Graph-Based Multimodal Data Processing

Geoffrey Iyer

University of California, Los Angeles

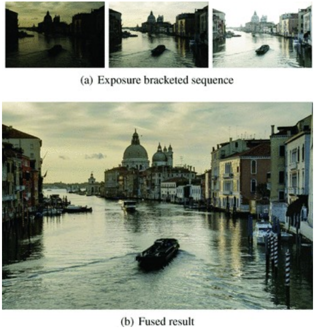
gsiyer@math.ucla.edu

November 9th, 2017

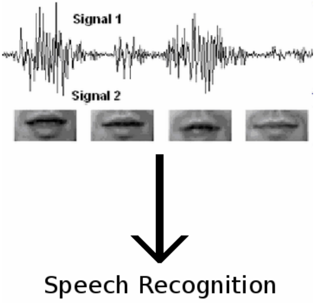
Overview

Multimodal Datasets

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



Exposure Fusion:
[Mertens et al, CGF, 2008]

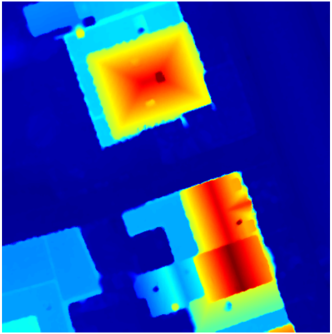


Speech Recognition:
[Datcu et al, IEEE CVPR 2007]

Example Multimodal Data

Remote sensing example: RGB + Elevation map.
From 2015 IEEE Data Fusion Contest.
[Bampos-Taberner et al, IEEE J-STARS 2016]

RGB Data



Lidar Data

Challenges in Multimodality

Most multimodal methods are developed specifically for one problem, BUT:

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

Our approach:
Represent data via a graph, compare graph representations.

Graph Representation: Background

For a dataset X with N elements

For each pair $x_i, x_j \in X$, define a *weight* $w_{ij} \geq 0$ 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

$$w_{ij} = \exp\left(-\left\|x_i - x_j\right\|^2 / \sigma\right).$$

Get a representation of X as a weighted graph (V, E).

Graph Laplacian

Perform analysis via the *Graph Laplacian* [Chung, AMS, 1997]

$$L = D - W.$$

$D =$ diagonal degree matrix, $d_{ii} = \sum_{j=1}^{N} w_{ij}$.

View L as an operator on the space of functions $f : V \to \mathbb{R}$.
Related to the standard differential Laplacian $\Delta = \nabla^2$.

A manifold $\mathcal{M}$ can be discretized and represented by a graph.

Under mild assumptions and some renormalization,
L converges to Δ as mesh size $\to 0$.

Graph Laplacian vs Differential Laplacian

Ex: A 1-D line graph. L is related to the discrete Laplacian.

$$L = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \\ \cdots & 2 & -1 & 0 & 0 & \cdots \\ \cdots & -1 & 2 & -1 & 0 & \cdots \\ \cdots & 0 & -1 & 2 & -1 & \cdots \\ \cdots & 0 & 0 & -1 & 2 & \cdots \\ & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

1-D discrete Laplacian with stepsize h:

$$\Delta f(x) \approx \frac{-f(x-h) + 2 \cdot f(x) - f(x+h)}{h^2}.$$

Laplacian Eigenfunctions and Fourier Modes

Eigenfunctions of differential Laplacian give modes of vibration.
Corresponding eigenvalues describe the energy of the state.

Ex: On the interval $[0, 1]$, eigenfunctions are solutions to

$$\Delta f = \frac{d^2 f}{dx^2} = -\lambda f.$$

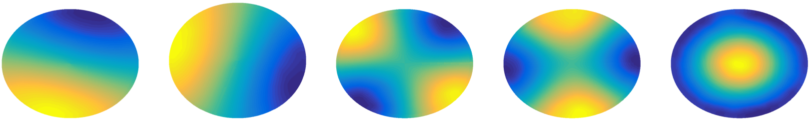
These are exactly the normal Fourier modes

$$\{\sin(k\pi x), \cos(\ell\pi x) \,:\, k \geq 1, \ell \geq 0\}.$$

Eigenvectors of Graph Laplacian give a discrete version.
Allow us to study symmetries and structure of the graph.

