

[Egyptian Informatics Journal (2012) 13, 85–94](http://dx.doi.org/10.1016/j.eij.2012.04.004)

Cairo University

Egyptian Informatics Journal

[www.elsevier.com/locate/eij](http://www.elsevier.com/locate/eij) [www.sciencedirect.com](http://www.sciencedirect.com/science/journal/11108665)

ORIGINAL ARTICLE

Learnable hyperspectral measures

A. Galal [a](#_bookmark0),[\*](#_bookmark3), H. Hasan [b](#_bookmark1), I.F. Imam [c](#_bookmark2)

a *Science and Technology Group, Arab Administrative Development Organization, Cairo, Egypt*

b *Computer Science Department, Faculty of Computers and Information, Cairo University, Cairo, Egypt*

c *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 trou- blesome process. To overcome this problem, the proposed approach aims to develop learnable hyperspectral measures. This is done through using hyperspectral measures values as similarity pat- terns 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 hyperspec- tral space and subspaces. Experiments were conducted on a challenging benchmark dataset. Exper- imental 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).

1110-8665 © 2012 Faculty of Computers and Information, Cairo University. Production and hosting by Elsevier B.V. All rights reserved.

Peer review under responsibility of Faculty of Computers and Information, Cairo University. <http://dx.doi.org/10.1016/j.eij.2012.04.004>

Production and hosting by Elsevier

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 trans- forming 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]](#_bookmark15). Minimum Noise Fraction (MNF) transforms the data according to Signal-to-Noise Ratio (SNR) [[2]](#_bookmark16). Independent Component Analysis (ICA) transforms the data into maximally independent components [[3]](#_bookmark17). However, feature selection and extraction techniques do not always suc- ceed 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]](#_bookmark18) for spectral signa- ture 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 per- formance 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 spec- trums. For example, in Spectral Angle Mapper (SAM), the similarity value ranges from 0 (highly similar) to 1 (highly dis- similar). 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 re- lax 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 simi- larity threshold. Two similarity patterns are proposed. The first pattern is the cosine similarity vector for the second spec- tral 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]](#_bookmark19) 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, measur- ing similarity for text documents is much similar in complexity to measuring similarity of spectral signatures. Bilenko and Mooney in [[5]](#_bookmark19) 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]](#_bookmark19) developed a composite vector

The proposed approach aims to develop learnable hyper- spectral measures as replacement for static threshold hyperspectral measures. This is done through using hyperspec- tral 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 vec- tor of different similarity measures values.

The proposed approach is applied on full hyperspectral space and subspaces. In full hyperspectral space, all spectral re- gions 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 sub- space. The second reason is to minimize the statistical depen- dence 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 evalu- ation of the proposed approach, and finally the conclusions.

1. Proposed approach

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

* 1. *Version 1.1: cosine similarity vector applied on full hyperspectral space*

Version 1.1 calculates the cosine similarity vectors for the sec- ond order derivatives of the spectral signature pairs. The vec- tors form similar and dissimilar patterns. The resulting patterns are classified by SVM that acts as an adaptive similar- ity 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 spec- tral 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:

P*n s*(k )

of different similarity measures values and employed a classi-

Es(k )= *i*=1 *i*

(1)

fier as an adaptive similarity threshold.

*j*

*n*

where *s(*k*i*) is the true signal of the spectrum, *Es(*k*i*) is the esti- mated 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 deriva- tives are calculated. The reasons for using second spectral derivatives are: (1) they are relatively insensitive to varia- tions in illumination intensity caused by changes in sun an- gle, cloud cover, or topography [[7]](#_bookmark19); (2) several interesting spectral features are apparent in the derivative spectra that were obscure in the original spectra. Second order deriva- tives 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*' = *xj*+1 — *xj*; *j* = 1; 2; ... ; *N* — 1 (2)

*j*

where *xj* 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:

