

COMPARISON OF RX-BASED ANOMALY DETECTORS ON SYNTHETIC AND REAL HYPERSPECTRAL DATA

Sefa Küçük, Seniha Esen Yüksel

Hacettepe University, Department of Electrical and Electronics Engineering, Ankara, Turkey

ABSTRACT

Anomaly detection refers to detecting the deviations from the normal background behavior without any prior information about the target or the background. For hyperspectral image analysis, Reed-Xiaoli (RX) algorithm is arguably the most popular anomaly detector. It models the background as a multidimensional Gaussian distribution and computes how much a test vector is deviating from the background model. Over the years, many versions of RX have been developed and compared on VNIR or SWIR data, but longwave-infrared (LWIR) data comparisons are very few. In this paper, a comprehensive comparison of six different anomaly detectors, namely the global RX, local RX, dual window RX, subspace RX, kernel RX and the global RX combined with a uniform target detector, have been presented. The comparisons have been made on real LWIR hyperspectral data and synthetic data with varying noise levels and target sizes. Several factors to consider such as parameter selection, resilience to noise, effect of window size, computational complexity have been discussed and the detection performance have been presented on receiver operating characteristic curves.

Index Terms— Anomaly detection, hyperspectral, LWIR.

1. INTRODUCTION

Due to the vast sizes of data but little targets of interest, anomaly detection holds an important part in hyperspectral image processing. The goal of anomaly detection is to detect the pixels in the scene whose spectra differ from their surroundings. Most anomaly detection algorithms in literature [1] are based on the family of Gaussian distributions, and the most commonly used anomaly detector, Reed-Xiaoli (RX) [2], can be considered as a benchmark.

The anomaly detection problem can be broken down into two sub-problems: how to characterize the background and how to measure the anomaly score. RX detector assumes that the background is homogeneous and models it with a multivariate Gaussian distribution. However, this hypothesis may not always be adequate. Therefore, several extensions to RX were developed that address the two sub-problems [1, 3]. In the rest of the paper, we compare six variations of RX in

terms of speed and accuracy on synthetic and real hyperspectral (LWIR) datasets and discuss parameter selection.

2. ANOMALY DETECTION METHODS

In this section, we provide the basics of anomaly detectors from the implementation point of view under a unified syntax.

2.1. Global Reed-Xiaoli Detector (GRX)

GRX [4], also called the Mahalanobis distance detector, models the background of the complete scene with a multivariate Gaussian distribution with mean $\hat{\mu}_b$ and covariance $\hat{\Sigma}_b$. It computes the Mahalanobis distance between a test pixel \mathbf{r} and the mean of the background vector as follows:

$$D_{GRX}(\mathbf{r}) = (\mathbf{r} - \hat{\mu}_b)^T \hat{\Sigma}_b^{-1} (\mathbf{r} - \hat{\mu}_b). \quad (1)$$

The test pixel is called an anomaly if this distance D is bigger than a defined threshold.

2.2. Local Reed-Xiaoli Detector (LRX)

LRX [2] is basically a spatially adaptive version of the GRX statistic that replaces the global background model with a local normal model [4]. To detect anomalies, a double concentric window is slid over every pixel in the image as illustrated in Fig.1. The size of inner window that is called guard band is assumed to be the size of target in the image. For the local RX detector, Eq.(1) is modified as follows:

$$D_{LRX}(\mathbf{r}) = (\mathbf{r} - \hat{\mu}_{local})^T \hat{\Sigma}_{local}^{-1} (\mathbf{r} - \hat{\mu}_{local}). \quad (2)$$

where \mathbf{r} is the test pixel, $\hat{\mu}_{local}$ is the mean vector and $\hat{\Sigma}_{local}$ is the covariance matrix of the local background, respectively.

2.3. Reed-Xiaoli and Uniform Target Detector (UTD)

In [5], it has been suggested that removal of the background and also noise has improved the performance of RX detector. A new type of anomaly detector was developed by incorporating the UTD into RX [3]. By subtracting the UTD from the RX, the equation is obtained as follows:

$$D_{RXD-UTD}(\mathbf{r}) = (\mathbf{r} - \mathbf{1}_{d \times 1})^T \hat{\Sigma}_b^{-1} (\mathbf{r} - \mu). \quad (3)$$

where $\mathbf{1}_{d \times 1} = (1, 1, \dots, 1)^T$ is the d dimensional unity vector where d is the number of spectral bands in the data.

