BINARY PARTITION TREES-BASED ROBUST ADAPTIVE HYPERSPECTRAL RX ANOMALY DETECTION

M.A. Veganzones[1], J. Frontera-Pons[2], F. Pascal[2], J.-P. Ovarlez[2,3] J. Chanussot[1,4]

[1]GIPSA-lab, Grenoble-INP, Saint Martin d'Hères, France
[2]SONDRA Research Alliance, Supélec, France
[3]French Aerospace Lab, ONERA DEMR/TSI, France
[4]University of Iceland, Reykjavik, Iceland

ABSTRACT

The Reed-Xiaoli (RX) is considered as the benchmark algorithm in multidimensional anomaly detection (AD). However, the RX detector performance decreases when the statistical parameters estimation is poor. This could happen when the background is non-homogeneous or the noise independence assumption is not fulfilled. For a better performance, the statistical parameters are estimated locally using a sliding window approach. In this approach, called adaptive RX, a window is centered over the pixel under the test (PUT), so the background mean and covariance statistics are estimated using the data samples lying inside the window's spatial support, named the secondary data. Sometimes, a smaller guard window prevents those pixels close to the PUT to be used, in order to avoid the presence of outliers in the statistical estimation. The size of the window is chosen large enough to ensure the invertibility of the covariance matrix and small enough to justify both spatial and spectral homogeneity. We present here an alternative methodology to select the secondary data for a PUT by means of a binary partition tree (BPT) representation of the image. We test the proposed BPT-based adaptive hyperspectral RX AD algorithm using a real dataset provided by the Target Detection Blind Test project.

Index Terms— Anomaly detection, RX AD, binary partition trees

1. INTRODUCTION

Target detection (TD) and anomaly detection (AD) of multidimensional signals have proved to be valuable techniques in a wide range of applications, including search-and-rescue, surveillance, rare minerals detection, mine detection, etc [1]. TD aims to discover the presence of a specific signal of interest (the target) among a set of signals. TD is based on the Neyman-Pearson (NP) criterion, which maximizes the probability of detection for a given probability of false alarm. AD is a special case of TD in which no a-priori target is provided. Hence, the goal of AD is to detect signals that are anomalous respect to the background [2].

The Reed-Xiaoli (RX) AD algorithm [3] is considered as the benchmark algorithm in multidimensional AD. However, the RX detector performance strongly relies on the statistical parameters estimation. Hence, when the background is non-homogeneous or the noise independence assumption is not fulfilled, the detector performance is deteriorated. The statistical parameters can be estimated globally or locally. The local version is called adaptive RX, since it adapts to the local properties of the data. The conventional approach to develop an adaptive RX algorithm is by using sliding windows. For a given pixel under test (PUT), a small window with the size of the expected anomalies maximum size, named the guard window, is centred on the PUT. A second large window, usually named outer window, is also centred on the PUT. The pixels inside the outer window, except those lying inside the guard window, compose the secondary data used to estimate the statistical parameters for the given PUT (see Fig. 1).
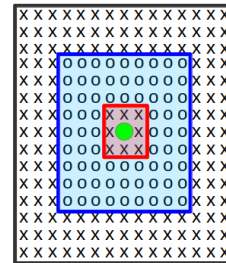

Fig. 1. Sliding window over the pixel under test (green circle): (red) 3×3 guard window and (blue) 15×15 outer window.

In this work we propose to use a binary partition tree (BPT)-based approach to define the secondary data. The BPT is a hierarchical tree representation of the data that exploits the spatial and spectral information contained on the image [4]. Analogously to the sliding window-based definitions of a guard and an outer windows, we define a guard and an outer nodes based on the BPT representation of the image that serves as an alternative definition of the secondary

data of a given PUT. This BPT-based definition of secondary data allows to overcome the limitations of the conventional sliding window approach. On one hand it helps to define more homogeneous background regions and on the other hand it is flexible in order to fit adequately the geometry of the background.

We assessed the performance of the proposed BPT-based approach compared to the conventional sliding windows-based approach on a real hyperspectral dataset provided by the *Target Detection Blind Test* project[1].

