Metric Learning for Hyperspectral Image Segmentation

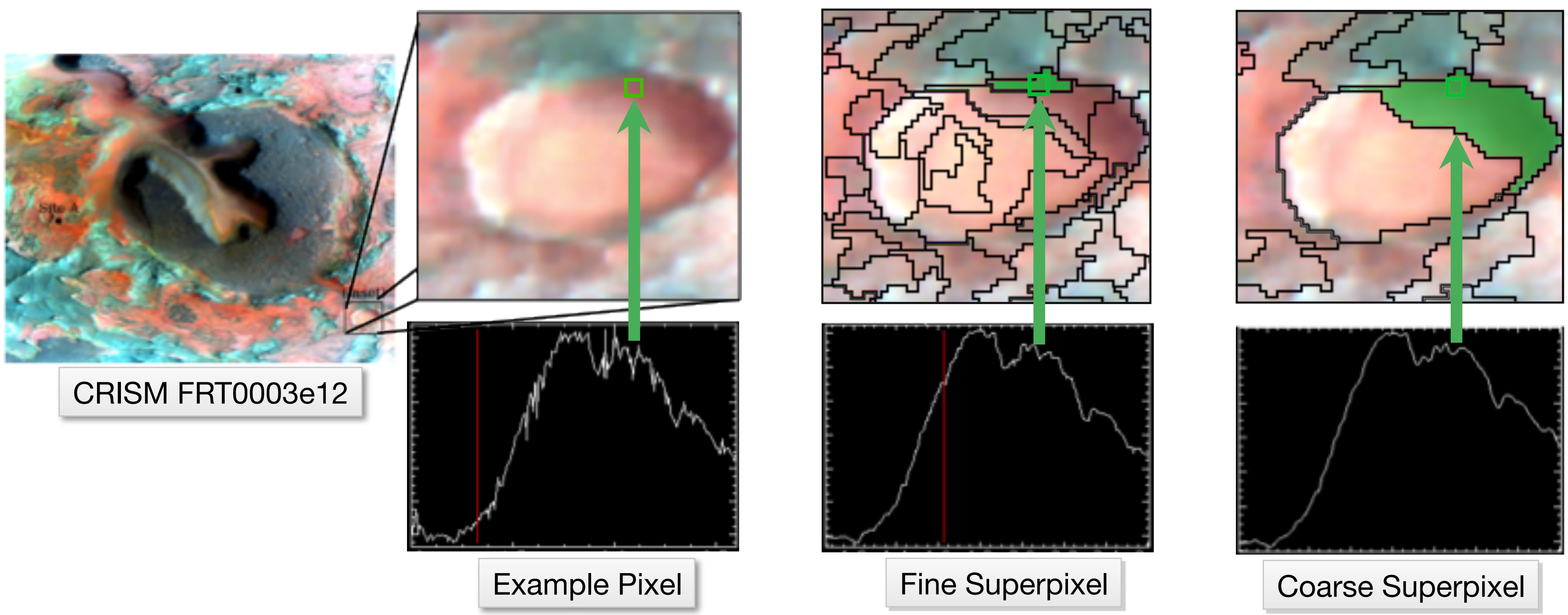
B.D. Bue[1], D.R. Thompson[2], M.S. Gilmore[3], R. Castaño[2]

[1] Rice University, Electrical and Computer Engineering, Houston, TX
[2] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA
[3] Weslayan University, Earth and Environmental Sciences, Middletown, CT

The work described involves research and development carried out for the AMMOS MultiMission Operations System and MGSS Instrument Operations Subsystem. AMMOS and U.S. Government Support Acknowledged.

