Weighted Chebyshev Distance Algorithms for Hyperspectral Target Detection and Classification Applications

S.Demirci[1], I.Erer[2], O.Ersoy[3]
[1] Turkish Air Force Academy, Yesilyurt, Istanbul, Turkey
[2] Istanbul Technical University, Faculty of Electrical and Electronics Engineering, Maslak, Istanbul, Turkey
[3] Purdue University, School of Electrical and Computer Engineering, W. Lafayette, Indiana, USA

Abstract. In this study, an efficient spectral similarity method referred to as Weighted Chebyshev Distance (WCD) method is introduced for supervised classification of hyperspectral imagery (HSI) and target detection applications. The WCD is based on a simple spectral similarity based decision rule using limited amount of reference data. The estimation of upper and lower spectral boundaries of spectral signatures for all classes across spectral bands is referred to as a vector tunnel (VT). To obtain the reference information, the training signatures are provided randomly from existing data for a known class. After determination of the parameters of the WCD algorithm with the training set, classification or detection procedures are accomplished at each pixel. The comparative performances of the algorithms are tested under various cases.

The decision criterion for classification of an input vector is based on choosing its class corresponding to the narrowest VT that the input vector fits in to. This is also shown to be approximated by the WCD in which the weights are chosen as an inverse power of the generalized standard deviation per spectral band. In computer experiments, the WCD classifier is compared with the Euclidian Distance (ED) classifier and the Spectral Angle Map (SAM) classifier.

The WCD algorithm is also used for HSI target detection purpose. Target detection problem is considered as a two-class classification problem. The WCD is characterized only by the target class spectral information. Then, this method is compared with ED, SAM, Spectral Matched Filter (SMF), Adaptive Cosine Estimator (ACE) and Support Vector Machine (SVM) algorithms. During these studies, threshold levels are evaluated based on the Receiver Operating Characteristic Curves (ROC).

Keywords: Hyperspectral Images, Supervised Classification, Target Detection, Weighted Chebyshev Distance Algorithm

1. Introduction

Data received from satellite or airborne sensors may be analyzed to extract spatial information of surface materials. One of the major problems in remote sensing is the lack of sufficient spatial information corresponding to the investigated material. To overcome this problem, hyperspectral remote sensors have been developed to record the image at a large number of wavelengths. The reflectance spectra obtained in this way are called spectral signature or spectra, and uniquely characterize materials [1-2].

There are a big variety of potential civilian and military applications for HSI remote sensing. Among these applications, the main objective of target detection is to search the pixels in a HSI for the existence of a specific material. For this purpose, the measured hyperspectral data is compared with the reflectance spectra of the material derived from field work or laboratory study to determine whether the given input scene contains a target [3-4]. If we have no a priori information about targets, the detection approach can be based on searching for pixels whose spectral content is significantly different from the spectral content of the background [5].

The other major topic involving pixel labeling in a hyperspectral data cube is classification. In supervised classification, measured hyperspectral testing data is compared with the reference reflectance spectra (spectral signature) of each class to determine whether the input scene contains similar spectral characteristics of specified classes. If a priori information about classes

does not exist, unsupervised classification usually in the form of a clustering method is utilized to determine regions which have similar spectral characteristics [6-7].

Among a number of distance measures [8-10], there are two widely used computationally simple deterministic metrics for identification of spectral similarity between unknown spectra and desired spectra. They are ED measure and SAM measure.

Due to inherent variations present in a class as well as presence of external disturbance factors in the recorded image, spectral signatures measured from samples of the same class have associated random variations. This leads to detection or classification algorithms based on statistical analysis. In statistical decision procedures, the normally distributed probabilistic model is widely used. Here the spectral signatures of all classes are modeled with multivariate normal vectors. The maximum likelihood (ML) classification algorithm is often used to assign an unknown pixel vector into a class. This decision is made by choosing the class that has the maximum probability. In the Gaussian maximum likelihood (GML) classifier, the natural logarithm of the probability density function leads to the statistical distance measure known as the Mahalanobis Distance [10]. It plays a major role in target detection and classification algorithms.

For target detection applications, background classes consist of different types of materials. So, target detection algorithms are different from classification algorithms by way of modeling background representations. If the target and background classes have the same covariance matrix, the quadratic detector becomes Fisher's Linear Discriminant Function, or the SMF [5]. After performing minimum total energy based optimization, normalized MF becomes Constrained Energy Minimization (CEM) filter [10]. Other detection algorithms have been proposed based on statistical analysis in [11-14]. Most of these detectors are optimal when the target and background classes follow multivariate normal distributions with the same covariance matrix. In practical applications, these quantities are unavailable, and have to be estimated from existing data. Unfortunately, there is usually not sufficient data to determine the covariance matrix of the target, especially with hyperspectral data. Hence, data cube covariance matrix and a target spectral signature from a library or the mean of a small number of known target pixels is used. The SMF is the most commonly used method where the data cube mean is usually removed from the target and test pixel spectra.

In Adaptive SMF (ASMF), the background clutter statistics is adapted locally by using a dual window centered at each pixel. The inner window is approximately the same size as the target. The outer window is larger and represents the local background clutter. The background covariance matrix is obtained from the pixels within the outer region. In this way, the spectral matched filter is adapted for every pixel.

The performance of SMF that uses global background clutter statistics has been found to be worse than ASMF that uses the local clutter statistics around the test pixels [15-16]. On the other hand, the ASMF method is very slow compared to the SMF method because of the locally calculated background covariance matrix. In addition, since a relatively small number of data points representing background information around test pixels exist, a regularization process is needed to make the background covariance matrix more stable. Moreover, due to the inner window, it is not suitable for targets which have unpredictable size and shape like forest, mine, crop, and in applications such as disturbed soil, liquid leakage and gas exposure detection, etc.

For sub-pixel targets, Kelly introduced a Generalized Likelihood Ratio (GLR) based detector by using unstructured background model [17]. ACE, ASMF detectors and their other versions,

Matched Subspace Detectors (MSDs) and Adaptive Subspace Detectors (ASDs) are also introduced [18-20]. By using structured background model, the orthogonal subspace projector (OSP) methods are widely employed [12, 21-22].

A new approach called kernel method has recently been used in classification, target detection, anomaly detection, and change detection with hyperspectral images [23-26]. The kernel approach is based on mapping the data using a non-linear transformation into a higher dimensional space, and then searching for a linear separation surface between the two classes. In this way, non-linear border surfaces in the original data space can be transformed to a hyperplane in a higher dimensional space. Replacing inner products by Mercer Kernel function inner product which can be computed in advance in terms of the input data is a major idea behind what is called the 'kernel trick' [25, 27].

The best known kernel method is SVM, especially due to its satisfactory results in high dimensional feature spaces [26, 28]. The SVM approach consists of finding the optimal hyperplane that maximizes the distance between the closest training samples and the separating hyperplane. In a binary classification problem, the aim is to place all the data belonging to class +1 lie on one side of the separating hyperplane and all the data belonging to class -1 lie on the other side with an optimal distance called the margin. Type of kernel function affects the discriminant function. Some popular kernel functions are linear, polynomial, Gaussian radial basis function (GRBF) and sigmoid kernels. The GRBF kernel is often the most popular one used in the literature.

