GEOMETRIC GRAPH-BASED METHODS FOR HIGH DIMENSIONAL DATA

Andrea Bertozzi

University of California, Los Angeles

UCLA

Former students: Tijana Kostic (Microsoft), Cristina Garcia (LANL), Justin Sunu (CGU)

Current students: Huiyi Hu, Ekaterina Murkerjev, Michael Luo, Gloria Meng

Former Postdoc: Y. van Gennip (Nottingham), B. Osting (Utah), N. Guillen (U Mass)

Collaborators: Arjuna Flenner (China Lake), Allon Percus (CGU), Mason Porter (Oxford), Thomas Laurent (Loyola Marymount),

Inspiration: earlier work of Stan Osher, Chris Anderson, Luminita Vese, and Tony Chan

Thanks to NSF, ONR, AFOSR for support.

Variational Functionals for Image Segmentation - sharp interfaces with penalty function restricting regularity of interface

$$E(u,\Gamma) = \int_{R^2} (u-f)^2 dx + \mu \int_{R^2-\Gamma} |\nabla u|^2 dx + \nu|\Gamma|$$

Mumford-Shah segmentation model 1989 CPAM

$$E(z) = \alpha \int z_{ss}^2 ds + \beta \int z_s^2 ds + \int (F(z)) ds$$

Terzopoulos snakes, Lagrangian curve attracted to edges, F is an environmental function that attracts to edges, Kass-Witkin-Terzopoulos *IJCV* 1987

$$E(C_1,C_2,\Gamma) = \int_{\Gamma_{in}} (f-C_1)^2 + \int_{\Gamma_{out}} (f-C_2)^2 + \nu|\Gamma|$$

Chan-Vese Segmentation – binary with sharp interface Gamma between regions, *IEEE Trans. Imag. Proc*. 2001. Solved using level sets and the TV functional via a gradient flow.

FROM EUCLIDEAN SPACE TO SIMILARITY GRAPHS FOR LARGE DATA

Euclidean space ↓

- Minimal surface problem
- Laplace operator
- Pseudo-spectral methods
- Fast Fourier Transform
- Uses all the modes

Similarity graphs ↓

- Graph mincut problem
- Graph Laplacian
- Projection to eigensubspace of graph Laplacian
- Nystrom extension/ Rayleigh-Chebyshev
- Often only needs a small percentage of spectral modes.

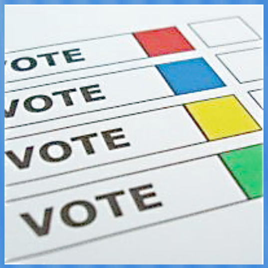
WEIGHTED GRAPHS FOR "BIG DATA"

$$w(x,y) = \exp(-||x-y||^2/\tau)$$

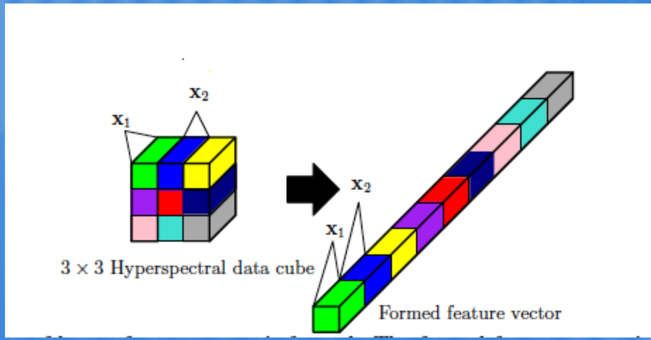
In a typical application we have data supported on the graph, possibly high dimensional. The above weights represent comparison of the data.

Examples include:

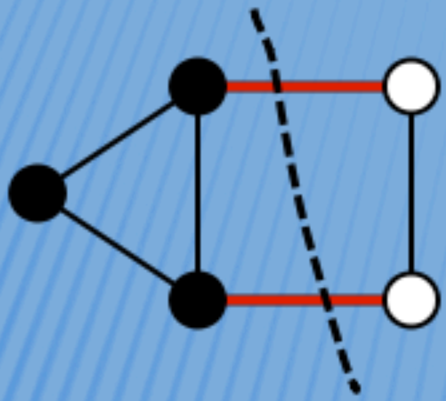
voting records of US Congress – each person has a vote vector associated with them.



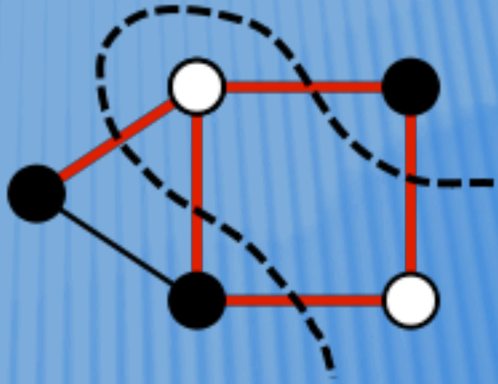
Nonlocal means image processing – each pixel has a pixel neighborhood that can be compared with nearby and far away pixels.



GRAPH CUTS AND TOTAL VARIATION



Minimum cut



Maximum cut

Total Variation of function *f* defined on nodes of a weighted graph:

$$\sum \omega_{ij} |f_i - f_j|$$

Min cut problems can be reformulated as a total variation minimization problem for binary/multivalued functions defined on the nodes of the graph.

