

Segmentation via Graph-spectral Methods and Riemannian Geometry

Antonio Robles-Kelly^{1,2}

¹NICTA, Locked Bag 8001, Canberra ACT 2601, Australia

²RSISE, Building 115, ANU, Canberra ACT 0200, Australia

¹antonio.robles-kelly@nicta.com.au ²antonio.robles-kelly@anu.edu.au

Abstract

In this paper, we describe the use of graph-spectral techniques and their relationship to Riemannian geometry for the purposes of segmentation and grouping. We pose the problem of segmenting a set of tokens as that of partitioning the set of nodes in a graph whose edge weights are given by the geodesic distances between points in a manifold. To do this, we commence by explaining the relationship between the graph Laplacian, the incidence mapping of the graph and a Gram matrix of scalar products. This treatment permits the recovery of the embedding coordinates in a closed form and opens up the possibility of improving the segmentation results by modifying the metric of the space in which the manifold is defined. With the set of embedding coordinates at hand, we find the partition of the embedding space which maximises both, the inter-cluster distance and the intra-cluster affinity. The utility of the method for purposes of grouping is illustrated on a set of shape silhouettes.

1. Introduction

Many problems in computer vision can be posed as ones of pairwise clustering. That is to say, they involve grouping objects together based on their mutual similarity rather than their closeness to a cluster prototype. Such problems naturally lend themselves to a graph-theoretic treatment in which the objects to be clustered are represented using a weighted graph. Here the nodes represent the objects to be clustered and the edge-weights represent the strength of pairwise similarity relations between them.

One of the most elegant solutions to the pairwise clustering problem comes from spectral graph theory, i.e. the characterisation of the eigenpairs of the graph Laplacian and the adjacency matrix. Along these lines, some of the earliest work was done by Scott and Longuet-Higgins [11] who developed a method for refining the block-structure of the affinity matrix by relocating its eigenvectors. In a related development, Sarkar and Boyer [10] presented a spectral method which locates clusters that maximise the average association. Perona and Freeman [8] have a similar method which uses the second largest eigenvector of the affinity matrix. The method of Shi and Malik [12], on the other hand, uses the normalised cut, which balances the cut and the association.

Whereas eigenvectors have been traditionally viewed in combinatorial terms from the characteristic polynomial perspective, its relationship with eigenfunction expansions in differential geometry has been often overlooked for the purposes of segmentation and grouping. Nonetheless, there has recently been renewed interest in the use of manifold learning theory for the purposes of graph embedding [9, 6], classification [1] and visualisation [13].

In this paper, we draw on the field of mathematics known as spectral geometry, which aims to characterise the properties of operators on Riemannian manifolds using the eigenvalues and eigenvectors of the Laplacian matrix [3]. We commence by viewing the weight for the edge between each pair of nodes as a squared distance in a Euclidean space. Viewed in this way, the graph Laplacian can be related to a Gram matrix of scalar products. This, in turn, allows the use of matrix factorisation techniques to recover the coordinates for the embedding of the nodes in the graph into a metric space. With the embedding coordinates at hand, the clustering process is posed as a recursive partition of the space based upon a set of functions which bisect the embedding so as to minimise the distance between members of the same cluster and maximise the distance between elements of different clusters.

2. Clustering via Graph Embedding

We cast the problem of clustering into a graph-theoretic setting where the set of objects to be clustered are abstracted using a weighted graph. The problem is characterised by the set of nodes V that represent the objects and the set of edge weights, which are the affinities or “distances” between them. Viewed in this way, the goal of computation is then to partition the weighted graph $G = (V, E, W)$, with index-set V , edge-set $E = \{(v, w) | (v, w) \in V \times V, v \neq w\}$ and edge-weight function set $W : E \rightarrow [0, 1]$, into disjoint and disconnected subgraphs. Since the aim is to recover disconnected subgraphs, we can perform a recursive bipartition of the node-set V . In this section, we motivate the relationship between the graph Laplacian, the incidence mapping and the embedding of graphs into metric spaces. This treatment leads naturally to the clustering, as presented in the next section, via the sign of a vector of cluster membership functions.

