



Cairo University  
**Egyptian Informatics Journal**

[www.elsevier.com/locate/eij](http://www.elsevier.com/locate/eij)  
[www.sciencedirect.com](http://www.sciencedirect.com)



ORIGINAL ARTICLE

# Learnable hyperspectral measures

A. Galal <sup>a,\*</sup>, H. Hasan <sup>b</sup>, I.F. Imam <sup>c</sup>

<sup>a</sup> *Science and Technology Group, Arab Administrative Development Organization, Cairo, Egypt*

<sup>b</sup> *Computer Science Department, Faculty of Computers and Information, Cairo University, Cairo, Egypt*

<sup>c</sup> *Computer Science Department, Virginia Tech., VA, USA*

Received 26 February 2012; revised 27 March 2012; accepted 29 April 2012

Available online 30 May 2012

## KEYWORDS

Hyperspectral measures;  
Support vector machine;  
Adaptive similarity  
threshold

**Abstract** Hyperspectral measures are used to capture the degree of similarity between two spectra. Spectral Angle Mapper (SAM) is an example of such measures. SAM similarity values range from 0 to 1. These values do not indicate whether the two spectra are similar or not. A static similarity threshold is imposed to recognize similar and dissimilar spectra. Adjusting such threshold is a troublesome process. To overcome this problem, the proposed approach aims to develop learnable hyperspectral measures. This is done through using hyperspectral measures values as similarity patterns and employing a classifier. The classifier acts as an adaptive similarity threshold. The derived similarity patterns are flexible as they are able to capture the specific notion of similarity that is appropriate for each spectral region. Two similarity patterns are proposed. The first pattern is the cosine similarity vector for the second spectral derivative pair. The second pattern is a composite vector of different similarity measures values. The proposed approach is applied on full hyperspectral space and subspaces. Experiments were conducted on a challenging benchmark dataset. Experimental results showed that, classifications based on second patterns were far better than first patterns. This is because first patterns were concerned only with the geometrical features of the spectral signatures, while second patterns combined various discriminatory features such as: orthogonal projections information, correlation coefficients, and probability distributions produced

\* Corresponding author.

E-mail address: [abdulrahman.galal@gmail.com](mailto:abdulrahman.galal@gmail.com) (A. Galal).



by the spectral signatures. The proposed approach results are statistically significant. This implies that using simple learnable measures overcomes complex and manually tuned techniques used in classification tasks.

© 2012 Faculty of Computers and Information, Cairo University.  
Production and hosting by Elsevier B.V. All rights reserved.

## 1. Introduction

Feature selection techniques seek finding the most informative hyperspectral bands to improve the classification task. Feature extraction techniques are superior to selection techniques as they handle mixed datasets as well. This is done through transforming spectral signatures into a new domain. The new domain features are arranged according to specific criterion. For example, Principal Component Analysis (PCA) transforms the data according to variance [1]. Minimum Noise Fraction (MNF) transforms the data according to Signal-to-Noise Ratio (SNR) [2]. Independent Component Analysis (ICA) transforms the data into maximally independent components [3]. However, feature selection and extraction techniques do not always succeed in deriving significant features. This is because the spectral signatures are always changing due to environmental factors. A promising trial was building a 3D model [4] for spectral signature changes across the seasons. Unfortunately, the 3D model is impractical, as it cannot detect every change happening to each material. Based on the previous facts, powerful classifiers based on feature selection and extraction techniques have variant performance on different datasets.

Hyperspectral measures are alternative approach to derive discriminatory information regarding spectral signatures. These measures are simple and computationally light. They are able to capture the degree of similarity between two spectrums. For example, in Spectral Angle Mapper (SAM), the similarity value ranges from 0 (highly similar) to 1 (highly dissimilar). The similarity value should, for instance, be less than 0.3 to recognize similar and dissimilar spectrums. Adjusting static thresholds requires extensive expert intervention. To relax such intervention, the proposed approach aims to develop learnable hyperspectral measures. This is done through using simple hyperspectral measures as similarity patterns and employing a classifier. This classifier acts as an adaptive similarity threshold. Two similarity patterns are proposed. The first pattern is the cosine similarity vector for the second spectral derivative pair. The second pattern is a composite vector of different similarity measures values. The idea was inspired from the highly successful researches [5,6] in measuring text documents similarity. A document may discuss an event with a certain vocabulary, while another document may discuss the same event with different vocabulary. Therefore, measuring similarity for text documents is much similar in complexity to measuring similarity of spectral signatures. Bilenko and Mooney in [5] developed a learnable text similarity measure. This measure is a cosine similarity vector based measure that employs SVM as an adaptive similarity threshold. The cosine similarity vector contains the term weights of the investigated document pair. Chen et al. in [6] developed a composite vector of different similarity measures values and employed a classifier as an adaptive similarity threshold.

The proposed approach aims to develop learnable hyperspectral measures as replacement for static threshold hyperspectral measures. This is done through using hyperspectral measures values as similarity patterns and employing a classifier. The classifier acts as an adaptive similarity threshold. The derived similarity patterns are flexible as they are able to capture the specific notion of similarity that is appropriate for each spectral region. Two similarity patterns are proposed. The first pattern is the cosine similarity vector for the second spectral derivative pair. The second pattern is a composite vector of different similarity measures values.

The proposed approach is applied on full hyperspectral space and subspaces. In full hyperspectral space, all spectral regions are treated as one domain. In hyperspectral subspaces, each spectral region is treated as a stand-alone domain. This process is called hyperspectral space decomposition. The decomposition is done for two reasons. The first reason is to maximize the information discrimination within each subspace. The second reason is to minimize the statistical dependence between subspaces. In doing so, potentially useful spectral signature information is not discarded. In addition, it overcomes the small-sample size problem, since the number of training signatures required per subspace is substantially low.

The paper is organized as follows: Section 2 describes the proposed approach; Section 3 presents the experimental evaluation of the proposed approach, and finally the conclusions.

## 2. Proposed approach

The proposed approach comes in two versions. Each version has been implemented twice using the full hyperspectral space and subspaces.

