Multimodal Data Processing

Geoffrey Iyer

University of California

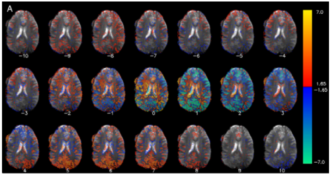
gsiyer@math.ucla.edu

June 5, 2017

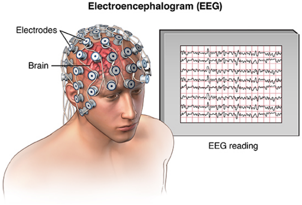
Overview

Multimodal datasets

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



FMRI



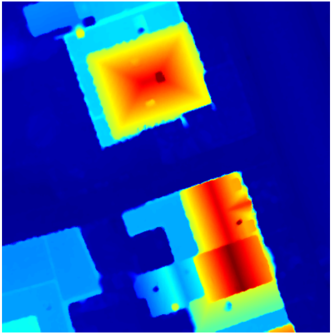
EEG

How to solve machine learning problems on multimodal data?

Example Multimodal Data

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

RGB Data



Lidar Data

Examples from the literature

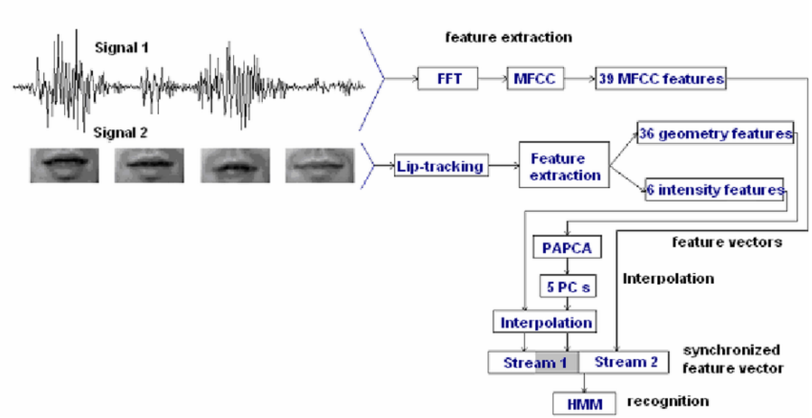
Exposure Fusion, from [Mertens et al, 2008].

(a) Exposure bracketed sequence

(b) Fused result

Examples from the literature

Audio-Visual speech recognition, from [Datcu et al, 2007].



Challenges in multimodality

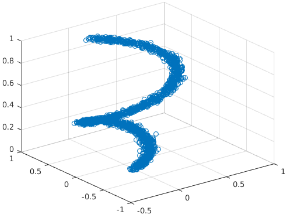
Most multimodal methods are developed specifically for one problem, BUT:

[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."

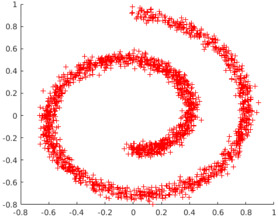
Manifold alignment

Attempt to address multimodality in general via manifold alignment.

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

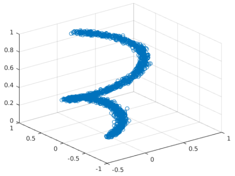


X^1

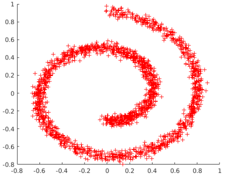


X^2

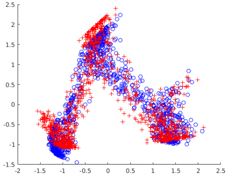
Create a *latent space* Y and maps $X^i \to Y$.



X^1



X^2



Images in latent space

Example from [Tuia et al, 2016]

Compare sets by using the latent space image.

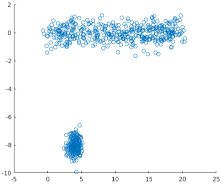
Manifold alignment: Methods from the literature

Some examples from the literature:

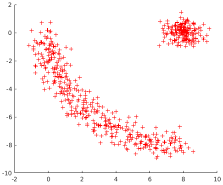
- [Yeh et al, 2014]: Canonical Correlation Analysis, linear or with nonlinear kernel (unsupervised)
- [Wang et al, 2013]: Graph-based methods (semi-supervised)
- [Tuia et al, 2016]: Similar to [Wang et al, 2013] with an added nonlinear kernel (semi-supervised)

Manifold alignment: Methods from the literature

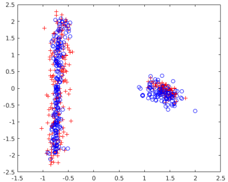
Common theme: Create the latent space by finding and correlating redundancies between sets.



