Combined Local-Global Background Modeling for Anomaly Detection in Hyperspectral Images

Research Thesis

In Partial Fulfillment of The Requirements for the Degree of Master of Science in Electrical Engineering

Eyal Madar

Submitted to the Senate of the Technion - Israel Institute of Technology

tevet, 5771 Haifa december, 2010

THIS RESEARCH THESIS WAS SUPERVISED BY PROFESSOR DAVID MALAH AND DOCTOR MEIR BARZOHAR UNDER THE AUSPICES OF THE ELECTRICAL ENGINEERING DEPARTMENT

Acknowledgements

I would like to express my deep and sincere gratitude to my supervisor, Prof. David Malah, for all his dedicated personal guidance, encouragement and support during all the years of this research.

I am heartily thankful to my supervisor, Dr. Meir Barzohar, for his detailed and constructive comments.

During this work I have collaborated with many colleagues for whom I have great regard, and I wish to extend my warmest thanks to Nimrod, Avi, Ziva and Yair, who have helped me with my work in the Signal and Image Processing Laboratory.

I wish to thank my parents. They bore me, raised me, supported me, taught me, and loved me.

Lastly, and most importantly, I owe my loving thanks to my son, Itay, and my wife, Johanna, for all her support and patience. To them I dedicate this thesis.

Contents

List of Figures **ix**

List of Tables **ix**

Abstract **x**

List of Symbols and Abbreviations **xiii**

1 Introduction **1**

1.1 Hyperspectral Imaging 1

1.2 Anomaly Detection 2

1.3 Background Modeling 3

1.4 Thesis Outline 5

2 Statistical Background Modeling Approaches in Hyperspectral Imaging **7**

2.1 Local Approach: RX 7

2.1.1 Theory 7

2.1.2 RX performance for Real Hyperspectral Data 10

2.2 Global Approach: Gaussian Mixture Model RX (GMM-RX) . . . 13

2.2.1 GMM-RX on Real Hyperspectral Data 15

2.3 Summary 17

3 Background Extreme Value Analysis (BEVA) for Anomaly Detection 18

3.1 Drawbacks of Local and Global Approaches 18

3.2 Proposed Local-Global Approach 20

3.2.1 Background cluster hypothesis test 20

3.2.2 Local background modeling 22

3.2.3 Interrelating local background models 27

3.3 Experiments with Real Data 29

3.4 Summary 30

4 Global Auxiliary Dictionary 32

4.1 BEVA Global Filter Drawback 32

4.2 Auxiliary Dictionary using a GMM Algorithm 33

4.3 Experiments with Real Data 34

4.4 Summary 35

5 Spectral Clustering for Anomaly Detection 37

5.1 BEVA's Local Part Drawback 37

5.2 Spectral Clustering 38

5.2.1 Graph Theory 39

5.2.2 Spectral Clustering Algorithm 41

5.2.3 Local Scaling 43

5.2.4 Segmentation Results for Real Hyperspectral Data 44

5.3 Anomaly Detection using Spectral BEVA 46

5.4 Experiments with Real Data 49

5.5 Summary 50

6 Non-Gaussian Background Modeling for Anomaly Detection 52

6.1 The Gaussian Assumption in BEVA 52

6.2 Gamma Fitting for Background Modeling 53

6.3 Experiments with Real Data 57

6.4 Summary 59

7 Conclusion 61

7.1 Summary 61

7.2 Future Work 66

References 68

A Experimental Data 73

B GMM Parameters Estimation 77

C Distribution of maximum-norm Gaussian realizations 80

D Derivation of conditional hypothesis probabilities 83

E Maximum Likelihood Estimation of Gamma Distribution 84

List of Figures

1.1 **Hyperspectral Imaging** 2

2.1 **Concentric windows for local anomaly detection.** 9

2.2 **ROC curve for RX - Data Set I.** 11

2.3 **ROC curve for RX - Data Set II.** 11

2.4 **A typical RX false alarm.** Right image - RX score map with a strong false alarm , Left image - An RGB representation of the area with the false alarm marked by a red ellipse 13

2.5 **ROC curve GMM-RX vs. RX - Data Set I.** 16

2.6 **ROC curve GMM-RX vs. RX - Data Set II.** 16

3.1 **Conditional hypotheses probabilities $P(H_0|\eta)$ and $P(H_1|\eta)$ obtained from $N = 10,000$ pixels of 65 spectral bands** . . . 22

3.2 **BEVA's Local Part - Block Diagram** 27

3.3 **ROC curve BEVA vs. RX, GMM-RX and MCD - Data Set I.** 29

3.4 **ROC curve BEVA vs. RX, GMM-RX and MCD - Data Set II.** 30

4.1 **ROC curve BEVA with the auxiliary dictionary vs. BEVA - Data Set I.** 34

4.2 **ROC curve BEVA with the auxiliary dictionary vs. BEVA - Data Set II.** 35

5.1 **Data Representation using Graph Theory** 39
5.2 **Local Scaling Effect (from [29])** 44
5.3 **Spectral Clustering Result - Global vs. Local Scaling Parameter.** (a) - Single hyperspectral band. (b) - Spectral Clustering Result with Global Scaling Parameter. (c) Spectral Clustering with Local Scaling Parameter 45
5.4 **Spectral BEVA's Local Part - Block Diagram** 48
5.5 **ROC curve Spectral BEVA vs. BEVA - Data Set I.** 49
5.6 **ROC curve Spectral BEVA vs. BEVA - Data Set II.** . . . 50

6.1 **Block diagram of the Local Part of Spectral NG-BEVA** . 57
6.2 **ROC curve Spectral NG-BEVA vs. Spectral BEVA - Data Set I.** 58
6.3 **ROC curve Spectral NG-BEVA vs. Spectral BEVA - Data Set II.** 59

7.1 **ROC curves of all the examined algorithms - Data Set 1** . 64
7.2 **ROC curves of all the examined algorithms - Data Set 2** . 65
7.3 **Computation Time (Matlab®)** 66

A.1 **Hyperspectral false RGB representation - cube 1.** 74
A.2 **Hyperspectral false RGB representation - cube 2.** 75
A.3 **High resolution CCD image of the analyzed scene,** used as a ground-truth indication The ground-truth anomalies are encircled by red ellipses. 76

List of Tables

3.1 First Stage of a background cluster estimation 25
3.2 Second Stage of a background cluster estimation 26
3.3 Global Filter 28

5.1 Spectral Clustering Algorithm [27] 42
5.2 Spectral BEVA - Background Parameter Estimation 47

6.1 NG-BEVA - First Stage 55
6.2 NG-BEVA - Second Stage 56

B.1 Hard-Clustering EM algorithm for Gaussian Distribution 79

Abstract

The detection of materials and objects using remotely sensed spectral information collected by hyperspectral sensors has many military and civilian applications. Detection algorithms exploit the spectral information present in hyperspectral data to detect and discriminate localized man-made targets, e.g., small buildings, vehicles, etc. In anomaly detection, no prior knowledge on the target spectral signature is assumed. Therefore, anomaly detection algorithms first model the abundant material spectra (background process). Then, every pixel or group of pixels spectrally different in a meaningful way from the background process are declared to be anomalies.

According to the hyperspectral literature, two major approaches to statistical background modeling can be distinguished. In the first approach, named "local", the background is modeled by a large number of local independent distributions, each of which is responsible to represent a different local region in the image. Local algorithms can tightly fit the background data, however they are subject to an overfitting problem, which may produce an excessive number of false-alarms. The second background modeling approach, denoted "global", is based on a global representation of the background process in the whole image. By design, this approach is more resistant to the overfitting problem. However, it has a limited ability to adapt to all nuances of the background process (an

underfitting problem), which may result in high false alarm rates, as well as low anomaly detection rates.

In this research, we propose a combination of the local and global background modeling approaches by introducing the BEVA (Background Extreme Value Analysis) algorithm. In the local part of BEVA, the local background is approximated using a greedy sequential estimation process. It is composed of a robust estimation of the Gaussian statistics and a background cluster hypothesis discriminator, which is based on Extreme Value Theory results. Then, in its global part, the obtained local background models are inter-related to reduce the number of false alarms. BEVA has the ability to adapt to all nuances of the background process like the local approach but avoids overfitting.

In the next part of the research, we propose several improvements to the BEVA algorithm. First, we reinforce BEVA's global filter performance by adding an auxiliary background dictionary to deal with spatially dispersed background pixels that are a source of false alarms in BEVA. Second, we improve BEVA's local part via a preprocessing segmentation that is based on Spectral Clustering. The Gaussian model used in BEVA, although efficient and mathematically tractable, is only partially adequate to represent real hyperspectral data. In order to overcome this drawback, we introduce the NG-BEVA (Non-Gaussian BEVA) algorithm, a non-Gaussian version of BEVA, which replaces the Gaussian assumption with Gamma distribution fitting.

The results strongly prove the effectiveness of the proposed "local-global" approach. On real hyperspectral data, our local-global algorithms perform better than other examined global or local anomaly detection techniques.

List of Symbols and Abbreviations

Symbols

$\mathbf{X}$	matrix with observed vectors as its columns
$\mathbf{x}_i$	column number i of $\mathbf{X}$
d_i	Mahalanobis distance of pixel $\mathbf{x}_i$
V_i	verticle of a pixel $\mathbf{x}_i$
w_{ij}	similarity between pixels $\mathbf{x}_i$ and $\mathbf{x}_j$
σ	scaling parameter
p	dimensionality of observations
$\boldsymbol{\mu}$	mean vector
$\boldsymbol{\Sigma}$	covariance matrix
$N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Gausssian distribution
$\chi^2(p)$	Chi-squared distribution of order p
k	shape parameter of a Gamma distribution
Θ	scale parameter of a Gamma distribution
$\boldsymbol{\Gamma}(k, \Theta)$	Gamma distribution
Γ	Gamma function
ℓ_2-norm	square root of sum of squares

Abbreviations

BEVA Background Extreme Value Analysis

EM Expectation Maximization

EVT Extreme Value Theory

GLRT Generalized Likelihood Ratio Threshold

GMM-RX Gaussian Mixture Model RX algorithm

HSI Hyperspectral Imaging

MCD Maximum Covariance Determinant

MOCA Maximum Orthogonal-Complements Algorithm

NG-BEVA Non Gaussian Background Extreme Value Analysis

ROC Receiver Operating Curves

RX Reed-Xiaoli algorithm, a benchmark anomaly detection algorithm for hyperspectral imagery

SNR Signal to Noise Ratio

SC Spectral Clustering

SVD Singular Value Decomposition

VNIR Visible and Near Infra Red

Chapter 1

Introduction

1.1 Hyperspectral Imaging

Hyperspectral remote sensing exploits the fact that all materials reflect, absorb, and emit electromagnetic energy, at specific wavelengths, in distinctive patterns related to their molecular composition. Compared to a typical color camera that uses three wavelength bands, corresponding to the red, green and blue colors, hyperspectral imaging (HSI) sensors acquire digital images in many contiguous and very narrow (nominally about 10 nm wide) spectral bands that typically span the visible, near-infrared, and mid-infrared portions of the spectrum (0.4-2.5 μm). This enables the construction of an essentially continuous radiance spectrum for every pixel in the scene. The end result of the high spectral resolution of HSI is that we can identify and classify the materials present in the scene and, as presented in this work, detect targets of interest.

In Fig. 1.1, the spectrum signatures of three different materials (soil, vegetation and water) are obtained from a hyperspectral image. Due to the high spectral resolution, it is easy to discriminate between each material.

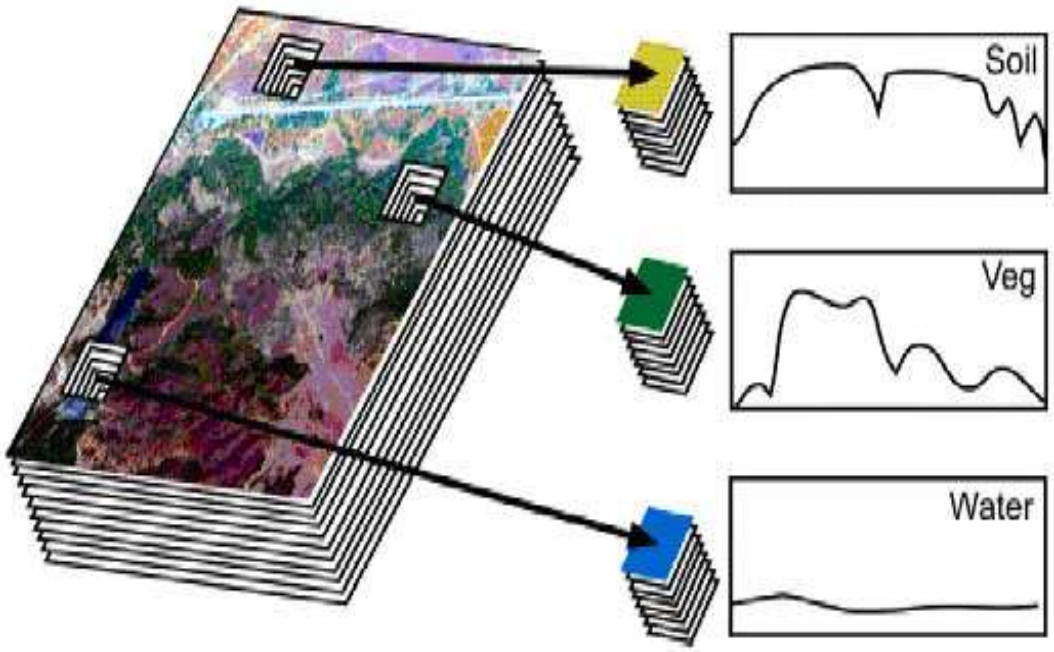

Figure 1.1: **Hyperspectral Imaging**

1.2 Anomaly Detection

The detection of materials and objects using remotely sensed spectral information has many military and civilian applications. Detection algorithms exploit the spectral information present in hyperspectral data to detect and discriminate localized man-made objects, e.g., small buildings, vehicles, etc.

Detection algorithms can be divided in two classes: supervised detection and unsupervised detection. In supervised detection (also called "Target Detection"), the main objective is to search for a specific given material spectrum (target) in a hyperspectral image. The spectral signature of the target is known "a-priori" from laboratory measurements.

In this research, we focus our attention on the problem of unsupervised detection in hyperspectral images also called "Anomaly Detection". Anomaly detection algorithms do not require any knowledge of the spectral signature of the targets. Anomalies are defined as patterns in data that do not conform to a well defined notion of normal behavior. In anomaly detection tasks, hyperspectral pixels have to be classified into either background (normal behavior) or anomalies. Every pixel or group of pixels that are spectrally different in a meaningful way from the abundant (background) material spectra are declared to be anomalies.

Since neither prior anomaly signatures nor their statistical model, are known, anomaly detection methods first model the background and then detect anomalies by finding pixels that are not well-described by the background model. It turns out that the problem of background pixels modeling is a critical and a subtle task. As a matter of fact, it poses a two-fold problem: On one hand, the model has to be general enough in order to accurately represent the wealth of background material spectra, so as to avoid false alarms due background pixel deviations from the model. On the other hand, the model has to be concise enough, limiting its ability to adapt to anomalies, and leaving anomalies to disagree with the model, which is essential for a high probability of detection.

1.3 Background Modeling

According to hyperspectral related publications, there are two main background modeling methodologies, one based on a linear mixture model ([1], [2], [3], [9]) where the background process is modeled using a small number of representative spectra known as "endmember", the second on statistical modeling ([4]). Two major approaches to statistical background modeling can be distinguished.

In the first approach, the background is modeled by a large number of local independent distributions, each of which is responsible to represent a different local region in the image. Therefore, we call this approach "local". A classical local background modeling algorithm, named RX [4], has become a benchmark anomaly detection algorithm in the remote sensing community. Many other local algorithms have also been proposed by the hyperspectral society like in [6], [7] or [8]. Most of local algorithms are based on the Gaussian model. Since the Gaussian assumption for modeling the local background process is in general not valid, Kwon and Nassrabadi proposed, in [5], a nonlinear version of the RX algorithm called kernel-RX that transforms each spectral pixel into a very high-dimensional feature space (could be of infinite dimensions) by a nonlinear mapping function. Detection is made by assuming a Gaussian distribution to the local background process in the high-dimensional feature space. Modeling the input data in the feature space by a Gaussian distribution is equivalent to representing the distribution of the input data with a much more complex model when defined in the original input space.

The second background modeling approach is based on a global representation of the background process in the whole image. Therefore, we call this approach "global". There are various statistical global modeling algorithms found in the literature ([11], [13], [15] or [16]). In the global approach, the background is modeled using a simple universal distribution, which is designed to represent the background process in the whole image. For example, the global background process is modeled by a mixture of Gaussians in the GMM-RX algorithm (Gaussian Mixture Model RX) proposed in [11].

Classical local and global methods rely on non-robust covariance matrix estimation which are highly sensitive to the presence of outliers. The background

modeling obtained by a non-robust statistic estimation can lead to a decrease in the anomaly detector's performance. The authors of [14] proposed a robust global anomaly detection based on the Minimum Covariance Determinant (MCD), a highly robust estimator of multivariate statistics introduced by Rousseeuw in [12]. MCD's objective is to find h (out of n) observations whose covariance matrix has the lowest determinant. In hyperspectral data, n is the number of pixels of a local region in the image. These observations are then used to estimate the background process by a Gaussian with robust parameters.

As explained in the sequel, in section 3.1, the local approach is subject to an overfitting problem, which may produce an excessive number of false-alarms. On their part, global algorithms have a limited ability to adapt to all nuances of the background process (an underfitting problem), which may result in high false alarm rates, as well as low anomaly detection rates.

In this research we present a novel statistical background modeling approach denoted "local-global". The local-global background model has the ability to adapt to all nuances of the background process like in the local approach but avoids overfitting. The local model, because it is composed of a number of distinct clusters, is capable of handling multiple types of terrain. Then a global filter is applied to alleviate the overfitting problem by inter-relating the independent local background models.

1.4 Thesis Outline

The thesis is organized as follows: In Chapter 2, we focus our attention on two representative anomaly detection methods: RX [4] as a local approach algorithm and GMM-RX [11] as a global one.

In Chapter 3, we present our proposed local-global background modeling algorithm named BEVA (Background Extreme Value Analysis). In the local part of BEVA, the local background is approximated using a greedy sequential estimation process. It is composed of a robust estimation of the Gaussian statistics and a background cluster hypothesis discriminator, which is based on Extreme Value Theory results (EVT) [24]. EVT was already applied to hyperspectral images for redundancy reduction that preserves anomalies [9]. Their results strongly prove the effectiveness of EVT in hyperspectral data. In the global part of BEVA, the obtained local background models are inter-related to reduce the number of false alarms.