In this study, an efficient spectral similarity method called WCD method for the supervised classification and detection of hyperspectral imagery is introduced. The estimation of upper and lower spectral boundaries of spectral signatures for reference classes across spectral bands is referred to as a VT. Detection or classification procedures can be carried out with VT parameters obtained from specifications of class references. In the case of learning with training data having labeled training vectors, this is also shown to be equivalent to the use of the WCD in which the weights are chosen as a power of the standard deviation per spectral band.

2 WCD ALGORITHMS FOR TARGET DETECTION AND CLASSIFICATION

Although various statistical detection and classification algorithms have been developed, insufficient training data and high dimensionality of spectra reduce the performance and effectiveness of these algorithms [29]. High-dimensional vector spaces have some unusual characteristics such as the Hughes effect [30]. Furthermore, high dimensionality may cause increasing the operation cost and reducing the separability of classes [31]. Some methods are developed to avoid increasing computation time and poor separation resulting from high dimensionality of hyperspectral data. One simple and effective way is to reduce the number of dimensions without sacrificing valuable information. This process is called feature extraction [32-34]. If not done properly, a reduction in the number of features may reduce discrimination power and lower the accuracy of the resulting recognition system.

Although SVM may perform well in high dimensional spaces, a number of issues such as training time, proper choice of parameters and kernel function are still challenging [35]. Furthermore, in real applications, the stated assumptions during optimization are not optimal.

It is clear from above evaluation that the amount of training time required for nonparametric methods rises with increase of dimensionality. In the proposed WCD method, the spectral

signatures of all known pixels are considered in the vector space, and the boundaries of the known spectral signatures are considered. This bounded region is to be called vector tunnel due to its tunnellike shape. The prediction of upper and lower spectral boundaries of all known classes spectral signatures constitutes the basic consideration leading to the proposed WCD target detection and classification algorithms.

In spectral similarity-based methods, input test pixels can be modeled as $\mathbf{x} = a\mathbf{s} + \mathbf{n}$ where $\mathbf{x} = (x_1, x_2, \cdots, x_L)^T$ is the spectral sample of the L-band test pixel, $\mathbf{s} = (s_1, s_2, \cdots, s_L)^T$ is the reference spectral signature, ***a*** is a scaling constant, and ***n*** is an L-dimensional additive parameter vector depending on a number of factors such as the illumination level. In this study, **x** is normalized by the inverse of the maximum value among the elements of the training vectors when ground truth is available.

When the length of vectors **x** and **s** are linearly scaled by the constant ***a***, the ED is also scaled by ***a*** while the SAM remains the same. Consequently, there is a high correlation between SAM and the reflectance spectra whereas ED is highly sensitive to illumination level. As a result, the variation of ***a*** affects ED while variation of ***n*** affects SAM substantially [13].

The distance measure used in the proposed WCD algorithm can trace both the total changes of illumination level and shape of spectra. The WCD algorithm is based on the estimation of upper and lower spectral boundaries of all class spectral signatures across whole spectral bands. Entire pixel vectors from a known class are used to estimate the natural boundaries containing all possible pixels from the target class. We name this bounded region as VT due to tunnel like shape of the plots. This leads to the use of a special WCD between an input vector to be classified and the mean vector of a target class. The minimum such distance decides the proper class of the input vector.

2.1 WCD Method for Classification Application

The L-dimensional mean vector $\boldsymbol{\mu}_i = (\mu_{i,1}, \mu_{i,2}, \cdots, \mu_{i,L})^T$ and the standard deviation vector $\boldsymbol{\sigma}_i = (\sigma_{i,1}, \sigma_{i,2}, \cdots, \sigma_{i,L})^T$ of the i^{th} class are estimated from the GT data [36-41]. Using these statistical parameters and the VT scaling parameter tr_i to be discussed, the initial decision rule is based on whether the test pixel vector $\mathbf{x} = (x_1, x_2, \cdots, x_L)^T$ is in the VT region given as

$$\boldsymbol{\mu}_i - tr_i \cdot \mathbf{std}_i \le \mathbf{x} \le \boldsymbol{\mu}_i + tr_i \cdot \mathbf{std}_i \qquad i = 1, 2, \cdots, c \tag{1}$$

where subscript *i* refers to class number. If $\mathbf{x}$ is in the i^{th} class, then each pixel component must satisfy (1).

In experimental work, we discovered that higher classification accuracy is achievable if the standard deviation vector is further generalized. Hence, the L-dimensional i^{th} class generalized standard deviation vector is defined by

$$\mathbf{std}_i = \boldsymbol{\sigma}_{i,j}^{p} = (\sigma_{i,1}^{p}, \ \sigma_{i,2}^{p}, \ \cdots, \ \sigma_{i,L}^{p})^T \qquad i = 1, 2, \cdots, c \tag{2}$$

where $\boldsymbol{\sigma}_{i,j}^{p}$ denotes the vector obtained by generating the *p*'th power of each component of the standard deviation vector. The value of p affects the performance results substantially, and the procedure for the estimation of its best value is discussed in Section 3.

Writing the upper and lower decision boundaries in (1) as $\mathbf{HB}_i = \boldsymbol{\mu}_i + tr_i \cdot \mathbf{std}_i$ and $\mathbf{LB}_i = \boldsymbol{\mu}_i - tr_i \cdot \mathbf{stdp}_i$ respectively, the classification rule can be stated as

$$\textit{min}\ \ tr_i \quad \text{subject to}: \quad \mathbf{LB}_i \le \mathbf{x} \le \mathbf{HB}_i \qquad i = 1,\ 2,\ \cdots,\ c \tag{3}$$

By increasing tr_i each VT can be broadened so that $\mathbf{x}$ lies totally inside the corresponding VT.

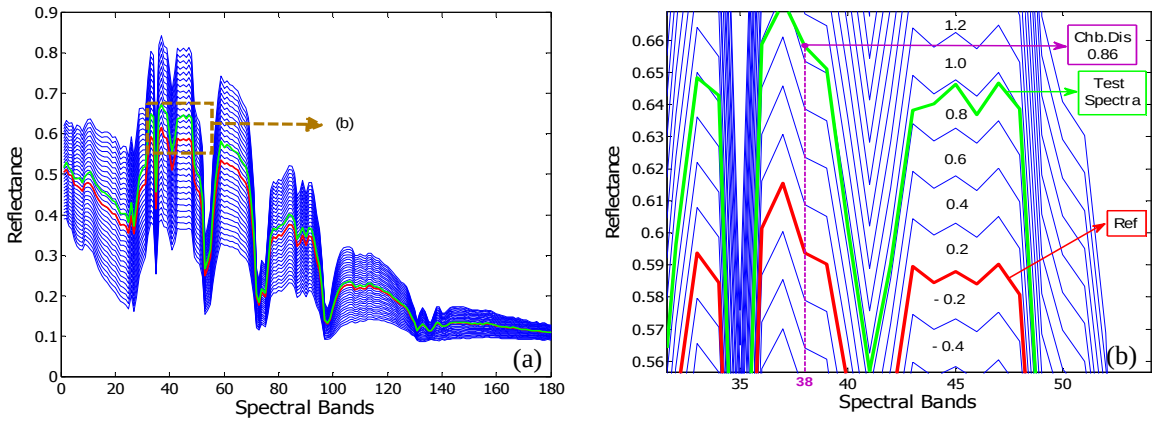

Figure 1. (a) VTs for different values of tr_i, (b) optimal VT defined by WCD $\eta_i(\mathbf{x})$ shown by green curve.