### 2.1. Version 1.1: cosine similarity vector applied on full hyperspectral space

Version 1.1 calculates the cosine similarity vectors for the second order derivatives of the spectral signature pairs. The vectors form similar and dissimilar patterns. The resulting patterns are classified by SVM that acts as an adaptive similarity threshold. Version 1.1 is applied on the full hyperspectral space of the spectral signature. In this section, we describe the steps of version 1.1.

*Step 1: Smoothing spectral signatures:* The mean filter has been used to minimize the random noise before analyzing spectral signatures vectors. The filter calculates the mean value of all points within a specified window as the new value of the midpoint of the window. The mean filter is defined as:

$$Es(\lambda_j) = \frac{\sum_{i=1}^n s(\lambda_i)}{n} \quad (1)$$

where  $s(\lambda_i)$  is the true signal of the spectrum,  $Es(\lambda_i)$  is the estimated noise-free spectrum,  $n$  (filter size) is determined by the half-bandwidth, and  $j$  is the index of the middle point of the filter.

**Step 2: Calculating spectral derivative features:** Once the spectral signatures are smoothed, the second spectral derivatives are calculated. The reasons for using second spectral derivatives are: (1) they are relatively insensitive to variations in illumination intensity caused by changes in sun angle, cloud cover, or topography [7]; (2) several interesting spectral features are apparent in the derivative spectra that were obscure in the original spectra. Second order derivatives swing with greater amplitude than the primary spectra. Consequently, derivative spectrums change from a positive slope to a negative slope at the peak of a narrow feature. These discriminatory derivatives are useful for separating out peaks of overlapping bands. The first spectral derivative is defined as:

$$x'_j = x_{j+1} - x_j, j = 1, 2, \dots, N - 1 \quad (2)$$

where  $x_j$  is the  $j$ th value of the raw spectral data,  $N$  is the total number of hyperspectral bands. The second spectral derivative is defined as:

$$x''_l = x'_{l+1} - x'_l, l = 1, 2, \dots, N - 2 \quad (3)$$

**Step 3: Calculating class mean vector:** Once the second spectral derivatives are calculated, the mean vector for each class is defined as:

$$\mu_c = \left[ \sum_{i=1}^y \frac{x''_{1i}}{y}, \sum_{i=1}^y \frac{x''_{2i}}{y}, \dots, \sum_{i=1}^y \frac{x''_{(N-2)i}}{y} \right] \quad (4)$$

where  $x''_{1i}$  is the first value of the second spectral derivative  $i$  in class  $c$ ,  $y$  is the number of samples in class  $c$  and  $N$  is the total number of hyperspectral bands.

**Step 4: Forming similarity patterns:** The cosine similarity vector is calculated for each second spectral derivative sample and class mean vectors as:

$$\cos(M, \mu_c) = \frac{M \cdot \mu_c}{\|M\| \|\mu_c\|} \quad (5)$$

where  $M$  is a second spectral derivative sample,  $\mu_c$  is a  $c$  class mean vector and  $\|\cdot\|$  is the L2 norm. Each similar or dissimilar pattern is a vector containing the cosine similarity values of size  $N - 2$ , where  $N$  is the number of hyperspectral bands. The resulting patterns are classified by SVM as shown in Fig. 1.

## 2.2. Version 1.2: cosine similarity vector applied on hyperspectral subspaces

Version 2.2 is the same as version 1.1 but it is applied on hyperspectral subspaces. Decomposing hyperspectral space is based on a knowledge derived from [8]. The following subspaces have been used: blue region (400–499 nm), green region (500–550 nm), red edge (650–750 nm), water absorption (900–1000 nm), and water content (1.35–2.4 nm). This version follows the same steps of version 1.1. The second spectral derivatives are calculated for the smoothed samples. The class subspace mean vectors are calculated. For each sample  $M$  in class  $C$ , the cosine similarity vector is calculated between the five subspaces of  $M$  and the corresponding five subspaces of

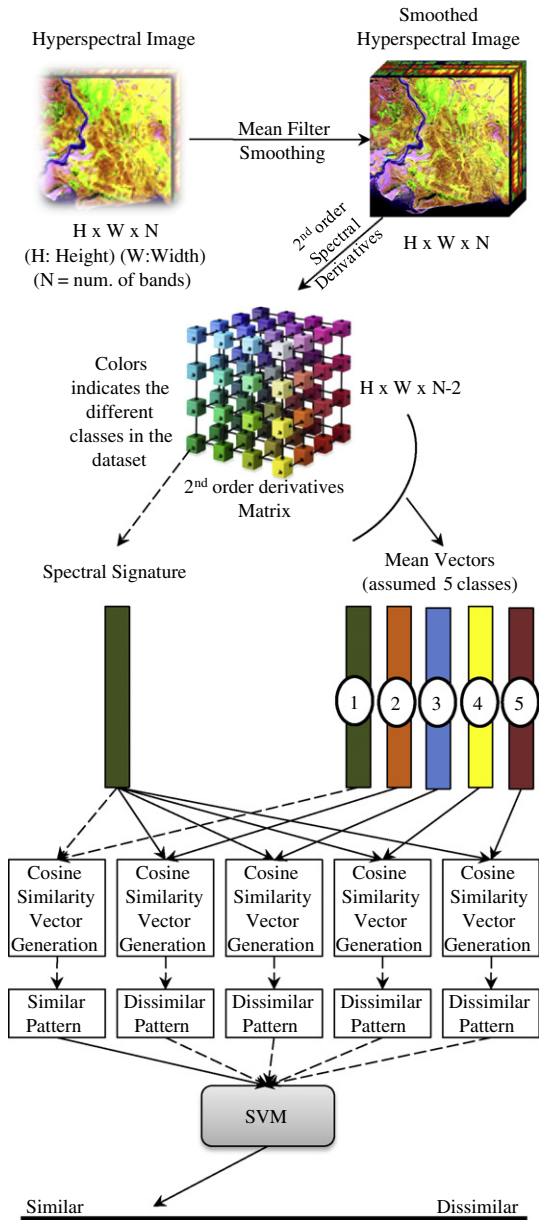


Figure 1 Proposed approach (Ver. 1.1).