In Chapter 4, we notice that BEVA's global part cannot deal with spatially dispersed background pixels which are a source of false alarms. We propose to reinforce BEVA's global filter performance by adding an auxiliary background dictionary. Chapter 5 deals with the improvement of the local part of BEVA via a segmentation that is based Spectral Clustering [29].

Finally, in Chapter 6 we develop a technique for non-Gaussian background model estimation, denoted NG-BEVA (Non-Gaussian BEVA). Indeed, as most local algorithms, BEVA's local part models each local background cluster by a Gaussian distribution, although the hyperspectral data is not following closely the Gaussian distribution. In NG-BEVA, we remove the Gaussian assumption by using Extreme Value Theory combined with Gamma distribution fitting.

All the algorithms developed in this thesis are tested on real hyperspectral data (see Appendix A) and their performance are compared using ROC (Receiver operating characteristic) curves.

Chapter 2

Statistical Background Modeling Approaches in Hyperspectral Imaging

In this chapter, we focus our attention on two anomaly detection algorithms derived from the statical background modeling methodology. First, we examine the classical "RX" algorithm [4] based on the local approach. The second algorithm we present is a global algorithm called "Gaussian Mixture Model RX" [11] (GMM-RX) where a global representation of the background process in the whole image is used to detect anomalies. The algorithms are explained bellow in detail and tested with real hyperspectral data.

2.1 Local Approach: RX

2.1.1 Theory

In this section we summarize the classical local anomaly detector proposed by Reed and Xiaoli in [4], named RX, which has become a benchmark anomaly detection algorithm in the remote sensing community.

In the RX algorithm, local background pixels are assumed to be independent, identically distributed, Gaussian random vectors. The model assumes that the data arises from two Gaussian pdfs with the same covariance matrix $\mathbf{\Sigma}$ but different means. Thus, two hypotheses can be formulated as:

$$\begin{aligned} H_0 &= N(\boldsymbol{\mu}, \mathbf{\Sigma}) \qquad \text{(Anomaly absent)} \\ H_1 &= N(\boldsymbol{\mu} + \mathbf{s}, \mathbf{\Sigma}) \qquad \text{(Anomaly present)} \end{aligned} \tag{2.1}$$

where $\boldsymbol{\mu}$ is the background mean vector and $\mathbf{s} = (s_1, s_2, ..., s_p)^T$ is the target mean spectra.

For each tested pixel $\mathbf{y}$, a double concentric square window is used to separate a local area into two regions - the inner window region (IWR) and the outer window region (OWR) as shown in Fig. 2.1. The IWR size is generally set so that it can fully enclose a target. Another concentric square centered at the tested pixel known as the "guard window region" can be utilized as well. The guard band is slightly larger in size than the IWR but smaller than the OWR. The main purpose of the guard band is to reduce the probability that some target pixels will be present in the OWR and hence affect the background statistics estimation.

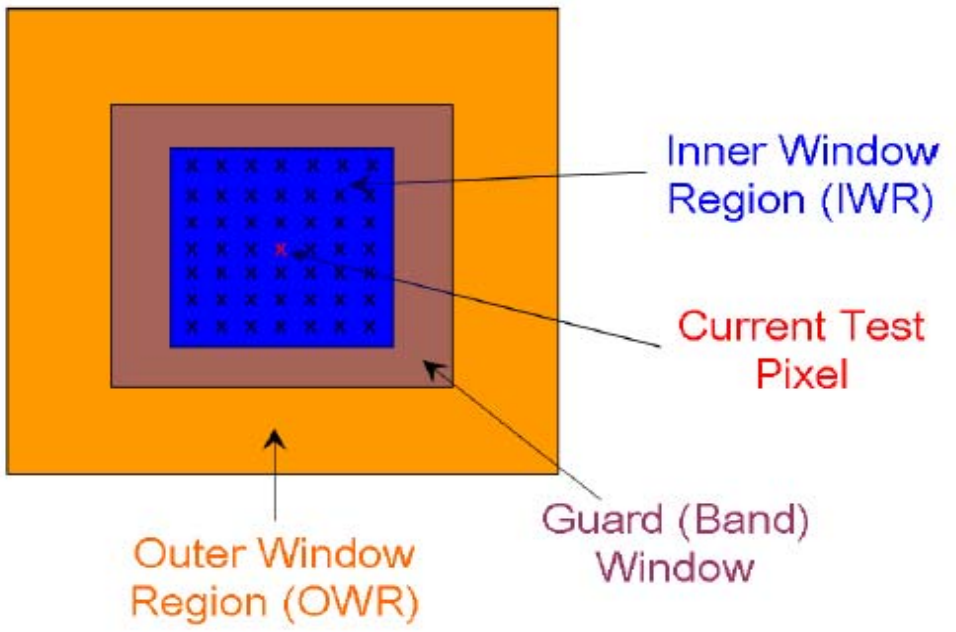

Figure 2.1: **Concentric windows for local anomaly detection.**

The background mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of the RX model are estimated from the OWR as follow:

$$\begin{aligned} \boldsymbol{\mu} &= \frac{1}{N_b}\sum_{i=1}^{N_b}\mathbf{x}_i \\ \boldsymbol{\Sigma} &= \frac{1}{N_b-1}\sum_{i=1}^{N_b}(\mathbf{x}_i-\boldsymbol{\mu})^T(\mathbf{x}_i-\boldsymbol{\mu}) \end{aligned} \tag{2.2}$$

where N_b is the number of pixels in the OWR.

After estimating the background mean vector and covariance matrix, the GLRT (Generalized Likelihood Ratio Threshold) is computed, which leads to a simple test. If the Mahalanobis distance between the tested pixel $\mathbf{y}$ and the

background mean vector $\boldsymbol{\mu}$ is greater than a user-defined threshold Th,

$$RX(y) = (\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) > Th, \tag{2.3}$$

the pixel is declared as an anomaly.

2.1.2 RX performance for Real Hyperspectral Data

The RX algorithm was applied to two data sets of real hyperspectral data (For a description of the hyperspectral data used, see Appendix A). The IWR size was set to 7×7, the OWR size to 15×15.

An anomaly is considered as detected if at least one of the detected pixels hits the corresponding marked segment. All pixels detected by the algorithms were grouped into connected objects using 8-connected object labeling. If an object doesn't intersect a marked anomaly, it is considered a false alarm object. This kind of anomaly detection/miss criteria is particularly suitable for applications that aim to alert the user on all anomalies of all sizes. Therefore, it is more important to detect at least one pixel on each anomaly, rather than many pixels on only some of the anomalies.

RX ROC curves on data set I and II are shown in Fig. 2.2 and 2.3, respectively.

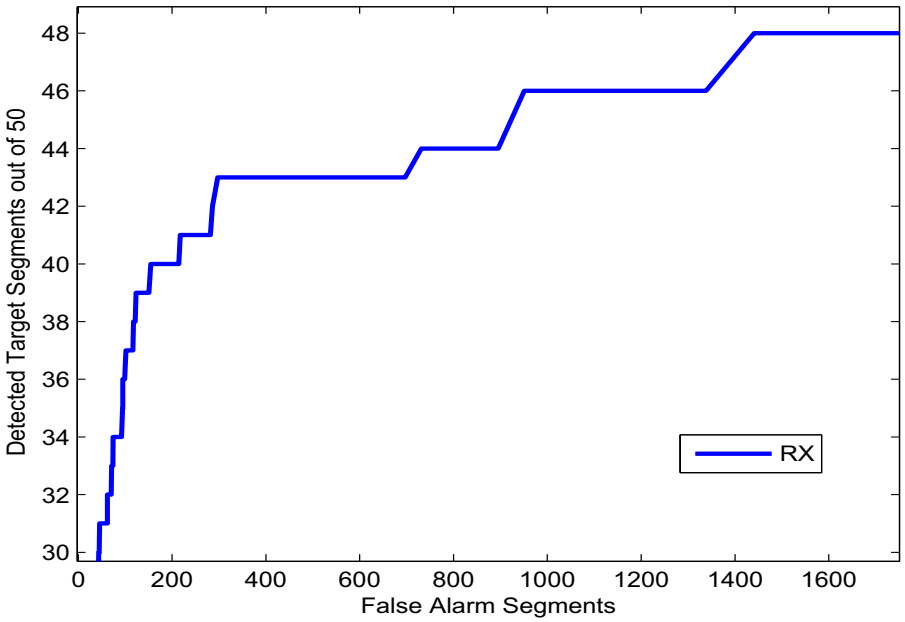

Figure 2.2: **ROC curve for RX - Data Set I.**

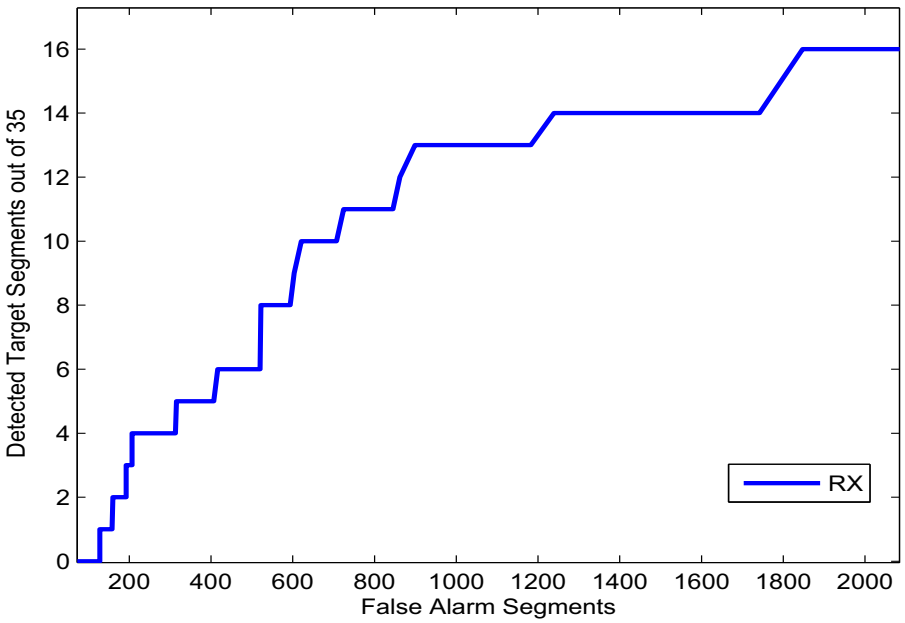

Figure 2.3: **ROC curve for RX - Data Set II.**

A typical false alarm of the RX algorithm (an isolated tree detected as an anomaly despite the large number of similar trees present in the image) is shown in Fig. 2.4. It reveals a major drawback of this algorithm: the RX model has a too high number of degrees of freedom in representing the background process. Indeed, the model is estimated for each pixel using a small local window. An overfitting problem is almost inevitable, leading to an excessive number of false alarms. Furthermore, anomaly pixels present in the OWR can severely affect the background statistic estimation. In fact, RX relies on a non-robust covariance matrix estimation which is highly sensitive to the presence of anomalies. This sensitivity manifests itself in distortions of the estimated covariance that are not representative of the true structure of the data. Two notable problems can result. First, a very strong anomaly present in the OWR may "inflate" the Gaussian distribution that approximates the background process to the point that other anomalies ar missed. Second, an increase in the false alarm rate can occur when the Gaussian distribution is shifted or rotated by such a degree that background pixels are no longer part of the background process.

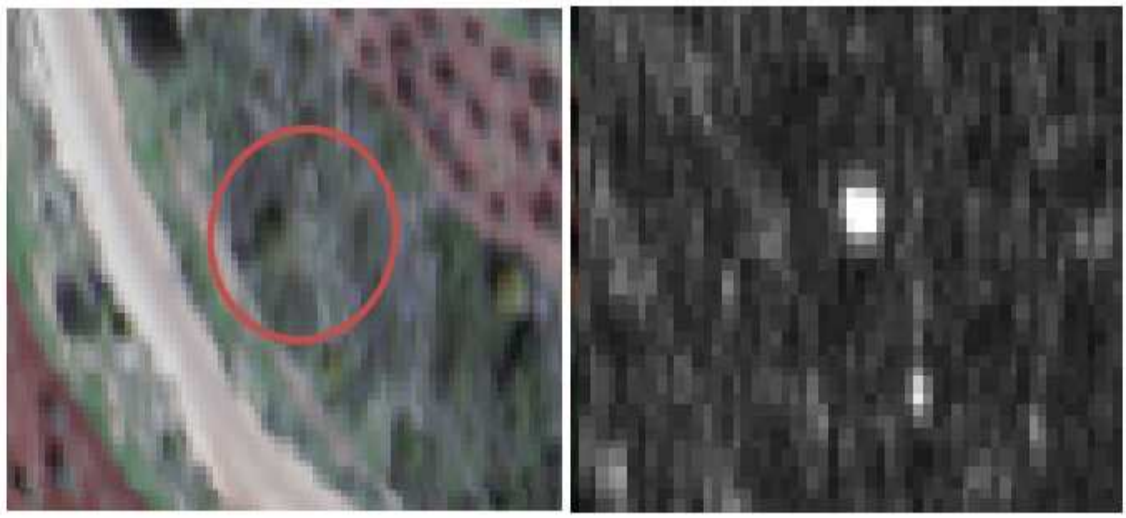

Figure 2.4: **A typical RX false alarm.** Right image - RX score map with a strong false alarm , Left image - An RGB representation of the area with the false alarm marked by a red ellipse

2.2 Global Approach: Gaussian Mixture Model RX (GMM-RX)

In this section, we present a global algorithm, based on a probabilistic model, known as "Gaussian Mixture Model" (GMM).

Unlike the local approach, a unique unimodal multivariate Gaussian distribution is not appropriate for modeling the background of a whole image since it contains multiple types of terrain. A mixture of Gaussian distributions can be used for better characterization of nonhomogeneous backgrounds.

The GMM approach ([16],[11]) assumes that the image data $\mathbf{X}$ in R^p (p-number of spectral bands) arises from a linear combination of Gaussian pdfs,

resulting in a mixture probability density function of the form

$$f(\mathbf{X}) = \sum_{k=1}^{K} \tau_k f_k(\mathbf{X}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{2.4}$$

where K is the number of Gaussians in the mixture and $\{\tau_k\}_{k=1}^{K}$ are the mixing proportions, with $\sum_{k=1}^{K} \tau_k = 1$ and $0 \leq \tau_k \leq 1$. $f_k(\mathbf{X}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ denotes the density of class k defined as a Gaussian distribution with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$.

A Gaussian mixture model can be used for anomaly detection in the same way as a unimodal Gaussian model. An anomaly is, as in local methods, a pixel that does not fit well to the background process. Here an anomaly is a vector that does not fit well to the GMM Model, so a test can be proposed to detect outliers that use the pdf as an anomaly score. If the pdf value of a tested pixel is lower than a user-defined probability threshold value P_{th}, the pixel will be considered as an anomaly:

$$\text{GMM-RX}(\mathbf{y}) = \sum_{k=1}^{K} \tau_k f_k(\mathbf{y}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) < P_{th}, \tag{2.5}$$

where $\mathbf{y}$ is a tested pixel.

While estimating the parameters of a single Gaussian is straightforward, estimating the parameters of a GMM requires more complex estimation procedures. GMM parameters can be estimated using the Expectation-Maximization (EM) algorithm (Appendix B).

In [11], a hard clustering GMM is proposed where a class label $c_k \in \{1, \cdots, K\}$ tells, for each vector $\mathbf{x}_i$, to which class it belongs. This method, which was chosen for our implementation, gives very similar results to the standard GMM, but avoids numerical implementation difficulties. In a hard clustering GMM, a pixel

will be considered as an anomaly if

$$\text{GMM-RX}(\mathbf{y}) = f_{c_k}(\mathbf{y}|\boldsymbol{\mu}_{c_k}, \boldsymbol{\Sigma}_{c_k}) < P_{th}, \tag{2.6}$$

where c_k is the background cluster to whom y was associated during the hard clustering GMM estimation.

Our GMM-RX algorithm was initialized by an excessive number of Gaussians using the k-means algorithm for initializing the Gaussian parameters. During the EM iterations of the GMM-RX, too small clusters, and hence unreliable, were eliminated.

2.2.1 GMM-RX on Real Hyperspectral Data

The GMM-RX algorithm was applied to two data sets of real hyperspectral data (For a description of the hyperspectral data used, see Appendix A) and its ROC curves are presented in Fig. 2.5 and 2.6 together with the RX ROC curves. The initial number of Gaussians was set to 10 for each hyperspectral image.

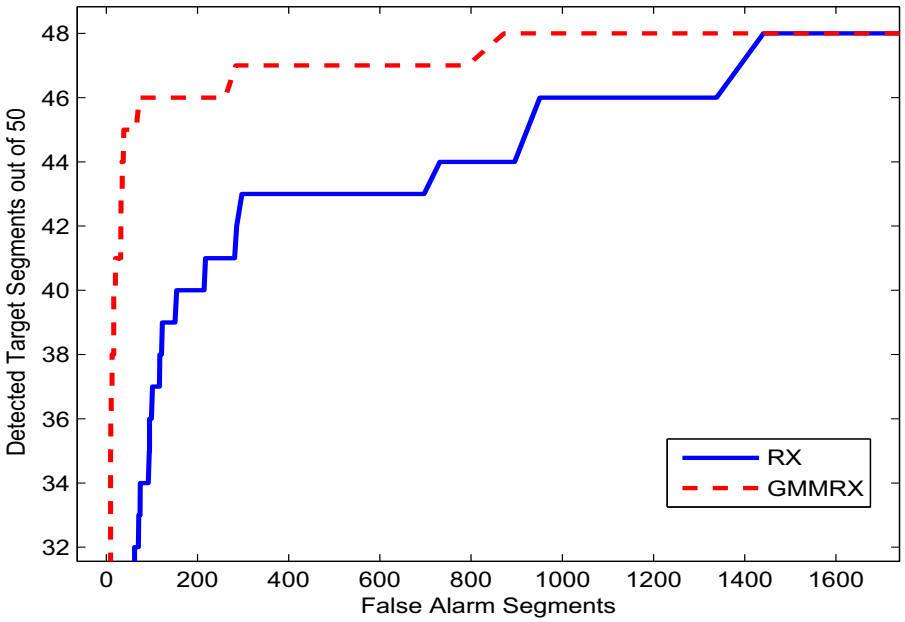

Figure 2.5: **ROC curve GMM-RX vs. RX - Data Set I.**

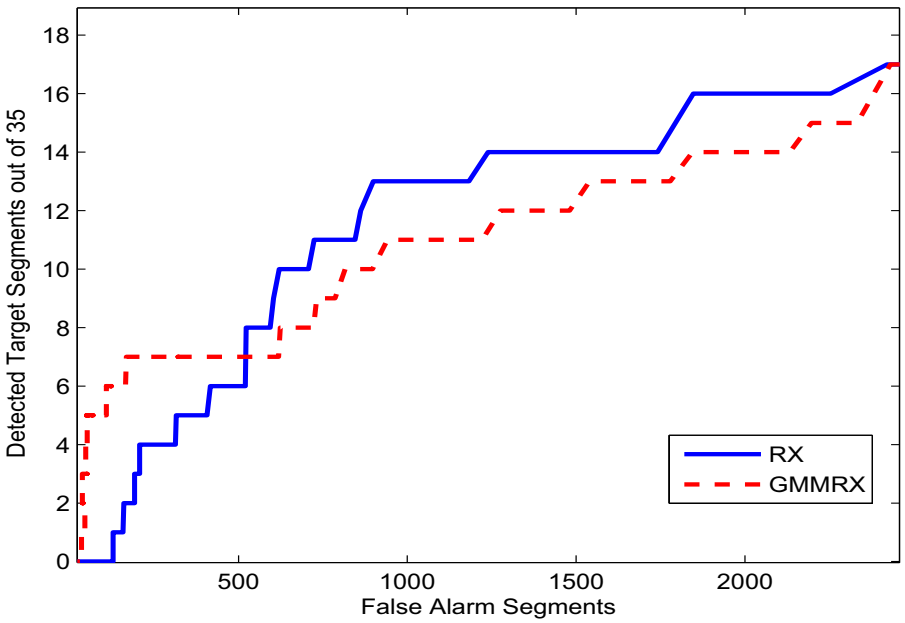

Figure 2.6: **ROC curve GMM-RX vs. RX - Data Set II.**

As it can be seen, the GMM-RX algorithm has fewer false alarm pixels than the RX algorithm for all probability of detection values in data set I. However, weak anomalies detection is difficult because the global modeling of GMM-RX has no ability to adapt to all the nuances of the background. It explains why GMM-RX is outperformed by RX for high value of probability of detection in data set II.