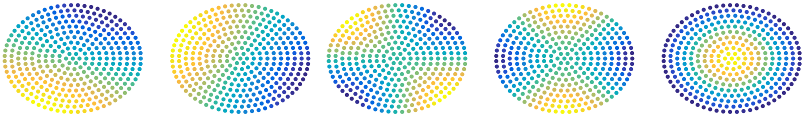
Laplacian Eigenfunctions and Fourier Modes

Example: A disc in $\mathbb{R}^2$, and a graph representation.



v_2 v_3 v_4 v_5 v_6

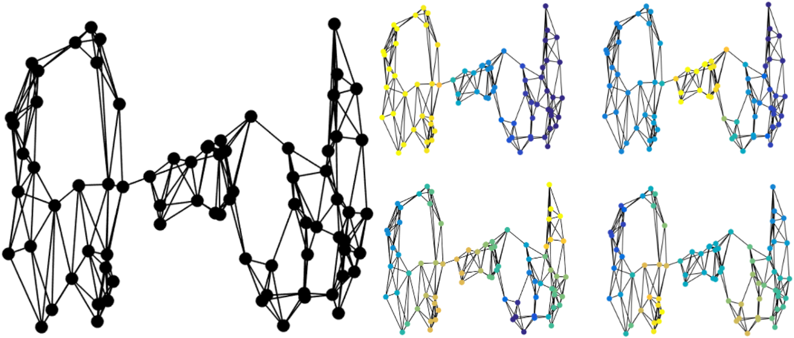
Differential Laplacian Eigenfunctions



v_2 v_3 v_4 v_5 v_6

Graph Laplacian Eigenvectors

Laplacian Eigenvectors and Fourier Modes



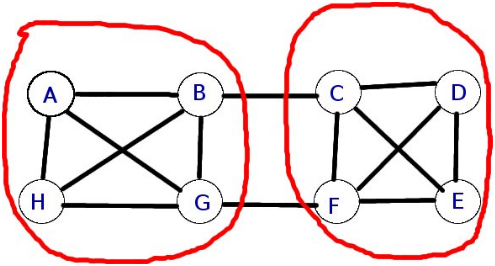
Synthetic Graph and Example Eigenvectors

Graph Min-Cut Problem

Graph Laplacian can also be understood via graph-cuts.
Partition V into k sets $A_1, \ldots, A_k$ minimizing graph-cut energy.

$$\text{Cut}(A_1, \ldots, A_k) = \frac{1}{2} \sum_{j=1}^{k} W(A_j, A_j^c).$$

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



Graph Laplacian and Min-Cut

Notation: $u = |V| \times k$ indicator matrix for $A_1, \dots, A_k$.

$$u_{ij} = \begin{cases} 1 & \text{if node } i \text{ is in set } A_j \\ 0 & \text{else.} \end{cases}$$

Then

$$Cut(A_1, \dots, A_k) = \operatorname{Tr}\left(u^T L u\right).$$

So we rephrase our min-cut problem as:

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

Laplacian Eigenvectors and Min-Cut

Exact min-cut solution is computationally infeasible.

One popular relaxation: find

$$\mathrm{argmin}_{u^T u = I} \mathrm{Tr}\left(u^T L u\right).$$

Solution: columns of u are eigenvectors of L corresponding to smallest eigenvalues.

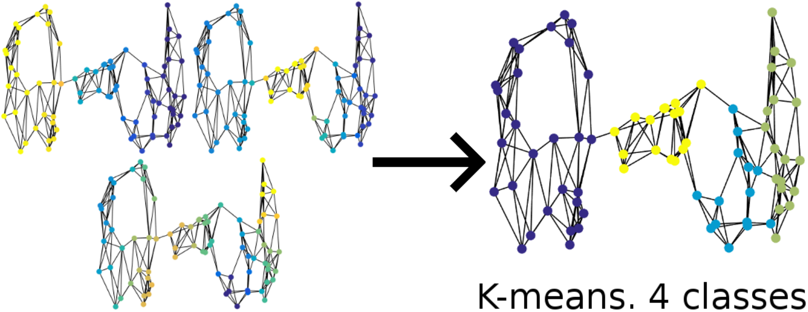
Heuristic [vonLuxburg, Stat Comput 2007]:

$$\text{eigenvectors of } L \iff \text{features extracted from data}$$

Laplacian Eigenvectors and Min-Cut

Can use eigenvectors for a variety of applications.

