IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Global Unsupervised Anomaly Extraction and Discrimination in Hypersprectral Images via Maximum Orthogonal-Complements Analysis

Oleg Kuybeda, David Malah, and Meir Barzohar

CCIT Report #684
February 2008

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

CCIT Report #684 February 2008

Global Unsupervised Anomaly Extraction and Discrimination in Hyperspectral Images via Maximum Orthogonal-Complements Analysis

Oleg Kuybeda, David Malah, and Meir Barzohar

Department of Electrical Engineering

Technion IIT, Haifa 32000, Israel

koleg@techunix.technion.ac.il, malah@ee.technion.ac.il,

meirb@visionsense.com

Abstract

In this paper we address the problem of global unsupervised detection, discrimination, and population estimation of anomalies of the same type, in hyperspectral images. The proposed approach, denoted as Anomaly Extraction and Discrimination Algorithm (AXDA), detects anomalies via analysis of a signal-subspace obtained by the recently developed Maximum Orthogonal Complement Algorithm (MOCA). MOCA is unique in providing an unsupervised combined estimation of signal-subspace that includes anomalies, and its rank. The main idea of AXDA is to iteratively reduce the anomaly vector subspace-rank, making the related anomalies to be poorly represented. This helps to detect them by a statistical analysis of the $\ell_{2,\infty}$-norm of data residuals. As a by-product, AXDA provides also an anomaly-free robust background subspace and rank estimation. We experimentally show that AXDA performs better than other global anomaly detection techniques, such as the Gaussian Mixture Model-based (GMRX) algorithm and the classical Matched Subspace Detector (MSD), in most of the range of the tested parameters. Since MSD requires prior knowledge of anomaly and background subspaces, which are unknown, the MSD was applied to the anomaly subspace obtained from MOCA and the anomaly-free background subspace obtained from AXDA.

Index Terms

Signal-subspace rank, Maximum Orthogonal-Complements Algorithm (MOCA), Unsupervised Anomaly Detection, Unsupervised Anomaly Discrimination, Hyperspectral Images, Background Modelling.

I. Introduction

In this work we address the problem of anomaly detection, as well as anomaly discrimination and population estimation of anomalies of the same type, in hyperspectral image cubes (denoted in this paper as hyperspectral images). The considered sceneries are composed of reflected spectra of abundant natural ground materials such as vegetation, soil, minerals, etc., along with anomalies such as localized man-made objects. E.g., small buildings, vehicles, etc. 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.

Since most often, 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 modelling 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 (e.g., in terms of its order/rank), limiting its ability to adapt to anomalies, and leaving anomalies to disagree with the model, which is essential for a high probability of detection.

A. Background-modelling literature review

A variety of background modelling methods appears in the literature. One type of these methods is based on estimating the underlying probability density function (pdf) of the background signature, and applying a threshold to the likelihood of tested pixels. The Reed-Xiaoli (RX) algorithm [3], is a benchmark anomaly detector for hyperspectral imagery. According to this algorithm, the background pixels in a local neighborhood of a tested pixel are assumed to be independent, identically distributed, Gaussian random vectors. After estimating the background mean vector and covariance matrix, the Mahalanobis distance between the tested pixel and the background mean vector is compared to a threshold to detect an anomaly [3]. Unfortunately, in many environments, it has been shown empirically that local background modelling by a single Gaussian provides an inadequate representation of the underlying distribution [8], leading to poor false alarm performance. This is especially true when the local background contains multiple classes of terrain.

To properly characterize nonhomogeneous backgrounds, researchers have employed a Gaussian Mixture Model (GMM) [6], [8]. This approach models the background signature distribution as a linear

combination of Gaussian distributions. The Gaussian Mixture distribution is applied as a global model since the parameters are estimated over large regions. Anomaly detection may be achieved by applying the generalized likelihood ratio test (GLRT) to the model. The authors of [8] denote the related approach by GMRX. While GMRX provides good performance, it is limited by the simplicity of Gaussian components. GMRX is further limited by the need to know or estimate a priori the number of terrain classes in the image.

Another approach to local background modelling corresponds to the so-called large-margin techniques, such as support vector machines (SVMs), which detect anomalies by directly estimating a decision boundary with maximal separability. The authors of [5] propose to determine the minimal enclosing hypersurface that contains a training set of background data pixels. A training set is sampled from a window enclosing the tested pixel, excluding pixels belonging to its adjacent neighborhood (which makes this method local) that is supposed to be large enough to contain a maximum-size anomaly. The anomaly is detected by thresholding the distance from the tested pixel to the obtained hypersurface.