2.3 Summary

In this chapter, we presented in detail two algorithms based on local and global background modeling approaches: the classical RX and GMM-RX. In the local algorithm (RX), the background is modeled by a large number of Gaussian distributions, each of which is responsible to represent a different local region in the image. The second algorithm, called GMM-RX, uses a simple universal distribution, consisting of a linear combination of Gaussian distributions, to represent the background process in the whole image.

For the two data sets of real hyperspectral images, GMM-RX has a better performance than RX in the detection of strong anomalies. However, weak anomalies detection is more difficult in GMM-RX. In the next chapter, we will highlight drawbacks of each approach and introduce a novel algorithm BEVA (Background Extreme Value Analysis) which is based on the combined local-global background modeling approach.

Chapter 3

Background Extreme Value Analysis (BEVA) for Anomaly Detection

3.1 Drawbacks of Local and Global Approaches

In the previous chapter we presented two approaches of background modeling for anomaly detection.

In the local approach, the background is modeled by a large number of local independent distributions, each of which is responsible to represent a different local region in the image. Local background models can be tightly fitted to the background data since each local region has its own estimated parameters. This allows obtaining high detection rate of anomalies. Unfortunately, in many cases, the strength of local models turns into a weakness. In fact, each local background model has a high number of degrees of freedom. In RX, for example, local model parameters are estimated using a small window (OWR). The number of degrees of freedom for the mean and covariance estimation is p and $\frac{p(p+1)}{2}$, respectively, where p is the number of spectral bands. This high number of

degrees of freedom may cause model overfitting. Over-fitting generally occurs when a model is excessively complex, such as having too many degrees of freedom, relative to the amount of data available. Generally, this can be explained by the "Hughes Phenomenon" [20], which states that the number of training data pixels has to be significantly higher than the number of the model degrees of freedom for a correct estimation. A model which over-fits the data will generally produces a high false-alarm rate. Since the number of free parameters in local background models is proportional to the data size, the model overfitting problem is almost inevitable. According to the "Hughes Phenomenon", performance of anomaly detection algorithms significantly deteriorates when the number of pixels is severely limited for an accurate learning of the local background models.

The global background modeling approach is based on a simple universal distribution, which is designed to represent the background process in the whole image. By design, these methods are more resistant to the overfitting problem. However, they have a limited ability to adapt to all nuances of the background process (an underfitting problem), which may result in high false alarm rates, as well as low anomaly detection rates.

Obviously, there is no ultimate answer how to completely avoid the overfitting or underfitting problems, however, one may significantly improve detector performance by a proper combination of the local and global background modeling principles. One way to accomplish this is to use local background models that are not independent, but are interrelated in some way. This construction may help to significantly reduce the vast number of degrees of freedom of the local method, while retaining the ability of local models to be intimately adjusted to the background. We call this approach "local-global".

3.2 Proposed Local-Global Approach

In this research, we propose BEVA (Background Extreme Value Analysis) an algorithm based on the local-global approach in which both the local part and the global part are novel. The local model that is capable of handling multiple types of terrain, is composed of a small number of distinct clusters, L (up to 3), ordered by size, each distributed as a separate Gaussian distribution.

$$\begin{cases} x \in C_k, \;\; 1 \leq k \leq L \\ C_k \sim N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \\ |C_1| \geq |C_2| \geq \cdots \geq |C_L| \end{cases} \tag{3.1}$$

where $|\cdot|$ denotes set cardinality

In the global part, we reduce the number of degrees of freedom by inter-relating the independent local background models. Each found local anomaly is compared to other local background models of a larger image area. If it can be associated to one of the background clusters, it will be removed from the anomaly set.

3.2.1 Background cluster hypothesis test

In this subsection we construct an automatic test that isolates pixels belonging to a specific background cluster in a local image area. It is based on examining the Mahalanobis distance $d = (x - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x - \boldsymbol{\mu})$ of a pixel x to the mean of a background cluster, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and covariance of a Gaussian $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ that approximates this background cluster.

A set of N data-vector indices of a local image area can hypothetically be

divided into two subsets:

$$B = \{\text{indices of realizations of a specific background cluster}\}$$

$$A = \{\text{indices of realizations}$$

$$\text{of other background clusters or anomalies}\}$$

Let be $\eta = \max_{i=1,\cdots,N}(d_i)$, the maximum Mahalanobis distance over this set obtained at index δ. Given η and δ, we formulate the following hypotheses:

$$\begin{aligned} H_0 &: \delta \text{ belongs to } B \\ H_1 &: \delta \text{ belongs to } A \end{aligned} \tag{3.2}$$

Denoting $\upsilon = \max_{i\in B}(d_i)$ and $\xi = \max_{i\in A}(d_i)$, as the maximum Mahalanobis distance of subset A and subset B, respectively, η can be expressed as:

$$\eta = \max(\upsilon, \xi) \tag{3.3}$$

In order to evaluate the conditional probabilities $p(H_0|\eta)$ and $p(H_1|\eta)$, one has to specify pdfs $f_\upsilon(\cdot)$ and $f_\xi(\cdot)$, or, equivalently, cdfs $F_\nu(\cdot)$ and $F_\xi(\cdot)$. While the probability of υ can be determined by Extreme Value Theory results [24], as shown in Appendix C, the pdf of ξ is generally unknown. A possible choice for $f_\xi(\cdot)$ is therefore,

$$\xi \sim U[0, \eta], \tag{3.4}$$

where U denotes a uniform distribution.

It can be shown (see Appendix D for details) that conditional probabilities are given by:

$$p(H_0|\eta) = \frac{\eta f_\upsilon(\eta)}{\eta f_\upsilon(\eta) + F_\upsilon(\eta)}, \tag{3.5}$$

$$p(H_1|\eta) = \frac{F_\upsilon(\eta)}{\eta f_\upsilon(\eta) + F_\upsilon(\eta)}, \tag{3.6}$$

Fig. 3.1 shows the two conditional hypotheses obtained from $N = 10,000$ pixels of 65 spectral bands:

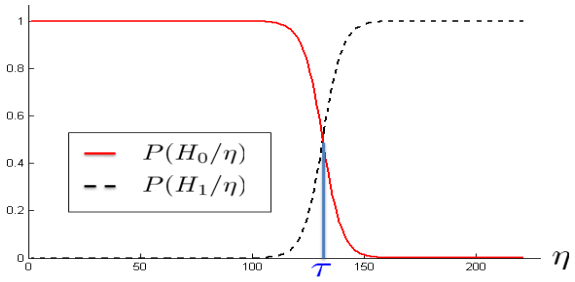

Figure 3.1: **Conditional hypotheses probabilities** $P(H_0|\eta)$ **and** $P(H_1|\eta)$ **obtained from** $N = 10,000$ **pixels of** 65 **spectral bands**

The crossing point τ of the two hypotheses, i.e., the Mahalanobis distance above which $P(H_0|\eta) < P(H_1|\eta)$, can be used as a threshold to isolate the background cluster realizations from other realizations in the data-set. Given a Gaussian statistics $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, which represents a background cluster pdf, a data pixel having a Mahalanobis distance that is below τ, will be declared as a background cluster pixel.

3.2.2 Local background modeling

In this subsection we describe in detail how the local background is estimated and local anomalies found. We develop a greedy sequential algorithm that estimates the unknown number and statistics of the background clusters. The estimation process is applied to a local image area composed of N data-vectors. Starting with the dominant cluster estimation, we initialize two distinct indices sets $A^0 = \emptyset$ and

$B^0 = \{1 : N\}$ and aim to get

$$\begin{aligned} B^f &= \text{Pixel indices of the dominant background cluster} \\ A^f &= \text{Pixels indices of other background clusters or anomalies} \end{aligned} \tag{3.7}$$

By re-initializing $B^0 = A^f$ and $A^0 = \emptyset$, the process is repeated to estimate a new background cluster in the same local image area. However, the remaining number of pixels in A^f has to be sufficient to allow correct estimation. Moreover, if a too small number of pixels remains in A^f, we could estimate a Gaussian pdf that approximates the anomaly class and we will miss anomalies. Thus, an estimation of a new background cluster is allowed only if the remaining number of pixels is greater than a threshold value that is usually dictated by the anomaly supposed size (10% of the block in our test).

Given that the Gaussian statistics of each cluster typically needs to be estimated in the presence of a large number of outliers (other background clusters and anomalies), we propose a two-stage iterative estimation process that combines robust estimation of the Gaussian statistics from [22] with the background cluster hypothesis testing as described in section 3.2.1. The robust estimation process of the Gaussian statistics gives a full weight to observations assumed to come from the background process, but reduces the influence to observation from the tails of the distribution (outliers). The weights calculation and statistics estimation are computed iteratively. For each iteration, the weigh of a pixel is calculated using the Mahalanobis distance of this pixel to the Gaussian distribution with the current estimated statistics. The Gaussian statistics of the next iteration are obtained by the maximum likelihood estimators weighted by the current iteration weights.

In the first stage, we obtain an intermediate set B^{tmp} from the set B^0 that is

exclusively composed of pixels of the dominant background cluster.

However, some pixels of the dominant background cluster may have been wrongly excluded in the previous stage. A second stage is necessary to introduce the excluded pixels back into B.

The two-stage estimation process is described in detail in Tables 3.1 and 3.2.

Inputs: $\{x_m\}_1^N$, p - # of spectral bands

Initialization: $A^0 = \emptyset,\ B^0 = \{1 : N\},\ \omega^0 = 1,$
$d_0 = (\sqrt{p} + \sqrt{2})^2,\ \gamma = 1.25$ and $i = 0$

Main Iteration: Perform the following steps:

1. *Robust estimation of Mean and Covariance* [22]

$$\boldsymbol{\mu}^i = \frac{\sum_{x_m \in B^i} \omega_m^i x_m}{\sum_{x_m \in B^i} \omega_m^i}$$
$$\boldsymbol{\Sigma}^i = \frac{\sum_{x_m \in B^i} (\omega_m^i)^2 (x_m - \boldsymbol{\mu}^i)^T (x_m - \boldsymbol{\mu}^i)}{\sum_{x_m \in B^i} (\omega_m^i)^2 - 1}$$

2. *Calculation of Mahalanobis distances in* B^i *:*

$$\forall x_m \in B^i :\ d_m = (x_m - \boldsymbol{\mu}^i)^T (\boldsymbol{\Sigma}^i)^{-1} (x_m - \boldsymbol{\mu}^i)$$

3. *Update weights* [22]

$$\omega_m^{i+1} = \begin{cases} 1 & if\ \ d_m \le d_0 \\ \frac{d_0}{d_m} \exp(-0.5(d_m - d_0)^2/\gamma^2) & if\ \ d_m \ge d_0 \end{cases}$$

4. *Update sets:*

Find data pixel indices $\{\delta^i\}$ having Mahalanobis distances that exceed the background cluster hypothesis threshold value τ^i which is a function of $p, |B^i|$ and $\boldsymbol{\Sigma}^i$, as explained in Appendix C

$$\{\delta^i\} = \{I(d_m \ge \tau^i)\}\ \ (I = Index\ of\ value)$$
$$B^{i+1} = B^i \backslash \{\delta^i\}$$
$$A^{i+1} = A^i \bigcup \{\delta^i\}$$

5. *Stopping rule:*

If $\{\delta^i\} = \emptyset$, stop. Otherwise increment i and go to 1

Output: $A^{tmp} = A^i, B^{tmp} = B^i, \boldsymbol{\mu}^{tmp} = \boldsymbol{\mu}^i\ and\ \boldsymbol{\Sigma}^{tmp} = \boldsymbol{\Sigma}^i$

Table 3.1: First Stage of a background cluster estimation

Initialization: $A^0 = A^{tmp},\ B^0 = B^{tmp},\ \boldsymbol{\mu}^0 = \boldsymbol{\mu}^{tmp},\ \boldsymbol{\Sigma}^0 = \boldsymbol{\Sigma}^{tmp}\ \ and\ \ j = 0$

Main Iteration: Perform the following steps:

1. *Calculation of Mahalanobis distances in* A^j

$$\forall x_m \in A^j :\ \ d_m = (x_m - \boldsymbol{\mu}^j)^T (\boldsymbol{\Sigma}^j)^{-1} (x_m - \boldsymbol{\mu}^j)$$

2. *Update sets:*

Find pixel indices $\{\delta^j\}$ having Mahalanobis distances that are below the background cluster hypothesis threshold value τ^j which is a function of $p, |B^j|$ and $\boldsymbol{\Sigma}^j$, as explained in Appendix C

$$\{\delta^j\} = \{I(d_m \leq \tau^j)\}$$
$$A^{j+1} = A^j \backslash \{\delta^j\}$$
$$B^{j+1} = B^j \bigcup \{\delta^j\}$$

4. *Stopping rule:*

If $\{\delta^j\} = \emptyset$, stop. Otherwise, estimate $\boldsymbol{\mu}^{j+1}$ and $\boldsymbol{\Sigma}^{j+1}$ using B^{j+1} like in the first stage, increment j by and go to 1

Output: $A^f = A^j,\ B^f = B^j,\ \boldsymbol{\mu} = \boldsymbol{\mu}^j,\ and\ \ \boldsymbol{\Sigma} = \boldsymbol{\Sigma}^j$

Table 3.2: Second Stage of a background cluster estimation

In summary, by applying a greedy sequential algorithm, the local background pdf is estimated using a small number of Gaussian pdfs and all the hyperspectral pixels of the local image area are classified as background or local anomalies. The block diagram of BEVA's local part is shown in Fig. 3.2

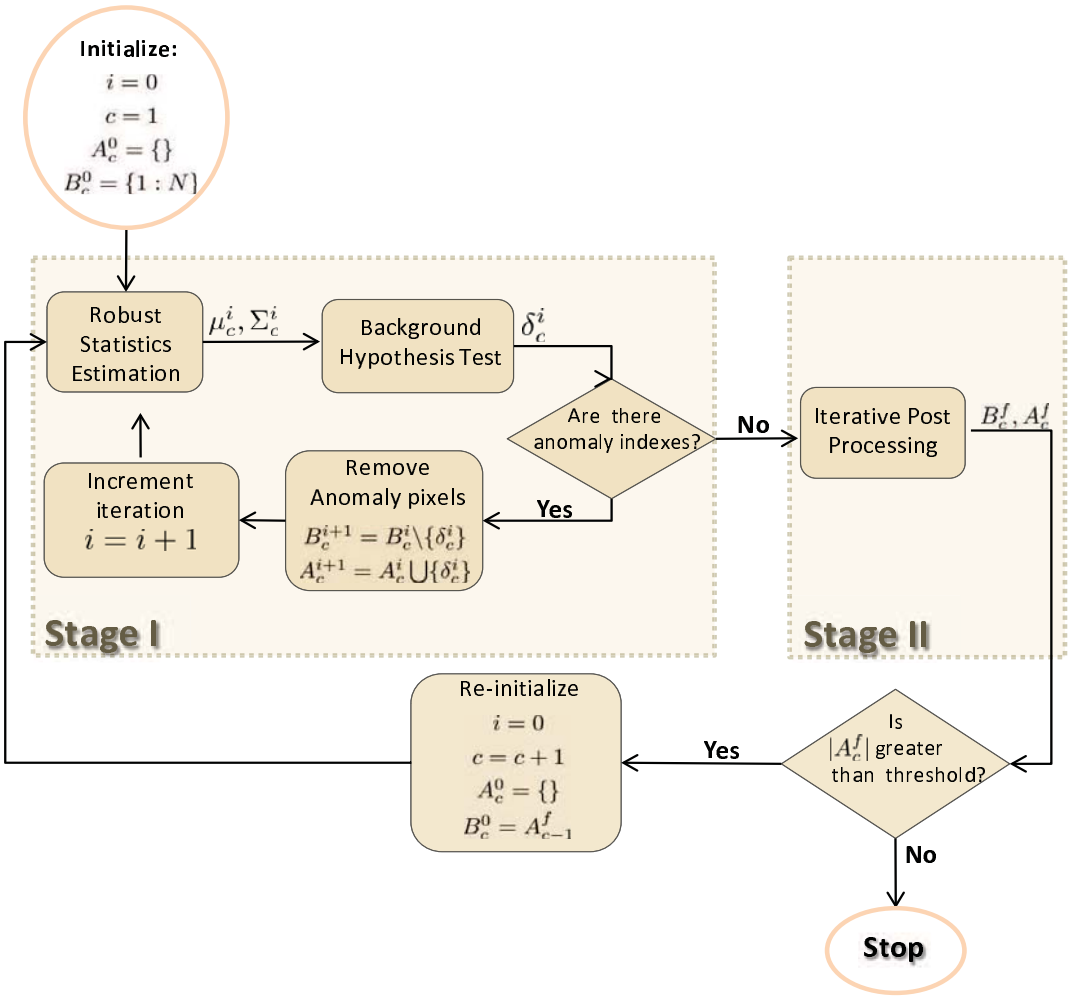

Figure 3.2: **BEVA's Local Part - Block Diagram**

3.2.3 Interrelating local background models

The proposed greedy algorithm for local background estimation allows the use of several Gaussian pdfs to accurately represent the wealth of the local background spectrum. However, the many degrees of freedom in the parameter selection can lead to a high false-alarms rate stemming from overfitting.

