NON-GAUSSIAN BACKGROUND MODELING FOR ANOMALY DETECTION IN HYPERSPECTRAL IMAGES

Eyal Madar, David Malah and Meir Barzohar

Department of Electrical Engineering, Technion IIT,
32000, Haifa, Israel
{eyalm@tx, malah@ee}.technion.ac.il, meirb@visionsense.com

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 and does not assume Gaussianity. The local-global background model has the ability to adapt to all nuances of the background process, like local models, but avoids overfitting that may result due a too high number of degrees of freedom, producing a high false alarm rate. This is achieved by globally combining the local background models into a "dictionary", which serves to remove false alarms. Experimental results strongly prove the effectiveness of the proposed algorithm. These results show that the proposed local-global algorithm performs better than several other local or global anomaly detection techniques, such as the well known RX or its Gaussian Mixture version (GMM-RX).

1. INTRODUCTION

The detection of materials and objects using remotely sensed spectral information collected by hyperspectral sensors has many civilian and military 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. In anomaly detection, no prior knowledge on the target spectral signature is assumed. Therefore, anomaly detection algorithms first model the abundant materials 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 models can tightly fit the background data, however they are subject to overfitting, 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 overfitting. 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.

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 a set of local models, while retaining the ability of local models to be intimately adjusted to the background. We call this approach "local-global".

In [1], we present an algorithm based on a "local-global" statistical background modeling approach. The local model in the proposed algorithm is composed of a small number of distinct clusters, each one assumed to have a Gaussian distribution. This configuration helps to handle multiple types of terrain but is subject to overfitting, which may lead to a large number of false alarms. Thus, a global filter is then applied to alleviate the overfitting problem. Each anomaly pixel detected in the local part of the proposed algorithm is compared to a "dictionary" where each "word" consists of estimated local background cluster statistical parameters of a larger image area. Only pixels that are determined as not being associated with any background clusters, are be declared as true anomalies. This proposed "local-global" algorithm was tested on real hyperspectral data and has shown better performance than several examined local or global anomaly detection algorithms in terms of Receiver Operation Characteristic (ROC) curves.

The choice of the Gaussian model in [1] is due to its efficient processing and mathematical tractability. In fact, it simplifies the derivation of decision rules and the evaluation of detector 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 ([2], [3]) that the Gaussian model fails in its representation of the distribution tails. In particular, distributions of 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 paper, we introduce a novel algorithm, denoted NG-BEVA (Non-Gaussian Background Extreme Value Analysis). It is an anomaly detection algorithm that is based on the combined local-global background modeling approach, but without the Gaussian assumption.

2. PROPOSED ALGORITHM DESCRIPTION

2.1 Local background modeling

In the local part of NG-BEVA, the hyperspectral image is partitioned into distinct local blocks. The background local model used for each local block is capable of handling multiple types of terrain. It is composed of L distinct clusters (typically, up to 3),

$$x \in C_k,\ 1 \leq k \leq L \quad (1)$$

Where x is a background pixel in the local block and C_k, the k-th cluster of the L background clusters.

Each local block is first clustered using the Spectral Clustering algorithm in [4]. Each obtained cluster is supposed to be a background cluster "polluted" by some anomalies. We estimate the statistics of the background cluster by using a greedy iterative estimation process based on Extreme Value Theory [7] results for the Gamma distribution, as detailed below.

2.1.1 Spectral Clustering

In recent years, Spectral Clustering (SC) has become one of the most popular modern clustering algorithms [5]. It is simple to implement, can be solved efficiently by standard linear algebra software, does not assume cluster Gaussianity, and very often outperforms traditional clustering algorithms such as K-means [6]. In order to cluster a given dataset, Spectral Clustering relies on the eigenstructure of an affinity matrix W, where an element w_{ij} represents a measure of the affinity between the vectors i and j in the dataset.

An outline of the Spectral Clustering algorithm is shown in Table 1.

In experiments with real hyperspectral data, we observed that instead of selecting a single scaling parameter σ, the use of a local scaling parameter σ_i for each data point x_i, as proposed by Zelnik-Manor and Perona in [4], improves greatly the clustering results.

