

Application of local supervised feature selection approach to target detection in hyperspectral imagery

Amir Moeini Rad ¹, Ali Akbar Abkar ², Barat Mojaradi ^{3*}

¹Department of Photogrammetry and Remote Sensing, Faculty of Geodesy and Geomatics Engineering, K.N. Toosi University of Technology, Tehran, Iran

²GeoInfoSolutions BV, Almelo, the Netherlands

³Department of Geomatics Engineering, School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran

Article history:

Received: 1 October 2018, Received in revised form: 25 March 2019, Accepted: 29 March 2019

ABSTRACT

Feature selection (FS) for target detection (TD) attempts to select features that enhance the discrimination between the target and the image background. Moreover, TD usually suffers from background interference. Therefore, features that help detectors suppress the background signals and magnify the target signal effectively are considered more useful. Accordingly, in this paper, a supervised FS method, called autocorrelation-based feature selection (AFS), is proposed based on the TD concept. This method uses the image autocorrelation matrix and the target signature in the detection space (DS) for FS. Features that increase the first-norm distance between the target energy and the mean energy of the background in DS are selected as the optimal features. To evaluate the proposed method and to explore the impact of FS on the TD performance, the target detection accuracy (TDA) measure is employed. The experiment shows that the proposed FS method outperforms the two existing FS methods used for comparison. In fact, AFS achieves the maximum TDA value of 19.02% using 58 features while, compared to FS, the other methods achieve much lower values. Furthermore, the effect of image partitioning on the TD performance in both full-band and reduced-dimensionality feature spaces is investigated. The experiment results show that partitioning, as a way of adding local spatial information to TD, dramatically improves the TD performance. For experiments, the HyMap dataset is employed.

KEYWORDS

Supervised Feature Selection
Target Detection
Background Suppression
Hyperspectral Imagery

1. Introduction

Hyperspectral imagery (HSI) provides scientists with various applications such as improved classification map production (Landgrebe, 1999), subpixel target and anomaly detection (Xue, et al., 2017; Manolakis, Siracusa, & Shaw, 2001; Chang & Chiang, 2002) and spectral unmixing (Bioucas-Dias, et al., 2012). However, due to the high number of spectral bands, huge data volumes are produced. Furthermore, much of the information supplied by hyperspectral data is redundant since the adjacent spectral bands are highly correlated. Hence, many studies have been conducted on reducing the data dimensionality by feature selection (FS) and feature extraction (FE) methods. In

general, the goal of dimensionality reduction (DR) is to reduce the data volume, enhance the computing process and the accuracy of analyses (Kuo & Landgrebe, 2004).

In general, there are three main categories of FS methods: filter, wrapper, and embedded methods. The filter methods select features independent of the subsequent image analysis to be conducted, such as classification. Several filter methods have been developed, which use criteria such as correlation coefficient, entropy, mutual information, first/second spectral derivative, contrast, and spectral ratio (Bajcsy & Groves, 2004). Other methods such as fuzzy logic (Basak, De, & Pal, 1998), Inf-FS (Roffo, Melzi, & Cristani, 2015), support vector machine recursive feature elimination (SVM-RFE), Relief-F and correlation-based feature selection (CFS)

* Corresponding author

E-mail addresses: amoinirad@mail.kntu.ac.ir (A. Moeini Rad); abkar@alumni.itc.nl (A.A. Abkar); mojaradi@iust.ac.ir (B. Mojaradi)

DOI: 10.22059/eoge.2019.279956.1047

have also been used for FS (Ma, et al., 2017). In another class of filter methods, optimal features are selected based on the geometrical properties of bands in the prototype space (PS) such as prototype space-based feature selection (PFS) and maximum tangent discrimination (MTD) (Mojaradi, Abrishami-Moghaddam, Zoj, & Duin, 2009; Ghamary Asl, Mobasheri, & Mojaradi, 2014).

Other than filter methods, there are wrapper FS methods. These methods iteratively select bands based on the classification results. Random Forest (RF) and SVM classifiers have been used as wrappers (Ma, et al., 2017). The genetic algorithm has also been used as a wrapper method (Raymer, Punch, Goodman, Kuhn, & Jain, 2000). In addition to the filter and wrapper FS methods, there are embedded FS methods. The embedded methods are quite similar to the wrapper methods. The difference to the wrapper methods is that an intrinsic model building metric is used during the learning process. An example of embedded methods is L1 regularization (LASSO) (Tibshirani, 1996). Furthermore, there is a new type of FS methods called deep learning methods, such as Convolutional Neural Network (CNN), which are used for both FS and FE to improve the classification accuracy (Chen, Jiang, Li, Jia, & Ghamisi, 2016). These methods can be regarded as wrapper or embedded.

It should be noted that the methods developed for FS/FE are mostly used for classification. However, classification and detection are conceptually different. In fact, these targets are composed of a small number of pixels. Moreover, the targets are usually smaller than the ground sampling distance (GSD). Hence, no spatial and statistical information can be practically extracted about them. Furthermore, a challenge is to suppress the background since it deteriorates the accuracy of target detection (TD). In this regard, features that further discriminate the targets from the background are regarded as optimal. Therefore, new DR methods are required to be developed for TD.

Filter FS methods, supervised and unsupervised, aimed at improving TD have been recently developed. In an unsupervised manner, In (Cao, Wu, Tao, & Jiao, 2016) regarded the bands that produce better edge maps as more informative. A band search strategy based on the Particle Swarm Optimization (PSO) along with a supervised target-background separation ratio is proposed by (Xu, Du, & Younan, 2017) to improve TD. Some other supervised FS methods use band clustering and ranking. For instance, in clustering-based band selection (CBS) (Datta, Ghosh, & Ghosh, 2013), band clustering is conducted using the density-based spatial clustering of applications with the noise (DBSCAN) algorithm (Ester, Kriegel, Sander, & Xu, 1996). Variable-number variable-band selection (VNVBS) (Wang & Chang, 2007) uses the concept of orthogonal

subspace projection (OSP) to select the bands containing the most discriminatory information.

Constrained energy minimization (CEM)-based constrained band selection (CCBS) (Chang & Wang, 2006) also exploits the CEM detector to select the bands. In this unsupervised filter method, each band vector is treated as the desired target vector and all of the remaining band vectors are regarded as undesired. The output of the FIR filter is used as the FS criterion. Constrained band subset selection (CBSS) (Yu, Song, & Chang, 2018) is another unsupervised FS method very similar to CCBS. The difference is that CBSS constrains multiple bands as a band subset as opposed to CCBS which constrains a single band as a singleton set. In band add-on (BAO), a decomposition of the spectral angle mapper (SAM) is presented to select optimal bands based on the angular separation (Keshava, 2004). (Du, 2003) proposes an unsupervised method that selects optimal features using the skewness or kurtosis concept. The highly correlated features are removed using the divergence of bands.

