

# ROBUST ANOMALY DETECTION IN HYPERSPECTRAL IMAGING

*J. Frontera-Pons<sup>1</sup>, M.A. Veganzones<sup>2</sup>, S. Velasco-Forero<sup>3</sup>, F. Pascal<sup>1</sup>, J.P. Ovarlez<sup>1,4</sup>, J. Chanussot<sup>2,5</sup>*

<sup>1</sup> SONDRRA Research Alliance, Supélec, France

<sup>2</sup>GIPSA-lab, Grenoble-INP, Saint Martin d’Hères, France

<sup>3</sup>Department of Mathematics, National University of Singapore, Singapore

<sup>4</sup>French Aerospace Lab, ONERA DEMR/TSI, France

<sup>5</sup>University of Iceland, Reykjavik, Iceland

## ABSTRACT

Anomaly Detection methods are used when there is not enough information about the target to detect. These methods search for pixels in the image with spectral characteristics that differ from the background. The most widespread detection test, the RX-detector, is based on the Mahalanobis distance and on the background statistical characterization through the mean vector and the covariance matrix. Although non-Gaussian distributions have already been introduced for background modeling in Hyperspectral Imaging, the parameters estimation is still performed using the Maximum Likelihood Estimates for Gaussian distribution. This paper describes robust estimation procedures more suitable for non-Gaussian environment. Therefore, they can be used as plug-in estimators for the RX-detector leading to some great improvement in the detection process. This theoretical improvement has been evidenced over two real hyperspectral images.

*Index Terms*— hyperspectral imaging, anomaly detection, elliptical distributions, M-estimators

## 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 mineral and land mines detection, etc (see for e.g. [1, 2]). TD aims to discover the presence of a specific signal of interest (the target) among a set of signals. Statistical 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. The Reed-Xiao (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. Accordingly, when the background is non-homogeneous or the noise independence assumption is not fulfilled, the detector performance can be deteriorated.

Here, we highlight a third drawback in the estimation problems: the presence of outliers in the secondary data used for the parameters estimation.

In hyperspectral imaging, the actual distribution of the background pixels differs from the theoretically predicted under Gaussian hypothesis. In fact, as stated in [4], the empirical distribution usually has heavier tails compared to the Gaussian distribution, and these tails strongly influence the observed false-alarm rate of the detector. One of the most general and acknowledged model for background statistics characterization is the family of Elliptically-Contoured Distributions (ECD). They account for non-Gaussianity providing a long tailed alternative to multivariate normal model. They are proven to represent a more accurate characterization of HSI than models based on Gaussian assumption [4]. Although non-Gaussian distributions have already been assumed for background modeling, the parameters estimation is still performed using classical Gaussian based estimators; as in the case of covariance matrix, generally determined by the Sample Covariance Matrix (SCM) and the mean vector with the Sample Mean Vector (SMV). These classical estimators correspond to the Maximum Likelihood Estimators (MLE) for Gaussian assumption. However, they lead to sub-optimal detection schemes when the noise is a non-Gaussian process. When working on ECD framework the model can be used to assess the robustness of statistical procedures and to derive alternative robust estimators of the parameters, the mean vector and the covariance matrix [5, 6]. These can then be used as plug-in estimators in place of the unknown mean vector or/and of the covariance matrix. This is a simple but often efficient method to obtain robust properties for signal processors derived under the Gaussian assumption.

## 2. ELLIPTICALLY CONTOURED DISTRIBUTIONS

Hyperspectral data have been proven not to be multivariate normal but long tailed distributed. In order to take into account these features, the class of elliptically-contoured distributions is considered to describe clutter statistical behavior. It

provides a multivariate location-scatter family of distributions that primarily serves as heavy tailed alternative to the multivariate normal model. An  $m$ -dimensional random complex vector  $\mathbf{y} = [y_1, y_2, \dots, y_m]^T$  with mean  $\boldsymbol{\mu}$  and scatter matrix  $\boldsymbol{\Sigma}$  has an elliptical distribution if its probability density function (PDF) has the form [7]:

$$f_{\mathbf{y}}(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1} h_m((\mathbf{y} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})) \quad (1)$$

where  $H$  denotes the conjugate transpose operator and  $h_m(\cdot)$  is any function such as (1) defines a PDF. If the second-order moment exists, then  $\boldsymbol{\Sigma}$  reflects the structure of the covariance matrix of the elliptically distributed random vector  $\mathbf{y}$ , i.e. the covariance matrix is equal to the scatter matrix up to a scalar constant. It serves to characterize the correlation structure existing within the spectral bands. It is worth pointing out that the ECD class includes a large number of distributions, notably the Gaussian distribution, multivariate  $t$  distribution,  $K$ -distribution or multivariate Cauchy. Thus, it allows for heterogeneity of the background power with the texture.

