

# An Adaptive Similarity Measure for Classification of Hyperspectral Signatures

Brian D. Bue and Erzsébet Merényi, *Senior Member, IEEE*

**Abstract**—Capturing both the shape of the spectral continuum and the positions/widths of absorption bands is essential to accurately measure similarity between hyperspectral signatures. Furthermore, the relative importance of these features is data dependent. In this letter, we present an adaptive version of our recently proposed continuum-intact (CI)/continuum-removed (CR) similarity measure which automatically determines a convex weighting between similarity measurements of CI and CR signatures according to input data. We describe an efficient technique to calculate an optimal weight for a linear combination of CI and CR similarity measurements. We evaluate the technique on the Airborne Visible/Infrared Imaging Spectrometer spectra sampled from a well-studied urban scene and show that our technique yields improved classification accuracy in comparison to CI or CR similarity measurements alone and performs comparably to calculating the weight via brute-force search, at a much reduced computational cost. A source code implementation of our algorithm is provided online.

**Index Terms**—Adaptive, continuum removal, hyperspectral, linear discriminant analysis (LDA), metric learning, similarity measure.

## I. SPECTRAL SIMILARITY MEASURES

THE widespread deployment of hyperspectral imaging systems on both terrestrial and planetary orbiters allows for highly detailed analysis of large-scale regional surveys. Improvements in both spectral and spatial resolutions of these systems drive the innovation of advanced image classification techniques. Such image classification techniques have proven to be an invaluable tool for analysts, as the sheer volume of the collected data renders exhaustive manual image interpretation impossible. Increasing data volumes demand algorithms which not only are sophisticated enough to yield high classification accuracies in various scenarios but which are also efficient, capable of processing large hyperspectral data sets quickly, for example, in ground-based archives or in resource-constrained/real-time onboard applications.

A problem central to image classification is assessing the similarity between spectral signatures (pixels), which is challenging due to numerous factors such as data dimensionality, noise, and environmental effects. Employing adaptive similarity measures that exploit knowledge of known classes can

Manuscript received February 1, 2012; revised May 21, 2012; accepted June 14, 2012. Date of publication August 10, 2012; date of current version October 22, 2012.

B. D. Bue is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 USA (e-mail: bbue@rice.edu).

E. Merényi is with the Department of Statistics, Rice University, Houston, TX 77005 USA (e-mail: erzsebet@rice.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LGRS.2012.2206011

help mitigate these effects and improve classification accuracy, potentially with fewer training samples than a task-agnostic measure may require to achieve the same accuracy. Robust spectral similarity measures must capture both the shape of the spectra and the positions/widths of absorption bands. Most widely used measures assume that all spectral bands are of equal importance, without specific emphasis on absorption features. These measures take *continuum-intact* (CI) reflectance signatures as input, which often poorly capture differences in absorption features. To compensate for this, one can employ measures which characterize absorption features using *continuum-removed* (CR) signatures. The CR representation measures spectral absorption features by “dividing out” a spectral continuum curve. The CR representation often better captures the composition and concentration of the material that a spectral signature represents than its CI counterpart. However, since signatures with considerably different continua can have equivalent CR representations (as observed in [1]–[3]), the CR representation alone can be an unreliable descriptor in classification settings.

In this letter, we present an adaptive similarity measure designed specifically for hyperspectral signatures. Our measure  $d_{\text{CICR}}$  finds a compromise between the use of the continuum shape and absorption features by calculating a scalar weight  $\alpha$  in a convex combination of CI and CR distance measurements that best separates a set of training classes. When spectral signatures contain significant absorption features, our measure better discriminates material classes in comparison to measures considering CI or CR signatures alone. We evaluate classification accuracy on real Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) imagery from a well-studied urban area, in scenarios consisting of prominent or subtle absorption features.