We propose here a global filtering approach, which reduces the number of

degrees of freedom by inter-relating the obtained local background pdfs. Given the anomaly subset A of a local block, we define a larger image area composed of T blocks around it. All the local anomaly pixel in A are compared to the T relevant backgrounds (each modeled by up to 3 clusters). For each anomaly pixel $x \in A$, we find the minimum Mahalanobis distance from it to all the clusters of the T backgrounds, $d = min_{i,j}\{(x - \boldsymbol{\mu}_{i,j})^T(\boldsymbol{\Sigma}_{i,j})^{-1}(x - \boldsymbol{\mu}_{i,j})\}$ where $i = 1, \cdots, T$ and j runs over the number of clusters in block i. If the minimum Mahalanobis distance is smaller than the background hypothesis threshold value of the cluster where the minimum was found , the local anomaly pixel will be removed from the anomaly set. At the end, A is composed just of global anomalies.

Table 3.3 presents a formal description of the global filter.

Task: Reduce the number of false alarms in a local block anomaly subset A

Loop: For $i = 1 : T$

1. *Background hypothesis testing on a local block i containing L_i background clusters,* $\forall x_m \in A$, $j = 1, \cdots, L_i$

$$d_m = min_j\{(x_m - \boldsymbol{\mu}_{i,j})^T(\boldsymbol{\Sigma}_{i,j})^{-1}(x_m - \boldsymbol{\mu}_{i,j})\}$$

2. *Update sets:*

Find data pixel indices having Mahalanobis distances that are below the background cluster hypothesis threshold value τ_i :

$$\{\delta_i\} = I(d_m \leq \tau_i)$$
$$A = A \backslash \{\delta_i\}$$

Output: A is composed just of global anomalies

Table 3.3: Global Filter

3.3 Experiments with Real Data

In this section we evaluate the performance of the proposed algorithm by applying it to real hyperspectral data (For a description of the hyperspectral data used, see Appendix A). We compare the BEVA algorithm to RX [4], GMM-RX [11] and MCD [14], in terms of Receiver Operation Characteristic (ROC) curves. For the experiment, the image was divided into non-overlapping local areas of size 35×35 . The global filter was applied using a large image area of size 525×300. The ROC curves are presented in Fig. 3.3 and 3.4

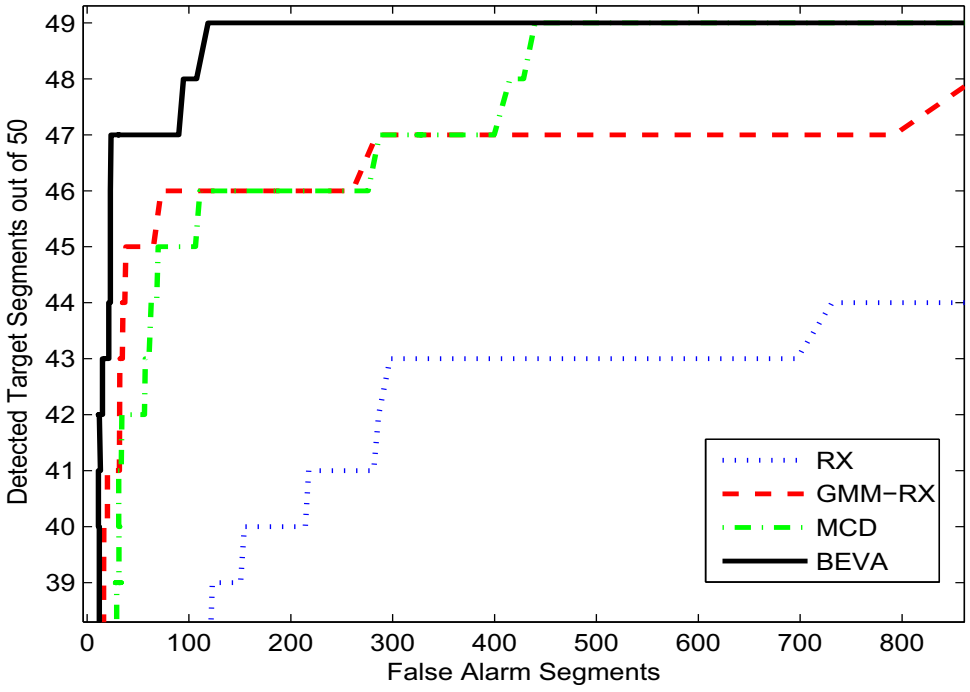

Figure 3.3: **ROC curve BEVA vs. RX, GMM-RX and MCD - Data Set I.**

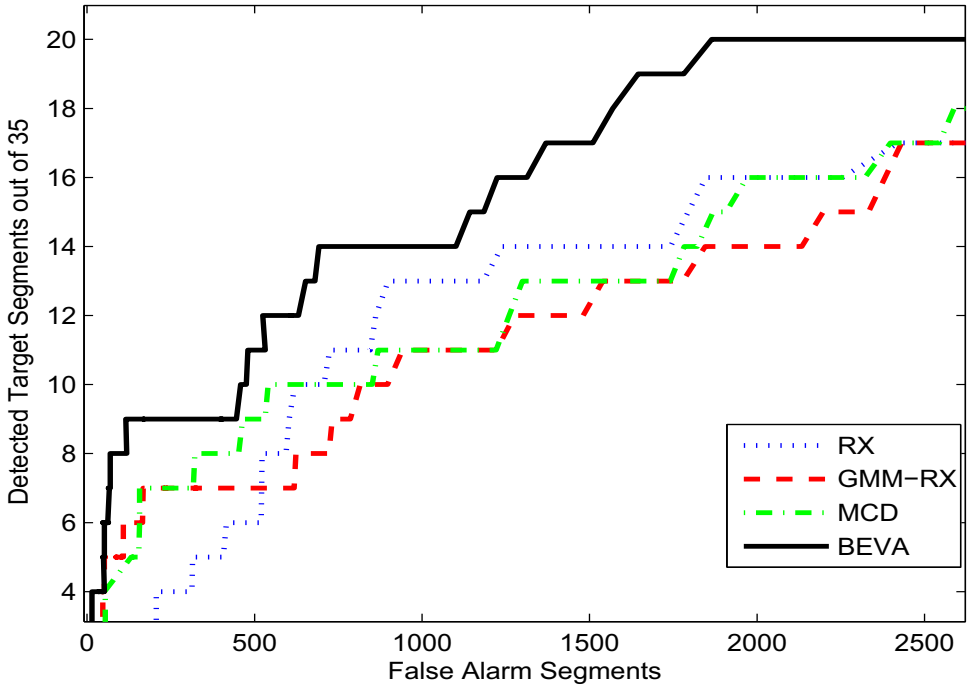

Figure 3.4: **ROC curve BEVA vs. RX, GMM-RX and MCD - Data Set II.**

For these two data sets of real hyperspectral images, the BEVA algorithm is seen to have a better performance than the other tested algorithms.

3.4 Summary

In this chapter we presented the BEVA algorithm for anomaly detection that is based on a local-global statistical background modeling approach. It significantly reduces the vast number of degrees of freedom of the local method, while retaining the ability of local models to be intimately adjusted to the background. In the local part, the local background is approximated using a greedy sequential estimation process that applies a robust estimation of the Gaussian statistics and

background cluster hypothesis testing based on Extreme Value Theory results. Then, in the global part, the obtained local background models are inter-related to reduce the number of false alarms. In experiments with real hyperspectral image cubes, BEVA was found to have a better performance than RX [4], GMM-RX [11] and MCD-based algorithm [14].

In the next chapters, we propose several improvements of both local and global parts of BEVA.

Chapter 4

Global Auxiliary Dictionary

In this chapter, we propose to reinforce BEVA's global filter performance by adding an auxiliary background dictionary.

4.1 BEVA Global Filter Drawback

BEVA combines a local anomaly detector with a global filter. In the global filter, for each pixel x of a local block, we define a larger image area composed of T blocks around it. x is then compared to the T relevant block backgrounds by computing the Mahalanobis distances $d_{i,j} = (x - \boldsymbol{\mu}_{i,j})^T (\boldsymbol{\Sigma}_{i,j})^{-1} (x - \boldsymbol{\mu}_{i,j})$ where $i = 1, \cdots, T$, and j runs over the number of clusters in block i (up to L_i). The minimum Mahalanobis distance over i and j is the final pixel score.

By inter-relating local blocks, we have obtained a global background model composed of several local background clusters. In other words, pixels are compared to a "dictionary" where each "word" is an estimated local background cluster. A pixel is declared as an anomaly if it does not fit well any "word" of the "dictionary".

However, this "dictionary" suffers from a drawback: it is based only on local

background clusters. Spatially dispersed background pixels like rocks in a vegetation area, wrongly detected in BEVA local part as anomalies, will not be removed by the global filter since they cannot be associated with any local background cluster of the larger image area.

4.2 Auxiliary Dictionary using a GMM Algorithm

We overcome the above drawback by adding an auxiliary background dictionary composed of global background clusters estimated using a global spectral clustering.

In the global clustering, the background process of the whole image is modeled by several clusters, each one distributed as a Gaussian distribution. The global clustering is initialized by an excessive number of Gaussians using the k-means algorithm [23] for initializing the Gaussian parameters. Then, we applied the Expectation-Maximization (EM) [25] algorithm on the initial partition (See Appendix B).

The EM algorithm estimates the Gaussian statistics $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ of each cluster k. As an improvement of the EM algorithm, we also remove too small clusters, and hence unreliable, during the EM iterations. The algorithm converges then to a local maximum with a smaller number of clusters, P, than in the initialization step. Each cluster is represented by a Gaussian distribution:

$$\begin{cases} C_k^{EM}, \ \ 1 \leq k \leq P \\ C_k^{EM} \sim N(\boldsymbol{\mu}_k^{EM}, \boldsymbol{\Sigma}_k^{EM}) \end{cases} \tag{4.1}$$

This dictionary of global background clusters is used in the global filter of BEVA. The tested pixel $\mathbf{x}$ is compared to the T relevant block backgrounds and also to the P clusters obtained from the global spectral clustering. The final pixel

score d is then given by the minimum distance over all the clusters:

$$d = min_{i,j,k}[\{(\mathbf{x} - \boldsymbol{\mu}_{i,j})^T(\boldsymbol{\Sigma}_{i,j})^{-1}(\mathbf{x} - \boldsymbol{\mu}_{i,j})\} \cup \{(\mathbf{x} - \boldsymbol{\mu}_k^{EM})^T(\boldsymbol{\Sigma}_k^{EM})^{-1}(\mathbf{x} - \boldsymbol{\mu}_k^{EM})\}] \tag{4.2}$$

where $k = 1, \cdots, P$, $i = 1, \cdots, T$ and j runs over the number of clusters in block i (up to L_i)

4.3 Experiments with Real Data

The improved algorithm was applied to the same two data sets (see Appendix A) and compared to BEVA performance. The auxiliary dictionary was initialized to 10 clusters for each image of the data sets. The ROC curves are shown in Fig. 4.1 and 4.2.

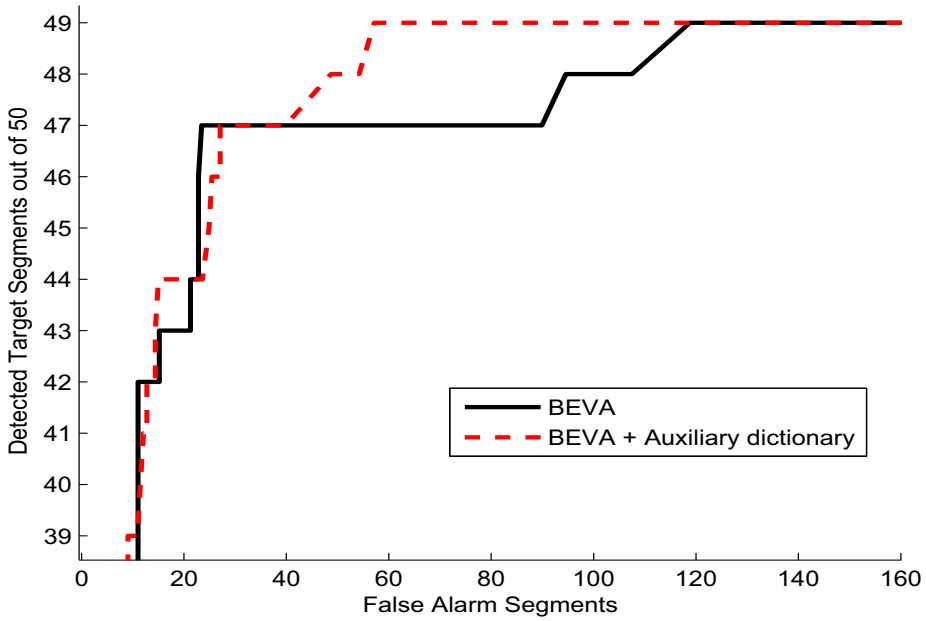

Figure 4.1: **ROC curve BEVA with the auxiliary dictionary vs. BEVA - Data Set I.**

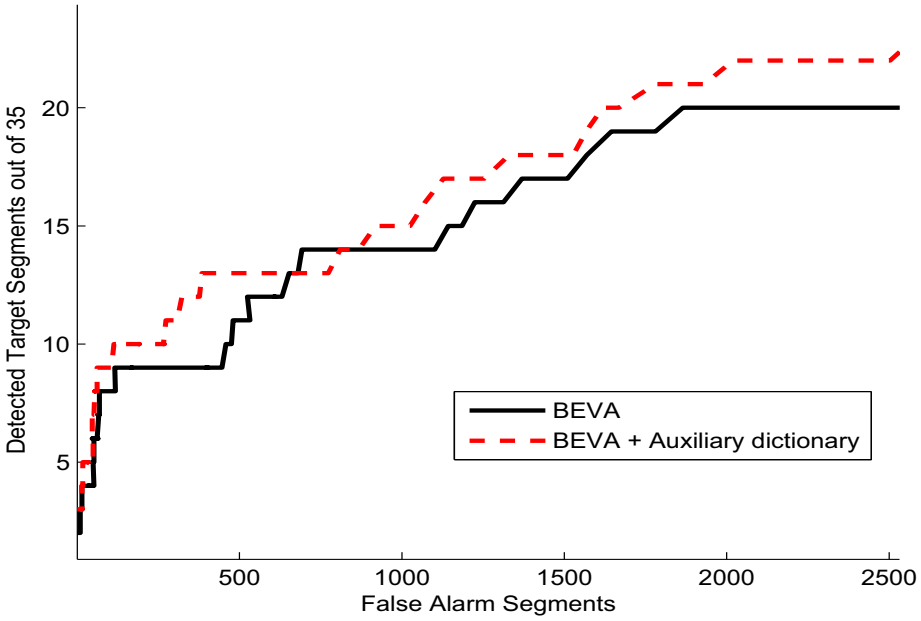

Figure 4.2: **ROC curve BEVA with the auxiliary dictionary vs. BEVA - Data Set II.**

Since there are almost no spatially dispersed background pixels wrongly detected by BEVA in data set I, the two ROC curves are close. However, the addition of the auxiliary dictionary improves BEVA's performance for the second data set of hyperspectral images where a lot of spatially dispersed background pixels were wrongly detected as anomalies.

4.4 Summary

In this chapter, we proposed an improved version of BEVA: we add an auxiliary background dictionary composed of global background clusters estimated using a GMM clustering method. Since this clustering is not influenced by the spatial configuration of the pixels, it helps to reduce false alarms on spatially dispersed

background pixels which was a source of false alarms in BEVA.

Another improvement of BEVA, denoted "Spectral BEVA" is introduced in the next chapter.

Chapter 5

Spectral Clustering for Anomaly Detection

In this chapter we introduce Spectral BEVA, an improvement of the local part of BEVA algorithm.

5.1 BEVA's Local Part Drawback

In BEVA's local part, the hyperspectral image is partitioned into distinct local blocks in which the background process is modeled by a small number of clusters, L (up to 3), ordered by size, each distributed as a separate Gaussian distribution:

$$\begin{cases} x \in C_k, \;\; 1 \leq k \leq L \\ C_k \sim N(\boldsymbol{\mu}_k, \Gamma_k) \\ |C_1| \geq |C_2| \geq \cdots \geq |C_L| \end{cases} \tag{5.1}$$

The cluster parameters are estimated using a greedy sequential estimation process that applies robust estimation of the Gaussian statistics and background cluster hypothesis testing based on Extreme Value Theory results. Starting with the dominant cluster estimation, this process is repeated until all the background clusters are estimated.

In BEVA's local part, segmentation of the background clusters and their estimation are combined together in one process that uses the background hypothesis test develop in section 3.2.1 to determine if a pixel belongs or not to a background cluster. The background hypothesis test, based on Extreme Value Theory, is not sensitive enough to give a correct segmentation. In our simulations with real hyperspectral data, it can be seen that most of the local blocks are segmented into just one cluster, which by viewing the images is clearly an under-estimation of the real number of background clusters.

In the next section, we propose an algorithm, denoted Spectral BEVA, which overcomes this weakness. In Spectral BEVA, each local block is first segmented using a Spectral Clustering algorithm ([27], [29]). Then, each segment is considered as a background cluster and its parameters are estimated using a simplified version of the estimation process in BEVA.

5.2 Spectral Clustering

In recent years, Spectral Clustering (SC) has become one of the most popular modern clustering algorithms ([27], [28]). It is simple to implement, can be solved efficiently by standard linear algebra software, and very often outperforms traditional clustering algorithms such as the k-means algorithm [23]. In this section, we present some graph theory concepts that are the basis of SC methods. Then, we describe in detail the SC algorithm that is used in Spectral BEVA [29] and show some segmentation results on real hyperspectral images.

5.2.1 Graph Theory

Given a set of data points $x_1, \cdots, x_N$ and some notion of similarity $w_{ij} \geq 0$ between all pairs of data points x_i and x_j , the intuitive goal of clustering is to partition the data points into several groups such that points in the same group are similar and points in different groups are dissimilar to each other. If we do not have more information than similarities between data points, a nice way of representing the data is in form of the similarity graph $G = (V, E)$. Each vertex $v_i \in V$ in this graph represents a data point x_i. Two vertices are connected with an edge $e_{ij} \in E$, weighted by the similarity w_{ij} between the corresponding data points x_i and x_j. Large weights mean that the adjacent vertices are very similar; small weights imply dissimilarity. In Fig.5.1, an example of a graph with 6 vertices and the similarity between them. The similarities not represented in this graph are assumed to be equal to zero.