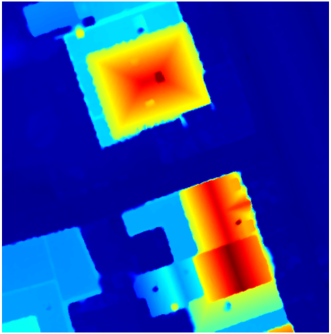
Ex: K-means on eigenvectors $\rightarrow$ segmentation:
(called Spectral Clustering [vonLuxburg, Stat Comput 2007])



Problem Setup

Perform segmentation on *coregistered*, multimodal datasets.
[Iyer et al., ICIP 2017]

RGB Data



Lidar Data

Example Multimodal Data

Problem Setup

From each modality, have a data set X^k. $N = \left|X^1\right| = \cdots = \left|X^k\right|$.

From co-registration assumption:
i-th point in X^{k_1} corresponds to i-th point in X^{k_2}.

x_i^k = element i from X^k.

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.
(often $\|\cdot\|$ is the L^2-norm)

Multimodal Weight Matrix

Scale each matrix by standard deviation (nondimensionalization)

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

Define

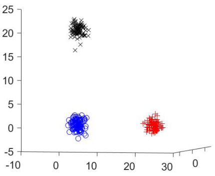
$$w_{ij} = \exp\left(-\max\left(\bar{E}^1_{ij}, \ldots, \bar{E}^k_{ij}\right)/\sigma\right).$$

Using this weight matrix and Graph Laplacian theory, perform *Spectral Clustering* and *Semisupervised Graph MBO* [Merkurjev, SIIMS 2013] to segment our datasets.

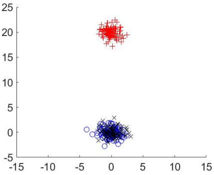
Synthetic Example: Data

Why use max?

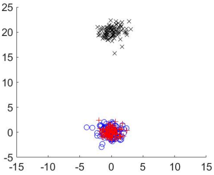
Ground truth = 3 point clouds in $\mathbb{R}^3$ (100 points per cloud).
Modality 1 = projection onto xy-plane.
Modality 2 = projection onto xz-plane.
Co-registration assumption: index is input to algorithm.



Underlying Data



Modality 1

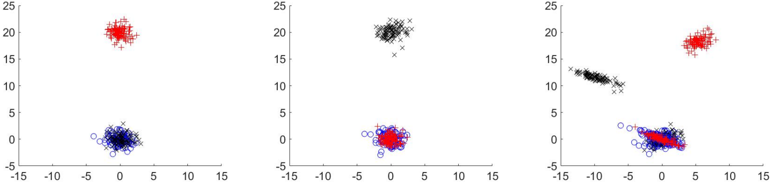


Modality 2

Synthetic Example: Result of CCA

Result of CCA algorithm from [Yeh et al, IEEE TIP 2014]:

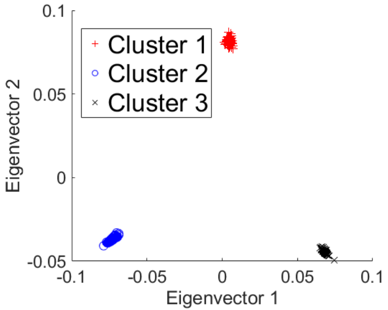
Both modalities are mapped to a common *latent space*.
(classification can then be done in this space.)



Modality 1 Modality 2 Image in latent space

Synthetic Example: Result of Our Method

Result of our method:
First 2 Laplacian Eigenvectors, plotted in $\mathbb{R}^2$.



2 eigenvectors of graph Laplacian

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

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

[Fowlkes et al., IEEE TPAMI 2004]:
Approximate graph Laplacian eigenvectors using only $W_{A,A}$, $W_{A^c,A}$.