## II. CICR MEASURE

Given a pair of hyperspectral signatures  $\mathbf{x}_i$  and  $\mathbf{x}_j$ ,<sup>1</sup> each in  $\mathbb{R}^D$ , we define the  $d_{\text{CICR}}$  similarity measure as follows:

$$d_{\text{CICR}}(\mathbf{x}_i, \mathbf{x}_j, \alpha) = (1 - \alpha)d_{\text{CI}}(\mathbf{x}_i, \mathbf{x}_j) + \alpha d_{\text{CR}}(\mathbf{x}_i, \mathbf{x}_j) \quad (1)$$

with

$$d_{\text{CI}}(\mathbf{x}_i, \mathbf{x}_j) = \left\| \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|} - \frac{\mathbf{x}_j}{\|\mathbf{x}_j\|} \right\| \quad (2)$$

$$d_{\text{CR}}(\mathbf{x}_i, \mathbf{x}_j) = \left\| \frac{\text{CR}(\mathbf{x}_i)}{\|\text{CR}(\mathbf{x}_i)\|} - \frac{\text{CR}(\mathbf{x}_j)}{\|\text{CR}(\mathbf{x}_j)\|} \right\|. \quad (3)$$

<sup>1</sup>In this and subsequent sections, we denote scalar variables in italics, vectors in bold, and matrices in bold caps.

Here,  $\|\cdot\|$  is the  $L^2$  norm,  $\alpha \in [0, 1]$  is a weighting parameter, and  $\text{CR}(\cdot)$  performs continuum removal. We estimate the continuum of a given spectrum by fitting a piecewise linear function to the local maxima using the procedure described in [4]. Observations on the continuum are assigned values of zero, and absorption features (observations between local maxima) are assigned values in the  $[0, 1]$  range, proportional to their relative distances from the estimated continuum. Because the continuum removal procedure is sensitive to spurious local maxima, we smooth each signature using a moving average filter before performing continuum removal. While smoothing may mask small absorption features, such features are often close to the noise floor of the sensor, and we accept this loss in specificity in favor of noise reduction. In our experiments using AVIRIS data, smoothing windows ranging from three to five bands ( $0.03\text{--}0.05 \mu\text{m}$ ) have performed well.

Dividing the CI and CR signatures by the  $L^2$  norm is important for several reasons. First, scaling each signature by its norm maps both CI and CR distances to a common range. This allows us to better tune the weight  $\alpha$  to combine the CI and CR similarity measurements according to input data. Second, for CI signatures, this scaling mitigates linear effects caused by differing illumination conditions (while preserving spectral angles). Lastly, scaling CR signatures by their respective norms can accentuate discriminative absorption features and can improve performance when classifying CR signatures.

The measure described in this work differs from our original formulation in [4]. First, the convex combination of CI and CR terms yields more consistent performance than applying  $\alpha$  to the CR term only. Second, due to the nature of continuum estimation, CR signatures contain many values near zero, which provides little discriminating information between signatures when combined with the CI distance measure. Scaling the CR signatures uniformly by the variance of all CR distances (as we originally described) does not resolve this issue, whereas scaling each signature by its norm allows the most prominent absorption features to play greater roles in discrimination.

### III. COMBINING CI AND CR DISTANCE MEASUREMENTS

Fig. 1 shows an overview of the methodology that we use to calculate the weight parameter  $\alpha$  in (1), which we describe in detail as follows.

Given a set of  $N$  vectors  $\{\mathbf{x}_i\}_{i=1}^N$ ,  $\mathbf{x}_i \in \mathbb{R}^D$  belonging to  $K$  classes, with labels  $y_i \in [1, K]$ , we calculate  $\alpha$  using a method inspired by multiclass linear discriminant analysis (LDA) ([5] and [6]). Multiclass LDA computes the vector  $\mathbf{w}$  that maximizes the Rayleigh quotient (using the formulation given in [7])

$$S = (\mathbf{w}^T \mathbf{M}_B \mathbf{w}) (\mathbf{w}^T \mathbf{M}_W \mathbf{w})^{-1} \quad (4)$$

where  $\mathbf{M}_B$  and  $\mathbf{M}_W$  are (symmetric, positive-definite) between-class separation and within-class scatter matrices. We form the  $\mathbf{M}_B$  and  $\mathbf{M}_W$  matrices according to the capability of each of the  $\{d_{CI}, d_{CR}\}$  measures to separate the given classes.