$C$  class mean vector. The result is five cosine similarity subspace vectors forming one combined similar pattern. For each sample  $K$  not in class  $C$ , the cosine similarity is calculated between the five subspaces of  $K$  and the corresponding five subspaces in  $C$  class mean vector. The result is five cosine similarity subspace vectors, forming one combined dissimilar pattern. The resulting combined patterns of size  $N - 2$ , where  $N$  is the number of hyperspectral bands are classified by SVM as shown in Fig. 2.

## 2.3. Version 2.1: similarity measures values vector applied on full hyperspectral space

Version 2.1 calculates different similarity measures values for each spectral pair. This is done through using nine hyperspec-

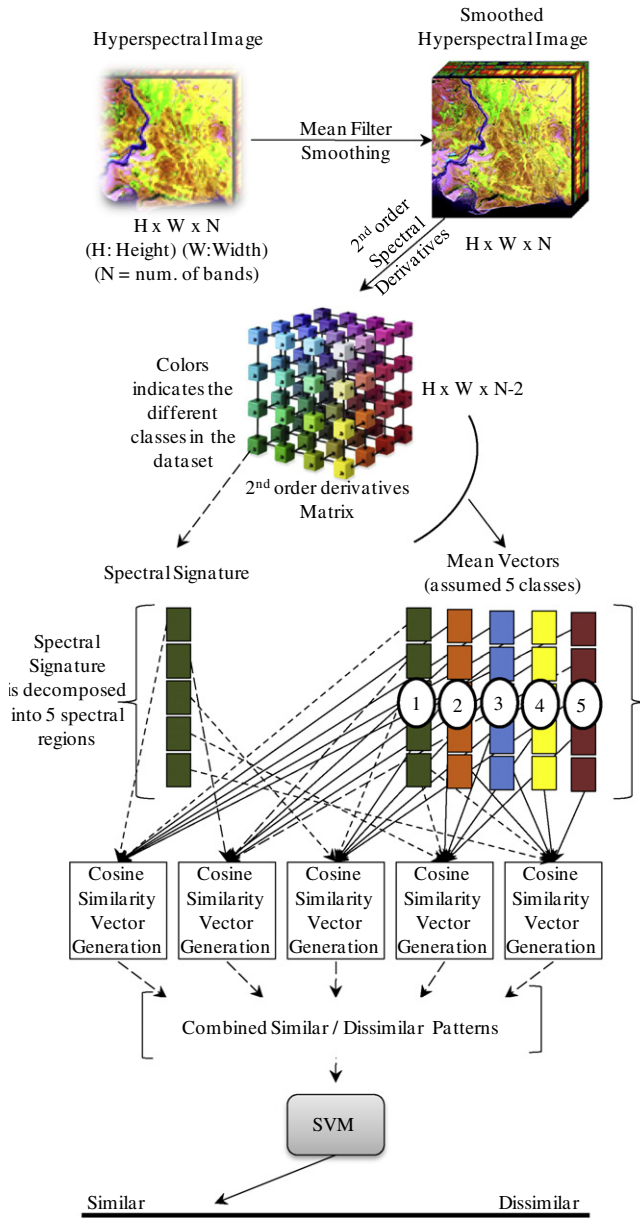


Figure 2 Proposed approach (Ver. 1.2).

tral measures to form similar and dissimilar patterns. The resulting patterns are classified by SVM that acts as an adaptive similarity threshold. Version 2.1 is applied on the full hyperspectral space of the spectral signature. Combining similarity values means consolidating the different statistics derived by similarity measures. The resulting composite vector of similarity values is used to discriminate each spectrum pair. In this section we describe the steps of version 2.1.

*Step 1: Smoothing spectral signatures:* The same as in step 1 of version 1.1.

*Step 2: hyperspectral measures:* The used measures for  $N$  (number of hyperspectral bands) spectral signatures are:

- (1) *Spectral Angle Mapper (SAM)* [9]: SAM measures the angle between two spectral signatures  $s_i$  and  $s_j$ . SAM is defined as:

$$\text{SAM}(s_i, s_j) = \cos^{-1} \left( \frac{\sum_{l=1}^N s_{il}s_{jl}}{[\sum_{l=1}^N s_{il}^2]^{1/2} [\sum_{l=1}^N s_{jl}^2]^{1/2}} \right) \quad (6)$$

- (2) *Orthogonal Projection Divergence (OPD)* [10]: OPD finds the residuals of orthogonal projections resulting from two spectral signatures  $s_i$  and  $s_j$ . OPD is defined as:

$$\text{OPD}(s_i, s_j) = (s_i^T P_{s_j}^\perp + s_j^T P_{s_i}^\perp) \quad (7)$$

where  $P_k^\perp = I - s_k(s_k^T s_k)^{-1} s_k^T$  for  $k = i, j$ ,  $s^T$  is the transpose of  $s$ , and  $I$  is the  $L \times L$  identity matrix.

- (3) *Spectral Correlation Mapper (SCM)* [9–11]: SCM partly takes into consideration brightness and shape differences between spectra. SCM is defined as:

$$\text{SCM}(s_i, s_j) = \frac{N \sum_1^N s_i s_j - \sum_1^N s_i \sum_1^N s_j}{\sqrt{[N \sum_1^N s_i^2 - (\sum_1^N s_i)^2] [N \sum_1^N s_j^2 - (\sum_1^N s_j)^2]}} \quad (8)$$

- (4) *Euclidean distance (ED)* [9]: ED takes into account the brightness difference between the two spectra, whereas SAM and SCM are invariant with brightness. ED is defined as:

$$\text{ED}(s_i, s_j) = 2\sqrt{1 - \cos(\text{SAM}(s_i, s_j))} \quad (9)$$

- (5) *Spectral Information Divergence (SID)* [9]: SID calculates the distance between the probability distributions produced by the spectral signatures  $s_i$  and  $s_j$ . SID is defined as:

$$\text{SID}(s_i, s_j) = D(s_i || s_j) + D(s_j || s_i) \quad (10)$$

$$D(s_j || s_i) = \sum_{l=1}^L q_l D_l(s_j || s_i) = \sum_{l=1}^L q_l (I_l(s_i) - I_l(s_j))$$

$$D(s_i || s_j) = \sum_{l=1}^L p_l D_l(s_i || s_j) = \sum_{l=1}^L p_l (I_l(s_j) - I_l(s_i))$$

$$p_k = \frac{s_{ik}}{\sum_{l=1}^L s_{il}}, q_k = \frac{s_{jk}}{\sum_{l=1}^L s_{jl}}, I_l(s_j) = -\log q_l, I_l(s_i) = -\log p_l$$

where  $p = (p_1, p_2, \dots, p_L)^T$  and  $q = (q_1, q_2, \dots, q_L)^T$  are the probabilities vectors for the spectral signatures of vectors,  $s_i$  and  $s_j$ .

