Multimodal Data Processing

Geoffrey Iyer

University of California

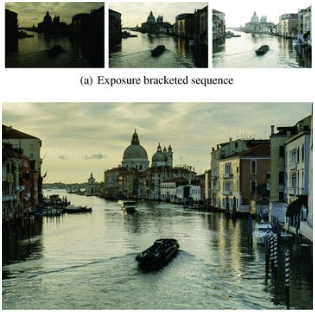
gsiyer@math.ucla.edu

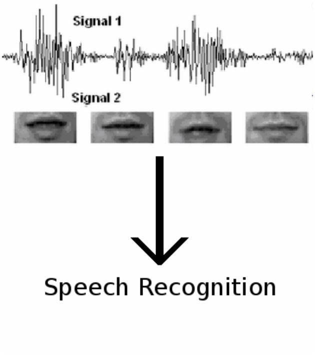
June 13th, 2017

Overview

Multimodal datasets

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


(a) Exposure bracketed sequence
(b) Fused result



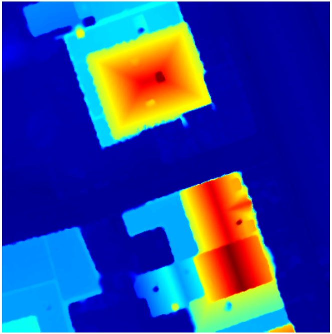
Exposure Fusion:
[Mertens et al, CGF, 2008]

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

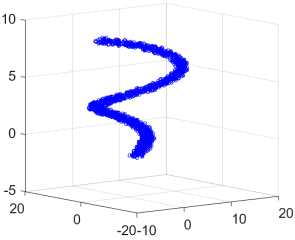
Most multimodal methods are developed specifically for one problem, BUT:

[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."

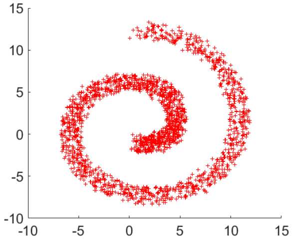
Manifold alignment

Attempt to address multimodality in general.

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

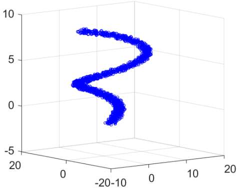


X^1

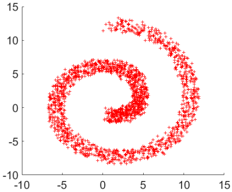


X^2

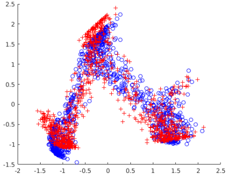
Create a *latent space* Y and maps $X^i \to Y$.



X^1



X^2



Images in latent space

Example from [Tuia et al, PLOS ONE 2016]

Compare sets by using the latent space image.

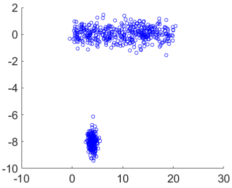
Manifold alignment: Methods from the literature

Some examples from the literature:

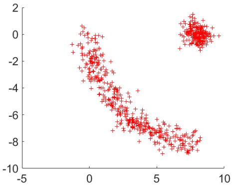
- [Yeh et al, IEEE TIP 2014]: Canonical Correlation Analysis, linear or with nonlinear kernel (unsupervised)
- [Wang et al, IJCAI 2013]: Graph-based methods (semi-supervised)
- [Tuia et al, PLOS ONE 2016]: Similar to the above with an added nonlinear kernel (semi-supervised)

Manifold alignment: Methods from the literature

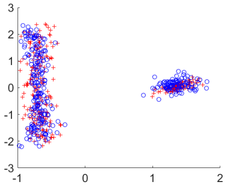
Common theme: Create the latent space by finding and correlating *redundancies* between sets.



X^1



X^2



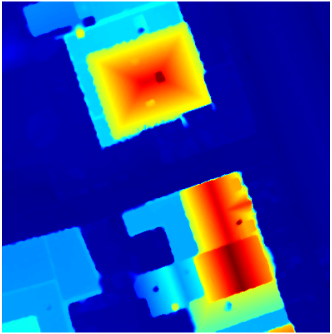
Images in latent space

Using code from [Tuia et al, PLOS ONE 2016]