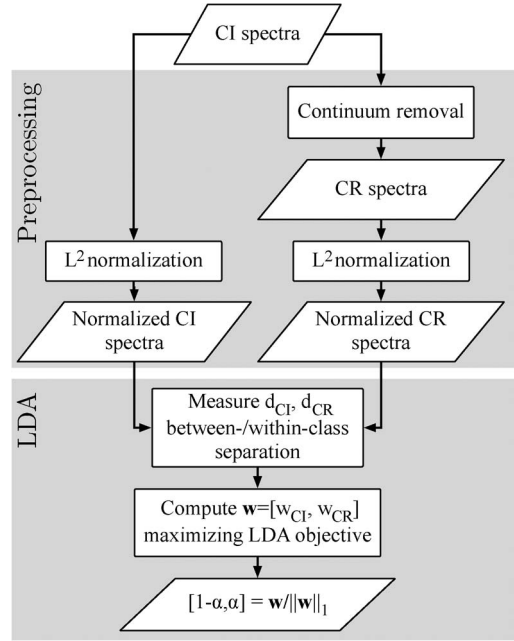


Fig. 1. Processing steps for calculating  $d_{CICR}$  weight parameter  $\alpha$ .

Specifically, we define the between-class separation  $s_b$  and within-class scatter  $s_w$  between measures  $d_1$  and  $d_2$  as

$$s_b(d_1, d_2) = \sum_{j=1}^K N_j d_1(\boldsymbol{\mu}_j, \bar{\boldsymbol{\mu}}) d_2(\boldsymbol{\mu}_j, \bar{\boldsymbol{\mu}}) \quad (5)$$

$$s_w(d_1, d_2) = \sum_{j=1}^K \sum_{i: y_i=j} d_1(\mathbf{x}_i, \boldsymbol{\mu}_j) d_2(\mathbf{x}_i, \boldsymbol{\mu}_j) \quad (6)$$

where  $\{\boldsymbol{\mu}_j\}_{j=1}^K$  denotes the mean vectors of each of the  $K$  classes,  $\bar{\boldsymbol{\mu}}$  is the mean of  $\boldsymbol{\mu}_j$ , and  $N_j$  is the number of samples in class  $j$ . We then form the  $\mathbf{M}_B$  and  $\mathbf{M}_W$  matrices as follows:

$$\mathbf{M}_B = \frac{1}{N} \begin{bmatrix} s_b(d_{CI}, d_{CI}) & s_b(d_{CI}, d_{CR}) \\ s_b(d_{CI}, d_{CR}) & s_b(d_{CR}, d_{CR}) \end{bmatrix} \quad (7)$$

$$\mathbf{M}_W = \frac{1}{N} \begin{bmatrix} s_w(d_{CI}, d_{CI}) & s_w(d_{CI}, d_{CR}) \\ s_w(d_{CI}, d_{CR}) & s_w(d_{CR}, d_{CR}) \end{bmatrix}. \quad (8)$$

The first (largest) eigenvector of  $\mathbf{M}_W^{-1} \mathbf{M}_B$ ,  $\mathbf{w}$ , maximizes (4), with separation  $S$  equal to the corresponding eigenvalue [7]. The components of  $\mathbf{w} = [w_{CI}, w_{CR}]$  provide a weighting of the CI and CR distances with good class separation on training data, which is not necessarily convex [as we require in (1)], and may not generalize well to test data. Because Rayleigh quotients are invariant with respect to scaling of  $\mathbf{w}$  [i.e., for any  $c > 0$ ,  $c\mathbf{w}$  also maximizes (4)] [8], we can scale the components of  $\mathbf{w}$  to a convex range by dividing each component by  $\|\mathbf{w}\|_1$ . This yields the convex pair  $\{w_{CI}/\|\mathbf{w}\|_1, w_{CR}/\|\mathbf{w}\|_1\} = \{(1-\alpha), \alpha\}$ , as desired.

As (4) may become ill posed, we regularize the within-class scatter matrix via a shrinkage operator

$$\mathbf{M}'_W = (1-\lambda)\mathbf{M}_W + \lambda\mathbf{I} \quad (9)$$

where  $\lambda \in (0, 1)$  is a regularization parameter and  $\mathcal{I}$  is the  $(2 \times 2)$  identity matrix. In practice, we select  $\lambda$  via cross-validation, using the methodology described in the next section. We provide implementations of our LDA-based algorithm and the continuum removal algorithm that we used at the following url: <http://www.ece.rice.edu/~bdb1/#code>.