### 3. ROBUST PARAMETERS ESTIMATION

Along with their well-known properties and their simplicity of analysis, the SCM and the SMV are the most extended estimates since they are the MLEs for Gaussian case.

$$\hat{\boldsymbol{\mu}}_{SMV} = \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i \quad \hat{\boldsymbol{\Sigma}}_{SCM} = \frac{1}{N} \sum_{i=1}^N (\mathbf{y}_i - \hat{\boldsymbol{\mu}})(\mathbf{y}_i - \hat{\boldsymbol{\mu}})^H \quad (2)$$

where  $N$  denotes the number of secondary data. However, such widespread techniques are suboptimal when the noise is a non-Gaussian stochastic process. This article reviews some robust procedures particularly suited for estimating the covariance matrix and the mean vector of elliptical populations.

The Fixed Point estimators, according to the definition proposed by Tyler in [8], satisfy the following equations:

$$\hat{\boldsymbol{\mu}}_{FP} = \frac{\sum_{i=1}^N \frac{\mathbf{y}_i}{\left( (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_{FP})^H \hat{\boldsymbol{\Sigma}}_{FP}^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_{FP}) \right)^{1/2}}}{\sum_{i=1}^N \frac{1}{\left( (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_{FP})^H \hat{\boldsymbol{\Sigma}}_{FP}^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_{FP}) \right)^{1/2}}} \quad (3)$$

$$\hat{\boldsymbol{\Sigma}}_{FP} = \frac{m}{N} \sum_{i=1}^N \frac{(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_{FP})(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_{FP})^H}{\left( (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_{FP})^H \hat{\boldsymbol{\Sigma}}_{FP}^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_{FP}) \right)} \quad (4)$$

The Fixed Point estimates have been widely investigated in statistics and signal processing literature. We refer to [9] for a detailed performance analysis. It is worth pointing out that  $\hat{\boldsymbol{\Sigma}}_{SCM}$  and  $\hat{\boldsymbol{\Sigma}}_{FP}$  have the same asymptotic Gaussian

distribution which differs on their second order moment by a factor  $\frac{m+1}{m}N$ , i.e. for  $N$  sufficiently large,  $\hat{\boldsymbol{\Sigma}}_{FP}$  behaves as a Wishart matrix with  $\frac{m}{m+1}$  degrees of freedom.

### 4. RX ADAPTIVE ANOMALY DETECTION

The RX algorithm was derived from the Likelihood Ratio assuming Gaussian hypothesis:

$$\begin{cases} \mathcal{H}_0 : \mathbf{y} = \mathbf{b} \\ \mathcal{H}_1 : \mathbf{y} = \mathbf{s} + \mathbf{b} \end{cases}, \quad (5)$$

where  $\mathbf{s}$  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 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 surrounding the pixel under test,  $\mathbf{y}_i$ ,  $i = 1, \dots, N$ . The resulting GLRT decision rule is the following:

$$\mathbf{t}_{RX}(\mathbf{y}) = (\mathbf{y} - \hat{\boldsymbol{\mu}}_{SMV})^H \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{y} - \hat{\boldsymbol{\mu}}_{SMV}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda. \quad (6)$$

and  $\lambda$  is a given threshold. When Gaussian assumption is valid, the quadratic form  $(\mathbf{y} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})$  follows a  $\chi^2$  distribution for  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\mu}$  perfectly known. This quadratic form is usually known as the Mahalanobis distance [10]. When the parameters and under Gaussian assumptions  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\mu}$  are replaced by their MLE parameters and under Gaussian assumptions (2), the distribution of the quadratic form

$$\frac{N - m + 1}{mN} (\mathbf{y} - \hat{\boldsymbol{\mu}}_{SMV})^H \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{y} - \hat{\boldsymbol{\mu}}_{SMV}) \sim F_{m, N-m+1}$$

