METRIC LEARNING FOR HYPERSPECTRAL IMAGE SEGMENTATION

Brian D. Bue[1],   *David R. Thompson*[2],   *Martha S. Gilmore*[3],   *Rebecca Castaño*[2]

[1] Rice University, Electrical and Computer Engineering, 6100 Main St., Houston, TX, 77006

[2] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, 91109

[3] Wesleyan University, Earth and Environmental Sciences, Weslayan Station, Middletown, CT 06459

ABSTRACT

We present a metric learning approach to improve the performance of unsupervised hyperspectral image segmentation. Unsupervised spatial segmentation can assist both user visualization and automatic recognition of surface features. Spatially-continuous segments can improve noise levels and localization of feature boundaries. However, existing segmentation approaches rely on generic measures of spectral similarity. Here we learn domain-specific distance metrics based on training data, improving segment fidelity to semantic categories of interest. Multiclass Linear Discriminant Analysis produces a linear transform that optimally separates a labeled set of training classes. This defines a distance metric that generalizes to new scenes, enabling graph-based segmentations that emphasize key spectral features. We describe tests on data from the Compact Reconnaissance Imaging Spectrometer (CRISM) in which learned metrics improve segment homogeneity with respect to mineralogical categories.

Index Terms— Segmentation, Metric Learning, CRISM

1. HYPERSPECTRAL IMAGE SEGMENTATION

Unsupervised hyperspectral image segmentation can reveal spatial trends that show the physical structure of the scene to an analyst. They highlight borders and reveal areas of homogeneity and change. Segmentations are independently helpful for object recognition, and assist with automated production of symbolic maps. Additionally, a good segmentation can dramatically reduce the number of effective spectra in an image, enabling analyses that would otherwise be computationally prohibitive. Specifically, using an oversegmentation of the image instead of individual pixels can reduce noise and potentially improve the results of statistical post-analysis [1].

Typical segmentation methods for hyperspectral imagery include the watershed transform [2], Markov Random Fields [3], and the Felzenszwalb graph segmentation algorithm [4]. Generally speaking, they cluster pixels based on spatial proximity and a measure of spectral similarity given as a distance metric or other pairwise function. Existing hyperspectral segmentation approaches generally use generic distance measures that treat all channels equally or weight channels based on global statistical properties of the dataset. Such metrics are often confused by noise, instrument artifacts, or spectral variations that are irrelevant to semantic categories of interest.

This work aims to improve automatic segmentations by learning a task-relevant measure of spectral distance from expert-labeled training data. We employ a multiclass Linear Discriminant Analysis (LDA) based approach to learn this measure. It produces segmentations that are not only are more visually cohesive, but also quantiatively more accurate in separating known materials into disjoint segments, in comparison to segmentations produced using the Euclidean metric. We evaluate this technique by comparing a set of expert-labeled mineral class maps to the segmentation maps produced by learned metrics, and provide a results on a case study focusing on several well-analyzed CRISM images [5]

2. METRIC LEARNING FOR HYPERSPECTRAL IMAGE SEGMENTATION

Our segmentation strategy employs the Felzenszwalb algorithm for its simplicity and computational efficiency [4]. The approach was first developed for computer vision applications and recently applied to hyperspectral data [1]. It represents the hyperspectral image as an 8-connected grid of pixels initialized as independent segments. Edges between nodes in the grid represent distances between neighboring pixels according to some pairwise similarity function. The algorithm iteratively joins neighboring pixels together into larger segments, and describes each segment by the minimum spanning tree of edges that joins all pixels in the segment. For nodes v_i and v_j, we define the maximum internal edge weight of a cluster S to be $\text{Int}(S)$ which is the largest edge weight in the segment's minimum spanning tree, $\text{MST}(S)$.

$$\text{Int}(S) \quad = \quad \max_{v_i, v_j} d(v_i, v_j) \, \forall \, v_i \in S, \, v_j \in S,$$
$$(v_i, v_j) \in \text{MST}(S) \quad (1)$$