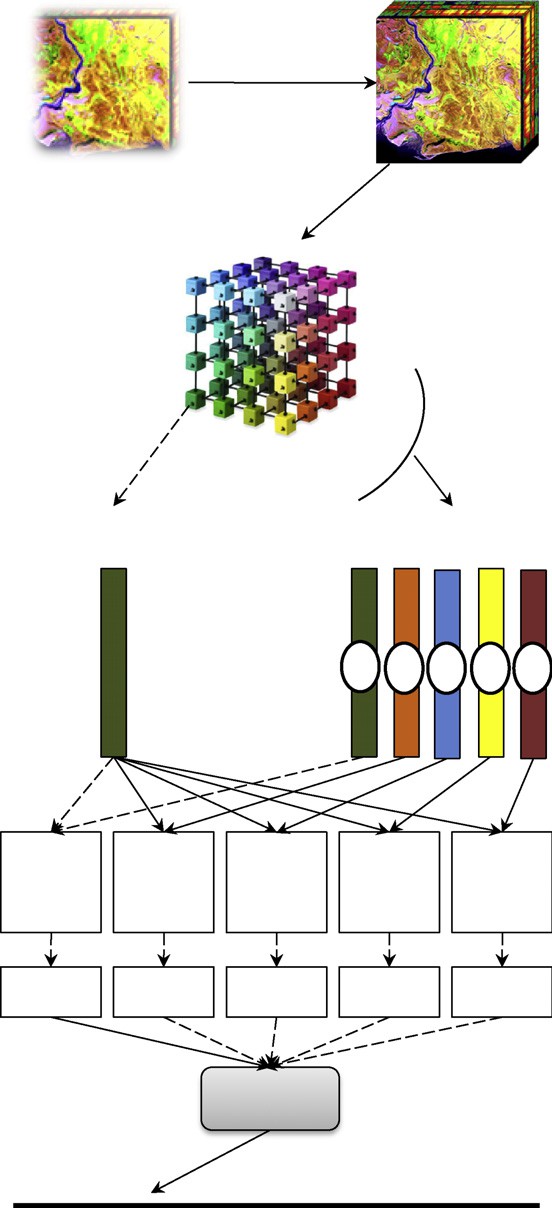
Hyperspectral Image

Smoothed Hyperspectral Image

*x*'' = *x*'

— *x*'; *l* = 1; 2; ... ; *N* — 2 (3)

*l l*+1 *l*



Mean Filter Smoothing

H x W x N

(H: Height) (W:Width) (N = num. of bands)

H x W x N

Colors indicates the different classes in the dataset

H x W x N-2

2nd order derivatives Matrix

Spectral Signature

Mean Vectors (assumed 5 classes)

1 2 3 4 5

Cosine Cosine Cosine Cosine Cosine Similarity Similarity Similarity Similarity Similarity Vector Vector Vector Vector Vector Generation Generation Generation Generation Generation

Similar Dissimilar Dissimilar Dissimilar Dissimilar Pattern Pattern Pattern Pattern Pattern

SVM

Similar

Dissimilar

*Step 3: Calculating class mean vector*: Once the second spectral

derivatives are calculated, the mean vector for each class is de- fined as:

*y*

2*i* ; ...

*y*

(*N*—2)*i*

*y*

(4)

"X*y*

l*c*

=

*x*'' X

*x*''

X *x*'' #

where *x*'' 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

*y*

1*i*

*i*=1

1*i* ;

*y*

*i*=1

*i*=1

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:

*M* · l

cos(*M*; l )= *c*

(5)

*c M* l 



*c*

where *M* is a second spectral derivative sample, l*c* is a *c* class mean vector and ||.|| is the L2 norm. Each similar or dissimilar

size *N* — 2, where *N* is the number of hyperspectral bands. pattern is a vector containing the cosine similarity values of The resulting patterns are classified by SVM as shown in

[Fig. 1](#_bookmark4).

* 1. *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 hyper- spectral subspaces. Decomposing hyperspectral space is based on a knowledge derived from [[8]](#_bookmark19). 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 sub- space vectors forming one combined similar pattern. For each sample K not in class C, the cosine similarity is calculated be- tween the five subspaces of K and the corresponding five sub- spaces in C class mean vector. The result is five cosine

