HYPERSPECTRAL CHANNEL REDUCTION FOR LOCAL ANOMALY DETECTION

Oleg Kuybeda, David Malah, and Meir Barzohar

Department of Electrical Engineering,
Technion IIT,
Haifa 32000, Israel
email:
koleg@techunix.technion.ac.il, malah@ee.technion.ac.il, meirb@visionsense.com

ABSTRACT

In this work we propose a novel technique for channel reduction in hyperspectral images designed to improve performance of local anomaly detection algorithms. The channel reduction is performed by replacing subsets of adjacent hyperspectral bands by their means. An optimal partition of hyperspectral bands is obtained by minimizing the Maximum of Mahalanobis Norms (MXMN) of errors, obtained due to missrepresentation of hyperspectral bands by constants. By minimizing the MXMN of errors, one reduces the anomaly contribution to the errors, which allows to retain more anomaly-related information in the reduced channels. We demonstrate that the proposed algorithm produces better results, in terms of the ROC curve of a benchmark anomaly detection algorithm (RX) - applied after the dimensionality reduction, as compared to two other dimensionality reduction techniques, including PCA.

1. INTRODUCTION

In this work we propose a novel unsupervised technique for Designing Multispectral Filters that facilitates an improved performance of local anomaly detection algorithms. The proposed approach is based on processing a sample hyperspectral image of a typical scene that is likely to be faced by anomaly detection algorithms. Here, the problem of Multispectral Filters design is formulated as a problem of Redundancy Reduction in Hyperspectral channels, which is performed by replacing adjacent spectral bands by their means. This is a real-world Redundancy Reduction problem that requires preservation of anomalies.

The wealth of spectral information in hyperspectral images provides plentiful amount of data for classification tasks. One such task relates to anomaly detection, in which hyperspectral pixels have to be classified into either background material spectra class or anomaly material spectra class. Generally, anomalies are defined with reference to a model of the background, i.e., the anomaly pixels are those that are not well-described by the background model. Background models are developed using reference data from either a *local* neighborhood of the test pixel or a large (*global*) region of the image. Both approaches have their merits and drawbacks [1].

A common problem of local anomaly detection algorithms is so-called *Hughes phenomenon* [3], according to which the perfomance of anomaly detection algorithms significantly deteriorates when the number of pixels is severely limited for an accurate learning of the local background models. In order to alleviate the effect of this phenomenon, one has to reduce the number of hyperspectral bands, since the complexity of background models is proportional to the hyperspectral data dimensionality.

Linear Dimensionality Reduction (LDR) is a widely used preprocessing technique for the alleviation of *Hughes phenomenon* in classification problems [4], [5]. LDR also allows to eliminate redundancies occuring due to high correlations among adjacet bands. Of particular interest are techniques that reduce the dimensionality of hyperspectral data by replacing subsets of adjacent bands by their means, since the resulting features can be physically interpreted as responses of *multispectral* filters, which may be tuned to aplication-dependent needs. Thus, the authors of [6] propose top-down and bottom-up algorithms designed to find subsets of bands yielding high Fisher discrimination among classes. In [4] one can find an approach that groups the channels into a partition that increases interclass distance computed on a training set. Another approach, based on dynamic programming, is proposed in [5]. It minimizes the mean squared error of representing all hyperspectral pixels in the image by piece-wise constant spectral segments.

Unfortunately, little attention has been drawn in the literature to channel reduction techniques designed to improve the performance of local anomaly detection algorithms. This problem is of high importance in applications that seek a technology to construct high performance multispectral filters for anomaly detection. An appealing approach for this purpose is proposed in [5], denoted as Fast Hyperspectral Feature Reduction (FFR). It looks for a best piece-wise constant represenation of the hypersectral data and does not assume any prior knowledge about the data. However, FFR is not well-taylored to data that contains anomalies, since it uses the mean squared error based (ℓ_2-norm based) criterion. As discussed in [7], this criterion is known to be insensitive to anomaly contributions and, as a result, may lead to a poor represention of anomalies. Moreover, the mean-squared error based criterion is biased to represent better background contributions, since they majorate the contributions of anomalies in the ℓ_2-sense. This may come in contradiction to the goal of anomaly detection-oriented channel reduction that should be designed to retain anomaly manifestations in the data.

