ANOMALY DETECTION ALGORITHMS FOR HYPERSPECTRAL IMAGERY

S.R.Soofbaf [1], H.Fahimnejad [2], M. J.Valadan Zoej [3], B. Mojaradi[4]
Geodesy and Geomatics Faculty, K.N.Toosi University of Technology,
Vali_Asr Street, Mirdamad Cross, Tehran, Iran
Tel: +98 9122016721, +98 21 8877 0218 Fax: +98 21 8878 6213

[1] sr.soofbaf@gmail.com , 2 hamed_fahimnejad@yahoo.com ,3 valadanzouj@kntu.ac.ir ,
[4] Mojaradi@yahoo.com

KEY WORDS: Hyperspectral, Target detection, Anomaly detection

ABSTRACT:

Nowadays the use of hyperspectral imagery specifically automatic target detection algorithms for these images is a relatively exciting area of research.
An important challenge of hyperspectral target detection is to detect small targets without any prior knowledge, particularly when the interested targets are insignificant with low probabilities of occurrence. The specific characteristic of anomaly detection is that it does not require atmospheric correction and signature libraries. Recently, several useful applications of anomaly detection approaches have been developed in remote sensing.
With this in mind, in this paper some anomaly detectors such as RX-based anomaly detectors (MRX, NRX,CRX,RX-UTD), as well as adaptive anomaly detectors such Nested Spatial Window-Based approach(NSW) and dual window-based eigen separation transform (DWEST) ,are compared. Finally the most efficient method is proposed for implementation in a planned software system.

INTRODUCTION:

Hyperspectral sensors are passive sensors that simultaneously record images for hundreds of contiguous and narrowly spaced regions of the electromagnetic spectrum, and group these spatially co-registered bands in a hyperspectral data cube. This type of data shows vast potential for use in automatic target detection and recognition for the reason that it provides both spatial features and important information about the spectral characteristics of the materials, targets and backgrounds in the imagery [1].
In other words hyperspectral target detection can locate targets that generally cannot be resolved by multispectral imaging sensors, as a consequence of high spectral resolution in hyperspectral sensors. In this case, the concept of a spectral signature, which uniquely characterizes any given material, is very interesting and generally used. The amount of variability is more important in remote sensing applications caused by the variations in atmospheric conditions, sensor noise, material composition, location, adjacent materials, and other factors. But if we have no prior information about the target we should work with radiance. In this respect the most practical approach is to search for anything that displays different spatial and/or spectral characteristics from its surroundings. This process is known in hyperspectral literature as "anomaly detection".
The main purpose of these unsupervised target detection methods is to locate targets which are commonly unknown, relatively small and only occur in the image scene with low probabilities.

These algorithms are powerful for detecting objects of interest for the reason that the difficult step of atmospheric compensation is not a prerequisite for them and they don't require training and signature libraries [2]. Hence they are usually simple algorithms to implement.
Originally anomaly detection was considered to be only a military purpose with the detection of man-made targets. But recently, it has found in a broad variety of applications ranging from defense, agriculture, geology, environmental monitoring and etc.

RX-BASED ALGORITHMS:

The basic RX algorithm is the benchmark anomaly detection algorithm, originally developed for multispectral imagery by Reed and Yu (1993) and is given by

$$\delta_{RX}(r) = (r-\mu)^T K_{L\times L}^{-1}(r-\mu) \qquad (1)$$

That is in fact the well-known Mahalanobis distance where r is the pixel spectral vector, μ is the mean spectral vector for the area of interest (the mean of each spectral band), L is the number of spectral bands, and K is the spectral covariance matrix [3].
The local mean is achieved by sliding a dual concentric window (a small interior window centered within a larger outer window) over every spectral pixel in the image and calculate the mean of the spectral pixels falling within the outer window. The size of the interior window is assumed to be the size of the typical target of interest in the image. The residual signal after mean subtraction is assumed to approximate a zero-mean pixel-to-pixel independent Gaussian random process.