Thus, VT which hosts the test pixel vector $\mathbf{x}$ with the smallest tr_i is decided to be the right VT and $\mathbf{x}$ is assigned to the class to which that VT belongs. In practical implementation, for each class, the optimization of tr_i can be achieved as follows: the value of tr_i is initialized with a small value and is increased by small increments until (1) is satisfied. Figure 1.(a) shows how VTs are generated for nonzero tr_i values around a class signature (red line in the figure). This procedure is repeated for all classes. We observe that minimum tr_i approximately corresponds to the narrowest VT (in terms of tr_i) according to (1). Hence, we choose this as the decision criterion: a pixel is assigned to the *i*'th class which has the minimum tr_i value representing the narrowest VT. In this way, all pixels in the image are assigned as being a member of only one class.

Writing (1) in another format and making proper arrangements show that the VT method is equivalent to minimizing WCD. We start by writing (1) as

$$-tr_i \cdot \mathbf{std}_i \le \mathbf{x} - \boldsymbol{\mu}_i \le tr_i \cdot \mathbf{std}_i \qquad i = 1,\ 2,\ \cdots,\ c \tag{4}$$

Eq (4) must be satisfied for each entry of the vectors $\mathbf{std}_i$, $\mathbf{x}$ and $\boldsymbol{\mu}_i$. Therefore, we can express the requirement in (4) for each entry as

$$-tr_i \le \frac{1}{\sigma_{ij}^{\,p}}(\mathbf{x}_j - \boldsymbol{\mu}_{ij}) \le tr_i \qquad \begin{matrix} i = 1,\ 2,\ \cdots,\ c \\ j = 1,\ 2,\ \cdots,\ L \end{matrix} \tag{5}$$

where *i* refers to the class number and *j* refers to the band number. After further arranging (5),

$$\left| \frac{1}{\sigma_{ij}^{\,p}}(x_j - \mu_{ij}) \right| \le tr_i \qquad \begin{matrix} i = 1,\ 2,\ \cdots,\ c \\ j = 1,\ 2,\ \cdots,\ L \end{matrix} \tag{6}$$

is obtained. If (6) is satisfied for maximum of left hand side, then it is satisfied for all *j*. Thus, we can combine the condition for all *j* in a single statement as

$$\max_{1 \le j \le L} \left| \frac{1}{\sigma_{ij}^{p}} (x_j - \mu_{ij}) \right| \le tr_i \qquad \begin{matrix} i = 1, 2, \cdots, c \\ j = 1, 2, \cdots, L \end{matrix} \tag{7}$$

For given two L-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$, the weighted l_∞, or the WCD, is defined by [21]

$$d_\infty(\mathbf{x}, \mathbf{y}) = \left\| diag\,(w_j)(\mathbf{x} - \mathbf{y}) \right\|_\infty = \max_{1 \le j \le L} (w_j \left| x_j - y_j \right|) \qquad j = 1, 2, \cdots, L \tag{8}$$

where $diag\ (w_j)$ is a diagonal matrix whose diagonal entries are weights w_js which multiply j[th] entry of vector $\mathbf{x} - \mathbf{y}$. By comparing (7) and (8), we observe that the left-hand side of (7) is the weighted Chebyshev distance given by

$$\eta_i(\mathbf{x}) = d_{i,\infty}(\mathbf{x}, \boldsymbol{\mu}_i) = \left\| diag(\frac{1}{\sigma_{ij}^{p}})(\mathbf{x} - \boldsymbol{\mu}_i) \right\|_\infty \qquad i = 1, 2, \cdots, c \tag{9}$$

Choosing the smallest tr_i among the classes to decide the class chosen corresponds approximately to the final decision criterion given by

$$Class\ (\mathbf{x}) = \arg \min_i \{\eta_i(\mathbf{x})\} \qquad i = 1, 2, \cdots, c \tag{10}$$

An example of this procedure is illustrated in Figure 1. The green curve based on $\eta_i(\mathbf{x})$ computed by (9) shows the chosen VT.

In summary, there are three steps of the proposed WCD algorithm as follows:

1) The mean and standard deviation vectors of the classes are calculated by using the training set.

2) The WCD measures of the test pixels are calculated for all classes by using (9).

3) Each test pixel is labeled to a class based on the minimum WCD measure according to (10).

A related supervised classification algorithm is the parallelepiped (PP) method [36-37]. The PP method does not use parameters as in the WCD method. The use of weighted Chebyshev distance in terms of the inverse *p* th power of feature standard deviations and optimization of *p* constitute the most important differences between the WCD and PP methods.

2.2 WCD Method for Target Detection Application

The VT consists of the L-dimensional target class mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_L)^T$ and the L-dimensional target class standard deviation vector $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \cdots, \sigma_L)^T$ estimated from the data [36-41]. Using these statistical parameters and the Constant False Alarm Rate (CFAR) threshold parameter η, the decision rule is based on whether the pixel is in the target VT region as

$$\boldsymbol{\mu} - \eta \cdot \boldsymbol{\sigma} \le \mathbf{x} \le \boldsymbol{\mu} + \eta \cdot \boldsymbol{\sigma} \tag{11}$$

meaning that each target pixel component must satisfy (11). Therefore, if the upper and lower decision boundaries are calculated as $\mathbf{HB} = \boldsymbol{\mu} + \eta \cdot \boldsymbol{\sigma}$ and $\mathbf{LB} = \boldsymbol{\mu} - \eta \cdot \boldsymbol{\sigma}$, the target VT region can be stated as

$$\mathbf{LB} \leq \mathbf{x} \leq \mathbf{HB} \tag{12}$$

The decision boundary generated by (12) is based on the L-dimensional VT. If the pixel value lies between low and high tunnel boundaries for all bands, it is accepted as target. If the unknown pixel value does not satisfy the criteria, it is assigned to the background category. Writing (2) in another form and making proper arrangements show that the VT can be characterized as a WCD. We start by writing (11) as

$$-\eta \cdot \boldsymbol{\sigma} \leq \mathbf{x} - \boldsymbol{\mu} \leq \eta \cdot \boldsymbol{\sigma} \tag{13}$$

Eq (13) must be satisfied for each entry of the vectors $\boldsymbol{\sigma}$, $\mathbf{x}$ and $\boldsymbol{\mu}$. Therefore, we can express the requirement in (13) for each entry as

$$-\eta \leq \frac{1}{\sigma_i}(x_i - \mu_i) \leq \eta \qquad i = 1, 2, \cdots, L \tag{14}$$

where i refers to the i th band number. (14) can also be written as

$$\left|\frac{1}{\sigma_i}(x_i - \mu_i)\right| \leq \eta \qquad i = 1, 2, \cdots, L \tag{15}$$

If (15) is satisfied for the maximum of the left-hand side, then it is satisfied for all i. Thus, we can combine the condition for all i in a single statement as

$$\max_{1 \leq i \leq L} \left|\frac{1}{\sigma_i}(x_i - \mu_i)\right| \leq \eta \qquad i = 1, 2, \cdots, L \tag{16}$$