X^1



X^2



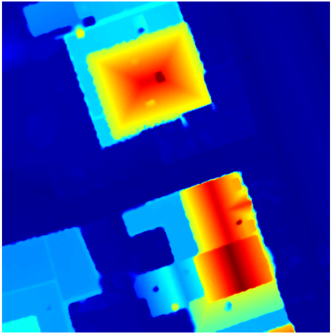
Images in latent space

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.

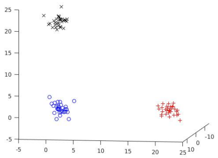
Distinguish road from grass



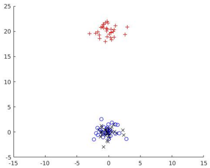
Distinguish roof from ground

Synthetic example: Data

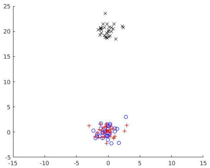
Synthetic example:
Ground truth = 3 point clouds in $\mathbb{R}^3$ (20 points per cloud).
Modality 1 = projection onto xy-plane.
Modality 2 = projection onto xz-plane.



Ground Truth



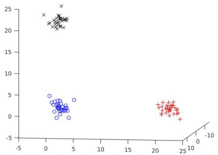
Modality 1



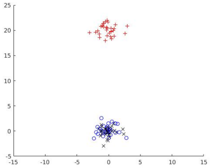
Modality 2

Synthetic example: Data

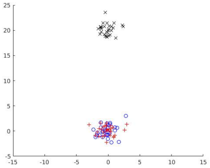
Assumption: Data is *co-registered*. i-th point from modality 1 corresponds to i-th point from modality 2.



Ground Truth



Modality 1



Modality 2

Synthetic Example: Result of CCA

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

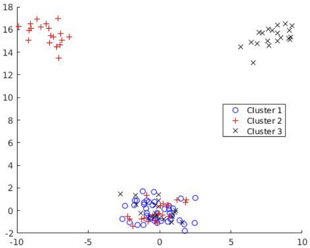


Image of clusters in latent space

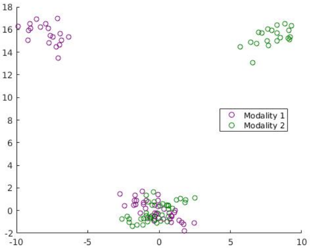
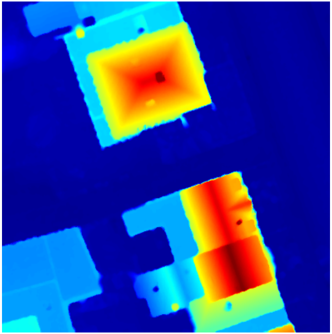


Image of data in latent space

Problem setup

We use co-registration assumption and Graph Laplacian theory for segmentation of multimodal datasets. [Iyer et al, 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.

Weight Matrix: Background

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

$\Longrightarrow$ 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

$$E^k_{ij} = \left\| x^k_i - x^k_j \right\|.$$

$\|\cdot\|$ chosen based on the details of the modality.

(in our examples $\|\cdot\|$ is 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 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

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

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

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

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

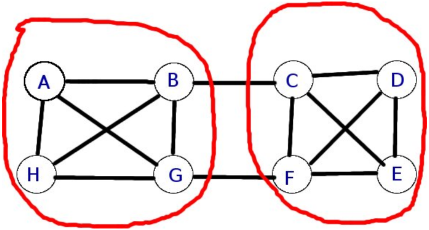
$$vol(A) = \sum_{i \in A, j \in \{1, \dots, n\}} w_{ij}.$$

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

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

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

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

Relaxed graph min cut

Optimal graph cut is

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

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

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

Solution:

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

Relaxed graph min cut

In relaxed problem,

$$\begin{aligned} \text{columns of } H &\iff \text{features} \\ \text{rows of } H &\iff \text{data points.} \end{aligned}$$

Can use features for a variety of applications.