Using a specific scaling parameter for each pixel allows self-tuning of the pixel-to-pixel distances according to the local statistics of the neighborhoods surrounding pixels i and j. The affinity between two pixels is now given by

$$w_{ij} = exp(-\frac{||x_i - x_j||_2^2}{\sigma_i \sigma_j}) \quad (2)$$

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

$$\sigma_i = d(x_i, x_M) \quad (3)$$

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

2.1.2 Local background model estimation

By applying Spectral Clustering as described above, we get for each local block L "coarse" clusters. The number of "coarse" clusters L is a user-defined parameter. We assume that each "coarse" cluster contains pixels of a background cluster together with some possible anomalies. The background pixels have then to be discriminated from other pixels so that the statistics of the background cluster could be reliably estimated. To perform this task, we initialize for each

Table 1: Spectral Clustering Algorithm [5]

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

Algorithm: Perform the following steps:

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

$$w_{ij} = exp(-\frac{||x_i - x_j||_2^2}{\sigma^2})$$

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_w:*

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

4. *Solve the following eigenvalue problem:*

$$L_w v = \lambda v$$

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

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

6. *Normalize the rows of V:*

$$\widetilde{V}_{ij} = V_{ij} / \sqrt{\sum_{j=1}^{L} V_{ij}^2}\ , \quad 1 \leq i \leq N$$

7. *Treat each row $\widetilde{V}_i$ of $\widetilde{V}$ as a point in $\mathbb{R}^L$ and cluster the N rows via K-means (K=L).*

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 L clusters.

cluster C of the L "coarse" clusters, two distinct indices sets $A_C^0 = \emptyset$ and $B_C^0 = \{m | x_m \in C\}$ and aim to get

$$\begin{aligned} B_C^f &= \{\text{Pixel indices of Background cluster } C\} \\ A_C^f &= \{\text{Pixel indices of local Anomalies}\} \end{aligned} \quad (4)$$

To remove the effect of outliers (local anomalies) on the estimation of background clusters statistics, we propose an iterative estimation process that combines robust mean and covariance estimation [9] together with a background cluster hypothesis test based on Extreme Value Theory results [7] for the Gamma distribution [8], as described bellow.

The background hypothesis test determines which pixels belong to a background cluster, and which are outliers in the particular block. It is based on examining the Mahalanobis distance $d = (x - \mu)^t \Sigma^{-1} (x - \mu)$ from a realization x to the mean of a background cluster, where μ and Σ are the first and second moments of the distribution that approximates this background cluster, and t denotes "transpose".

Let $\eta = \max_{i=1,\cdots,N}(d_i)$ be, the maximum Mahalanobis distance over the N data-vector indices of the"coarse" cluster

C, obtained at index δ. Given η and δ, we formulate the following hypotheses:

$$\begin{aligned} H_0 &: \delta \text{ belongs to } B_C^f \\ H_1 &: \delta \text{ belongs to } A_C^f \end{aligned} \tag{5}$$

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

$$\eta = \max(\upsilon, \xi) \tag{6}$$

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 $F_V(\cdot)$ and $F_\xi(\cdot)$.

In [1], each background cluster was approximated by a unimodal Gaussian pdf $N(\mu, \Sigma)$ with μ and Σ being the mean and covariance, respectively. For larger values of N, the Mahalanobis distances can be approximated by a Chi-squared distribution of order p, denoted by $\chi^2(p)$, where p is equal to the number of spectral bands.

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

$$f(u) = \frac{1}{\Theta^k \Gamma(k)} u^{k-1} e^{-u/\Theta} \qquad \text{with} \quad u \geq 0 \tag{7}$$

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

By using the Gamma distribution, we have relaxed the Gaussian model constraint. 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 $\Gamma(k, \Theta)$, obtained for $k = \frac{p}{2}$ and $\Theta = 2$.

Given the pdf and cdf of the Gamma distribution, the conditional hypotheses probabilities are given by [8]:

$$P(H_0|\eta) = \frac{\eta f_\upsilon(\eta)}{\eta f_\upsilon(\eta) + F_\upsilon(\eta)}, \tag{8}$$

$$p(H_1|\eta) = \frac{F_\upsilon(\eta)}{\eta f_\upsilon(\eta) + F_\upsilon(\eta)}, \tag{9}$$

where f_υ and F_υ are the pdf and cdf, respectively, of the Gamma distribution with estimated max-likelihood parameters $\hat{k}$ and $\hat{\Theta}$.

Fig. 1 shows the two conditional hypothesis probabilities obtained for $N = 10,000$ pixels of $p = 65$ spectral bands and a Gamma distribution with $k = 25$ and $\Theta = 2.5$. Theses values are close to the values one obtains for real hyperspectral data.

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. A data pixel having a Mahalanobis distance that is below τ, will be declared as a background cluster pixel. The transition region between hypotheses is steep and narrow. Actually, its width is inversely proportional to the number of spectral bands p and to the square of N, the number of pixels.

The iterative estimation process that combines Robust statistics estimation with the background hypothesis test is described in detail in Table 2.