In this work we propose a novel approach based on the *Mahalanobis norm* [2] which, unlike the ℓ_2-norm, is not dependent on the scale and/or abundace of measurements. *Mahalanobis norm* is widely used in anomaly detection-related literature [9]. It is also known as a good measure to assess multivariate normality [2], [10]. Both these virtues make the proposed approach to be better-taylored to data that may contain anomalies. I.e., on one hand, the algorithm should be aware of anomalies, if they are present in the hyperspectral data on which the multispectral filters are trained. Such "awareness" can be naturally supported by the anomaly-detection ability of the *Mahalanobis norm*.

On the other hand, if there are *no anomalies* in the training data, then one still needs to make an appropriate tradeoff during the allocation of spectral intervals based on the background information only. In this case, it is desirable to make a coarse partition in spectral regions in which the background behavior is too noisy and is difficult to predict. Thus, these spectral regions should be suppressed, since even if anomaly contributions were present there, they would be likely to be masked by the background. Whereas spectral regions, in which the background behavior is more predictable, should obtain a more dense partition. In these regions anomaly contributions (if they are present) are easier to be captured by anomaly detection algorithms. Therefore, we empoy the ability of the *Mahalanobis norm* to assess multivariate normality in order to specify spectral regions in which the background is more "Gaussian", and, therefore, is less predictable. In the sequel, we refer to these regions as *background clutter*.

More specifically, the optimal partition of the spectrum is obtained by *Minimizing the Maximal Mahalanobis Norm* of errors, obtained due to the misrepresentaion of spectral intervals by constants. Therfore, we denote the proposed technique as Min-Max MN or, in short, MXMN.

We compare MXMN with other dimensionality reduction techniques, such as classical principal components analysis (PCA) and FFR, by examining the results of the Reed-Xiaoli (RX) algorithm [9], [12], a benchmark anomaly detector for hyperspectral imagery, applied after the dimensionality reduction.

We demonstrate that the proposed approach results in a better ROC curve, as compared to PCA and FFR, for a wide range of false alarm rates, and even better than obtained by applying RX on the original data (without the dimensionality reduction) for the important range of low false-alarm rates.

This paper is orgainzed as follows: In section 2 we develop the proposed MXMN algorithm. Then, in section 3, we compare the results of the RX algorithm applied on multispectral data obtained by MXMN, FFR and Principal Component Analysis (PCA) algorithms and discuss the obtained results. Finally, in section 4, we draw conculsions about the proposed method and the obtained results.

2. ANOMALY PRESERVING PIECEWISE CONSTANT REPRESENTATION

2.1 Problem statement

Let $x_{i,j}$ denote the ith hyperspectral band of an observed hyperspectral pixel j, where $i = 1,\dots,M$ and $j = 1,\dots N$. The piecewise constant representaion model consists of a vector of $K < M$ breakpoints,

$$\mathbf{b}_K \triangleq \{b_1,\dots,b_K\}, \tag{1}$$

corresponding to $K-1$ contiguous intervals

$$I_k = [b_k, b_{k+1}), \qquad k = 1,\dots,K-1. \tag{2}$$

Each observed hyperspectral pixel $\mathbf{x}_j$ is approximated by a set of constants $\{\mu_{k,j}\}_{k=1}^{K-1}$, obtained by averaging its values in the spectral intervals I_k, as follows:

$$\mu_{k,j} = \frac{1}{|I_k|}\sum_{i\in I_k} x_{i,j}, \qquad k = 1,\dots,K-1 \tag{3}$$

where $|I_k|$ denotes the cardinality of the interval I_k. As a matter of fact, the constants $\{\mu_{k,j}\}$ minimize the mean squared error $S_{k,j}$ in each interval k, defined as follows:

$$S_{k,j} = \sum_{i\in I_k}(x_{i,j} - \mu_{k,j})^2. \tag{4}$$

Thus, the partition of spectral bands into $K-1$ intervals by the breakpoints $\mathbf{b}_K$ uniquely determines the piecewise constant representation/approximation of each pixel. The goal is to determine a partition that facilitates good performance of anomaly detection algorithms when applied to the obtained constants $\{\mu_{k,j}\}$.

2.2 Objective function

The general idea of the proposed anomaly preserving channel reduction algorithm is to minimize an objective function $J(\mathbf{b}_K)$ that penalizes partitions which may potentially lead to the loss of anomalies during the channel reduction process. We choose the function $J(\mathbf{b}_K)$ to be of the following form:

$$J(\mathbf{b}_K) = \max{}_{k=1}^{K-1} D_k, \tag{5}$$

where by D_k we denote the Potential Anomaly Loss (PAL) measure correspoding to the interval I_k. Thus, by minimizing $J(\mathbf{b}_K)$, one minimizes the worst case PAL measure.