For given two L-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$, the weighted l_∞ distance, or the WCD, is defined by [40]

$$d_\infty(\mathbf{x}, \mathbf{y}) = \left\| diag\ (w_i)(\mathbf{x} - \mathbf{y}) \right\|_\infty = \max_{1 \leq i \leq L} (w_i |x_i - y_i|) \tag{17}$$

where $diag(w_i)$ is the diagonal matrix whose diagonal entries are weights w_i's. By comparing (16) and (17), we observe that the left-hand side of (16) is the WCD $d_c(\mathbf{x})$,

$$d_c(\mathbf{x}) = d_\infty(\mathbf{x}, \boldsymbol{\mu}) = \left\| diag\ \left(\frac{1}{\sigma_i}\right)(\mathbf{x} - \boldsymbol{\mu}) \right\|_\infty \qquad i = 1, 2, \cdots, L \tag{18}$$

Using $d_c(\mathbf{x})$ distance measure, the weighted Chebyshev decision rule becomes

$$d_c(\mathbf{x}) \leq \eta \qquad i = 1, 2, \cdots, L \tag{19}$$

Thus, a test pixel is classified as belonging to the target class if (19) is satisfied. The threshold parameter η can be estimated with the aid of ROC curves as explained in the next section. It can also be selected according to the detection and false alarm rates by the user, such as max detection rate or min false alarm rate.

3 EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

The three hyperspectral images used in the experiments are Airborne Visible/ Infrared Imaging Spectrometer (AVIRIS) Indiana's Pine, Botswana and Kennedy Space Center (KSC).

AVIRIS Indiana's Pine

The AVIRIS data was acquired by the National Aeronautics and Space Administration (NASA) on 12 June 1992. It consists of 145 by 145 pixels (20m spatial resolution) and 220 spectral bands, with about two-thirds agriculture, and one-third forest or other long lasting vegetation. A total of 16 cover classes are identified in the ground truth data set. After removing the absorption and low-SNR bands, 180 remaining spectral channels were used for the study [42].

Botswana Hyperion (BOTS)

Hyperion data with 9 identified classes of complex natural vegetation were acquired over the Okavango Delta, Botswana, in May 2001 [43]. The general class groupings include seasonal swamps, occasional swamps, and woodlands. Signatures of several classes are spectrally overlapped, typically resulting in poor classification accuracies. After removing water absorption, noisy, and overlapping spectral bands, 145 bands were used for classification experiments. The embedding and classification results are reported for all 9 classes.

Kennedy Space Center (KSC)

Airborne hyperspectral data were acquired by NASA AVIRIS sensor at 18-m spatial resolution over KSC during March 1996 [43]. Noisy and water absorption bands were removed, leaving 176 features for 13 wetland and upland classes of interest. Cabbage palm hammock (c3) and broad leaf/oak hammock (c6) are upland trees; willow swamp (c2), hardwood swamp (c7), graminoid marsh (c8), and spartina marsh (c9) are trees and grasses in wetlands. Their spectral signatures are mixed and often exhibit only subtle differences. The results for all 13 classes including these “difficult” classes are reported for the embedding and classification experiments.

Table 1. AVIRIS, BOTS and KSC data sets.

	AVIRIS			**BOTS**			**KSC**	
No	**Target Classes**	**Size**		**Target Classes**	**Size**		**Target Classes**	**Size**
c1	Alfalfa	54		Water	158		Scrub	761
c2	Corn-no till	1434		Floodplain	228		Willow swamp	243
c3	Corn-min till	834		Riparian	237		Cabbage hamm	256
c4	Corn	234		Firescar	178		Cabbage palm	252
c5	Grass/Past.	497		Island interior	183		Slash pine	161
c6	Grass/Trees	747		Woodlands	199		Oak	229
c7	Grass/Past. Moved	26		Savanna	162		Hardwood swamp	105
c8	Hay-Windowed	489		Short mopane	124		Graminoid marsh	431
c9	Oats	20		Exposed soils	111		Spartina marsh	520
c10	Soybeans-no till	968					Cattail marsh	404
c11	Soybeans-min till	2468					Salt marsh	419
c12	Soybeans-clean till	614					Mud flats	503
c13	Wheat	212					Water	927
c14	Woods	1294						
c15	Bldg-Grass-Tree	380						
c16	Stone Steel Tower	95						

Classes and sample sizes for every data set are presented in Table 1.

3.1 Classification Results

The labeled classes of the AVIRIS, BOTS and KSC HSI date are divided in to three equal portions. The means of the first portion spectral signatures are used as training spectral signatures for classes. The second portion spectral signatures are used as validation data, and the last portion spectral signatures are used for testing purposes. Considering the various power of standard deviation (p) in (1), the WCD method is tested by using the reference, test and validation data given in Table 2.

Table 2. AVIRIS, BOTS and KSC training, testing and validation data sets.

Amount of data													
AVIRIS					**BOTS**					**KSC**			
Class	**Ref**	**Test**	**Val**		**Class**	**Ref**	**Test**	**Val**		**Class**	**Ref**	**Test**	**Val**
c1	18	18	18		c1	52	54	52		c1	253	255	253
c3	278	278	278		c2	76	76	76		c2	81	81	81
c6	249	249	249		c3	79	79	79		c3	85	86	85
c8	163	163	163		c4	59	60	59		c5	53	55	53
c9	6	8	6		c5	61	61	61		c7	35	35	35
c13	70	72	70		c7	54	54	54		c9	173	174	173
c14	431	432	431		c8	41	42	41		c11	139	141	139
c16	31	33	31		c9	37	37	37		c12	167	169	167
										c13	309	309	309

According to the results given in Figure 2, the best performance rates are obtained with 0.5-0.8 values of p. In the experimental study, p was selected as 0.6 as being the best value of three data. After this determination, the test and validation pixels in the image were combined to generate a larger testing set.

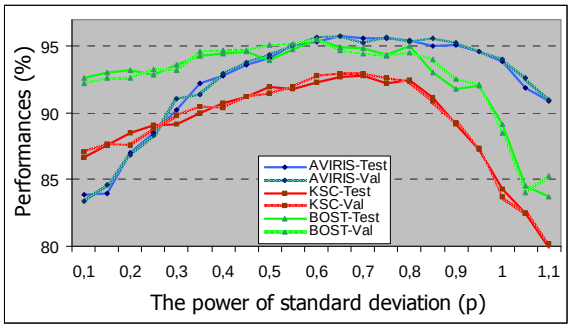

Figure 2. The variation of classification accuracy referred to as performances with respect to the power of standard deviation denoted by p.

The classification performance results of the WCD method with respect to the investigated classes are compared with the SAM and ED methods quantitatively in Figure 3. The performance values are defined as the percentage of correct labels for every class.

According to the results shown in Figure 3, the performance of the WCD method is better since the WCD performances of the classes individually are better than the SAM and ED methods for most of the classes. Considering the total classification results (overall performances) for the SAM, ED and WCD methods respectively, the WCD method is considerably better than SAM and ED methods in the experiments performed.