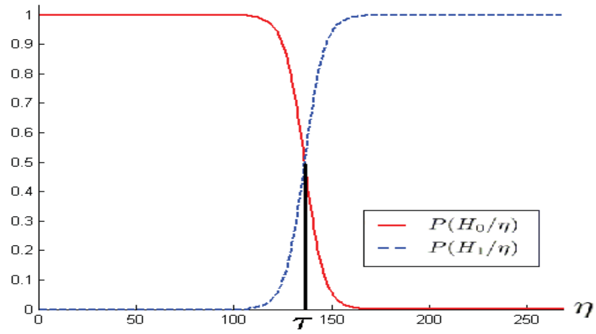

Figure 1: **Conditional hypotheses probabilities $P(H_0|\eta)$ and $P(H_1|\eta)$ obtained for $N = 10,000$ pixels of $p = 65$ spectral bands. Gamma distribution parameters used are $k = 25$ and $\Theta = 2.5$.**

Table 2: NG BEVA - Background Parameter Estimation

Inputs: $In = \{m | x_m \in C\}$, $x_m : [p \times 1]$, 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* [9] :

$$\mu_C^i = \frac{\sum_{x_m \in B_C^i} \omega_m^i x_m}{\sum_{x_m \in B_C^i} \omega_m^i}$$

$$\Sigma_C^i = \frac{\sum_{x_m \in B_C^i} (\omega_m^i)^2 (x_m - \mu_C^i)^t (x_m - \mu_C^i)}{\sum_{x_m \in B_C^i} (\omega_m^i)^2 - 1}$$

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

$$\forall x_m \in B_C^i : \; d_m = (x_m - \mu_C^i)^t (\Sigma_C^i)^{-1} (x_m - \mu_C^i)$$

3. *Update weights* [9]:

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

4. *Gamma Distribution Fitting using the estimated Maximum Likelihood parameters:*

$$\forall x_m \in B_C^i : d_m \rightarrow \Gamma(\hat{k}_C^i, \hat{\Theta}_C^i)$$

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 $p, |B_C^i|, \Sigma_C^i, \hat{k}_C^i$ and $\hat{\Theta}_C^i$:

$$\{\delta^i\} = \{I(d_m \geq \tau^i)\} \text{ where } I = Index\,of\,value$$
$$B_C^{i+1} = B_C^i \backslash \{\delta^i\}$$
$$A_C^{i+1} = A_C^i \bigcup \{\delta^i\}$$

6. *Stopping rule:*

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

Output: $A_C^f = A_C^i$, $B_C^f = B_C^i$, $\mu_C^f = \mu_C^i$, $\Sigma_C^f = \Sigma_C^i$, $\hat{k}_C^f = \hat{k}_C^i$ *and* $\hat{\Theta}_C^f = \hat{\Theta}_C^i$.

2.2 Global Filtering

The proposed algorithm for local background estimation allows the use of several background cluster 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.

As in our work in [1], we apply a global filtering approach, which reduces the number of degrees of freedom by inter-relating the obtained local background pdfs. Given a local block, having A as the subset of indices of local anomaly pixels in it, we consider a larger image area composed of T blocks around it. All the local anomaly pixels whose indices are in A are compared to the T relevant local background models (each consisting of up to 3 clusters). For every local anomaly pixel $\{x_m | m \in A\}$, we find the minimum Mahalanobis distance from it to all the clusters centroids of the T backgrounds, $d_m = min_{i,j}\{(x_m - \mu_{i,j})^t (\Sigma_{i,j})^{-1} (x_m - \mu_{i,j})\}$ 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 in which the minimum was found, the index is removed from A. At the end, A contains just the indices of global anomaly pixels present in the examined local block.

By combining 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" consists of estimated local background cluster statistical parameters. A pixel is declared as an anomaly if it does not fit well any "word" in the "dictionary". Additional words, based on global learning (or even supervised learning) of the background clusters, can be added to the "dictionary" to improve its performances.

Table 3 presents a formal description of the global filter.

Table 3: Global Filter

Task: Reduce the number of false alarms in a given local block.

Input: Subset A of indices of local anomaly pixels in the given local block.

Loop: For each $\{x_m | m \in A\}$
1. *Calculate the minimal Mahalanobis distance to the centroids of the L_i background clusters over the blocks $i = 1, \cdots, T$:*

$$d_m = min_{j,i}\{(x_m - \mu_{i,j})^t (\Sigma_{i,j})^{-1} (x_m - \mu_{i,j})\}, \quad j = 1, \cdots, L_i$$

2. *Update A:*
Compare d_m to τ, the crossing point of the two conditional hypotheses (eq. (9)) corresponding to the closest background cluster:

If $d_m \leq \tau$
Then $A = A \backslash \{m\}$

Output: A contains just the indices of global anomaly pixels present in the examined local block.