In contrast, there are wrapper FS methods designed for TD. Curve area and genetic theory (CAGT) (Wang, Huang, Liu, & Wang, September 2014) use the genetic algorithm as the search strategy to find the optimal feature subset. The receiver operating characteristics (ROC) (Fawcett, 2006) area under curve (AUC) is used as the supervised criterion to evaluate each band subset. Furthermore, there are embedded FS methods designed for TD. A supervised sparsity-based method called LASSO-based band selection (LBS) has been developed to select different band subsets based on the CEM detector (Du, Ren, & Chang, 2003; Sun, Geng, & Ji, 2015). The band subset minimizing the finite impulse response (FIR) filter in CEM is selected as optimal.

An ideal FS method is one that helps target detectors discriminate the target and the image background more effectively. Therefore, developing an FS method based on the TD concept would help select more discriminative bands. In this regard, the proposed supervised method called autocorrelation-based feature selection (AFS) innovatively utilizes the image autocorrelation matrix as well as the target signature to simultaneously include the target and the background separation information in the FS process. This contradicts many unsupervised FS methods that rely only on the background information and many supervised methods that use only the target information and limited background information. The other innovative feature of this method is that it introduces the detection space (DS) to select optimal features. Indeed, the performance of an FS method depends on both the FS criterion and efficient space for data representation. Therefore, it must be emphasized that AFS has two key characteristics, i.e., (1) it implements a geometrical FS criterion in DS and (2) exploits the target and the background information together as opposed to almost all of the existing FS methods.

This paper is organized as follows. In Section 2, the datasets used for experiments are introduced. Then, three existing TD-based FS methods, i.e., BAO, LBS, and SKBS are briefly introduced for comparison. Furthermore, the concept behind the proposed FS method is explained completely followed by the explanation of the evaluation measures. Section 3 presents the experiments, results, and discussions in detail. The CEM detector is employed for detection. Finally, the conclusion is presented in Section 4.

2. Data and Methods

2.1. Hyperspectral Data

The HyMap dataset was collected in July 2006 in Cooke City, Montana, USA (Snyder, et al). The GSD of the image is approximately 3 meters. The data contains 126 spectral bands in the VNIR-SWIR range. Water absorption and low signal-to-noise bands were removed, leaving 112 bands. Civilian vehicles and small fabric panels are used as targets. The locations of the ground truth objects or targets are shown in circles in Figure 1. Zoomed views of the target locations can be seen in Figure 2. The details of the number, size, and type of targets are given in Table 1. The HyMap dataset and all the information regarding the ground truth given in Figure 1, Figure 2 and Table 1 are available online at (Target Detection Blind Test).



Figure 1. Part of the HyMap color image of Cook City, Montana, USA



Figure 2. Zoomed views of the fabric panels (left) and vehicles (right) used as targets in the HyMap dataset

It must be added that the ground truth samples consist of pure and mixed pixels as well as guard pixels that separate the target panels from the surrounding areas. In our experiments, the guard pixels were removed from the targets list. Furthermore, for targets F1 and F2, the spectra of the pure pixels were given as input to target detectors. F3 and F4 are arranged in two sizes. Therefore, they are considered as

F3a, F3b, F4a, F4b. F3a and F4a have central pure pixels, but F3b and F4b only consist of mixed pixels. As a result, for F3a and F4a, pure pixels and, for F3b and F4b, mixed pixels are used as the target spectra. For targets V1, V2 and V3, only mixed pixels were available, which were used inevitably as the input to the detectors. Meanwhile, the HyMap dataset is geometrically and radiometrically corrected, and the pixel values demonstrate reflectance spectra.

Table 1. Details of the fabric panels and vehicles used as targets in the HyMap Dataset

Names	Materials and Sizes	Number of Pixels
F1	3m Red Cotton Panel	9
F2	3m Yellow Nylon Panel	9
F3a	2m Blue Cotton Panel	9
F3b	1m Blue Cotton Panel	1
F4a	2m Red Nylon Panel	9
F4b	1m Red Nylon Panel	1
V1	1993 Chevy Blazer	1
V2	1997 Toyota T100	1
V3	1985 Subaru GL Wagon	1

Figure 3 shows the reflectance spectra of the targets. As seen, most of the targets have spectral responses similar to those of the background vegetation. This implies a challenging condition to separate targets from the background.

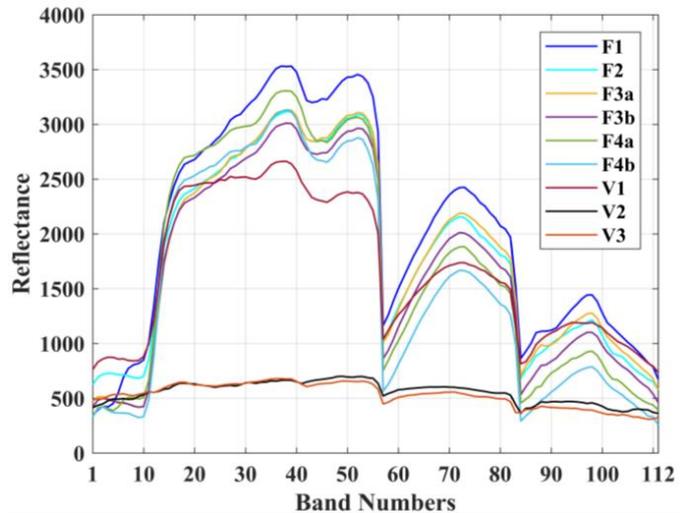


Figure 3. The reflectance spectra of targets in the HyMap data

2.2. BAO

This supervised method is developed based on the SAM concept, where bands producing the maximum angles between the target and a reference signature are selected as optimal. The image mean vector is used as the reference signature. FS starts from an initial band subset, and the remaining bands are added one by one to the initial subset. The bands that minimize the cost function are the ones which maximize the discrimination between the target and the background (Keshava, 2004).

This method suffers from the fact that it neglects the background information. Indeed, the image mean vector is not a good representative for the background, especially in a case such as TD where subtle spectral information is required to separate the target and the background.

2.3. LBS

This method operates in a supervised manner based on the TD concept, where the difference between the target abundance map in the full dimension and the abundance map produced by the selected bands is used as the criterion for FS. The CEM vector in the reduced dimension is estimated using the linear regression model with L1 regularization (LASSO). The indices of the non-zero elements in the estimated CEM vector are interpreted as the optimal band numbers (Du, Ren, & Chang, 2003; Sun, Geng, & Ji, 2015).