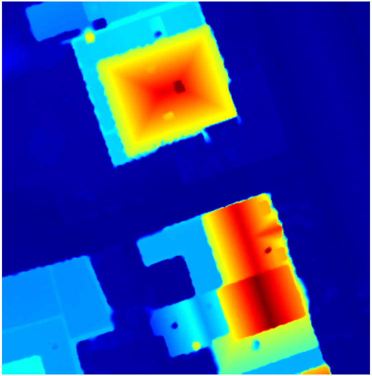
$$W \approx \begin{pmatrix} W_{A,A} \\ W_{A^c,A} \end{pmatrix} W_{AA}^{-1} \begin{pmatrix} W_{A,A} & W_{A,A^c} \end{pmatrix}.$$

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

DFC2015 Data

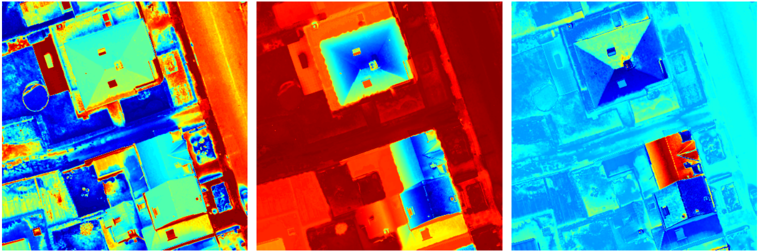
Our algorithm applied to DFC 2015
[Bampos-Taberner et al, IEEE J-STARS 2016].

RGB Modality



Lidar Modality

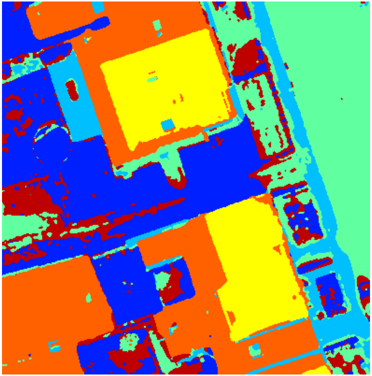
Results



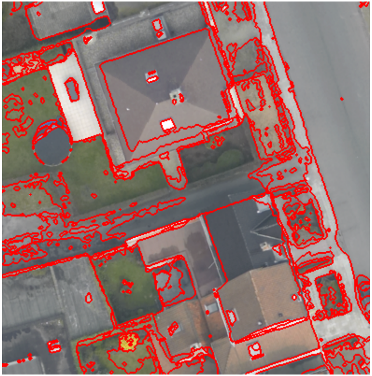
Example eigenvectors of graph Laplacian

Results

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



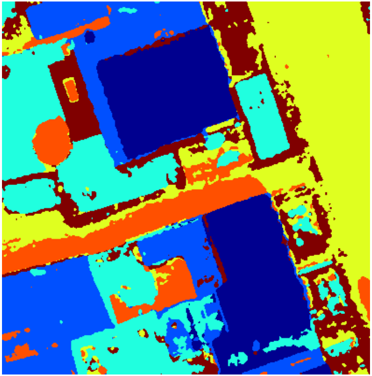
Classes



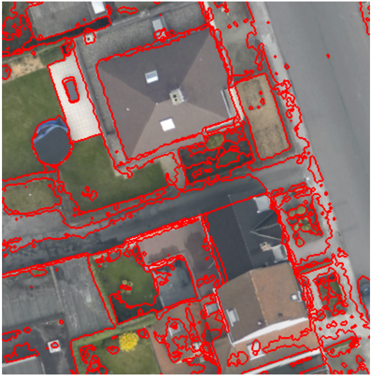
Regions on original image

Results

Graph MBO (7% supervised). $m = 6$ classes.