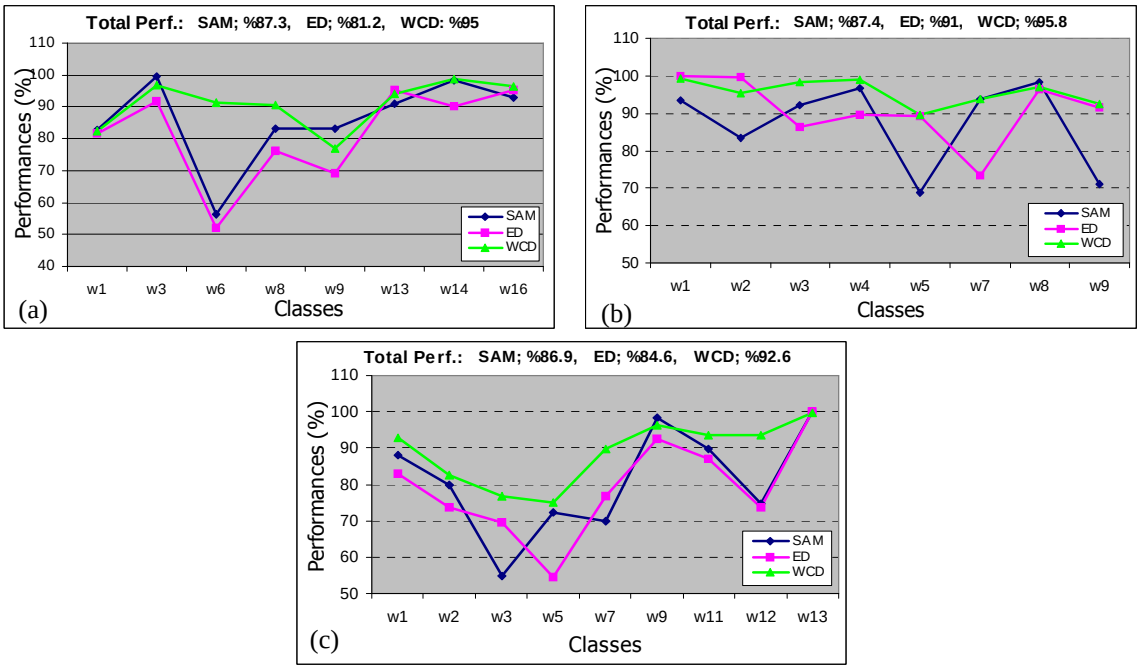

Figure 3. The performance results of the SAM, ED and WCD using (a) AVIRIS, (b) BOTS and (c) KSC data.

3.2 Target Detection Results

For investigation with the AVIRIS, BOTS and KSC data sets, representative %2, %5, %10 and %20 of target pixels (min 10 samples) were selected for each class as training set. The means of the pixel spectra in the training set were used as target spectral signatures, and all pixels in the image were utilized to determine classification accuracy. For the ED, SAM, SMF and ACE methods, corresponding equations are used respectively. Whole data cube is used during estimation of the global background covariance matrix for the SMF and ACE. All pixels in the image were checked to decide if the ED, SAM, SMF, ACE and WCD measures from the target signatures were smaller/bigger than their respective thresholds (TR).

In the experimental study, the results of the WCD detector were compared with the ED, SAM, SMF, ACE and SVM detectors quantitatively by the ROC curves. The ROC curves describe True Positive Rate (TPR) as a function of False Positive Rate (FPR). To visualize the tradeoff between TPR representing the Detection Rate (DR) and FPR standing for the False Alarm Rate (FAR), a ROC curve can be constructed by plotting TPR versus FPR for a range of TR. Due to a big amount of TR values for all cases, it is redundant to plot all possible ROC curves. Hence, the estimation of the area under the ROC curve corresponding to average detection rate (ADR) is used for capturing the detector performance [44]. One of the threshold values can be used to determine optimal detector performance according to user preferences of TPR or FPR.

The example of optimal TR values and corresponding FPR's by using ROC curves for the data sets corresponding to %90-91 TPR is given in Table 3.

Table 3 The example of optimal TR values and FPR by using ROC curves corresponding to %90-91 TPR.

Target Class	AVIRIS			BOTS			KSC		
	FPR (%)	TPR (%)	TR	FPR (%)	TPR (%)	TR	FPR (%)	TPR (%)	TR
c1	3,07	90,91	4,47	0,00	90,54	7,64	6,27	90,88	2,90
c2	28,72	90,94	2,35	3,11	90,37	3,65	11,80	90,13	3,39
c3	29,21	90,81	2,52	10,20	90,31	2,37	3,55	90,24	2,89
c4	40,71	90,05	2,43	1,36	90,48	3,07	12,74	90,08	4,31
c5	25,18	90,60	2,35	2,36	90,17	3,36	14,04	90,07	2,61
c6	11,83	91,37	2,69	21,29	91,01	1,97	18,53	90,87	3,16
c7	1,31	93,75	4,41	2,68	90,13	1,81	2,66	90,53	2,67
c8	0,99	91,59	3,12	5,77	90,35	3,72	20,69	90,02	2,45
c9	2,43	90,00	3,01	9,60	90,10	1,87	5,44	90,39	2,67
c10	24,89	91,04	2,43				16,87	90,10	2,35
c11	24,16	90,14	2,77				1,17	90,22	3,00
c12	35,84	90,60	2,35				14,93	90,06	2,94
c13	0,54	90,58	3,68				0,00	90,42	2,67
c14	11,49	90,64	3,22						
c15	39,10	90,64	2,27						
c16	3,55	90,59	4,37						

The performance results of the AVIRIS, BOTS and KSC data sets are given in Fig.4 and Fig.5 In these figures, performance is defined as the area under the ROC curve. According to Fig.4, the WCD method is better than the ED, SAM, SMF and ACE methods for most of the classes when using a small number of samples.

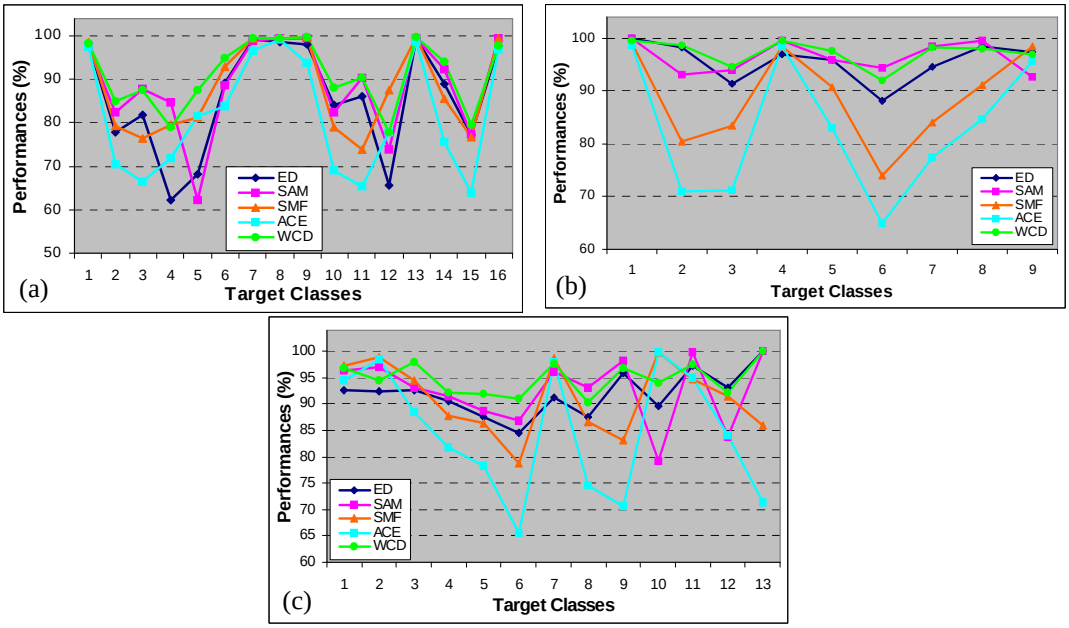

Fig.4 The SMF, ACE, ED, SAM and WCD % performances of the AVIRIS (a), BOTS (b) and KSC (c) data when using %2 training set.