In this paper we adopt the following linear mixing model:

$$\mathbf{x}_i = \mathbf{A}\mathbf{s}_i + \mathbf{z}_i, \qquad i = 1, \ldots, N, \tag{1}$$

where, $\mathbf{x}_i \in \mathbb{R}^p$ is the observed pixel, $\mathbf{z}_i \in \mathbb{R}^p$ is additive noise, columns of $\mathbf{A}$ are the pure materials spectra (endmembers) and $\mathbf{s}_i \in \mathbb{R}^r$, $r \leq p$, their corresponding abundances. Using this model, authors of [28] propose a class of Generalized Likelihood Ratio Test (GLRT) - based algorithms to detect anomalies. The authors of [28] assume a priori known target signatures, whereas the the background subspace is estimated via SVD for a priori known background subspace rank. Using Orthogonal Subspace Projection, the authors of [4] propose first to select target pixels from the data, then to detect the presence of anomaly signatures in the data pixels. Unfortunately, the approaches of [4] and [28] lack a systematic way to estimate the anomaly and the background subspace dimensionality. Moreover, they require ancillary information to identify signatures belonging to anomalies.

There is a variety of rank determination methods that can be considered for determination of the number of endmembers [9], [29]. The classical methods for rank determination such as MDL and AIC [10] were shown as not reliable when applied to real hyperspectral data [12]. The authors of [12] propose a new virtual dimensionality (VD) concept that is defined as “the minimum number of spectrally distinct signal sources that characterize the hyperspectral data from the perspective view of target detection and classification”. The drawback of the VD concept is that it is based on analyzing an eigen-structure

of the sample covariance and correlation matrices. As experimentally shown in [2], the eigen-structure of high-dimensional data containing energetically (related to ℓ_2-norm) insignificant (rare or anomalous) vectors, is extremely vulnerable to high-dimensional noise. The problem is that such weak contributions, like that of single-pixel anomalies, met in practice, to the sample eigenvalues are typically masked by noise perturbations with probability close to 1. Thus, the probability that anomaly-related eigenvalues will be interchanged with the noise-related eigenvalue is extremely high. Therefore, VD, as well as other ℓ_2-related methods, work only when the anomalies are stronger than noise perturbations in the ℓ_2-norm sense. A recent rank-estimation approach, named MOCA [2], was shown to be much more robust to noise perturbations, since it is based on a norm (denoted as $\ell_{2,\infty}$-norm), which is much more sensitive to individual pixel contributions. The proposed approach for anomaly detection, being based on MOCA, inherits this property.

B. Maximum Orthogonal Complement Algorithm (MOCA) principles

In this paper, we propose to adapt MOCA [2] for anomaly detection and discrimination in hyperspectral images. Originally, MOCA was developed by the authors of this paper for signal-subspace and rank estimation in high-dimensional noisy signals that may contain anomaly vectors, which MOCA intends to preserve. Similar to the approach proposed in [4], MOCA assumes the linear mixing model (1). It is important to note that MOCA assumes $\mathbf{z}_i$ to be Gaussian and *spectrally white* with a known variance. As discussed in Appendix I, this assumption doesn't actually hold in hyperspectral images. Therefore, a preprocessing procedure is proposed in Appendix I for estimation of the noise variance and spectrally whitening it before applying MOCA. The signal-subspace in MOCA is estimated by minimizing the *maximum* of misrepresentation-residual ℓ_2-norms denoted as $\ell_{2,\infty}$-norm. The rank is determined by applying Extreme Value Theory results [22], [23], [25] to model the distribution of the misrepresentation $\ell_{2,\infty}$-norm. Since $\ell_{2,\infty}$ penalizes individual data-vector misrepresentations, it helps to represent well not only background vectors, but also anomaly vectors. As a matter of fact, MOCA proposes an appropriate compromise between the following two approaches: The first approach is based on selecting the signal-subspace basis vectors directly from the data as presented in [4]. This approach is good for representing anomalies, since it is capable of selecting anomalies from the data. However, due to noise in the obtained basis vectors, it may perform poorly in representing background pixels. The second approach is based on SVD, which, as shown in [2], represents well the background pixels. Yet, it may perform poorly in

representing anomalies [2]. Thus, the signal-subspace estimated by MOCA admits the following form:

$$\hat{\mathcal{S}}_k = range\ [\mathbf{\Psi}_{k-h}|\mathbf{\Omega}_h]\,, \tag{2}$$