NONLOCAL MEANS GRAPHS AND TOTAL VARIATION

- Buades Coll and Morel (2006)– introduced the NL Means functional for imaging applications – patch comparisons between pixels
- Osher and Gilboa (2007-8)– developed the Nonlocal TV functional for imaging applications- very effective for image inpainting applications with texture
- Drawback with Osher-Gilboa is slowness of algorithm
- We will accomplish these results with much faster run time and extend to general Machine Learning problems
- Suggests an alternative to the NL means calculus of Gilboa-Osher

DIFFUSE INTERFACE METHODS ON GRAPHS

Bertozzi and Flenner MMS 2012. SIAM Oustanding Paper Prize 2014

Arjuna Flenner
China Lake

$$L(\nu,\mu) = \begin{cases} d(\nu) & \text{if } \nu = \mu, \\ -w(\nu,\mu) & \text{otherwise.} \end{cases}$$

$$\langle u, Lu \rangle = \frac{1}{2} \sum_{\mu,\nu \in V} w(\nu,\mu)(u(\nu) - u(\mu))^2$$

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

$$E(u) = \frac{\epsilon}{2}\langle u, L_s u \rangle + \frac{1}{4\epsilon} \sum_{z \in Z} (u^2(z) - 1)^2 + \sum_{z \in Z} \frac{\lambda(z)}{2} (u(z) - u_0(z))^2.$$

CONVERGENCE OF GRAPH GL FUNCTIONAL

van Gennip and ALB Adv. Diff. Eq. 2012

Yves
Van Gennip

$$f_\varepsilon(u) := \chi \sum_{i,j=1}^{m} \omega_{ij}(u_i - u_j)^2 + \frac{1}{\varepsilon}\sum_{i=1}^{m} W(u_i),$$

$$\frac{1}{2}\|\nabla u\|_{\mathcal{E}}^2 = \frac{1}{4}\sum_{i,j\in I_m} \omega_{ij}(u_i - u_j)^2.$$

Theorem 3.1 (Γ-convergence). *$f_\varepsilon \xrightarrow{\Gamma} f_0$ as $\varepsilon \to 0$, where*

$$f_0(u) := \begin{cases} \chi \sum_{i,j\in I_m} \omega_{ij}|u_i - u_j| & \text{if } u \in \mathcal{V}^b, \\ +\infty & \text{otherwise} \end{cases} = \begin{cases} 2\chi TV_{a1}(u) & \text{if } u \in \mathcal{V}^b, \\ +\infty & \text{otherwise.} \end{cases}$$