We would like to HAVELSAN Inc., Turkey for providing the LWIR data.

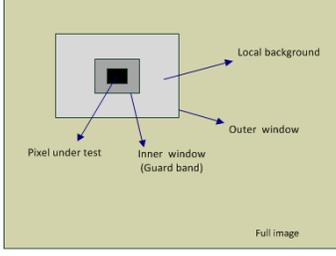


Fig. 1. Double concentric window surrounding the test pixel.

2.4. Dual Window Reed-Xiaoli Detector (DWRX)

In LRX, guard band is used to prevent the target pixels from penetrating into the calculation of background statistics. Unlike the LRX, DWRX [6] uses this guard band information. Instead of using the spectra of the test pixel under test, it uses the mean of the spectral pixels falling into the guard band. DWRX distance is computed as follows:

$$D_{DWRX}(\mathbf{r}) = (\hat{\mu}_{diff}(\mathbf{r}))^T \hat{\Sigma}_{local}^{-1} (\hat{\mu}_{diff}(\mathbf{r})). \quad (4)$$

$$\hat{\mu}_{diff}(\mathbf{r}) = \hat{\mu}_{inner}(\mathbf{r}) - \hat{\mu}_{outer}(\mathbf{r}). \quad (5)$$

where $\hat{\mu}_{inner}$ is the mean of the spectral pixels in the inner window, $\hat{\mu}_{outer}$ is the mean vector and $\hat{\Sigma}_{local}$ is the covariance matrix of the local background region, respectively.

2.5. Subspace Reed-Xiaoli Detector (SSRX)

SSRX [7] assumes that target pixels and background statistics can be expressed in different subspaces. The background subspace is projected onto a subspace that provides suppression of the background statistics. SSRX is computed as follows:

$$D_{SSRX}(\mathbf{r}) = (\mathbf{r} - \hat{\mu}_b)^T (I - WW^\dagger) (\mathbf{r} - \hat{\mu}_b). \quad (6)$$

where I is the identity matrix, $\hat{\mu}_b$ is the mean of the background samples, W is a matrix consisting of eigenvectors corresponding to large eigenvalues of background covariance matrix and W^\dagger denotes the pseudo-inverse of W . The product WW^\dagger is the projection operator.

2.6. Kernel Reed-Xiaoli Detector (KRX)

RX assumes that the background is homogeneously distributed. However, there may be a non-linear relationship between the spectral bands. KRX [8] has been developed for the case that the samples need more complex decision boundaries. KRX uses non-linear models and maps the input space to a high dimensional feature space using kernel functions. A possible kernel function is the Gaussian RBF, given as

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / (c))$$

With the kernel function defined, the KRX value for the test pixel \mathbf{r} can be computed as follows:

$$D_{KRX}(\mathbf{x}) = (\mathbf{K}_r^T - \mathbf{K}_{\hat{\mu}_b}^T)^T \hat{\mathbf{K}}_b^{-1} (\mathbf{K}_r^T - \mathbf{K}_{\hat{\mu}_b}^T). \quad (7)$$

where $\hat{\mathbf{K}}_b$ is the Gram (kernel) matrix of the M background pixels, obtained as follows:

$$\hat{\mathbf{K}}_b = \mathbf{K}(\mathbf{X}_b, \mathbf{X}_b) = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_M) \\ k(\mathbf{x}_2, \mathbf{x}_1) & \cdots & k(\mathbf{x}_2, \mathbf{x}_M) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_M, \mathbf{x}_1) & \cdots & k(\mathbf{x}_M, \mathbf{x}_M) \end{bmatrix}$$

The Gram matrix is also denoted by $\hat{\mathbf{K}}_b = \mathbf{K}(\mathbf{X}_b, \mathbf{X}_b)$ where $\mathbf{X}_b = [\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(M)]$ is the background matrix whose columns are the spectral pixels of d dimensions. In Eq.7, \mathbf{K}_r^T is a vector and equals the kernel map of the test pixel \mathbf{r} as follows:

$$\mathbf{K}_r^T \equiv \mathbf{k}(\mathbf{X}_b, \mathbf{r})^T - \frac{1}{M} \sum_{i=1}^M k(\mathbf{x}(i), \mathbf{r}). \quad (8)$$

where the first term in Eq.8 is given as:

$$\mathbf{k}(\mathbf{X}_b, \mathbf{r}) = [k(\mathbf{x}(1), \mathbf{r}) \quad \cdots \quad k(\mathbf{x}(M), \mathbf{r})]$$

and second term in Eq.8 is the scalar mean of $\mathbf{k}(\mathbf{X}_b, \mathbf{r})$. Finally, $\mathbf{K}_{\hat{\mu}_b}^T$ in Eq.7 equals:

$$\mathbf{K}_{\hat{\mu}_b}^T \equiv \frac{1}{M} \sum_{i=1}^M \mathbf{k}(\mathbf{X}_b, \mathbf{x}(i)) - \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M k(\mathbf{x}(i), \mathbf{x}(j)). \quad (9)$$