Classes

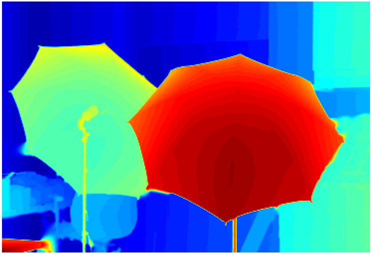


Regions on original image

Umbrella Data

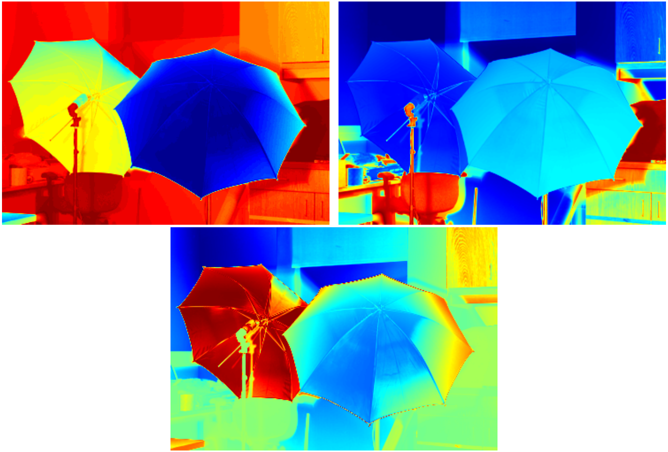
Our algorithm applied to [Scharstein et al., GCPR 2014] dataset.

RGB Modality



Lidar Modality

Results

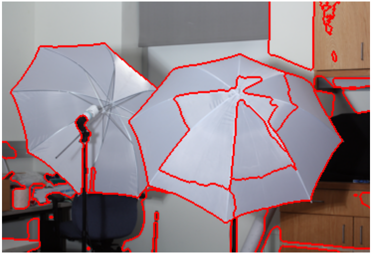


Example eigenvectors of graph Laplacian

Results

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

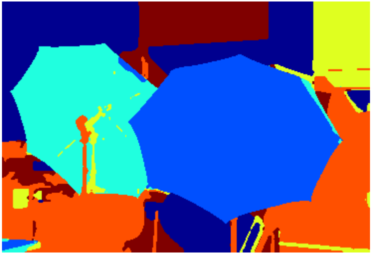
Classes



Regions on original image

Results

Graph MBO result (5% supervised). $m = 6$ classes.



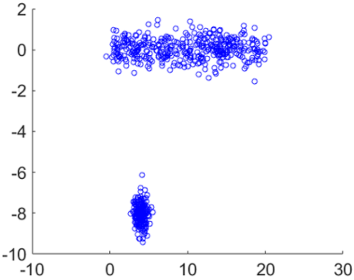
Classes

Regions on original image

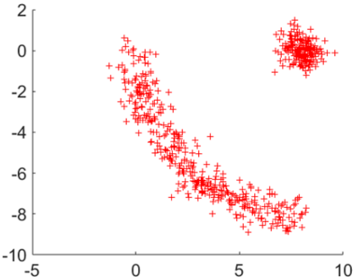
Graph Matching

Goal: Remove or weaken the coregistration assumption.
Create our own registration using topological qualities of data.

View each dataset as a (weighted) graph.
Try to match nodes with similar structure.



$\approx$

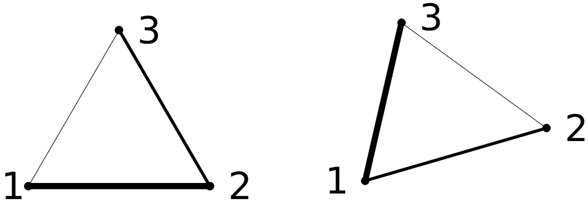


Problem Setup

Two weighted graphs, G_1, G_2, with weight matrices W_1, W_2.

For convenience, assume $|G_1| = |G_2| = N$.

Search for a graph isomorphism $G_1 \to G_2$ preserving edge weights.



Best isomorphism is $1 \to 3$, $2 \to 1$, $3 \to 2$.

Problem Setup

Isomorphism $G_1 \to G_2$ corresponds to a permutation on nodes. Have P the corresponding permutation matrix. Want to find

$$\text{argmin}_{P \text{ a Permutation matrix}} \left\| W_1 - P W_2 P^T \right\|_F^2 .$$

Exact solution is too expensive. Approximate via Graph Laplacian. [Umeyama, IEEE TPAMI 1988, Knossow et al., GbRPR 2009].