The distance between two neighboring segments S_a and S_b is given by the smallest edge weight that joins them, i.e. the most similar pixel pair on their border.

$$\text{Dif}(S_a, S_b) \quad = \quad \min_{v_i, v_j} d(v_i, v_j) \, \forall \, v_i \in S_a, v_j \in S_b,$$
$$(v_i, v_j) \in E \quad (2)$$

The merge criterion tests this cross-segment distance to see if it is larger than the minimum of both internal weights, weighted by a constant k and inversely proportional to a segment's area $|S|$:

$$\text{MInt}(S_a, S_b) = \min \left(\text{Int}(S_a) + \frac{k}{|S_a|}, \text{Int}(S_b) + \frac{k}{|S_b|} \right) \quad (3)$$

Finally, any small remaining regions below a minimum size threshold are merged with their spectrally-closest neighbors. The resulting segmentation is fast to compute and accomodates any desired measure of spectral distance. Previous studies have used spectral angle distance and Euclidean (Euc) distance to compute the segmentation. We attempt a *superpixel* segmentation in which the image is conservatively oversegmented; that is, we accept that single surface features may be split into multiple segments, but try to ensure that each individual segment - or superpixel - has homogeneous mineralogy [1].

In this work we augment the segmentation algorithm with a task-specific Mahalanobis distance metric learned from training data. The Mahalanobis distance between samples $\{\mathbf{x}_i, \mathbf{x}_j\} \in \mathbb{R}^d$ is $d(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{M}(\mathbf{x}_i - \mathbf{x}_j)$, where $\mathbf{M} = \mathbf{A}^T\mathbf{A}$ is a $d \times d$ linear transformation matrix. We seek to learn the matrix $\mathbf{A}$ which best separates a set of N samples belonging to C classes. Our approach employs multiclass linear discriminant analysis (LDA) to maximize the ratio of between-class vs. within-class separation S, defined by the Rayleigh quotient

$$S = \frac{\boldsymbol{\alpha}^T \boldsymbol{\Sigma}_b \boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \boldsymbol{\Sigma}_w \boldsymbol{\alpha}} \qquad (4)$$

where $\boldsymbol{\Sigma}_b$ and $\boldsymbol{\Sigma}_w$ are the between and within class scatter matrices, respectively. By selecting the top $C - 1$ eigenvectors of $\boldsymbol{\Sigma}_w^{-1}\boldsymbol{\Sigma}_b$ to define the $d \times C - 1$ transformation matrix $\mathbf{A}$, we define a projection for d-dimensional spectra into a $C - 1$ dimensional subspace that captures the variability between features with respect to training data [6]. As the number of training samples is often less than the dimensionality of the data – which causes the maximization of Equation 4 to become ill-posed – we regularize $\boldsymbol{\Sigma}_w$ by a parameter $\gamma_{\mathrm{LDA}} \in [0, 1]$ according to: $\boldsymbol{\Sigma}_w = (1 - \gamma)\boldsymbol{\Sigma}_w + \gamma \mathrm{I}^C$, where $\mathbf{I}^C$ is the $C \times C$ identity matrix. We select γ_{LDA} via cross-validation.

For comparison, we also learn a Mahalanobis distance via the technique proposed by Davis et al. in [7], Information Theoretic Metric Learning (ITML). ITML calculates the matrix $\mathbf{M}$ by maximizing the relative entropy between a multivariate Gaussian parameterized by a set of training samples, and another multivariate Gaussian belonging to a known, well-behaved Mahalanobis distance function. This maximization is constrained such that similar classes remain nearby one another and dissimilar classes remain far apart in the space defined by the learned metric. The ITML algorithm takes a parameter γ_{ITML} which controls the tradeoff between satisfying the similarity/dissimilarity constraints and maximizing the divergence between the Gaussians. We learn the metric and select this parameter via cross-validation, using the code provided by the authors [8].