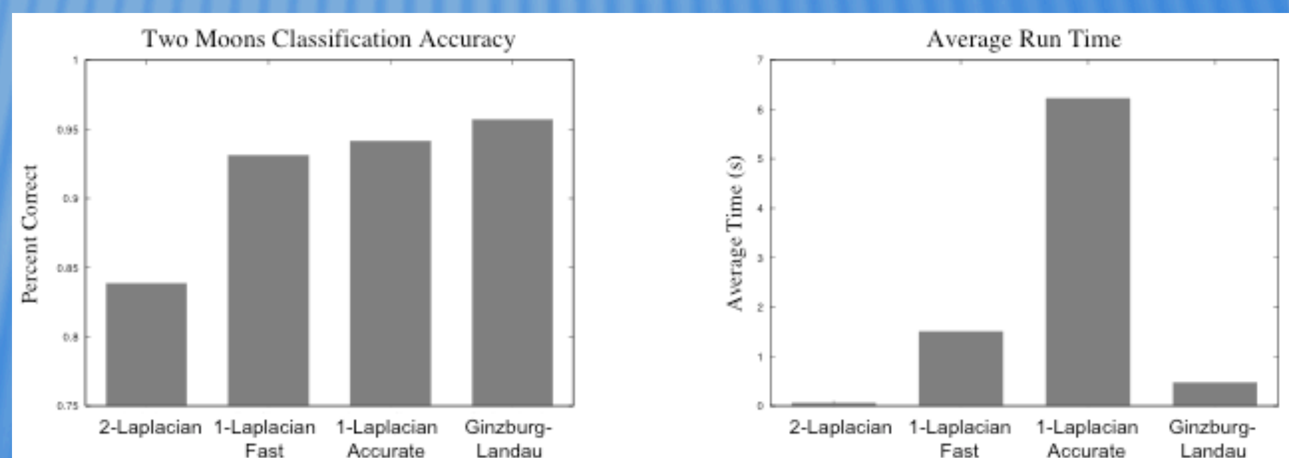
DIFFUSE INTERFACES ON GRAPHS

An Example: two moons

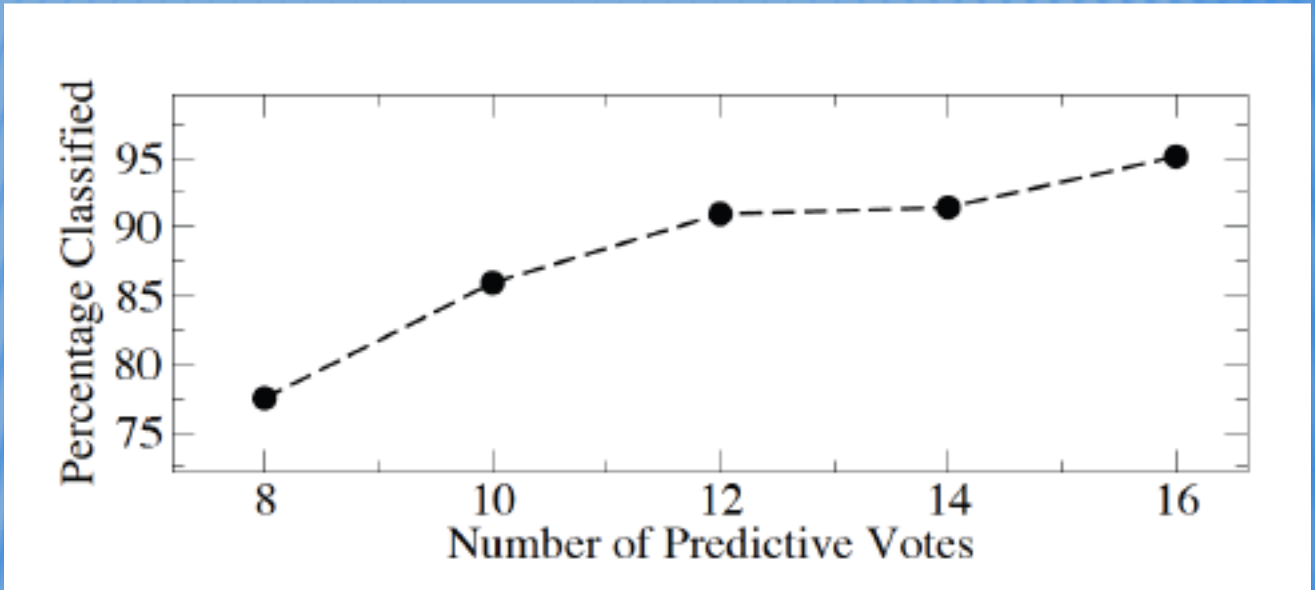
Replaces Laplace operator with a weighted graph Laplacian in the Ginzburg Landau Functional

Allows for segmentation using L1-like metrics due to connection with GL

Comparison with Hein-Buehler 1-Laplacian 2010.

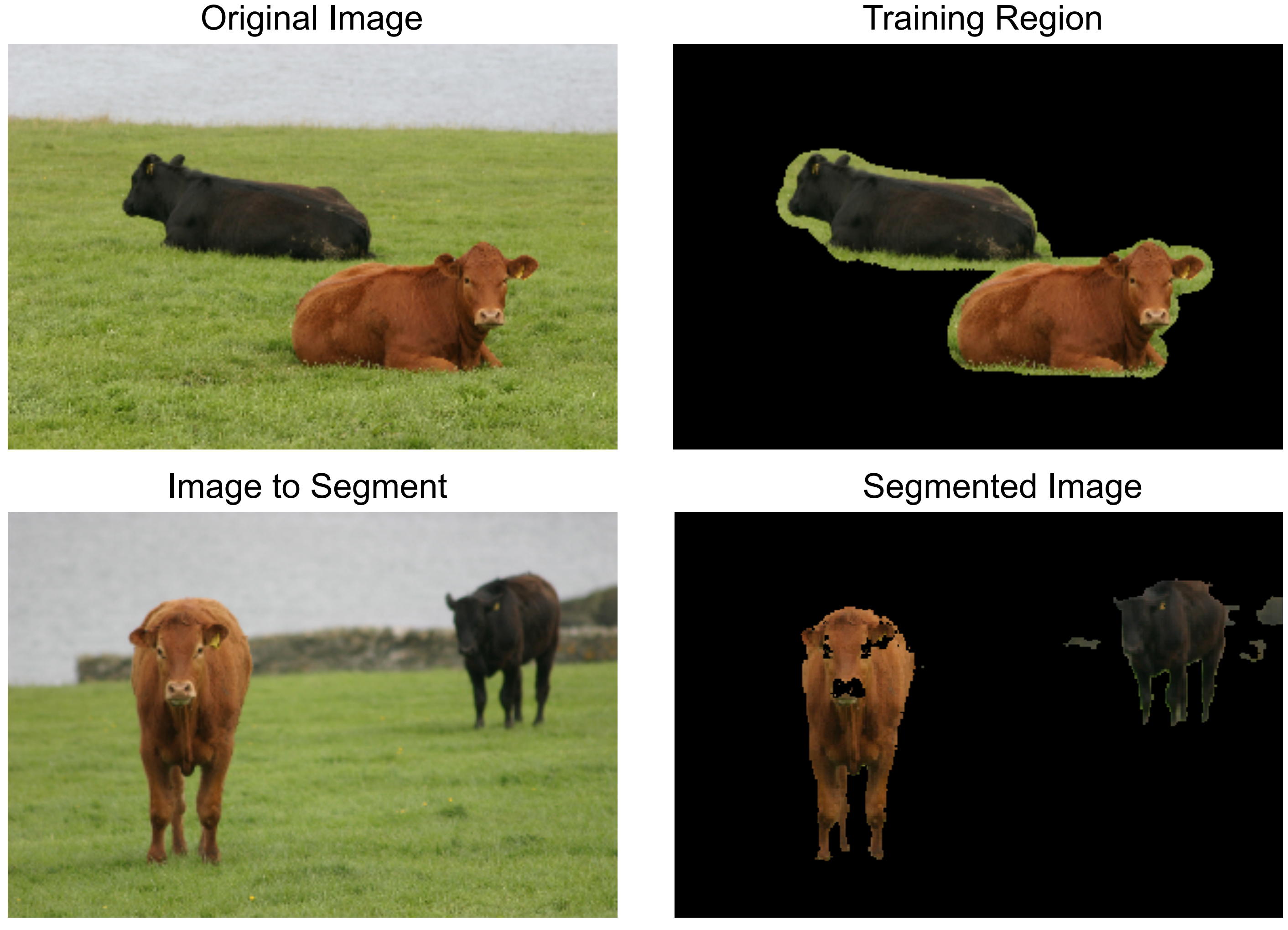


US HOUSE OF REPRESENTATIVES VOTING RECORD CLASSIFICATION OF PARTY AFFILIATION FROM VOTING RECORD



98th US Congress 1984
Assume knowledge of party affiliation of 5 of the 435 members of the House
Infer party affiliation of the remaining 430 members from voting records
Gaussian similarity weight matrix for vector of votes (1, 0, -1)

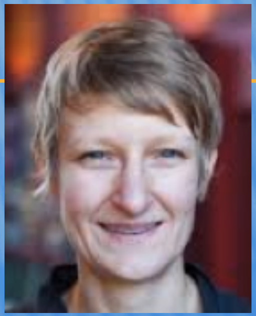
MACHINE LEARNING IDENTIFICATION OF SIMILAR REGIONS IN IMAGES



High dimensional fully connected graph – use Nystrom extension methods for fast computation methods.