Relaxation

Relax problem to

$$Q^* = \operatorname{argmin}_{QQ^T=I} \left\| W_1 - QW_2Q^T \right\|_F^2 .$$

Let L_1, L_2 the graph Laplacians corresponding to W_1, W_2

U_1, U_2 the corresponding matrices of eigenvectors.

Then $Q^* = U_1 S U_2^T$.

S is diagonal matrix with entries ± 1 to handle sign ambiguity.

Heuristics

Recall from Laplacian Theory

$$\begin{aligned} \text{column of } U_i &\iff \text{feature extracted from data} \\ \text{row of } U_i &\iff \text{image of data point in new feature space.} \end{aligned}$$

Match rows of U_1 to rows of U_2 by considering $U_1 U_2^T$.

Currently, S is determined in a semisupervised manner, align eigenvectors using a few ($< 1\%$) known matches.

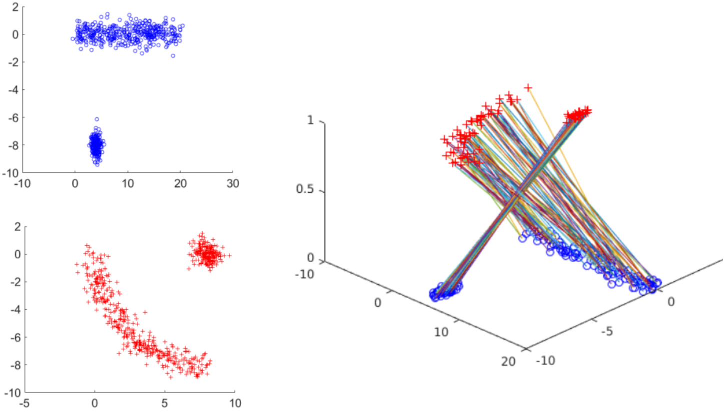
Matching Algorithm

Q^*_{ij} gives the similarity between node i of G_1 and node j of G_2.

Many choices for how to complete the matching:

- Match 1-to-many via maximum similarity.
- Match 1-to-1 via *Hungarian Algorithm*.
- Match many-to-many via a threshold on Q^*.
- **Hierarchical Matching:**
  Match some landmark nodes, extend to a full match.
  Similar to Nystöm, avoid creating the full $N \times N$ graph.

Example Matching

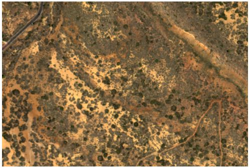


Graph Match on Synthetic Data

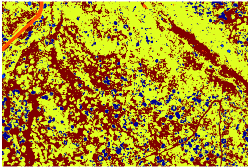
Knowledge Transfer

Knowledge transfer via Graph Matching.

- 2 hyperspectral images. 68 bands. (ARTEMISAT-2 Project)
- Area 1 is segmented into 6 classes. (Given with data)
- Match Area 1 to Area 2. Transfer classification via match.



Area 1 (RGB bands)

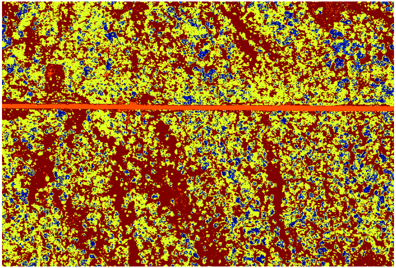


Area 1 Classification

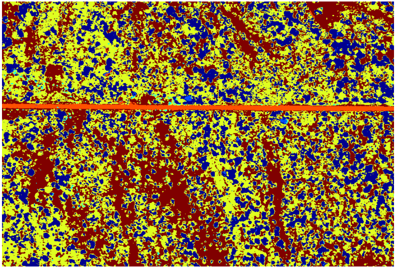
Area 2 (RGB bands)

Knowledge Transfer

Correctly classify 72% of pixels.



Transfer to Area 2



Area 2 Ground Truth

Change Detection

Change detection via Graph Matching:

Given *coregistered* data X and Y, compare the match via registration against the match via graphs.

