LOCAL-GLOBAL BACKGROUND MODELING FOR ANOMALY DETECTION IN HYPERSPECTRAL IMAGES

Eyal Madar, Oleg Kuybeda, David Malah, and Meir Barzohar

Department of Electrical Engineering
Technion IIT, Haifa 32000, Israel

ABSTRACT

In this paper, we address the problem of unsupervised detection of anomalies in hyperspectral images. Our proposed method is based on a novel statistical background modeling approach that combines local and global approaches. The local-global background model has the ability to adapt to all nuances of the background process like local approaches but avoids over-fitting due to a too high number of degrees of freedom, which produces a high false alarm rate. This is done by constraining the local background models to be interrelated. The results strongly prove the effectiveness of the proposed algorithm. We experimentally show that our local-global algorithm performs better than several other global or local anomaly detection techniques, such as the well known RX or its Gaussian Mixture version (GMRX).

***Index Terms*—** Background Modeling, Unsupervised Anomaly Detection, Hyperspectral Images

1. INTRODUCTION

In this work we address the problem of anomaly detection in hyperspectral images. The considered sceneries are composed of abundant natural materials such as vegetation, soil, minerals, etc., along with rare material such as localized man-made objects, e.g., small buildings, vehicles, etc.

In anomaly detection tasks, hyperspectral pixels have to be classified into either background or anomalies. An anomaly can be defined as a singular set of hyperspectral pixels spectrally different in a meaningful way from the abundant (background) material spectra. If no prior knowledge about anomaly signatures is available, anomaly detection algorithms first model the image background process. Then, every pixel whose spectrum is not well-described by the background model, is declared to be an anomaly.

A common background modeling methodology is based on statistical modeling. There are two major approaches in statistical background modeling. According to the first approach, the background is modeled by a large number of mutually 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 approach, named RX [1], has become a benchmark anomaly detection algorithm in the remote sensing community. Due to the many degrees of freedom, local background models can be tightly fitted to the background data. This allows obtaining high detection rate of anomalies, since local models are able to discern subtle differences between anomalies and the background more easily. Unfortunately, in many cases, the strength of local models turns into a weakness, since the high number of degrees of freedom may cause model over-fitting, which, in turn, produces a high false-alarm rate. Generally, this happens due to the well known phenomenon that the number of training data pixels has to be significantly higher than the number of the model degrees of freedom. Since the number of free parameters in local background models is proportional to the data size, the model over-fitting problem is almost inevitable.

The second background modeling approach is based on a simple universal distribution, which is designed to represent the background process in the whole image. Therefore, we call this approach ”global”. There are various global modeling approaches found in the literature like GMM [2]. By design, these methods are more resistant to the over-fitting problem. However, they have a limited ability to adapt to all nuances of the background process (an under-fitting problem), which may result in both - high false alarm rates, as well as low anomaly detection rates.

Obviously, there is no ultimate answer how to completely avoid the over-fitting or under-fitting 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 constrained to be interrelated in some way. This construction may 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. Therefore, we call this approach as ”local-global”.

In this work we propose a local-global approach based on a greedy algorithm 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.

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

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.

2. PROPOSED ALGORITHM DESCRIPTION

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 $d = (x - \mu)\Gamma^{-1}(x - \mu)^T$ of a realization x to the mean of a background cluster, where μ and Γ are the mean and covariance of a Gaussian $N(\mu, \Gamma)$ that approximates this background cluster.

A set of N data-vector indices of a local image area can hypothetically be divided into two subsets:

$$B = \{\text{indices of realizations of a specific background cluster}\}$$
$$A = \{\text{indices of realizations of other background clusters or anomalies}\}$$

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:

$$\begin{aligned} H_0 &: \delta \ \text{belongs to } B \\ H_1 &: \delta \ \text{belongs to } A \end{aligned} \tag{2}$$

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:

$$\eta = \max(\upsilon, \xi) \tag{3}$$

Since the model supposes that the background cluster pdf can be approximated by a Gaussian pdf, the value of υ is governed by the extreme value statistics of maximum-norm Gaussian realizations [3]. Since the pdf shape of ξ is unknown, is set $\xi \sim U[0, \eta]$.