Our code: K-means on feature vectors $\rightarrow$ classification

(this is called 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

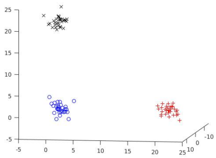
Nyström: Approximate Graph Laplacian eigenvectors using only $W_{A,A}$, $W_{A^c,A}$.

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

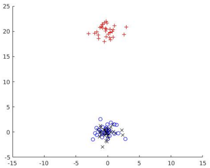
Compute and store matrices of size at most $|X| \times |A|$.

Synthetic example: Data

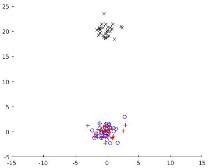
Synthetic example:
Ground truth $=$ 3 point clouds in $\mathbb{R}^3$ (20 points per cloud).
Modality 1 $=$ projection onto xy-plane.
Modality 2 $=$ projection onto xz-plane.



Ground Truth



Modality 1



Modality 2

Synthetic Example: Result of Our Method

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

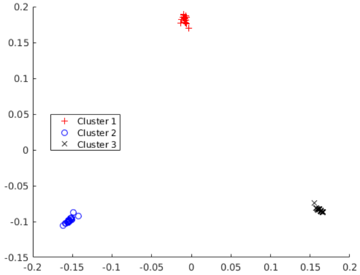
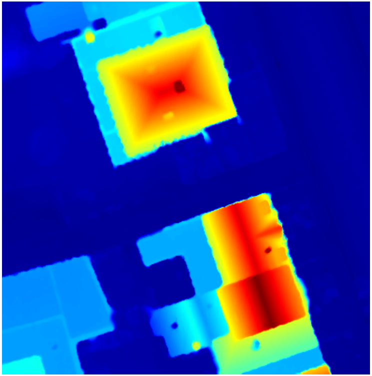


Image of clusters in latent space

Data

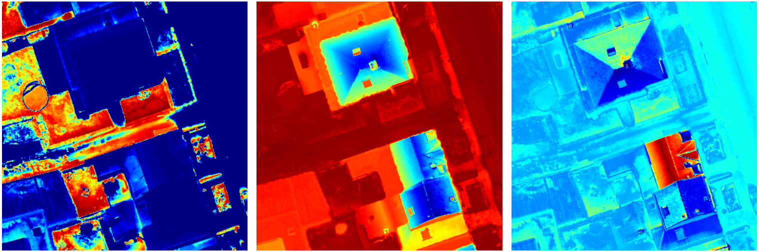
Our algorithm applied to [Bampos-Taberner et al, 2016] dataset.

RBG Modality



Lidar Modality

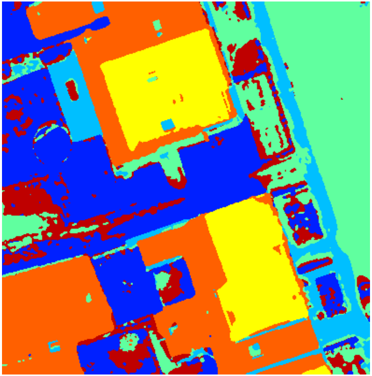
Results



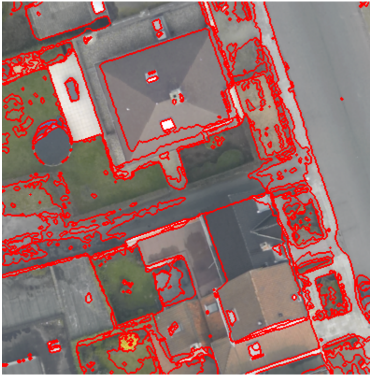
Example eigenvectors of Graph Laplacian

Results

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



Classes

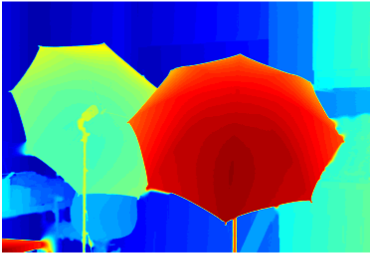


Regions on original image

Data

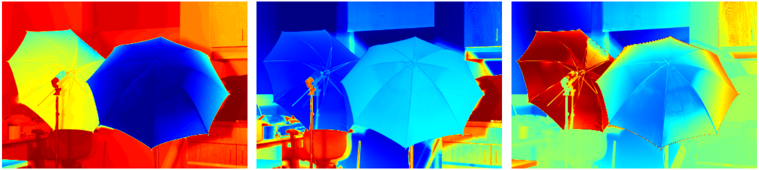
Our algorithm applied to [Scharstein et al. 2014] dataset.