- (6) *SAM-SID measure* [12]: SAM-SID is a combination of probability and geometry spaces. Such combination made two similar spectral signatures more similar, while two dissimilar spectral signatures more distinct. SAM-SID is defined as:

$$\text{SAMSID} = \text{SID} \times \tan(\text{SAM}) \quad (11)$$

- (7) *Pearson correlation coefficient (PCC)* [11]: This is standard version of PCC. It standardizes the data by centralizing itself in the mean of the spectral signatures  $s_i$  and  $s_j$ . PCC is defined as:

$$\text{PCC}(s_i, s_j) = \frac{\sum_{l=1}^N (s_{il} - \mu_i)(s_{jl} - \mu_j)}{[\sum_{l=1}^N (s_{il} - \mu_i)^2 \sum_{l=1}^N (s_{jl} - \mu_j)^2]^{1/2}} \quad (12)$$

where  $\mu$  is spectral signature mean value.



- (8) *Spectral Similarity Value (SSV)* [13]: SSV combines brightness and shape similarity. It is a combined measure of PC and ED measures. SSV is defined as:

$$SSV = \sqrt{ED^2 + (1 - PC)^2} \quad (13)$$

- (9) *Mahalanobis distance (MD)* [10,14]: MD takes the correlation between spectral signatures into account when computing statistical distances. The Mahalanobis distance has the following properties: (1) it accounts for the fact that the variances in each direction are different, (2) it accounts for the covariance between signatures, and (3) it reduces to the familiar Euclidean distance for uncorrelated variables with unit variance.

$$MD(s_i, s_j) = \sqrt{(s_i - s_j)^T Q^{-1} (s_i - s_j)} \quad (14)$$

$$Q = \frac{1}{n-1} \sum_{x=1}^n (s_x - \bar{s})(s_x - \bar{s})^T, \quad \bar{s} = \frac{1}{n} \sum_{x=1}^n s_x$$

where  $Q$  is the estimated covariance matrix computed with  $n$  training data samples.

*Step 3: Calculating class mean vector:* The same as step 3 of version 1.1 but it is applied on the smoothed data directly.

*Step 4: Forming similarity patterns:* Nine similarity measures values are calculated between spectral signatures and class mean vectors. The derived similarity values are combined in one vector, forming similar and dissimilar patterns. The resulting patterns of size 9 (number of similarity measures values) are classified by SVM as shown in Fig. 3. SVM acts as an adaptive similarity threshold.

#### 2.4. Version 2.2: similarity measures values vector applied on hyperspectral subspaces

Version 2.2 is similar to version 2.1 but it is applied on hyperspectral subspaces. The same subspaces used in version 1.2 are used in this version. This version follows the same steps of version 2.1. The subspace mean vectors are calculated once the data samples are smoothed. For each sample  $M$  in class  $C$ , 45 different similarity values (nine similarity measures values  $\times$  five subspaces) are calculated between the five subspaces of  $M$  and the corresponding five subspaces of  $C$  class mean vector. The result is 45 similarity measures values vector, forming one combined similarity pattern. For each sample  $K$  not in class  $C$ , 45 different similarity measures values are calculated between the five subspaces of  $M$  and the corresponding five subspaces of class  $C$  mean vector. The result is 45 similarity measures values vector, forming one combined dissimilar pattern. The resulting combined patterns of size 45 are classified by SVM as shown in Fig. 4. SVM acts as an adaptive similarity threshold.

### 3. Experimental evaluation

In this section, we present the used dataset, describe the experimental methodology and analyze the experimental results.

#### 3.1. Dataset

The dataset represents an Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) image. This image was taken from an

