A GRAPH-BASED APPROACH FOR FEATURE EXTRACTION AND SEGMENTATION OF MULTIMODAL IMAGES

Geoffrey Iyer[1,2], **Jocelyn Chanussot**[1,2], **and Andrea Bertozzi**[1]

1. University of California, Los Angeles
2. Univ. Grenoble Alpes, GIPSA-lab

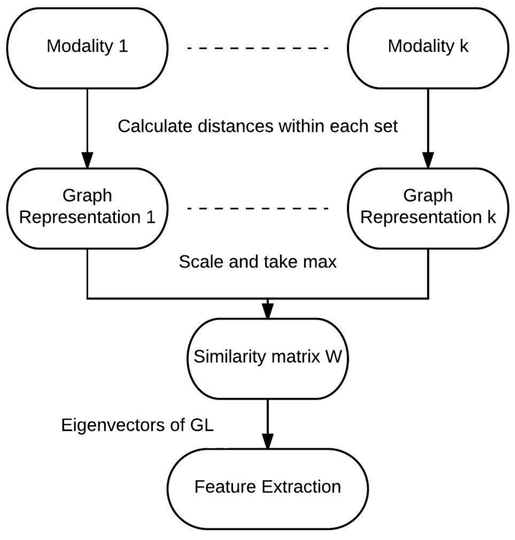
Contact Information:
Geoffrey Iyer
UCLA Mathematics Department
Box 951555
Los Angeles, CA 90095
Email: gsiyer@math.ucla.edu

Introduction

With the increasing availability of data we often come upon problems that collect data from more than once source, or *modality*. To properly handle these problems, we need to compare data across the different modalities. The step presents a lot of difficulty, as it requires some understanding of the format of the data [5]. Our goal is to create a general algorithm for processing multiple modalities simultaneously.

In this paper, we assume our datasets are co-registered (each modality contains the same number of points, and they share a common indexing), as is often the case in image processing problems. Our method compares graph representations of each modality, extracts features in the form of eigenvectors of the graph Laplacian, then applies standard data-segmentation algorithms on these features to obtain a final classification.

1 Feature Extraction



Graph Representation

- Datasets $X^1, X^2, \ldots, X^k$, with $\left|X^1\right| = \cdots = \left|X^k\right| = m$.
- E^ℓ = *distance matrix*. $e^\ell_{ij} = \left\|x^\ell_i - x^\ell_j\right\|$.
- $\lambda_\ell = \text{stdev}\left(E^\ell\right)$, scaling factor.
- $X = (X^1, X^2, \ldots, X^k) \subseteq \mathbb{R}^{n\times(dim_1+\cdots+dim_k)}$ the concatenated dataset.
- W = *similarity matrix*. w_{ij} = similarity between x_i and x_j.

$$w_{ij} := \exp\left(-\max\left(e^\ell_{ij}/\lambda_\ell \mid 1 \le \ell \le k\right)\right). \tag{1}$$

With the $\max$ norm, two nodes are considered similar only if they are similar in every individual modality. Heuristically, this emphasizes the unique information that each dataset brings.

The Graph Laplacian (GL) and Eigendecomposition

Once we have created the weights, we define the *normalized graph Laplacian* (GL) [9].

$$L_{sym} = I - D^{-1/2}WD^{-1/2}, \tag{2}$$

Where D is the diagonal matrix consisting of degrees of nodes. We use the graph Laplacian in our energy minimizations below.

We use the GL to solve the *graph N-cut* problem. Given a partition of our into subsets $A_1, A_2, \ldots, A_m$, we define

$$\text{NCut}(A_1, \ldots, A_m) = \frac{1}{2}\sum_{i=1}^{m}\frac{W(A_i, A_i^c)}{vol(A_i)}. \tag{3}$$

Minimizing the NCut separates dissimilar nodes (the $W(A_i, A_i^c)$ term) and groups similar nodes (the $vol(A_i)$ term). Solving the graph min-cut problem is equivalent to finding an $n \times m$ indicator matrix u, where $u_{ij} = 1$ if $x_i \in A_j$, and $u_{ij} = 0$ otherwise. Note that

$$\text{NCut}(A_1, \ldots, A_m) = \text{Tr}\left(u^T L_{sym} u\right). \tag{4}$$

Minimizing this energy is computationally infeasible [4]. We relax the problem, allowing u to be an orthogonal matrix. We find

$$\text{argmin}_{u\in\mathbb{R}^{n\times m}}\text{Tr}\left(u^T L_{sym} u\right) \quad \text{where } u^T u = I. \tag{5}$$

This problem is solved by choosing the columns of u to be the m eigenvectors of L_{sym} corresponding to the m smallest eigenvalues. These eigenvectors are features extracted from the original dataset X.