To commence, we note that the weight matrix W is related to the normalised Laplacian $\mathcal{L} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{D} - W)\mathbf{D}^{-\frac{1}{2}} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathcal{L}\mathbf{D}^{-\frac{1}{2}}$, where \mathbf{D} is a diagonal matrix such that $\mathbf{D} = \text{diag}(\text{deg}(1), \text{deg}(2), \dots, \text{deg}(|V|))$. Consider the mapping \mathcal{I} of all functions over the set of vertices V to the functions $g(e)$ over the set of edges E . The incidence mapping \mathcal{I} is then an operator such that $\mathcal{I}g(e) = f(e_+) - f(e_-)$, where the nodes $v = e_+$ and $w = e_-$ are the head and tail, respectively, of the edge $e \in E$. As a result, \mathcal{I} is a $|V| \times |E|$ matrix which satisfies

$$\mathcal{L} = \mathcal{I}\mathcal{I}^T \tag{1}$$

The expression in Equation 1 opens-up the possibility of relating the graph Laplacian \mathcal{L} to a matrix $\mathbf{H} = \mathbf{J}\mathbf{J}^T$, which can then be viewed as a matrix of scalar products. We view the scalar products as the sums of squared, geodesic pairwise distances on a manifold. Hence, the problem of embedding the graph reduces itself to that of finding

the coordinates that satisfy, in an optimum manner, a set of known scalar products. We solve the problem in two stages. First, we find the matrix $\mathbf{H} = \mathbf{J}\mathbf{J}^T$, which is the matrix of pairwise sums of squares and scalar products. Second, we factorise it to find J .

With this in mind, we write the matrix H in terms of the graph Laplacian \mathcal{L} as follows

$$\mathbf{H} = -\frac{1}{2}\mathbf{B}\mathcal{L}\mathbf{B} \quad (2)$$

Further, by introducing the vector \mathbf{c} , whose entries are the diagonal elements of $\mathbf{J}\mathbf{J}^T$, and the all-ones vector \mathbf{e} , i.e. a vector whose coefficients are all unity, into the equation above, we can write

$$\mathbf{H} = -\frac{1}{2}\mathbf{B}[\mathbf{c}\mathbf{e}^T + \mathbf{e}\mathbf{c}^T - 2\mathbf{J}\mathbf{J}^T]\mathbf{B} \quad (3)$$

Note that, in order to have $\mathbf{H} = \mathbf{J}\mathbf{J}^T$, the vectors $(\mathbf{e}^T\mathbf{B})^T$ and $\mathbf{B}\mathbf{e}$ must be null and $\mathbf{B}\mathbf{J}\mathbf{J}^T\mathbf{B} = \mathbf{J}\mathbf{J}^T$. If \mathbf{B} is a centering matrix, this is the case and then the matrix \mathbf{H} becomes the double-centered graph Laplacian. Double centering is a standard procedure in classical scaling [14, 5] which introduces a linear dependency over the columns of the matrix \mathbf{H} [2]. As a result, the element indexed v, w of the matrix \mathbf{H} is given by

$$\mathbf{H}(v, w) = -\frac{1}{2}[\mathcal{L}(v, w)^2 - \mathcal{A}^2 - \mathcal{B}^2 + \mathcal{C}^2], \quad (4)$$

where

$$\mathcal{A} = \frac{1}{|\Upsilon_v|} \sum_{w \sim v} \mathcal{L}(v, w), \quad \mathcal{B} = \frac{1}{|\Upsilon_w|} \sum_{v \sim w} \mathcal{L}(v, w), \quad \mathcal{C} = \frac{1}{|V|^2} \sum_{w, v \in V} \mathcal{L}(v, w)$$

and Υ_v is the set of first-neighbours of the node $v \in V$. From the above equations, $\mathcal{A}f(v)$ can be regarded as the average value of the functions $\mathcal{I}g(e) = f(e_+) - f(e_-)$ over those nodes w adjacent to the node v . Similarly, $\mathcal{B}f(w)$ is the average value of the functions $-\mathcal{I}g(e) = f(e_-) - f(e_+)$ over those nodes $v \sim w$. The average value over the set of functions $\mathcal{I}g$ is given by $\mathcal{C}f$.