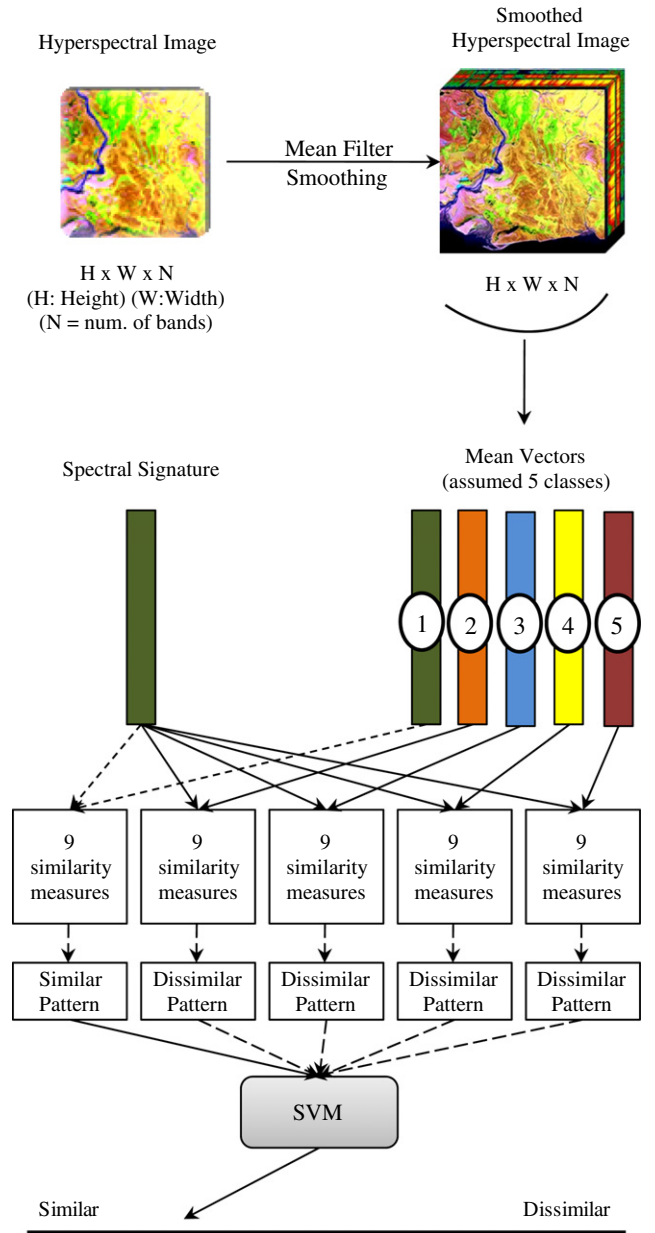
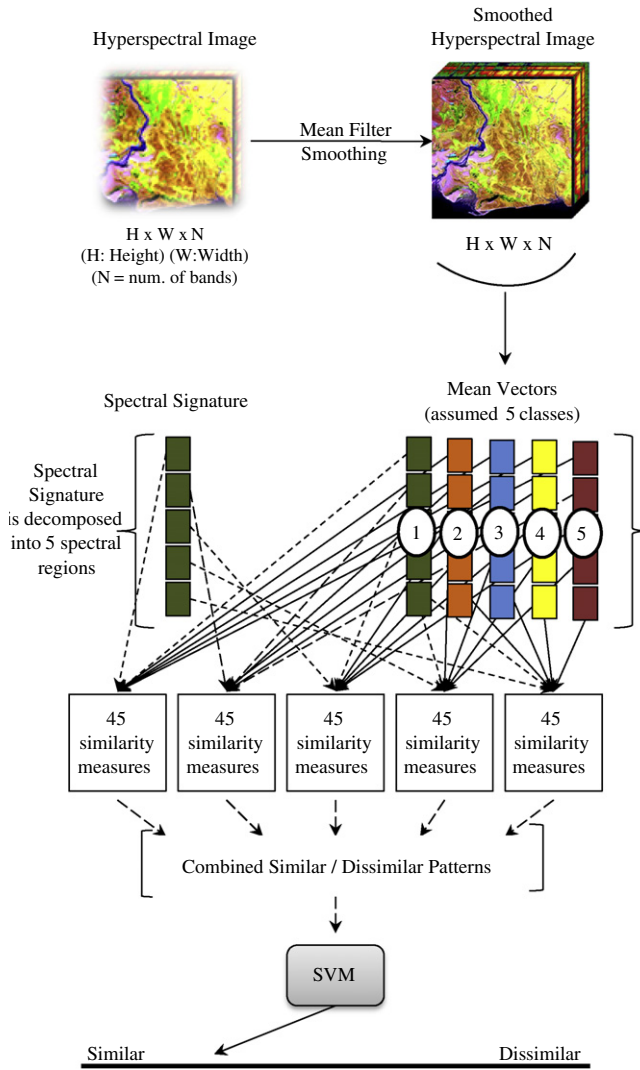


Figure 3 Proposed approach (Ver. 2.1).

area of mixed agriculture and forestry in Northwestern Indiana, USA. The data was recorded in June 1992 with 220 bands. Water absorption bands, bands 104–108 and 150–162 are removed leaving only 202 bands. The dataset was calibrated and hosted at: [https://gridsphere.rcac.purdue.edu:10443/irods-Portal/FileDownload?filename=av920612\\_NS\\_line.lan&home\\_dir=/rcacZone/home/lars/DVR\\_021/av920612\\_NS\\_line&username=biehl](https://gridsphere.rcac.purdue.edu:10443/irods-Portal/FileDownload?filename=av920612_NS_line.lan&home_dir=/rcacZone/home/lars/DVR_021/av920612_NS_line&username=biehl).

The test dataset is accompanied by a reference map, indicating partial ground truth, whereby pixels are labeled as belonging to one of 16 classes of vegetation or other land types. The ground truth data is found at [http://cobweb.ecn.purdue.edu/~biehl/av920612\\_NS\\_line\\_gr.zip](http://cobweb.ecn.purdue.edu/~biehl/av920612_NS_line_gr.zip). The scene is categorized into 17 classes as shown in Fig. 5. All competitive approaches used nine classes out of 17. The used classes were 2, 3, 5, 6, 8, 10, 11, 12, and 14. This dataset has been chosen



**Figure 4** Proposed approach (Ver. 2.2).

because it has been studied extensively in hyperspectral image classification field. Many classification methods consider it a big challenge as pixels are highly mixed [10]. Consequently, any spectral similarity measure may consider pixels in different classes belong to the same class. Fig. 6 shows a Google image for the test area.

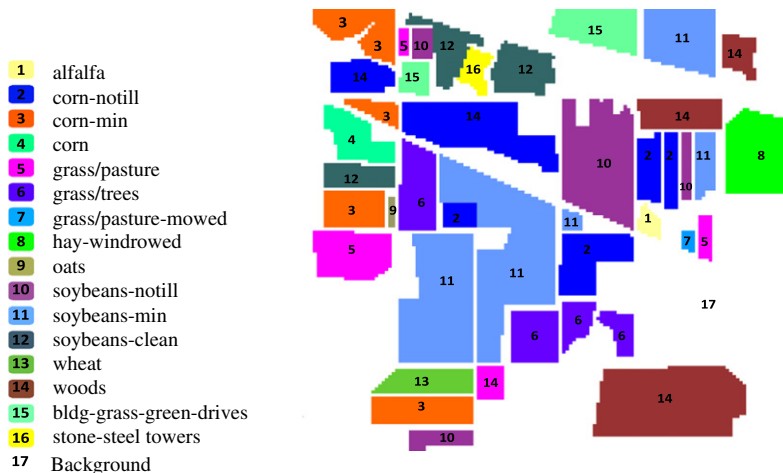
### 3.2. Experimental methodology

Experiments were conducted to (1) compare the performance of different multi-class SVM types utilized by the proposed approach; and (2) compare the performance of the proposed approach to [15,16] approaches. In this section, each competitive approach is briefly discussed.