In order to properly define the PAL measure, D_k, let's explore statistical properties of the errors $e_{i,j,k}$ obtained due to the misrepresentation of hyperspectral pixel entries belonging to the interval I_k:

$$e_{i,j,k} = x_{i,j} - \mu_{k,j}, \qquad i \in I_k. \tag{6}$$

Denoting all error entries that belong to the same pixel j and correspond to an interval I_k, ordered in a vector form, by $\mathbf{e}_{j,k}$, we assume that all random vectors $\mathbf{e}_{j,k}$ corresponding to the non-anomalous (background) vectors are i.i.d. At this point, we observe that anomaly manifestations in an interval k, which were not represented well by the corresponding constants $\mu_{k,j}$, are likely to produce anomalous error realizations. Eventually, anomalous error realizations are those that do not agree well with the pdf of the background-related errors $\mathbf{e}_{j,k}$. Therefore, D_k, as a PAL measure, should measure the deviation of the obtained error statistics from a background statistical model. Now, if one models the background-related errors $\mathbf{e}_{j,k}$ by a zero-mean Gaussian pdf, then D_k can be obtained by measuring the deviation of error realizations from the Gaussian model.

This approach is quite reasonable, since the larger is the deviation of the error statistics from being Gaussian, the more signal structure is absorbed by the error and the larger is the likelihood that some important information is lost by channel reduction. A widely used criterion for anomaly detection is the Mahalanobis distance between a tested pixel and the background mean vector [1], [9]. This criterion has also been extensively used for assessing multivariate normality [10]. For a *zero mean* Gaussian random vector $\mathbf{e}$, the Mahalanobis distance or, equivalently, the Mahalanobis norm is defined as:

$$G(\mathbf{e}) \triangleq \sqrt{\mathbf{e}^{\top}\mathbf{\Sigma}^{-1}\mathbf{e}}, \quad (7)$$

where $\mathbf{\Sigma}$ is the covariance matrix of the random vector $\mathbf{e}$.

Intuitively, the Mahalanobis norm of vectors $\mathbf{e}$ that contain outlying signal contributions and, therefore, are not properly normalized by $\mathbf{\Sigma}^{-1}$ in (7), is expected to be larger than obtained for vectors that obey the Gaussian paradigm. Thus, in the Reed-Xiaoli (RX) algorithm [9], a benchmark anomaly detector for hyperspectral imagery, the Mahalanobis distance between a tested pixel and the background mean vector is used to detect anomalies by comparing it to a threshold.

It turns out that if the realizations $\mathbf{e}_j$ are contaminated by anomaly or other Non-Gaussian signal contributions, they are likely to produce large Mahalanobis norms. Therefore, we define D_k as follows:

$$D_k \triangleq \max{}_{j=1}^{N}\, G(\mathbf{e}_{j,k}) \quad (8)$$

This completes the definition of the objective function $J(\mathbf{b}_K)$ in (5) that penalizes partitions that may cause a PAL.

2.3 Minimizing the objective function

In order to minimize $J(\mathbf{b}_K)$, over the set of breakpoints $\{b_1,\ldots,b_K\}$, we apply a dynamic programming algorithm based on [5] and [11]. Let's redefine D_k as $D_{[g,h]}$, where g and h are interval boundaries which can be equivalently used to specify intervals instead of using their corresponding indices $\{k\}$. Throughout the minimization process, we iteratively calculate $J(k,p)$, where $J(k,p)$ is the objective function defined using only the first k breakpoints , $\{b_1,\ldots,b_k\}$, $1<k\leq K$, and the first p spectral bands, $(k-1)\leq p\leq(M-K+k)$.

Initially, we set

$$J(2,p) = D_{[1,p]}, \qquad p=1,\ldots,(M-K+2). \quad (9)$$

I.e., in the first level of recursion we set an objective function value for all possible combinations of allocating only one spectral interval corresponding to its all possible right boundary positions denoted here by p.