#### IV. CASE STUDY: AVIRIS SPECTRAL SIGNATURES

The starting point of the work described here is a set of reflectance spectra sampled across distinct material species from a low-altitude AVIRIS image of Ocean City, MD. This image (acquired on November 5, 1998, with 4-m/px spatial resolution, in 220 spectral bands from 0.4–2.5  $\mu\text{m}$ ) was analyzed in previous work to capture spectral clusters, verify them against field knowledge, and identify the materials that they represent, as reported in detail in [9]. The 35 clusters resulting from [9] guided the extraction of a trustworthy representative subset of spectra for this study, by stratified random sampling across 14 of those 35 clusters for which material identification was unambiguous, and which served the methodology design for this work. The design is explained as follows. The work in [9] performed all the necessary preprocessing of this image, including atmospheric correction, conversion of radiances to reflectances, and normalization to cancel illumination effects. These preprocessing steps are duly described in [9]. For this work, the reflectance spectra were extracted from the already preprocessed Ocean City image. The number of samples per class is constrained by the size of the smallest cluster ( $\approx 145$  samples), and we fix the number of samples per class to 100 for most of our analysis and characterize classification accuracy versus training set size later in this section (Fig. 4).

This work is a deeper analysis of previous work ([3], [4]) that examines three different spectral scenarios specifically constructed to contrast the performance of the proposed method. In the first, all samples contain only *minor* absorptions, where we define a “minor” absorption as one with no CR band depths greater than threshold  $\tau$ ; we use  $\tau = 0.1$  (10% absorption with respect to the continuum) in this work. In this case, we expect similar classification accuracies from the  $d_{\text{CICR}}$  and  $d_{\text{CI}}$  measures, since the CR signatures lack prominent absorption features (and therefore are flat and uninformative). The classes in this scenario consist of asphalt rooftop materials (class A), roads/parking lots (classes I, J, T, W, and h), and dry beach sand (class e). Fig. 2 (top) shows the CI and CR mean signatures for these classes. In the second scenario, all signatures contain one or more *major* absorptions, where we define major absorptions as those with CR values greater than  $\tau$ . Here, we expect a more significant boost in accuracy in comparison to the minor absorption scenario, as the CR signatures are more informative in this case. The subset of data in this scenario consists of vegetation (classes L and M), a tennis court (class C), wet sand (classes O and Q), and composite rooftop materials (classes D and U). Fig. 2 (bottom) shows the CI and CR mean signatures for these classes. The seven spectral species in each of the major and minor absorption categories are relatively “pure” representatives of their respective species. The last scenario, i.e., *combined*, consists of all classes from both major and minor