From graph matching, get a permutation

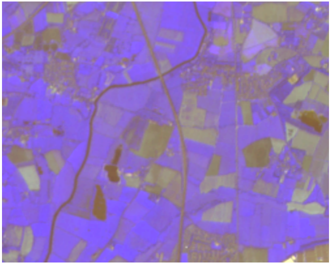
$$\rho : \{1, \ldots, n\} \to \{1, \ldots, n\}.$$

Look at $\left\| x_i - x_{\rho^{-1}(i)} \right\|$ and $\left\| y_i - y_{\rho(i)} \right\|$.

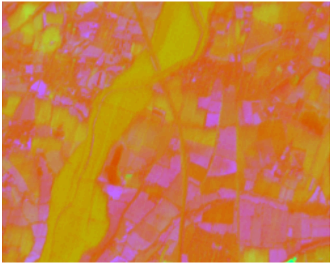
Change Detection Example

Flood data: Data Fusion Contest 2010
[Longbotham et al, IEEE,2012].

Satellite images before and after flood.
3 bands, near-infrared (displayed in false color).

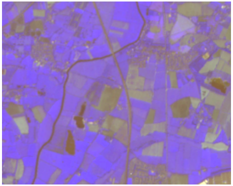


Before Flood

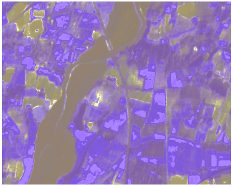


After Flood

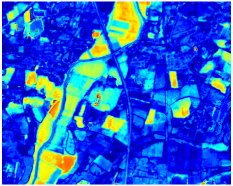
Change Detection Example



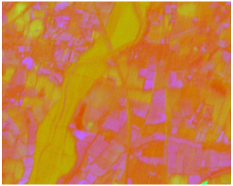
Data X



Approx Y by X

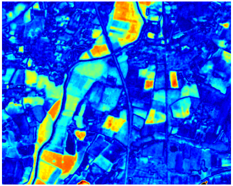


$\|X - \text{perm}(X)\|$



Data Y

Approx X by Y

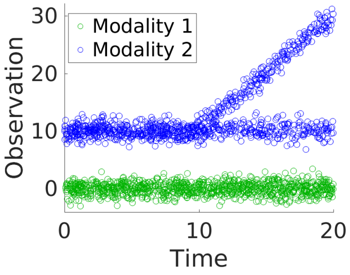


$\|Y - \text{perm}(Y)\|$

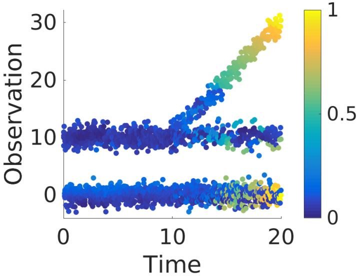
Change Detection

Change detection to find unique information in different modalities.

Synthetic Example: 1-D observations over time.



Input Data

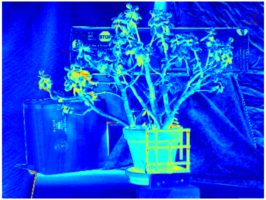


Norm Comparison (scaled)

Change detection to find unique information in different modalities.

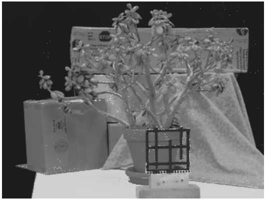
Data X

Approx Y by X



$\|X - \mathrm{perm}(X)\|$

Data Y



Approx X by Y

$\|Y - \mathrm{perm}(Y)\|$

References I

Longbotham, Nathan and Pacifici, Fabio and Glenn, Taylor and Zare, Alina and Volpi, Michele and Tuia, Devis and Christophe, Emmanuel and Michel, Julien and Inglada, Jordi and Chanussot, Jocelyn and Du, Qian
Multi-modal change detection, application to the detection of flooded areas: outcome of the 2009-2010 data fusion contest
IEEE J. Sel. Topics Appl. Earth Observ. 5(1):331-342

Fan R. K. Chung
Spectral Graph Theory
CBMS Reg. Conf. Ser. Math. 92, Providence, RI, 1997