The remainder of the paper is as follows: in Sec. 2 and Sec. 3 we overview the hyperspectral RX AD algorithm and BPT representation respectively. In Sec. 4 we introduce the proposed BPT-based approach. In Sec. 5 we provide the experimental validation and finally, in Sec. 6 we give some conclusion remarks.

2. RX ADAPTIVE ANOMALY DETECTION

The RX algorithm was derived from the Generalized Likelihood Ratio Test (GLRT) assuming Gaussian hypothesis [3]:

$$\begin{cases} \mathcal{H}_0 : \mathbf{y} = \mathbf{b} \\ \mathcal{H}_1 : \mathbf{y} = \mathbf{s} + \mathbf{b} \end{cases}, \tag{1}$$

where $\mathbf{b}$ represents the background and $\mathbf{s}$ denotes the presence of an anomalous signal. The adaptive detector is obtained by replacing the unknown parameters by their estimates. For example, an estimate may be obtained from the range cells surrounding the cell under test. The size of the cell has to be chosen large enough to ensure the invertibility of the covariance matrix and small enough to justify both spectral homogeneity (stationarity) and spatial homogeneity. The use of a sliding mask (adaptive RX) provides a more realistic scenario than when estimating the parameters using all the pixels in the image. Thus, the mean vector $\boldsymbol{\mu}$ and the background covariance matrix $\boldsymbol{\Sigma}$, are estimated from N signal free secondary data, $\mathbf{y}_i = 1, \ldots, N$. The resulting GLRT decision rule is the following:

$$\mathbf{t}_{\text{RX}}(\mathbf{y}) = (\mathbf{y} - \hat{\boldsymbol{\mu}}_{\text{SMV}})^T \hat{\boldsymbol{\Sigma}}_{\text{SCM}}^{-1} (\mathbf{y} - \hat{\boldsymbol{\mu}}_{\text{SMV}}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda, \tag{2}$$

where $\hat{\boldsymbol{\mu}}_{\text{SMV}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i$ and $\hat{\boldsymbol{\Sigma}}_{\text{SCM}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i \mathbf{y}_i^{\mathbf{T}}$ denote the sample mean vector (SMV) and sample covariance matrix (SCM) estimators respectively.

For Gaussian distribution, the quadratic form $(\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})$ follows a χ^2 distribution for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ perfectly known. This quadratic form is usually known as the Mahalanobis distance [5]. When the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are replaced by their maximum likelihood estimate parameters, $\hat{\boldsymbol{\mu}}_{\text{SMV}}$ and $\hat{\boldsymbol{\Sigma}}_{\text{SCM}}$, the distribution of the quadratic form:

$$\frac{N - m + 1}{Nm} (\mathbf{y} - \hat{\boldsymbol{\mu}}_{\text{SMV}})^T \hat{\boldsymbol{\Sigma}}_{\text{SCM}}^{-1} (\mathbf{y} - \hat{\boldsymbol{\mu}}_{\text{SMV}}) \sim F_{m, N-m+1}, \tag{3}$$

can be approximated by the non-central F-distribution with m and $N - m + 1$ degrees of freedom [6]. For high values of N (i.e. $N > 10m$), the distribution can be approximated by the χ^2 distribution.

3. BINARY PARTITION TREES

In the BPT representation, the leaf nodes correspond to an initial partition of the image, which can be the individual pixels, or a coarser segmentation map [4]. From this initial partition, an iterative bottom-up region merging algorithm is applied until only one region remains. This last region represents the whole image and corresponds to the root node. All the nodes between the leaves and the root result of the merging of two adjacent children regions. An example of BPT is displayed in Fig. 2. If the initial partition contains n leaf nodes, the final BPT contains $2n - 1$ nodes.
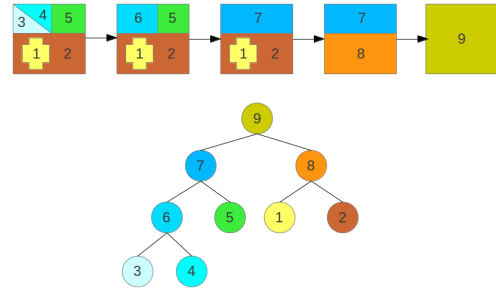

Fig. 2. Construction of the Binary Partition Tree (BPT).

