EFFICIENT DETECTION OF ANOMALIES IN HYPERSPECTRAL IMAGES

S. R. Soofbaf [a], M. J. ValadanZoej [b, *], H. Fahimnejad [c], H. Ashoori [b]

[a] ISA (Iran Space Agency), Sayeh Alley, Africa St., Tehran, Iran – sr.soofbaf@gmail.com
[b] K.N.Toosi University of Tech. Geomatics Eng., Mirdamad Cross, Valiasr St., Tehran, Iran - (valadanzouj@kntu.ac.ir, Hamed_Ashoori@yahoo.com)
[c] Sepehr Geomatics Engineering Co., East Hoveizeh St., North Sohrevardi Ave, Tehran, Iran - Hamed_Fahimnejad@yahoo.com

Commission VI, WG VI/4

KEY WORDS: Hyperspectral, Target detection, Anomaly detection, RX algorithm, ROC

ABSTRACT:

By reason of recent advances in airborne and ground-base hyperspectral imaging technology, many applications have been developed.
One of the most important hyperspectral images applications involves automatic detection of hidden objects without any prior knowledge about them such as man-made targets, rare minerals in geology, vegetation stresses in agriculture, poisonous wastes in environments, cancerous cells or tumors in medical imaging, etc.
The most robust class of algorithms for detection of this type of targets is arguably the one that searches the pixels of image cube for rare pixels whose information significantly differs from their surrounding pixels and local background. These targets are known as "Anomaly" in image processing and remote sensing literature.
Considering mentioned concepts, in this research, a host of different anomaly detectors such as RX-base anomaly detectors (Basic RX, Modified RX, Normalized RX, Weighted RX, Causal RX, UTD, RX-UTD, and ACAD), Dual window-base Eigen Separation Transform (DWEST) method, Nested Spatial Window-base Target detector (NSWTD) and Combined F-Test (CFT) algorithm are investigated and compared.

1. INTRODUCTION

Hyperspectral imaging sensors acquire images in many contiguous and very narrow spectral bands in visible, near-infrared, and mid-infrared portions of electromagnetic spectrum. This type of digital data shows vast potential for use in automatic target detection since it provides useful information about the spectral characteristics of the materials and targets in the image.
This means that many targets that generally cannot be resolved by multispectral images can be located as a consequence of high spectral resolution in hyperspectral images according to the concept of a spectral signature which uniquely characterizes any given material. In this process, many factors should be considered such as variations in atmospheric conditions, location, noise of sensor, material composition, adjacent materials, etc. But in many cases there is not any prior information about the target. In this respect the most practical approach is to search for anything that displays significantly different spectral characteristics from its surroundings. This process is known as "anomaly detection" in remote sensing literature.
The main purpose of these target detection methods is to locate targets which are commonly unknown, relatively small and only occur in the image scene with low probabilities. Also these algorithms can be applied directly to the radiance at the sensor level. Thus they do not require any training or difficult step of atmospheric correction and they are usually simple to implement, even in a real-time or near real-time manner. Nowadays anomaly detection has found in a broad variety of applications ranging from defence, agriculture, geology, environmental monitoring, medical imaging and etc.
In this research, a group of different anomaly detectors are investigated and compared to select the best algorithm according to purpose of aimed application, data specifications and its quality (number of spectral bands, signal-to-noise ratio, etc) and other effective parameters.

2. ANOMALY DETECTION ALGORITHMS

In this section most important anomaly detection algorithms are briefly described theoretically which will be used for comparative analysis.

2.1 Basic RX algorithm

The basic RX algorithm is the benchmark anomaly detection algorithm, originally developed for multispectral images by Reed and Yu (1993) that is formulated based on two hypotheses. The first one models the image background as a Gaussian distribution with zero mean and an unknown covariance matrix which is estimated globally or locally from the data ($N(0,\Sigma)$). The second hypothesis models the target as a linear combination of a target signature and background noise. So, under a spectral vector is represented by a Gaussian distribution with a mean equal to the signature of the target (s) and an additive noise equal to the background covariance matrix in hypothesis ($N(s,\Sigma)$).

$$\begin{aligned} H_0 &: r = n \\ H_1 &: r = \alpha s + n \end{aligned} \qquad (1)$$

In this case detection process is based on exploiting the difference between the spectral signatures of an input pixel and its surrounding pixels that is very similar to the well-known Mahalanobis distance and given by (Chang, 2003):

$$\delta_{RX}(r) = (r-\mu)^T K_{LxL}^{-1}(r-\mu) \quad (2)$$

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.