It is also observed in Fig.5 that the SMF and ACE methods are becoming relatively better with good representative reference spectra. It is clear from this result that the SMF and ACE are very dependent on the amount of reference data for representative spectral signature. On the contrary, the more stable results are obtained from the ED, SAM and WCD methods by using %2, %5, %10 and %20 training sets.

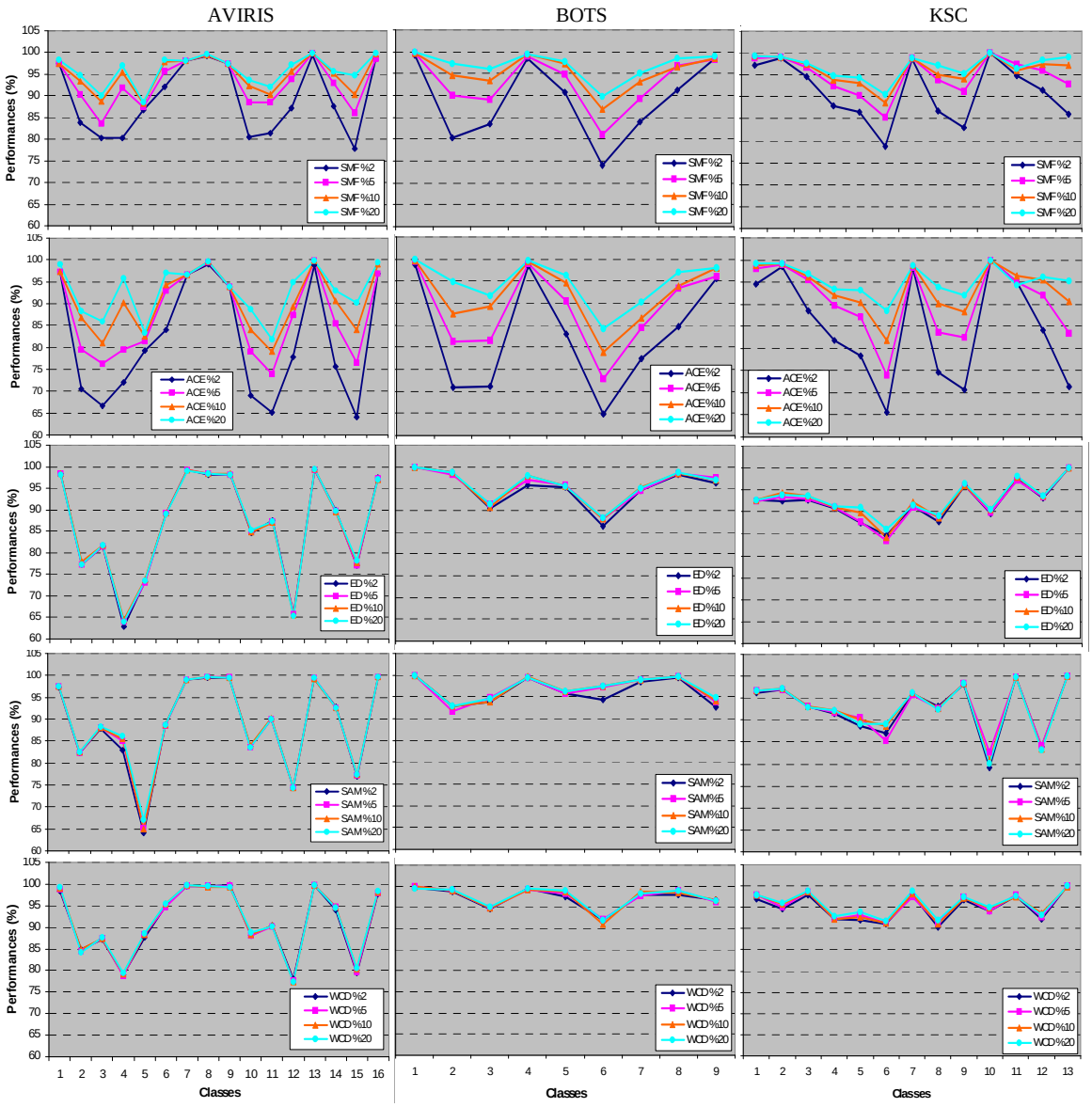

Fig.5 The comparison of the SMF, AMF, ED, SAM and WCD performances with %2, %5, %10 and %20 training sets for AVIRIS, BOTS and KSC.

The results of the SMF and ACE are becoming better for some classes using good representative data, whereas the WCD method may be preferable in most cases considering easy implementation and shorter computation times shown in Fig.6.

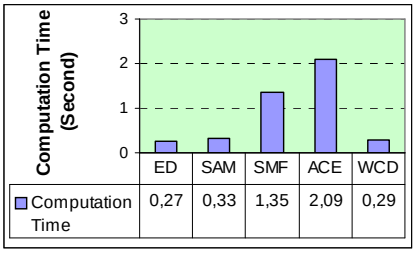

Fig.6 The computation time of the ED, SAM, SMF, ACE and WCD methods with AVIRIS data.

Next the SVM method was compared with the WCD method in several ways. For implementation of the SVM method in C, SVM^{light} software was used [45].

Table 4. CB and SB combinations of AVIRIS, BOTS and KSC data based on chosen classes.

Target Class	AVIRIS		BOTS		KSC	
	CB	SB	CB	SB	CB	SB
c1	2, 5, 11, 14, 16	2, 11	3, 5, 6, 8	3, 6	2, 4, 6, 8, 10, 12	2, 6, 12
c2	8, 5, 11, 14, 16	11, 16	3, 5, 6, 8	1, 6	1, 3, 5, 7, 9, 11	3, 7, 11
c3	8, 5, 11, 14, 16	11, 16	2, 5, 6, 8	2, 6	1, 5, 7, 9, 11, 13	3, 7, 11
c4	8, 5, 11, 14, 16	11, 16	2, 5, 6, 8	1, 3	3, 5, 7, 9, 11, 13	3, 7, 11
c5	2, 8, 11, 14, 16	14, 16	2, 4, 6, 8	1, 9	2, 3, 5, 8, 9, 11	3, 7, 11
c6	2, 8, 11, 14, 16	13, 14	2,4, 5, 8	1, 4	1, 3, 5, 8, 10, 12	3, 7, 11
c7	2, 8, 11, 14, 16	2, 11	2, 4, 6, 8	1, 4	1, 3, 5, 9, 11, 13	2, 3, 8
c8	2, 5, 11, 14, 16	3, 11	1, 3, 4, 6	1, 3	3, 5, 7, 9, 11, 13	3, 7, 11
c9	2, 8, 11, 14, 16	13, 14	2, 4, 6, 8	1, 4	2, 4, 6, 8, 10, 12	3, 7, 11
c10	2, 5, 8, 14, 16	2, 16			3, 5, 4, 7, 9, 11	3, 5, 7
c11	2, 5, 8, 14, 16	2, 16			1, 3, 5, 7, 9, 10	3, 5, 9
c12	2, 5, 8, 14, 16	2, 16			3, 5, 7, 9, 11, 13	1, 5, 9
c13	2, 8, 11, 14, 16	4, 12			1, 3, 5, 7, 9, 11	3, 7, 11
c14	2, 5, 8, 11, 16	2, 5				
c15	2, 8, 11, 14, 16	14, 16				
c16	2, 5, 8, 11, 14	6, 9				

During the implementation of the SVM, one of the classes is selected as target and three different types of background representations were considered. The first case was the full background (FB) with all classes, the second case was the complex background (CB) representation, and the last one was the simple background (SB) representation as shown in Table 4. 30 reference pixels (10 pixels for classes less than 50 pixels) were randomly chosen for target and every corresponding class of background.