Two notions are of prime importance when defining a BPT: i) the *region model* $\mathcal{M}_{\mathcal{R}}$ which specifies how a region $\mathcal{R}$ is modeled, and ii) the *merging criterion* $\mathcal{O}(\mathcal{M}_{\mathcal{R}_\alpha}, \mathcal{M}_{\mathcal{R}_\beta})$, which is a distance measure between the region models of any two regions $\mathcal{R}_\alpha$ and $\mathcal{R}_\beta$. Each merging iteration involves the search of the two adjacent regions which achieve the lowest pair-wise similarity among all the pairs of adjacent regions in the current segmentation map. Those two regions are consequently merged.

Given a hyperspectral region $\mathcal{R}$, with $N_{\mathcal{R}}$ hyperspectral samples $\mathbf{y_j} \in \mathbb{R}^{\mathbf{q}}$, $j \in 1 \ldots N_{\mathcal{R}}$, the first-order parametric model $\mathcal{M}_{\mathcal{R}}$ is defined by the sample mean vector of the hyperspectral samples $\hat{\boldsymbol{\mu}}_{\mathcal{R}}$:

$$\mathcal{M}_{\mathcal{R}} : \hat{\boldsymbol{\mu}}_{\mathcal{R}} = \frac{1}{N_{\mathcal{R}}} \sum_{j=1}^{N_{\mathcal{R}}} \mathbf{y}_j. \tag{4}$$

Using the first-order parametric model (4), a merging criterion is defined as the spectral angle distance, d_{SAM}, between the sample mean vectors of any two adjacent regions [7]:

$$\mathcal{O}(\mathcal{M}_{\mathcal{R}_\alpha}, \mathcal{M}_{\mathcal{R}_\beta}) : d_{\text{SAM}}(\hat{\boldsymbol{\mu}}_{\mathcal{R}_\alpha}, \hat{\boldsymbol{\mu}}_{\mathcal{R}_\beta}), \tag{5}$$

where $d_{\text{SAM}}(\mathbf{a}, \mathbf{b}) = \arccos\left(\frac{\mathbf{ab}}{\|\mathbf{a}\| \|\mathbf{b}\|}\right)$.

The building of a BPT may suffer from small and meaningless regions resulting in a spatially unbalanced tree. To overcome this limitation, a priority term is included in the merging criterion that forces those regions smaller than a given percentage of the average region size to be merged first [8].

4. BPT-BASED ADAPTIVE RX

Given a BPT representation of a hyperspectral image as it is explained in Sec. 3. Let a branch of a PUT, $\mathbf{x}$, denoted as $\mathcal{B}(\mathbf{x})$, define the sequence of nodes ascending on the BPT representation from the leaf containing the PUT up to the root node. For instance, given the BPT representation depicted in Fig. 2, the branch of a PUT in the leaf number 5 is: $\mathcal{B}(\mathbf{x}) = \{5 \rightarrow 7 \rightarrow 9\}$. This branch definition is a sorted list of nodes starting in the leaf node and ending in the root node. Then, the guard and outer nodes as defined as follows:

- The *guard node* of a PUT, $G(\mathbf{x})$, is the first node in the PUT's branch, $\mathcal{B}(\mathbf{x})$, containing the PUT and at least, a given number of guard pixels.

- The *outer node* of a PUT, $O(\mathbf{x})$, is the first node in the PUT's branch, $\mathcal{B}(\mathbf{x})$, containing the PUT, the guard pixels and at least, a given number of secondary pixels.