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 K. It offers explanation of why RX algorithm works for anomaly detection. Other types of RX-base algorithms derived and developed from this basic algorithm to improve detection performance which will be discussed.

2.2 Normalized RX (NRX) & Modified RX (MRX)

It is interesting to see that the RX equation by (2) performs some kind of a matched filter that its performance is entirely depend on two parameters: the matched signal ($(r-\mu)^T K_{L\times L}^{-1}$) and the scale constant ($\kappa=1$) which seems before the matched filter. According to this advantage, two alternatives of the RX referred as normalized RX and modified RX, those are denoted by (3) and (4) can be developed by setting $\kappa = [(r-\mu)^T(r-\mu)]^{-1}$ and $\kappa = [(r-\mu)^T(r-\mu)]^{-1/2}$ in this way (Chang, 2002):

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

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

2.3 Weighted RX algorithm (WRX)

In view of the fact that anomaly targets are generally small and the background is homogeneous, the sample covariance matrix of the entire image can be viewed as background sample covariance matrix. Also the anomalies or small man-made targets can be separated as outlier in image data cube. Consequently, the RX algorithm uses Mahanalobis distance to find them by using sample covariance matrix to whiten the background pixel, then those anomaly pixels become outliers. It would not be a problem when the number of anomaly pixels is few. But since this algorithm assumes Gaussian noise and uses sample covariance matrix for data whitening, when percentage of the anomaly pixels is relatively large, the sample covariance matrix cannot represent the background distribution. In this case the RX algorithm will not perform well. So to solve this problem, it is proposed to use weighted covariance matrix (Hsuan, 2005).

In weighted RX algorithm, proper weight is assigned to each pixel in the sample covariance matrix using its distance to the data center for better representation of background distribution. So weighted matrix can be written as:

$$K_W = \frac{\sum_1^N w_i (r_i-\mu_W)(r_i-\mu_W)^T}{\sum_1^N w_i} \quad (5)$$

$$\mu_W = \frac{\sum_1^N q_i r_i}{\sum_1^N q_i} \quad (6)$$

Where w_i and q_i are weight scalars for each pixel in the image defined by:

$$q_i = \frac{1}{1+\|r_i-\mu\|}, \qquad w_i = \frac{1}{1+\|r_i-\mu_W\|} \quad (7)$$

Finally RX filter with weighted covariance matrix can be applied to the image.

$$\delta_{WRX}(r) = (r-\mu_W)^T K_W^{-1}(r-\mu_W) \quad (8)$$

2.4 Uniform Target Detector algorithm (UTD)

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

$$\delta_{LPTD}(r) = 1_{L\times L}{}^T R_{LxL}^{-1} \; r \quad (9)$$

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

$$\delta_{UTD}(r) = (1_{L\times 1} - \mu)^T K_{LxL}^{-1}(r-\mu) \quad (10)$$

Where $1_{L\times 1}$ is the L dimensional unity vector with ones in all the elements. 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.

In this case it is remarkable to note that the background subtraction could enhance the RX detection algorithm performance as shown by Ashton and Schaum (1998). By incorporating the UTD into basic RX, the background can be removed 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 (Chang, 2003):

$$\delta_{RX-UTD}(r) = (r-\mu)^T K_{LxL}^{-1}(r-\mu) \quad (11)$$

2.5 Causal RX algorithm (CRX)

Since RX detector involves mean and covariance matrix computation, it can not be implemented in real-time. Hence a real-time processing version of the RX is introduced where the sample correlation matrix (R) is used instead of the sample covariance matrix (K). It is called "Causal" which means that the information used for data processing is up to the pixel being processed and updated only based on the pixels that were already processed.