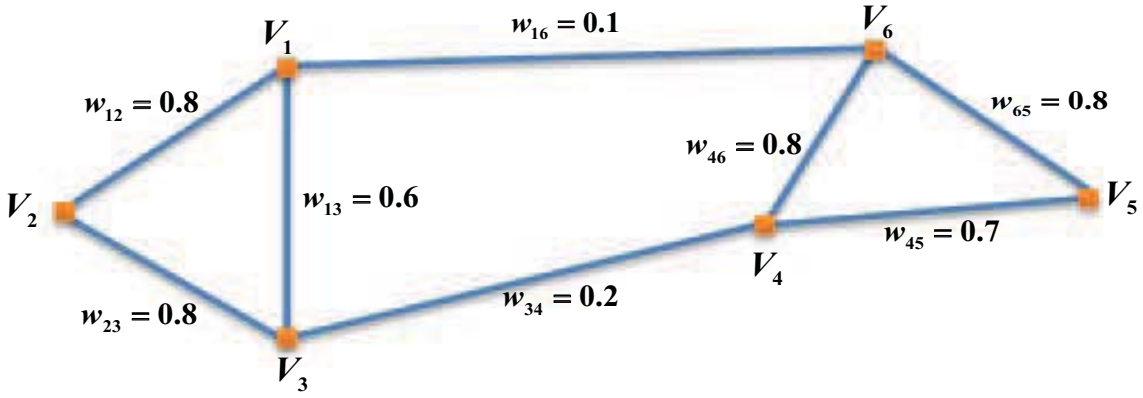

Figure 5.1: **Data Representation using Graph Theory**

The problem of clustering can now be reformulated using the similarity graph: we want to find a partition of the graph such that the edges between different groups have very low weights and the edges within a group have high weights. In other words, given a similarity graph with affinity matrix $W = \{w_{i,j}\}_1^N$, a way to construct a partition $\{A_1, \cdots, A_C\}$ of the graph is to solve the Normalized Cut [26] (Ncut) problem. The Ncut is defined as:

$$Ncut(A_1, \cdots, A_C) = \sum_{i=1}^{C} \frac{cut(A_i, \bar{A}_i)}{vol(A_i)} \tag{5.2}$$

where $\bar{A}$ is the complement of A

The partition $\mathbf{A}^* = \{A_1^*, \cdots, A_C^*\}$ that solves the Ncut problem is given by:

$$\mathbf{A}^* = min_{A_1, \cdots, A_C} Ncut \tag{5.3}$$

Here, the cut between two partitions A and B is defined by

$$cut(A, B) = \sum_{i \in A, j \in B} w_{i,j} \tag{5.4}$$

and the volume of a graph subset A by

$$vol(A) = \sum_{i \in A} \sum_{j=1}^{N} w_{ij} \tag{5.5}$$

The Ncut tries to achieve a minimal cut together with "balanced" clusters, as measured by the number of vertices weights. Unfortunately, the Ncut problem is "NP-Hard" and hence intractable. Spectral Clustering used in Spectral BEVA is based on a standard relaxation procedure that transforms the Ncut problem into a tractable eigenvector problem [27].

5.2.2 Spectral Clustering Algorithm

Given N data points $x_1, \cdots, x_N$, the spectral clustering (SC) algorithm [27] constructs an affinity matrix $W \in \mathbb{R}^{N \times N}$, where

$$w_{ij} = exp(-\frac{||x_i - x_j||^2}{\sigma^2}) \tag{5.6}$$

reflects the relationships between x_i and x_j, with σ as a user-defined parameter. The similarity between x_i and itself is assumed to be 0. SC uses the similarity information to group $x_1, \cdots, x_N$ into C clusters.

The algorithm is described in detail in Table 5.1

Inputs: $\{x_m\}_1^N$, $x_m : p \times 1$, p - # of spectral bands,
Number of clusters C , σ - Scaling parameter

Algorithm: Perform the following steps:

1. *Compute the Affinity Matrix* $W \in \mathbb{R}^{N \times N}$ *where*

$$w_{ij} = exp(-\frac{||x_i - x_j||^2}{\sigma^2})$$

2. *Compute the Degree Matrix* D *to be the diagonal matrix with:*

$$D_{ii} = \sum_{j=1}^{N} w_{ij}$$

3. *Compute the Normalized Laplacian Affinity Matrix* L*:*

$$L = D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$$

4. *Solve the following eigenvalue problem:*

$$Lv = \lambda v$$

5. *Find the eigenvectors corresponding to the* C *largest eigenvalues of* L*:*

$$V = \{v_1, v_2, \cdots, v_C\} \in \mathbb{R}^{N \times C}$$

6. *Normalize the rows of* V*:*

$$\widetilde{V}_{ij} = V_{ij} / \sqrt{\sum_j V_{ij}^2}$$

7. *Treat each row of* $\widetilde{V}$ *as a point in* $\mathbb{R}^C$ *and cluster via k-means*

8. *Assign the original point* x_i *to cluster* c *if and only if the corresponding row* i *of the matrix* $\widetilde{V}$ *was assigned to cluster* c

Output: Each data point x_i belongs to one of the C clusters

Table 5.1: Spectral Clustering Algorithm [27]

In Table 5.1, the scaling parameter σ is set a priori by the user. σ is some measure of when two points are considered similar, and it can be difficult to tune. Moreover, when the input data includes clusters with different local statistics there may not be a single value of σ that works well for all the data. In the next subsection, we introduce an automatic local scaling parameter as proposed in [29].

5.2.3 Local Scaling

Zelnik-Manor and Perona proposed in [29] to calculate a local scaling parameter σ_i for each data point x_i instead of selecting a single scaling parameter σ. Using a specific scaling parameter for each point allows self-tuning of the point-to-point distances according to the local statistics of the neighborhoods surrounding points i and j. The affinity between two points is now given by

$$w_{ij} = exp(-\frac{||x_i - x_j||^2}{\sigma_i \sigma_j}) \tag{5.7}$$

The selection of the local scaling parameter σ_i is done by studying the local statistics of the neighborhood of point x_i:

$$\sigma_i = d(x_i, x_K) \tag{5.8}$$

i.e., the local scaling parameter σ_i is the distance from x_i to its K'th nearest neighbor x_K. The choice of K is data dependent. A too small K will not represent well the local statistics of the neighborhood of x_i. On the other hand, segmentation using a too large K would be affected by the presence of outliers (anomalies). For our data, K between 10 to 50 gives similar segmentation results.

Fig. 5.2, provides a visualization of the effect of the local scaling.

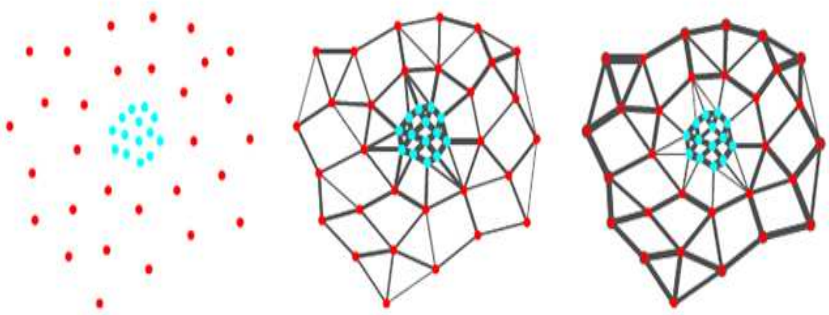

Figure 5.2: **Local Scaling Effect (from [29])**

In Fig. 5.2, a cluster resides within a bigger cluster. The affinity between each point and its surrounding neighbors is indicated by the thickness of the line connecting them. In the center image, the affinities across clusters are larger than the affinities within the background cluster leading to a wrong segmentation. In the right image, the local scaling was applied. The affinities across clusters are now significantly lower than the affinities within any single cluster, which should improve the segmentation results.

5.2.4 Segmentation Results for Real Hyperspectral Data

The Spectral Clustering algorithm was applied to real hyperspectral data. The results with a global scaling parameter $\sigma = 0.5$ and using local scaling parameters with $K = 20$ are shown in Fig. 5.3. The number of clusters is set to 3. As we can see, the segmentation using local scaling parameters is more accurate.

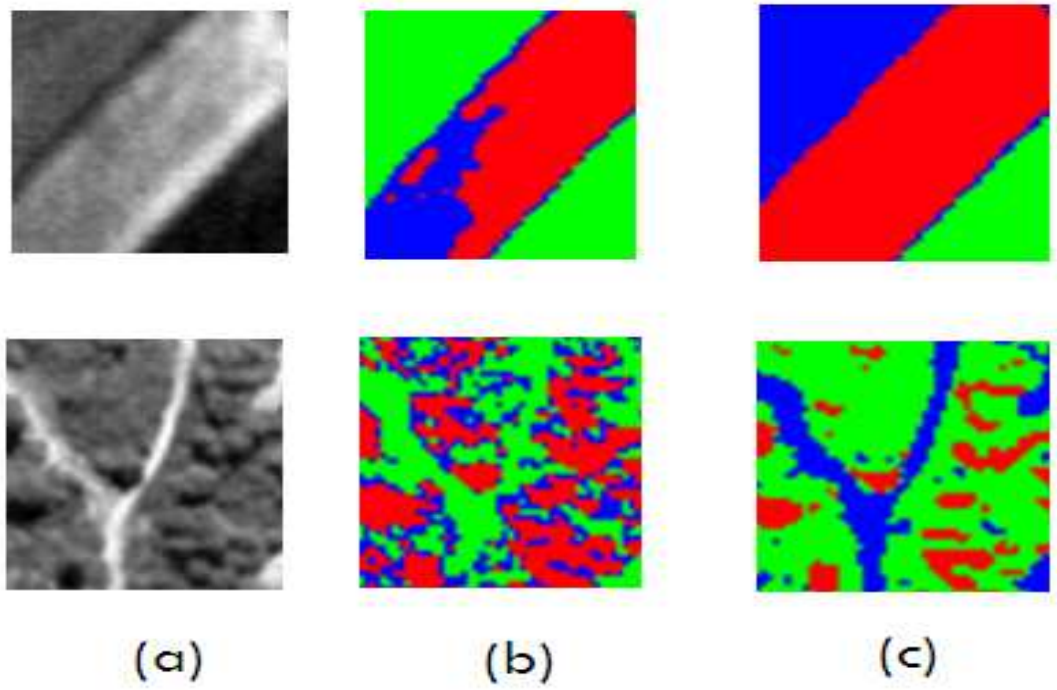

Figure 5.3: **Spectral Clustering Result - Global vs. Local Scaling Parameter.** (a) - Single hyperspectral band. (b) - Spectral Clustering Result with Global Scaling Parameter. (c) Spectral Clustering with Local Scaling Parameter

5.3 Anomaly Detection using Spectral BEVA

In Spectral BEVA, the local block is first segmented into C clusters using the Spectral Clustering algorithm proposed in the previous section. The number of clusters C is a user-defined parameter. In our experiments, C is set to 3 for a [35×35] local block. Then, each segment is considered as a background cluster and its parameters are estimated using a simplified version of the BEVA estimation process detailed in Table 5.2. In order to estimate the statistics of a cluster $c \in C$, we initialize two distinct indices sets $A_c^0 = \emptyset$ and $B_c^0 = \{m | x_m \in c\}$ and aim to get

$$\begin{aligned} B_c^f &= \text{Pixel indices of background cluster } c \\ A_c^f &= \text{Pixel indices of anomalies} \end{aligned} \tag{5.9}$$

From B_c^f, we also estimate the statistics $\boldsymbol{\mu}_c^f$ and $\boldsymbol{\Sigma}_c^f$ of cluster c.

Inputs: $In = \{m|x_m \in c\}$, $c \in C$, $x_m : [p \times 1]$, p - # of spectral bands

Initialization: $A_c^0 = \emptyset,\ B_c^0 = In,\ \omega^0 = 1,$
$d_0 = (\sqrt{p} + \sqrt{2})^2,\ \gamma = 1.25$ and $i = 0$

Main Iteration: Perform the following steps:

1. *Robust estimation of Mean and Covariance* [22]

$$\boldsymbol{\mu}_c^i = \frac{\sum_{x_m \in B_c^i} \omega_m^i x_m}{\sum_{x_m \in B_c^i} \omega_m^i}$$
$$\boldsymbol{\Sigma}_c^i = \frac{\sum_{x_m \in B_c^i} (\omega_m^i)^2 (x_m - \boldsymbol{\mu}_c^i)^T (x_m - \boldsymbol{\mu}_c^i)}{\sum_{x_m \in B_c^i} (\omega_m^i)^2 - 1}$$

2. *Calculation of Mahalanobis distances in* B_c^i *:*

$$\forall x_m \in B_c^i :\ \ d_m = (x_m - \boldsymbol{\mu}_c^i)^T (\boldsymbol{\Sigma}_c^i)^{-1} (x_m - \boldsymbol{\mu}_c^i)$$

3. *Update weights* [22]:

$$\omega_m^{i+1} = \begin{cases} 1 & if\ \ d_m \le d_0 \\ \frac{d_0}{d_m} \exp(-0.5(d_m - d_0)^2/\gamma^2) & if\ \ d_m \ge d_0 \end{cases}$$

4. *Update sets:*

Find data pixel indices $\{\delta^i\}$ having Mahalanobis distances that exceed the background cluster hypothesis threshold value τ^i which is a function of $p, |B_c^i|$ and $\boldsymbol{\Sigma}_c^i$, as explained in Appendix C:

$$\{\delta^i\} = \{I(d_m \ge \tau^i)\}\ \ (I = Index\ of\ value)$$
$$B_c^{i+1} = B_c^i \backslash \{\delta^i\}$$
$$A_c^{i+1} = A_c^i \bigcup \{\delta^i\}$$

5. *Stopping rule:*

If $\{\delta^i\} = \emptyset$, stop. Otherwise increment i and go to 1

Output: $A_c^f = A_c^i,\ B_c^f = B_c^i,\ \boldsymbol{\mu}_c^f = \boldsymbol{\mu}_c^i$ *and* $\boldsymbol{\Sigma}_c^f = \boldsymbol{\Sigma}_c^i$

Table 5.2: Spectral BEVA - Background Parameter Estimation

The block diagram of Spectral BEVA's local part is shown in Fig. 5.4:

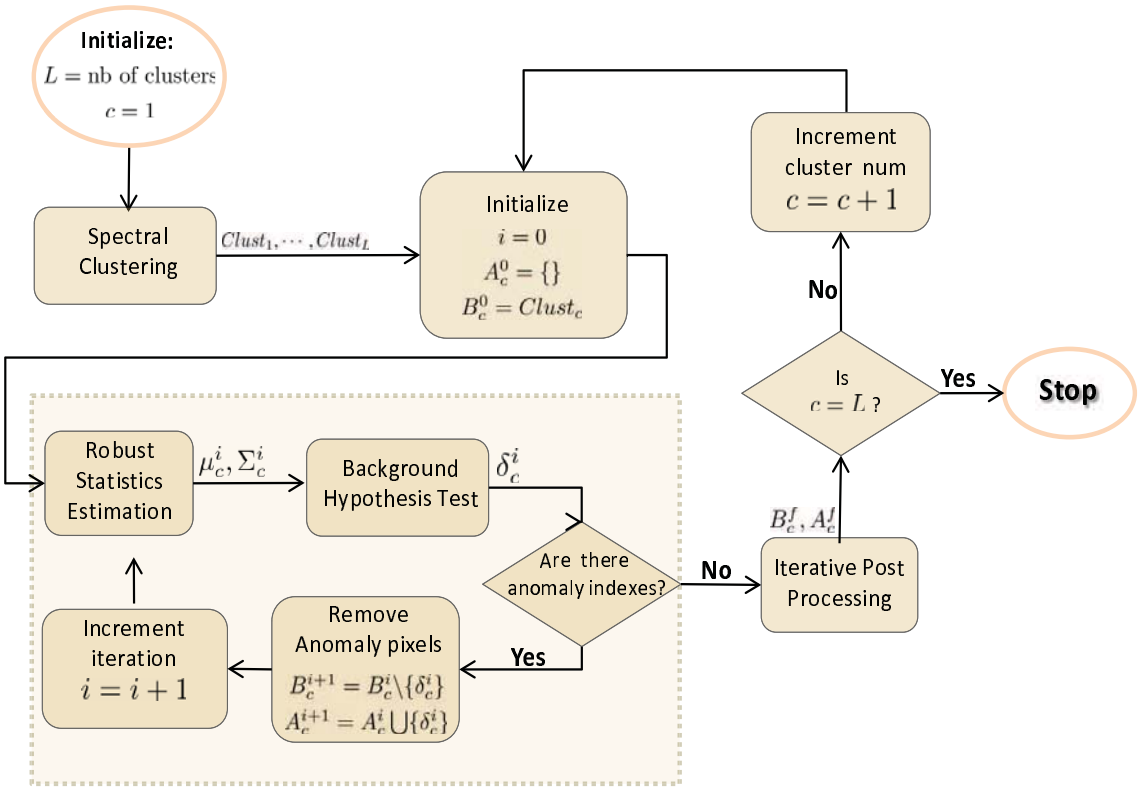

Figure 5.4: **Spectral BEVA's Local Part - Block Diagram**

Anomalies are then detected using a global filter identical to BEVA with the auxiliary dictionary. The tested pixel $\mathbf{x}$ is compared to T relevant block backgrounds and also to the P clusters of the auxiliary dictionary obtained from the global segmentation. The final pixel score d is then given by the minimum distance over all the clusters:

$$d = min_{i,j,k}[\{(\mathbf{x} - \boldsymbol{\mu}_{i,j})^T(\boldsymbol{\Sigma}_{i,j})^{-1}(\mathbf{x} - \boldsymbol{\mu}_{i,j})\} \cup \{(\mathbf{x} - \boldsymbol{\mu}_k^{EM})^T(\boldsymbol{\Sigma}_k^{EM})^{-1}(\mathbf{x} - \boldsymbol{\mu}_k^{EM})\}] \tag{5.10}$$

where $k = 1, \cdots, P$, $i = 1, \cdots, T$ and j runs over the C clusters in block i.