RBG Modality



Lidar Modality

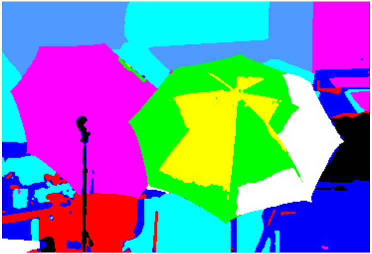
Results



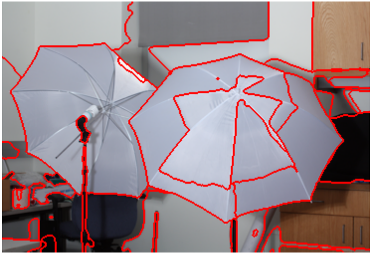
Example eigenvectors of Graph Laplacian

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.

Graph Matching Example

Graph 1 Graph 2

Any reasonable matching sends $1 \to 2$.

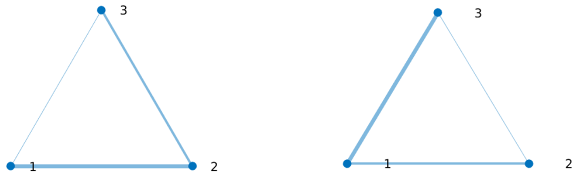
Other nodes can be matched in any way (symmetry).

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 $1 \to 3$, $2 \to 1$, $3 \to 2$.

Problem Setup

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

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

Exact solution is too expensive. Can solve using Graph Laplacian trick from [Umetama 1988, Knossow et al. 2009].

Relaxation

Relax problem to

$$Q^* = \text{argmin}_{QQ^T=I} \left\| QW_1Q^T - W_2 \right\|_F^2 .$$

Let L_1, L_2 the Graph Laplcians 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

$$\begin{aligned} \text{columns of } U_i &\iff \text{features} \\ \text{rows of } U_i &\iff \text{data points.} \end{aligned}$$

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 $O(N^3)$.

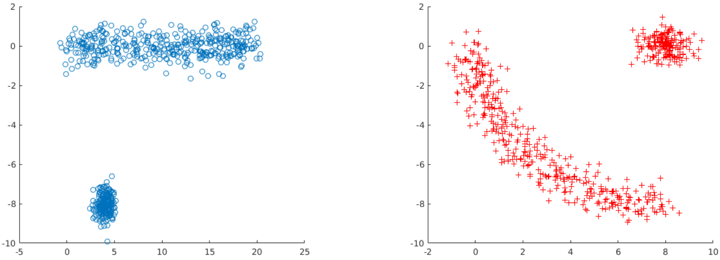
Benefits of Graph Matching

Benefits of Graph Matching

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

Example Matching

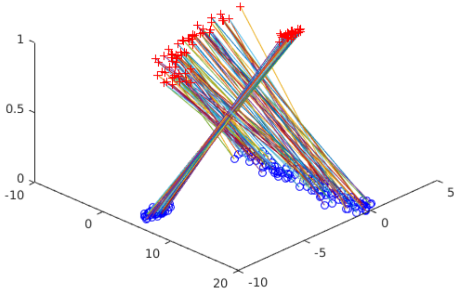
Recall from earlier.



X^1 $\qquad$ X^2

Synthetic Dataset

Example Matching



Result of our code

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.

Change Detection

Let $X = \{x_1, x_2, \ldots, x_n\}$, $Y = \{y_1, y_2, \ldots, y_n\}$.

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 $y_{\rho(i)}$.

A poor match $\implies$ some change occured.

Change Detection Example

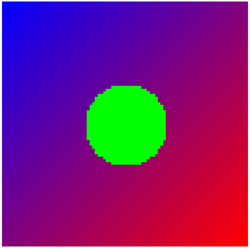


Image X

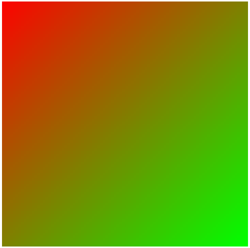
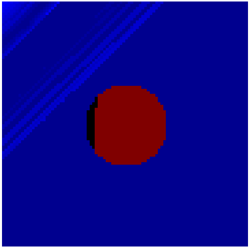


Image Y



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

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