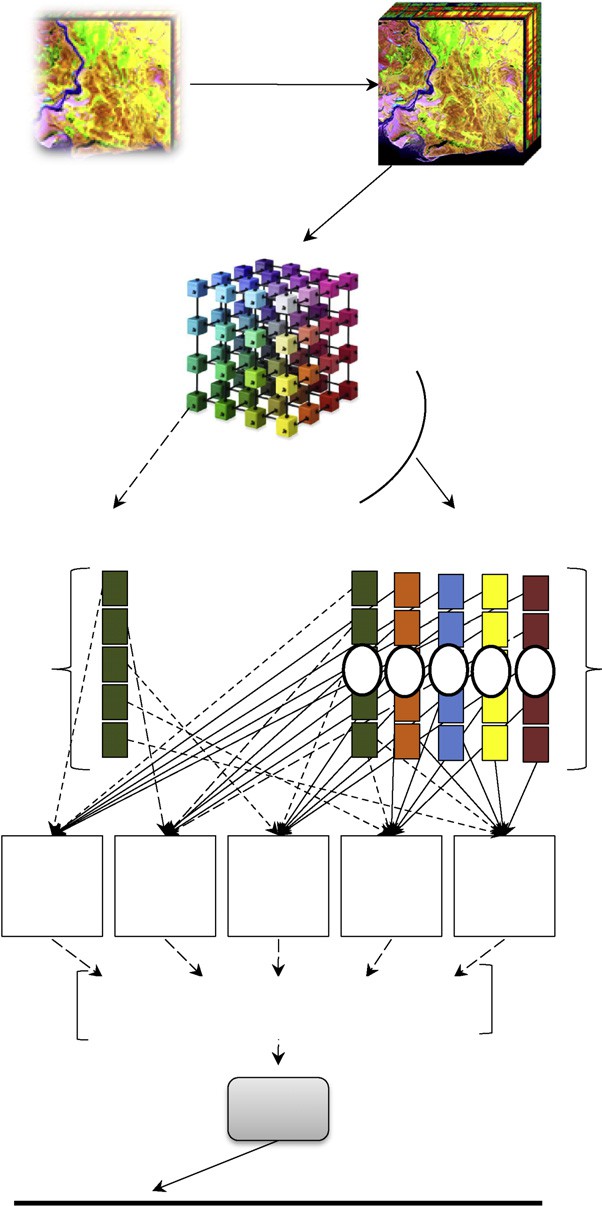
pattern. The resulting combined patterns of size *N* — 2, where similarity subspace vectors, forming one combined dissimilar *N* is the number of hyperspectral bands are classified by SVM

as shown in [Fig. 2](#_bookmark5).

* 1. *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-

Hyperspectra l Image



*s*2

Smoothed

Hyperspectral Image

SAM(*s* , *s* )= cos—1 B0

*N*

*l*=1 *il jl*

P *s s*

P*N*

*s*2

1/2hP*N*

*l*=1 *il*

*l*=1 *jl*

C1 (6)

H x W x N

(H: Height) (W:Width)

Mean Filter Smoothing

H x W x N

1. *Orthogonal Projection Divergence (OPD)* [[10]](#_bookmark19): OPD finds the residuals of orthogonal projections resulting from two spectral signatures *si* and *sj*. OPD is defined as:

OPD(*si*, *sj*)= (*sTP*⊥*si* + *sTP*⊥*sj*) (7)

@

A

*i*

*j*

i1/2

(N = num. of bands)

*i sj*

*j si*

where *P*⊥ = *I* — *s* (*sTs* )—1*sT* for *k* = *i*, *j*, *sT* is the transpose of *s*,

*sk k k k k*

Colors indicates the different

classes in the dataset

nd

H x W x N-2

and *I* is the *L* · *L* identity matrix.

1. *Spectral Correlation Mapper (SCM)* [[9–11]](#_bookmark19): SCM partly takes into consideration brightness and shape differ- ences between spectra. SCM is defined as:

*N N N*

*N*P *s s* — P *s* P *s*

SCM(*s* , *s* )= 1 *i j* 1 *i* 1 *j*

2 order deriva tives

Matrix

*N*

*Ns*2 —

*Ns*

*N*

*Ns*2 —

*Ns*

*i j* rhﬃﬃﬃﬃﬃﬃPﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃÿﬃﬃPﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃ ﬃﬃ2ﬃiﬃﬃﬃhﬃﬃﬃﬃﬃﬃPﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃÿﬃﬃPﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃ ﬃﬃ2ﬃﬃi**ﬃ**

Spectra l Signa ture

Mean Vectors

(a ssumed 5 cla sses)

1 *i* 1 *i*

1 *j* 1 *j*

(8)

Spectra l Signa ture

s decomposed

into 5 spectral

1 2 3 4 5

1. *Euclidean distance (ED)* [[9]](#_bookmark19): ED takes into account the

brightness difference between the two spectra, whereas SAM and SCM are invariant with brightness. ED is defined as:

regions

*i*

*j*

*i*

*j*

ED(*s* , *s* )= 2q1ﬃﬃﬃﬃ—ﬃﬃﬃﬃﬃcﬃﬃoﬃﬃﬃsﬃ(ﬃﬃSﬃﬃﬃAﬃﬃﬃﬃMﬃﬃﬃﬃ(ﬃﬃ*s*ﬃﬃﬃ,ﬃﬃ*s*ﬃﬃﬃ)ﬃﬃ)**ﬃ**

