Multimodal Data Processing

Geoffrey Iyer

University of California

gsiyer@math.ucla.edu

April 12, 2017

Overview

1. Introduction: Multimodal Data
   - Multimodality
   - Manifold Alignment
   - Synthetic Example
2. Our Method
   - Problem setup
   - Multimodal Weights
   - Graph Laplacian Theory
   - Nyström Extension
3. Results
   - DFC2015 Data
   - Umbrella Data
4. Future Work
   - Manifold Alignment
5. References

Multimodal datasets

With the increasing availability of data, many applications involve data drawn from more than one source (called *modalities*).

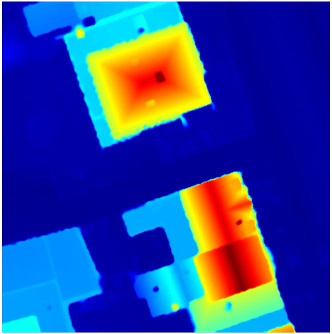
Ex: In medicine, a patient is often tested and monitored in several different ways. This data is then synthesized (usually by a doctor) to make the final diagnosis and treatment decision.

The existence of multimodal data raises a new topic: how to solve standard machine learning problems with these sets?

Example Multimodal Data

Ex: RGB + Elevation map of residential neighborhood in Belgium. Found in [Bampos-Taberner et al, 2016].

RGB Data



Lidar Data

Examples from the literature

Placeholder slide. Will fill in later.

I'm interested in citing [Song et al, 2012] and [Correa et al, 2010] and [Sedighin et al, 2016].

For each paper, give a short explanation of how multimodality is important.

Challenges in multimodality

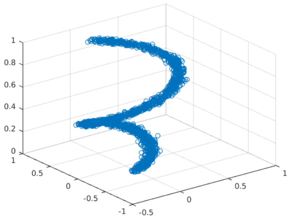
Most multimodal methods are developed specifically for one problem.

[Lahat et al, 2015]: "... a solution that is based on a sufficiently data-driven, model-free approach may turn out to be useful in very different domains."

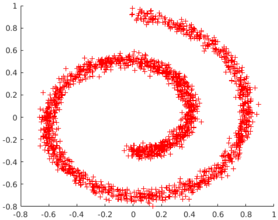
Manifold alignment

Attempt to address multimodality in general via manifold alignment.

For each modality, view the data as a manifold (have sets $X^1, X^2, \ldots, X^k$. $k =$ number of modalities).

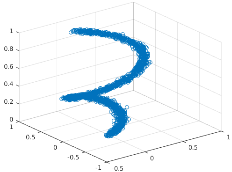


X^1

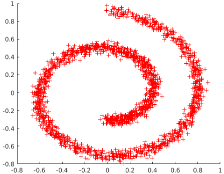


X^2

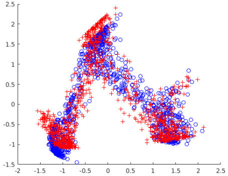
Create a *latent space* Y and maps $X^i \to Y$.



X^1



X^2



Images in latent space

Example from [Tuia et al, 2016]

Compare sets by using the latent space image.

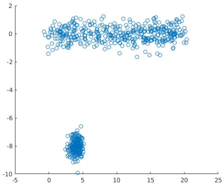
Manifold alignment: Methods from the literature

Some examples from the literature:

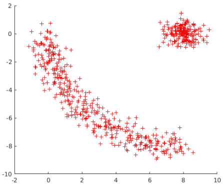
- [Yeh et al, 2014]: Canonical Correlation Analysis, linear or with nonlinear kernel (unsupervised)
- [Wang et al, 2013]: Graph-based methods (semi-supervised)
- [Tuia et al, 2016]: Similar to [Wang et al, 2013] with an added nonlinear kernel (semi-supervised)

Manifold alignment: Methods from the literature

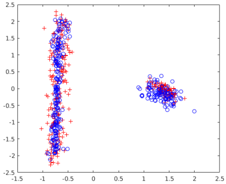
Common theme: Create the latent space by finding and correlating redundancies between sets.