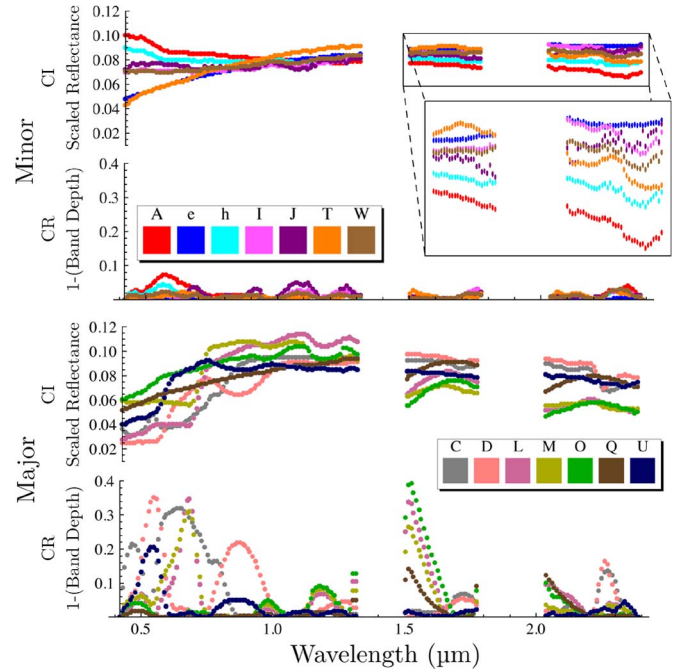


Fig. 2. (Top) Mean CI and CR signatures for minor absorption classes. (Top inset) Detail view of minor absorption signatures, wavelengths 1.5–2.5  $\mu\text{m}$ . (Bottom) Mean CI and CR signatures for major absorption classes. The large disconnected regions near 1.3–1.5 and 1.7–2.0  $\mu\text{m}$  consist of bands removed due to water saturation. CI signatures are scaled by their  $L^2$  norms to compensate for varying illumination conditions.

absorption scenarios. We expect to see notable performance gains with the  $d_{\text{CICR}}$  measure in this case, as both the CI and CR signatures provide information to distinguish between classes.

We present results using a minimum-distance-to-class-means (MinDist) classifier with fivefold random stratified sampling, using 50% for training and the remaining 50% for testing. We calculate  $\alpha$  by maximizing  $S(\alpha)$  as described in Section III. We compare this  $\alpha$  value to the  $\alpha_{\text{LS}}$  value obtained by line search (LS) on a uniformly spaced range of 100 points  $\alpha_{\text{LS}} \in (0, 1)$ , which yields the highest classification accuracy. We calculate classification accuracy according to  $\text{accuracy} = (\# \text{ of True Positives}) / (\# \text{ of Samples})$ , and we report accuracy on test data only. Accuracies produced via LS are an approximate upper bound on achievable accuracy. We choose the regularization parameter  $\lambda$  for each scenario on a hold-out set comprising 50 samples per class, which is not included in the training/testing samples. For our data, we select the  $\lambda$  with the best classification accuracy on the hold-out set from ten uniformly spaced values in  $[0.001, 0.1]$ . We chose this range because smaller  $\lambda$  values tended to yield ill-posed solutions and larger values did not improve classification accuracy in any of the three scenarios—regardless of  $\alpha$ . We calculate  $\lambda$  once for each of the minor, major, and combined scenarios and use the same value for each cross-validation fold. We also reject any  $\lambda$  values which produce solutions to  $\mathbf{M}_W^{-1} \mathbf{M}_B$  with no positive eigenvalues, as such  $\lambda$  values yield rank-deficient  $\mathbf{M}'_w$  (9).

Fig. 3 shows the overall and per-class classification accuracies for  $\alpha \in [0, 1]$ . The vertical magenta dashed line marks the  $\alpha$  value determined by maximizing (4), and the black vertical



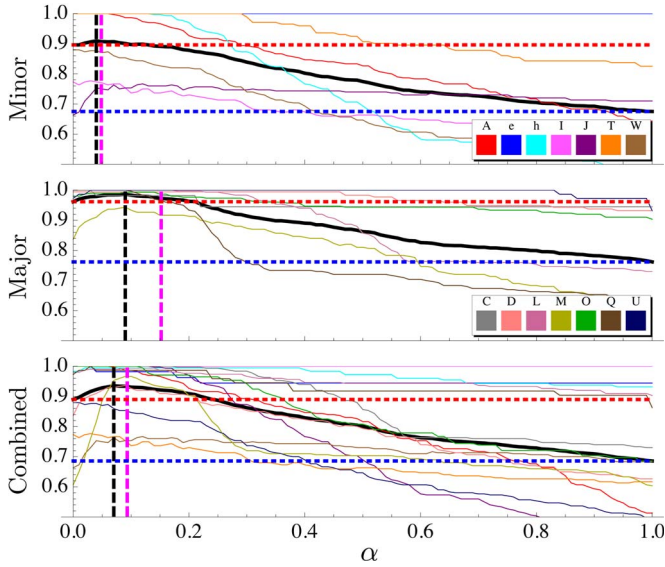


Fig. 3.  $\alpha$  versus per-class  $d_{CICR}$  classification accuracy for (top) minor, (middle) major, and (bottom) combined absorption classes. The colored lines indicate the per-class accuracies, and the black solid line gives the overall classification accuracy. The black vertical bar gives the  $\alpha_{LS}$ , and the magenta vertical bar gives the  $\alpha$ . The horizontal lines give the (red) CI ( $\alpha = 0$ ) and (blue) CR ( $\alpha = 1$ ) classification accuracies. Because the CI representation is generally more informative than the CR one,  $\alpha$  values tend toward zero, but larger values occur in cases when the CR representation provides additional discrimination information (as in the major and combined absorption class scenarios).