3. EVALUATING SEGMENTATION RESULTS WITH RESPECT TO CLASS KNOWLEDGE

Hyperspectral segmentation algorithms partition images into spectrally homogenous regions. However, the exact definition of homogeneity is dependant on the chosen similarity metric. By leveraging a small set of labeled pixels with known mineralogical interpretations, we learn a metric that suppresses uninformative spectral content. To assess the quality of the segmentation with respect of known material species, we compare superpixels produced using the selected metric to a set of expert-labeled classes defined by a planetary geologist. The geologist identified the primary constituents in each of image, along with the image pixels containing the purest examples of each mineral, and defined class maps for each material using the ENVI spectral angle mapper function [9]. A minor fraction of image pixels contain ambiguous or mixed materials. We mark these pixels as unclassified and exclude them from the following performance evaluation.

We define two measures to quantify the degree which the resulting segments partition distinct mineralogical species. The first measure is the conditional entropy of the class map given the segmentation map, H(class|segment). The conditional entropy $H(X|Y)$ quantifies the amount of remaining uncertainty for a random variable X (in our case, the distribution of material classes), given the value of another random variable Y is known (here, the partitioning produced by a segmentation algorithm). In the case of a perfect segmentation of the classes, H(class|segment) will be zero, as the

segmentation perfectly reconstructs the class map. Thus, we prefer smaller values of H(class|segment). Our second measure of segmentation quality, the "impurity ratio," is the ratio of "impure" vs. "pure" segments with respect to the class map. A "pure" segment consists of pixels belonging to a single class, whereas an "impure" segment consists of pixels belonging to multiple classes. Because the size of a segment can bias this score, we scale the impurity ratio for each segment by its size in pixels. As with the H(class|segment) measure, smaller impurity ratios are better.

We evaluate the quality of segmentations produced by each metric learning algorithm by segmenting spatially contiguous halves of each image. We sample 100 spectra from each class from the first half of the image (subsequently referred to as the "train" image), and use these points to train each metric learning algorithm. We then segment the train image and the remaining half of the image (the "test" image) using the Euclidean distance as a baseline in addition to the learned metrics. In order to objectively compare results between several metrics, we must compare segmentations that produce a similar number of superpixels. For a given distance metric, the internal bias k (Equation 3) alters the size – and subsequently the quantity – of the resulting superpixels. To acheive this balance, we segment each image using a range of k values in $[10^{-3}, 10^1]$ and provide overall statistics for each algorithm on that range. We chose this range because the number of superpixels produced by each metric followed a similar trend for all of the images we studied. Also, we focus on segmentations that produce 200-1250 superpixels, as segmentations with few superpixels tend to inadequately capture morphological characteristics of our imagery, while segmentations with large quantities of superpixels are more sensitive to noise and insignificant differences in spectra. Finally, we ignore superpixels consisting of less than 50 pixels, as they tend to be unstable and noisy with respect to the training classes.

4. CASE STUDY: CRISM IMAGERY

We examine three well-studied CRISM scenes: 3e12, 3fb9, and 863e (omitting the frt0000 catalog prefix). We use the Brown CRISM Analysis Toolkit [10] to perform radiometric correction and atmospheric calibration, and remove noisy bands in the extreme short and long wavelengths, leaving a total of 231 bands in the in the [1.06, 2.58] μm range for analysis. Figure 1 gives the mean spectra of the most pure material samples for the classes within each image we consider in this work. See [1] for further details regarding these images and their constituent material classes.

Figures 2 and 3 give the H(class|segment) and purity scores, respectively, vs. the number of segments produced by each metric. LDA outperforms both the baseline Euclidean metric and ITML, occasionally dramatically (e.g. on images 863e and 3fb9). The Euclidean metric performs worst, which is not surprising since it is more susceptible to noise that a learned metric can suppress. ITML yields about the same performance as the Euclidean distance for train images 3e12 and 3fb9, which is likely due to the fact that these images only contain two and three material classes, respectively. The quantity of training samples is small for these two images, and ITML inadequately determines which spectral bands are the most prominent. However, ITML still exhibits improved generalization performance on test data over the baseline Euclidean distance, indicating that some noise characteristics are potentially captured. This is reflected in the summary statistics per-image for each segmentation metric given in Table 1, as well.