Then, for an increasing number $k=3,\ldots,K$, of breakpoints $\{b_k\}$, we define a recursion that calculates the correspoding minimal objective function values $J(k,p)$ as follows:

$$J(k,p) = \min{}_{r=k-1}^{p-1}\, (D_{[r+1,p]} + J(k-1,r)). \quad (10)$$

I.e., $\{J(k,p)\}_{k=3}^{K}$ minimize the sum of PAL measure $D_{[r+1,p]}$ corresponding to all possible variants of allocating the last interval $I_{k-1,k}$, and the minimal objective function value obtained in a previous recursion level $k-1$.

At the end of the iterative process, the resulting $J(K,M)$ gives the minimal value of the objective function $J(\mathbf{b}_K)$ defined in (5). The optimal partition in terms of the breakspoints $\{b_1,\ldots,b_K\}$ is obtained by recursively backtracking the minimizers r^* for which the minimal objective functions were obtained throughout the recursion sequence

$$\{J(K,M), J(K-1,r_K^*),\ldots,J(2,r_3^*)\}. \quad (11)$$

3. EXPERIMENTS WITH REAL DATA

In this section we evaluate the performance of the RX algorithm, which, as mentioned, is a benchmark anomaly detector for Hyperspectral Imagery [9]. We applied it to Hyperspectral Data before and after the dimensionality reduction by PCA, FFR and the proposed MXMN algorithm. To demonstrate the results, the RX algorithm was applied to 6 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, a pixel resolution corresponds to $(0.8m)^2$. The obtained image cubes are $b\times r\times c = 65\times 300\times 479$ hyperspectral images, where b,r and c denote the number of hyperspectral bands, the number of rows and the number of columns in the image, respectively.

In Fig. 1, we show the 30th band of a typical hyperspectral image cube. The image contains ground-truth anomalies (vehicles and small agriculture facilities, which occupy a few pixel segments marked in white and encircled by red ellipses), manually identified using side information collected from high resolution RGB images of the corresponding scenes, as shown in Fig. 2. All 6 images are not shown here just because of space limitations.

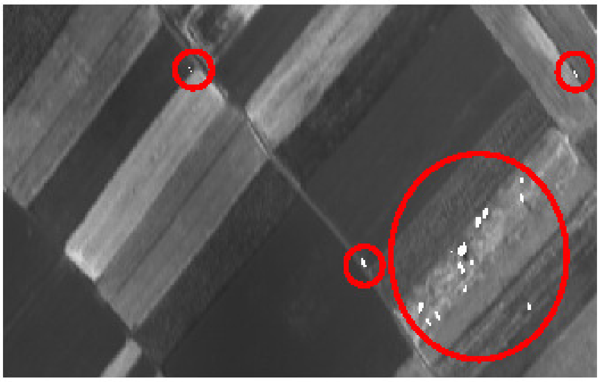

Figure 1: 30th band of a hyperspectral image cube with anomalies marked in white and encircled by red ellipses.

Figure 2: An example of high resolution RGB image used for manual target identification.

We applied FFR and the proposed MXMN algorithms to an image cube (the 30th band of which is shown in Fig. 3) that *does not contain anomalies* to reduce the hyperspectral dimensionality from 65 to 10 by the corresponding piecewise constant spectral segments. We also applied PCA to obtain an ℓ_2 optimal 10-dimensional basis.

Figure 3: 30th band of a hyperspectral image cube used for training MXMN.

In Fig. 4, one can see the obtained piece-wise constant approximations by FFR (cyan (bright) thick line) and MXMN (blue (dark) thick line) for 3 selected hyperspectral pixels (blue thin lines). The uppermost graph corresponds to anomaly pixel, whereas the other two lower graphs correspond to pixels that were selected from different background regions. As can be seen from the figure, the partition obtained by MXMN has a denser granularity in bands $[1-35]$, in which the anomaly is expressed. This is on the expense of other bands, which, in spite of being energetically prominent, are less important for anomaly detection, since, as it turns out, they correspond to the background clutter. This observation is also supported by ROC results, which follow below. On the contrary, FFR adapts better to the energetical bands and, as a result, assigns less channels to bands $[1-35]$ which makes it prone to loosing anomalies.

In Fig. 5, we compare FFR, MXMN and PCA in terms of Receiver Operation Characteristic (ROC) curves obtained by applying the RX algorithm on hyperspectral data after the dimensionality reduction. For the purpose of ROC curves generation, all 6 hyperspectral images were used, in which the total number of anomaly segments count is 25. We assume that it is appropriate to include the training image in the evaluation of the anomaly detection performance in terms of ROC, since no side information about anomalies was used during the training. Moreover, the training image did not include any known anomalies.