(9)

Cosine Simila rity Vector

Genera tion

Cosine Simila rity Vector

Genera tion

Cosine Simila rity Vector

Genera tion

Cosine Simila rity Vector

Generation

Cosine Simila rity Vector

Genera tion

1. *Spectral Information Divergence (SID)* [[9]](#_bookmark19): SID calcu-

lates the distance between the probability distributions produced by the spectral signatures *si* and *sj*. SID is defined as:

SID(*si*, *sj*)= *D*(*si*||*sj*)+ *D*(*sj*||*si*) (10)

X

X

Combined Simila r / Dissimila r Patterns

SVM

*D*(*sj*||*si*)=

*D*(*si*||*sj*)=

*L*

*l*=1 *L*

X

*l*=1

*ql Dl*(*sj*||*si*)=

*pl Dl*(*si*||*sj*)=

*L*

*l*=1 *L*

X

*l*=1

*ql*(*Il*(*si*)— *Il*(*sj*))

*pl*(*Il*(*sj*)— *Il*(*si*))

Simila r Dissimila r

*p* = *sik*

, *q* = *sjk* , *I* (*s* )= — log *q* , *I* (*s* )= — log *p*

*k* P*L s*

Figure 2 Proposed approach (Ver. 1.2).

*l*=1 *il*

*l*=1 *jl*

*k* P*L s l j*

*l l i l*

tral measures to form similar and dissimilar patterns. The resulting patterns are classified by SVM that acts as an adap- tive similarity threshold. Version 2.1 is applied on the full hyperspectral space of the spectral signature. Combining sim- ilarity values means consolidating the different statistics de- rived 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:

where *p* = (*p*1, *p*2,.. ., *pL*)*T* and *q* = (*q*, *q*2,.. ., *qL*)*T* are the prob- abilities vectors for the spectral signatures of vectors, *si* and *sj*.

1. *SAM-SID measure* [[12]](#_bookmark20): 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:

SAMSID = SIDx tan(SAM) (11)

1. *Pearson correlation coefficient (PCC)* [[11]](#_bookmark19): This is stan-

dard version of PCC. It standardizes the data by central- izing itself in the mean of the spectral signatures *si* and *sj*. PCC is defined as:

P*N* (*s* — l )(*s* — l )

PCC(*s* , *s* )= *l*=1 *il i jl j*

(12)

hP

(1) *Spectral Angle Mapper (SAM)* [[9]](#_bookmark19): SAM measures the angle between two spectral signatures *si* and *sj*. SAM is

*i j*

*N* (*sil* — l )

*l*=1 *i*

*l*=1 *j*

2P*N* (*sjl* — l )