Thus, in order to define the guard and outer nodes it is necessary to set the number of guard and secondary pixels, which work in a similar fashion to the guard and outer windows size respectively. The secondary data of a PUT, $S(\mathbf{x})$, is then defined by the pixels contained in the outer node once the pixels in the guard node has been removed:

$$S(\mathbf{x}) = O(\mathbf{x}) \backslash G(\mathbf{x}), \tag{6}$$

where $A \backslash B$ denotes the complement operation between sets A and B.

5. EXPERIMENTAL METHODOLOGY AND RESULTS

5.1. Methodology

In order to test the proposed approach, we used a hyperspectral dataset provided by the *Target Detection Blind Test* project. The dataset includes a high-resolution hyperspectral image, spectral libraries of targets in the scene, and the location of targets in the scene. The hyperspectral imagery was collected by the HyMap sensor operated by HyVista[2]. The georegistered HyMap imagery has approximately 3 meter ground resolution and it is available both in calibrated spectral radiance as well as in spectral reflectance after atmospheric compensation [9]. Fig. 3 depicts a false color image of the dataset. Three civilian vehicles and four small ($1m - 3m$) fabric panels were used as targets, for a total of 129 target pixels in the image. Fig. 4 shows the location of the targets in the map.

Fig. 3. Hyperspectral dataset provided by the *Target Detection Blind Test* project.
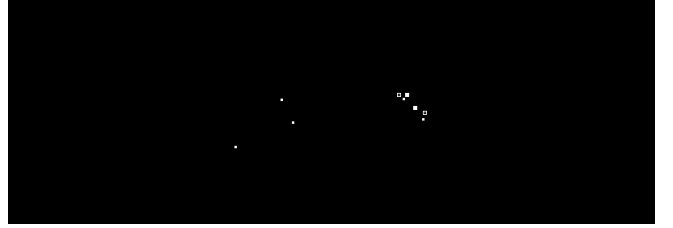

Fig. 4. Location of the targets on the image.

The output of the RX adaptive detector, either using the sliding window or the BPT-based approaches, has been calculated using equation (3). For the sliding window approach, the guard window is of size 3×3 and the outer window is of size 21×21. For the BPT-based approach, the guard node should have at least 9 pixels while the outer node should have at least 400 pixels. For each of the two approaches, the probability of detection (P_{D}) has been calculated for different values of the probability of false alarm (P_{FA}), using as ground-truth the location of the 129 target pixels given by the dataset provider.

5.2. Results

Fig. 5 shows the P_{FA}-P_{D} plot (ROC curves) comparing the window-based and the BPT-based RX adaptive hyperspectral anomaly detectors over the dataset. The plot shows that
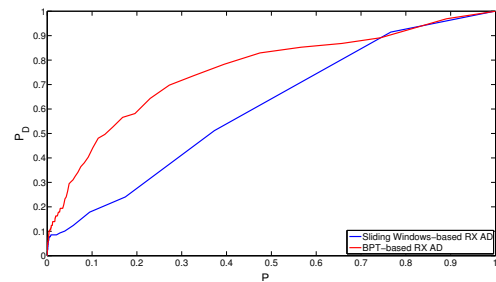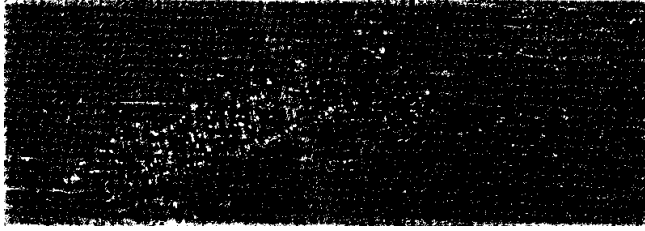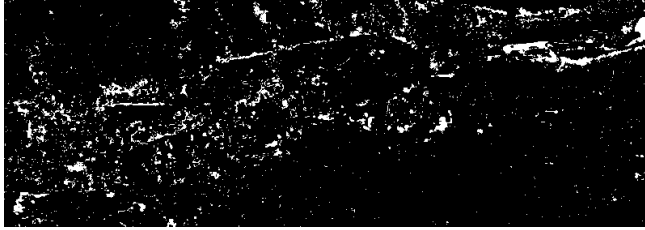
---
[2]http://www.hyvista.com/

Fig. 5. P_{FA} - P_{D} plot comparing the sliding window-based and the BPT-based RX adaptive hyperspectral anomaly detectors over the *Target Detection Blind Test* project's hyperspectral dataset.