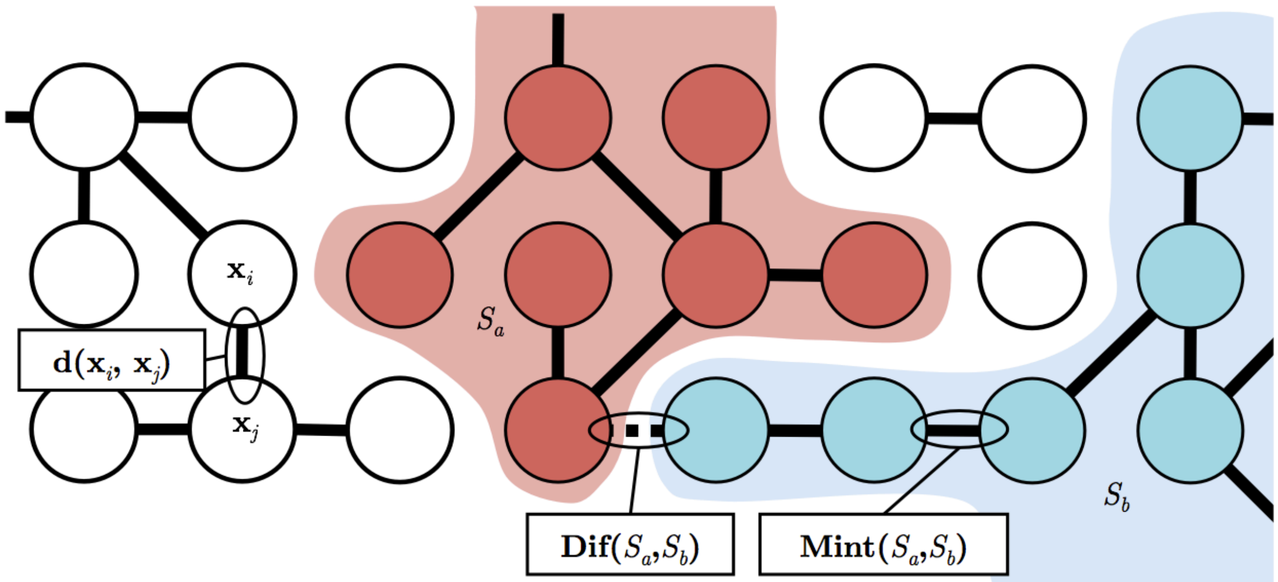
Application: Superpixel Segmentation [Thompson et al., 2010]

- Find spatially contiguous, spectrally homogeneous regions (“superpixels”) corresponding to physical features
- Reduces processing time of subsequent analyses
- Yields theoretical noise improvement of order $n^{1/2}$ for a superpixel of size n



Graph-based Segmentation Algorithm [Felzenszwalb]

- Image = 8-connected graph weighted by distances $\mathrm{d}(\mathrm{x_i}, \mathrm{x_j})$ between adjacent pixels $\mathrm{x_i}$ and $\mathrm{x_j}$
- Agglomerative clustering iteratively connects segments by growing minimum spanning trees



- Segment merging criterion:

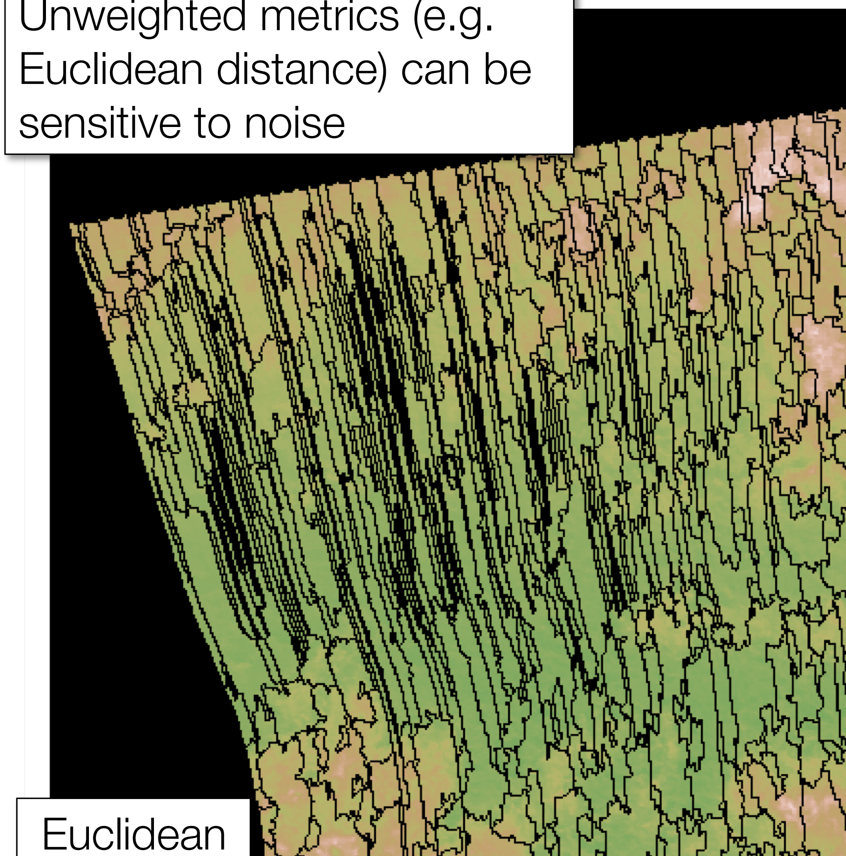
$$\mathrm{Dif}(S_a, S_b) > \mathrm{MInt}(S_a, S_b) = \min\left(\mathrm{Int}(S_a) + \frac{k}{|S_a|}, \mathrm{Int}(S_b) + \frac{k}{|S_b|}\right)$$