RECALL CONVEX SPLITTING SCHEMES

Schoenlieb and Bertozzi, *Comm. Math. Sci. 2011*
Analysis of convex splitting schemes for higher order PDE in image processing



Carola Schoenlieb

Basic idea:

$$E(u) = E_c(u) - E_e(u)$$

$$U_{k+1} - U_k = -\Delta t(\nabla E_c(U_{k+1}) - \nabla E_e(U_k))$$

Project onto Eigenfunctions of the gradient (first variation) operator

For the GL functional the operator is the graph Laplacian

REMOVE THE DIFFUSE INTERFACE: MBO SCHEME ON GRAPHS

Merkurjev, Kostic, and ALB, SIIMS 2013

- 1) propagation by graph heat equation + forcing term

$$\frac{\partial z}{\partial t} = -L_s z - C_1 \lambda(x)(z - z_0)$$

- 2) thresholding

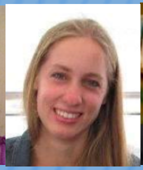
$$u^{n+1}(x) = \begin{cases} 1, & \text{if } y(x) \geq 0 \\ -1, & \text{if } y(x) < 0 \end{cases}$$

- Simple! And often converges in just a few iterations (e.g. 4 for MNIST dataset)

ALGORITHM

- I) Create a graph from the data, choose a weight function and then create the symmetric graph Laplacian.
- II) Calculate the eigenvectors and eigenvalues of the symmetric graph Laplacian. *It is only necessary to calculate a portion of the eigenvectors*.*
- III) Initialize u.
- IV) Iterate the two-step scheme described above until a stopping criterion is satisfied.
- *Fast linear algebra routines are necessary – either Raleigh-Chebyshev procedure or Nystrom extension.

GENERALIZATION MULTICLASS MACHINE LEARNING PROBLEMS (MBO)

Garcia, Merkurjev, Bertozzi, Percus, Flenner, *IEEE TPAMI, 2014*

Semi-supervised learning

$$E(u) = \frac{\epsilon}{2}\langle u, L_s u\rangle + \frac{1}{2\epsilon}\sum_{i\in V}\prod_{j=1}^{K}\frac{1}{4}\|\vec{u}_i - \vec{s}_j\|^2_{L_1} + \sum_{i\in V}\frac{\lambda_i}{2}\left\|\vec{u}_i - \vec{u}_i^{\,0}\right\|^2, \quad (13)$$

where

$$u = \begin{bmatrix}\vec{u}_1\\ \dots\\ \vec{u}_{N_D}\end{bmatrix} \text{ with } \vec{u}_i = [(u_i)_1, \dots, (u_i)_K]$$

$$\langle u, L_s u\rangle = \text{trace}(u^T L_s u)$$

$$\|\vec{u}_i - \vec{s}_j\|_{L_1} = \sum_{m=1}^{K}|(u_i)_m - \delta_{jm}|$$

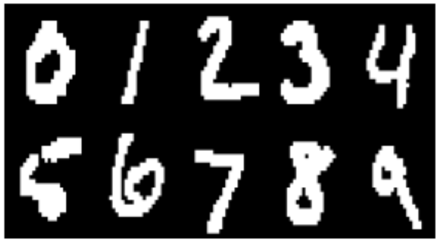


Fig. 4: Examples of digits from the MNIST data base

Instead of double well we have N-class well with Minima on a simplex in N-dimensions

MNIST DATABASE

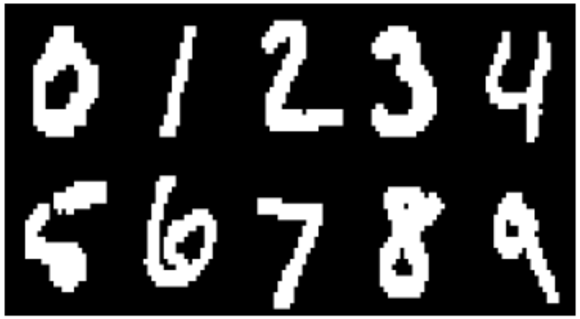


Fig. 6: Examples of digits from the MNIST data base

MNIST

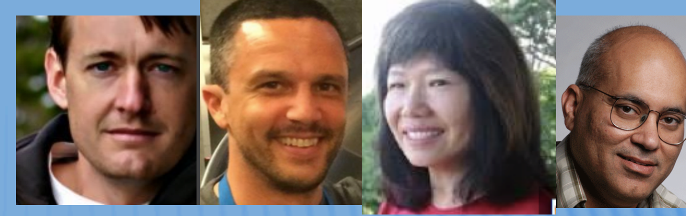
Method	Accuracy
p-Laplacian	87.1%
multicut norm. 1-cut	87.64%
Cheeger cut	88.2%
linear classifiers	88%
nonlinear classifiers	96.4%-96.7%
k-NN	95.0%- 97.17%
boosted stumps	92.3%- 98.74%
neutral/convolution nets	95.3-99.65%
SVM	98.6%-99.32%
multiclass GL	96.8%
MBO reduction	96.91%

We use local rescaled graph as in Zelnik-Manor&Perona

Comparisons
Semi-supervised learning
Vs Supervised learning

We do semi-supervised with only 3.6% of the digits as the Known data.