X^1



X^2



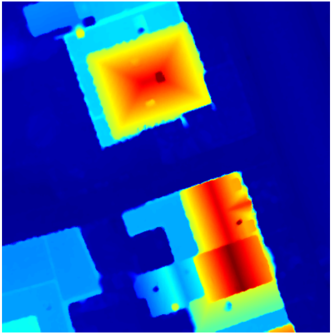
Images in latent space

Using code from [Tuia et al, 2016]

Manifold alignment: Our goal

Our idea: Can improve on these methods. Find and exploit the unique information that each modality brings.

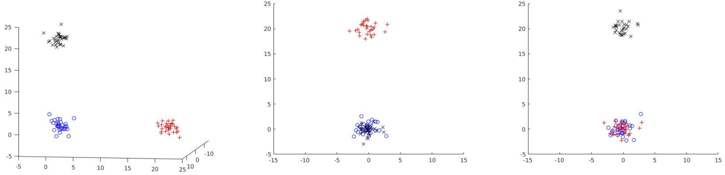
Distinguish road from grass



Distinguish roof from ground

Synthetic example: Data

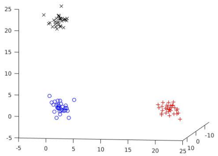
Synthetic example:
Ground truth = 3 point clouds in $\mathbb{R}^3$ (20 points per cloud).
Modality 1 = projection onto xy-plane.
Modality 2 = projection onto xz-plane.



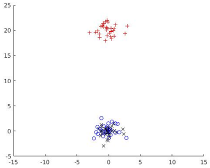
Ground Truth | Modality 1 | Modality 2

Synthetic example: Data

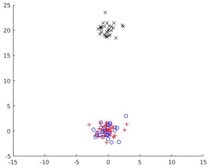
Data is *co-registered*. *i*-th point from modality 1 corresponds to *i*-th point from modality 2. This is used in the algorithm.



Ground Truth



Modality 1



Modality 2

Synthetic Example: Result of CCA

Result of CCA algorithm from [Yeh et al, 2014] applied to the data:

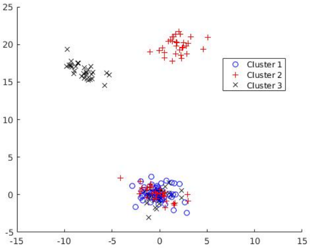


Image of clusters in latent space

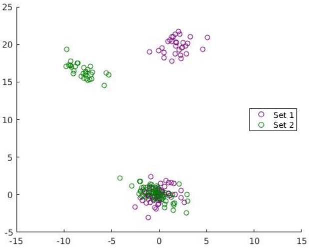


Image of data in latent space

Synthetic Example: Result of Our Method

Result of our multimodal graph-based algorithm applied to the data:

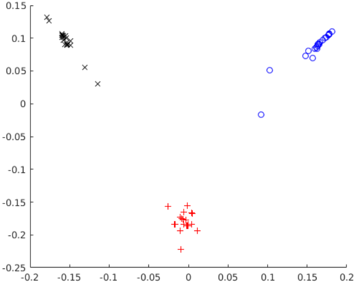
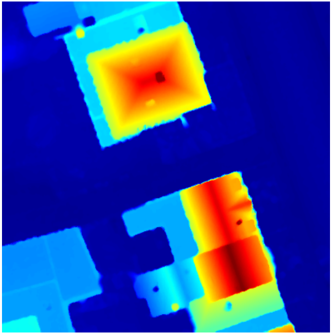


Image of clusters in latent space

Problem setup

We use co-registration assumption and Graph Laplacian theory for segmentation of multimodal datasets.

RGB Data



Lidar Data

Notation

From each modality, have a data set X^k. $m =$ number of modalities.

$N =$ number of observations.

$d_j =$ dimension of set X^k. (Can view $X^k \in \mathbb{R}^{N \times d_k}$).

From co-registration assumption: i-th point in X^{k_1} corresponds to i-th point in X^{k_2}. Create concatenated set $X = (X^1, X^2, \dots, X^\ell) \subseteq \mathbb{R}^{N \times (d_1 + \cdots + d_\ell)}$.

$x_i =$ element i from X. $x_i^k =$ element i from X^k.

Weight Matrix: Background

For each pair $x_i, x_j \in X$, define a *weight* w_{ij} that measures the similarity between the points.