In order to obtain a matrix of embedding coordinates \mathbf{J} from the matrix \mathbf{H} , we perform an eigenvector analysis on \mathbf{H} . The validity of this procedure is based upon the Young-Householder factorisation theorem [15]. The Young-Householder theorem states that if \mathbf{J} is of the form $\mathbf{J} = [\psi_1 \mid \psi_2 \mid \dots \mid \psi_k]$ for all $i = 1, 2, \dots, k$, we have $(\mathbf{J}\mathbf{J}^T)\phi_i = \sqrt{\lambda_i}\psi_i$, where ψ_i is the eigenvector corresponding to the i^{th} eigenvalue λ_i .

Let ϕ_i be the i^{th} eigenvector scaled so its sum of squares is equal to λ_i (i.e. $\phi_i = \sqrt{\lambda_i}\psi_i$). Since $\mathbf{H}\phi_i = \lambda_i\phi_i$ and $(\mathbf{J}\mathbf{J}^T)\phi_i = \mathbf{H}\phi_i$, it follows that $\mathbf{H} = \mathbf{J}\mathbf{J}^T$. As a result, the vector of coordinates $\varphi(v)$ for the node $v \in V$ is given by $\varphi(v) = [\phi_1(v) \mid \phi_2(v) \mid \dots \mid \phi_{|V|}(v)]^T$, where the eigenvalues λ_i of the matrix \mathbf{H} are arranged according to their magnitude order so as to satisfy the condition $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_{|V|}| > 0$.

3. Partitioning of the Embedding Space

There are a number of important consequences that result from both, the double-centering operation on the graph Laplacian, and the Young-Householder theorem itself.

In this section, we use the properties of the embedding to recover the cluster membership functions $x(v) : v \in V \mapsto \mathfrak{R}$. The cluster membership functions are such that a pair of nodes v, w belong to the i^{th} cluster ω_i if and only if $\text{sgn}(x(v)) = \text{sgn}(x(w))$, where $\text{sgn}(\cdot)$ is the sign function. We recover the cluster membership functions as follows. Firstly, we show that the leading eigenvector of the double-centered Laplacian \mathbf{H} maximises the intra-cluster affinity. We then show that the leading eigenvector of the double-centered Laplacian also minimises the inter-cluster proximity. At this point, is also worth noting that, in order to be able to recover the set of membership functions $x(v) \forall v \in V$ in a closed form and, following Fiedler [4], we impose the constraints $\sum_{v \in V} x(v) = 0$ and $\sum_{v \in V} x(v)^2 = 1$ on the elements of the vector of cluster membership functions \mathbf{x} .

3.1. Maximising Intra-cluster Affinity

Recall that, in the previous section, we commenced by relating the embedding coordinates to the graph Laplacian using the matrix of scalar products $\mathbf{H} = \mathbf{J}\mathbf{J}^T$. As a result, the scalar product between the vectors of coordinates for the nodes v and $w \in V$ is given by

$$\langle \varphi(v), \varphi(w) \rangle = \sum_{l=1}^{|V|} \lambda_l \phi_l(v) \phi_l(w) \quad (5)$$

whereas the squared distance between the same pair of nodes is

$$\| \varphi(v) - \varphi(w) \|^2 = \sum_{l=1}^{|V|} \lambda_l (\phi_l(v) - \phi_l(w))^2 = \mathbf{H}(v, v) + \mathbf{H}(w, w) - 2\mathbf{H}(v, w) \quad (6)$$

It is also worth noting that, from Equations 5 and 6, it follows that $\langle \varphi(v), \varphi(v) \rangle \equiv 0$ for any node $v \in V$. This, in turn, implies that $\mathbf{H}(v, v) \equiv 0$.

With these ingredients, the problem of finding the functions $x(v)$ such that the distances between members of the same cluster is minimum can be viewed as that of minimising the quantity

$$\epsilon = \sum_{\omega_i \in \Omega} \sum_{v, w \in \omega_i \subset V} \|x(v)\varphi(v) - x(w)\varphi(w)\|^2 \quad (7)$$

where Ω is the set of all clusters ω_i . Thus, only pairwise distances corresponding to nodes in the same cluster contribute to the quantity ϵ . To take our analysis further, we use Equation 6 and, after some algebra, write

$$\epsilon = \sum_{v, w \in V} (x(v)^2 \mathbf{H}(v, v) + x(w)^2 \mathbf{H}(w, w) - 2x(v)x(w)\mathbf{H}(v, w)) \quad (8)$$

but, since $\mathbf{H}(v, v) \equiv 0$, the equation above can be rewritten, in matrix notation, as follows

$$\epsilon = -2\mathbf{x}\mathbf{H}\mathbf{x} \quad (9)$$

where \mathbf{x} is the vector of order $|V|$ whose i^{th} element is given by $x(v)$. Thus, minimising ϵ is equivalent to maximising $\mathbf{x}\mathbf{H}\mathbf{x}$ and, therefore, \mathbf{x} is given by the leading eigenvector of \mathbf{H} , i.e. $\mathbf{x} = \phi_1$.