Given the pdfs $f_\upsilon(\cdot)$ and $f_\xi(\cdot)$, the conditional hypotheses $P(H_0|\eta)$ and $P(H_1|\eta)$ can be evaluated (for details see [4]).

In Fig. 1 , the two conditional hypotheses are represented. The crossing point τ of the two hypotheses, i.e. the Mahalanobis distance above which $P(H_0|\eta) \leq P(H_1|\eta)$, can be used as a threshold to isolate the background cluster realizations from other realizations in the data-set. Given a Gaussian statistics $N(\mu, \Gamma)$, which represents a background cluster pdf, a data pixel having a Mahalanobis distance that is below τ, will be declared as background cluster pixel.

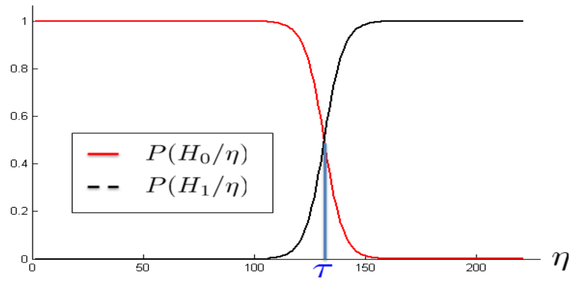

Fig. 1. Conditional hypotheses probabilities $P(H_0|\eta)$ and $P(H_1|\eta)$ obtained from $N = 10,000$ pixels of 65 spectrals bands

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 on 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

$$\begin{aligned} B^f &= \text{Pixels of the dominant background cluster} \\ A^f &= \text{Pixels of other background clusters or anomalies} \end{aligned} \tag{4}$$

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 (5% 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 Gaussian statistics estimation from [5] with the background cluster hypothesis testing as described in section 2-1. 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 Figs. 2 and 3.

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

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

Main Iteration: Perform the following steps:

1. *Robust estimation of Mean and Covariance (see [5])*

$$\mu^i = \frac{\sum_{x_m \in B^i} \omega_m^i x_m}{\sum_{x_m \in B^i} \omega_m^i}$$

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

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

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

3. *Update weights (see [5]):*

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

4. *Update sets:*

Find data pixel indices $\{\delta^i\}$ having Mahalanobis distances that exceed the background cluster hypothesis threshold value $\tau^i = f(p, |B^i|, \Gamma^i)$, obtained from [4]:

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

5. *Stopping rule:*

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

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

Fig. 2. First Stage of a background cluster estimation

In summary, by 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.

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 an over-fitting problem.

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

Main Iteration: Perform the following steps:

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

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

2. *Update sets:*

Find pixel indices $\{\delta^j\}$ having Mahalanobis distances that are below the background cluster hypothesis threshold value $\tau^j = f(p, |B^j|, \Gamma^j)$, obtained from [4]:

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

4. *Stopping rule:*

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

Output: $A^f = A^j,\ \ B^f = B^j,\ \ \mu = \mu^j,\ \ and\ \ \Gamma = \Gamma^j$

Fig. 3. Second Stage of a background cluster estimation

We propose here a global filtering approach, which reduces the number of degrees of freedom by inter-relating the found 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, $d = min_{i,j}\{(x - \mu_{i,j})(\Gamma_{i,j})^{-1}(x - \mu_{i,j})^T\}$ where $i = 1 \cdots T$ and j runs over the number of clusters in block i. If the found minimum Mahalanobis distance is smaller than the background hypothesis threshold value, the local anomaly pixel will be removed from the anomaly set. At the end, A is composed just of global anomalies.

3. RESULTS WITH REAL HYPERSPECTRAL DATA

In this section we evaluate the performance of the proposed algorithm by applying it to real hyperspectral data. To demonstrate the results, the algorithm was applied to 5 real hyperspectral image cubes collected by an AISA airborne sensor configured to 65 spectral bands, uniformly covering VNIR range of 400nm - 1000nm wavelengths. At 4 km altitude pixel resolution corresponds to $(0.8m)^2$. The total covered area of the 5 cubes is approximately $1.2km^2$. For the experiment, the image is divided into non-overlapping local areas of size 40×40 . The global filter is applied to each local anomaly found using a large image area of size of 400×400.