Manifold alignment: Our goal

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

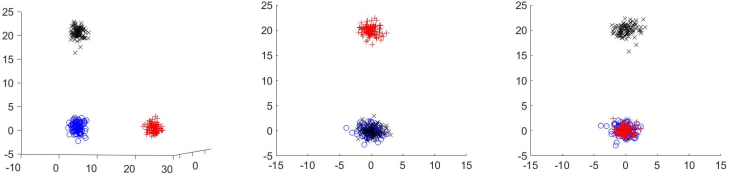
Distinguish road from grass



Distinguish roof from ground

Synthetic example: Data

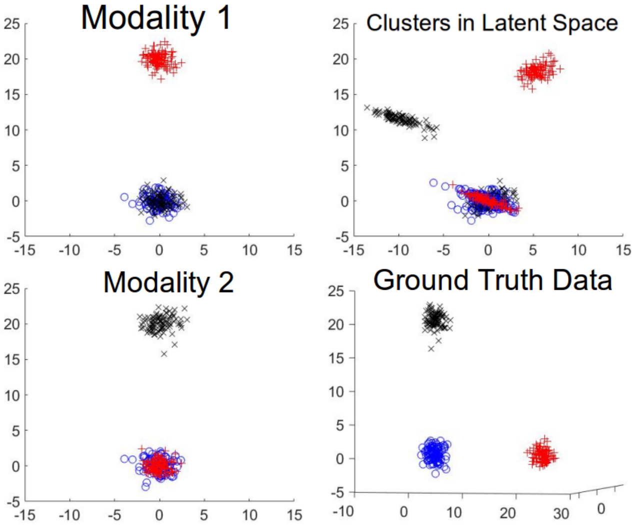
Ground truth = 3 point clouds in $\mathbb{R}^3$ (100 points per cloud).
Modality 1 = projection onto *xy*-plane.
Modality 2 = projection onto *xz*-plane.
Co-registration assumption: index is input to algorithm.



Underlying Data Modality 1 Modality 2

Synthetic Example: Result of CCA

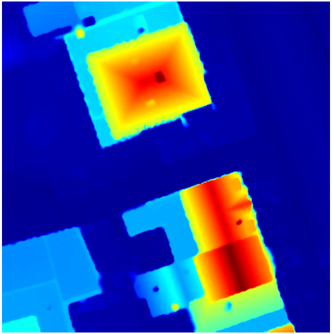
Result of CCA algorithm from [Yeh et al, IEEE TIP 2014]:



Problem setup

We use co-registration assumption and graph Laplacian theory for segmentation of multimodal datasets.
[Iyer et al., Accepted Paper, ICIP 2017]

RGB Data



Lidar Data

Notation

From each modality, have a data set X^k. $\ell =$ 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 $X = (X^1, X^2, \ldots, X^\ell) \subseteq \mathbb{R}^{N \times (d_1 + \cdots + d_\ell)}$.

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

Graph Representation: Background

When using a single modality:

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

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

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 $\|\cdot\|$ is the 2-norm)

Scale each matrix by standard deviation (nondimensionalization)

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

Multimodal Weight Matrix

Define

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

Heuristics:

- Standard deviation 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

Use weights W to segment X into $A_1, \dots, A_k$. We want to

- group nodes with high similarity (weight) together
- ensure each set is a reasonable size

Use the *Normalized graph-cut*

$$\mathrm{Ncut}(A_1, \dots, A_k) = \frac{1}{2} \sum_{j=1}^{k} \frac{W(A_j, A_j^c)}{vol(A_j)}.$$

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

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

Exact min-cut solution is computationally infeasible.

Graph Laplacian

A well-known approximation for graph min-cut:
eigenvectors of the graph Laplacian L.

$$L = I - D^{-1/2} W D^{-1/2},$$

where $D = N \times N$ diagonal matrix (degree matrix), with

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

Feature extraction

Result of [vonLuxburg, Stat Comput 2007]

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

Can use eigenvectors for a variety of applications.

Our results: K-means on eigenvectors $\rightarrow$ segmentation
(this is called Spectral Clustering).

Future work: Apply graph MBO [Merkurjev, SIIMS 2013] instead.

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

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

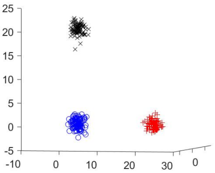
[Fowlkes et al., IEEE TPAMI 2004]:
Approximate graph Laplacian eigenvectors using only $W_{A,A}$, $W_{A^c,A}$.

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

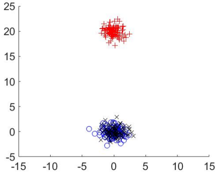
Compute and store matrices of size at most $|X| \times |A|$.

Synthetic example: Data

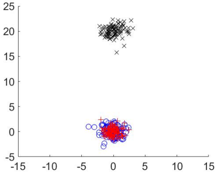
Synthetic example:
Ground truth = 3 point clouds in $\mathbb{R}^3$ (100 points per cloud).
Modality 1 = projection onto xy-plane.
Modality 2 = projection onto xz-plane.