- Small k = many segments, large k = few segments, dependant on $\mathrm{d}(\mathrm{x_i}, \mathrm{x_j})$

Metric Learning for Hyperspectral Image Segmentation

- **Segmentation quality dependent on robustness of spectral similarity measure**

Unweighted metrics (e.g. Euclidean distance) can be sensitive to noise



Learned metrics can emphasize spectral bands relevant to training classes

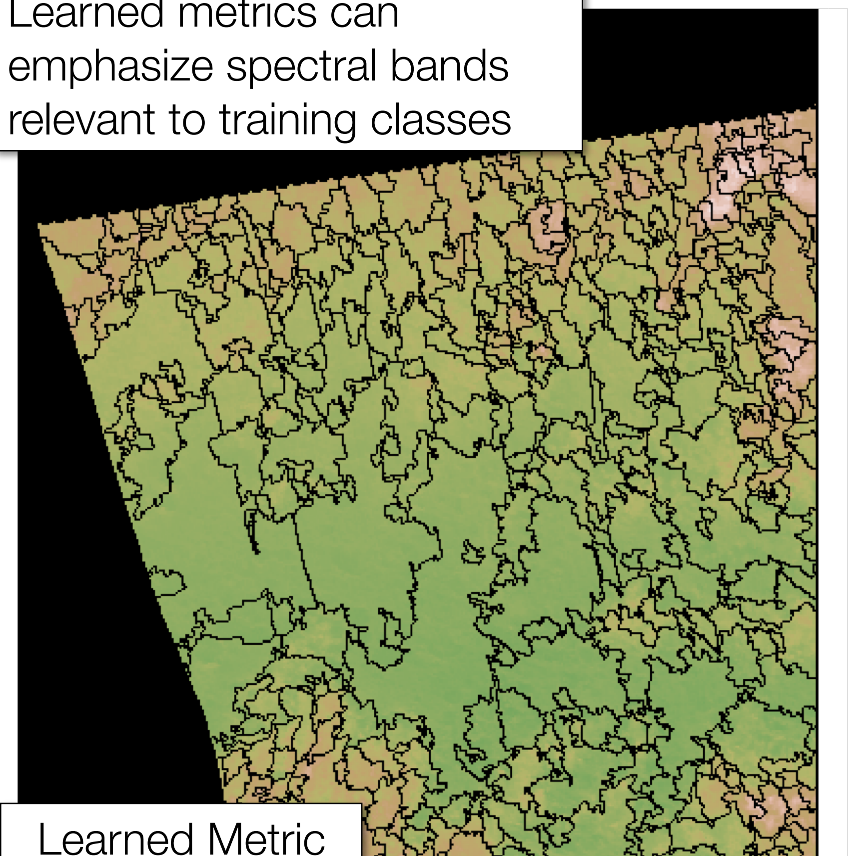


Image: CRISM FRT000863e

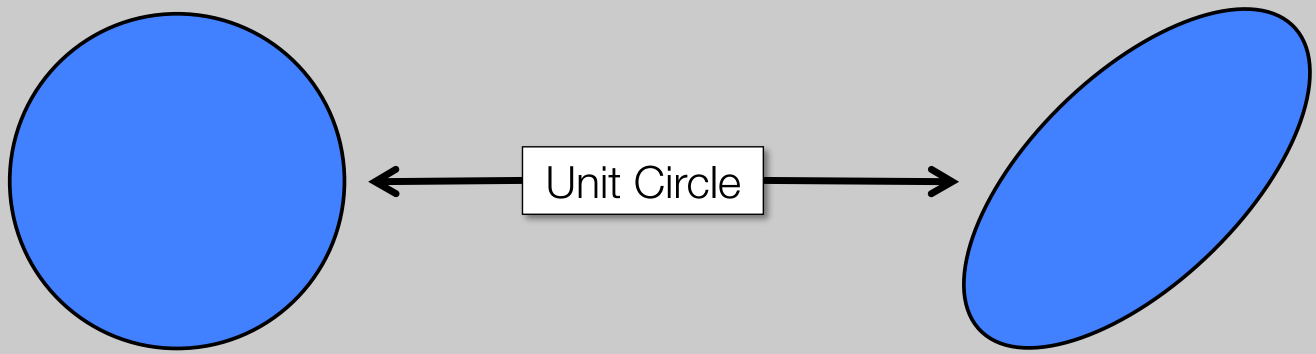
Mahalanobis Metric Learning

- Goal: learn a task-specific Mahalanobis metric given labeled data

$$\{\mathbf{x}_i, y_i\}_{i=1}^N \qquad \mathbf{x}_i \in \mathbb{R}^d,\; y_i \in [1, k]$$