TABLE I

AVERAGE CLASSIFICATION ACCURACY OBTAINED WITH EACH OF THE  $d_{CI}$ ,  $d_{CR}$ , AND  $d_{CICR}$  MEASURES SHOWN IN FIG. 3. MEAN AND STANDARD DEVIATION ( $\sigma$ ) OF  $\alpha$  VALUES FOR  $d_{CICR}$  MEASURES ARE GIVEN IN PARENTHESES. THE MOST ACCURATE MEASURE IS GIVEN IN BOLD TEXT

	$d_{CI}$	$d_{CR}$	$d_{CICR}(\alpha \pm \sigma_\alpha)$	$d_{CICR}(\alpha_{LS} \pm \sigma_{\alpha_{LS}})$
<b>Minor</b>	88.5	66.1	<b>90.1</b> (0.0510±0.0375)	90.4 (0.0485±0.0003)
<b>Major</b>	95.9	76.7	<b>97.4</b> (0.1493±0.0041)	98.4 (0.0770±0.0195)
<b>Combined</b>	88.2	66.4	<b>92.4</b> (0.0903±0.0006)	92.6 (0.0670±0.0057)

line gives  $\alpha_{LS}$ . Table I provides the average accuracies for each measure. In all three scenarios, small alpha values ( $< 0.3$ ) yield the highest classification accuracies (although we do not constrain the search to this range). This indicates that, for this data set, CI signatures are more robust descriptors than CR signatures for classification. This is particularly obvious in the minor absorption scenario (Fig. 3, top), where the CR signatures lack discriminative features. Here, classification accuracy using  $d_{CI}$  is close to  $d_{CICR}$ , and both our method and the LS produce  $\alpha$  values near zero.

For major absorption classes (Fig. 3, middle), note that the rate of decrease in classification accuracy is less dramatic as  $\alpha$  approaches one, by comparison to the minor (top) and combined (bottom) absorption scenarios. This indicates that the CR signatures provide additional discriminating information, which increases the  $\alpha$  values, yielding higher classification accuracy. Correspondingly, the maximum separation also shifts toward larger  $\alpha$  values. While  $\alpha$  and  $\alpha_{LS}$  differ the most in this scenario, their corresponding classification accuracies are not far apart (97.4% versus 98.4%). Both are improvements over

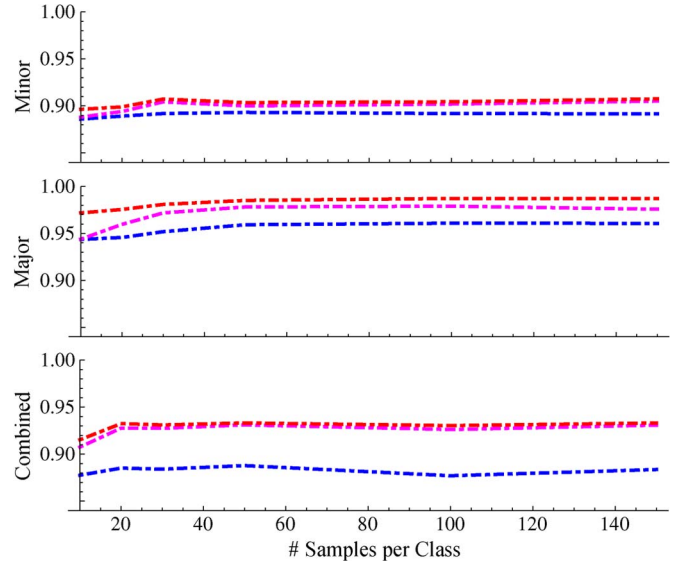


Fig. 4. Classification accuracy versus the number of samples per class for (top) minor absorption, (middle) major absorption, and (bottom) combined scenarios. In each scenario, (magenta line) the LDA-based  $d_{CICR}$  measure outperforms (blue line) the baseline  $d_{CI}$ -based classifier and, with a sufficient number of training samples ( $\sim 20$ -30, scenario dependent), achieves classification accuracy comparable to (red line) LS.  $d_{CR}$ -based classification accuracy not shown above due to significantly lower accuracies ( $\sim 65\%$ - $77\%$ ) in comparison to the  $d_{CI}$ - and  $d_{CICR}$ -based classifiers.