$$R(r_k) = \frac{1}{k}\Sigma_{i-1}^{k} r_i r_i^T \tag{12}$$

Since the computation of the inverse of a sample correlation matrix can be carried out in parallel via QR matrix decomposition method, this algorithm can be implemented in a real-time manner (Chang, 2003).

$$\delta_{CRX}\ (r_k) = (r_k)^T R_{(r_k)}^{-1}(r_k) \tag{13}$$

2.6 Adaptive Causal Anomaly Detector algorithm (ACAD)

ACAD algorithm is a developed version of causal RX model. In this algorithm, strong signatures of detected anomalies are removed during detection process due to their undesirable effects on detection of subsequent anomalies. Because one major problem encountered in CRX algorithms is that if an earlier detected anomaly has an intense signature it may have considerable impact on the detection of later anomalies. This occurrence is mainly caused by an inappropriate use of sample correlation matrix. According to Chang (2003), a proper sample correlation matrix should be one that removes all the earlier detected anomaly pixels being included in the sample correlation matrix. For this reason, the *(R)* in Causal RX equation should be replaced with a sample correlation matrix that removes all detected anomalies defined by (Hsueh, 2004):

$$\begin{aligned} \tilde{R}(r_k) &= R(r_k) - \Sigma_{t_j \in \Delta(k)}\ t_j t_j^T \\ \delta_{r_k}^{ACAD} &= r_k^T\ \tilde{R}_{r_k}^{-1}\ r_k \end{aligned} \tag{14}$$

Where *Δ(k)* is the set of earlier detected anomalous target pixels t_j prior to the currently being processed image pixel (r_k).

Also the mentioned Rx-base algorithms are known as Global or Local anomaly detector if the mean spectrum is derived from the full image data or from a local window around each pixel during detection process.

2.7 Dual Window-base Eigen Separation Transform anomaly detector (DWEST)

DWEST model implements two local windows, entitled inner and outer windows which are used to maximize the separation between anomalies and background. The idea of using the inner window is to detect an anomaly present in it, whereas the purpose of the outer window is to model the background of the anomaly assumed in the inner window[5]. By moving these two local windows entire the image, local mean (m_{in}, m_{out}) and covariance matrix (C_{in}, C_{out}) of each window and their differences are calculated as below:

$$\begin{aligned} m_{diff} &= m_{out} - m_{in} \\ C_{diff} &= C_{in} - C_{out} \end{aligned} \tag{15}$$

Consequently anomalies can be extracted by projecting the differential mean between two windows on to the eigenvector associated with the largest positive eigenvalue of differential covariance matrix (Kwon, 2003) by:

$$\delta^{DWEST} = \left| \Sigma_{v_i} V_i^T m_{diff}(r) \right| \tag{16}$$

Also in order to implement the RX for the dual windows, the RX in equation is modified as:

$$\delta^{RX-DW}(r) = \left|\ m_{diff}(r)^T \left[C_{out}^{-1}(r) \right] m_{diff}(r)\ \right| \tag{17}$$

2.8 Nested Spatial Window-base Target Detector (NSWTD)

NSWTD model implements a nested three local windows, entitled inner, middle and outer windows where the first two windows are used to extract smallest and largest anomalies respectively, while the outer window is used to model the local background. Moreover the other main distinction of this model from the DWEST and RX-base algorithms is using the Orthogonal Projection Divergence (OPD) as measurement criterion instead of eigenvector projection or sample covariance matrix (Liu, 2004) by:

$$\begin{aligned} OPD\ (s_i, s_j) &= \sqrt{(s_i\ P_{s_j}{}^{\perp} s_i\ +\ s_j\ P_{s_i}{}^{\perp} s_j\ )} \\ P_{s_k}^{\perp} &= I_{L\times L} - s_k (s_k^T s_k)^{-1} s_k^T \end{aligned} \tag{18}$$

Since three nested windows used in this algorithm, 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

$$\delta_1^{2W-NSW}(r) = \ OPD\left(\ m_{in}(r), m_{diff,\ 1}(r)\right) \tag{19}$$