Supervised uses 60000 digits for training and tests on 10000 digits.



NYSTROM EXTENSION

Fowlkes Belongie Chung and Malik, IEEE T. PAMI 2004.

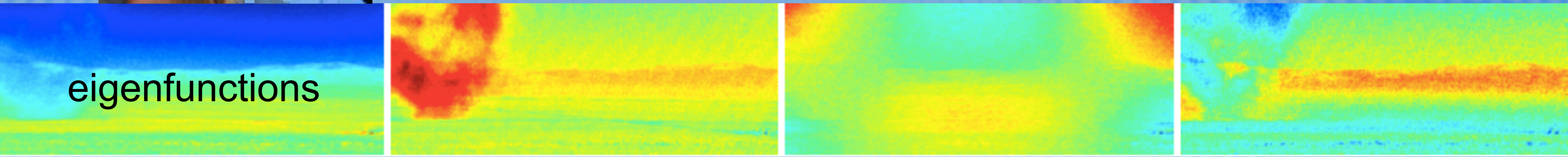
$$W = \begin{pmatrix} W_{XX} & W_{XY} \\ W_{YX} & W_{YY} \end{pmatrix}, \quad W \sim \begin{pmatrix} W_{XX} \\ W_{YX} \end{pmatrix} W_{XX}^{-1} \begin{pmatrix} W_{XX} & W_{XY} \end{pmatrix}.$$

Computing $W_{XX}, W_{XY} = W_{YX}^T$ requires only $(|X| \cdot (|X| + |Y|)$ computations versus $(|X| + |Y|)^2$ for the whole similarity matrix. The method approximates W_{YY} by $W_{YX} W_{XX}^{-1} W_{XY}$ and the error is determined by how much the rows of W_{XY} span the rows of W_{YY}.

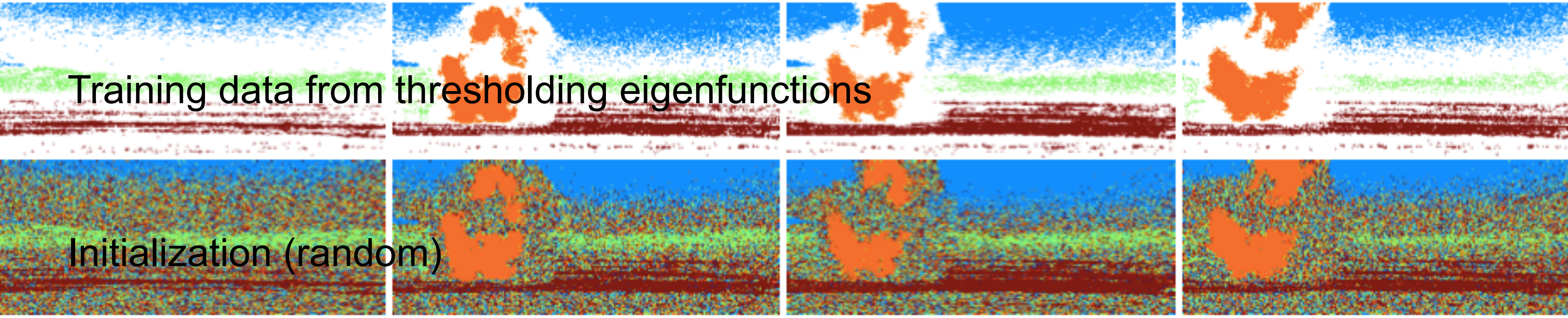
HYPERSPECTRAL VIDEO SEGMENTATION – SEMI SUPERVISED

Merkurjev, Sunu, and Bertozzi, 2014, ICIP Paris 2014

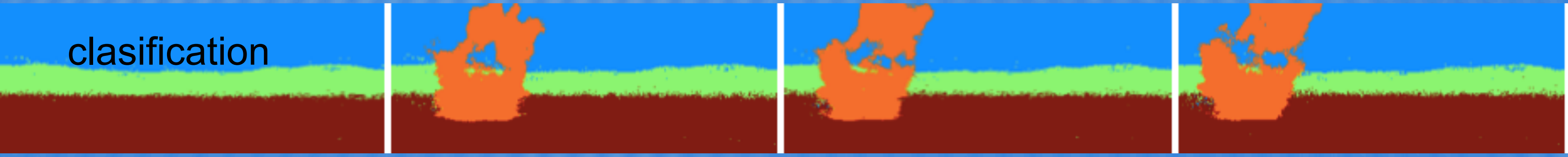
Eigenfunctions computed using Nystrom



“ground truth obtained from thresholding eigenfunctions; random initialization otherwise