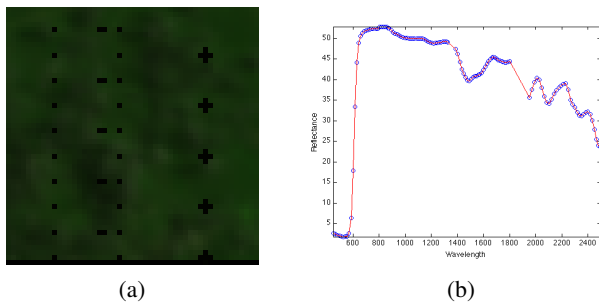
follows a Hotelling  $T^2$  distribution  $F_{m, N-m+1}$  which is the non-central  $F$ -distribution with  $m$  and  $N - m + 1$  degrees of freedom [11]. For high values of  $N$ , ( $N > 10m$ ), the distribution can be approximated by the  $\chi^2$  distribution. However, real hyperspectral scenes can not be described only with Gaussian distribution, as mentioned above. In this work we explore the use of *Fixed Point* estimators in the classical RX detector :

$$\mathbf{t}_{RX-FP}(\mathbf{y}) = (\mathbf{y} - \hat{\boldsymbol{\mu}}_{FP})^H \hat{\boldsymbol{\Sigma}}_{FP}^{-1} (\mathbf{y} - \hat{\boldsymbol{\mu}}_{FP}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda. \quad (7)$$

It is important to highlight that the distribution of this detector is still an open question, as far as the authors are aware.

## 5. RESULTS

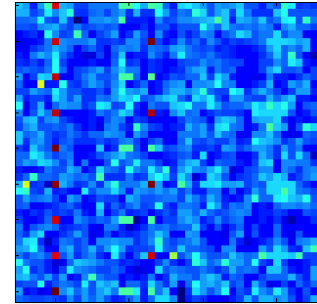
The experiments were conducted firstly on a real hyperspectral image where artificial targets with known spectral signature were introduced as anomalies in the background, see Fig. 1. The original data set consist in  $50 \times 50$  pixels with 126 bands. Most of the theory on covariance matrix estimation have been recently extended to complex value signals [12]. Since hyperspectral data are real and positive, we proposed to use a Hilbert filter in order to render them complex. However, it is important to note that the real component after Hilbert transform is still the original signal. To avoid the well-known problem due to high dimensionality we have chosen sequentially eleven bands in the complex representation. In this approach, both covariance matrix and mean vector are estimated using a sliding window of size  $9 \times 9$ , having  $N = 80$  secondary data. The results for this image are shown on the Fig. 2.



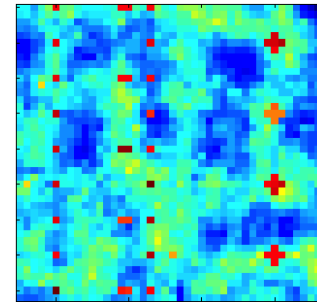
**Fig. 1.** (a) Original background image with artificial anomalies, (b) Endmember used in the experiment.

The results obtained with (4) show that the robust detector  $t_{RX-FP}$  is capable of locate all the artificial targets and present a lower number of false alarms. This improvement is due to the fact that *Fixed Point* estimators treat the outliers and impulsive samples in order for them to have a smaller contribution to the background characterization process, while the SMV-SCM estimates suffer from the presence of strong reflectance pixels in the secondary data.

The algorithm has also been applied for galaxy detection on the MUSE data cube. The Multi Unit Spectroscopic Explorer (MUSE) project (see [13]) aims to provide astronomers with a new generation of optical instrument, capable of simultaneously imaging the sky (in 2D) and measuring the optical spectra of the light received at a given position on the sky. MUSE was installed on the VLT telescope and operational in 2013, and its performances are expected to allow observation of far galaxies up to 100 times fainter than those presently detectable. MUSE will deliver a 3D data-cube made of a stack of images recorded at 3578 different wavelengths over the range 465- 930 nm. Each monochromatic image repre-