3.2. Minimising Inter-cluster Proximity

Having shown that the leading eigenvector of the double-centered Laplacian \mathbf{H} is the maximiser of intra-cluster affinity, we now proceed to prove that it also minimises the inter-cluster proximity. We do this by making use of Lagrange multipliers to show that the distance between the centers of mass for the clusters ω_i and ω_j is maximised when the set of nodes in the graph are partitioned using the leading eigenvector of \mathbf{H} . We commence by noting that the squared distance between cluster mass centers is given by

$$\rho(\omega_i, \omega_j) = \left\| \frac{1}{|\omega_i|} \sum_{v \in \omega_i \subset V} \varphi(v) - \frac{1}{|\omega_j|} \sum_{w \in \omega_j \subset V} \varphi(w) \right\|^2 \quad (10)$$

At this point, we note that, as a consequence of the matrix \mathbf{H} being double-centered, we have $\frac{1}{|V|} \sum_{v \in V} \varphi(v) \equiv 0$. In other words, the center of mass for the embedding is at the origin and, therefore, the squared distance $\rho(\omega_i, \omega_j)$ can be expressed in terms of the embedding coordinates as follows

$$\rho(\omega_i, \omega_j) = 2 \sum_{l=1}^{|V|} \left\{ \frac{\lambda_l}{|V|^2} \left(\sum_{v \in \omega_i \subset V} \phi_l(v) + \sum_{w \in \omega_j \subset V} \phi_l(w) \right)^2 \right\} \quad (11)$$

The importance of this observation resides in the fact that it enables us to make use of the cluster memberships to introduce the weighted squared distance

$$\varrho(\omega_i, \omega_j) = \tau \sum_{l=1}^{|V|} \left\{ \lambda_l \left(\sum_{v \in \omega_i \subset V} x(v) \phi_l(v) + \sum_{w \in \omega_j \subset V} x(w) \phi_l(w) \right)^2 \right\} \quad (12)$$

where $\tau = \frac{2}{|V|^2}$. The quantity above is the one we aim to maximise. Through the use of Lagrange multipliers, we can use the norm constraint $\|\mathbf{x}\|^2 = 1$ for the vector of cluster membership functions and write

$$\tau \frac{\partial \varrho(\omega_i, \omega_j)}{\partial x(w)} = \zeta \frac{\partial \|\mathbf{x}\|^2}{\partial x(w)} \quad (13)$$

where ζ is a Lagrange multiplier. After some algebra, it can be shown that the system of equations obtained in this manner can be cast as an eigenvalue problem of the form

$$\tau \mathbf{H} \mathbf{x} = \zeta \mathbf{x} \quad (14)$$

Therefore, the maximum of $\varrho(\omega_i, \omega_j)$ is reached when \mathbf{x} corresponds to the leading eigenvector of the double-centered Laplacian \mathbf{H} .

4. Deforming the Embedding Space

In this section, we explore a means to improve the clustering results by modifying the intrinsic geometric properties of the embedding space \mathcal{S} for the graph G . The goal of computation here is, therefore, to compute an improved matrix of edge weights \hat{W} that is both, representative of the space in which the nodes of the graph G are to be embedded and, most importantly, better suited for the purposes of clustering. We do this

by requiring the embedding space to have constant positive curvature. This approach hinges in the properties of spherical spaces. On a sphere, the optimum separation of the space corresponds to the hyperplane bisecting the sphere through the equator, i.e. the centers of mass for the clusters are at either opposite pole of the sphere.