Figure 4 shows a set of resulting segmentation maps for which the Euclidean metric, and LDA/ITML-learned metrics produced a
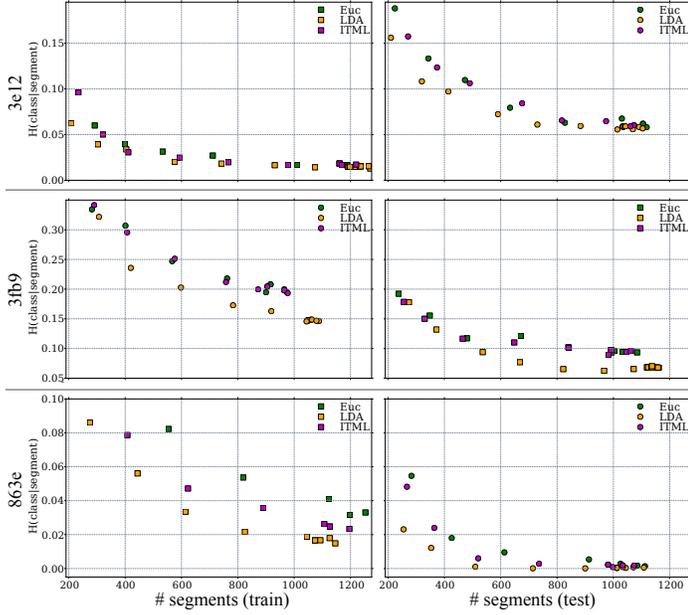
Fig. 2. H(class|segment) values for Euclidean (green markers), LDA (yellow markers) and ITML (magenta markers) based segmentations vs. number of segments on train image (left) and testing image (right).
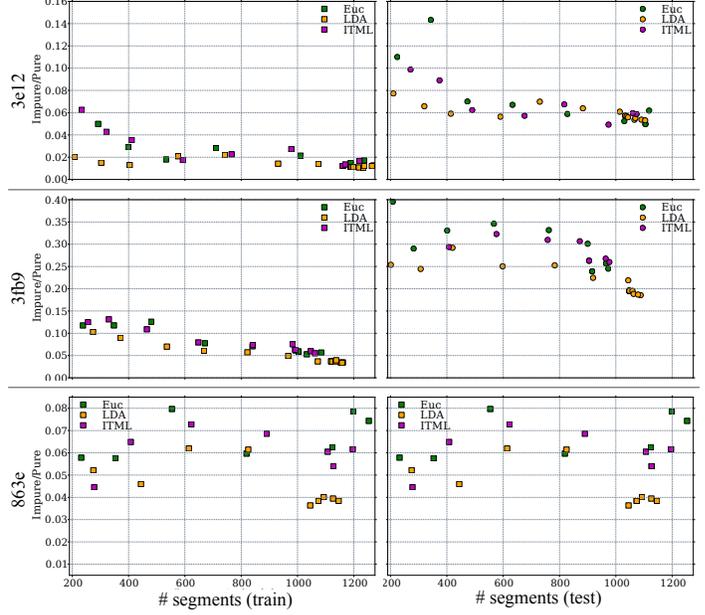

Fig. 3. Impure/Pure ratio values for Euclidean (green markers), LDA (yellow markers) and ITML (magenta markers) based segmentations vs. number of segments on train image (left) and testing image (right).

		Euc	LDA	ITML
863e	H(class|segment)	0.0473 / 0.00403	0.0184 / 0.000584	0.031 / 0.00228
	Purity	0.0684 / 0.032	0.0398 / 0.0124	0.0611 / 0.0266
3e12	H(class|segment)	0.0169 / 0.0676	0.0148 / 0.0588	0.0191 / 0.0655
	Purity	0.018 / 0.0619	0.0116 / 0.0573	0.02 / 0.0596
3fb9	H(class|segment)	0.0884 / 0.378	0.0497 / 0.242	0.0972 / 0.354
	Purity	0.0661 / 0.296	0.0368 / 0.195	0.0745 / 0.294