Example: Data Fusion Contest 2015 [1]

Figure 1: RGB Data

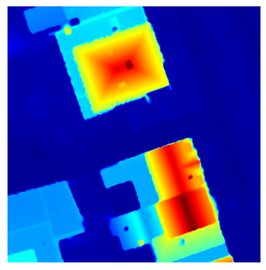


Figure 2: Lidar Data

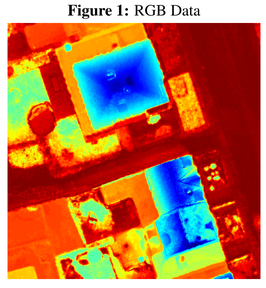


Figure 3: Eigenvector 1

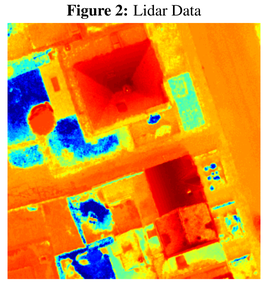


Figure 4: Eigenvector 2

Segmentation

We apply two different segmentation algorithms to these features. The first is *Spectral Clustering*, in which we directly apply k-means to the feature vectors. The second is *Graph MBO*, explained below.

Graph MBO

Here we minimize a Ginzburg-Landau energy with a semisupervised term [3, 7, 8]. Have u an $n \times m$ assignment matrix with

$$u_{ij} \ge 0 \;\forall i, j, \quad \sum_{j=1}^{m} u_{ij} = 1. \tag{6}$$

The final output of the algorithm will be a matrix u where each value is either 0 or 1. The energy we minimize is

$$E(u) = \epsilon \cdot \text{Tr}\left(u^T L_{sym} u\right) + \frac{1}{\epsilon}\sum_i W(u_i) + \sum_i \frac{\mu}{2}\lambda(x_i)\left\|u_i - \hat{u}_i\right\|^2_{L_2}. \tag{7}$$

The first term is the graph cut energy, similar to (3). The second term is the multiwell potential $W(u_i) = \prod_{k=1}^{m}\frac{1}{4}\left\|u_i - e_k\right\|^2_{L_1}$ where e_k is the k-th standard basis vector. The last term includes the fidelity, where $\hat{u}$ represents the semisupervised input,

$$\lambda(x_i) = \begin{cases} 1 & \text{if } x_i \text{ is part of fidelity input} \\ 0 & \text{else} \end{cases}, \tag{8}$$

We minimize this via iteritive diffusion and thresholding. If u^n represents the n-th iterate, then to calculate u^{n+1} we first diffuse

$$\frac{u^{n+\frac{1}{2}} - u^n}{dt} = -L_{sym}u^n - \mu\lambda(x)(u^n - \hat{u}). \tag{9}$$

Then threshold each row

$$u_i^{n+1} = e_r \quad \text{where } r = \text{argmax}_j u_{ij}^{n+\frac{1}{2}}. \tag{10}$$

The diffusion calculation can be done very efficiently by using the eigendecomposition of L_{sym} (the feature vectors described in (5)). If we write change coordinates to the eigenbasis, then the diffusion step reduces to solving for coefficients

$$a_k^{n+1} = (1 - dt\cdot\lambda_k)\cdot a_k^n - dt\cdot d_k^n. \tag{11}$$

where λ_k is the k-th eigenvalue of L_{sym}, in ascending order.

Results: Data Fusion Contest 2015 [1]

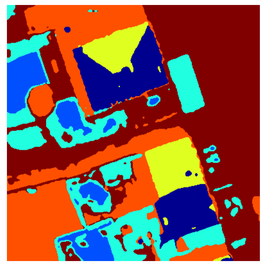


Figure 5: Spectral Clustering

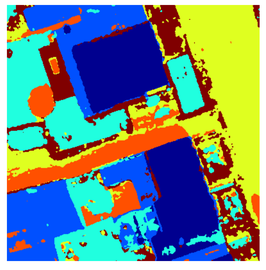


Figure 6: Graph MBO

Results: Umbrella Data [11]

Figure 7: RGB Data

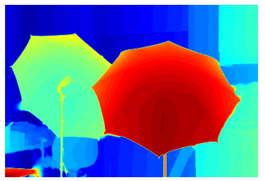


Figure 8: Lidar Data

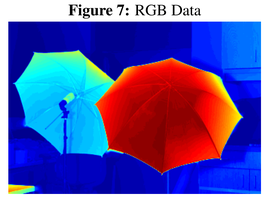


Figure 9: Eigenvector 1

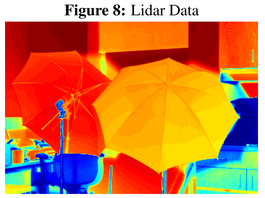


Figure 10: Eigenvector 2

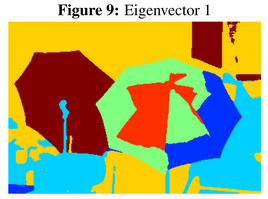


Figure 11: Spectral Clustering

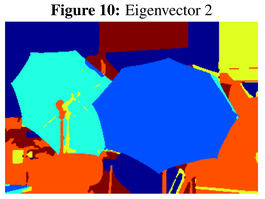


Figure 12: Graph MBO

Results: Jade Plant Data [11]