*Approach 1:* Demir and Ertürk in [15] used SVM to classify the following: (1) magnitude features (raw spectral values); (2) a vector containing magnitude features and its first order derivatives; (3) PCA of a vector containing magnitude features; (4) PCA of vector containing magnitude features and its first order derivatives; and (5) PCA of a vector containing magnitude features and its first and second order derivatives.

*Approach 2:* Weizman and Goldberger in [16] used Neighborhood Component Analysis (NCA) to extract discriminatory features of the spectral signatures. *K*-nearest neighbor was used to classify the resulting features.

*The proposed approach:* it comes in two versions. The first version calculates the cosine similarity vector for the second spectral derivatives to form similar and dissimilar patterns. The second version calculates different similarity values using nine similarity measures to form similar and dissimilar patterns. SVM classifies the resulting patterns to act as adaptive similarity threshold. Each version has been implemented twice using full hyperspectral space and subspaces. Mathworks Matlab version R2009b has been used for implementing the hyperspectral measures. LIBSVM [17], a support vector machines tool, has been used to handle the multi-class SVM types. The used SVM parameters have been derived from a research



**Figure 5** Test area classes distribution.



**Figure 6** Google image for the test area.

conducted by Watanachaturaporn et al. in [18] on the same test dataset. These parameters are: (1) Kernel Function = Radial Basis Function; and (2) Penalty Value  $C = 1000$ . All approaches are trained using 4757 samples and tested using 4588 samples. Training and test samples were selected randomly from the previously mentioned nine classes. The distribution of classes is shown in Fig. 5.

### 3.3. Results

According to analysis conducted by Wu and Chang in [10] on the test dataset, the spectral signatures of classes (2, 3, 4, 7, 10 and 12) are so close to each other, and the same condition for classes (1, 8, and 11). For classes (5, 14, and 15), they have less similar signatures. For classes (6, 13 and 16), their signatures are dissimilar. Classes 5 and 11 are highly mixed. Signal-to-Noise Ratio (SNR) at the time of data acquisition was lower than current AVIRIS standards. This means the noise level is high.

#### 3.3.1. Multi-class SVM types comparison

Table 1 shows the performance of each multi-class SVM type. One-against-One (OvO) was the best while One-against-All (OvA) was the worst.

OvA separated each class from the rest classes, and developed a classification model. Such procedure was not appropriate for highly mixed classes. Many of the separated classes contained spectral signatures that were close to spectral signatures of other classes. Therefore, SVM failed to discriminate the similarity patterns efficiently. The training complexity was high as each OvA classifier was trained using all available samples. As a result, the performance of OvA was poor.

OvO was much better than OvA as each OvO classifier was trained using samples of two classes only. The low number of samples causes smaller nonlinearity, shorter training times and significant information discrimination. As a result, OvO achieved better results than OvA.

#### 3.3.2. Performance comparison

Table 3 shows the average classification accuracies achieved by all approaches.

In Approach 1, we considered the classification accuracy of magnitude features – (denoted by version A in Table 3), the baseline for all the upcoming comparisons. The classification accuracy of approach 1 version A was 92.56% with 200 features.

- For combining magnitude features with their 1st order derivatives (1OD) – (denoted by version B in Table 3), the classification accuracy increased by +1.29% with 399 features. The reason for such enhancement is using 1OD. First, and second derivatives (2OD) swing with greater amplitude than the primary spectra. Consequently, derivative spectrums change from a positive slope to a negative slope at the peak of a narrow feature. These discriminatory derivatives are useful for separating out peaks of overlapping bands. 1OD acted as metadata for each raw spectral signature enabling SVM to better classify the mixed signatures.
- For applying Principle Component Analysis (PCA) on magnitude features – (denoted by version C in Table 3), the classification accuracy decreased by –3.53% with 20 features. PCA arranged its derived features according to variance. This means the first PCA bands contained the largest percentage of data variance. PCA was not appropriate for this test dataset. This is because we seek finding the subtle changes that discriminate the spectral signatures rather than finding pixels with strong variance. The presence of high noise misled PCA calculations and that was transparent by comparing standard deviation and eigenvalues of both PCA and MNF bands in Table 2. Consequently, the discriminatory information with less variance was not in higher order PC components but in lower order components. Besides, there were some classes with small number of training samples. These classes were not captured by the second-order statistics-based PCA. As a result, PCA was not able to correctly preserve the information of interest.
- For applying the Principle Component Analysis (PCA) on (magnitude features and 1OD) and (magnitude features, 1OD and 2OD) – (denoted by versions D and E respectively in Table 3), the classification accuracy decreased by –3.2% and –2.9% with 20 and 25 features respectively. Attaching 1OD and 2OD to magnitude features acted as metadata for the raw spectral signatures and magnified the subtle differences of the narrow features. Applying PCA on such com-

**Table 1** SVM types classification accuracies.

SVM type	# Training samples	# Test samples	Proposed approach			
			Ver. 1.1	Ver. 1.2	Ver. 2.1	Ver. 2.2
OvA	4757	4588	80.01	87.02	90.20	91.60
OvO	4757	4588	82.01	89.81	94.19	96.25



**Table 2** PCA bands vs. MNF bands.

Bands	PCA		MNF	
	St. dev.	Eigenvalue	St. dev.	Eigenvalue
Band 1	5038.16	25383111.68	6.82	46.54
Band 2	2937.63	8629687.52	4.4	19.34
Band 3	778.74	606428.62	3.85	14.81
Band 4	358.52	128537.97	3.57	12.75
Band 5	263.2	69275.83	3.45	11.92
Band 6	230.81	53271.49	2.99	8.96
Band 7	161.2	25984.19	2.85	8.12
Band 8	116.7	13617.98	2.55	6.51
Band 9	112.57	12670.98	2.42	5.84
Band 10	91.3	8336.14	2.19	4.79

binned features boosted the performance of PCA. The first PCA bands combined high variance bands and bands containing relative weak signal samples.