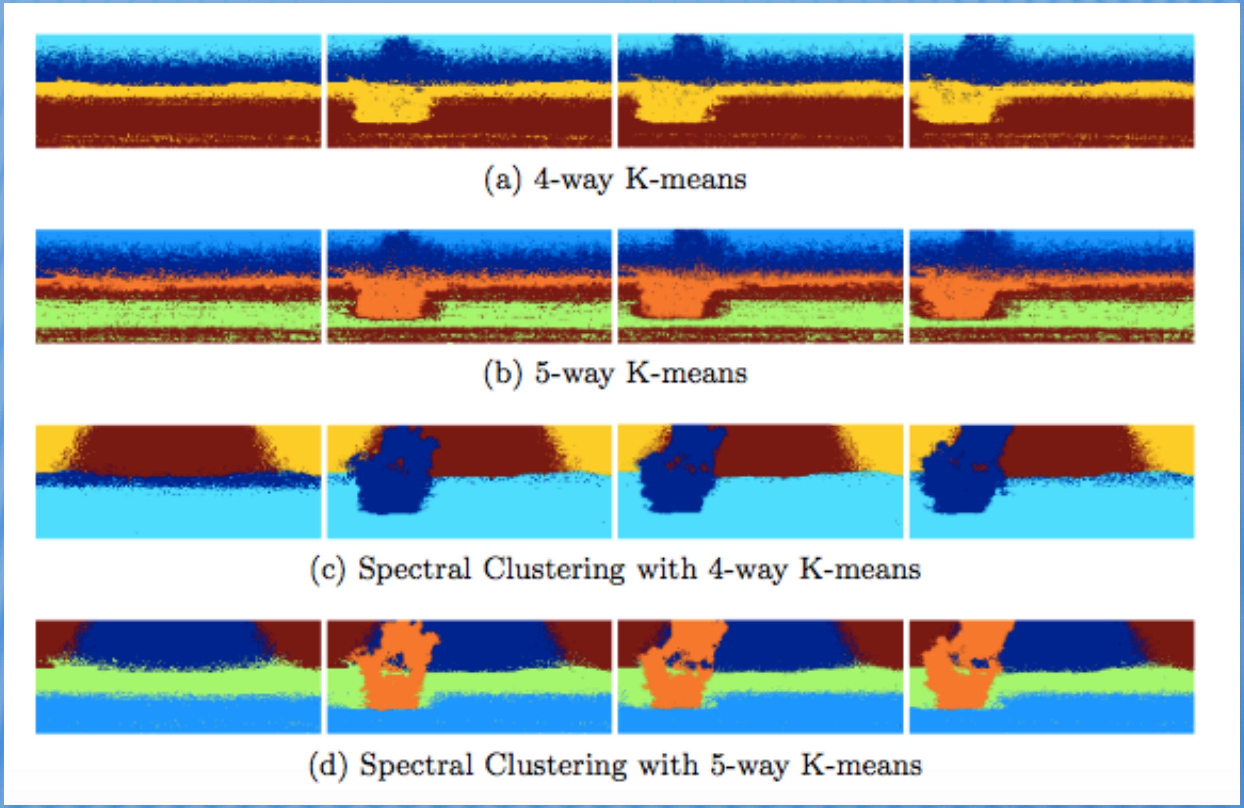
Four class hyperspectral pixel segmentation of gas plume, ground, mountain, and sky



COMPARISON TO KMEANS AND SPECTRAL CLUSTERING - UNSUPERVISED

EMMCVPR 2015 Hu, Sunu, and ALB

K-means
And
Spectral
Clustering



(a) 4-way K-means

(b) 5-way K-means

(c) Spectral Clustering with 4-way K-means

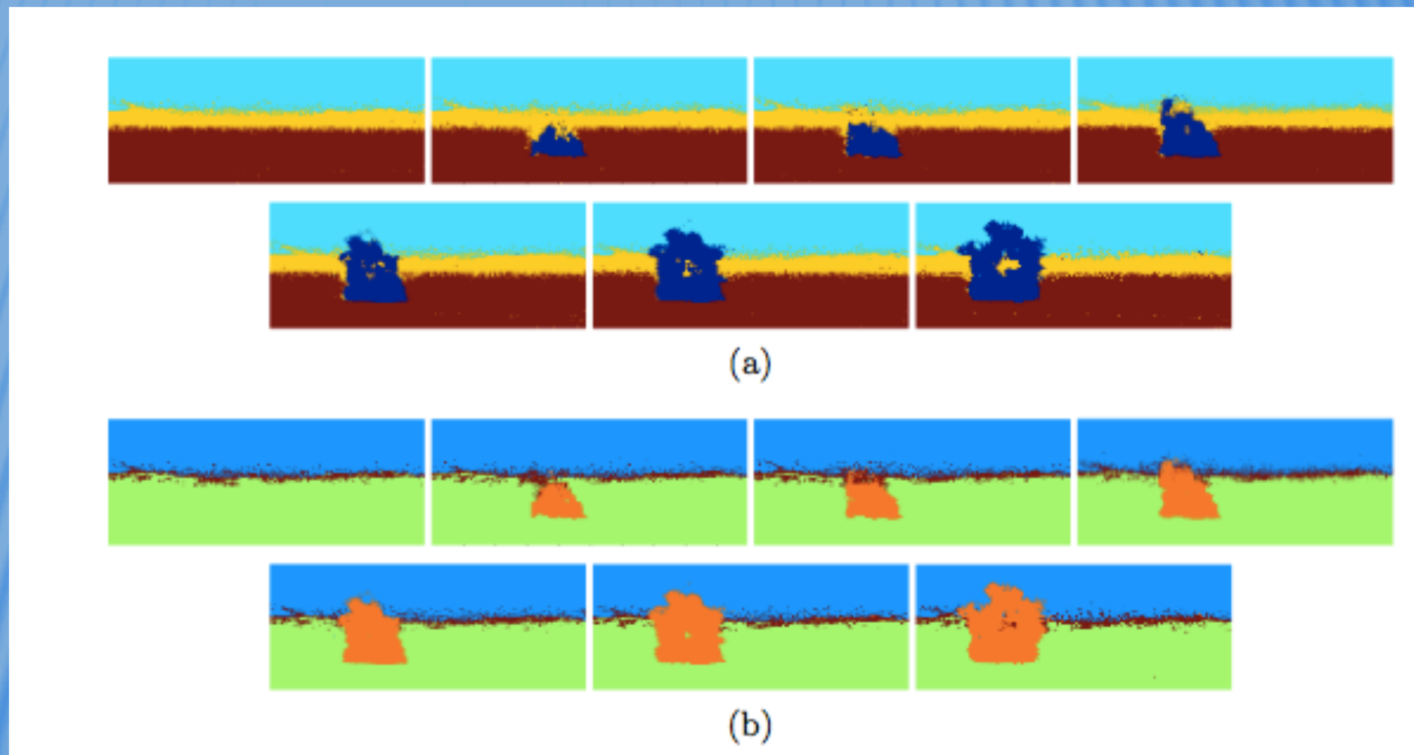
(d) Spectral Clustering with 5-way K-means