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

$$\delta_2^{2W-NSW}(r) = \ OPD\left(\ m_{mid}(r), m_{diff,\ 2}(r)\right) \tag{20}$$

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

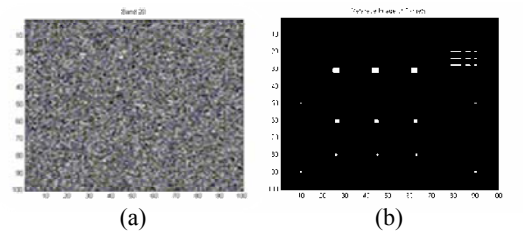
$$\delta^{3W-NSW}(r) = \ \max_{i=1,2}\ \{\delta_i^{2W-NSW}(r)\} \tag{21}$$

2.9 Combined F-Test anomaly detector (CFT)

CFT method is based on a nonparametric model that compares two sets of data (random variable), background data and target data by using a typical inner/outer window mechanism to sample the image for local detection, or using full image data to global detection. In this method after combining two mentioned samples, statistical parameters of them (mean and variance) are estimated. Then ratio of these parameters are calculated (Z_{CFT}) and tested by Fisher distribution of (1,1) degree of freedom.
A decision threshold T is determined via defining a type I error for F distribution function and compared with the result of ratio. If Z_{CFT} is greater than T, it means that these two sets are most likely sampled from different distributions. Therefore they are anomalous to each other. If not, they are likely sampled from the same distribution.
The main assumption in this method is asymptotic behaviour of Fisher's F distribution for data sets which are examined by a common statistical test (Rosario, 2005).

3. EXPERIMENTAL RESULTS

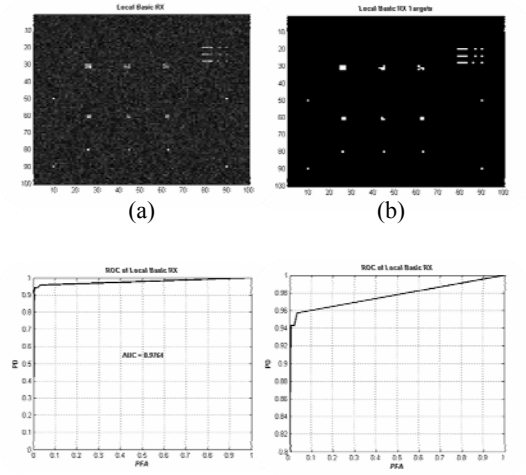
Two sets of hyperspctral data have been used for these experiments.
The first one is a real hyperspectral data (AVIRIS image of Cuprite region which is available in ENVI sample data) for visual inspection of detection results. The second one is a synthesized hyperspectral data for quantitative evaluation which was simulated by sampling of contiguous spectral curve with 20 nanometre spectral resolution for some natural and man-made targets such as dry grass, sandy loam, sagebrush, galvanized iron and aluminium metal using ENVI spectral libraries and MATLAB software. This synthetic hypercube's size was 100×100 pixels which made up linear combination of some above signatures for background region and anomaly panels.
In background region first pixel is started with 100 percent dry grass and 0 percent sandy loam. Then by moving to next pixel percentage of dry grass is decreased and percentage of sandy loam is increased. This process is repeated until the last pixel has 0 percent of dry grass and 100 percent sandy loam. Then anomaly pixels are added to this data in various locations. Anomaly panels in each column have the same pure signature with various sizes and they have the same signature in each row. All of anomaly pixels have 10 percent abundance of anomaly signatures such as galvanized iron and 90 percent abundance of background signatures. Also Gaussian noise is added to each pixel to achieve 30:1 signal-to-noise ratio.
Figure (1) shows the simulated hypercube in band number 20 (800 nanometre) and location of considered anomaly pixels in it.