$\implies$ represent data as $N \times N$ weight matrix W.

Common similarity measure from the literature: RBF kernel

$$w_{ij} = \exp\left(- \left\| x_i - x_j \right\| / \sigma\right).$$

Need to adapt this to multimodal data.

Multimodal Weight Matrix

For each modality X^k, calculate the distance matrix E^k via

$$E_{ij}^k = \left\| x_i^k - x_j^k \right\|.$$

$\|\cdot\|$ chosen based on the details of the modality. (in our examples we use the 2-norm)

Scale each distance matrix by standard deviation

$$\bar{E}^k = \frac{E^k}{\operatorname{std}\left(E^k\right)}.$$

Multimodal Weight Matrix

Define

$$w_{ij} = \exp\left(-\max\left(\bar{E}_{ij}^{1}, \ldots, \bar{E}_{ij}^{k}\right)/\sigma\right).$$

Heuristics:

- Standard devation scaling allows us to directly compare $\bar{E}^{k_1}, \bar{E}^{k_2}$ with reasonable results.
- Because of the max, elements are similar under this measure only if they are similar in each modality.

Graph min cut

Using W, state the problem as graph-cut minimization.

Given a partition of X into subsets $A_1, A_2, \ldots, A_m$, we define the *normalized graph-cut*

$$\mathrm{Ncut}(A_1, \ldots, A_m) = \frac{1}{2} \sum_{k=1}^{m} \frac{W(A_k, A_k^c)}{vol(A_k)}.$$

$$W(A, B) = \sum_{i \in A, j \in B} w_{ij}.$$

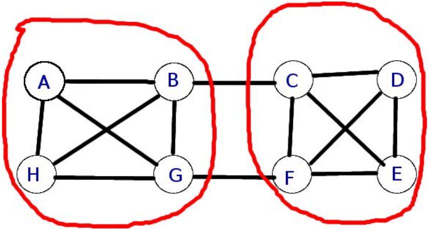
$$vol(A) = \sum_{i \in A, j \in \{1, \ldots, n\}} w_{ij}.$$

Graph min cut

$$\text{Ncut}(A_1, \ldots, A_m) = \frac{1}{2}\sum_{k=1}^{m} \frac{W(A_k, A_k^c)}{vol(A_k)}.$$

Minimize graph cut $\implies$ segment set. Compare the edges cut as a fraction of total edges.

Solving exactly is $O\left(|X|^{m^2}\right)$.



Example graph cut. $m = 2$

Graph Laplacian

Let $D = N \times N$ diagonal matrix, with

$$d_{ii} = \sum_{j=1}^{n} w_{ij}.$$

Graph Laplacian

$$L = D - W.$$

Graph Laplacian

From $A_1, \dots, A_m$, get $H = N \times m$ indicator matrix.

$$H_{ij} = \begin{cases} \frac{1}{vol(A_j)} & \text{if } x_i \in A_j \\ 0 & \text{else} \end{cases}$$

Columns of $H \iff$ classes. Rows of $H \iff$ data points.

$$\begin{aligned} \mathrm{Ncut}(A_1, \dots, A_m) &= \frac{1}{2} \sum_{i=1}^{m} \frac{W(A_i, A_i^c)}{vol(A_i)} \\ &= \mathrm{Tr}\left(H^T L H\right). \end{aligned}$$

Relaxed graph min cut

Optimal graph cut is

$$\text{argmin}_{H \text{ an indicator matrix}} \operatorname{Tr}\left(H^T L H\right).$$

This is an $O\left(|X|^{m^2}\right)$ problem. Instead we solve the relaxed problem:

$$\text{argmin}_{H \in \mathbb{R}^{n \times m},\ H^T H = I} \operatorname{Tr}\left(H^T L H\right).$$

Solution: Columns of $H =$ eigenvectors of L with smallest eigenvalues.

Relaxed graph min cut

In relaxed problem, columns of $H \iff$ features. Rows of $H \iff$ data points.

Can use features for a variety of applications.

Our code: K-means on feature vectors $\rightarrow$ classification. (Spectral Clustering)

Nyström Extension

As $|X|$ becomes large, computing the $|X| \times |X|$ weight matrix W becomes prohibitive.