$$\mathrm{d}_{\mathrm{Euc}}(\mathbf{x}_i, \mathbf{x}_j)^2 = (\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j)$$

$$\mathrm{d}_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j)^2 = (\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{M}(\mathbf{x}_i - \mathbf{x}_j)$$



- $\mathrm{M} = \mathrm{A}\mathrm{A}^{\mathrm{T}} =$ positive semi-definite transformation matrix
    - Squashes uninformative / emphasizes informative dimensions w.r.t. classes

Image credit: Weinberger et al. NIPS 2010

Multiclass Linear Discriminant Analysis [Fisher. 1934]

- Maximize separation ratio $S = \dfrac{\boldsymbol{\alpha}^T \boldsymbol{\Sigma}_b \boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \boldsymbol{\Sigma}_w \boldsymbol{\alpha}}$, where:

$$\boldsymbol{\Sigma}_w = \frac{1}{NK} \sum_{i=1}^{K} \sum_{j=1}^{N} (\mathbf{x}_j - \boldsymbol{\mu}_i)(\mathbf{x}_j - \boldsymbol{\mu}_i)^T \qquad \mu_i = \mathrm{E}\left[x_j | y_j = i\right]$$

$$\boldsymbol{\Sigma}_b = \frac{1}{K} \sum_{i=1}^{K} (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T \qquad \mu = \mathrm{E}\left[\mu_i\right]$$

- S maximized when α the top eigenvector of $\boldsymbol{\Sigma}_w{}^{-1} \boldsymbol{\Sigma}_b$
- A = top (k-1) eigenvectors of $\boldsymbol{\Sigma}_w{}^{-1} \boldsymbol{\Sigma}_b$
- To prevent degenerate solutions, regularize:

$$\boldsymbol{\Sigma}_w = (1 - \gamma) \boldsymbol{\Sigma}_w + \gamma \mathbf{I}, \ \gamma \in [0, 1]$$

Information Theoretic Metric Learning (ITML) [Davis et al. 2007]

- Bijection between Mahalanobis distances and multivariate Gaussians

$$\mathcal{N}(x|\mu, \mathbf{M}) = \frac{1}{Z}\exp(-\frac{1}{2}\mathrm{d}_{\mathbf{M}}(x, \mu))$$ (assume fixed μ)

- Solve:

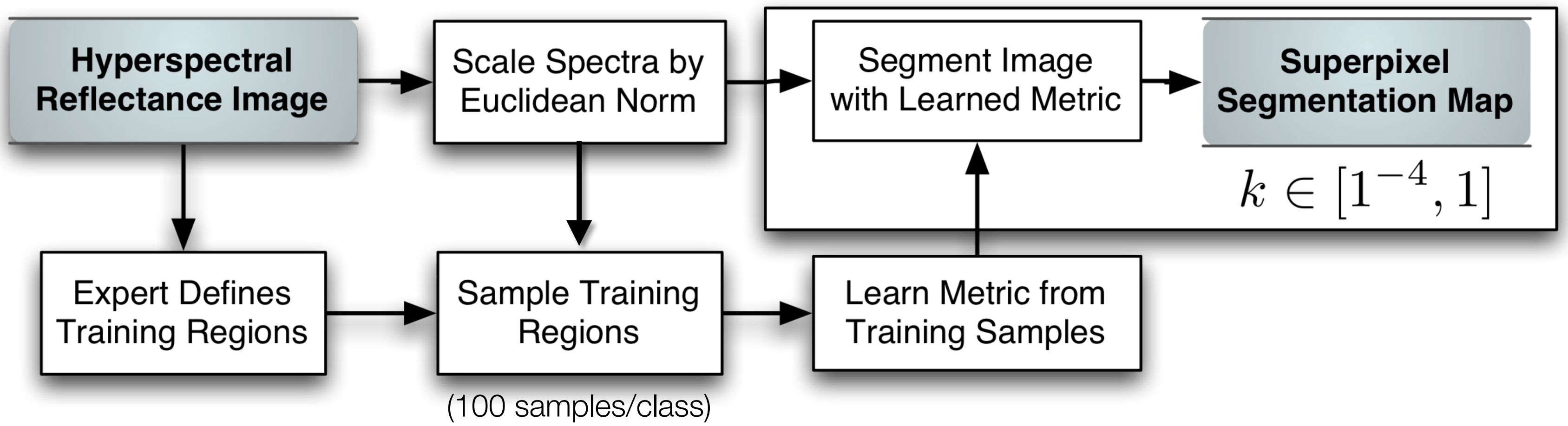
$$\min_{\mathbf{M}} \int \mathcal{N}(x|\mu, \mathbf{M}) \log\left(\frac{\mathcal{N}(x|\mu, \mathbf{M_0})}{\mathcal{N}(x|\mu, \mathbf{M})}\right) dx$$