the baseline  $d_{CI}$  accuracy (1.5% and 2.5% relative improvements for our LDA-based  $\alpha$  and  $\alpha_{LS}$ , respectively).

In the combined scenario (Fig. 3, bottom), due to potentially increased class confusion between signatures (compared to the other two scenarios), locating a compromise between the CI and CR terms is challenging. As we see in Fig. 3, the mean classification accuracy for this scenario generally falls between the mean accuracies of the minor and major absorption scenarios. However, we see the most significant improvement, over the baseline  $d_{CI}$  method, in classification accuracy in this scenario (4.5%, by comparison of the thick black line to the horizontal red dashed line in Fig. 3, bottom), versus the other two scenarios, since both the CI and CR representations provide complementary information to discriminate the classes. This is noteworthy given that the CI and CR classification accuracies in the combined scenario are close to those of the minor absorption scenario (88.5% versus 88.2% and 66.1% versus 66.4%, in the combined versus minor scenarios, respectively), yet the relative improvement in the minor scenario is, not surprisingly, lower by 1.7%.

Fig. 4 characterizes the relationship between the number of labeled samples available for training and classification accuracy. As before, classification accuracy is measured over five stratified 50/50 splits into training/test data. In every case, the  $d_{CICR}$ -based classifiers match or outperform the  $d_{CI}$ -based classifier. Additionally, our LDA-based technique for calculating  $\alpha$  performs comparably to brute-force search when a sufficient quantity (about 50/class) of training samples is available. We observe the most significant performance gains of the three scenarios in the combined scenario, where the  $d_{CICR}$  measure can exploit absorption features to separate the classes belonging to the major and minor absorption scenarios and also can capitalize on the absorption characteristics of individual

classes.  $d_{CR}$  performs the worst in all three cases and is not shown in Fig. 4 to emphasize the performance of the better performing measures.

Determining  $\alpha$  using our LDA-based method is significantly less computationally expensive than via brute-force search (i.e.,  $\alpha_{LS}$ ). Quantitatively, assuming  $N$  samples of dimensionality  $D$  belonging to  $K$  classes, we first compute the CR representation of each spectrum using our piecewise linear continuum estimation procedure—an  $O(D)$  operation per spectrum. Then, given the set of (precomputed) class means, a MinDist classifier must compare each signature to each class mean, an  $O(DNK)$  operation. Let  $A$  be the number of values that  $\alpha_{LS}$  can take in  $(0, 1)$  (in this work, we choose  $A = 100$ ). Using brute-force search, we apply the  $O(DNK)$  MinDist classifier  $A$  times. With the LDA-based method, calculating the (symmetric)  $M_B$  involves three  $O(DK)$  operations, calculating  $M_W$  involves three  $O(DN)$  operations, and calculating the eigendecomposition of the  $(2 \times 2)$   $M_W^{-1}M_B$  matrix can be performed in constant time. This amounts to roughly an  $A$ -fold improvement in performance by the LDA-based method over LS. Because  $A$  must be large enough to adequately cover the weight parameter space, our method is an order of magnitude faster than the brute-force search.

## V. DISCUSSION AND FUTURE WORK

As hyperspectral sensors improve in sensitivity and spectral resolution, the choice of spectral similarity measure will play a significant role in resolving subtle differences between classes. We have shown in this letter that, with a small amount of preprocessing, we can improve classification performance over traditional task-agnostic similarity measures using our adaptive hyperspectral domain-specific similarity measure. We have also provided an efficient method to calculate the weight parameter used in our similarity measure which yields classification accuracies within 1% of brute-force search over the range of possible weight values.