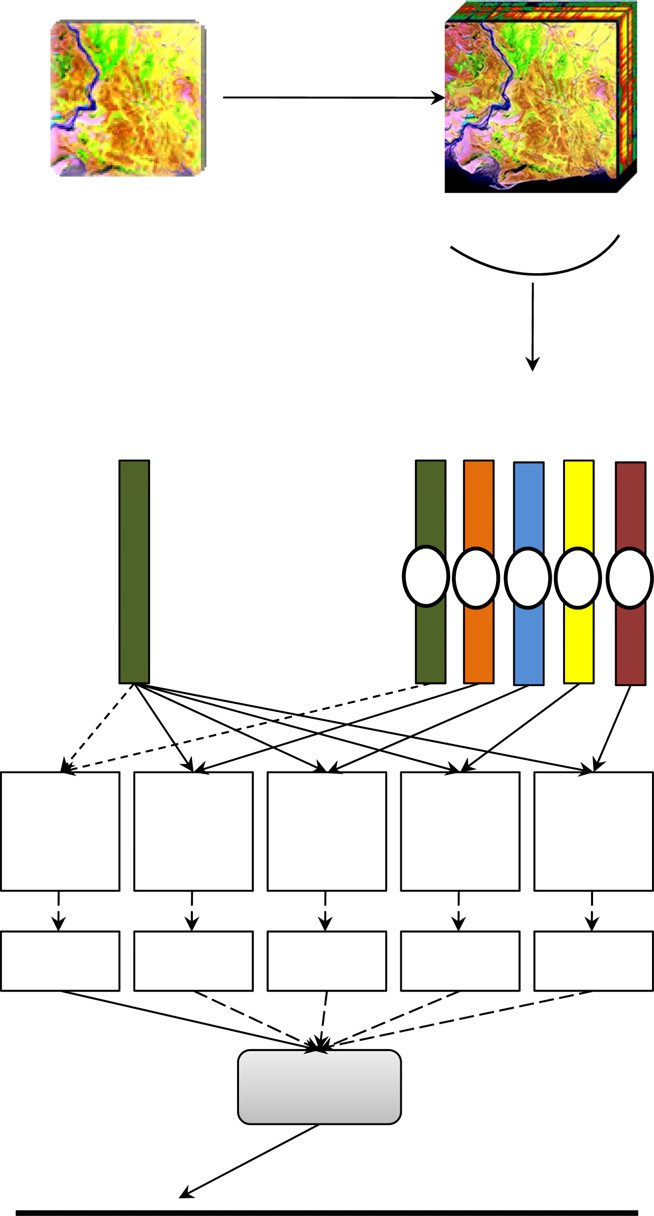
2i1/2

defined as:

where l is spectral signature mean value.

1. *Spectral Similarity Value (SSV)* [[13]](#_bookmark20): SSV combines brightness and shape similarity. It is a combined mea- sure of PC and ED measures. SSV is defined as:

Hyperspectral Image



Mean Filter Smoothing

H x W x N

(H: Height) (W:Width) (N = num. of bands)

H x W x N

Spectral Signature

Mean Vectors (assumed 5 classes)

1 2 3 4 5

SVM

Similar

Dissimilar

Smoothed Hyperspectral Image

SSV = qEﬃﬃﬃﬃDﬃﬃﬃﬃ2ﬃﬃﬃ+ﬃﬃﬃﬃ(ﬃﬃ1ﬃﬃﬃﬃ—ﬃﬃﬃﬃPﬃﬃﬃCﬃﬃﬃﬃ)ﬃﬃ2ﬃ

(13)

1. *Mahalanobis distance (MD)* [[10,14]](#_bookmark19): MD takes the cor-

relation between spectral signatures into account when computing statistical distances. The Mahalanobis dis- tance 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* , *s* )= q(ﬃﬃ*s*ﬃﬃﬃﬃ—ﬃﬃﬃﬃﬃ*s*ﬃﬃﬃ)ﬃﬃ*T*ﬃﬃ*Q*ﬃﬃﬃﬃ—ﬃﬃ1ﬃﬃ(ﬃﬃ*s*ﬃﬃﬃﬃ—ﬃﬃﬃﬃﬃ*s*ﬃﬃﬃ)**ﬃ**

*i*

*j*

*i*

*j*

*i*

*j*

(*sx* — *s*¯)(*sx* — *s*¯) , *s*¯ = *n*

*sx*

(14)

1 X*n*

*Q* = *n* — 1

*x*=1

*T* 1 X*n*

*x*=1

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 mea- sures 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](#_bookmark6). SVM acts as an adaptive similarity threshold.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 9 | 9 | 9 | 9 | 9 |
| similarity | similarity | similarity | similarity | similarity |
| measures | measures | measures | measures | measures |
| Similar | Dissimilar | Dissimilar | Dissimilar | Dissimilar |
| Pattern | Pattern | Pattern | Pattern | Pattern |

* 1. *Version 2.2: similarity measures values vector applied on hyperspectral subspaces*

Version 2.2 is similar to version 2.1 but it it is applied on hyper- spectral subspaces. The same subspaces used in version 1.2 are used in this version. This version follows the same steps of ver- sion 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 · five subspaces) are calculated between the five subspaces of M and the corresponding five subspaces of C class mean vector. The re- sult is 45 similarity measures values vector, forming one com- bined 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](#_bookmark7). SVM acts as an adaptive similarity threshold.

1. Experimental evaluation

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

* 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 Indi- ana, USA. The data was recorded in June 1992 with 220 bands. Water absorption bands, bands 104–108 and 150–162 are re- moved 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&user name=biehl.

The test dataset is accompanied by a reference map, indi- cating partial ground truth, whereby pixels are labeled as belonging to one of 16 classes of vegetation or other land types. The gound truth data is found at [http://cobweb.ecn.pur-](http://cobweb.ecn.purdue.edu/~biehl/av920612_NS_line_gr.zip) [due.edu/~biehl/av920612\_NS\_line\_gr.zip](http://cobweb.ecn.purdue.edu/~biehl/av920612_NS_line_gr.zip). The scene is catego- rized into 17 classes as shown in [Fig. 5](#_bookmark8). 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

Hyperspectral Image

H x W x N

(H: Height) (W:Width) (N = num. of bands)