RELATIONSHIP BETWEEN RX AND PCA ALGORITHMS:

Principal components analysis (PCA) algorithm de-correlates the data matrix in such a manner that different amounts of the image information can be conserved in separate components images, each of which correspond to a different portion of uncorrelated information. Thus, PCA algorithm has been commonly used to compress image information into a few main principal components specified by the eigenvectors of that correspond to large eigenvalues, but it is not designed to be used for detection or classification process. Though, if the image data have interesting target pixels which occur with low probabilities in the data, it is clear that these targets will not be shown in main principal components, but rather in minor components specified by the eigenvectors of that are associated with small eigenvalues. It is interesting to note that mathematically, RX algorithm can be considered to be an inverse procedure of the PCA algorithm which searches for targets in minor components. In this case, a small eigenvalue will create a large value of $\delta_{RX}(r)$. This is comparable to searching for minor components by finding smaller eigenvalue of $\boldsymbol{K}$. It offers explanation of why RX algorithm works for anomaly detection [4].

It is interesting to see that the RX equation by (1) performs some kind of a matched filter specified by

$$M_d(r) = \kappa . d^T r \qquad (2)$$

Where d is the matched signal and κ is a constant, but can be also a function of r . The performance of (2) is entirely determined by two parameters: the matched signal and the scale constant κ that seems before the matched filter. By using (2), we can interpret the RX equation as a matched filter operating on $(r-\mu)$ with the matched signal $d = (r-\mu)^T K_{LxL}^{-1}$ by setting $\kappa = 1$. By taking this advantage, two alternatives of the RX referred to as normalized RX and modified RX denoted by (3) and (4) can be developed from (1) and (2) for anomaly detection by setting $\kappa =[(r-\mu)T(r-\mu)]^{-1}$ and $\kappa =[(r-\mu)T(r-\mu)]^{-1/2}$ in this way

$$\delta_{NRX}(r)=\frac{(r-\mu)^T K_{L\times L}^{-1}(r-\mu)}{(r-\mu)^T(r-\mu)} \quad (3)$$

$$\delta_{MRX}(r)=\frac{(r-\mu)^T K_{L\times L}^{-1}(r-\mu)}{\sqrt{(r-\mu)^T(r-\mu)}} \quad (4)$$

This algorithm is a maximum likelihood anomaly detection process that simplifies the clutter to being spatially white. The RX algorithm uses a binary hypothesis approach to detection, and implements a generalized likelihood ratio test. Assessment of the maximum likelihood detection statistic requires full spectral sample covariance matrices to be estimated and then inverted, or the evaluation of their determinants.
Extending the RX algorithm from multispectral to hyperspectral imagery suffers from some major restrictions. First, the clutter model implemented in this algorithm is limited to being spatially uncorrelated, or spatially white. This model neglects the potentially important spatial correlation information of the clutter. The second restriction is its computational cost arising from the expensive inversion or determinant evaluation of the covariance matrix under each of the hypotheses.

CAUSAL RX:

Since RX detector involves the computation of the mean and covariance matrix, it can not be implemented in real-time. Hence a real-time processing version of the RX is introduced where the sample correlation matrix instead of the sample covariance matrix is used[4]. It is referred to as "Causal" RX (CRX) that borrow s from digital signal processing terminology which means that the information used for data processing is up to the pixel being processed and updated solely based on the pixels that were already processed.
Real-time process can be implemented in two ways: line-by-line and pixel-by-pixel. Two advantages can be benefited from the CRX. Since the computation of the inverse of a sample correlation matrix can be carried out in parallel via a QR-decomposition, the CRX has an ability of processing data in a real-time manner.

$$\delta_{CRX}(r_k)=r_k^T\ R^{-1}(r_k)\ r_k \quad (5)$$

Where:

$$R(r_k)=\frac{1}{k}\sum_{i-1}^{k} r_i r_i^T \quad (6)$$

It should be noted that the sample data correlation matrix $R(r_k)$ formed by the sample vectors $\{r_1,r_2,\ldots,r_k\}$ up to the pixel to be processed,r_k.
In other words, unlike the RX, which requires the knowledge of all the data samples to structure the sample covariance matrix prior to processing, the CRX processes and updates data either line-by-line or sample-by-sample. A second advantage results from the fact that the sample correlation matrix accounts for both the first-order and second-order statistics. For most remotely sensed images which are generally considered to be non-stationary, the CRX can capture spectral variability more efficiently than the RX, which only takes care of second-order statistics.