(a) (b)

Figure 1. Simulated image in band number 20 (a), location of anomalies (b)

All mentioned algorithms (in global or local manner with various window sizes) are implemented in MATLAB software and their performances are compared by discrete Receiver Operating Characteristic (ROC) curves and their area under curves (AUC) as appropriate criterions for evaluation of detection algorithms. Therefore seven threshold values are used to achieve confidence levels about 93% to 99% for calculating probability of correct detection and false alarm.
For example figure (2) shows the visual result of local RX method with local window size 15, correct detected targets and ROC curve of this algorithm

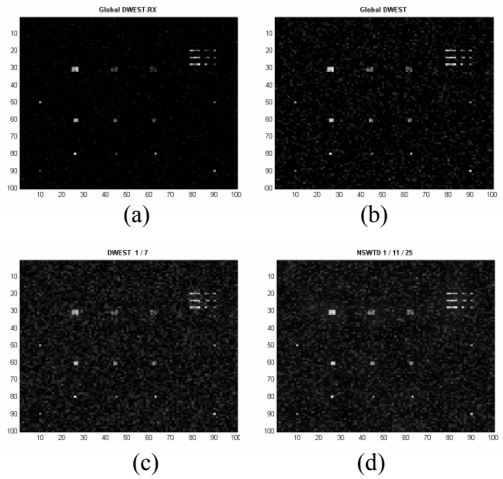


(a) (b)

(c) (d)

Figure 2. Result of local RX (a), correct detected anomalies in 99% confidence level (b), full ROC curve (c), comparable part of ROC curve (d)

Local DWEST and NSWTD algorithms are tested with different combination of local window sizes. Figure (3) shows the best visual results of these methods.



(a) (b)

(c) (d)

Figure 3. Global DWEST-RX (a), global DWEST (b), local DWEST with in/out window size 1/7 (c), NSWTD with in/mid/out window size 1/11/25 (d)

Also figure (4) shows the visual results of global RX-base methods.

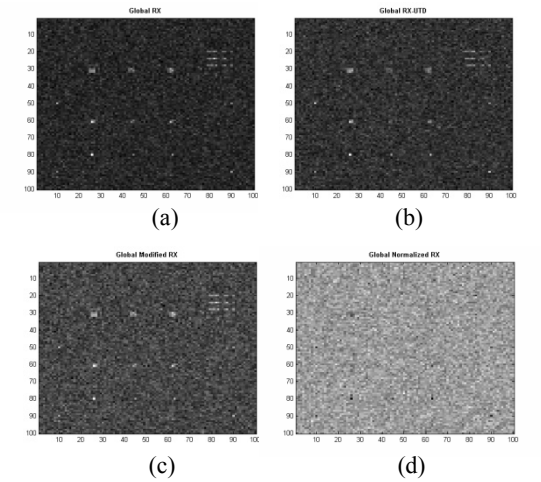


Figure 4. Global RX (a), Global RX-UTD (b), Global modified RX (c), Global normalized RX (d)

Figure (5) shows the visual results of global weighted RX-base methods.

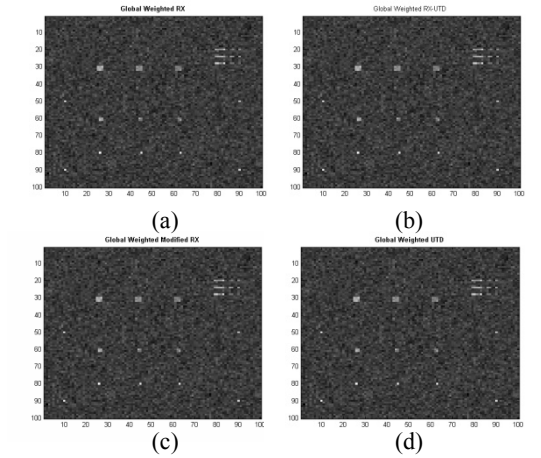


Figure 5. Weighted RX (a), weighted RX-UTD (b), Weighted modified RX (c), weighted normalized RX (d)

Figure (6) shows the visual results of local RX-base algorithms.