Table 1. Average H(class|segment) and Purity Statistics Per-Image. Green and red fonts indicate the best and worst performing metric, respectively
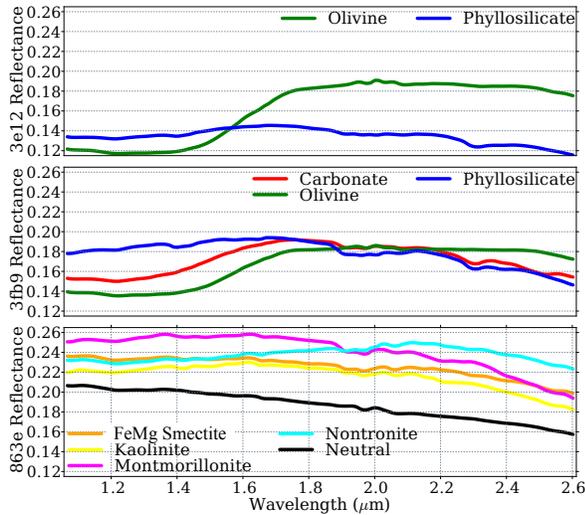

Fig. 1. Mean spectra of samples from most pure material classes in images 3e12, 3fb9 and 863e. The "neutral" class in image 863e is a mostly featureless, dark material which is spectrally dissimilar from each of the other material species.

comparable number of segments. Visually, the LDA-based segmentation produces segments that better match the underlying morphology of the image data. The Euclidean-based segmentation, and to a lesser degree, the ITML-based segmentation, both suffer significantly from column striping artifacts.

Table 2 gives the percentages of pure segments for each material species for the three segmentation metrics. Both learned metrics outperform the baseline, with LDA improving over the Euclidean metric for material classes FeMg Smectite, Montmorillonite and Nontronite. ITML gives comparable performance to LDA for most materials, but the gains are not as significant for the Montmorillionite and Nontronite classes.

5. DISCUSSION AND FUTURE WORK

The superior performance of LDA over ITML is somewhat surprising, considering the simplicity of the LDA projection in comparison to the expectedly more robust optimization performed by ITML. An issue with ITML (as observed by Parameswaran et al. in [11]) is that the (global) metric is not optimized locally, which can cause problems with overfitting to multi-modal data distributions. Conversely, (regularized) LDA does not suffer from such overfitting issues. Alternative regularization schemes may improve ITMLs generalization capabilities.
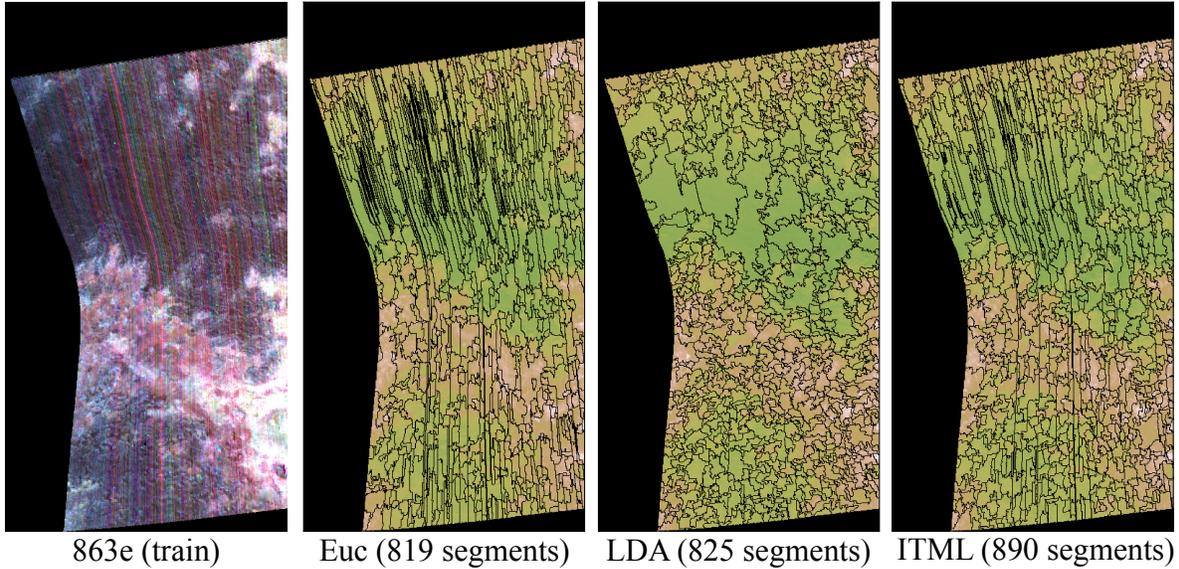
863e (train)   Euc (819 segments)   LDA (825 segments)   ITML (890 segments)