In Approach 2, Neighborhood Component Analysis (NCA) outperformed approach 1. It increased the classification accuracy by +2.14% with 15 features. Unlike PCA which is not directly related with the final classification performance, NCA was designed to directly optimize the expected leave-one-out (LOO) classification error on the training data. NCA aims at learning a distance metric by finding a linear transformation of input data to enable  $K$ -nearest neighbor to perform well in this transformed space. Although NCA performance was good, it is computationally expensive. This implies that NCA only suitable for small-scale classification tasks.

In the proposed approach, two different similarity patterns have been proposed. The similarity patterns were derived from simple hyperspectral measures. The resulting patterns were classified by SVM.

- For version 1.1, the classification accuracy decreased by -10.55% with 200 features. Although version 1.1 was applied on 2OD, it failed to classify some of the highly mixed data samples. This is because the cosine weights of the spectral features were normalized across the full hyperspectral space. This means the cosine weight values indicating similar spectral regions tend to be low to approach dissimilar regions. As a result, version 1.1 performance was poor.
- For version 1.2, the classification accuracy decreased by -2.75% with 200 features. The reason for such enhancement was decomposing the hyperspectral space into subspaces. The calculated cosine weights for spectral features kept its power as they were normalized across small spectral regions. Both versions 1.1 and 1.2 were concerned only with the geometry of the spectral signatures. They did not capture any other discriminatory information such as: orthogonal projections information, correlation coefficients, and probability distributions produced by the spectral signatures. Versions 2.1 and 2.2 have combined all of these characteristics.
- For version 2.1, the classification accuracy increased by +1.63% with nine features. The reason for such enhancement was the consolidation of different discriminatory statistics powered by nine different similarity measures. The composite vector of the similarity values enables SVM to discriminate the mixed classes.

- For version 2.2, the classification accuracy increased by +3.69% with 50 features. The reasons for such enhancement were the consolidation of different similarity measures and the decomposition of hyperspectral space.

All hyperspectral subspace versions performed better than their counterparts applied on the full hyperspectral space. This is because decomposing the hyperspectral into subspaces maximized the information discrimination within each subspace, and minimized the statistical dependence between subspaces. In doing so, potentially useful spectral response information was not discarded. In addition, it overcame the small-sample size problem, since the number of training signatures required per subspace was substantially low. Fig. 7. depicts the performance of all approaches.

### 3.3.3. Training time comparison

Fig. 8 shows the training time comparison of all approaches exceeding the baseline classification accuracy 92.56% (approach 1 version A). Version 2.2 of the proposed approach was the lowest complexity and the highest accuracy with moderate number of features. Version 2.2 decreased the training time by 56.21% compared to NCA approach 2 – the best competitive approach.

### 3.3.4. Statistical significance

The previous experiments have been applied on nine classes out of 17. This is because PCA approach avoids classifying the remaining eight classes as the samples of these classes were relatively small. By applying NCA on the neglected classes (1, 4, 7, 9, 13, 15, 16 and 17) using DistLearnKit,<sup>1</sup> the classification accuracy for each neglected class was calculated. The following hypotheses have been set for right-tail Z-test:  $H_0$  ( $P_1 \leq P$ ) and  $H_1$  ( $P_1 > P$ ).  $P$  and  $P_1$  are the average classification accuracies for the 17 class samples achieved by NCA approach and version 2.2 of the proposed approach respectively.  $H_0$  is accepted when the calculated  $Z$  ( $Z_c$ )  $\leq$  the tabular  $Z$  ( $Z_T$ ).  $H_1$  is accepted when the calculated  $Z$  ( $Z_c$ )  $>$  the tabular  $Z$  ( $Z_T$ ). The calculated  $Z$  ( $Z_c$ ) is defined as:

$$Z_c = \frac{P_1 - P}{\sqrt{\frac{P(1-P)}{n}}} \quad (15)$$

where  $n$  is the number of samples. By conducting the right-tail Z-test with confidence level 99% assuming unequal variance, the  $Z_c$  value equals 2.36 and the  $Z_T$  value equals 2.32 for  $P_1 = 94.25\%$ ,  $P_2 = 93.69\%$  and  $n = 10,500$ . This means the two classification accuracies are significantly different. The reason for conducting right-tail test is that we know the direction of test as we compare the increase significance in classification accuracy.

## 4. Conclusion

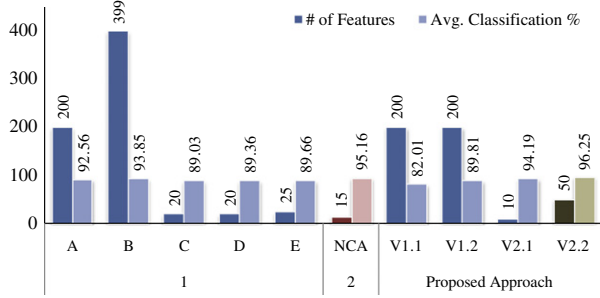
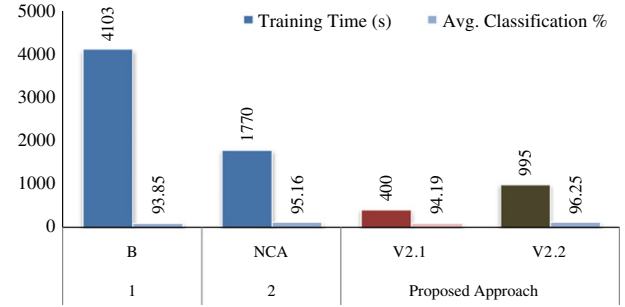
Hyperspectral similarity measures are static threshold based measures. Such measures require extensive expert intervention. The proposed approach developed learnable hyperspectral measures to relax expert engagement. This is done through

<sup>1</sup> [www.cs.cmu.edu/~liuy/distlearn.htm](http://www.cs.cmu.edu/~liuy/distlearn.htm) – School of Computer Science in Carnegie Mellon.