(a) Classical RX detector



(b)RX detector built with the FP estimates

**Fig. 2.** Careful readers should remark that the targets are easily observed with a naked eye in (b).

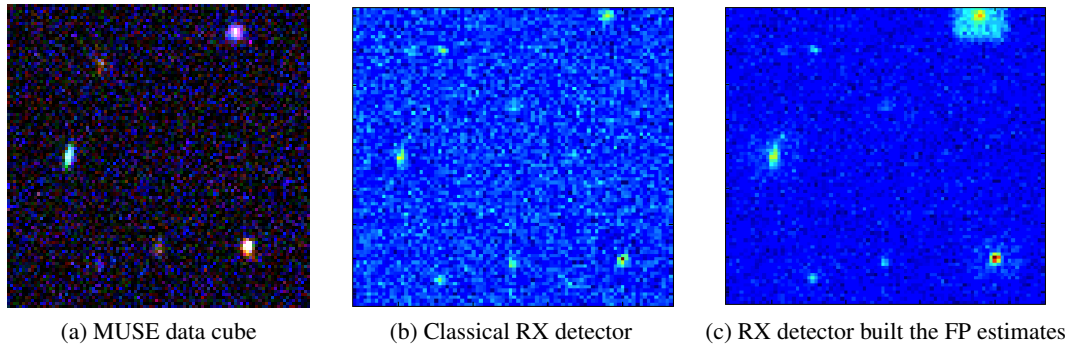
sents a field of view of  $60 \times 60$  arcsec, recorded with a spatial sampling of 0.2 arcsec. Each record results in a data cube of size 1570 MB encoding 3578 images of  $300 \times 300$  pixels, possibly containing thousands of objects (galaxies) existing over different subsets of wavelengths!

An example of MUSE data cube image is displayed in Fig. 3, from the 3578 available bands, we have chosen one band of each 100 after Hilbert transformation. The results for anomaly detection are presented in Fig.3. Note that detection with fixed-point estimators (c) provides results with lower false alarm rate than classical ones (b).

These examples illustrate the robust behavior of *Fixed Point* estimators in non-Gaussian environments or for close targets detection problems.

## 6. CONCLUSIONS

The family of elliptical distributions is considered for impulsive background characterization in hyperspectral imaging. In this context, robust estimation methods for mean vector and covariance matrix are used to overcome the non-Gaussianity of the background and the presence of outliers or strong scatters in the secondary data. Moreover, the use of the robust *Fixed Point* estimators for anomaly detection purposes has been discussed and compared to the classical SMV-SCM Gaussian estimators. The theoretical improvement provided by the robustness of the estimators is borne out through two real hyperspectral images.



**Fig. 3.** Classical and Fixed-point anomaly detection in a hyperspectral image of  $300 \times 300$  in 3578 channels. See details in the text of Section 5.

## 7. REFERENCES

- [1] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aero. and Electr. Systems Mag.*, vol. 25, no. 7, pp. 5–28, July 2010.
- [2] D.W.J. Stein, S.G. Beaven, L.E. Hoff, E.M. Winter, A.P. Schaum, and A.D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 58–69, Jan 2002.
- [3] I.S. Reed and X. 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] D. Manolakis and D. Marden, "Non gaussian models for hyperspectral algorithm design and assessment," in *Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International*. IEEE, 2002, vol. 3, pp. 1664–1666.
- [5] F. Gini and M. V. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Processing, special section on SP with Heavy Tailed Distributions*, vol. 82, no. 12, pp. 1847–1859, December 2002.
- [6] J. Frontera-Pons, M. Mahot, JP Ovarlez, F. Pascal, SK Pang, and J. Chanussot, "A class of robust estimates for detection in hyperspectral images using elliptical distributions background," in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*. IEEE, 2012, pp. 4166–4169.
- [7] Douglas Kelker, "Distribution theory of spherical distributions and a location-scale parameter generalization," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 419–430, 1970.
- [8] D.E. Tyler, "A distribution-free  $M$ -estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.
- [9] F. Pascal, P. Forster, J.-P. Ovarlez, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *IEEE Trans.-SP*, vol. 56, no. 6, pp. 2206–2217, June 2008.
- [10] Prasanta Chandra Mahalanobis, "On the generalized distance in statistics," *Proceedings of the National Institute of Sciences (Calcutta)*, vol. 2, pp. 49–55, 1936.
- [11] Eric W Weisstein, *CRC concise encyclopedia of mathematics*, CRC press, 2010.
- [12] Melanie Mahot, Frédéric Pascal, Philippe Forster, and Jean-Philippe Ovarlez, "Asymptotic properties of robust complex covariance matrix estimates," 2013.
- [13] "Official website of the muse project "http://muse.univ-lyon1.fr/"", .