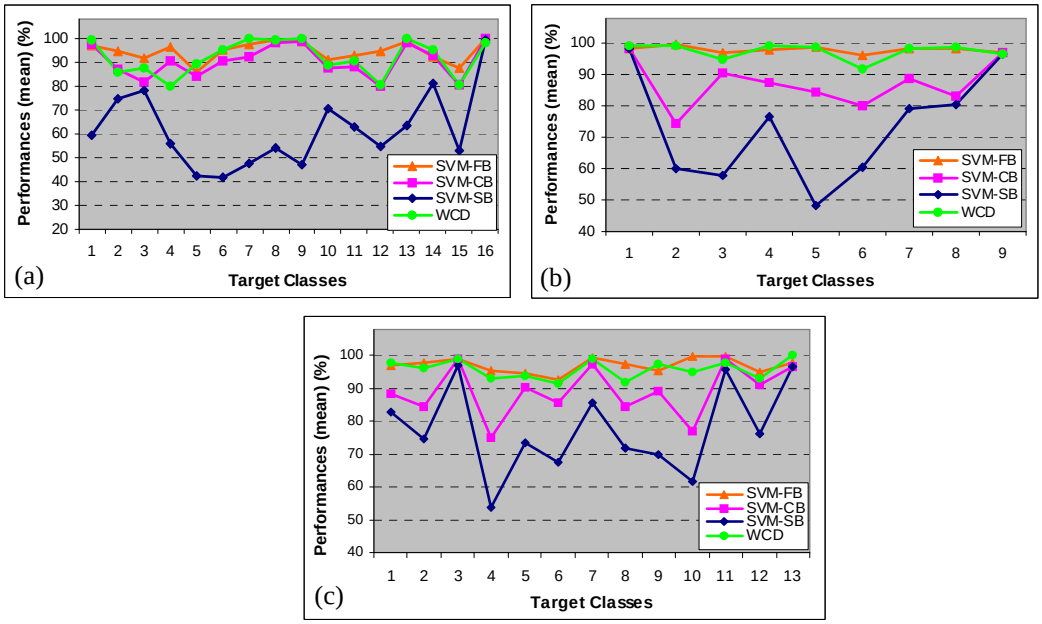

Fig.7 The WCD and SVM % performances for the AVIRIS (a), BOTS (b) and KSC (c) data when using FB, CB and SB.

With the full and complex representation of background classes, better detection performances were obtained as shown in Fig.7. With the simple representation of background classes, substantially different performance results were obtained. According to the results, the best detection rates were obtained with the SVM method using FB as illustrated in Fig.4. The performance of the WCD method is slightly better than the SVM in the case of CB for most classes. With the SB representation, the WCD method is much better than SVM for all classes. From this result, it is observed that the SVM method is highly sensitive to background selection whereas the WCD method is much more robust with different types of background because it is using only target information during the detection process.

4. CONCLUSIONS

In this study, an efficient and computationally fast spectral similarity method called WCD algorithm is introduced for HSI supervised classification and target detection applications. Estimation of upper and lower spectral boundaries of all target class spectral signatures across spectral bands constitutes the basic consideration underlying the WCD algorithm. The data sets used were the AVIRIS, BOTS and KSC data.

The VT concept leads to the use of a WCD measure per class which is defined in terms of the generalized class standard deviation. The minimum Chebyshev distance decides the chosen class of an input pixel vector. The method was investigated in supervised classification of labeled hyperspectral image data in comparison to SAM and ED methods. It is observed from the classification results that the performances of the WCD method with the AVIRIS, BOTS and KSC data are better than the classification results of the SAM and ED methods for most of the classes.

In addition, the resulting decision rule is that a test pixel belongs to the target class if its WCD measure is less than a CFAR threshold chosen by the user. The proposed algorithm was investigated with AVIRIS, BOTS and KSC data. The detection performance results were compared with the ED, SAM, SMF, ACE and radial basis SVM detectors quantitatively by the areas under the ROC curves. According to these results, the performances of the WCD algorithm are better than the ED, SAM, SMF and ACE for most of the classes in case of limited amount of reference data. The SMF and ACE methods are considerably dependent on representative reference data, and the SVM is highly sensitive to different background representation whereas the WCD is much more stable and robust in all cases.

For classification purpose, the proposed technique is not dependent on a priori statistical information about classes to a high degree except for mean vectors and feature standard deviations, or dimensional reduction. For target detection applications, it is not dependent on background reference data, second order statistical information about target and background except for target feature standard deviations estimated from the training data, or dimensional reduction.

Simplicity, high speed of computation, stability in the case of decreasing numbers of reference sample data, and robustness against different types of background make the WCD algorithm attractive for use in high dimensional target detection applications, especially with targets which are unpredictable in size and shape. It is also a computationally efficient method of classifying hyperspectral images with high accuracy.

References

[1] J.B.Campbell, and R.H.Wynne, *Introduction to Remote Sensing (5th ed)*. The Guilford Press, 667 p., 2011.

[2] Q.Wang, Q.Guo, J.Zhou and Q.Lin, "Nonlinear Joint Fractional Fourier Transform Correlation for Target Detection in Hyperspectral Image," Optics&Laser Techn., 44, p.1897–1904, Elsevier, 2012.

[3] D.Manolakis, C.Siracusa and G.Shaw, "Hyperspectral Subpixel Target Detection Using the Linear Mixing Model," IEEE Trans. on Geosc. and Rem. Sens., 39 (7), 2001.

[4] M.S.Alam, M.N.Islam, A.Bal and M.A.Karim, "Hyperspectral target detection using Gaussian filter and post-processing," Opt. and Lasers in Eng., 46 (11), p.817-822, Elsevier, 2008.

[5] D.Manolakis and G.Shaw, "Detection Algorithms for Hyperspectral Imaging Applications," IEEE Sign. Processing Magazine, 19 (1), p.29–43, 2002.

[6] A.G.S.Filho and A.C.Frery, "Hyperspectral Images Clustering on Reconfigurable Hardware Using the K-means Algorithm," IEEE SBCCI, 99–104, 2003.

[7] S.M.Hsu and H.K.Burke, "Multisensor Fusion with Hyperspectral Imaging Data: Detection and Classification," Lincoln Laboratory Journal 14(1), 145-159, 2003.