LPTD & UTD:

Another type of anomaly detector, referred to as the low probability target detector (LPTD), was developed by Harsanyi (1993) and given by

$$\delta_{\text{LPTD}}(\text{r}) = 1_{L\times 1}{}^{T}\, R_{L\times L}^{-1}\ \text{r} \qquad (7)$$

which was designed base on the sample correlation matrix *R*. If we replace *R* with the sample covariance matrix *K*, we can develop an alternative LPTD using sample covariance matrix *K*, referred to as uniform target detector (UTD) which is given by

$$\delta_{UTD}(r) = (1_{L\times 1} - \mu)^{T} K_{L\times L}^{-1}(r - \mu) \qquad (8)$$

Where $1_{L\times 1} = (1, 1, \ldots, 1)^{T}$ is the L dimensional unity vector with ones in all the elements. As it can be seen, UTD algorithm uses the unity vector $1_{L\times 1}$ as its matched signature. Because there isn't any prior knowledge available, thus an anomalous target is assumed to have uniform distribution of radiance over all the spectral bands. Therefore it is predictable to extract background signatures which are uniformly distributed in the image scene. However, if there is some limited information available, the mentioned unity vector can be replaced by a certain particular vector. For example, if we are interested in near infrared wavelengths, we can set zeroes for all visible bands while assigning ones to all the near-infrared bands.

In this case it is remarkable to note that the background subtraction could enhance the RX detection algorithm performance as shown by Ashton & Schaum (1998). By incorporating the UTD into RX detector we can remove the background as well as noise to improve the performance of basic RX detector. This advantage enables us to develop a new type of anomaly detector by subtracting UTD from RX as follows:

$$\delta_{RX-UTD}(r) = (r - 1)^{T} K_{L\times L}^{-1}(r - \mu) \qquad (9)$$

For Hyperspectral data, where the number of spectral bands runs into the hundreds, the RX-based algorithm rapidly becomes unfeasible.

Dual Window-based Eigen Separation Transform (DWEST):

An adaptive anomaly detector, referred to as dual window-based eigen separation transform (DWEST) was developed by Kwon et al [5]. It implements two windows, called inner and outer windows (figure 1) which are used to maximize the separation between two-class data: target class and background class. The purpose of using the inner window is to capture a target present in the window, while the goal of the outer window is to model the background of the target assumed in the inner window so that the target can extracted by projecting the differential mean between the two windows onto the eigenvectors associated with the first few largest eigenvalues of $\boldsymbol{C}_{diff}$.
Assume that $\boldsymbol{r}$ is an image pixel vector at which the inner and outer windows are centered. Let $\boldsymbol{m}_{out}(\boldsymbol{r})$ and $\boldsymbol{m}_{in}(\boldsymbol{r})$ be the means of the outer and inner windows respectively, and $\boldsymbol{C}_{out}$ and $\boldsymbol{C}_{in}$ be

their respective covariance matrices and $\boldsymbol{C}_{diff} = \boldsymbol{C}_{in} - \boldsymbol{C}_{out}$ defined as difference covariance matrix between $\boldsymbol{C}_{out}$ and $\boldsymbol{C}_{in}$.

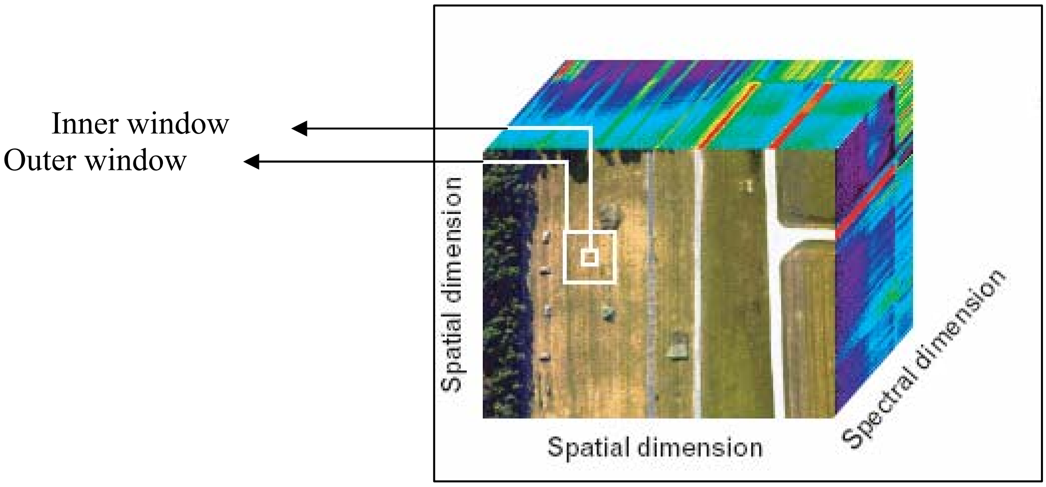


Figure (1): Dual Window-based Eigen Separation Transform process in Hyperspectral cube