Fig. 4. False-colored composite image of image 863e (left), and the segmentation maps produced using Euclidean distance (left), LDA (center), and ITML (right). We overlay each segmentation map on a single color-scaled band from image 863e (wavelength 1.25μm) illustrate image morphology. The LDA-based segmentation is less susceptible to column striping artifacts, and better characterizes image morphology.

Class (# pixels)	Euc	LDA	ITML
FeMg Smectite (9628)	26	49	48
Kaolinite (88642)	98	99	99
Montmorillonite (2029)	11	31	17
Nontronite (1836)	37	52	40
Neutral Region (112415)	97	99	98
Average	53	66	60

Table 2. Pure pixels / segment for Euclidean, LDA and ITML-based segmentations shown in Figure 4. Best and worst average per-class accuracy given in green and red font, respectively.

One avenue which we are currently exploring is the potential for learning class structure across multiple, related images. We have developed an technique, Multi-Domain/Multi-Class LDA (MDMC-LDA) and a corresponding regularization scheme which asllows LDA to exploit class structure local to an individual image while simultaneously capturing class relationships common to other images with related classes [12].

6. REFERENCES

[1] D. R. Thompson, L. Mandrake, M.S. Gilmore, and R. Castaño, "Superpixel Endmember Detection," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 11, pp. 4023–4033, 2010.

[2] Y. Tarabalka, J. Chanussot, and J.A. Benediktsson, "Segmentation and classification of hyperspectral data using watershed transformation," *Pattern Recognition*, vol. 43, no. 7, pp. 2367–2379, 2010.

[3] A. Mohammadpour, O. Féron, and A. Mohammad-Djafari, "Bayesian segmentation of hyperspectral images," *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, vol. 735, pp. 541–548, 2004.

[4] Pedro F. Felzenszwalb and Daniel P. Huttenlocher, "Efficient graph-based image segmentation," *Intl. J. Computer Vision*, vol. 59:2, September 2004.

[5] S. Murchie et al., "CRISM (Compact Reconnaissance Imaging Spectrometer for Mars) on MRO (Mars Reconnaissance Orbiter)," *J. Geophys. Res*, vol. 112, no. E05, 2007.

[6] R.A Fisher, "The statistical utilization of multiple measurements.," *Annals of Eugenics*, vol. 8, pp. 376–386, 1938.

[7] J Davis, B Kulis, P Jain, S Sra, and I Dhillon, "Information-theoretic metric learning," *Proceedings of the 24th international conference on Machine learning*, Jan 2007.

[8] Jason V. Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S. Dhillon, *Information Theoretic Metric Learning*, UT, Austin, http://www.cs.utexas.edu/users/pjain/itml/.

[9] Research Systems Inc, *ENVI 4.6 Users Guide*, 2008, 1196 pp.

[10] F. Morgan, F. Seelos, and S. Murchie, "Cat tutorial," in *CRISM Data Users Workshop, Lunar Planetary Sci. Conf.*, 2009.

[11] S Parameswaran and Kilian Q Weinberger, "Large margin multi-task metric learning," *Proceedings of NIPS 2010*, 2010.

[12] D. Hayden, S. Chien, D. Thompson, and R. Casta no, "Using clustering and metric learning to improve science return in remotely sensed imagery," in *In Proceedings, ACM Transactions on Intelligent Systems and Technology*. 2011, ACM, (submitted).