[8] J.N.Sweet, "The spectral similarity scale and its application to the classification of hyperspectral remote sensing data," IEEE WS on Adv. in Tech. for Analysis of Remotely Sensed Data, 92–99, 2003.

[9] M.E.Hodgson, "Reducing Computational Requirements of Minimum-Distance Classifier," Remt. Sens. of Envirt., 25, p.117-128, 1998.

[10] D.Manolakis, D.Marden and G.A.Shaw, "Hyperspectral Image Processing for Automatic Target Detection Applications," Linc. Lab. Journ., 14 (1), p.79-116, 2003.

[11] D.Manolakis, "Detection Algorithms for Hyperspectral Imaging Applications: A Signal Processing Perspective," IEEE WS on Advances in Tech. for Analysis of Remotely Sensed Data, p.378–384, 2003.

[12] S.A.Robila, "Spectral Screened Orthogonal Subspace Projection for Target Detection in Hyperspectral Imagery," Adv. in Pattern Recog., p.173-194, Springer LINK, 2009.

[13] D.Manolakis, "Hyperspectral Signal Models and Implications to Material Detection Algorithms," IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings (ICASSP '04), 3 (iii), p.117-20, 2004.

[14] M.D.Farrell and R.M.Mersereau, "Application of a Two-Stage Algorithm for Adaptive Detection in Hyperspectral Imaging," IEEE Worksh. on Statis. Sign. Processing, p.465-470, 2005.

[15] N.M.Nasrabadi, "Regularized Spectral Matched Filter for Target Recognition in Hyperspectral Imagery," IEEE Signal Processing Letters, 15, p.317-320, 2008.

[16] P.Bajorski, "Practical Evaluation of Max-Type Detectors for Hyperspectral Images," IEEE JSTARS, 5(2), p.462-469, 2012.

[17] E.J. Kelly, "An Adaptive Detection Algorithm," IEEE Trans. Aerosp. Electron. Syst. 22 (1), p.115–127. 1986.

[18] S.Kraut and L.L.Scharf, “The CFAR Adaptive Subspace Detector Is a Scale-Invariant GLRT,” IEEE Trans. Signal Process. 47 (9), pp. 2538–2541, 1999.

[19] S.Kraut, L.L.Scharf and L.T.McWhorter, “Adaptive Subspace Detectors,” IEEE Trans. Sig. Proc. 49(1), p.1–16, 2001.

[20] L.L.Scharf and B.Friedlander, “Matched Subspace Detectors,” IEEE Trans. Signal Process. 42 (8), p.2146–2157, 1994.

[21] J.C.Harsanyi and C.-I.Chang, “Hyperspectral Image Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Approach,” IEEE Trans. Geosci. Remote Sens. 32 (4), p.779–785, 1994.

[22] P.Bajorski, “Target Detection Under Misspecified Models in Hyperspectral Images,” IEEE JSTARS, 5(2), p.470-477, 2012.

[23] H.Kwon, N.M.Nasrabadi, “Kernel Orthogonal Subspace Projection for Hyperspectral Signal Classification,” IEEE Trans. on Geosc. and Remote Sensing, 43 (12), 2005.

[24] J.A.Gualtieri, “Hyperspectral Analysis, the Support Vector Machine, and Land and Benthic Habitats,” IEEE Workshop on Adv. in Tech. for Anls. of Rem. Sens. Data, p.354-363, 2003.

[25] V.N.Vapnik, Statistical Learning Theory. NY: Wiley, 1998.

[26] G.Matasci, D.Tuia, and M.Kanevski, “SVM-Based Boosting of Active Learning Strategies for Efficient Domain Adaptation,” IEEE JSTARS, 5 (5), p.1335-1343, 2012.

[27] N.G.Kasapoğlu, Border Feature Detectıon and Adaptatıon: A New Algorıthm for Classıfıcatıon of Remote Sensıng Images. PhD Thesis, Istanbul Technical University, 2007.

[28] F.Melgani and L.Bruzzone, “Classification of Hyperspectral Remote Sensing Images With Support Vector Machines,” IEEE Trans. on Geos. and Remote Sensing, 42 (8), 2004.

[29] J.B.MacQueen, "Some Methods for classification and Analysis of Multivariate Observations," Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability 1, 281-297, 1967.

[30] D.Landgrebe, “Hyperspectral Image Data Analysis,” IEEE Signal Processing Magazine. (02): 1053-5888, 2002.

[31] D.A.Landgrebe, Signal Theory Methods in Multispectral Remote Sensing. J. Wiley&Sons, Inc., 508, Hoboken, New Jersey, USA, 2003.

[32] P.H.Hsu, “Feature extraction of hyperspectral images using wavelet and matching pursuit,” ISPRS Jour. of Photo. and Remote Sensing, 62 (2), p.78-92, June 2007.

[33] P.H.Hsu, Spectral Feature Extraction of Hyperspectral Images Using Wavelet Transform. Ph.D Thesis, Department of Surveying Engineering, National Cheng Kung University, Taiwan, 2003.

[34] P.H.Hsu, Y.H.Tseng, P.Gong, “Dimension reduction of hyperspectral images for classification applications,” Geographic Information Sciences 8 (1), 1–8, 2002.

[35] P.Du, K.Tan and X.Xing, “Wavelet SVM in Reproducing Kernel Hilbert Space for Hyperspectral Remote Sensing Image Classification,” Opt. Comn., 283 (24), p.4978-4984, 2010.

[36] J.R.Jensen, Introductory Digital Image Processing, A Remote Sensing Perspective. Prentice Hall Inc., Upper Saddle River USA, 316, 1996.

[37] S.Demirci, B.Yazgan, and O.Ersoy, “Multispectral target detection by statistical methods,” IEEE RAST, 653-659, 2005.

[38] S.Demirci, I.Erer, “Unsupervised classification of hyperspectral images using an Adaptive Vector Tunnel classifier,” Proc. SPIE8537 Image and Signal Processing for Rem. Sen., 85371L, 2012.

[39] S.Demirci, I.Erer, and N.Unaldi, “Multi-scale vector tunnel classification algorithm for hyperspectral images”, Proc. SPIE8743-24 Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery, 2013.

[40] S.Demirci, I.Erer, and O.Ersoy, “Vector tunnel algorithm for hyperspectral target detection”, Proc. SPIE9088-12, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery, 2014.

[41] S.Theodoridis, and K.Koutroumbas, Pattern Recognition, Fourth Edition, Elsevier Inc., 525 B Street, Suite 1900, San Diego, California 92101-4495, USA, 984 (2009).

[42] AVIRIS Indiana's Pine hyperspectral data cube is available online at website: https:// engineering.purdue.edu/~biehl/ MultiSpec/hyperspectral.html (12.11.2013).

[43] D.Lunga, and O.Ersoy, “Multidimensional Artificial Field Embedding with Spatial Sensitivity,” IEEE Transactions on Geoscience and Remote Sensing, 52 (2), p.1518-1532, 2014.

[44] P.Bajorski, E.Ientilucci, and J.Schott. “Comparison of Basis-Vector Selection Methods for Target and Background Subspaces as Applied to Subpixel Target Detection”, Proc. SPIE 5425, Algorithms and Techn. for Multispectral, Hyperspectral, and Ultraspectral Imagery X, 2004.

[45] SVMlight software is available online at website (17.02.2013): http://svmlight.joachims.org/ .