Leveraging techniques employed by spectroscopists (i.e., continuum removal) is a natural approach to characterize spectral similarity. However, determining the best manner in which to exploit this information is challenging. First, numerous methods for continuum estimation and removal exist, ranging from fitting piecewise linear functions [1] to more involved techniques that exploit high-order derivatives of spectral signatures [10]. Determining which method performs best in terms of classification accuracy has not been widely explored in the literature (to the best of our knowledge) and is a subject of future work.

We experimented with other methods for learning the  $\alpha$  weighting in the CI/CR (CICR) measure, but our initial experiments in those directions produced mixed results. We believe that an issue with such methods (e.g., [11]) is that the learning problem is transformed from a multiclass classification problem to a binary classification problem where spectra from “similar” versus “dissimilar” classes are grouped into two meta-classes, thereby discarding discrimination information with respect to individual classes. We observed a similar issue with other multiclass metric learning techniques in previous work [12].

When continuum removal detects spurious absorption features,  $L^2$  normalization may exacerbate noise. In such cases, it may be advantageous to adopt alternative normalization schemes to map the CI and CR signatures to a common range, such as scaling signatures by their respective (CI or CR) standard deviations. Using such a “global” scaling factor, we would not accentuate noise on individual signatures. However, this would not provide the same degree of contrast between classes as  $L^2$  normalization provides. The tradeoffs involved with these normalization schemes are a subject of future investigation.

As mentioned in Section III, normalization of spectral signatures plays a significant role in the performance of this technique. In particular, scaling CR signatures by their  $L^2$  norms often improved the classification accuracy by giving additional weight to discriminative absorption features. However, in some cases, this normalization may not be desirable, for example, if the CR signatures fail to capture the most relevant absorptions, perhaps due to noise and/or poorly estimated continua. Other normalization techniques such as scaling by the within-class variances may be beneficial in such cases.

## ACKNOWLEDGMENT

The authors would like to thank Dr. D. Thompson, Dr. K. Wagstaff, and Prof. D. Subramanian for their invaluable input and helpful discussions and Prof. B. Csathó for providing the Ocean City imagery and help with the interpretation of spectral clusters.

## REFERENCES

- [1] R. N. Clark and T. Rousch, “Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications,” *J. Geophys. Res.*, vol. 89, no. B7, pp. 6329–6340, Jul. 1984.
- [2] E. Howell, E. Merényi, and L. Lebofsky, “Classification of asteroid spectra using a neural network,” *J. Geophys. Res.*, vol. 99, no. E5, pp. 10 847–10 865, 1994.
- [3] B. Bue, E. Merényi, and B. Csathó, “Automated labeling of materials in hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4059–4070, Nov. 2010.
- [4] B. D. Bue, E. Merényi, and B. Csathó, “Automated labeling of segmented hyperspectral imagery via spectral matching,” in *Proc. IEEE WHISPERS*, Aug. 2009, pp. 1–4.
- [5] R. Fisher, “The use of multiple measurements in taxonomic problems,” *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, Sep. 1936.
- [6] C. Rao, “The utilization of multiple measurements in problems of biological classification,” *J. R. Stat. Soc.*, vol. 10, no. 2, pp. 159–203, 1948.
- [7] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*, 2nd ed. New York: Springer-Verlag, Feb. 2011.
- [8] R. A. Horn and C. A. Johnson, *Matrix Analysis*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 1985.
- [9] E. Merényi, B. Csathó, and K. Tasdemir, “Knowledge discovery in urban environments from fused multi-dimensional imagery,” in *Proc. 4th IEEE GRSS/ISPRS Joint Workshop Remote Sens. Data Fusion over Urban Areas*, 2007, pp. 1–13.
- [10] A. Brown, “Spectral curve fitting for automatic hyperspectral data analysis,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1601–1608, Jun. 2006.
- [11] A. Woznica, A. Kalousis, and M. Hilario, “Learning to combine distances for complex representations,” in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 1031–1038.
- [12] B. Bue, D. Thompson, M. Gilmore, and R. Castaño, “Metric learning for hyperspectral image segmentation,” in *Proc. 3rd IEEE WHISPERS*, Jun. 2011, pp. 1–4.