The first term in Eq.9 is a row vector, containing the mean of each column of the Gram matrix. The second term is the scalar mean of the Gram matrix. This section provides the sufficient details to implement KRX. Detailed derivations to arrive at these equations can be found in [8].

3. EXPERIMENTAL RESULTS

SYNTHETIC DATA: Two synthetic datasets were generated with different noise levels and various target sizes. The first dataset is a datacube of size 75x75x10 and has two anomaly regions of size 5x5x10. The anomaly regions were generated from a normal distribution with mean 1.5 and variance 0.1; and the background was sampled from a normal distribution with mean 1 and variance 0.1. The detection results for this dataset is given in Fig.2. The second dataset had background mean 1, variance 0.2; and anomaly mean 1.2 with variance 0.2. Therefore, anomalies and the background were barely distinguishable as shown in Fig 2(a). In this second dataset, targets were of size 5x5, 7x7 and 9x9 from left to right to investigate the relations between the window and target size.

On the first dataset, window-based anomaly detectors were run with a 5x5 inner window and 9x9 outer window. For KRX, using the Gaussian RBF kernel, c was set to 40. Fig.2 shows the results of the detection without thresholding. For SSRX, the number of eigenvectors were selected as the principal components that account for about fifty percent of

the variances. For KRX, the size of outer window determines the dimension of Gram matrix and is a very critical parameter in terms of computational complexity. Table 1 shows the elapse time of detectors on both synthetic data and real data. The second dataset presents a much harder problem as the targets and the background have very similar distributions. Therefore, the detections decreased significantly and it became more important to select the best window parameters.

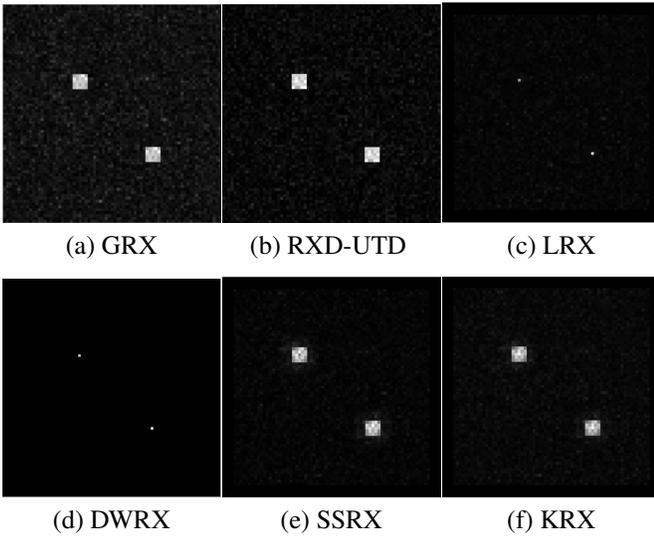


Fig. 2. Detection results for the first synthetic data. GRX generated the most false alarms, followed by RXD-UTD, KRX and SSRX. The LRX and DWRX detected only the center of anomalies due to the window size (5x5), but generated very few false alarms.

For LRX and DWRX, targets could not be detected if the inner window was smaller than 5x5. Thus, the size of inner window should not be smaller than the size of the target. For the outer window, results were enhanced with increasing outer window size as long as the window did not include any other targets. Therefore, an estimate about the distance between the targets can be used to determine the size of the outer window.

The GRX and RXD-UTD have no parameters. However, looking at Fig.3, it is clear that none of them are resilient to noise. This is partly due to the target signatures effecting the background model. However, local RX and SSRX also seems to fail under noise, and the selection of window parameters becomes a rather important issue. The only algorithm that stayed resilient to noise was DWRX.