To this end, we view the embedding space \mathcal{S} as an n -dimensional Riemannian manifold M and express the weight $W(v, w)$ as the energy $\mathcal{E}(p_v, p_w)$ over the parametric geodesic curve $\gamma : t \in [\alpha, \beta] \mapsto M$ intersecting the pair of points $p_v, p_w \in M$. To establish a relationship between the energy $\mathcal{E}(p_v, p_w)$, the geodesic γ and the curvature tensor, we employ the theory of Jacobi vector fields [3]. A Jacobi field along γ is the differentiable vector field Y in the tangent space to M , orthogonal to γ' , satisfying Jacobi's equation, i.e. $\nabla_t^2 Y + R(\gamma', Y)\gamma' = 0$. With these ingredients, we can obtain a bilinear form, i.e. the sectional curvature, from the curvature tensor $R(\gamma', Y)\gamma'$. The sectional curvature $\mathcal{K}(\gamma', Y)$ along γ is, hence, given by

$$\mathcal{K}(\gamma', Y) = \frac{\langle R(\gamma', Y)\gamma', Y \rangle}{|\gamma'|^2 |Y|^2 - \langle \gamma', Y \rangle^2} \quad (15)$$

Because Y is orthogonal to γ' and, due to the fact that Y is a Jacobi field, i.e. it satisfies the condition $\nabla_t^2 Y = -R(\gamma', Y)\gamma'$, we can write

$$\mathcal{K}(\gamma', Y) = \frac{\langle -\nabla_t^2 Y, Y \rangle}{\langle Y, Y \rangle} \quad (16)$$

where we have set $|\gamma'| = 1$.

This suggests a way of formulating the energy over the geodesic $\gamma \in M$ connecting the pair of points corresponding to the nodes indexed v and w . Consider the geodesic γ subject to the Jacobi field Y . The energy over the geodesic γ can be expressed making use of the following equation

$$\mathcal{E}(p_v, p_w) = \int_{\gamma} |\gamma' + \nabla_t^2 Y|^2 dt = \int_{\gamma} |\gamma' - \mathcal{K}(\gamma', Y)Y|^2 dt \quad (17)$$

In practice, as stated at the beginning of the section, we will confine our attention to the problem of embedding the nodes into a manifold of constant sectional curvature. For such a manifold, the sectional curvature is constant i.e. $\mathcal{K}(\gamma', Y) \equiv \kappa$. Under this restriction, the Jacobi field equation becomes $\nabla_t^2 Y = -\kappa Y$. With the boundary conditions $Y(0) = 0$ and $|\nabla_t Y(0)| = 1$, the solution is given by $Y(t) = \frac{\sin(\sqrt{\kappa}t)}{\sqrt{\kappa}}\eta$, where the vector η is in the tangent space of M at p_v and is orthogonal to γ' at the point indexed v , i.e. $\eta \in M_{p_v}$ and $\langle \eta, \gamma' |_{p_v} \rangle = 0$. Further, by rescaling the parameter t so that $|\gamma'| = a$, we can express the element of the improved weight matrix \hat{W} as the energy over the geodesic connecting the points p_v and p_w as follows

$$\hat{W}(v, w) = \int_0^1 \left(a_{v,w}^2 + \kappa \left(\sin(\sqrt{\kappa}a_{v,w}t) \right)^2 \right) dt \quad (18)$$

where $a_{v,w} = W(v, w)$ is the Euclidean distance between each pair of nodes when they have been embedded into a “flat” space, i.e. $a_{v,w} = \|p_v - p_w\|$ for $\kappa = 0$.

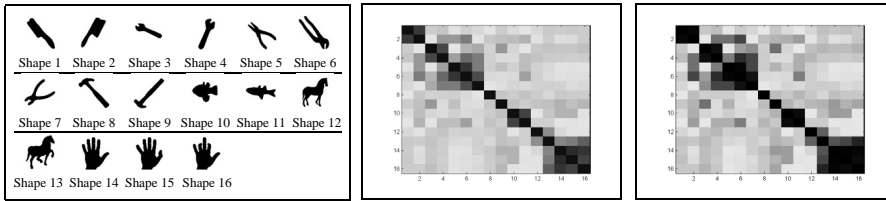
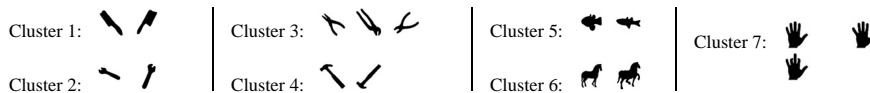


Figure 1: From left-to-right: Set of silhouettes used in our experiments, matrix of raw similarities for the shapes in the left-hand panel and matrix \hat{W} computed from the raw similarities.