(a)

(b)

Fig. 6. Detections of the adaptive RX algorithm ($P_{FA} = 0.05$): a) conventional windowing approach and, b) proposed BPT approach.

the BPT-based approach outperforms the window-based approach in almost all the probability of false alarm range. This improvement is due to the selection of more homogeneous regions to perform the estimation stage. These preliminary results support the use of BPT-representations for an improved estimation of the statistical parameters need in RX adaptive hyperspectral anomaly detection.

Figs. 6 and 7 show the detections obtained by the adaptive hyperspectral RX algorithm over the test dataset for probabilities of false alarm set to 0.05 and 0.10 respectively, using the conventional windowing approach and the proposed BPT-based approach. It is possible to notice that the detected anomalies are more homogeneous and less noisy on the proposed BPT approach, what it suggests that this general approach to identify the secondary data could successfully be extended to other anomaly and target detection techniques.
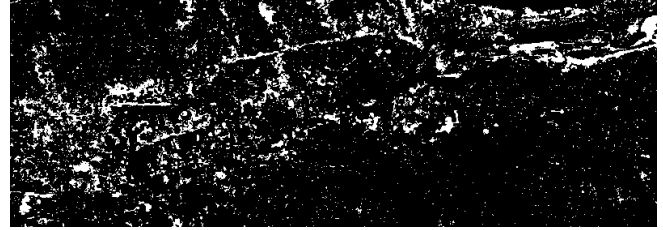
6. CONCLUSIONS

The proposed BPT-based approach to estimate the statistical parameters need for the RX adaptive hyperspectral AD show an improvement over the results obtained by the conventional sliding window-based approach on the *Target Detection Blind Test*'s hyperspectral dataset. These preliminary results encourage the authors to extend this approach to other adaptive anomaly and target detector algorithms. Further experiments will introduce robustness of the proposed approach in terms of the CFAR property of the detection algorithms.

7. REFERENCES

[1] N.M. Nasrabadi, "Hyperspectral target detection : An overview of current and future challenges," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 34–44, Jan 2014.

[2] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerospace and Electronic Systems Magazine*, vol. 25, no. 7, pp. 5–28, July 2010.

[3] 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 Transactions on*, vol. 38, no. 10, pp. 1760–1770, 1990.

[4] Philippe Salembier and Luis Garrido, "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," *Image Processing, IEEE Transactions on*, vol. 9, no. 4, pp. 561–576, 2000.

[5] P. C. Mahalanobis, "On the generalised distance in statistics," in *Proceedings National Institute of Science, India*, 1936, vol. 2, pp. 49–55.

[6] E.W. Weisstein, *CRC Concise Encyclopedia of Mathematics, Second Edition*, Chapman and Hall/CRC, 2002.

[7] M.A. Veganzones, G. Tochon, M. Dalla Mura, A. Plaza, and J. Chanussot, "Hyperspectral image segmentation using a new spectral mixture-based binary partition tree representation," in *2013 IEEE International Conference on Image Processing (ICIP)*, 2013.

[8] G. Tochon, J. Feret, R.E. Martin, R. Tupayachi, J. Chanussot, and G.P. Asner, "Binary partition tree as a hyperspectral segmentation tool for tropical rainforests," in *2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2012, pp. 6368–6371.

[9] D. Snyder, J. Kerekes, I. Fairweather, R. Crabtree, J. Shive, and S. Hager, "Development of a web-based application to evaluate target finding algorithms," in *2008 IEEE International Geoscience and Remote Sensing Symposium*, July 2008, vol. 2, pp. II–915–II–918.

(a)

(b)

Fig. 7. Detections of the adaptive RX algorithm ($P_{FA} = 0.10$): a) conventional windowing approach and, b) proposed BPT approach.