In Fig. 4, one can see results of anomaly detection and discrimination. The left image contains ground-truth anomalies (marked in red and encircled by ellipses), which were manually identified using high resolution CCD images of the corresponding scenes. The ground truth anomalies consist of vehicles and small agriculture facilities, which occupy few-pixel segments.

In Fig. 5, we compare the proposed algorithm to RX [1], GMRX [2] and FastMCD [6], in terms of Receiver Operation Characteristic (ROC) curves. For the purpose of ROC curves generation, all hyperspectral images were used, having a total number of 50 anomalies. 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.

Clearly, the proposed approach has a better performance than RX, GMRX and FastMCD for the examined images.

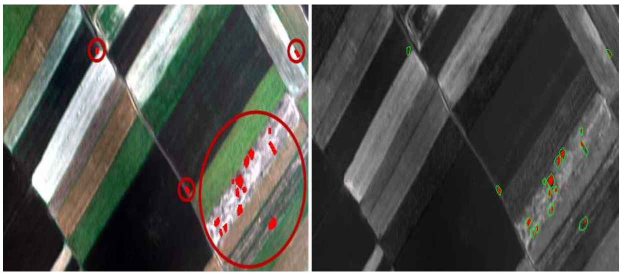

Fig. 4. Proposed algorithm results at the nominal operating point; Left image - manually identified ground-truth anomalies ; Right image - detected anomalies (marked in color)

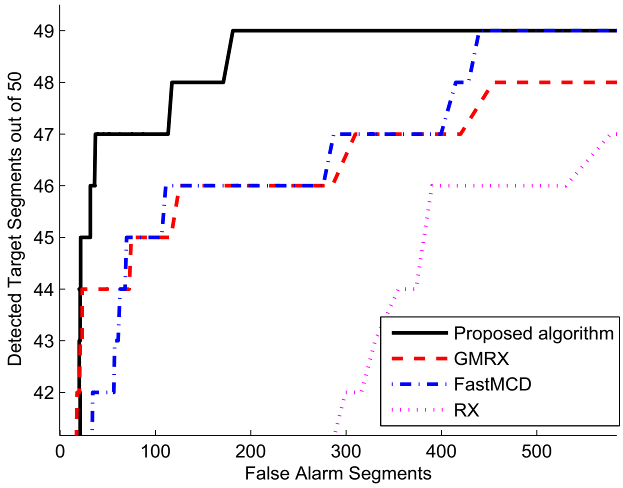

Fig. 5. ROC curves corresponding to RX, GMRX, FastMCD and proposed algorithm

4. CONCLUSION

In this work we presented an anomaly detection algorithm based on a local-global statistical background modeling approach that helps 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. In the local part, the local background is approximated using a greedy sequential estimation process that applies a robust Gaussian statistics estimation and background cluster hypothesis testing based on extreme value theory results. Then, in the global part, the found local background models are constrained to be inter-related to reduce the number of false alarms.

In experiments with real hyperspectral image cubes the proposed algorithm was shown to have a better performance than RX [1], GMRX [2] and FastMCD [6].

5. REFERENCES

[1] 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, pp. 1760 1770, Oct. 1990.

[2] 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.

[3] Stuart Coles, "An introduction to statistical modeling of extreme values," *Springer Series in Statistics.*, 2001.

[4] 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. 12, pp. 5579 5592, Dec. 2007.

[5] N.A. Campbell, "Robust procedures in multivariate analysis - robust covariance estimation," *Applied Statistics*, vol. 29, pp. 231–237, 1980.

[6] K.W. Smetek, T.E. Bauer, "Finding hyperspectral anomalies using multivariate outlier detection," *IEEE Aerospace Conf.*, pp. 1–24, March. 2007.