5.4 Experiments with Real Data

In this section we evaluate the performance of Spectral BEVA by applying it to real hyperspectral data (For a description of the hyperspectral data used, see Appendix A). We compare it to the BEVA algorithm, in terms of Receiver Operation Characteristic (ROC) curves. For the experiment, the image was divided into non-overlapping local areas of size 35×35 . The global filter was applied using a large image area of size 525×300. Each local area was segmented into 3 clusters with a local scaling parameter K set to 20. An auxiliary dictionary, initialized to 10 clusters for each image of the data sets, was added to the two algorithms. The ROC curves are presented in Fig. 5.5 and 5.6

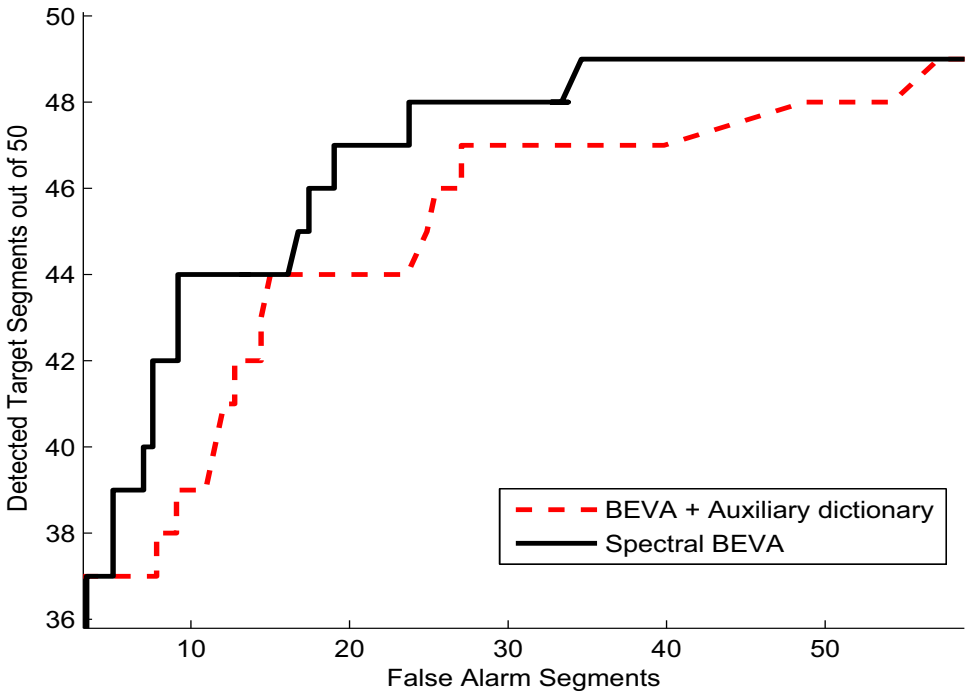

Figure 5.5: **ROC curve Spectral BEVA vs. BEVA - Data Set I.**

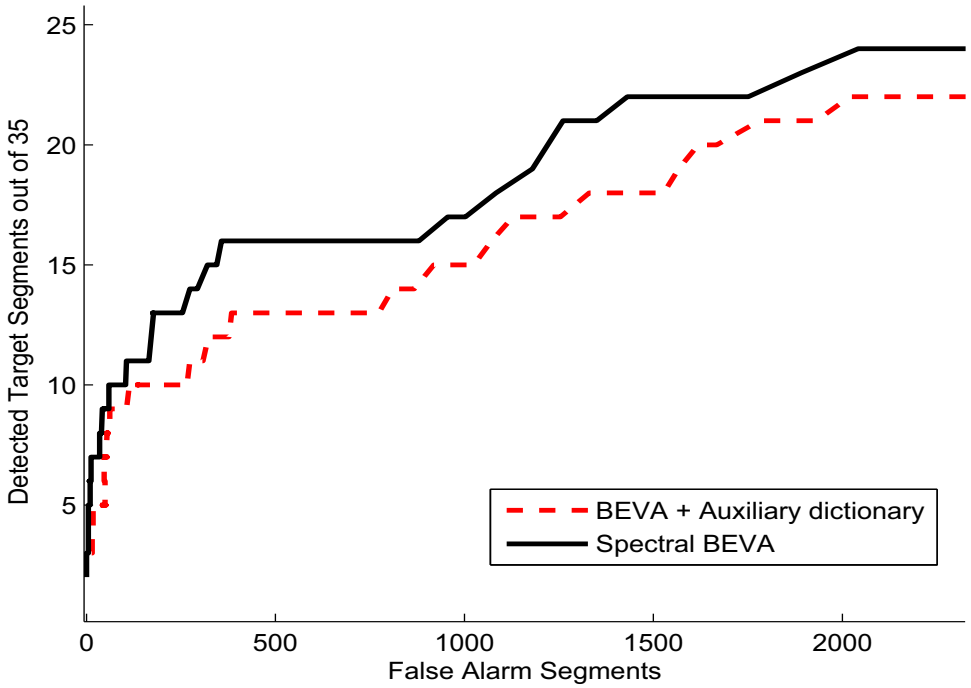

Figure 5.6: **ROC curve Spectral BEVA vs. BEVA - Data Set II.**

For these two data sets of real hyperspectral images, the Spectra BEVA algorithm is seen to have a better performance than the BEVA algorithm for all probability of detection values.

5.5 Summary

In this chapter, we proposed Spectral BEVA, an improvement of the BEVA algorithm. In the local part of Spectral BEVA, the local background is approximated using a Spectral Clustering algorithm [27] together with a greedy sequential estimation process based on robust estimation of the Gaussian statistics and Extreme Value Theory results. The global part is the same as in the BEVA algorithm: a filter using large image area statistics is applied to reduce the number of false

alarms.

Despite of its good performance, Spectral BEVA suffers from a drawback: The background process is still approximated with linear Gaussian distributions. In the next chapter, we propose to overcome this drawback by proposing a non-linear local-global based algorithm where no Gaussian assumption is made.

Chapter 6

Non-Gaussian Background Modeling for Anomaly Detection

6.1 The Gaussian Assumption in BEVA

In the different variations of the BEVA algorithm presented in the previous chapters, the local background model is composed of C distinct clusters, each approximated by a separate Gaussian distribution. During the local background estimation process, the background clusters pixels are discriminated from other pixels. For this purpose, we developed in section 3.2.1 a background hypothesis test that determines which pixels are belonging to a background cluster. It is based on examining the Mahalanobis distance $d = (x - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x - \boldsymbol{\mu})$ of a realization x to the mean of a background cluster, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and covariance of a Gaussian Gaussian $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ that approximates this background cluster. Assuming this Gaussian model, the Mahalanobis distances have a Chi-squared distribution of order p equal to the number of spectral bands.

Two conditional hypotheses are then obtained using Extreme Value Theory on a Chi-squared distribution (see appendix C). The crossing point τ of the two hypotheses can be used as a threshold to discriminate the background cluster

realizations from other realizations in the local block.

The choice of the Gaussian model is due to its efficient processing and mathematically tractability. In fact, it simplifies the derivation of decision rules and the evaluation of the detectors' performance. Unfortunately, the Gaussian model is not sufficiently adequate to represent the statistical behavior of a background cluster in real hyperspectral images. It has been shown ([31], [32], [33]) that the Gaussian model fails in its representation of the distribution tails. In particular, distributions of hyperspectral data have longer tails than the Gaussian pdf. This is a critical issue in BEVA. Indeed, anomalies are isolated using the Extreme Value Theory results. Since the Gaussian model underestimates the distribution tails, it can lead to an excess number of false alarms.

To overcome the limits of the Gaussian model, we propose in the next section a non-Gaussian version of the BEVA algorithm, named NG-BEVA (Non Gaussian Background Extreme Value Analysis).

6.2 Gamma Fitting for Background Modeling

In the BEVA algorithm, each background cluster is approximate by a unimodal Gaussian pdf $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ being the mean and covariance, respectively. The Mahalanobis distances to a background cluster in BEVA,

$$d_i = (x_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x_i - \boldsymbol{\mu}) \qquad \text{with} \quad i = 1, \ldots, N \tag{6.1}$$

have then a Chi-squared distribution of order p equal to the number of spectral bands, denoted by $\chi^2(p)$,

$$f(u) = \frac{1}{2^{p/2}\Gamma(p/2)} u^{(p/2)-1} e^{-u/2} \qquad \text{with} \quad u \geq 0 \tag{6.2}$$

where Γ is the Gamma function.

In the proposed NG-BEVA, we assume that the Mahalanobis distances have a Gamma distribution instead of a Chi-squared distribution. The Gamma distribution $\mathbf{\Gamma}(k, \Theta)$ is a two-parameter distribution. Its pdf has the following form:

$$f(u) = \frac{1}{\Theta^k \Gamma(k)} u^{k-1} e^{-u/\Theta} \qquad \text{with} \quad u \geq 0 \tag{6.3}$$

where Θ is the scale parameter and k the shape parameter.

The Gamma distribution is more general than the Chi-squared distribution. In fact, the Chi-squared distribution $\chi^2(p)$ is a special case of the Gamma distribution $\mathbf{\Gamma}(k, \Theta)$, obtained for $k = \frac{p}{2}$ and $\Theta = 2$.

By using the Gamma distribution, we have relaxed the Gaussian model constraint. However, the parameters k and Θ have to be estimated. The maximum likelihood estimators $\hat{k}$ and $\hat{\Theta}$ are given in appendix C.

As in BEVA (see section 3.2.1), the conditional hypotheses probabilities are given by:

$$p(H_0|\eta) = \frac{\eta f_v(\eta)}{\eta f_v(\eta) + F_v(\eta)}, \tag{6.4}$$

$$p(H_1|\eta) = \frac{F_v(\eta)}{\eta f_v(\eta) + F_v(\eta)}, \tag{6.5}$$

However, f_v and F_v are now the pdf and cdf, respectively, of the Gamma distribution with the estimated max-likelihood parameters $\hat{k}$ and $\hat{\Theta}$.

The two stages of the estimation process are described in more details in Table 6.1 and Table 6.2.

Inputs: $\{x_m\}_1^N$, p - # of spectral bands

Initialization: $A^0 = \emptyset,\ B^0 = \{1 : N\},\ \omega^0 = 1,$
$d_0 = (\sqrt{p} + \sqrt{2})^2,\ \gamma = 1.25$ and $i = 0$

Main Iteration: Perform the following steps:

1. *Robust estimation of Mean and Covariance* [22]

$$\boldsymbol{\mu}^i = \frac{\sum_{x_m \in B^i} \omega_m^i x_m}{\sum_{x_m \in B^i} \omega_m^i}$$
$$\boldsymbol{\Sigma}^i = \frac{\sum_{x_m \in B^i} (\omega_m^i)^2 (x_m - \boldsymbol{\mu}^i)^T (x_m - \mu^i)}{\sum_{x_m \in B^i} (\omega_m^i)^2 - 1}$$

2. *Calculation of Mahalanobis distances in* B^i *:*

$$\forall x_m \in B^i :\ d_m = (x_m - \boldsymbol{\mu}^i)^T (\boldsymbol{\Sigma}^i)^{-1} (x_m - \boldsymbol{\mu}^i)$$

3. *Update weights* [22]

$$\omega_m^{i+1} = \begin{cases} 1 & if\ \ d_m \le d_0 \\ \frac{d_0}{d_m} \exp(-0.5(d_m - d_0)^2/\gamma^2) & if\ \ d_m \ge d_0 \end{cases}$$

4. *Gamma Distribution Fitting using Maximum Likelihood Estimation* (Appendix E)

$$\forall x_m \in B^i : d_m \to \boldsymbol{\Gamma}(\hat{k}^i, \hat{\Theta}^i)$$

5. *Update sets:*

Find data pixel indices $\{\delta^i\}$ having Mahalanobis distances that exceed the background cluster hypothesis threshold value τ^i which is a function of $p, |B^i|, \boldsymbol{\Sigma}^i, \hat{k}^i$ and $\hat{\Theta}^i$

$$\{\delta^i\} = \{I(d_m \ge \tau^i)\}\ \ (I = Index\ of\ value)$$
$$B^{i+1} = B^i \backslash \{\delta^i\}$$
$$A^{i+1} = A^i \bigcup \{\delta^i\}$$

6. *Stopping rule:*

If $\{\delta^i\} = \emptyset$, stop. Otherwise increment i and go to 1

Output: $A^{tmp} = A^i, B^{tmp} = B^i, \boldsymbol{\mu}^{tmp} = \boldsymbol{\mu}^i, \boldsymbol{\Sigma}^{tmp} = \boldsymbol{\Sigma}^i, \hat{k}^{tmp} = \hat{k}^i\ \ and\ \hat{\Theta}^{tmp} = \hat{\Theta}^i$

Table 6.1: NG-BEVA - First Stage

Initialization: $A^0 = A^{tmp}$, $B^0 = B^{tmp}$, $\boldsymbol{\mu}^0 = \boldsymbol{\mu}^{tmp}$, $\boldsymbol{\Sigma}^0 = \boldsymbol{\Sigma}^{tmp}$, $\hat{k}^0 = \hat{k}^{tmp}$, $\hat{\Theta}^0 = \hat{\Theta}^{tmp}$ *and* $j = 0$

Main Iteration: Perform the following steps:

1. *Calculation of Mahalanobis distances in* A^j

$$\forall x_m \in A^j : \; d_m = (x_m - \boldsymbol{\mu}^j)^T (\boldsymbol{\Sigma}^j)^{-1} (x_m - \boldsymbol{\mu}^j)$$

2. *Update sets:*

Find pixel indices $\{\delta^j\}$ having Mahalanobis distances that are below the background cluster hypothesis threshold value τ^j which is a function of $p, |B^j|, \boldsymbol{\Sigma}^j, \hat{k}^j$ and $\hat{\Theta}^j$

$$\{\delta^j\} = \{I(d_m \leq \tau^j)\}$$
$$A^{j+1} = A^j \backslash \{\delta^j\}$$
$$B^{j+1} = B^j \bigcup \{\delta^j\}$$

4. *Stopping rule:*

If $\{\delta^j\} = \emptyset$, stop. Otherwise, estimate $\boldsymbol{\mu}^{j+1}$, $\boldsymbol{\Sigma}^{j+1}$, $\hat{k}^{j+1}$ and $\hat{\Theta}^{j+1}$ using B^{j+1} like in the first stage, increment j by and go to 1

Output: $A^f = A^j$, $B^f = B^j$, $\boldsymbol{\mu} = \boldsymbol{\mu}^j$, *and* $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^j$

Table 6.2: NG-BEVA - Second Stage

The global filter in NG-BEVA is similar to the one proposed in Chapter 4.

The NG-BEVA algorithm can be combined with Spectral Clustering as in Spectral BEVA. Fig. 6.1 summarizes the complete algorithm that includes improvements of Chapter 5 with NG-BEVA, named Spectral NG-BEVA.

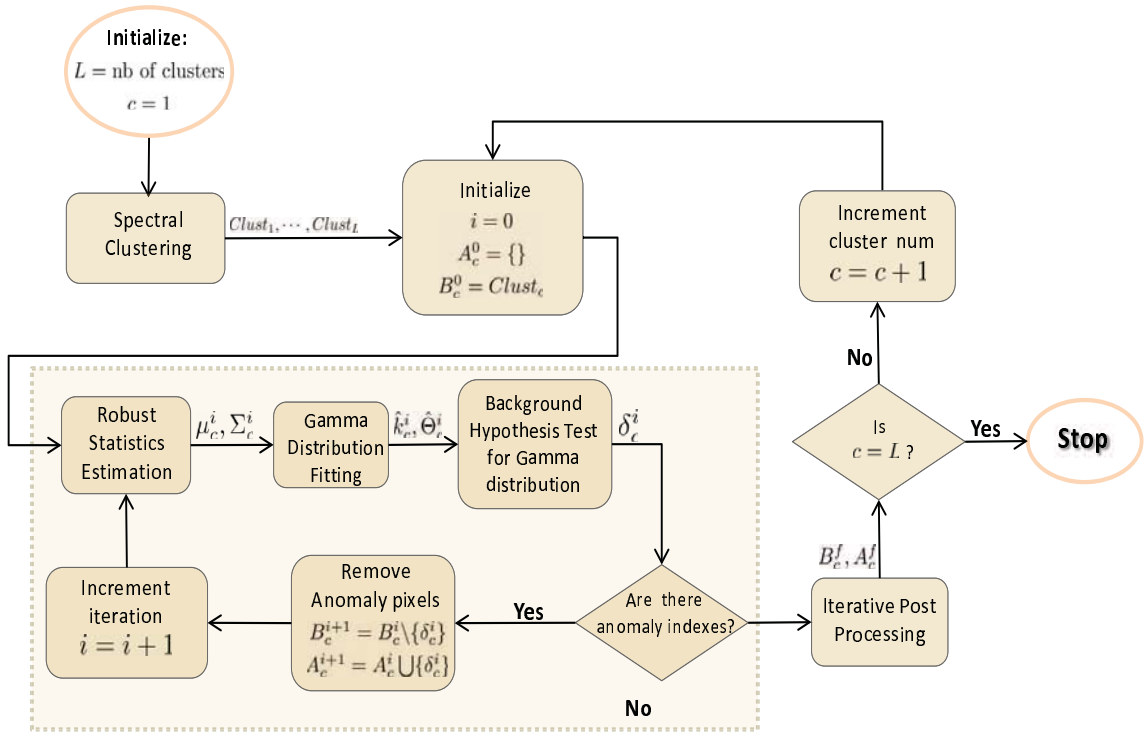

Figure 6.1: **Block diagram of the Local Part of Spectral NG-BEVA**

6.3 Experiments with Real Data

In this section we evaluate the performance of Spectral NG-BEVA by applying it to real hyperspectral data (For a description of the hyperspectral data used, see Appendix A). We compare it to the Spectral BEVA algorithm, in terms of Receiver Operation Characteristic (ROC) curves. For the two algorithms, the image was divided into non-overlapping local areas of size 35×35 . The global filter was applied to a larger image area of size 525×300. Each local area was segmented into 3 clusters with a local scaling parameter K set to 20. An auxiliary dictionary, initialized to 10 clusters for each image of the data sets, was added to