Consequently, the eigenvalues of $\boldsymbol{C}_{diff}$ can be divided into two groups, negative values and positive values that the eigenvectors associated with a small number of the large positive eigenvalues of $\boldsymbol{C}_{diff}$ could effectively extract the spectrally distinctive materials that are present in the inner window. If the eigenvectors represented by the positive eigenvalues in this small set are referred to as $\{\mathbf{v}i\}$, the anomaly detector derived by the DWEST, $\delta^{DWEST}(r)$, projects the differential means of two windows, $\boldsymbol{m}_{diff}(\mathbf{r}) = \boldsymbol{m}_{out}(\mathbf{r}) - \boldsymbol{m}_{in}(\mathbf{r})$ onto $\{\mathbf{v}_i\}$ specified by

$$\delta^{DWEST}(r) = \left| \sum_{v_i} v_i^T m_{diff}(r) \right| \quad (10)$$

In order to implement the RX for the dual windows, the RX in equation (1) is modified as

$$\delta^{RX}(r) = m_{diff}(r)^T \left[C_{outer}^{-1}(r) \right] \left(- m_{diff}(r) \right) \quad (11)$$

Nested Spatial Window-based (NSW):

In this section, we describe other adaptive approach, called Nested Spatial Window-based target detection (NSW). The idea is to use a nested three windows where the first two windows, inner and middle windows to extract targets with smallest size and largest size respectively, while the outer window is to describe the background for suppression. There is any more main distinction of the NSW from the DWEST and RX. Instead of using the eigenvector projection in the DWEST and the sample covariance matrix in RX, the following measure, called "Orthogonal Projection Divergence" [3] will be used in the NSW:

$$OPD\,(s_i,s_j) = \left(s_i\, P_{s_j}^{\perp} s_i \;+\; s_j\, P_{s_i}^{\perp} s_j\right)^{1/2} \qquad (12)$$

Where $P_{s_k}^{\perp}$ for $k=i,j$ is defined by:

$$P_{s_k}^{\perp} = I_{L\times L} - s_k (s_k^T s_k)^{-1} s_k^T \qquad (13)$$

Since three nested windows used in the NSW, the inner window implanted in the middle window which is in turn nested in outer window, the OPD must be implemented twice. First between inner and middle windows is specified by

$$\delta_1^{2W-NSW}(r) = OPD\left(m_{in}(r), m_{diff,\,1}(r)\right) \qquad (14)$$

Where $\mathbf{m}_{diff,1}$ is the mean of the outer window with subtraction of the inner window. The second one is between the middle and outer windows is specified by

$$\delta_2^{2W-NSW}(r) = OPD\left(m_{mid}(r), m_{diff,\,2}(r)\right) \qquad (15)$$

Where $\mathbf{m}_{\mathrm{diff},2}$ is the mean of the outer window with subtraction of the middle window. Then, a 3-window NSW, denoted by $\delta^{3W-NSW}(r)$ is defined by

$$\delta^{3W-NSW}(r) = max_{i=1,2}\,\{\delta_i^{2W-NSW}(r)\} \qquad (16)$$

As you can see when the NSW is implemented with the inner and outer windows, it is like the DWEST with equation (10) replaced by