This method uses a potentially convincing FS criterion in terms of TD. However, the proposed criterion is mathematically complex to be solved and, unfortunately, does not precisely yield the desired solution, i.e., the number of bands selected is not the same as the number of bands intended to be selected. In other words, the method cannot select the desired number of bands given as input to the algorithm.

2.4. SKBS

Skewness-based band selection (SKBS) is an unsupervised FS method that uses the skewness or kurtosis or a combination of these two statistical parameters to select the optimal features. This method sorts the spectral bands based on these parameters. It is assumed that the bands with greater skewness or kurtosis values contain more useful information. Then, the sorted bands are analyzed in pairs using the divergence concept. This step is used to remove the bands that are very similar or highly correlated (Du, 2003).

Although this FS method is conceptually simple and can be easily implemented, it does not take into account the target information. It only attempts to extract information on the background. Therefore, SKBS may not be able to select informative bands specifically helpful for TD.

2.5. The Proposed Autocorrelation-based Feature Selection (AFS) Algorithm

Feature Selection (FS) based on the target detection (TD) concept can help select more discriminative bands by using a projection vector consisting of the image autocorrelation matrix and the target signature. Moreover, it must be emphasized that using an appropriate space for the representation of bands and employing an effective criterion developed specifically for TD-based FS can lead to more informative bands to be selected. In this regard, the proposed FS method is presented as follows.

Conceptually, TD generally occurs in two steps. In the first step, the target signal is magnified prior to background

suppression. Then, in the second step, the background is suppressed by projecting the image into a new subspace. Based on this concept, in order to develop the autocorrelation-based feature selection (AFS) method, we introduced a new vector called the ‘TD projector’ defined as follows:

$$\mathbf{k}_{L \times 1} = \mathbf{R}_{L \times L}^{-1} \mathbf{d}_{L \times 1} \quad (1)$$

where \mathbf{k} is the TD projector, L is the number of spectral bands, \mathbf{d} is the desired target signature and \mathbf{R}^{-1} is the inverse of the image autocorrelation matrix \mathbf{R} :

$$\mathbf{R}_{L \times L} = \frac{1}{N} \sum_{i=1}^N \mathbf{r}_i \mathbf{r}_i^T \quad (2)$$

where N is the number of image pixels and \mathbf{r}_i is an $(L \times 1)$ -dimensional image pixel vector.

The TD projector in equation (1) is embedded in the CEM and adaptive matched filter (AMF) detectors. Since the background suppression is accompanied by the target energy being reduced, the first term is introduced by multiplying the image pixel vectors by \mathbf{d} . This term magnifies the target energy, which helps increase the probability of the target to be detected.

In the second term, the image pixels are transferred into the subspace spanned by \mathbf{R}^{-1} . It must be emphasized that in the TD concept, \mathbf{R} is regarded as the mean energy or power of the image. Hence, in this step, the approximately largest amount of the image background energy is suppressed.

Two points must be noted here. The first point is that the TD projector is defined based on the detection concept. Therefore, the values of the image pixel vectors, after being multiplied in an element-wise fashion by equation (1) and projected into a new subspace, change from reflectance or radiance into new values. Hence, we name this new subspace as the Detection Space (DS) hereafter, and the FS criterion for AFS will be implemented in this space. The second point is that since the TD projector uses the target signature, and the AFS method is also defined using this projector; therefore, the FS process is conducted in a supervised manner.

One other important point in equation (1) is that the inverse of the sample autocorrelation matrix plays a key role in suppressing the image background. In fact, in \mathbf{R} , each diagonal element i can be regarded as the mean energy of the i^{th} band in the original feature space. Therefore, \mathbf{R}^{-1} in \mathbf{k} minimizes the energy of individual bands. In this regard, if one does an element-wise product of \mathbf{k} and the diagonal in \mathbf{R} , the product can be considered as the mean value of the background in different bands in DS. Furthermore, the detectors such as CEM use the second power of \mathbf{k} to suppress \mathbf{R} ; hence, we also used the second power of \mathbf{k} in order to be consistent with the TD concept:

$$\mathbf{s}_{L \times 1} = \text{diag}(\mathbf{R}) \quad (3)$$

$$\mathbf{e}_{L \times 1} = (\mathbf{k}^{\wedge 2}) * \mathbf{s} \quad (4)$$

where s is the vector containing the diagonal elements of \mathbf{R} and e is the mean energy of the image bands in DS. The asterisk in equation (4) is used to denote the element-wise multiplication of vectors. $\mathbf{k}^{\wedge 2}$ also means the element-wise second power. Likewise, the target can be multiplied by the projector:

$$\mathbf{t}_{L \times 1} = |\mathbf{k} * \mathbf{d}| \quad (5)$$

where \mathbf{t} indicates the target values in each band in DS. The signs of the values in \mathbf{t} are not important as they only denote the direction of autocorrelation. Therefore, the absolute values of \mathbf{t} are employed in equation (5). Since \mathbf{R} is a positive matrix, e also has positive elements in equation (4). Theoretically, target detectors exploit the distance between the target and the background pixels in DS to conduct the detection process. Therefore, the higher the difference between the absolute values of \mathbf{t} and e , the higher the discrimination between the target and the background. Hence, the FS criterion is defined as follows:

$$\mathbf{a}_{L \times 1} = |\mathbf{t} - e| \quad (6)$$

where \mathbf{a} is the vector representing the separability between the target and the background. Indeed, a_i means the first-norm distance between the target and the background in the i^{th} feature. Features that yield the biggest values for a_i are regarded as optimal since they provide the highest amount of separability between the target and the background in DS.

To further clarify the FS criterion, Figure 4 displays the a_i values for all features or the difference in absolute values between \mathbf{t} and e in the full-dimensional DS, i.e., L original features. As it is seen, the sixth feature (f_6) provides the greatest absolute distance between the target and the background. Therefore, it can be selected as the first optimal feature in the full-dimensional DS. However, our analysis demonstrated that feature selection based on a backward elimination strategy, i.e., removing the features with the minimum a_i values in equation (6), which are the least discriminative ones, leads to better results. Therefore, we used the backward strategy for selecting features by AFS.

The steps of the AFS method are as follows:

- 1) For target $\mathbf{d}_{L \times 1}$, the TD projector $\mathbf{k}_{L \times 1}$ in equation (1) is computed using the autocorrelation matrix $\mathbf{R}_{L \times L}$ defined in equation (2).
- 2) $e_{L \times 1}$ and $\mathbf{t}_{L \times 1}$ are computed based on equations (4) and (5).

- 3) For feature (f_i), the a_i is computed as follows:

$$a_i = |t_i - e_i| \quad (7)$$

This step is conducted for all of the features.