Figure 13: RGB Data

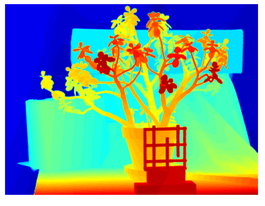


Figure 14: Lidar Data

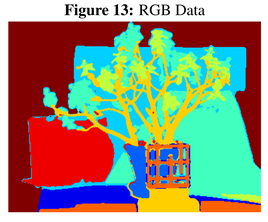


Figure 15: Spectral Clustering

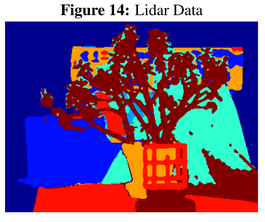


Figure 16: Graph MBO

Acknowledgments

This work was supported by NSF grant DMS-1118971, ONR grant N00014-16-1-2119, NSF grant DMS-1417674, European Research Council (Grant no. 320684 - CHESS project), and CNRS (Grant no. PICS-USA 263484)

References

[1] Campos-Taberner, M., Romero-Soriano, A., Gatta, C., Camps-Valls, G., Lagrange, A., Saux, B. L., Beaupre, A., Boulch, A., Chan-Hon-Tong, A., Herbin, S., Randrianarivo, H., Ferecatu, M., Shimoni, M., Moser, G., and Tuia, D. Processing of extremely high-resolution lidar and rgb data: Outcome of the 2015 ieee grss data fusion contest #8211;part a: 2-d contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9*, 12 (Dec 2016), 5547–5559.

[2] Fowlkes, C., Belongie, S., Chung, F., and Malik, J. Spectral grouping using the nystrom method. *IEEE Transactions on Pattern Analysis and Machine Intelligence 26*, 2 (February 2004).

[3] Garcia-Cardona, C., Merkurjev, E., Bertozzi, A., Flenner, A., and Percus, A. Multiclass data segmentation using diffuse interface methods on graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence 36*, 8 (2014), 1600–1613.

[4] Goldschmidt, O., and Hochbaum, D. S. A polynomial algorithm for the k-cut problem for fixed k. *Mathematics of Operations Research 19*, 1 (1994), 24–37.

[5] Lahat, D., Adali, T., and Jutten, C. Challenges in Multimodal Data Fusion. In *22nd European Signal Processing Conference (EUSIPCO-2014)* (Lisbonne, Portugal, Sept. 2014), pp. 101–105.

[6] Meng, Z., Merkurjev, E., Koniges, A., and Bertozzi, A. L. Hyperspectral image classification using graph clustering methods. Note: This is a preprint, 2017.

[7] Merkurjev, E., Garcia-Cardona, C., Bertozzi, A. L., Flenner, A., and Percus, A. G. Diffuse interface methods for multiclass segmentation of high-dimensional data. *Applied Mathematics Letters 33* (2014), 29 – 34.

[8] Merkurjev, E., Kostic, T., and Bertozzi, A. L. An mbo scheme on graphs for classification and image processing. *SIAM Journal on Imaging Sciences 6* (October 2013), 1903–1930.

[9] Mohar, B. The laplacian spectrum of graphs. *Graph Theory, Combinatorics, and Applications 2* (1991), 871–898.

[10] Potamianos, G., Neti, C., Gravier, G., Garg, A., and Senior, A. W. Recent advances in the automatic recognition of audiovisual speech. *Proceedings of the IEEE 91*, 9 (Sept 2003), 1306–1326.

[11] Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X., and Westling, P. High-resolution stereo datasets with subpixel-accurate ground truth. In *Proceedings of the 36th German Conference on Pattern Recognition* (september 2014).

[12] Sedighin, F., Babaie-Zadeh, M., Rivet, B., and Jutten, C. Two Multimodal Approaches for Single Microphone Source Separation. In *24th European Signal Processing Conference (EUSIPCO 2016)* (Budapest, Hungary, Sept. 2016), pp. 110–114.

[13] Woodworth, J. T., Mohler, G. O., Bertozzi, A. L., and Brantingham, P. J. Non-local crime density estimation incorporating housing information. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 372*, 2028 (2014).