Tom Mertens and J. Kautz and Frank Van Reeth
Exposure Fusion: A Simple and Practical Alternative to High Dynamic Range Photography
Computer Graphics Forum 28(1):161 - 171

References II

D. Datcu and Z. Yang and L. Rothkrantz (2007)
Multimodal workbench for automatic surveillance applications
2007 IEEE Conference on Computer Vision and Pattern Recognition 1-2

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

References III

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

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

Iyer, Geoffrey and Chanussot, Jocelyn and Bertozzi, Andrea (2017).
A Graph-Based Approach for Feature Extraction and Segmentation of Multimodal Images
Preprint

References IV

- 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

- Umeyama, S. (1988)
  An Eigendecomposition Approach to Weighted Graph Matching Problems
  IEEE Trans. Pattern Anal. Mach. Intell., 10(5), 695-703

- Knossow, David and Sharma, Avinash and Mateus, Diana and Horaud, Radu (2009)
  Inexact Matching of Large and Sparse Graphs Using Laplacian Eigenvectors
  Graph-Based Representations in Pattern Recognition: 7th IAPR-TC-15 International Workshop, 2009. Proceedings

References V

vonLuxburg, Ulrike (2007)
A tutorial on spectral clustering
Statistics and Computing, 17(4), 395-416

Ekaterina Merkurjev and Tijana Kostic and Andrea L Bertozzi (2013)
An MBO scheme on graphs for classification and image processing
SIAM Journal on Imaging Sciences, 6(4), 1903-1930

Charless Fowlkes and Serge Belongie and Fan Chung and Jitendra Malik (2004)
Spectral Grouping Using the Nystrom Method
IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(2)

J. Munkres (1957)
Algorithms for the Assignment and Transportation Problems
Journal of the Society for Industrial and Applied Mathematics, 5(1):3238

Hungarian Algorithm

Say we have $N = 6$ and calculated:

$$Q^* = \begin{pmatrix} -0.1629 & -0.1711 & -0.1703 & 0.3426 & 0.3717 & -0.2100 \\ -0.1647 & -0.1662 & -0.1677 & 0.2966 & 0.3192 & -0.1172 \\ -0.1660 & -0.1653 & -0.1657 & -0.1477 & -0.1861 & 0.8308 \\ -0.4579 & 0.6860 & 0.2665 & -0.1787 & -0.1480 & -0.1678 \\ 0.4939 & -0.1039 & 0.1196 & -0.6689 & 0.3080 & -0.1486 \\ 0.4577 & -0.0795 & 0.1176 & 0.3561 & -0.6647 & -0.1872 \end{pmatrix}$$

Then

$$P^* = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Graph Matching Example

Graph 1 Graph 2

Any reasonable matching sends $1 \to 2$.

Other nodes can be matched in many ways (symmetry).

Change Detection Example 2

Synthetic Example: Graph representation is useful!

Continuous transform ($\mathbb{R}^3 \to \mathbb{R}^3$) applied to most pixels,
One extra artifact added.
Simulate data captured from different sources.

Original Picture

Altered Picture

Norm Comparison

Change Detection Example

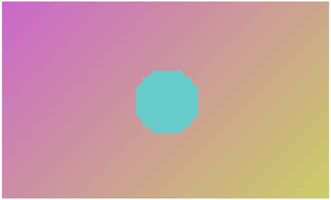


Image X

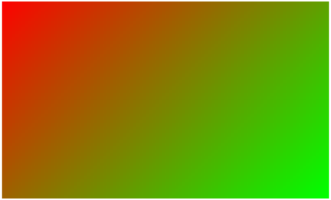
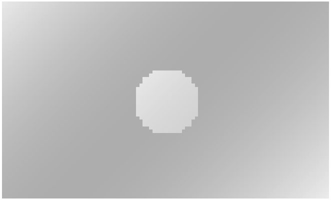
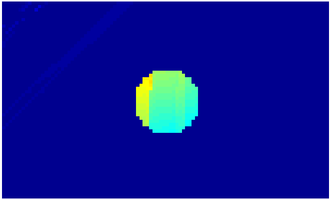


Image Y



Naive difference $\|X - Y\|$



$\left\| x_i - x_{\rho(i)} \right\|$