- 4) Using a backward elimination strategy, the feature (f_u) with the minimum a_u value is removed from the original set of features $\mathbf{f} = \{f_1, f_2, \dots, f_L\}$:

$$f_u = \underset{f_i}{\operatorname{argmin}}(a_i), \quad i = 1, \dots, L; 1 \leq u \leq L \quad (8)$$

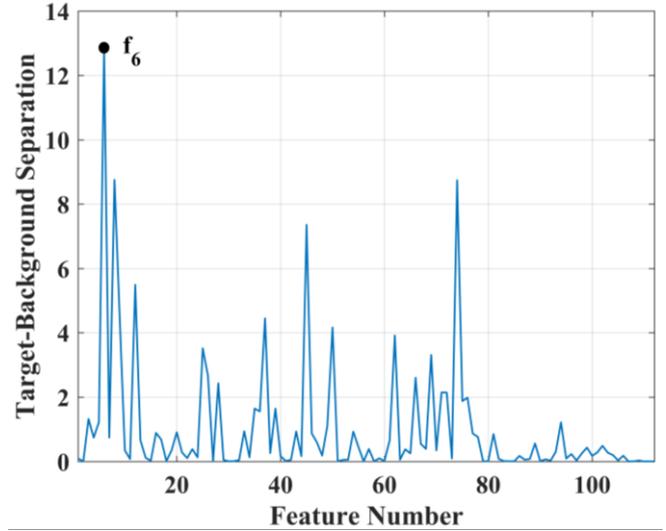


Figure 4. The absolute values of the difference between \mathbf{t} and e in all features obtained by the autocorrelation-based feature selection (AFS) method in the full-dimensional detection space (DS). (f_6) is the first optimal feature.

- 5) \mathbf{d} and \mathbf{R} are updated by removing the feature (f_u) determined in step 4.
- 6) \mathbf{k} is updated in the reduced-dimensionality DS using the new \mathbf{d} and \mathbf{R} .
- 7) The vector \mathbf{a} is computed for all of the new features in the reduced-dimensionality DS using the updated set of features \mathbf{f}' . It must be emphasized that the two feature sets \mathbf{f} and \mathbf{f}' are different in both the number of features and the features themselves.

Steps 3 to 7 are repeated until a stop criterion is satisfied, e.g., $x < L$ features are left. These features have the maximum a values in equation (7) and thus provide the highest amount of discrimination between the target and the background. The x remaining features are then used for TD. Furthermore, x can be determined based on the application or the user's needs. For example, it can be set when the number of FAs for a given feature subset is minimum compared with that of other subsets. Apart from FA, other criteria such as the number of truly detected targets or true positives (TPs) can also be employed to stop the FS process.

Therefore, it must be emphasized that AFS introduces a new space based on the TD concept in which the discriminative features can be selected based on a simple geometric criterion, i.e., the first-norm distance between the target and the background in DS.

Furthermore, as mentioned previously, it must be noticed that unsupervised FS methods only rely on the background data to select optimal features ignoring the target information while a few existing supervised FS methods mainly focus on the target information ignoring the background. In contrast, AFS simultaneously employs the full information regarding the target and the background separation using the general TD concept to select features without directly resorting to the detector feature vector. It does so by simultaneously

exploiting the autocorrelation matrix as well as the target signature.

Moreover, it must be also emphasized that many target signatures are spectrally mixed at the subpixel level, which can be detected only using subtle spectral information; hence, only a supervised TD-oriented approach to FS leads to better results.

2.6. Evaluation Measure: Target Detection Accuracy (TDA)

In this research, the focus is on both FA and true positive (TP) pixels by simultaneously minimizing FAs and maximizing TPs. Hence, after the FS process is terminated, subsets containing different numbers of selected features are employed to conduct TD by the CEM and AMF detectors. To do so, the number of features in the optimal feature subset increases progressively from 10 to 70. Then, we introduced the following equation to be applied to the detection abundance image in order to obtain the best FAs and TPs for each subset:

$$TDA_{i,t,max} = \frac{TP_{i,t}}{n_t + FA_{i,t}} \times 100, i = 10, \dots, 70 \quad (9)$$

where $TDA_{i,t,max}$ is the maximum target detection accuracy (TDA) obtained for target t using the subset containing i optimal features, $TP_{i,t}$ is the number of true positive pixels, $FA_{i,t}$ is the number of false alarm pixels, and n_t is the number of target pixels. Ideally, when $FA_{i,t} = 0$ and $TP_{i,t} = n_t$, i.e., all target pixels are detected with no FAs, then, $TDA_{i,t,max} = 100$. In the worst case, when $TP_{i,t} = 0$, i.e., the detector cannot detect any target pixels, $TDA_{i,t,max} = 0$. To be specific, in order to determine $FA_{i,t}$ and $TP_{i,t}$, a threshold must be used to convert a detection abundance map to a binary classification map. In this regard, for each target in each selected feature subset, a full range of thresholds from the minimum abundance to the maximum abundance in the detection map is used for the conversion. For each threshold, FAs and TPs are counted and $TDA_{i,t}$ is calculated. The threshold giving the maximum value, i.e., $TDA_{i,t,max}$ or the best trade-off between FAs and TPs, is used to produce the classification map. Then, the $FA_{i,t}$ and $TP_{i,t}$ values corresponding to $TDA_{i,t,max}$ are regarded as the best result accomplished by a detector using the i^{th} feature subset for target t based on equation (9).

It must be mentioned that equation (9) is designed so that both FAs and TPs are considered to evaluate the effect of different feature subsets produced by the FS methods. It must be also emphasized that, for each feature subset, the best FA and TP values are obtained for each target separately. Then, for each feature subset, the sum of best FAs and TPs of all targets are given as the final result achieved by a detector using an FS method:

$$\begin{aligned} TFA_i &= \sum_{t=1}^g FA_{i,t}, i = 1, \dots, L \\ TTP_i &= \sum_{t=1}^g TP_{i,t}, i = 1, \dots, L \end{aligned} \quad (10)$$

In equation (10), TFA_i is the total number of FAs regarding all targets, $FA_{i,t}$ is the best FA obtained for target t in equation (9), TTP_i is the total number of TPs considering all targets, $TP_{i,t}$ is the best TP gained for target t in equation (9) and g is the number of targets. i is the selected feature subset.

In order to have a better comparison of the effect of FS methods on TD, both FAs and TPs are considered as the measure of performance. In this regard, the TFA and TTP values obtained by equation (10) are used to determine the best TDA for each feature subset regarding all targets as follows:

$$TDA_i = \frac{TTP_i}{NT + TFA_i} \times 100, i = 1, \dots, L \quad (11)$$

where TFA_i and TTP_i are obtained for the i^{th} feature subset by equation (10), NT is the total number of pixels for all targets and TDA_i is the best TDA obtained for all of the targets using the i^{th} feature subset.

2.7. Evaluation Measure: Computing Speed (CS)

Finally, the time taken by the proposed FS method and the other FS methods to select optimal features are compared with each other.