It is clearly seen from the figure that the MXMN algorithm corresponds to a better ROC curve (blue solid line) compared to other dimensionality reduction techniques such as FFR (cyan dashed line) or PCA (red solid line with solid circles) for all tested parameters. It is important to note that the performance of the RX algorithm applied to the data obtained by the MXMN is even better than applying RX to the full-dimensional (original) images (green dot-dashed line), for the range of low false-alarm rates. This can be explained as follows:

Since $J(\mathbf{b}_K)$ of (5) is designed to favor partitions that produce errors which are "more Gaussian", its minimization using typical data *without anomalies*, results in a coarse partition in the spectral bands containing background clutter. As was already noted above, the background clutter relates to bands that produce "more Gaussian" piece-wise constant approximation errors. Generally, the behavior of such bands is much more difficult to predict by a local statistical model, since the "non-Gaussianity" primarily stems from a non-organized process - neither globally, nor locally. As a result, they may mask out subtle anomaly-related contributions that may appear in them and in other bands. Therefore, background clutter bands are considered to have a lower discriminative power.

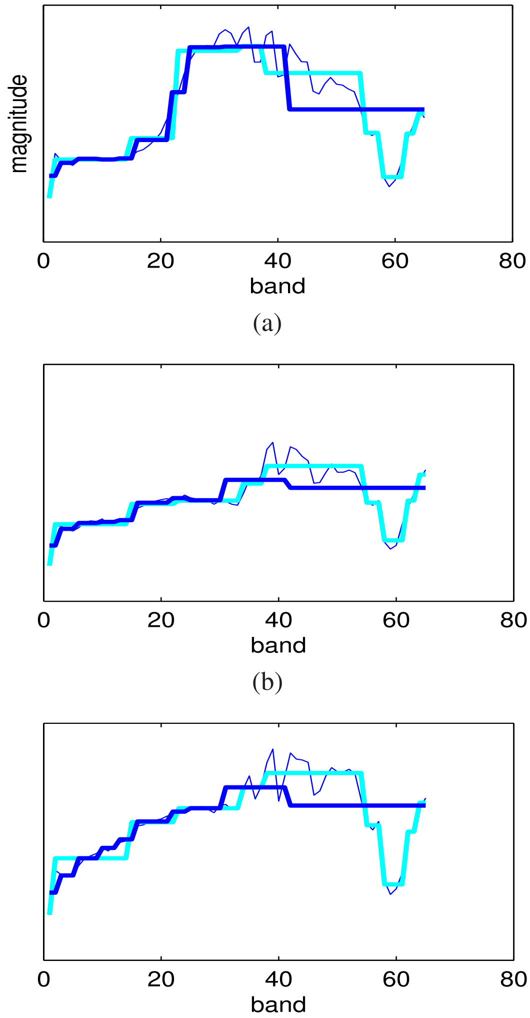

Figure 4: **Piecewise constant approximation.** (a) anomaly pixel; (b) and (c) background pixels. Original spectrum is in blue (dark) thin line, MXMN approximation is in blue (dark) thick line, FFR approximation is in cyan (bright) thick line.

The fine partition is obtained in the spectral bands that are "less Gaussian". Generally, it is easier to model such bands using local statistical models and, therefore, anomaly contributions in these bands may become more apparent to anomaly detection algorithms. This may explain why the proposed algorithm corresponds to a better ROC curve compared to the other algorithms, although the optimal partition was obtained using *an image that does not contain anomalies*.

Nevertheless, it is important to note that for best performance, one should consider training the multispectral filters on an image that contains typical anomalies that are anticipated to be faced in a real situation (if such prior information is available). In this case, the optimal partition will be primarily driven by anomalies in the data. Still, the approach is unsupervised, since, in any scenario, it does not require any ancillary knowledge about anomaly location in the data and/or anomaly spectrum.