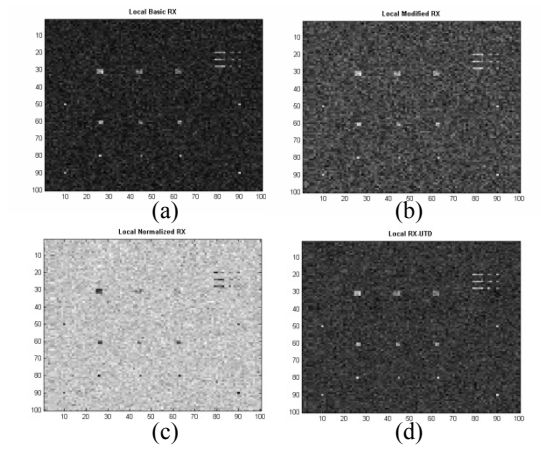


Figure 6. Local RX (a), local modified RX (b), local normalized RX (c), local RX-UTD (d)

Figure (7) shows the visual results of causal RX, ACAD and CFT algorithms.

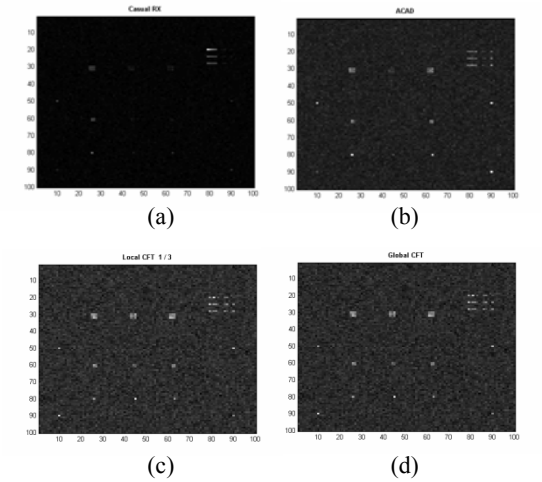


Figure 7. Causal RX (a), ACAD (b), local CFT (c), global CFT (d)

For comparing all implemented algorithms together, their area under ROC curves, number of correct detections (from 70 target pixels) and false alarm detections in 99% confidence level displayed in table (1).

	Method	*AUC*	*Correct Detected*	*False Alarm*
1	Global NRX	0.6429	11	201
2	Global UTD	0.8237	29	189
3	Local NRX	0.8981	47	143
4	Causal RX	0.9248	39	141
5	Global RX-UTD	0.9413	46	133
6	ACAD	0.9464	54	33
7	Local DWEST	0.9569	59	106
8	Local MRX	0.9625	55	99
9	Global MRX	0.9641	49	132
10	Local CFT	0.9666	58	52
11	Global RX	0.9686	53	108
12	Local UTD	0.9698	59	113
13	Global DWEST-RX	0.9706	60	***14***
14	Global DWEST	0.9719	63	110
15	NSWTD	0.9731	***64***	89
16	Local RX	0.9764	59	67
17	Local RX-UTD	0.9801	***64***	86
18	Global CFT	0.9823	60	41
19	Global Weighted MRX	0.9861	60	89
20	Global Weighted UTD	0.9934	60	81
21	Global Weighted RX	0.9944	61	71
22	Global Weighted RX_UTD	***0.9946***	61	68

Table 1. Comparative result of anomaly detection algorithms

Moreover, all of these algorithms are compared with computational complexity point of view and they are tested by simulated hyperspectral data with various additive Gaussian noises (20:1, 10:1 and 5:1 signal-to-noise ratio) to investigate noise sensitivity of them. For example figure (8) shows the

visual results of weighted RX using simulated image with various SNR. This algorithm is selected, because it has more sensitivity of noise increase.

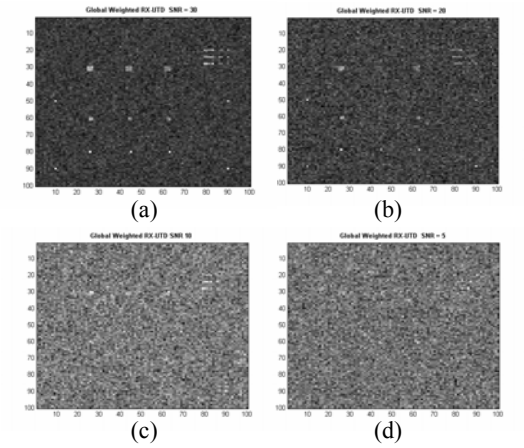


Figure 8. SNR 30:1(a), SNR 20:1(b), SNR 10:1(c), SNR 5:1(d)

Table (2) and figure (9) show the effect of Signal-to-noise ratio on detection performance (AUC).