3. Results and Discussions

3.1. Feature Selection: TDA Curves

Figure 5 displays the performance of CEM using the proposed AFS method for the HyMap dataset in terms of TDA. The TDA values are displayed as curves. The black horizontal line shows the full-band TDA value. FB in the figure's legend stands for 'full-band.' Meanwhile, it must be noted that, in general, the detection accuracy values are low compared with the traditional measures often employed in classification studies such as the overall accuracy and the kappa coefficient. This is because the number of target pixels is usually much lower than that of FA pixels. Therefore, the values obtained by equation (11) are mostly influenced by the number of FAs.

Figure 5 shows that the proposed AFS method has a better performance in all selected feature subsets compared with BAO, LBS, and SKBS. Therefore, regarding this experiment, it must be pointed out that the CEM detector using AFS performed much better compared when employed with other FS methods. The TDA values achieved by AFS were much higher than those of other methods in all feature subsets, especially in ones containing only a few features. The maximum TDA obtained by AFS was about 19.02% with 58 features. Furthermore, AFS also succeeded in generating TDA values higher than those of the full-band detection from subset 53 to subset 63. CEM did not obtain results better than the full-band detection using BAO, LBS, and SKBS.

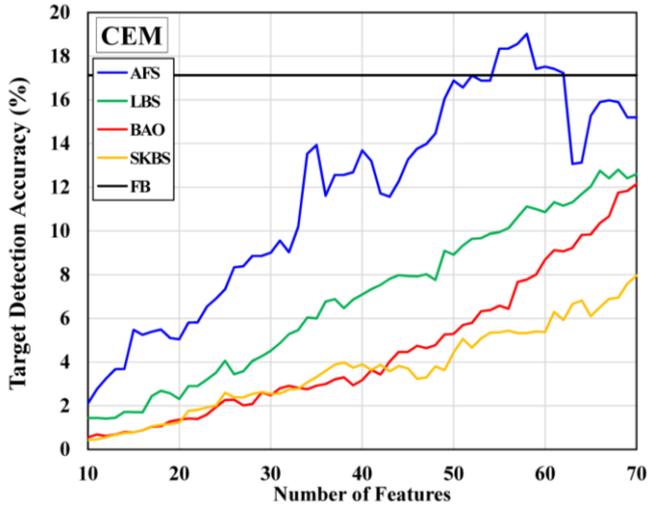


Figure 5. The best target detection accuracy (TDA) obtained by the CEM detector using four feature selection (FS) methods for the HyMap dataset. The black horizontal line indicates the full-band (FB) result for comparison with FS-based results.

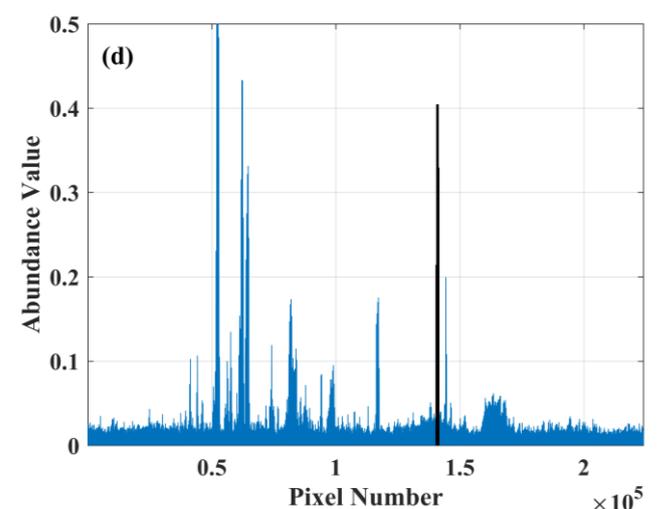
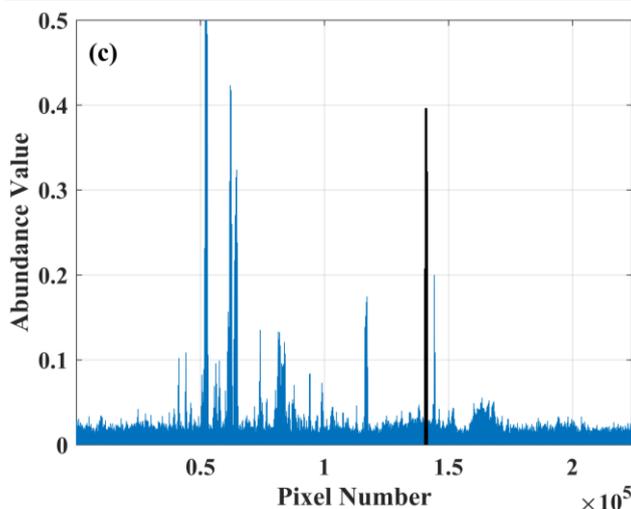
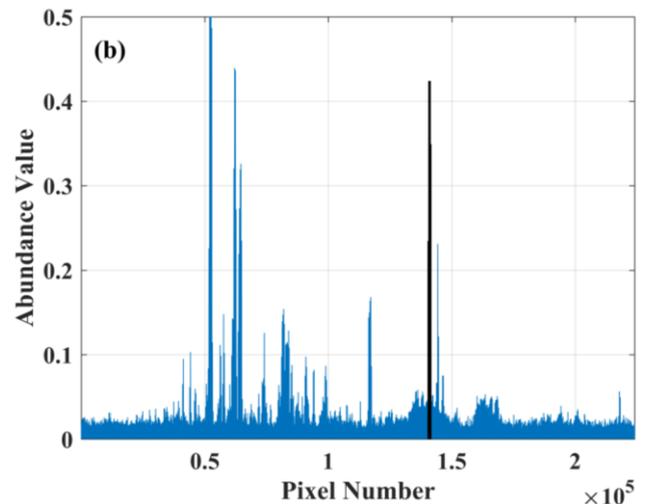
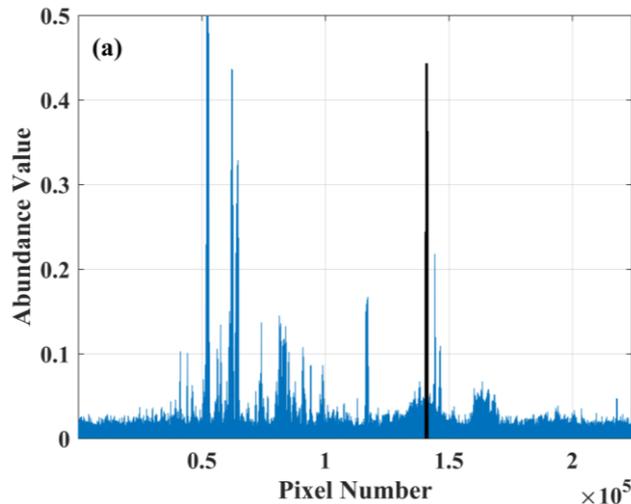


Figure 6. Pixel abundances of the detection map of the first target in the HyMap dataset obtained by CEM using 58 features. These features correspond to the maximum TDA value achieved by AFS. The FS methods are (a) AFS, (b) LBS, (c) BAO and (d) SKBS. The black vertical bars indicate the position of target pixels. AFS has help CEM gain higher values for the target pixels.