Ground Truth



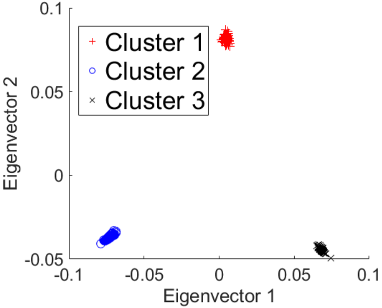
Modality 1



Modality 2

Synthetic Example: Result of Our Method

Result of our multimodal graph-based algorithm on earlier example: Plotting first 2 eigenvectors.

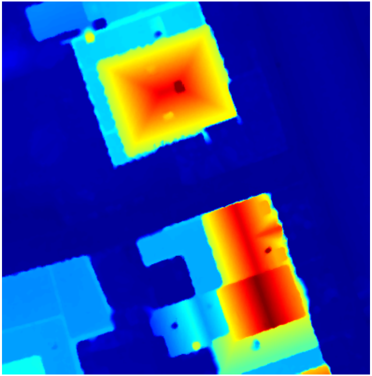


2 eigenvectors of graph Laplacian

Data

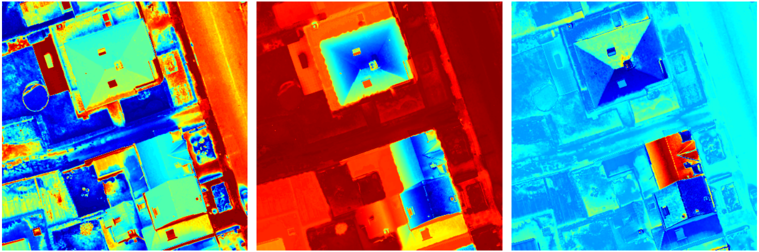
Our algorithm applied to
[Bampos-Taberner et al, IEEE J-STARS 2016] dataset.

RGB Modality



Lidar Modality

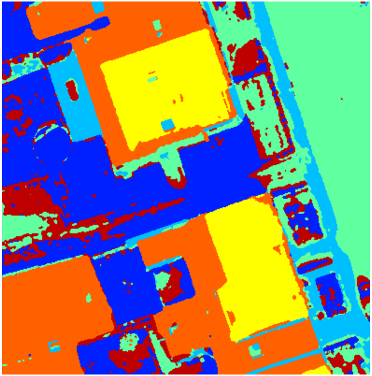
Results



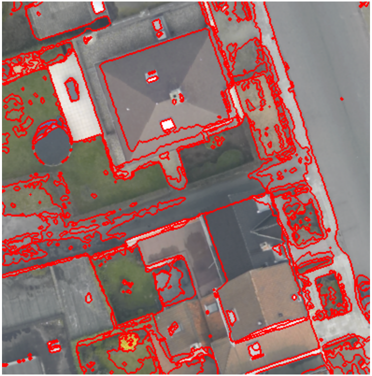
Example eigenvectors of graph Laplacian

Results

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



Classes

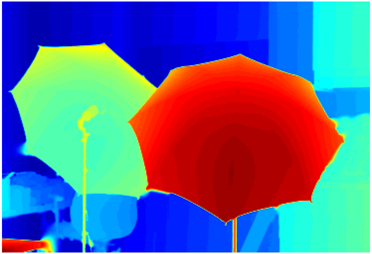


Regions on original image

Data

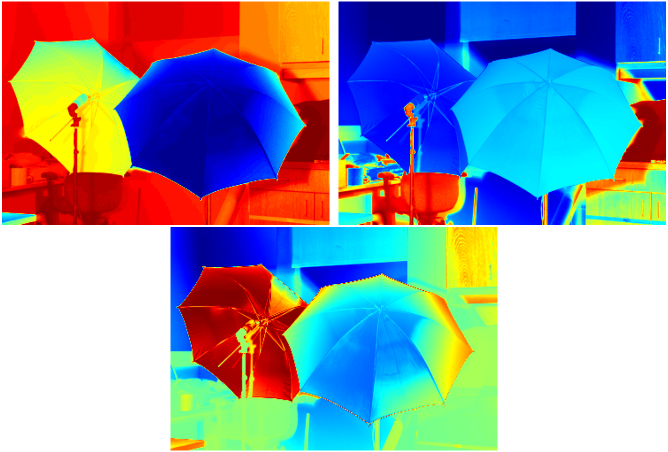
Our algorithm applied to [Scharstein et al., GCPR 2014] dataset.

