methods for mesh processing and analysis. In: *Eurographics Symposium on Geometry Processing*. (2007)
11. Biyikoglu, T., Leydold, J., Stadler, P.F.: *Laplacian Eigenvectors of Graphs*. Springer (2007)
12. Fraley, C., Raftery, A.: *MCLUST version 3 for R: Normal mixture modeling and model-based clustering*. Technical Report 504, Department of Statistics, University of Washington (2006)
13. Mateus, D., Horaud, R., Knossow, D., Cuzzolin, F., Boyer, E.: Articulated shape matching using Laplacian eigenfunctions and unsupervised point registration. In: *CVPR*. (2008)