the two algorithms. The ROC curves are presented in Fig. 6.2 and 6.3

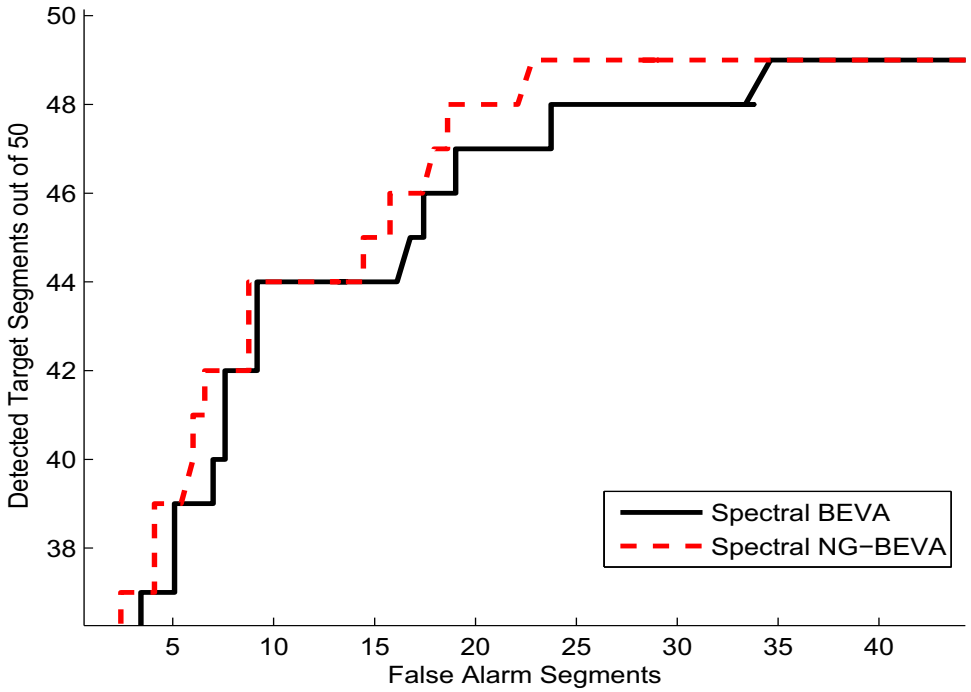

Figure 6.2: **ROC curve Spectral NG-BEVA vs. Spectral BEVA - Data Set I.**

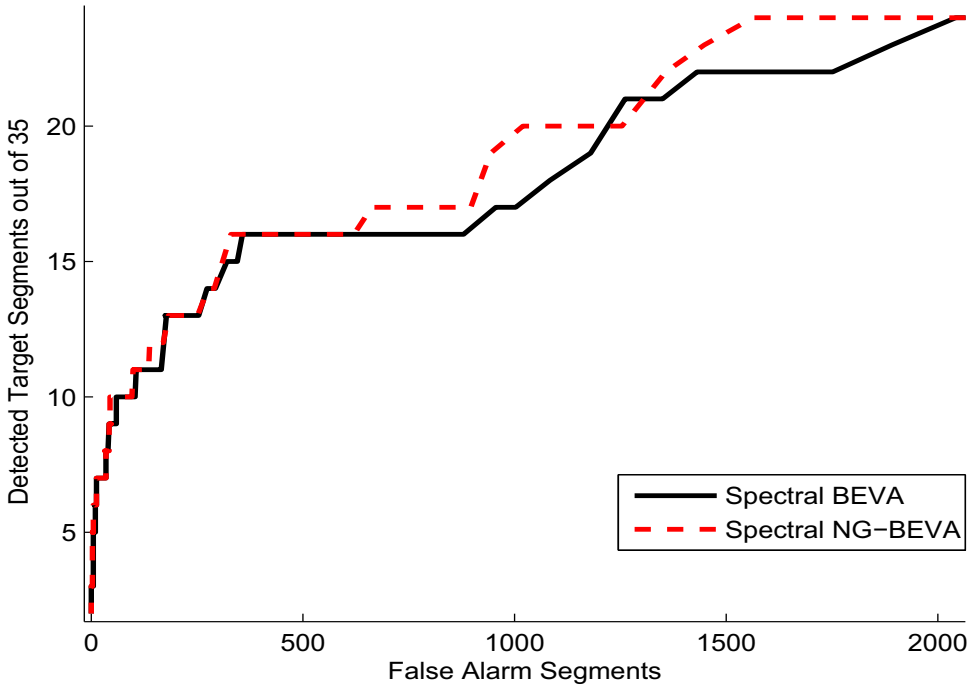

Figure 6.3: **ROC curve Spectral NG-BEVA vs. Spectral BEVA - Data Set II.**

For these two data sets of real hyperspectral images, the Spectra NG-BEVA algorithm is seen to have a better performance than the Spectral BEVA algorithm for all probability of detection values.

6.4 Summary

The Gaussian model used in BEVA, although efficient and mathematically tractable, is not sufficiently adequate to represent real hyperspectral data. In fact, background cluster distribution in hyperspectral data have heavier tails than the Gaussian pdf. Since the Gaussian model underestimates the distribution tails, it can lead to an excess number of false alarms. In this chapter we introduced

NG-BEVA, a non-Gaussian version of BEVA. In NG-BEVA, the Mahalanobis distances are no longer assumed to have a Chi-squared distribution, which was derived from the Gaussian assumption. Instead, a Gamma distribution $\mathbf{\Gamma}(\hat{k}, \hat{\Theta})$ is fitted to the Mahalanobis distances. In other words, the local background model is now more general and can better adjust to real hyperspectral data. In experiments with real hyperspectral image cubes, Spectral NG-BEVA, which combined Spectral Clustering improvement with NG-BEVA algorithm, was shown to have a better performance than any of the other tested algorithms in the whole range of probability of detection values.

Chapter 7

Conclusion

7.1 Summary

In this thesis we have studied how to detect anomalies in hyperspectral images. According to the hyperspectral literature, there are two major approaches to background modeling and anomaly detection. In the first approach, called "local", the background is modeled by a large number of local independent distributions, each of which is responsible to represent a different local region in the image. The second background modeling approach, called "global" is based on a global representation of the background process in the whole image.

In the second chapter, we focused our attention on two anomaly detection algorithms based on the local approach (RX [4]) and the global approach (GMM-RX [11]). These algorithms were implemented and tested on real hyperspectral data. For our data, the global algorithm GMM-RX performs better than RX.

Local algorithms like RX can approximate the background process by a model with a large number of degrees of freedom. Due to the many degrees of freedom, local background models may be tightly fitted to the background process. This allows obtaining high detection rate of anomalies. Unfortunately, in many cases,

the strength of local models turns into a weakness. This high number of degrees of freedom may cause model overfitting. Since the number of free parameters in local background models is proportional to the data size, the overfitting problem is almost inevitable. A model which over-fits the data will generally produces a high false-alarm rate.

On the contrary, the global background modeling algorithm GMM-RX is based on a universal multivariate distribution estimation. GMM-RX is more resistant to the overfitting problem, however, it is subject to an underfitting problem, which may result in high false alarm rates or/and low anomaly detection rates.

In order to overcome these major drawbacks, we propose in Chapter 3 the BEVA algorithm that properly combines local and global approaches. BEVA's local part consists of a two-stage iterative estimation process that combines robust estimation of the Gaussian statistics [22] with background cluster hypothesis testing. BEVA was found to perform better on real hyperspectral data than the other algorithms we have examined.

In BEVA, we use independent local background models that are interrelated in the global part. This construction help us to significantly reduce the vast number of degrees of freedom of the local method, while retaining the ability of local models to be intimately adjusted to the background. The essence of the local-global approach is to explore more complex local models that fit better to the background local disparities even at a price of a high false alarm rate, but alleviate the false alarm problem by using a global filter.

In Chapter 4, we notice a drawback of BEVA's global part: Spatially dispersed background pixels like rocks in a vegetation area, wrongly detected in BEVA local's part as anomalies, may not be removed by the global filer. This is because they cannot be associated with any local background cluster of the

global image area. A solution was given by adding an auxiliary background dictionary composed of global background clusters estimated using a global spectral segmentation algorithm. An anomaly pixel detected in BEVA's local part is now compared to both the local background clusters of the global image area and to the auxiliary background dictionary. Since the global segmentation is not influenced by the spatial configuration of the pixels, it helps to reduce false alarms for spatially dispersed background pixels which was a source of false alarms in BEVA.

In Chapter 5 we propose the Spectral BEVA algorithm to overcome a drawback of BEVA's local part. In BEVA's local part, segmentation of the background clusters and their estimation are combined together in one process that uses a background hypothesis test to determine if a pixel belongs or not to a background cluster. This background hypothesis test, based on Extreme Value Theory, is not sensitive enough to give a correct segmentation. In Spectral BEVA, the local block is first segmented using an improve Spectral Clustering algorithm ([27], [29]). Then, each segment is considered as a background cluster and its parameters are estimated using a simplified version of the estimation process in BEVA.

The Gaussian model used in BEVA, although efficient and mathematically tractable, is not sufficiently adequate to represent real hyperspectral data. In fact, background cluster distributions in hyperspectral data have heavier tails than the Gaussian pdf. Since the Gaussian model underestimates the distribution tails, it can lead to an excess number of false alarms. In order to overcome the Gaussian assumption, we introduce in Chapter 6 the NG-BEVA algorithm, a non-Gaussian version of the BEVA algorithm. In NG-BEVA, the Mahalanobis distances are no longer assumed to have a Chi-squared distribution, which was derived from the Gaussian assumption. Instead, a Gamma distribution $\mathbf{\Gamma}(\hat{k}, \hat{\Theta})$ is fitted to the

Mahalanobis distances. Then, a background hypothesis test is applied based on the Gamma distribution extreme values to detect anomalies.

In Fig. 7.1 and 7.2, the ROC curves of RX, GMM-RX, BEVA and Spectral NG-BEVA are represented. The Spectral NG-BEVA has the best performance in the whole range of relevant probability of detection values for the two data sets.

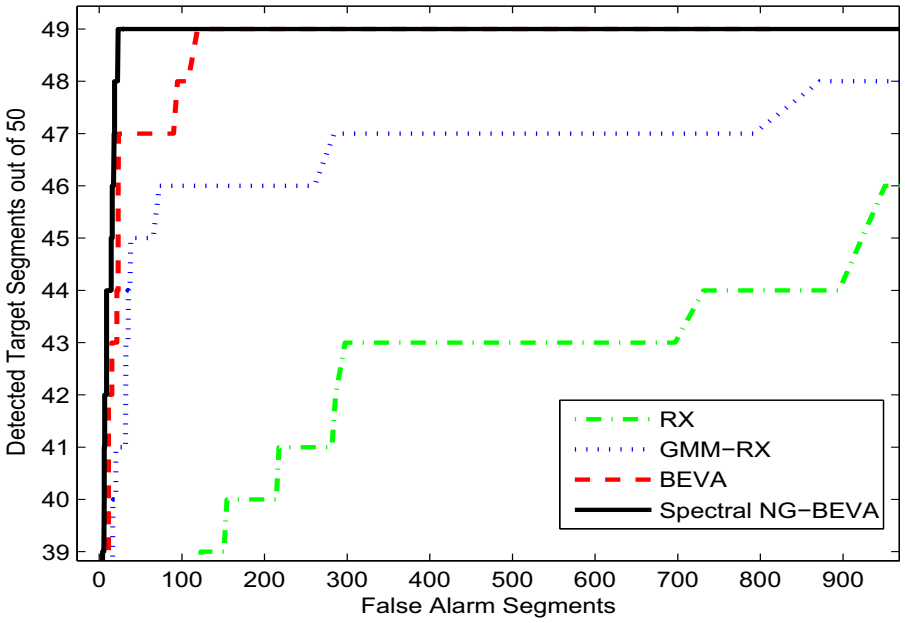

Figure 7.1: **ROC curves of all the examined algorithms - Data Set 1**

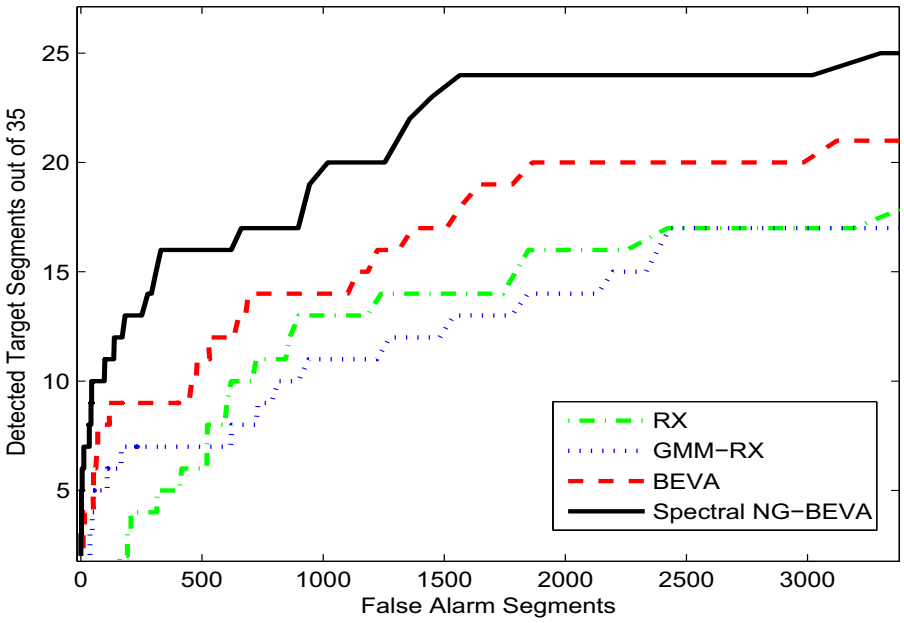

Figure 7.2: **ROC curves of all the examined algorithms - Data Set 2**

In Fig. 7.3, we compare the computation time of RX, GMM-RX, MCD and the different BEVA versions. The algorithms were applied to a 350×300 hyperspectral image of 65 spectral bands. The computer used was an Intel Core 2 duo 2Ghz with 2GB of RAM and all the algorithms were simulated in Matlab® environment. The algorithms' parameters were the same as in the previous chapters.

Algorithms	Computation Time
RX	770 s
GMM-RX	27 s
MCD	23 s
BEVA	61 s
BEVA+Auxilary Dictionary	98 s
Spectral BEVA	413 s
Spectral NG-BEVA	491 s

Figure 7.3: **Computation Time (Matlab®)**

The different variations of the BEVA algorithm are faster than RX but slower than GMM-RX and MCD. The Spectral NG-BEVA algorithm has the slower computation time of all BEVA versions but also gives the best performance.

7.2 Future Work

As discussed in Chapter 6, the Gaussian assumption is not sufficiently adequate to represent real hyperspectral data. In the NG-BEVA algorithm, a non-Gaussian version of the BEVA algorithm, a Gamma distribution $\mathbf{\Gamma}(\hat{k}, \hat{\Theta})$ is fitted to the Mahalanobis distances allowing to overcome the Gaussian model constraint. Other algorithms were proposed in hyperspectral literature to remove the Gaussian assumption. For example, Kernel-RX [5] proposed a non-linear version of the RX algorithm. Detection is made by assuming a Gaussian distribution for the two hypotheses H_0 and H_1 in a high-dimensional feature space obtained by a nonlinear mapping. Modeling the input data in the feature space by a Gaussian distribution is equivalent to representing the distribution of the input data with a much more complex model when defined in the original input space. The mapping is done by applying the "Kernel Trick" on the RX algorithm. The kernel defines

the distance between pixels in the non-linear feature space. In Spectral BEVA, the affinity map can also be considered as a kernel that defines distance between pixels. Therefore, in a future research, one may develop a non-linear local-global algorithm based on a kernelization using the affinity map obtained by Spectral Clustering.

In Chapter 5, we have proposed Spectral BEVA, an improvement of the BEVA algorithm. In the local part of Spectral BEVA, the local background is approximated using a Spectral Clustering algorithm [27]. Despite of its good performance, Spectral Clustering suffers from a serious drawback: the number of local background clusters is fixed and user defined for all the local blocks. The selection of a proper number of clusters for each local block may be a subject for future research.

The proposed algorithm BEVA and its variations combine local and global background modeling for anomaly detection in hyperspectral imaging. Although BEVA significantly improves detector performance compared to the local and global approaches, it is not the ultimate answer to completely avoid the overfitting or underfitting problems. Another way to accomplish this is to first reduce the data dimensionality before applying an anomaly detector. Therefore, a future research subject can be the development of an unsupervised algorithm for dimensionality reduction as a preprocessing stage of the BEVA algorithm.

References

[1] J.M. Grossmann, J. Bowles, D. Haas, J.A. Antoniades, M.R. Grunes, P. Palmadesso, D. Gillis, K.Y. Tsang, M. Baumback, M. Daniel, J. Fisher, and I. Triandaf *"Hyperspectral analysis and target detection system for the adaptive spectral reconnaissance program (ASRP)"*, SPIE, Algorithms for Multispectral and Hyperspectral Imagery IV, vol. 3372, Apr. 1998.

[2] M.E. Winter *"Fast autonomous spectral endmember determination in hyperspectral data"*, Proc. 13th Int. Conf. Applied Geologic Remote Sensing, vol. 2, pp. 337-344, 1999.

[3] R.A. Neville, K. Staenz, T. Szeredi, J. Lefebvre, and P. Hauff *"Automatic endmember extraction from hyperspectral data for mineral exploration"*, 4th Int. Airborne Remote Sensing Conf. Exhibition/21st Canadian Symposium on Remote Sensing, Jun. 1999.

[4] I. S. Reed and X. Yu, *"Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution"*, IEEE Trans. Acoust., Speech, Signal Process., vol. 38, no. 10, pp. 1760-1770, Oct. 1990.

[5] H. Kwon and N. M. Nasrabadi, *"Kernel RX-algorithm: A Nonlinear Anomaly Detector for Hyperspectral Imagery"*, IEEE Trans. on Geoscience and Remote Sensing, vol. 43, no. 2, pp. 388-397, Feb. 2005.

[6] M. Hsueh and C.-I. Chang *"Adaptive Causal Anomaly Detection for Hyperspectral Imagery"*, IEEE International Geoscience and Remote Sensing Symposium, vol. 5, 2004.

[7] A. Banerjee, P. Burlina, and C. Diehl *"A Support Vector Method for Anomaly Detection in Hyperspectral Imagery"*, IEEE Trans. on Geoscience and Remote Sensing, vol. 44, no. 8, pp. 2282-2291, Aug. 2006 .

[8] D. Rosario *"A Nonparametric F-Distribution Anomaly Detector for Hyperspectral Imagery"*, IEEE Aerospace Conference, vol. 5, no. 12, pp. 2022-2029, Mar. 2005.

[9] O. Kuybeda, D. Malah and M. Barzohar, *"Rank Estimation and Redundancy Reduction of High-Dimensional Noisy Signals with Preservation of Rare Vectors"*, IEEE Trans. Signal Proc., vol. 55, no. 12, pp. 5579-5592, Dec. 2007.

[10] O. Kuybeda, D. Malah and M. Barzohar, *"Global Unsupervised Anomaly Extraction and Discrimination in Hyperspectral Images via Maximum- Orthogonal Complement Analysis"*, EUSIPCO - European Signal Processing Conference, Aug. 2008, Lausanne Switzerland.

[11] D. Stein, S. Beaven, L. E. Hoff, E. Winter, A. Schaum and A. D. Stocker, *"Anomaly detection from hyperspectral imagery"*, IEEE Signal Process. Mag., vol. 19, pp. 58-69, Jan. 2002.

[12] P. J. Rousseuw *"Least Median of Squares Regression"*, Journal of the American Association, vol. 79, pp. 871-880, Dec. 1984.

[13] P. J. Rousseuw and K. van Driessen *"A Fast Algorithm for the Minimum Covariance Determinant Estimator"*, Technometrics, vol. 44, 1999.