Instead choose $A \subseteq X$ *landmark nodes* with $|A| \ll |X|$. Up to permutation, we have

$$W = \begin{pmatrix} W_{A,A} & W_{A,A^c} \\ W_{A^c,A} & W_{A^c,A^c} \end{pmatrix}.$$

Nyström Extension

Nyström: Approximate Graph Laplacian eigenvectors using only $W_{A,A}$, $W_{A^c,A}$.

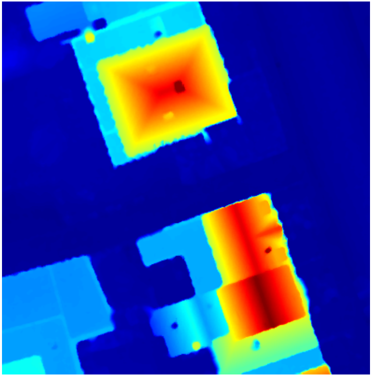
$$W \approx \begin{pmatrix} W_{A,A} \\ W_{A^c,A} \end{pmatrix} W_{AA}^{-1} \begin{pmatrix} W_{A,A} & W_{A,A^c} \end{pmatrix}.$$

Compute and store matrices of size at most $|X| \times |A|$.

Data

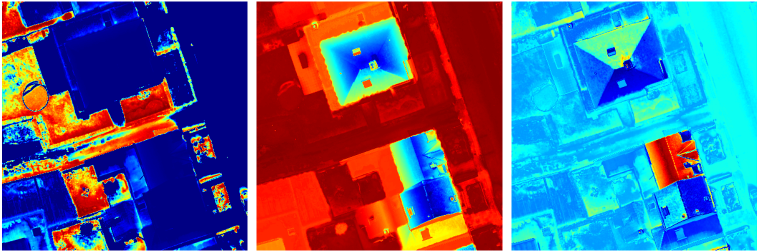
Our algorithm applied to [Bampos-Taberner et al, 2016] dataset.

RBG Modality



Lidar Modality

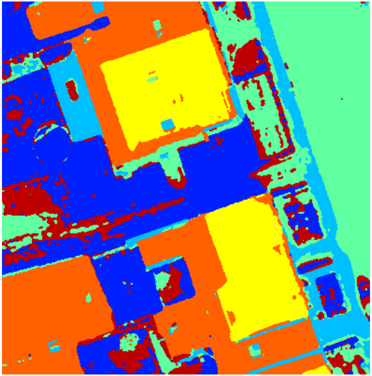
Results



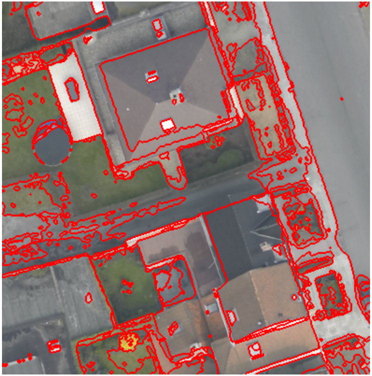
Example eigenvectors of GL

Results

Spectral Clustering result (unsupervised). $m = 6$ classes.



Classes

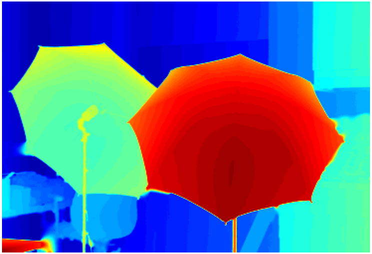


Regions on original image

Data

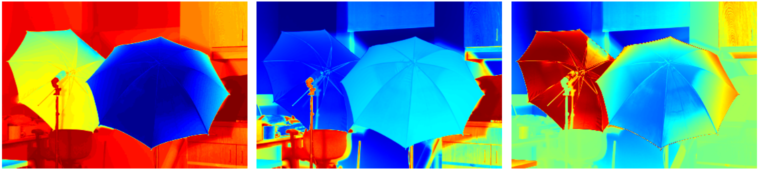
Our algorithm applied to [Scharstein et al. 2014] dataset.