**Table 3** Average classification accuracy of the proposed approach and the competitive ones.

Approach	1	1	1	1	1	2	Proposed approach			
Versions	A	B	C	D	E	NCA	Ver.1.1	Ver.1.2	Ver.2.1	Ver.2.2
# Features	200	399	20	20	25	15	200	200	10	50
Avg. %	92.56	93.85	89.03	89.36	89.66	95.16	82.01	89.81	94.19	96.25

**Figure 7** Average classification accuracy of investigated approaches (the red bars present the best competitive approach – NCA) (the brown bars present the best proposed approach version – version 2.2).**Figure 8** Training time of approaches exceeding the baseline average classification accuracy 92.56% (the red bars present the best competitive approach – NCA) (the brown bars present the best proposed approach version – version 2.2).

using Hyperspectral measures values as similarity patterns and employing a classifier. The classifier acts as an adaptive similarity threshold. Two similarity patterns are proposed. The first pattern is the cosine similarity vector for the second spectral derivative pair. The second pattern is a composite vector of different similarity measures values. The resulting patterns are classified by SVM. The proposed approach is applied on full Hyperspectral space and sub-spaces.

The experiments have been applied on one of the most challenging Hyperspectral datasets. This is done to test the robustness of the proposed approach compared to the best competitive approaches applied on the same dataset. The experimental evaluation showed that the proposed approach outperformed PCA and NCA approaches. By conducting a right-tail Z-test to compare the significance of version 2.2 of the proposed approach to the best competitive approach (NCA approach), the calculated Z value was 1.7047 and the one-tailed  $p$ -value was 0.0441. This means the two classification accuracies were significantly different.

PCA performance was poor. This is because PCA kept high variance bands and ignored low order bands containing discriminatory information. In addition, PCA failed to classify small size classes of the test dataset. Unlike PCA which is not directly related with the final classification performance, NCA was designed to directly optimize the expected leave-one-out (LOO) classification error on the training data. Therefore, NCA performance was far better than PCA. NCA developed a learnable distance metric by finding a linear transformation of input data to enable KNN to perform well in this transformed space. Although NCA achieved good results, it is computationally expensive.

The proposed approach versions were capable of capturing the specific notion of similarity that is appropriate for each spectral region. In addition, they were computationally light. The different similarity measures values vector versions performed better than cosine similarity vector versions, as they

were able to combine different discriminatory characteristics powered by different similarity measures. The proposed versions applied on hyperspectral subspace performed better than their counterparts applied on the full Hyperspectral space. This is because decomposing the hyperspectral into subspaces maximized the information discrimination within each subspace, and minimized the statistical dependence between subspaces. In doing so, potentially useful spectral response information was not discarded. In addition, it overcame the small-sample size problem, since the number of training signatures required per subspace was substantially low. Utilizing One-against-One SVM and RBF Kernel boosted the classification accuracies of the proposed approach versions.

The training time of PCA and NCA was so high compared to the proposed approach versions. Therefore, the larger the number of training samples, the longer the time needed to build classification models for both PCA and NCA. The results imply that using simple learnable hyperspectral measures overcome complex or manually tuned techniques used in classification tasks.

## References

- [1] Li W, Prasad S, Fowler J, Bruce L. Class dependent compressive-projection principal component analysis for hyperspectral image reconstruction In: Proceeding of workshop on hyperspectral image and signal processing: evolution in remote sensing; 2011.
- [2] Denghui Z, Le Y. Support vector machine based classification for hyperspectral remote sensing images after minimum noise fraction rotation transformation. In: Proceeding of internet computing and information services conference, September; 2011.
- [3] Dalla Mura M, Villa A, Benediktsson JA, Chanussot J, Bruzzone L. Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis. IEEE Geosci Remote Sens Lett 2011;8(3):541–5.
- [4] Hemissi S, Ettabaa K, Farah I, Soulaïman B. Towards multi-temporal hyperspectral images classification based on 3D signa-

- ture model and matching. In: Proceeding of hyperspectral 2010 workshop; 2010.
- [5] Bilenko M, Mooney R. Adaptive duplicate detection using learnable string similarity measures. In: Proceeding of ninth ACM SIGKDD international conference on knowledge discovery and data mining; 2003.
- [6] Chen F, Farahat A, Brants T. Multiple similarity measures and source-pair information in story link detection. In: Proceeding of human language technology conference of the north american chapter of the association for computational linguistics, May; 2004.
- [7] Tsai F, Philpot W. Derivative analysis of hyperspectral data. *Remote Sens Environ* 1998;66(1):41–51.
- [8] Castro-Esau K, Sa'chez-Azofeifa A, Rivard B, Wright J, Quesada M. Variability in leaf optical properties of mesoamerican trees and the potential for species classification. *Am J Bot* 2006;4:517–30.
- [9] Van der Meer F. The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. *Int J Appl Earth Observ Geoinform* 2006;8(1):3–17.
- [10] Wu C, Chang C. Soft decision-made hyperspectral measures for target discrimination and classification. *Proc Soc Photo-Opt Instrum Eng* 2009;7457:74570S.
- [11] De Carvalho OA, Meneses PR. Spectral correlation mapper (SCM): an improvement on the Spectral Angle Mapper (SAM). In: Proceeding of ninth annual JPL airborne earth science, workshop; 2000.
- [12] Du Y, Chang C-I, Ren H, Chang C-C, Jensen JO, D'Amico FM. New hyperspectral discrimination measure for spectral characterization. *J Opt Eng* 2004;43(8):1777–86.
- [13] Farifteh J, Van der Meer F, Carranza EJM. Similarity measures for spectral discrimination of saline soils. *Int J Remote Sens* 2007;28(23):5273–93.
- [14] Lo E, Ingram J. Hyperspectral anomaly detection based on minimum generalized variance method. *Proc Soc Photo-Opt Instrum Eng* 2008;6966:696603.
- [15] Demir B, Ertürk S. Spectral magnitude and spectral derivative feature fusion for improved classification of hyperspectral Images. In: Proceeding of international conference on geoscience and remote sensing symposium, July; 2008.
- [16] Weizman L, Goldberger J. A classification based linear-projection of labeled hyperspectral data. In: Proceeding of IEEE geoscience and remote sensing symposium, July; 2007.
- [17] Chih-Chung C, Chih-Jen L. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2011;2(3): 1–27.
- [18] Watanachaturaporn P, Varshney P, Arora M. Evaluation of factors affecting support vector machines for hyperspectral classification. in: Proceedings of American society for photogrammetry and remote sensing conference, May; 2004.