REAL DATA: Real hyperspectral data was acquired by a LWIR (8–12 μm) imager and was calibrated with Telops Reveal Air Calibrate software. The dataset contains 150x320 pixels and 81 spectral bands. In the scene, four targets composed of different materials including a wooden crate (T1), metallic bucket (T2), plastic jerrycan (T3), and glass jug (T4) were laid on a surface covered with sparse grass. The 3rd

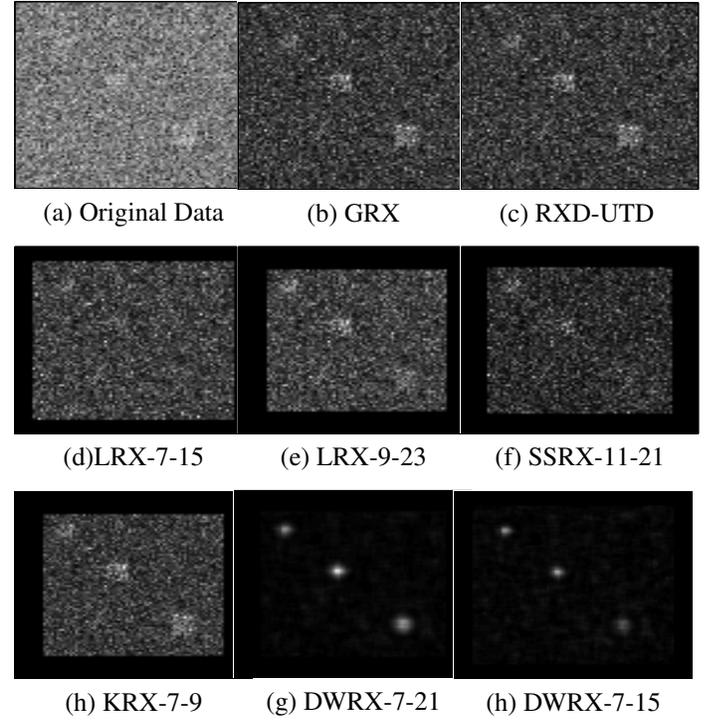


Fig. 3. Detection results for the second synthetic data for varying window sizes. LRX-7-15 means that a 7x7 inner window and a 15x15 outer window was used for local RX.

band image (898.48 cm^{-1}) and its ground truth are shown in Fig.4(a) and (b), respectively.

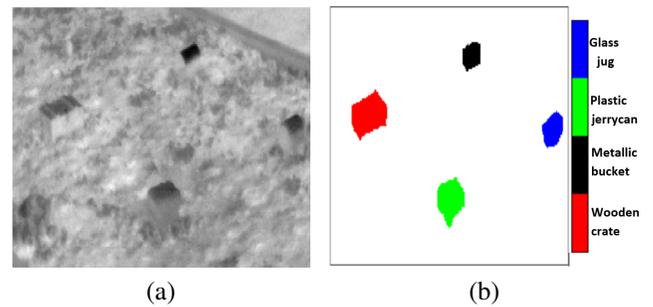


Fig. 4. LWIR data with four targets. (a) A snapshot of the 3rd spectral band (898.48 cm^{-1}) image. (b) The ground truth.

In the scene, the wooden crate (T1) is too large to be an anomaly. Therefore, for the LRX, DWRX and SSRX detectors, the size of inner and outer windows was selected as the size of plastic jerrycan (25x31 pixels) that is the second largest target in the scene. The size of inner and outer window was set to 25x31 and 49x61 pixel areas, respectively. Fig.5 shows the detection results of the compared anomaly detectors on the LWIR data, without thresholding. LRX and DWRX detected the targets (T2-T4) with a few false alarms,

while the GRX generated much more false alarms. The RXD-UTD suppressed the background. The SSRX detected all targets including T1 with many false alarms. The performance of detectors was compared using the receiver operating characteristic (ROC) in Fig.6. Note that the DWRX is an outstanding detector among five of them, similarly. KRX was not used here as it takes about 2 seconds to compute and invert the Gram matrix for every text pixel in the image for windows of this size.