Method	*SNR = 5*	*SNR = 10*	*SNR = 20*	*SNR = 30*
Global RX	0.8528	0.9095	0.9285	0.9686
Local RX_UTD	0.8568	0.8806	0.9189	0.9801
Global Weighted RX-UTD	0.6575	0.7210	0.8326	0.9946
DWEST	0.6711	0.8689	0.9642	0.9719
NSWTD	0.7863	0.8364	0.9061	0.9731
CFT	0.7010	0.8724	0.9210	0.9823

Table 2. AUC in various SNR

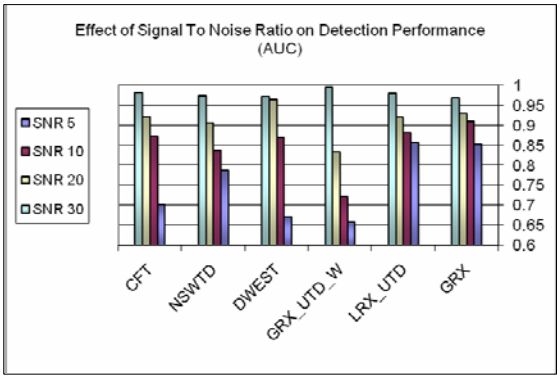


Figure 9. Effect of SNR on AUC

Also table (3) and figure (10) show the effect of signal-to-noise ratio on number of correct detections.

Method	*SNR = 5*	*SNR = 10*	*SNR = 20*	*SNR = 30*
Global RX	38	44	49	53
Local RX_UTD	37	38	41	64
Global Weighted RX-UTD	3	11	29	61
DWEST	14	34	61	63
NSWTD	24	33	45	64
CFT	17	40	49	60

Table 3. Number of correct detections in various SNR

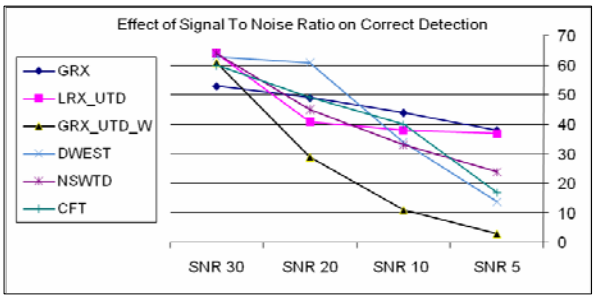


Figure 10. Effect of SNR on correct detection

4. CONCLUSION

Finally our experimental results show that weighted RX-UTD is the highest performance algorithm (maximum AUC), NSWTD method has maximum number of correct detections, DWEST-RX method has minimum number of false alarms, global RX algorithm has less sensitivity of various SNR and NSWTD is the highest speed method and it is suitable for processing of ultraspectral images.

REFERENCES

Chang, C., 2003. *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Kluwer Academic/Plenum, New York, pp. 89-02.

Chang, C., 2002. Anomaly detection and classification for hyperspectral imagery. *IEEE*, 40(6), pp. 1314-1325.

Hsuan, R., 2005. Weighted anomaly detection for hyperspectral remotely sensed images. *SPIE,* 5995(1).

Hsueh, M., 2004. Adaptive causal anomaly detection for hyperspectral imagery. *IEEE*. 5(20), pp. 3222-3224.

Kwon, H., 2003. Adaptive anomaly detection using subspace separation for hyperspectral imagery. *Optical Engineering*, 42(11), pp. 3342-3351.

Liu, W., 2004. A nested spatial window-based approach to target detection for hyperspectral Imagery. *IEEE*, 1(20), pp. 264-268.

Rosario, D., 2005. A nonparametric F-distribution anomaly detector for hyperspectral imagery. *IEEE*, 5(12), pp. 2022-2029.