C-V SEGMENTATION ON GRAPHS USING MBO SCHEME FOR UNSUPERVISED CLUSTERING OF HYPERSPECTRAL PIXELS

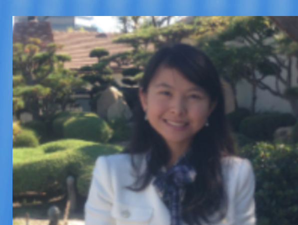
Multiclass MBO
with different Initializations.

7 video frames
280K pixels

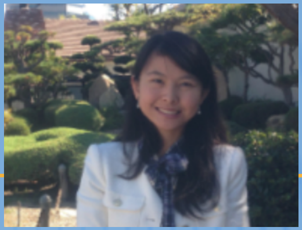
Each pixel is
128 dimensions



EMMCVPR 2015 Hu, Sunu, and ALB



COMMUNITY DETECTION – MODULARITY OPTIMIZATION



Joint work with Huiyi Hu (UCLA), Thomas Laurent (Loyola Marymount), and Mason Porter (Oxford) SIAP 2013.

$$\text{Modularity: } Q = \frac{1}{2m}\sum_{ij}(w_{ij} - \gamma P_{ij})\delta(g_i, g_j)$$

Newman, Girvan, Phys. Rev. E 2004.

The modularity of a partition measures the fraction of total edge weight within each community minus the edge weight expected if edges were placed randomly using some null model.

$[w_{ij}]$ is graph adjacency matrix
P is probability nullmodel (Newman-Girvan) $P_{ij}=k_i k_j/2m$
$k_i = \text{sum}_j\ w_{ij}$ (strength of the node)
Gamma is the resolution parameter
g_i is group assignment
2m is total volume of the graph = $\text{sum}_i\ k_i = \text{sum}_{ij}\ w_{ij}$

This is an optimization (max) problem. Combinatorially complex – optimize over all possible group assignments. Very expensive computationally.

EQUIVALENCE TO L1 COMPRESSIVE SENSING

Thus modularity optimization restricted to two groups is equivalent to

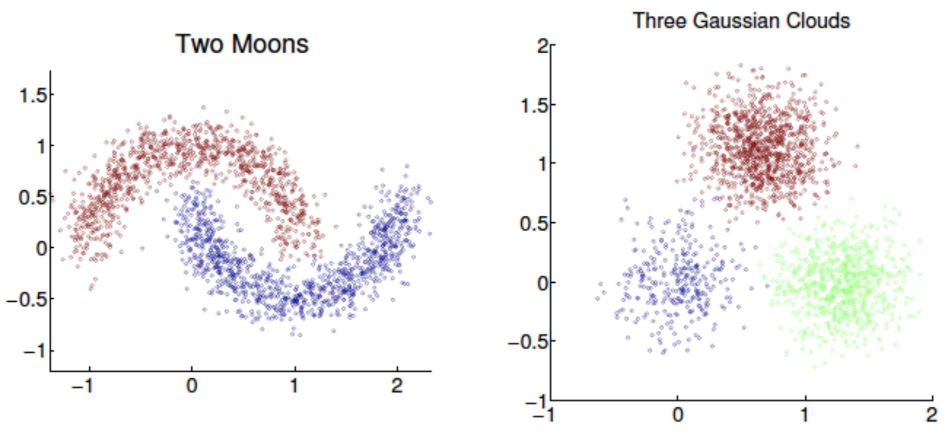
$$\boxed{\text{Minimize}_{\{f:G\to\{\pm 1\}\}} |f|_{TV} - \frac{\gamma}{2}\|f - m_2(f)\|^2_{L_2}}$$

This generalizes to n class optimization quite naturally

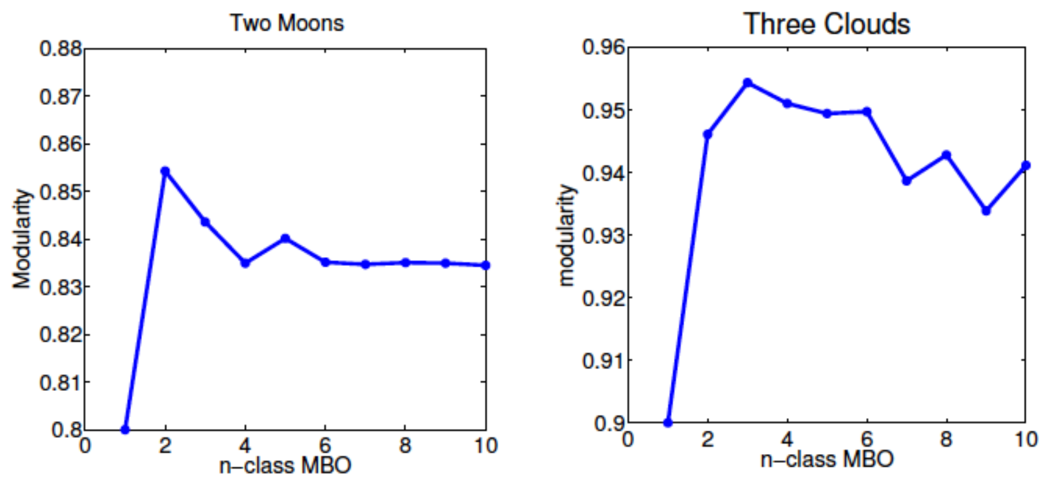
$$\boxed{\text{Minimize}_{\{f:G\to V^n\}}\ E(f) := |f|_{TV} - \gamma\|f - m_2(f)\|^2_{L_2}}$$