RBG Modality



Lidar Modality

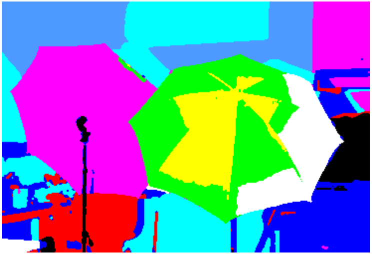
Results



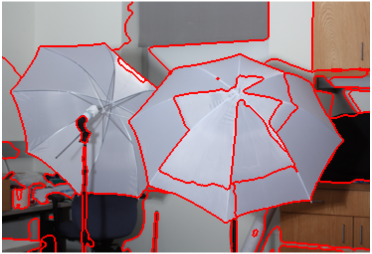
Example eigenvectors of GL

Results

Spectral Clustering result (unsupervised). $m = 8$ classes.



Classes



Regions on original image

Future Work

Remove or weaken the coregistration assumption.

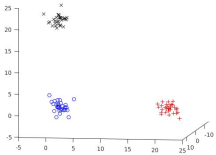
Semisupervised Method: Partial coregistration and/or class labeling.

Used in [Wang et al, 2013, Tuia et al, 2016], among others.

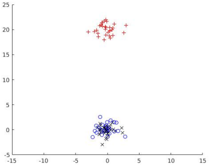
Future Work

Remove or weaken the coregistration assumption.

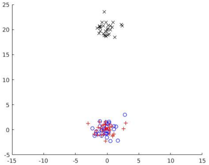
Information theory: minimal set of assumptions.



Ground Truth



Modality 1

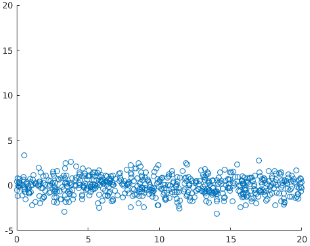


Modality 2

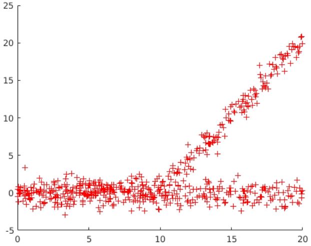
Future Work

Remove or weaken the coregistration assumption.

Work with geometry/topology of manifolds.



Modality 1



Modality 2

References I

M. Song and D. Tao and C. Chen and J. Bu and J. Luo and C. Zhang (2012)

Probabilistic Exposure Fusion

IEEE Transactions on Image Processing 21(1), 341-357

Correa, N. M. and Eichele, T. and Adali, T. and Li, Y. O. and Calhoun, V. D. (2010)

Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI

Neuroimage 50(4) 1438-1445

Sedighin, Farnaz and Babaie-Zadeh, Massoud and Rivet, Bertrand and Jutten, Christian (2016)

Two Multimodal Approaches for Single Microphone Source Separation

24th European Signal Processing Conference 110-114

References II

- Lahat, Dana and Adalı, Tülay and Jutten, Christian (2015)
  Multimodal Data Fusion: An Overview of Methods, Challenges and Prospects
  Proceedings of the IEEE 103(9), 1449-1477

- Yi-Ren Yes and Chun-Hao Huang and Yu-Chiang Frank Wang
  Heterogeneous Domain Adaptation and Classification by Exploiting the Correlation Subspace
  IEEE Transactions on Image Processing 23(5), 2009-2018

- Wang, Chang and Mahadevan, Sridhar (2013)
  Manifold Alignment Preserving Global Geometry
  Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence IJCAI '13, 1743-1749

- Tuia, Devis AND Camps-Valls, Gustau (2016)
  Kernal Manifold Alignment for Domain Adaptation
  PLOS ONE 11, 1-25

References III

M. Campos-Taberner and A. Romero-Soriano and C. Gatta and G. Camps-Valls and A. Lagrange and B. Le Saux and A. Beaupre and A. Boulch and A. Chan-Hon-Tong and S. Herbin and H. Randrianarivo and M. Ferecatu and M. Shimoni and G. Moser and D. Tuia (2015)
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), 5547-5559

Daniel Scharstein and Heiko Hirschmüller and York Kitajima and Greg Krathwohl and Nera Nešić and Xi Wang and Porter Westling (2014)
High-resolution stereo datasets with subpixel-accurate ground truth
Proceedings of the 36th German Conference on Pattern Recognition 31-42