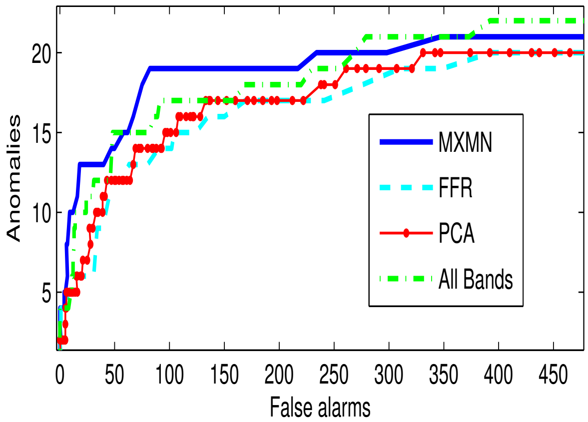

Figure 5: **ROC curves.**

4. CONCLUSION

In this work we proposed a novel multispectral filters design that is tailored to improve the performance of local anomaly detection algorithms. The filter design is based on processing a hypespectral image that contains a typical spectral content to be faced by the multispectral anomaly detection algorithms. The resulting multispectral filters are obtained by replacing subsets of adjacent hyperspectral bands by their means, producing piecewise constant pixel approximations. The optimal partition of hyperspectral bands is obtained by *Minimizing the Maximum of Mahalanobis Norms* (MXMN) of missrepresentation errors. Hence, the proposed algorithm is denoted as MXMN. The minimization of MXMN allows filtering out backgound clutter contributions. At the same time, it enables adaptation to spectral contributions of anomalies (if they are present in the training image). The minimization is performed by a dynamic programming technique, as used by the Fast Hyperspectral Feature Reduction (FFR) algorithm proposed in [5]. MXMN was compared with FFR and PCA by examining the results of the RX algorithm [9] applied after the dimensionality reduction. It was demonstrated that the proposed algorithm results in a better ROC curve in the whole range of false alarm values, and even better than applying RX on the original data without the dimensionality reduction in the important range of low false-alarm rates. The latter can be explained by the MXMN ability to filter out background clutter-related channels, while preserving resolution in those spectral channels in which anomalies are more easy to detect.

REFERENCES

[1] D. Stein, S. Beaven, L. E. Hoff, E. Winter, A. Shaum, and A. D. Stocker "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag., vol. 19, no. 1, pp. 58 – 69, Jan. 2002.*

[2] R. De Maesschalck, D. Jouan-Rimbaud, D.L. Massart "The Mahalanobis distance." *Chemometrics and Intelligent Laboratory Systems, no. 50 pp. 1 – 18, 2000.*

[3] G.F. Hughes "On The Mean Accuracy Of Statistical Pattern Recognizers," *IEEE Trans. Infor. Theory, vol. IT-14, NO. 1 , pp 55 – 63, 1968.*

[4] S. B. Serpico, M. DInca, and G. Moser, "Design of spectral channels for hyperspectral image classification," *Proc. IGARSS, vol. 2, pp. 956 – 959, 2004.*

[5] A. C. Jensen and A. S. Solberg, "Fast Hyperspectral Feature Reduction Using Piecewise Constant Function Approximations," *IEEE Geoscience and Remote Sensing Letters, vol. 4, no. 4, Oct. 2007.*

[6] S. Kumar, J. Ghosh, and M. Crawford, "Best-bases feature extraction algorithms for classification of hyperspectral data," *IEEE Trans. Geoscience and Remote Sensing, vol. 39, no. 7, pp. 1368 – 1379, Jul. 2001.*

[7] 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.*

[8] 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.*

[9] I. S. Reed and Xiaoli Yu "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Process., vol. 38, no. 1, pp. 1760 – 1770, Oct. 1990.*

[10] D. Ververidis and C. Kotropoulos "Gaussian Mixture Modeling by Exploiting the Mahalanobis Distance," *IEEE Trans. Sig. Proc., vol. 56, no. 7., July 2008.*

[11] G. Winkler and V. Liebscher "Smoothers for discontinuous signals," *J. Nonparametric Statist., vol. 14, no. 1/2, pp. 203 – 222, 2002.*

[12] X. Lu, L. E. Hoff, I. S. Reed, M. Chen, and L. B. Stotts "Automatic target detection and recognition in multiband imagery: A unified ML detection and estimation approach," *IEEE Trans. on circ. and syst.II: analog and digital signal processing, vol. 46, no. 1, pp. 143156, Jan. 1997.*

[13] H. Ren and C.I. Chang "Automatic spectral target recognition in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst., vol. 39, pp. 1232-1249, Oct. 2003.*

[14] D. Manolakis and G. Shaw "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Proc. Mag., pp. 29-43, Jan. 2002.*

[15] N.I. Ramey and M. Scoumekh "Hyperspectral anomaly detection within the signal subspace,", *IEEE Trans. Geoscience and Remote Sens. Let., vol. 3, no. 3, pp. 312-316, July 2006.*