  - $\mathbf{M}_0$ = regularization term - known, well-behaved Mahalanobis matrix (e.g., identity or sample covariance matrix)

- Subject to $\mathbf{M} \succeq \mathbf{0}$ and pairwise similarity/dissimilarity constraints:

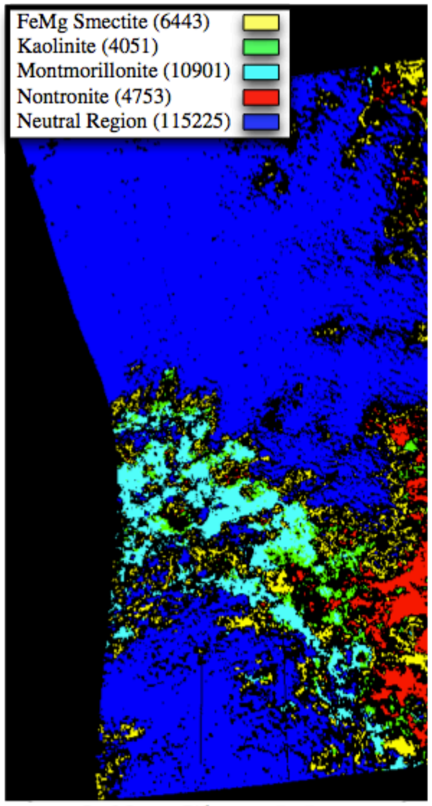
$$\mathrm{d}_{\mathbf{M}}(x_i, x_j) \leq u \rightarrow (i, j) \in S$$
$$\mathrm{d}_{\mathbf{M}}(x_i, x_j) \geq l \rightarrow (i, j) \in D$$

Evaluation Methodology

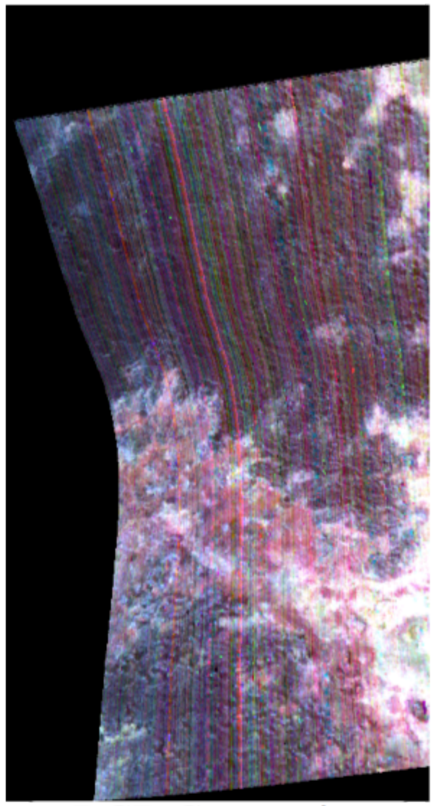


- # of segments (for a fixed image) dependant on (1) similarity metric and (2) segmentation parameter k
  - Vary k to compare segmentation maps with similar # of segments for each measure