$$\delta^{2W-NSW}(r) = OPD\left(m_{in}(r), m_{diff}(r)\right) \qquad (17)$$

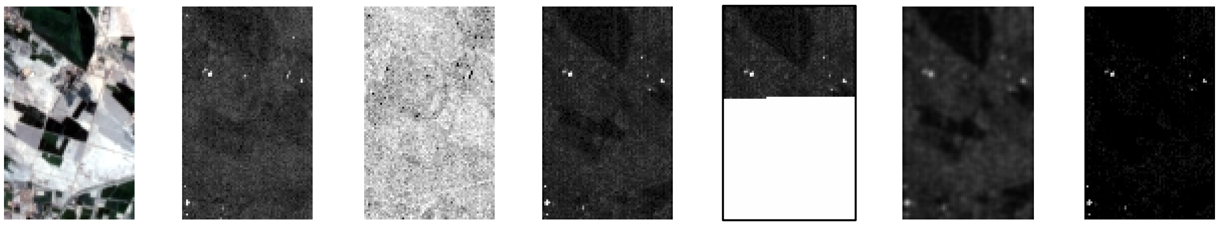
IMPLEMENTATION:

In this paper we use Hyperion data consist of 242 spectral bands covering the range *0.4 to 2.5 μm* to evaluate the performance of RX-based, DWEST and NSW algorithms. At first, a set of 155 stable spectral bands have been choused (because many of the bands suffer from low signal or very high noise). The selected bands are listed in table (1):

Selected Bands number
10 – 57
81-97
101-119
134-164
182-221

Table 1: number of selected spectral bands

Then a region of interest was selected as a sub-scene with 106 lines and 65 samples to run the implemented algorithms (MATLAB Code) much more efficient and faster.
Since many different combinations is used for dual windows, the results are presented in this example were obtained by choosing the best selections of dual windows based on our extensive experiments. In the case of the DWEST, we used the 3/9 dual window where 3 and 9 indicate the sizes of the inner and outer windows.
To implementing the NSW algorithm an appropriate size to generate good results was found to be 3/7/11 where 3, 7 and 11 indicate the sizes of the inner, middle and outer windows.
The result of basic RX, UTD, RX-UTD, DWEST, NSW algorithms which implementing on a Hyperion data of an agricultural region in south of Tehran are shown in figure (2):



(a) (b) (c) (d) (e) (f) (g)

Figure(2): (a) sub-scene of Hyperion data,(b) basic RX,(c) UTD, (d) RX-UTD, (e) CRX, (f) DWEST ,(g) NSW

CONCLUSIONS AND SUGGESTIONS:

As you can see in NSW approach, background data has been removed better than RX-based and DWEST algorithms and targets data has been detected with precise spatial location unlike DWEST model that have blurring artifacts around the targets which where caused by the use of dual windows.
As well one of major advantages of using the NSW is its simple computation complexity compared to the RX-based and DWEST because of using Orthogonal Projection Divergence instead of sample covariance matrix and eigenvector projection.
Another benefit is the NSW can be also extended to any *N* nested windows for target detection. It will be interested to investigate how much advantage can be gained by using more three windows, $N > 3$.
In future researches using some weighted sample covariance matrix in RX-based algorithms and high order statistics (skewness and kurtosis) of hyperspectral data for anomaly detection and the effect of window size(inner, middle and outer) in DWEST and NSW to detecting targets with various sizes efficiently, will be investigated.

REFERENCES:

[1] Dimitris Manolakis, David Marden, and Gary A. Shaw. 2003,"Hyperspectral Image Processing for Automatic Target Detection Applications" MIT University, Lincoln Laboratory Journal Volume 14, Number 1.

[2] Nicola Acito, Giovanni Corsini, Marco Diani, Andrea Cini, 2004," Experimental performance analysis of hyperspectral anomaly detectors", Image and Signal Processing for Remote Sensing, SPIE Vol. 5573, p. 41-51.
[3] Chein-I Chang. 2003" Hyperspectral Imaging: Techniques for Spectral Detection and Classification". Publisher: Kluwer Academic/Plenum Publishers, New York, N.Y., ISBN: 0-306-47483-2, chapter 6.

[4] Chein-I Chang, 2002," Anomaly Detection and Classification for Hyperspectral Imagery". IEEE Transactions On Geoscience And Remote Sensing, Vol. 40, No. 6.

[5] H. Kwon, S.Z. Der and N.M. Nasrabadi, "Adaptive anomaly detection using subspace separation for hyperspectral imagery," Optical Engineering, vol. 42,no. 11, pp. 3342-3351, Nov. 2003.

[6] Weimin Liu, Chein-I Chang, IEEE 2004,"A Nested Spatial Window-Based Approach to Target Detection for Hyperspectral Imagery"