[14] K.W. Smetek and T.E. Bauer *"Finding hyperspectral anomalies using multivariate outlier detection"*, IEEE Aerospace Conf., vol. 12, Mar. 2007.

[15] G. J. McLachlan and T. Krishnan *"SEM Algorithm and Unsupervised Statistical Segmentation of Satellite Images"*, IEEE Trans on Geoscience and Remote Sensing, vol. 31, no. 3, pp. 618-633, May 1993.

[16] S. G. Beaven, D. Stein, and L. E. Hoff *"Comparison of Gaussian mixture and linear mixture models for classification of hyperspectral data"*, Proc. IGARSS, Honolulu, HI, pp. 1597-1599, Jul. 2000.

[17] A. Hyvarinen, J. Karhunen, and E. Oja. *"Independent Component Analysis"*, Series on Adaptive and Learning Systems for Signal Processing, Communications, and Control. Wiley, 2001.

[18] G. J. McLachlan and T. Krishnan *"The EM Algorithm and Extensions"*, John Wiley and Sons, 1997.

[19] J. Bilmes *"A gentle tutorial on the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models"*, Technical Report, University of Berkeley, ICSI-TR-97-021, 1997.

[20] G. F. Hughes *"On the mean accuracy of statistical pattern recognizers"*, IEEE Trans. Inform. Theory, vol. 14, no. 1, pp. 55-63, Jan. 1968.

[21] E. Madar, O. Kuybeda, D. Malah, and M. Barzohar, *"Local-Global Background Modeling for Anomaly Detection in Hyperspectral Images"*, IEEE Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing - WHISPER, Aug. 2009.

[22] N. A. Campbell, *"Robust procedures in multivariate analysis I: Robust covariance estimation"*, Applied Statistics, vol. 29, no. 3, pp. 231-237, Jan. 1980.

[23] J. B. MacQueen, *"Some Methods for classification and Analysis of Multivariate Observations"*, Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, 1967.

[24] S. Coles, *"An Introduction to Statistical Modeling of Extreme Values"*, Springer Series in Statistics, 2001.

[25] G. J. McLachlan and T. Krishnan, *"The EM Algorithm and Extensions"*, John Wiley and Sons, 1997.

[26] J. Shi and J. Malik, *"Normalized cuts and image segmentation"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 888-905, 2000.

[27] A. Ng, M. Jordan and Y. Weiss, *"On spectral clustering: Analysis and an algorithm"*, Advances in Neural Information Processing Systems, vol. 14, 2001.

[28] D. Verma and M. Meila, *"A comparison of spectral clustering algorithms"*, Technical report, 2003 UW CSE Technical report 03-05-01.

[29] L. Zelnik-Manor and P. Perona, *"Self-Tuning Spectral Clustering"*, Advances in Neural Information Processing Systems, vol. 17, pp. 1601-1608, 2004.

[30] J. Wang, B. Thiesson, Y. Xu and M. Cohen, *"Image and Video Segmentation by Anisotropic Kernel Mean Shift"*, Lecture Notes in Computer Science, vol. 3022, pp. 238-249, 2004.

[31] D. A. Landgrebe, *"Signal Theory Methods in Multispectral Remote Sensing"*, JohnWiley and Sons, Hoboken, NJ, USA, 2003.

[32] D. Manolakis, D. Marden, J. Kerekes and G. Shaw, *"Statistics of hyperspectral imaging data"*, Proceedings of SPIE of Hyperspectral and Ultraspectral Imagery VII, vol. 4381, pp. 308-316, Apr. 2001.

[33] D. Manolakis and D. Marden, *"Non Gaussian models for hyperspectral algorithm design and assessment"*, Proceedings of IEEE International Geosciences and Remote Sensing Symposium (IGARSS 02), vol. 3, pp. 1664-1666, Jun. 2002.

[34] S. C. Choi and R. Wette, *"Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias"*, Technometrics, vol. 11, no. 4, pp. 683-690, Nov. 1969.

Appendix A

Experimental Data

In order to compare the performance of different algorithms developed in this thesis, we use ROC (Receiver operating characteristic) curves obtained by applying the different algorithms to two different data sets. The first one ("data set I") is composed of 5 real hyperspectral image cubes collected at 3 km altitude by an AISA airborne sensor configured to 65 spectral bands, uniformly covering VNIR range of 400nm - 1000nm wavelengths. Pixel resolution corresponds to $(0.8m)^2$. The total covered area of the 5 cubes is approximately $1.2km^2$. The second data set ("data set II") contains 2 real hyperspectral cubes collected at 4 km altitude by an AISA airborne sensor configured to 47 spectral bands, uniformly covering VNIR range of 400nm - 1000nm wavelengths. Pixel resolution corresponds to $(1.2m)^2$. The total covered area of the two cubes is approximately $0.6km^2$.

The images in the data sets contain ground-truth anomalies, which were manually identified using side information collected from high resolution CCD images of the corresponding scenes. For the purpose of ROC curves generation, all hyperspectral images were used, having a total number of 50 anomalies for data set I and 35 anomalies for data set II (vehicles, small agriculture facilities and small plastic containers). Anomalies in data set II are smaller (some of them are sub-

pixel anomalies) and are partially shaded compared to "data set I" anomalies. Thus their detection is more difficult.

In Fig. A.1 and A.2, one can see a false RGB representation of hyperspectral cubes used in this research. The false RGB image was obtained by combining three spectral bands each corresponding to the blue, green and red colors (bands 8, 18 and 27 corresponding to $461nm$, $552nm$ and $637nm$ respectively). The first hyperspectral cube (Fig. A.1) is of size $900 \times 400 \times 65$, the second (Fig. A.2) is of size $300 \times 400 \times 65$ where 65 is the number of spectral bands. The ground truth (in red) is composed of 16 and 11 anomalies, respectively.

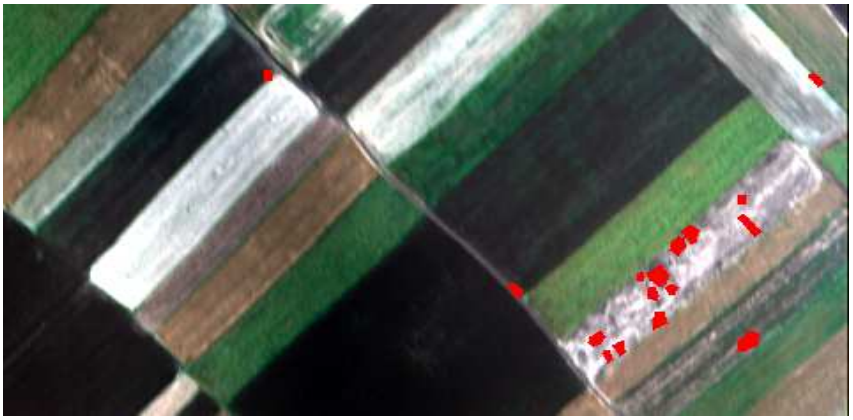

Figure A.1: **Hyperspectral false RGB representation - cube 1.**

Figure A.2: **Hyperspectral false RGB representation - cube 2.**

In Fig. A.3, we show one of the CCD images used for identifying the ground-truth anomalies.

Figure A.3: **High resolution CCD image of the analyzed scene,** used as a ground-truth indication The ground-truth anomalies are encircled by red ellipses.

Appendix B

GMM Parameters Estimation

GMM parameters can be estimated using a **hard** clustering version of the Expectation-Maximization (EM) algorithm ([18],[19]).

The EM algorithm is used for finding maximum likelihood estimates of parameters in probabilistic models, where the model depends on unobserved latent variables. In our case, we associate an unobserved vector $\mathbf{z}_i$ for each pixel vector $\mathbf{x}_i$. $\mathbf{z}_i$ is of length J, the number of background clusters, and its components are all zero except for one at index j indicating that $\mathbf{x}_i$ is belonging to cluster j.

The hard clustering version of the EM algorithm for GMM iterates between performing an expectation (E) step and a maximization (M) step. In the E step of iteration t, we compute an expectation of the likelihood with respect to the current estimated parameters $\mathbf{\Theta}$. The likelihood L is given as

$$L(\mathbf{x}, \mathbf{z}, \mathbf{\Theta}^{t-1}) = \prod_{i=1}^{N} \sum_{j=1}^{J} \mathbf{z}_i f_j(\mathbf{x}_i, \Theta_j^{t-1}) \tag{B.1}$$

where f_j is a Gaussian distribution with parameters Θ_j^{t-1} and $\mathbf{\Theta}^{t-1} \equiv \{\Theta_j^{t-1}\}_1^J$

The unobserved vectors $\mathbf{z}_i^t$ at iteration t are estimated from the data by the Bayes theorem with the Gaussian parameters $\boldsymbol{\mu}^{t-1}$ and $\mathbf{\Sigma}^{t-1}$ of the previous

iteration as follow (hard clustering case)

$$\mathbf{z}_{i,\widehat{j}}^{t} = \begin{cases} 1, & if\ \widehat{j} = argmin_j(\mathbf{x}_i - \boldsymbol{\mu}_j^{t-1})^T(\boldsymbol{\Sigma}_j^{t-1})^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_j^{t-1}) \\ 0, & otherwise \end{cases} \tag{B.2}$$

In the M step, we compute the parameters which maximize the likelihood found in the E step. The maximization is given as

$$\boldsymbol{\Theta}^t = argmax_{\boldsymbol{\Theta},\mathbf{z}^t} L(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{t-1}, z^t) \tag{B.3}$$

In GMM, the parameters $\boldsymbol{\mu}^t$ and $\boldsymbol{\Sigma}^t$ of each cluster j obtained at the M step are

$$\begin{aligned} \boldsymbol{\mu}_j^t &= \frac{\sum_{i=1}^N z_{i,j}^t \mathbf{x}_i}{\sum_{i=1}^N z_{i,j}^t} \\ \boldsymbol{\Sigma}_j^t &= \frac{\sum_{i=1}^N z_{i,j}^t (\mathbf{x}_i - \boldsymbol{\mu}_j^t)^T(\mathbf{x}_i - \boldsymbol{\mu}_j^t)}{\sum_{i=1}^N z_{i,j}^t} \end{aligned} \tag{B.4}$$

These parameters are then used to determine the distribution of the unobserved variables $\mathbf{z}_i$ in the next E step.

The initial partitioning of the data has a crucial importance on the output of the EM stage, as EM will converge to a local maximum in the neighborhood of the starting point. K-means clustering is usually used as an initialization of $\widehat{z_{ij}}$.

An outline of the hard clustering version of the EM algorithm is given in the table B.1.

Initialization: $\widehat{z}_{ik}^0$ is the output of a K-Means clustering. Iteration index: $t = 1$

Main iteration: Repeat:
1. *M-Step* Maximize given $\widehat{z}_{ij}^{t-1}$

$$\begin{aligned} \widehat{n}_j^t &= \sum_{i=1}^{N} \widehat{z}_{ij}^{t-1} \\ \widehat{\boldsymbol{\mu}}_j^t &= \frac{\sum_{i=1}^{N} \widehat{z}_{ij}^{t-1} \mathbf{x}_i}{\widehat{n}_j^t} \\ \widehat{\boldsymbol{\Sigma}}_j^t &= \frac{\sum_{i=1}^{N} \widehat{z}_{ij}^{t-1} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_j^t)^T (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_j^t)}{(\widehat{n}_j^{t-1})^2 - 1} \end{aligned} \tag{B.5}$$

2. *E-Step* Compute $\widehat{z}_{ij}^t$ given the parameter estimates from the M-Step

$$z_{i,\widehat{j}}^t = \begin{cases} 1, & if\ \widehat{j} = argmin_j (\mathbf{x}_i - \boldsymbol{\mu}_j^{t-1})^T (\boldsymbol{\Sigma}_j^{t-1})^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_j^{t-1}) \\ 0, & otherwise \end{cases} \tag{B.6}$$

3. *Stopping rule:*
If the l_2-norm of the representation error converges to a value smaller than a defined threshold T, stop. Otherwise increment j and go to 1

Output: $\widehat{z}_{ij} = \widehat{z}_{ij}^t,\ \widehat{\boldsymbol{\mu}}_j = \widehat{\boldsymbol{\mu}}_j^t\ and\ \widehat{\boldsymbol{\Sigma}}_j = \widehat{\boldsymbol{\Sigma}}_j^t$

Table B.1: Hard-Clustering EM algorithm for Gaussian Distribution

Appendix C

Distribution of maximum-norm Gaussian realizations

In this appendix we characterize the pdf $f_\nu(\cdot)$ of section 3.2.1. We assume that the background cluster of N pixels is approximated by a Gaussian with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Then, the Mahalanobis distances

$$d_i = (x_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x_i - \boldsymbol{\mu}) \tag{C.1}$$

$i = 1, \ldots, N$, have a Chi-squared distribution of order p equal to the number of spectral bands, denoted by $\chi^2(p)$,

$$f(u) = \frac{1}{2^{p/2}\Gamma(p/2)} u^{(p/2)-1} e^{-u/2}. \tag{C.2}$$

where Γ is the Gamma function.

For large p, the Central Limit Theorem can be used to obtain the following approximation:

$$d_i \sim \chi^2(p) \approx \mathcal{N}(p, 2p). \tag{C.3}$$

Now, the limiting distribution of v, which satisfies

$$v_k = \max_{i=1,\ldots,N} d_i, \tag{C.4}$$

can be obtained using the following Extreme Value Theory result [24]:

Theorem 1

If $\{d_i\}_{i=1}^N$ is i.i.d., with absolutely continuous distribution $F(x)$ and density $f(x)$, and letting

(i) $h(x) = f(x)/(1 - F(x))$

(ii) $b_N = F^{-1}(1 - \frac{1}{N})$

(iii) $a_N = h(b_N)$

(vi) $\omega = lim_{x \to x^*} \frac{dh(x)}{dx}$,
where x^* is the upper end-point of F,

then, for $M_N = \max\{\zeta_1 \ldots \zeta_N\}$,

$$P(a_N(M_N - b_N) \leq u) \underset{N \to \infty}{\longrightarrow} \begin{cases} \exp(-e^{-u}), & if\ \omega = \infty \\ \exp\{-[1 + \frac{u}{\omega}]^{\omega}\}, & if\ \omega < \infty \end{cases}, \tag{C.5}$$

The proof is found in [24].

In words: Theorem 1 says that the maximum of N i.i.d random variables has a limiting distribution that depends on ω - a parameter derived from their individual distributions. For the purpose of the present work, we consider normal and chi-squared distributions, which lead to $\omega = \infty$.

Therefore, from (C.5), the limiting distribution of interest is

$$G(u) \triangleq \exp(-e^{-u}), \tag{C.6}$$

also known as the Gumbel distribution. The mean and std of a variable distributed as (C.6) are $\eta = 0.5772$ and $\gamma = 1.6450$, respectively. The normalizing coefficients a_N and b_N are also functions of the ζ_i distribution. Theorem 1 also describes how to calculate the normalizing coefficients given the distribution function of d_i.

Appendix D

Derivation of conditional hypothesis probabilities

In the following, we derive the conditional probabilities $p(H_0|\eta)$ and $p(H_1|\eta)$ in (6.4) and (6.5), based on pdfs f_υ and f_ξ:

$$\begin{aligned}
f(H_0,\eta) &= f_\upsilon(\eta)p(\xi<\eta) = \\
&\quad f_\upsilon(\eta)F_\xi(\eta) = f_\upsilon(\eta), \\
f(H_1,\eta) &= f_\xi(\eta)p(\upsilon<\eta) = \\
&\quad f_\xi(\eta)F_\upsilon(\eta) = F_\upsilon(\eta)\frac{1}{\eta}, \\
f_\eta(\eta) &= f(H_0,\eta)+f(H_1,\eta) = \\
&\quad \frac{1}{\eta}\left[\eta f_\upsilon(\eta)+F_\upsilon(\eta)\right], \\
p(H_0|\eta) &= \frac{f_\upsilon(\eta)F_\xi(\eta)}{f_\eta(\eta)} = \frac{\eta f_\upsilon(\eta)}{\eta f_\upsilon(\eta)+F_\nu(\eta)}, \\
p(H_1|\eta) &= \frac{f_\xi(\eta)F_\upsilon(\eta)}{f_\eta(\eta)} = \frac{F_\upsilon(\eta)}{\eta f_\upsilon(\eta)+F_\upsilon(\eta)},
\end{aligned}$$

which are the expressions shown in (6.4) and (6.5).

Appendix E

Maximum Likelihood Estimation of Gamma Distribution

In the following, we derive Maximum Likelihood estimators $\hat{k}$ and $\hat{\Theta}$ of the Gamma distribution $\mathbf{\Gamma}(k, \Theta)$, with the following pdf:

$$f(u) = \frac{1}{\Theta^k \Gamma(k)} u^{k-1} e^{-u/\Theta} \qquad \text{with} \quad u \geq 0 \tag{E.1}$$

The likelihood function for N iid observations $(x_1, \cdots, x_N)$ is

$$L(k, \Theta) = \prod_{i=1}^{N} f(x_i; k, \Theta) \tag{E.2}$$

from which we calculate the log-likelihood function

$$l(k, \Theta) = (k-1) \sum_{i=1}^{N} ln(x_i) - \sum_{i=1}^{N} \frac{x_i}{\Theta} - Nkln(\Theta) - Nln\Gamma(k) \tag{E.3}$$

Finding the maximum with respect to Θ by taking the derivatives and setting it equal to zero gives the maximum likelihood estimator for the Θ parameter:

$$\hat{\Theta} = \frac{1}{kN} \sum_{i=1}^{N} x_i \tag{E.4}$$

Substituting this into the log-likelihood function gives

$$l(k) = (k-1)\sum_{i=1}^{N} ln(x_i) - Nk - Nkln(\frac{\sum x_i}{kN}) - Nln\Gamma(k) \tag{E.5}$$

Finding the maximum with respect to k by taking the derivative and equating it to zero yields

$$ln(k) - \psi(k) = ln(\frac{1}{N}\sum_{i=1}^{N} x_i) - \frac{1}{N}\sum_{i=1}^{N} ln(x_i) \tag{E.6}$$

where $\psi(k) = \frac{\Gamma'(k)}{\Gamma(k)}$.

There is no closed-form solution for k. An initial value of k can be found by using the approximation

$$k \approx \frac{3 - s + \sqrt{(s-3)^2 + 24s}}{12s} \tag{E.7}$$

where $s = ln(\frac{1}{N}\sum_{i=1}^{N} x_i) - \frac{1}{N}\sum_{i=1}^{N} ln(x_i)$.

The value of k can be obtained by an iteration process using this initial guess. The iteration update, proposed by Choi and Wette [34], has the following expression:

$$k \leftarrow k - \frac{ln(k) - \psi(k) - s}{\frac{1}{k} - \psi'(k)} \tag{E.8}$$