Case Study: CRISM Imagery

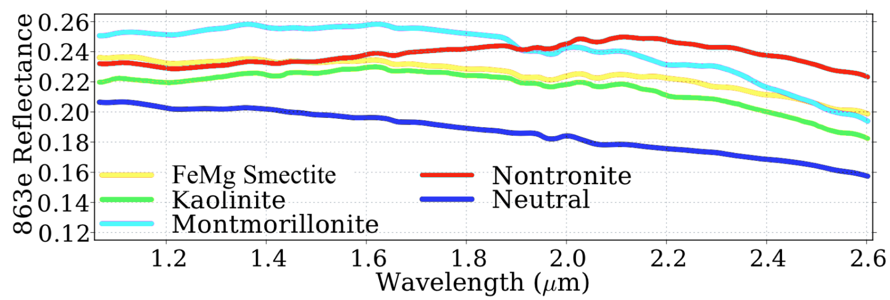


863e Class Map



RGB Composite

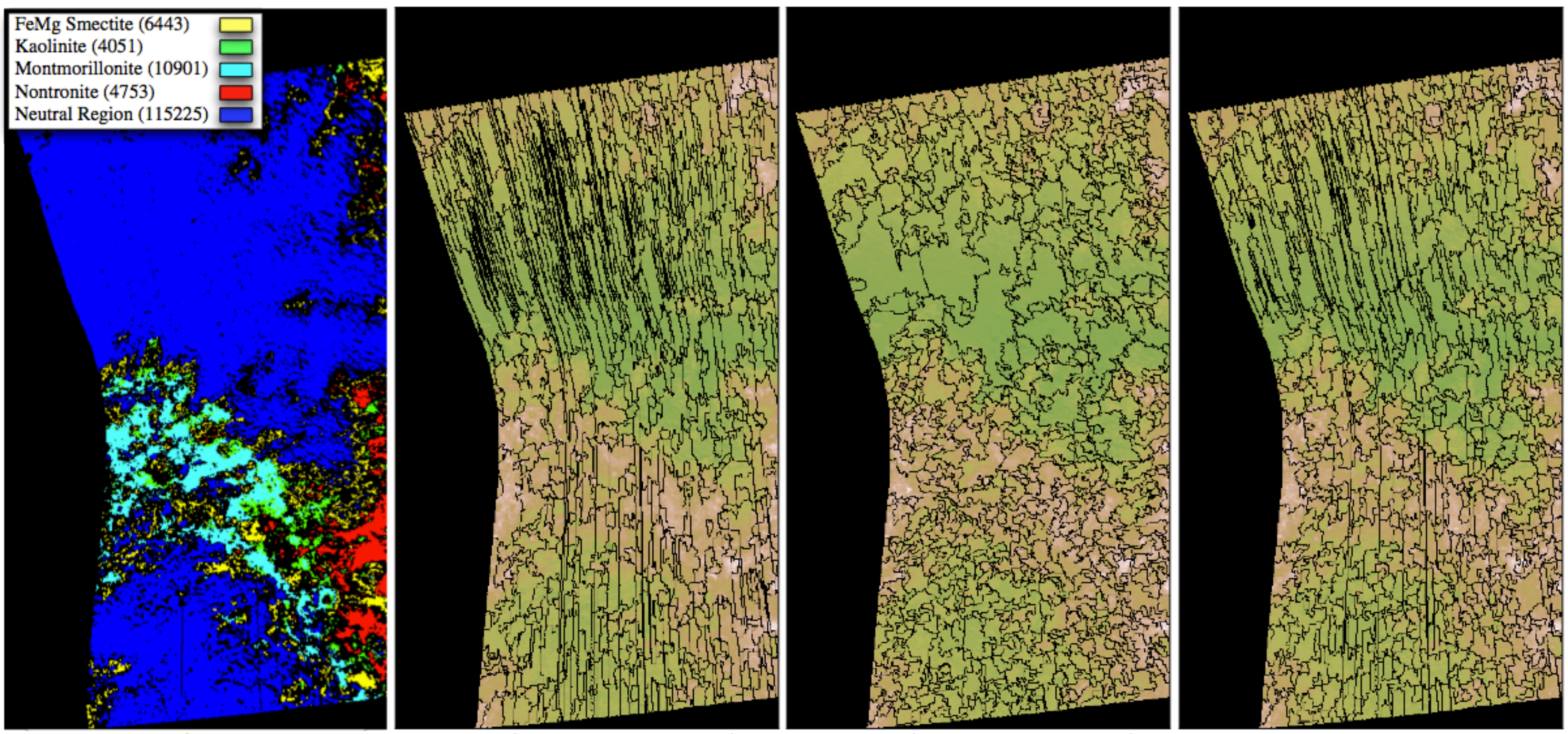
- Images: FRT 3e12, 3fb9, 863e
- 231 bands in [1.06, 2.58] μm
- Class maps defined and verified by geologist using ENVI Spectral Angle Mapper
- Unlabeled materials excluded from performance analysis