Therefore, it must be emphasized that AFS introduces a new space for FS. In contrast, BAO, LBS, SKBS, and other

existing FS methods reviewed in the introduction section use the traditional feature as well as the spectral and prototype

spaces. The proposed DS in this research is based on the detection concept and is hence better suited for TD. Therefore, the appropriate space employed to represent and analyze the features is the superiority of our proposed FS method. The other factor that differentiates AFS from the existing FS methods is the FS criterion. However, as our experiments show, the first step of defining the appropriate space in which the FS cost function is applied is more important. Other FS methods also possess acceptable FS criteria, but the space in which features are selected is not entirely suitable for TD. Moreover, one significant advantage of our research is that we have also compared our TD results with the full-band detection accuracy. This is an issue less noticed in other studies.

3.2. Feature Selection: CS

In this experiment, the amounts of time spent by the proposed FS method along with other FS methods to select and rank the complete set of optimal features are displayed as computing speeds (CS) in Table 2. The proposed AFS method consumed much less time compared with LBS and SKBS. In contrast, BAO performed FS is a slightly shorter time compared with AFS. The reason is that it only uses the target and the image mean vectors to conduct FS. However, the performance of BAO is much weaker than that of AFS in terms of FA and TDA.

Table 2. The time in seconds spent by the feature selection (FS) methods to select optimal features in the HyMap dataset.

Existing FS Methods			Proposed FS Method
LBS	BAO	SKBS	AFS
55	4	19	8

3.3. Full-Band Image Partitioning

It must be noticed that the image autocorrelation matrix R in CEM plays a significant role in suppressing the image background. Traditionally, this matrix is built globally using all image pixels. In this way, only the global spectral characteristics of the background are considered. However, subtle local spectral characteristics of the background are ignored. Therefore, locally modeling the image background can help better suppress it and consequently improve the TD performance.

In this regard, in this experiment, we intend to investigate the effect of image partitioning - as a way of localizing the image background - on TD. Particularly, the effect of the number and size of the partitions on the CEM performance in terms of both FA and TDA is studied. In order to divide the image into partitions, the MATLAB code starts with the original image and then, in each iteration i , the image is divided into $n = i^2$ partitions. The loop continues for ten times. Therefore, the number of partitions equals 1, 4, 9, 16, ..., 100. Obviously, as the number of partitions increases, the sizes of partitions decrease. In each iteration,

TD is conducted in each partition locally. Then, the global abundance map is produced by stitching partitions together. Finally, a global binary classification is run using the threshold determined by the maximum TDA.

Figure 7 demonstrates the total number of FA pixels generated by CEM for all targets in the full-band HyMap dataset. The first iteration ($i = 1$), shows the result ($n = 1$, $FA = 140$) for the original unpartitioned image. As the number of partitions increases or as the size of partitions reduces, the number of FAs decreases. This indicates that using a local R instead of a global R has a direct positive impact of the TD performance. Furthermore, Figure 8 displays the TDA values obtained by CEM with different numbers of partitions for the HyMap dataset. The first value ($n = 1$, $TDA = 17.13\%$) shows the result for the original unpartitioned dataset. It is seen that the TDA value increases sharply as the number of partitions increases. This also proves the constructive influence of local background modeling on the TD accuracy.

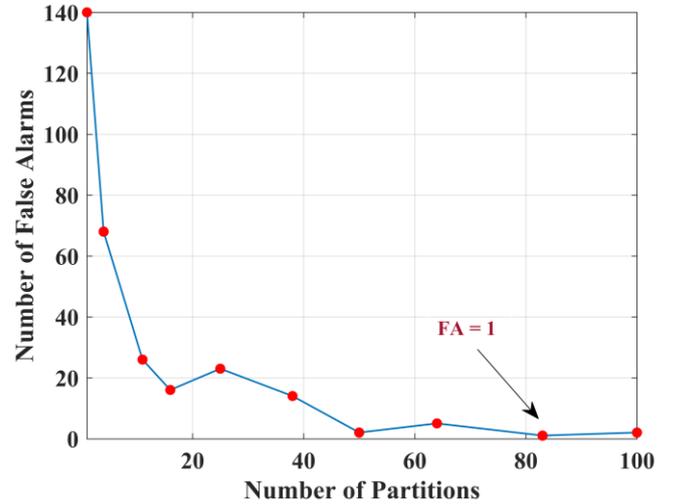


Figure 7. The number of false alarm (FA) pixels generated by the CEM detector versus the number of image partitions for the full-band HyMap dataset.

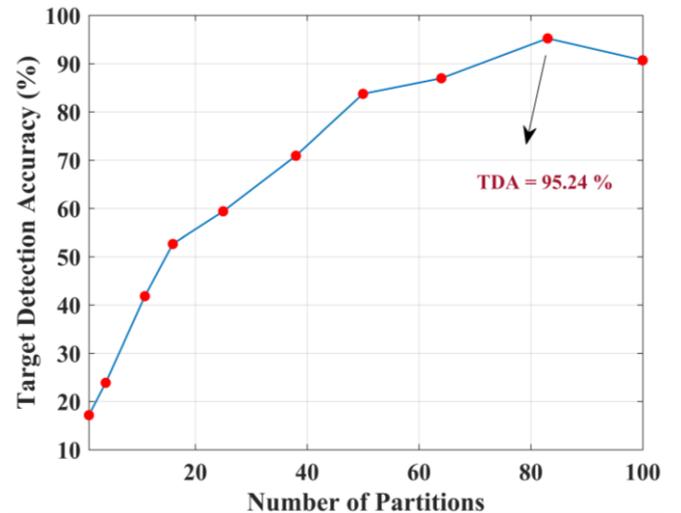


Figure 8. Target detection accuracy (TDA) achieved by the CEM detector versus the number of image partitions for the full-band HyMap dataset.

3.4. Feature Selection and Image Partitioning

Generally, in DR, whether FS or FE, some spectral information is lost. This reduction in data dimensionality can lead to some improvement in TD performance (Figure 5), since several of the highly correlated and redundant features are removed, and a subset of the most informative features are selected. In Figure 5, in some of the feature subsets, TDA values higher than the full-band result are obtained by CEM using features selected by AFS. However, in other feature subsets, especially the ones containing a few features, the TDA values are much lower than the full-band result. Therefore, we intend to demonstrate that adding spatial information to spectral information can lead to increased information content. In other words, this added spatial information can compensate the reduced spectral information specifically in the subsets containing only a few optimal features and improve TD accuracy.