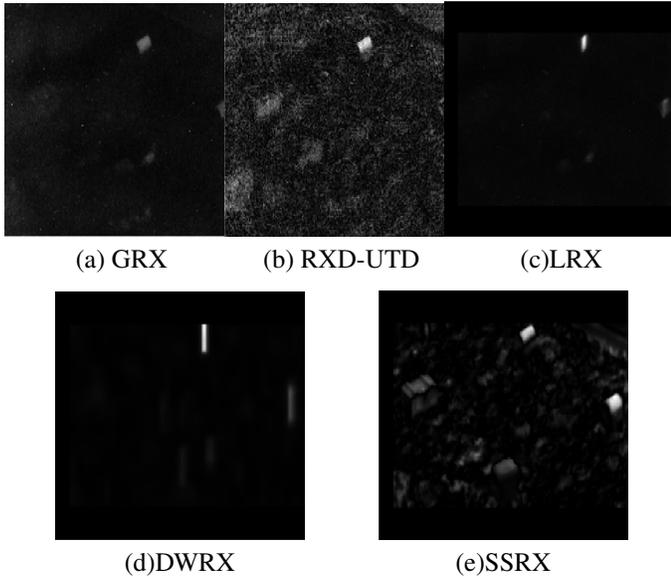


Fig. 5. Detection results for LWIR data.

AD	On Synt. Data (in sec.)	On Real Data (in sec.)
GRX	0.057361	0.443171
RXD-UTD	0.057646	0.480788
LRX	1.215343	303.109355
DWRX	3.963727	475.239997
SSRX	1.182056	256.108270
KRX	21.555117	—

Table 1. The elapse time of detectors on both synthetic and real data for a fixed window size. Anomaly detectors were run on a PC with Windows 8.1 core i7 2.4GHZ-16GB RAM.

4. CONCLUSIONS

We compared the performance of six RX-based anomaly detectors comprehensively and pointed out the issues in parameter selection. When we consider that it is not time-consuming and generates less false alarms, DWRX is an outstanding anomaly detector among six variations of the RX.

When the inner window size was small or equal to the target, DWRX detected only center of anomaly regions as in Fig.2. When window sizes were increased as in Fig.3; it found the whole target. This is also a pretty interesting result since

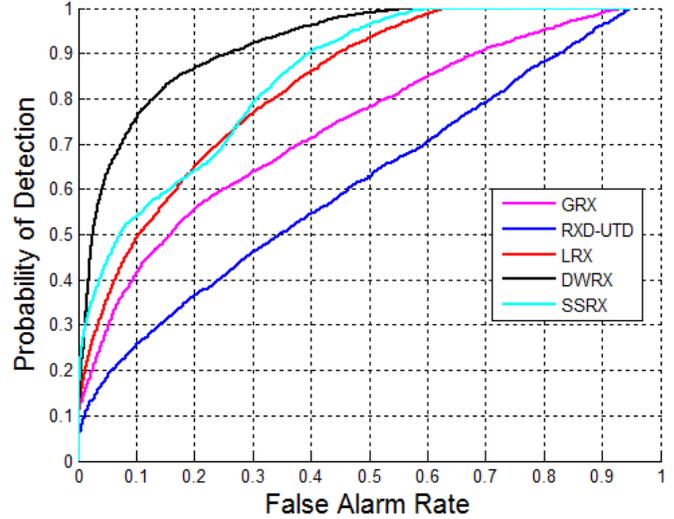


Fig. 6. ROC curves of the anomaly detectors on LWIR data.

in some applications, we are really interested in the center of the target and not the whole target itself.

References

- [1] Stefania Matteoli, Marco Diani, and Giovanni Corsini, “A tutorial overview of anomaly detection in hyperspectral images,” *Aerospace and Electronic Systems Magazine, IEEE*, vol. 25, no. 7, pp. 5–28, 2010.
- [2] Irving S Reed and Xiaoli Yu, “Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution,” *Acoustics, Speech and Signal Processing, IEEE Trans. on*, vol. 38, no. 10, pp. 1760–1770, 1990.
- [3] Chein-I Chang and Shao-Shan Chiang, “Anomaly detection and classification for hyperspectral imagery,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 40, no. 6, pp. 1314–1325, 2002.
- [4] Michael Theodore Eismann, “Hyperspectral remote sensing,” SPIE, 2012.
- [5] Edward A Ashton and Alan Schaum, “Algorithms for the detection of sub-pixel targets in multispectral imagery,” *Photogrammetric Engineering & Remote Sensing*, vol. 64, no. 7, pp. 723–731, 1998.
- [6] Weimin Liu and Chein-I Chang, “A nested spatial window-based approach to target detection for hyperspectral imagery,” in *IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2004, vol. 1.
- [7] N Nasrabadi, “Hyperspectral target detection: An overview of current and future challenges,” *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 34–44, 2014.
- [8] Heesung Kwon and Nasser M Nasrabadi, “Kernel rx-algorithm: a nonlinear anomaly detector for hyperspectral imagery,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 2, pp. 388–397, 2005.