Because the TV minimization problem involves functions with values on the simplex we can directly use the MBO scheme to solve this problem.

MODULARITY OPTIMIZATION MOONS AND CLOUDS



(a) (b)



(a) (b)

MNIST DIGIT CLASSIFICATION USING MODULARITY – UNSUPERVISED

Binary segmentation of 4 and 9:
13782 handwritten digits. Graph created based on similarity score between each digit. Weighted graph with 194816 connections.

	N_c	Q	NMI	Purity	Time (seconds)
GenLouvain	2	0.9305	0.85	0.975	110 s
Modularity MBO ($\hat{n} = 2$)	2	0.9316	0.85	0.977	11 s
Multi-$\hat{n}$ MM ($\hat{n} \in \{2, 3, \dots, 10\}$)	2	0.9316	0.85	0.977	25 s
Spectral Clustering (k-Means)	2	NA	0.003	0.534	1.5 s

TABLE 4.2

Full multiclass Segmentation of all 70K digits

TABLE 4.3
Computation times and network diagnostics for partitions of the MNIST 70k data set.

	N_c	Q	NMI	Purity	Time (s)
GenLouvain	11	0.93	0.92	0.97	10900
Multi-$\hat{n}$ MM ($\hat{n} \in \{2, 3, \dots, 20\}$)	11	0.93	0.89	0.96	290 / 212 *
Modularity MBO 3% GT ($\hat{n} = 10$)	10	0.92	0.95	0.96	94.5 / 16.5 *

* *Calculated with the RC procedure.*

GLOBAL METHOD

Global binary optimization on graphs for classification of high dimensional data
Ekaterina Merkurjev, Egil Bae, Andrea L. Bertozzi, Xue-Cheng Tai

	max-flow	primal augmented Lagrangian	binary MBO	binary GL
MNIST (3.6% fidelity) random initialization, random fidelity	98.48%	98.44%	98.37%	98.29%
MNIST (3.6% fidelity) 2nd eigenvector initialization, random fidelity	98.48%	98.43%	98.36%	98.25%
MNIST (3.6% fidelity) random initialization, corner fidelity	98.47%	98.40%	62.35%	64.39%
MNIST (3.6% fidelity) 2nd eigenvector initialization, corner fidelity	98.46%	98.40%	63.87%	63.19%
Banknote Authentication Data Set (5.1% fidelity)	99.09%	98.75%	95.43%	97.76%
Banknote Authentication Data Set (3.6% fidelity)	98.83%	98.29%	93.48%	96.10%
two moons (5% fidelity)	97.10%	97.07%	98.41%	98.31%
two moons (2.5% fidelity)	97.05%	96.78%	97.53%	98.15%

Table 2 Number of Iterations and Timing

Number of iterations	max-flow	primal augmented Lagrangian	binary MBO	binary GL
MNIST	426	2709	10	52
Banknote Authentication Data Set	314	725	7	449
two moons	1031	451	8	108
Timing (s)	**max-flow**	**primal augmented Lagrangian**	**binary MBO**	**binary GL**
MNIST [a]	2.88	43.21	0.52	0.78
Banknote Authentication Data Set	1.21	3.76	0.90	0.95
two moons	4.13	5.23	2.30	2.98

REPRINTS

- A. L. Bertozzi and A. Flenner, *Multiscale Modeling and Simulation*, 10(3), 2012.
- Tijana Kostic and Andrea Bertozzi, *J. Sci Comp.*, 2012
- Y van Gennip and ALB *Adv. Diff. Eq*. 2012
- H. Hu, Y. van Gennip, B. Hunter, A.L. Bertozzi, M.A. Porter, I*EEE ICDM'12*, 2012.
- Y. van Gennip et al *SIAP* (spectral clustering gang data) 2013
- E. Merkurjev, T. Kostic, and A. L. Bertozzi, *SIAM J. Imag. Proc. 2013.*
- Huiyi Hu, Thomas Laurent, Mason A. Porter, Andrea L. Bertozzi, *SIAM J. Appl. Math., 2013.*
- C. Garcia-Cardona, E. Merkurjev, A. L. Bertozzi, A. Flenner and A. G. Percus,, *IEEE Trans. PAMI 2014*
- Y. van Gennip, N. Guillen, B. Osting, and A. L. Bertozzi, Mean curvature, threshold dynamics, and phase field theory on finite graphs*, Milan J. Math. 2014.*
- Huiyi Hu, Justin Sunu, and Andrea L. Bertozzi, Multi-class Graph Mumford-Shah Model for Plume Detection using the MBO scheme, *Proc. EMMCVPR* Hong Kong 2015, pp. 209-222.
- E. Merkurjev, E. Bae, A. L. Bertozzi, X-C. Tai, Global binary optimization on graphs for classification of high dimensional data, *JMIV* 2015.