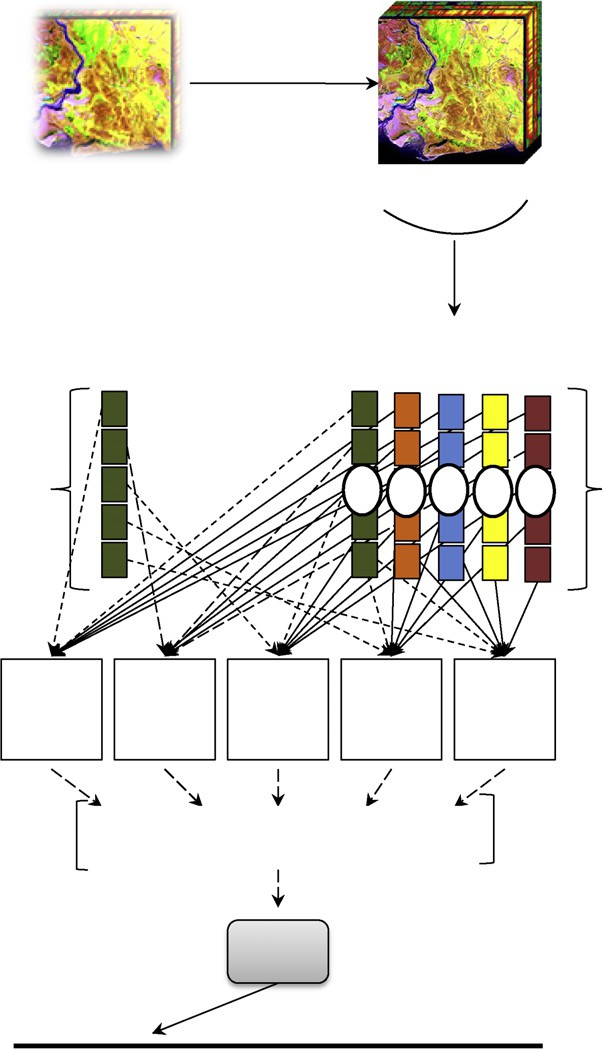
Spectral Signature

Spectral Signature

s decomposed into 5 spectral regions

Mean Filter Smoothing

Smoothed Hyperspectral Image

H x W x N

Mean Vectors (assumed 5 classes)

1 2 3 4 5

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]](#_bookmark19). Consequently, any spectral similarity measure may consider pixels in different classes belong to the same class. [Fig. 6](#_bookmark9) shows a Google image for the test area.

* 1. *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]](#_bookmark20) approaches. In this section, each competi- tive approach is briefly discussed.

*Approach 1*: Demir and Ertu¨rk in [[15]](#_bookmark20) used SVM to clas- sify the following: (1) magnitude features (raw spectral val- ues); (2) a vector containing magnitude features and its first order derivatives; (3) PCA of a vector containing mag- nitude features; (4) PCA of vector containing magnitude fea- tures 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]](#_bookmark20) used Neigh- borhood Component Analysis (NCA) to extract discrimina-

45

similarity measures

45

similarity measures

45

similarity measures

45

similarity measures

45

similarity measures

tory 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.

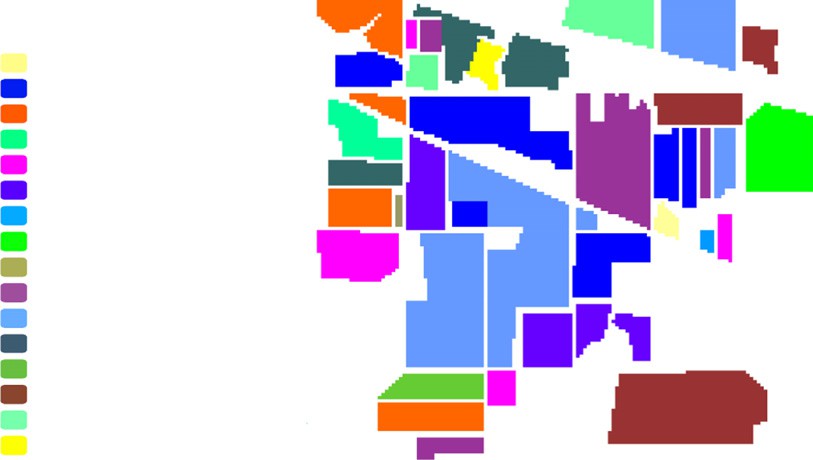
Combined Similar / Dissimilar Patterns

SVM

Similar Dissimilar

Figure 4 Proposed approach (Ver. 2.2).