3. PERFORMANCE EVALUATION

In this section we evaluate the performance of NG-BEVA 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, each image was divided into non-overlapping local blocks of size 35×35. Each block was partitioned into 3 clusters and the local scaling parameter σ_i was obtained with M set to 20 . The global filter is applied to each local anomaly found, using a larger image area of size 350×350.

In Fig. 2, one can see the ground-truth anomalies on a hyperspectral image (marked in red and encircled by ellipses), which were manually identified using high resolution CCD images of the corresponding scenes. In Fig. 3, we show the CCD image corresponding to the hyperspectral image in Fig. 2, which was used for identifying the ground-truth anomalies. The ground truth anomalies consist of vehicles and small agriculture facilities, which occupy few-pixel segments.

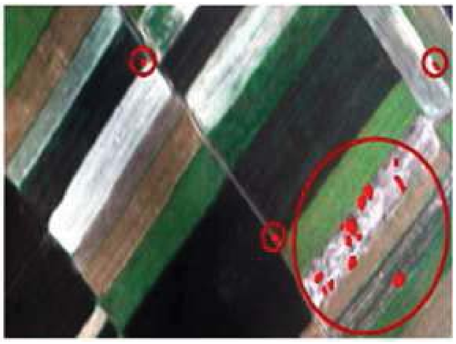

Figure 2: **Manually identified ground-truth anomalies**, marked in red and encircled by ellipses.

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

In Fig. 4, we compare NG-BEVA to the algorithm proposed in [1], RX [10], GMM-RX [11] and FastMCD [12], 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 (ground truth) 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 the other examined algorithms. The time of computation of NG-BEVA is slightly higher than BEVA, twice smaller than RX and larger than GMM-RX or FastMCD by a factor of six.

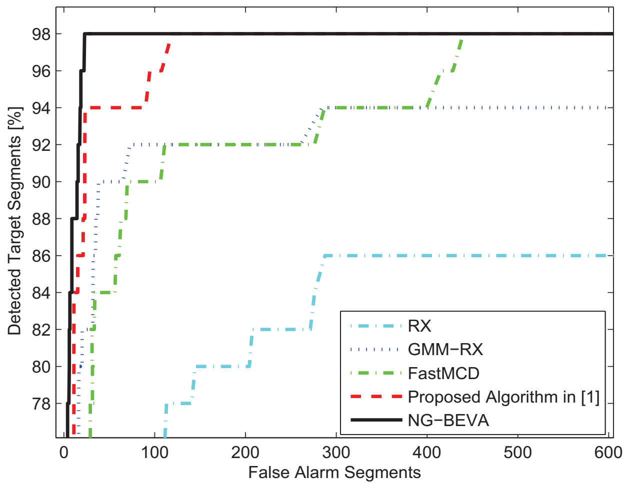

Figure 4: **ROC curves** corresponding to NG-BEVA, RX, GMM-RX, FastMCD and the algorithm proposed in [1].

4. CONCLUSION

In this work we presented an anomaly detection algorithm, denoted NG-BEVA, 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 intimately adjust to the background. The Gaussian model, although efficient and mathematically tractable, is not sufficiently adequate to represent real hyperspectral data. Thus, no Gaussianity assumption is made on the background cluster model in NG-BEVA.

In the local part of NG-BEVA, the hyperspectral image is partitioned into distinct local blocks. The data of each block is clustered into a fixed number of "coarse" clusters using Spectral Clustering, a non-Gaussian clustering method. (An adaptive selection of the number of clusters for each local block may be a subject for future research). Then, the local background is modeled using a greedy iterative estimation process that applies background cluster hypothesis testing based on extreme value theory results. The Mahalanobis distances of the background cluster pixels to the cluster centroid are assumed to have a Gamma distribution instead of a Chi-squared distribution used in the Gaussian case. Thus, by fitting a Gamma distribution to the Mahalanobis distances, we relaxed the Gaussianity assumption. Then, in the global part, the number of false alarms is reduced by comparing the found local anomalies to a "dictionary" of background local models in a large image area. Local anomalies that are too close to a background cluster centroid in any other block, are eliminated.

In experiments with real hyperspectral image cubes the proposed algorithm was shown to have a better performance than RX [10], GMM-RX [11], FastMCD [12] and the algorithm proposed in [1].

REFERENCES

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

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

[3] 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.

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

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

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

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

[8] E. Madar, *Combined Local-Global Background Modeling for Anomaly Detection in Hyperspectral Images*, MSc. Thesis, Technion, Israel Institute of Technology, Dec. 2010, *http://sipl.technion.ac.il/siglib/FP/Eyal-Madar_Thesis.pdf*.

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

[10] 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.

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