5. Experiments

In this section, we demonstrate the utility of the algorithm for purposes of unsupervised learning of shape-categories by means of grouping. This involves the abstraction of 2D binary shapes using shock-trees. Commencing from a data-base of silhouettes, the Hamilton-Jacobi skeleton is extracted and shocks, which correspond to singularities in the evolution of the object boundary under the eikonal equation, are located. The similarity of the shapes is then computed using the weighted tree edit distance developed by Luo *et al.* [7]. This is a structural method which hinges in augmenting the information given by the skeleton topology and the relative time of shock formation with a measure of feature importance based upon the rate of change of boundary length with respect to the overall distance along the skeleton. The problem of computing distances between pairs of shapes is then cast as that of finding the tree edit distance between the weighted graphs for their corresponding shock graphs.

In the left-hand panel of Figure 1, we show the shapes used in our experiments. The remaining two panels show the matrix of raw similarities, as computed using the algorithm of Luo *et al.* and the matrix \hat{W} obtained by setting $\kappa = 5$. Here, the row and column indexes for both matrices have been set to those in the left-hand panel of Figure 1. The clusters recovered by the algorithm are the following



From the matrices in Figure 1, we can conclude that the block-diagonal structure of the matrix of raw shape similarities has been enhanced in the the matrix \hat{W} . Furthermore, the shape classes recovered by the algorithm are in good accordance with the silhouette categories in the database.

6. Conclusions

In this paper, we have cast the clustering problem in a graph theoretical setting. This opens up the possibility of embedding the objects to be clustered, abstracted as nodes in the graph, into a metric space. We have shown how the incidence mapping and the graph Laplacian can be used to pose the problem of embedding as that of recovering a Gram matrix of scalar products. Further, we have illustrated how the topology of the

embedding space can be altered to improve the separation between clusters and, hence, the grouping results. We illustrated the utility of the method for purposes of grouping a set of shape silhouettes.

Aknowldegment

The author is indebted with Dr. Andrea Torsello (e-mail: torsello@dsi.unive.it) from the Univeristà Ca' Foscari di Venezia, Italy, for facilitating him the shape silhouettes and similarities used in the experimental section of this paper.

References

- [1] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Neural Information Processing Systems*, number 14, pages 634–640, 2002.
- [2] I. Borg and P. Groenen. *Modern Multidimensional Scaling, Theory and Applications*. Springer Series in Statistics. Springer, 1997.
- [3] I. Chavel. *Riemannian Geometry: A Modern Introduction*. Cambridge University Press, 1995.
- [4] M. Fiedler. A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czech Math. Journal*, (25):619–633, 1975.
- [5] J. C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53:325–328, 1966.
- [6] J. Ham, D. D. Lee, S. Mika, and B. Scholkopf. A kernel view of the dimensionality reduction of manifolds. In *In Proc. Int. Conf. Machine Learning*, page 369376, 2004.
- [7] B. Luo, A. Robles-Kelly, A. Torsello, R. C. Wilson, and E. R. Hancock. A probabilistic framework for graph clustering. In *EEE International Conference on Computer Vision and Pattern Recognition*, pages I:912–919, 2001.
- [8] P. Perona and W. T. Freeman. Factorization approach to grouping. In *Proc. ECCV*, pages 655–670, 1998.
- [9] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.
- [10] S. Sarkar and K. L. Boyer. Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors. *Computer Vision and Image Understanding*, 71(1):110–136, 1998.
- [11] G. L. Scott and H. C. Longuet-Higgins. Feature grouping by relocalisation of eigenvectors of the proximity matrix. In *British Machine Vision Conference*, pages 103–108, 1990.
- [12] J. Shi and J. Malik. Normalized cuts and image segmentations. In *Proc. of the IEEE Conf. on Comp. Vision and Pattern Recognition*, pages 731–737, 1997.
- [13] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
- [14] W. S. Torgerson. Multidimensional scaling I: Theory and method. *Psychometrika*, 17:401–419, 1952.
- [15] G. Young and A. S. Householder. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3:19–22, 1938.