The second version calculates different similarity values using nine similarity measures to form similar and dissimilar pat- terns. SVM classifies the resulting patterns to act as adaptive similarity threshold. Each version has been implemented twice using full hyperspectral space and subspaces. Mathworks Mat- lab version R2009b has been used for implementing the hyper- spectral measures. LIBSVM [[17]](#_bookmark20), 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



alfalfa corn-notill corn-min corn

grass/pasture grass/trees grass/pasture-mowed hay-windrowed

oats

soybeans-notill soybeans-min soybeans-clean wheat

woods

bldg-grass-green-drives stone-steel towers Background

Figure 5 Test area classes distribution.



Figure 6 Google image for the test area.

conducted by Watanachaturaporn et al. in [[18]](#_bookmark20) on the same test dataset. These parameters are: (1) Kernel Func- tion = 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](#_bookmark8).

* 1. *Results*

According to analysis conducted by Wu and Chang in [[10]](#_bookmark19) 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.

* + 1. *Multi-class SVM types comparison*

[Table 1](#_bookmark10) 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 devel- oped a classification model. Such procedure was not appropri- ate for highly mixed classes. Many of the separated classes contained spectral signatures that were close to spectral signa- tures 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.

* + 1. *Performance comparison*

[Table 3](#_bookmark12) 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](#_bookmark12)), 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 deriv- atives (1OD)* – (denoted by version B in [Table 3](#_bookmark12)), the clas-

sification 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*

classification accuracy decreased by —3.53% with 20 fea- *magnitude features* – (denoted by version C in [Table 3](#_bookmark12)), the tures. PCA arranged its derived features according to vari-

ance. This means the first PCA bands contained the largest percentage of data variance. PCA was not appropri- ate 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 trans- parent by comparing standard deviation and eigenvalues of both PCA and MNF bands in [Table 2](#_bookmark11). Consequently, the discriminatory information with less variance was not in higher order PC components but in lower order compo- nents. 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,*

in [Table 3](#_bookmark12)), the classification accuracy decreased by —3.2% and —2.9% with 20 and 25 features respectively. Attaching *1OD and 2OD)* – (denoted by versions D and E respectively

1OD and 2OD to magnitude features acted as metadata for

the raw spectral signatures and magnified the subtle differ- ences 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 |
|  |  |  |  |  |  |  |

* + - * + *For version 2.2*, the classification accuracy increased by

Table 2 PCA bands vs. MNF bands.

Bands PCA MNF

St. dev. Eigenvalue St. dev. Eigenvalue

+3.69% with 50 features. The reasons for such enhance-

ment were the consolidation of different similarity measures and the decomposition of hyperspectral space.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

All hyperspectral subspace versions performed better than their counterparts applied on the full hyperspectral space. This is because decomposing the hyperspectral into subspaces max- imized 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](#_bookmark13). depicts the perfor- mance of all approaches.

bined features boosted the performance of PCA. The first PCA bands combined high variance bands and bands con- taining relative weak signal samples.

*In Approach 2*, Neighborhood Component Analysis (NCA) outperformed approach 1. It increased the classification accu- racy 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 transfor- mation 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

* + 1. *Training time comparison*

[Fig. 8](#_bookmark14) shows the training time comparison of all approaches exceeding the baseline classification accuracy 92.56% (ap- proach 1 version A). Version 2.2 of the proposed approach was the lowest complexity and the highest accuracy with mod- erate number of features. Version 2.2 decreased the training time by 56.21% compared to NCA approach 2 – the best com- petitive approach.

* + 1. *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,1 the classifica- tion accuracy for each neglected class was calculated. The following hypotheses have been set for right-tail *Z*-test: H0 (*P*1 6 *P*) and H1 (*P*1 > *P*). *P* and *P*1 are the average classifica- tion accuracies for the 17 class samples achieved by NCA ap- proach and version 2.2 of the proposed approach respectively. H0 is accepted when the calculated *Z* (*Zc*) 6 the tabular *Z* (Z*T*). H1 is accepted when the calculated *Z* (*Zc*) > the tabular *Z* (Z*T*). The calculated *Z* (Z*c*) is defined as:

features were normalized across the full hyperspectral space. This means the cosine weight values indicating simi- lar spectral regions tend to be low to approach dissimilar

*P* — *P*

*c* = qﬃ*P*ﬃﬃ(ﬃ1ﬃ—ﬃﬃﬃ*P*ﬃﬃ)ﬃ

*n*

*Z* 1

(15)

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 enhance- ment was decomposing the hyperspectral space into

subspaces. The calculated cosine weights for spectral fea- tures 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 coeffi- cients, and probability distributions produced by the spec- tral 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 enhance-

ment was the consolidation of different discriminatory sta-

where *n* is the number of samples. By conducting the right- tail *Z*-test with confidence level 99% assuming unequal vari- ance, the Z*c* value equals 2.36 and the *ZT* 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 clas- sification accuracy.