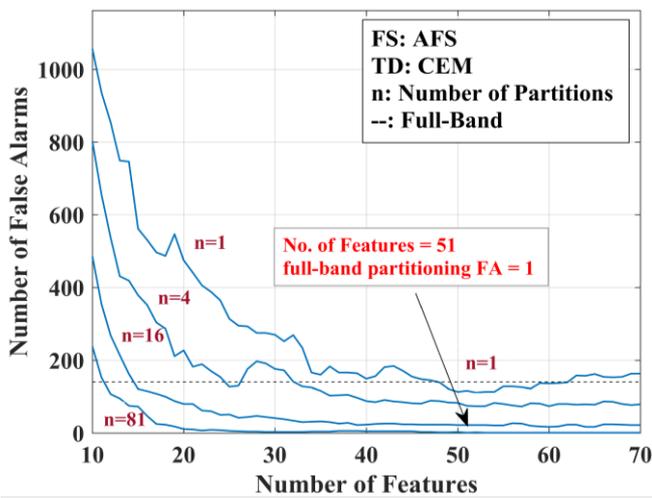


Figure 9. The number of false alarm (FA) pixels generated by the CEM detector using AFS versus the number of optimal features displayed for different numbers of partitions. n indicates the number of partitions. The horizontal dashed line demonstrates the FA value for the full-band image without partitioning.

In this regard, in this experiment, we combined image partitioning, as a way of extracting spatial information on the image background, with FS. Figure 9 displays the total number of FA pixels generated by CEM using the optimal features selected by AFS. The horizontal dashed line demonstrates the FA value for the full-band image without partitioning ($n = 1$). Four FA curves are shown corresponding to four different numbers of partitions ($n = 1, 4, 16$ and 81). $n = 1$ means the unpartitioned image. In each curve, as the number of features increases, the number of FA pixels decreases. Moreover, a comparison of the curves indicates that as the number of partitions increases, the number of FA pixels further decreases. This proves that image partitioning adds useful spatial information to FS so that the TD performance improves substantially in the reduced-dimensionality feature space.

One crucial point to be noticed is that CEM, using the full-band partitioned image, gains the minimum FA value of a single pixel with 81 partitions as seen in Figure 7. Furthermore, CEM, using AFS along with image partitioning, also yields the minimum FA value of a single pixel. This is shown in Figure 9 in a box saying that the full-band partitioning-based FA value is achieved using 51 optimal features with 81 partitions. This means that in the reduced-dimensionality image with 81 partitions, CEM, with the help of AFS, has achieved the result equal to that of the full-band image, also with 81 partitions, using only 45 % of the original features. In other words, AFS has accomplished a noticeable reduction in data volume without losing accuracy. Moreover, the result is much better in comparison with that of the full-band unpartitioned TD displayed by the dashed horizontal line.

Finally, in order to give an overall view of the information displayed in Figure 9, it must be mentioned that these comparisons are demonstrated: the FA value generated by the full-band unpartitioned TD (dashed horizontal line, $n = 1$), FA values generated in the reduced-dimensionality partitioned TD ($n = 4, 16, 81$) using different numbers of features and FA values generated in the reduced-dimensionality unpartitioned TD ($n = 1$) using different numbers of features. The arrow compares the minimum FA value of the reduced-dimensionality partitioned TD with that of the full-band partitioned TD (both FAs equal one pixel).

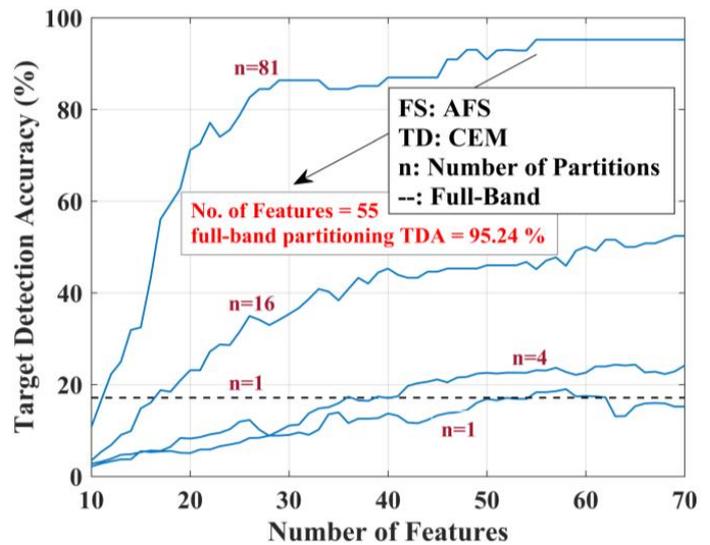


Figure 10. The target detection accuracy (TDA) achieved by the CEM detector using AFS versus the number of optimal features displayed for different numbers of partitions. n indicates the number of partitions. The horizontal dashed line demonstrates the TDA value for the full-band image without partitioning.

Figure 10 demonstrates the TDA values achieved by CEM using the optimal features selected by AFS for the HyMap dataset with different numbers of partitions ($n = 1, 4, 16$ and 81). $n = 1$ means the unpartitioned image. The horizontal dashed line demonstrates the TDA value for the full-band

image without partitioning $n = 1$. As it is seen, in each curve, the TDA values increase as the number of selected features increases. Moreover, as the number of partitions increases, the TDA value further increases. With a high number of partitions, i.e., ($n = 81$), the TDA values increase sharply with a considerable difference compared with low numbers of partitions, i.e., ($n = 4$ and 16). This shows the noticeable impact of local spatial information added to spectral information in improving the TD performance. Besides, the effect of partitioning in the feature subsets containing a low number of features compared with the full-band unpartitioned TD (dashed horizontal line, $n = 1$) and unpartitioned reduced-dimensionality TD ($n = 1$) is extraordinary. One last point about Figure 10 is that CEM has achieved a superior maximum TDA value of 95.24 % in the image with 81 partitions using only 55 optimal features, selected by AFS, or only 49 % of the original features. This TDA value equals the maximum value achieved by CEM in the full-band partitioned image with 81 partitions (Figure 8). Therefore, AFS provides CEM with a considerable amount of reduction in the data size as well as maintaining the TD accuracy.