Evaluation Measures

- For a set of classes C and a set of segments S
- $H(C|S) = \sum_{c \in C, s \in S} p(c,s) \log \frac{p(c)}{p(c,s)}$
  - Measures remaining uncertainty in class map given segmentation partitions
  - If segmentation reproduces class map, $H(C|S)$ = 0
- $\text{purity}(S, C) = \sum_{s \in S} \frac{\text{pure}(s, C)}{|s|}$
  - $\text{pure}(s,C) = 1$ if all pixels in segment s belong to a single class c in C

Image 863e Segmentation Results

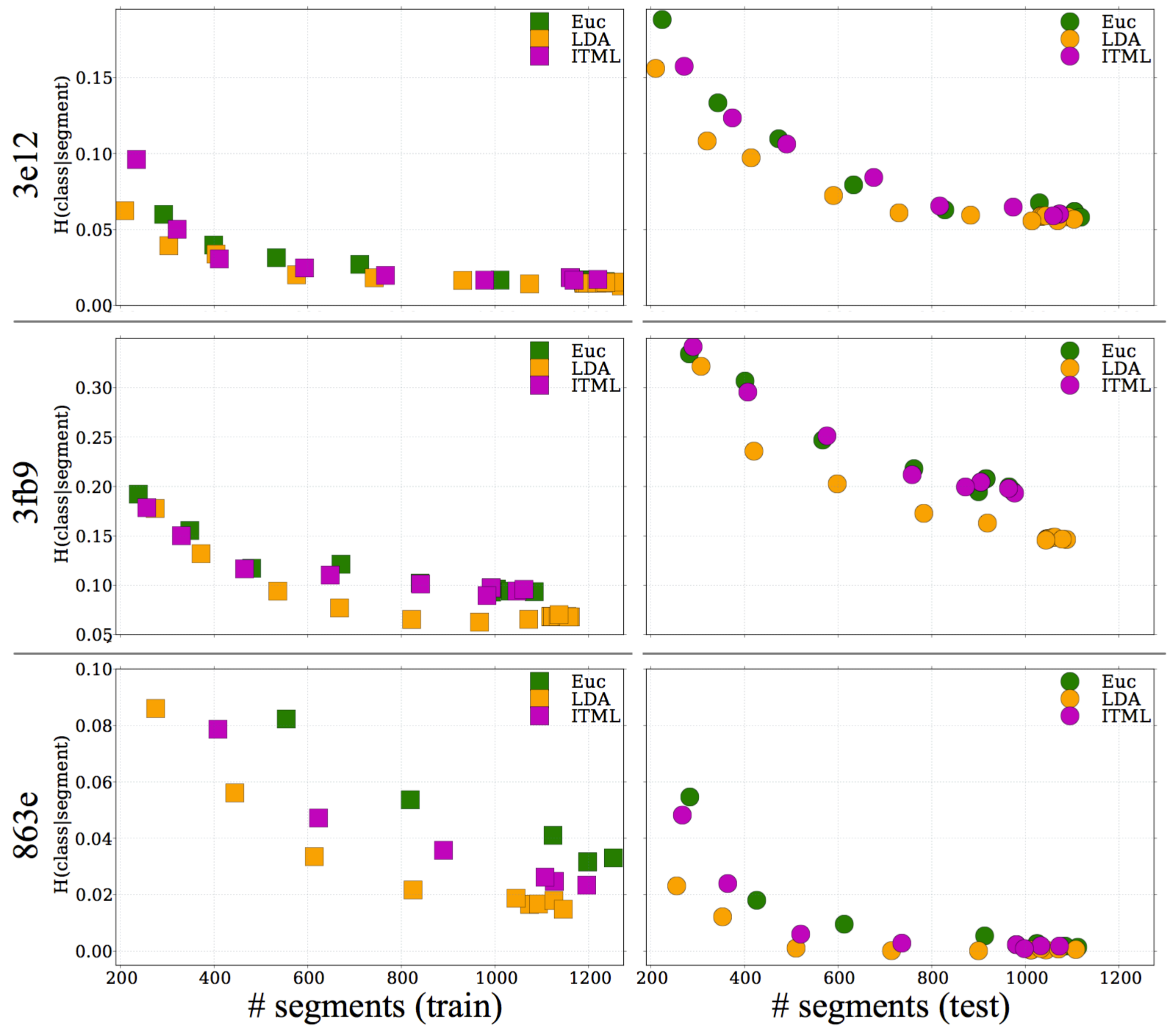


863e Class Map | Euc (819 segments) | LDA (825 segments) | ITML (890 segments)