1. 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

tistics powered by nine different similarity measures. The

composite vector of the similarity values enables SVM to discriminate the mixed classes.

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

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 | |
|  |  |  |  |  |  |  |  |  |  |  | |

400

399

300

200

200

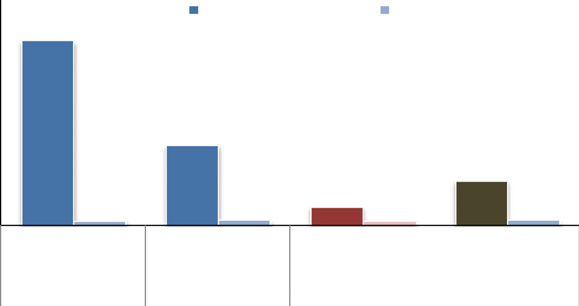
92.56

93.85

82.01

5000

4000



Training Time (s) Avg. Classification %

B

NCA

V2.1

V2.2

1 2 Proposed Approach

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

4103

3000

1770

2000

20

20

25

15

10

50

96.25

995

100

0

1000

200

200

0

93.85

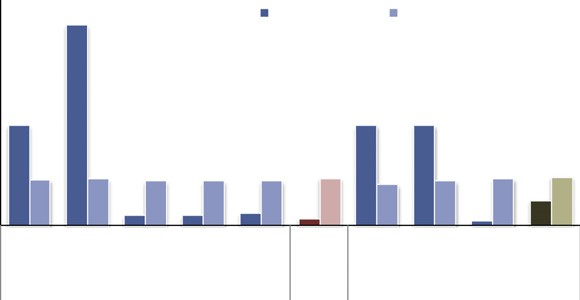
95.16

400

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).



# of Features Avg. Classification %

A B C D E NCA V1.1 V1.2 V2.1 V2.2

1 2 Proposed Approach

89.03

89.36

89.66

95.16

89.81

94.19

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 simi- larity threshold. Two similarity patterns are proposed. The first pattern is the cosine similarity vector for the second spec- tral 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 chal- lenging Hyperspectral datasets. This is done to test the robust- ness 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 classifica- tion 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 trans- formed space. Although NCA achieved good results, it is com- putationally 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 per- formed better than cosine similarity vector versions, as they

were able to combine different discriminatory characteristics powered by different similarity measures. The proposed ver- sions 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 signa- tures required per subspace was substantially low. Utilizing One-against-One SVM and RBF Kernel boosted the classifica- tion 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 re- sults imply that using simple learnable hyperspectral measures overcome complex or manually tuned techniques used in clas- sification 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.

1. Hemissi S, Ettabaa K, Farah I, Soulaiman B. Towards multi- temporal hyperspectral images classification based on 3D signa-

ture model and matching. In: Proceeding of hyperspectral 2010 workshop; 2010.

1. 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.
2. 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.
3. Tsai F, Philpot W. Derivative analysis of hyperspectral data. Remote Sens Environ 1998;66(1):41–51.
4. 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.

1. 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.
2. Wu C, Chang C. Soft decision-made hyperspectral measures for target discrimination and classification. Proc Soc Photo-Opt Instrum Eng 2009;7457:74570S.
3. 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.
4. Du Y, Chang C-I, Ren H, Chang C-C, Jensen JO, D’Amico FM. New hyperspectral discrimination measure for spectral character- ization. J Opt Eng 2004;43(8):1777–86.
5. 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.
6. Lo E, Ingram J. Hyperspectral anomaly detection based on minimum generalized variance method. Proc Soc Photo-Opt Instrum Eng 2008;6966:696603.
7. Demir B, Ertu¨rk 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.
8. Weizman L, Goldberger J. A classification based linear-projection of labeled hyperspectral data. In: Proceeding of IEEE geoscience and remote sensing symposium, July; 2007.
9. Chih-Chung C, Chih-Jen L. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2011;2(3): 1–27.
10. Watanachaturaporn P, Varshney P, Arora M. Evaluation of factors affecting support vector machines for hyperspectral classification. in: Proceedings of American society for photogram- metry and remote sensing conference, May; 2004.