4. Conclusions

In this research, a new feature selection method called AFS aimed at improving TD in hyperspectral imagery was proposed. For TD, CEM was employed as the detector. AFS was developed based on the image autocorrelation matrix and the target signature in DS. Three existing TD-oriented FS methods, BAO, LBS, and SKBS, were used for comparisons. For evaluation, the TDA measure was proposed and utilized with the HyMap dataset. As the experiment showed, AFS-based CEM achieved higher TDA values using different feature subsets compared with the processes in which BAO, LBS, and SKBS were used for FS. CEM, using AFS, also obtained TDA values higher than that of the full-band TD in some feature subsets. In fact, in contrast to existing FS methods, AFS introduced a new space, called DS, for feature selection, which was specifically defined for TD. This new space and the implementation of the proposed FS criterion in this space helped CEM achieve higher detection accuracy values compared with the process that used other spaces and criteria for FS.

Moreover, the effect of image partitioning was investigated on the TD accuracy using the full-band HyMap dataset. It was observed that as the number of partitions increased, the FA values decreased and, in contrast, the TDA values increased considerably. Also, the complementary effect of local spatial information on the reduced-dimensionality TD performance was explored, an issue that has not been much studied beforehand. Accordingly, partitioning was combined with AFS using four numbers of

partitions. It was observed that partitioning helped AFS achieve TDA values much higher than those of the full-band unpartitioned TD and reduced-dimensionality unpartitioned TD. Finally, it must be emphasized that the main achievement of this experiment was that AFS-based CEM managed to achieve a considerable data reduction volume of about 55 % and 51 %, regarding the FA and TDA values while maintaining the minimum FA and maximum TDA values achieved by the full-dimensional partitioned TD.

References

- Bajcsy, P., & Groves, P. (2004). Methodology for hyperspectral band selection. *Photogrammetric Engineering & Remote Sensing*, 70, 793-802.
- Basak, J., De, R., & Pal, S. (1998). Unsupervised feature selection using a neuro-fuzzy approach. *Pattern Recognition Letters*, 19, 997-1006.
- Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., & Chanussot, J. (2012). Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2), 354-379.
- Cao, X., Wu, B., Tao, D., & Jiao, L. (2016). Automatic Band Selection Using Spatial-Structure Information and Classifier-Based Clustering. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(9), 4352-4360. doi: 10.1109/JSTARS.2015.2509461
- Chang, C.-I., & Chiang, S.-S. (2002). Anomaly detection and classification for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 40, 1314-1325.
- Chang, C.-I., & Wang, S. (2006). Constrained Band Selection for Hyperspectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6), 1575-1585. doi:10.1109/TGRS.2006.864389
- Chen, Y., Jiang, H., Li, C., Jia, X., & Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54, 6232-6251.
- Datta, A., Ghosh, S., & Ghosh, A. (2013). Clustering based band selection for hyperspectral images. *IEEE 2012 International Conference on Communications, Devices and Intelligent Systems (CODIS)*. doi: 10.1109/CODIS.2012.6422146
- Du. (2003). 'Band Selection and Its Impact on Target Detection and Classification in Hyperspectral Image Analysis' *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, pp. 374-377.
- Du, Q., Ren, H., & Chang, C.-I. (2003). A Comparative Study for Orthogonal Subspace Projection and Constrained Energy Minimization. *IEEE Transactions on*

- Geoscience and Remote Sensing*, 41(6), 1525–1529. doi:10.1109/TGRS.2003.813704
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). 'A density-based algorithm for discovering clusters in large spatial databases with noise' *2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD-96)*, pp. 226–231. Portland, Oregon: IEEE Computer Society.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874.
- Ghamary Asl, M., Mobasheri, M., & Mojaradi, B. (2014). Unsupervised feature selection using geometrical measures in prototype space for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 52, 3774–3787.
- Keshava, N. (2004). Distance Metrics and Band Selection in Hyperspectral Processing with Applications to Material Identification and Spectral Libraries.”. *IEEE Transactions on Geoscience and Remote Sensing*, 42(7), 1552–1565. doi:10.1109/TGRS.2004.830549
- Kuo, B.-C., & Landgrebe, D. (2004). Nonparametric weighted feature extraction for classification. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 1096–1105.
- Landgrebe, D. (1999). Information extraction principles and methods for multispectral and hyperspectral image data. *Information processing for remote sensing*, 82, 3–38.
- Ma, L., Fu, T., Blaschke, T., Li, M., Tiede, D., Zhou, Z., . . . Chen, D. (2017). Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using random forest and support vector machine classifiers. *ISPRS International Journal of Geo-Information*, 6, 51.
- Manolakis, D., Siracusa, C., & Shaw, G. (2001). Hyperspectral subpixel target detection using the linear mixing model. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 1392–1409.
- Mojaradi, B., Abrishami-Moghaddam, H., Zojj, M., & Duin, R. (2009). Dimensionality reduction of hyperspectral data via spectral feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 47, 2091–2105.
- Raymer, M. L., Punch, W. F., Goodman, E. D., Kuhn, E. D., & Jain, L. A. (2000). Dimensionality reduction using genetic algorithm. *IEEE Transaction on Evolutionary Computation*, 4(2), 164–171.
- Roffo, G., Melzi, S., & Cristani, M. (2015). 'Infinite Feature Selection' In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*.
- Snyder, D., Kerekes, J., Fairweather, I., Crabtree, R., Shive, J., & Hager, S. (n.d.). 'Development of a web-based application to evaluate target finding algorithms' In *Proceedings of IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium*, (pp. II-915-II-918).
- Sun, K., Geng, X., & Ji, L. (2015). A New Sparsity-Based Band Selection Method for Target Detection of Hyperspectral Image. *IEEE Geoscience and Remote Sensing Letters*, 12(2). doi: 10.1109/LGRS.2014.2337957
- Target Detection Blind Test. (n.d.). Retrieved from <http://dirsapps.cis.rit.edu/blindtest/>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58, 267–288.
- Wang, S., & Chang, C.-I. (2007). Variable-Number Variable-Band Selection for Feature Characterization in Hyperspectral Signatures. *IEEE Transactions on Geoscience and Remote Sensing*, 2979–2992.
- Wang, Y., Huang, S., Liu, D., & Wang, H. (September 2014). A novel band selection method based on curve area and genetic theory. *Springer Journal of Optics*, 43(3), 193–202. doi:10.1007/s12596-014-0199-4
- Xu, Y., Du, Q., & Younan, N. H. (2017). Particle Swarm Optimization-Based Band Selection for Hyperspectral Target Detection. *IEEE Geoscience and Remote Sensing Letters*, 14(4). doi:0.1109/LGRS.2017.2658666
- Xue, B., Yu, C., Wang, Y., Song, M., Li, S., Wang, L., . . . Chang, C.-I. (2017). A subpixel target detection approach to hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55, 5093–5114.
- Yu, C., Song, M., & Chang, C.-I. (2018). Band subset selection for hyperspectral image classification. *Remote Sensing*, 10, 113.