Image 863e Segment Purity Scores

Class (# pixels)	Pure Segments (%) Euc	LDA	ITML
FeMg Smectite (6443)	26	49	48
Kaolinite (4051)	98	99	99
Montmorillonite (10901)	11	31	17
Nontronite (4753)	37	52	40
Neutral Region (115225)	97	99	98
Average (141373)	53	66	60

$H(C|S)$ Results: Images 3e12, 3fb9, 863e



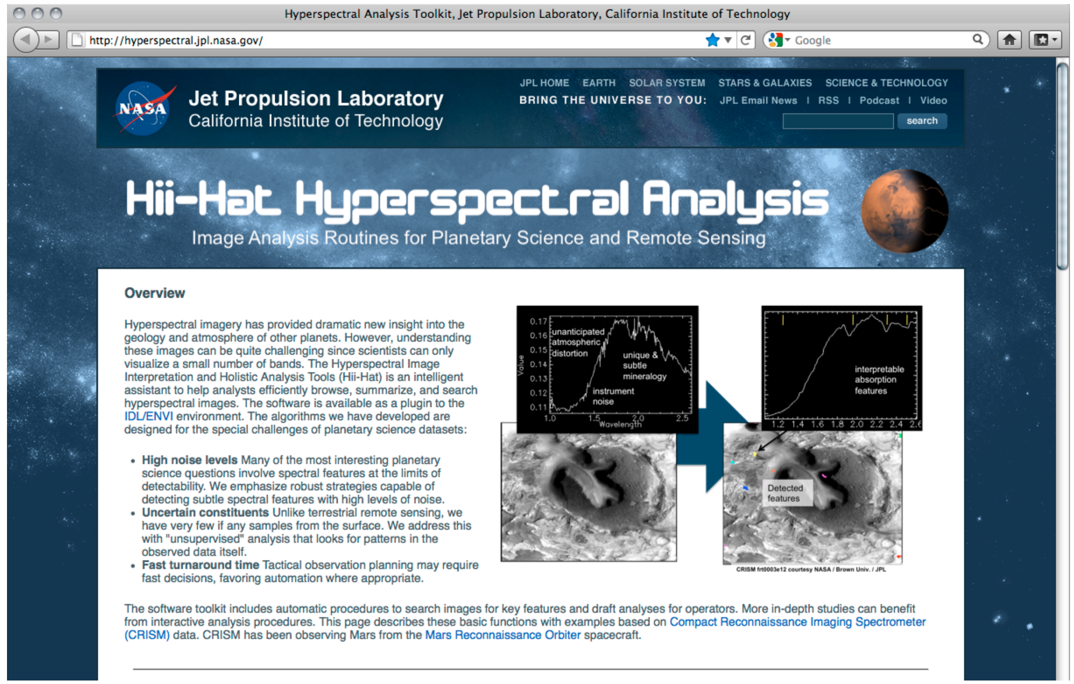
Conclusions / Future Work

- Learned metrics can significantly improve the quality of hyperspectral segmentation results
- Simple techniques (e.g., LDA) with only a few training samples often outperform more computationally expensive metric learning methods
    - Additional samples may improve ITML accuracy
- Future work: comparison to additional metric learning algorithms (neighborhood components analysis, relevant components analysis), and transfer learning scenarios
    - Initial results indicate LDA competitive with state of the art Mahalanobis metric learning algorithms

HiiHAT IDL/ENVI Toolkit

http://hyperspectral.jpl.nasa.gov

- ENVI Toolkit for hyperspectral image analysis
- Includes superpixel segmentation, endmember detection, feature enhancement and metric learning functions



- Free, non-commercial research licenses available

References

- Ronald Fisher (1936) The Use of Multiple Measurements in Taxonomic Problems In: Annals of Eugenics, 7, p. 179--188
- Jason V. Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S. Dhillon. "Information-theoretic Metric Learning." Proc. 24th International Conference on Machine Learning (ICML), 2007.
- David R. Thompson and Lukas Mandrake and Martha Gilmore and Rebecca Castano. Superpixel Endmember Detection. *Transactions on Geoscience and Remote Sensing*, 48(11):4023-4033, Nov. 2010.
- J. Goldberger, S. Roweis, G. Hinton, R. Salakhutdinov. (2005) Neighbourhood Component Analysis. Advances in Neural Information Processing Systems. 17, 513-520.
- A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, "Learning Distance Functions using Equivalence Relations," Proc. International Conference on Machine Learning (ICML), 2003, pp. 11-18.