RGB Modality



Lidar Modality

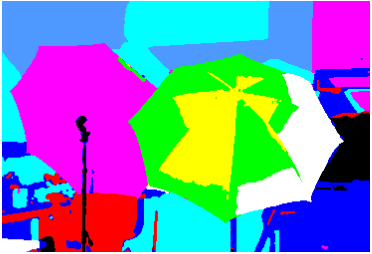
Results



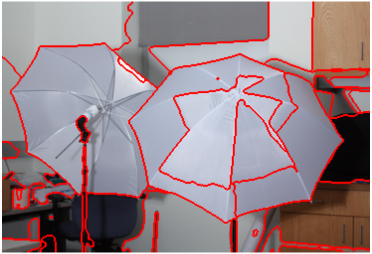
Example eigenvectors of graph Laplacian

Results

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



Classes



Regions on original image

Graph Matching

Goal: Remove or weaken the coregistration assumption.

Current idea: Graph matching.

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

Problem Setup

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

For now, $|G_1| = |G_2| = N$.

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

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

Problem Setup

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

$$\left\| W_1 - P W_2 P^T \right\|_F^2 .$$

Exact solution is too expensive. Can solve using graph Laplacian trick from
[Umeyama, IEEE TPAMI 1988, Knossow et al., GbRPR 2009].

Relaxation

Relax problem to

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

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 a diagonal matrix with entries of ± 1 to account for sign ambiguity in eigenvectors.

Heuristics

Recall from graph Laplacian

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

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

Matching Algorithm

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

Choose a permutation $p : \{1, 2, \ldots, N\} \to \{1, 2, \ldots, N\}$ via

$$\text{argmax}_{\text{permutations } p} \sum_{i=1}^{N} Q^*_{i,p(i)}.$$

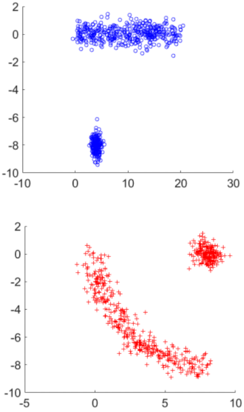
Hungarian algorithm finds this in $O(N^3)$ [Munkres, SIAM 1957].

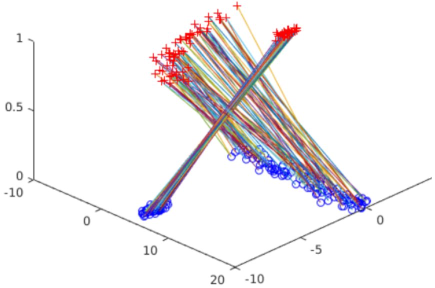
Benefits of Graph Matching

Benefits of Graph Matching

1. Invariant under conformal maps.
   - scaling, shifts, rotations, etc.
   - robust to continuous deformation.
2. A precise number representing similarity between nodes gives us many options.
   - Thresholding
   - Hierarchical matching
3. Easy extension to the case $|G_1| \neq |G_2|$.

Example Matching





Graph Match on Synthetic Data

Change Detection

One possible application: Change detection.

Given images X and Y of the same scene, compare coregistration against results of graph matching. Use this to pick out large changes between X, Y.

From graph matching, get a permutation

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

Compare x_i to $x_{\rho(i)}$, and y_i to $y_{\rho(i)}$.

Change Detection Example

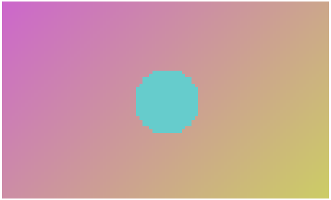


Image X

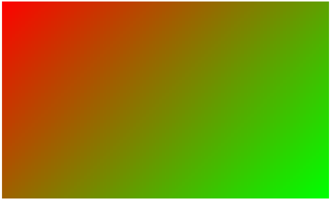
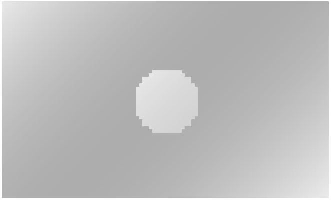
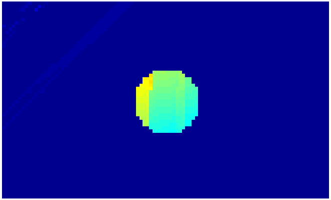


Image Y



Naive difference $\|X - Y\|$



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

Future directions

Possible directions for future work

1. Improve image segmentation using graph MBO.
2. Change detection using graph matching.
3. Push performance of graph matching.

References I

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

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

References II

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

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

References III

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

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

References IV

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

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

References V

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