i.e., $\hat{\mathcal{S}}_k$ is the space linearly spanned by columns of matrices $\mathbf{\Omega}_h$ and $\mathbf{\Psi}_{k-h}$, where $\mathbf{\Omega}_h$ is a matrix composed of h linearly independent columns selected from the data, k is the estimated signal rank and $\mathbf{\Psi}_{k-h}$ is a matrix with $k-h$ orthogonal columns, obtained via SVD of the data residuals $\mathcal{P}_{\mathbf{\Omega}_h^{\perp}}\mathbf{x}_i$, i=1,...,N, where $\mathcal{P}_{\mathbf{\Omega}_h^{\perp}}$ is a projection onto $(range\ \mathbf{\Omega}_h)^{\perp}$. This notation is equivalent to $\mathcal{P}^{\perp}_{\mathbf{\Omega}_h}$ used in [4]. The matrix $\mathbf{\Omega}_h$ provides a natural way to identify and discriminate anomalies, since it collects individual data pixels that are poorly represented by SVD and, therefore, are considered anomalous (see [2] for details). Whereas, columns of $\mathbf{\Psi}_{k-h}$ span the background pixels projections lying in the $(range\ \mathbf{\Omega}_h)^{\perp}$.

C. Anomaly detection via MSD

At first glance, once MOCA estimates the anomaly and background subspaces, one may apply to the result the classical Matched Subspace Detection (MSD) [1] for detection of anomalies. This method is widely used in the literature for anomaly detection in hyperspectral images when anomaly and background subspaces are known in advance (see for example references [13], [14], [15], [16], and there exist many more).

According to the MSD method, two hypotheses are defined:

$$H_0 \ : \ \mathbf{x}_i \sim \mathcal{N}[\mathbf{B}\mathbf{b}_i, \sigma^2\mathbf{I}], \tag{3}$$

$$H_1 \ : \ \mathbf{x}_i \sim \mathcal{N}[\mathbf{B}\mathbf{b}_i + \mathbf{T}\boldsymbol{\theta}_i, \sigma^2\mathbf{I}], \tag{4}$$

where $\mathcal{N}$ denotes the normal distribution and σ corresponds to the noise std; $\mathbf{B}$ and $\mathbf{T}$ are background and anomaly subspace bases with $\mathbf{b}_i$ and $\boldsymbol{\theta}_i$ background and anomaly subspace expansion coefficients of data vector $\mathbf{x}_i$, respectively. The matrices $\mathbf{B}$ and $\mathbf{T}$, comprising the signal-subspace basis, are not necessarily orthogonal each other (i.e. $\mathbf{B}^T\mathbf{T} \neq 0$), but they are linearly independent, meaning that there is no element in $\mathbf{B}$ that can be represented as a linear combination of vectors in $\mathbf{T}$. The hypothesis H_0 corresponds to the case in which the observed vector is drawn from the interference/background subspace, contaminated by white Gaussian noise. Whereas, the hypothesis H_1 corresponds to the case in which the observed vector is a superposition of a vector from the interference/background subspace and a vector from the anomaly subspace, contaminated by white Gaussian noise.

The Generalized Log-Likelihood Ratio (GLR) is given by

$$L(\mathbf{x}) = \frac{1}{\sigma^2}\mathbf{x}^T\mathcal{P}_{\mathbf{B}^\perp\mathbf{T}}\mathbf{x}, \tag{5}$$

where $\mathcal{P}_{\mathbf{B}^\perp\mathbf{T}}$ is a projection onto $(range\ \mathbf{B})^\perp \bigcap range\ \mathbf{T}$ - a low-rank anomaly-matched subspace, such that interference contribution contained in $range\ \mathbf{B}$, and noise contribution contained in both $range\ \mathbf{B}$ and $(range\ \mathbf{T})^\perp$, are removed. This filter is usually called a *matched subspace filter* or a *matched field filter*. The energy of the filter output (corresponding to $L(\mathbf{x})$) is computed and compared to a threshold. The problem with this approach is that it is not well-adapted to the $\ell_{2,\infty}$ optimality criterion of the signal-subspace basis $[\mathbf{\Psi}_{k-h}|\mathbf{\Omega}_h]$ found by MOCA. Although, according to MOCA, the maximum of the data residual norms $\eta = \|\mathcal{P}_{[\mathbf{\Psi}_{k-h}|\mathbf{\Omega}_h]^\perp}\mathbf{X}\|_{\ell_{2,\infty}}$ is minimized, it may still leave large residuals (with norms below η) belonging to $(range\ [\mathbf{\Psi}_{k-h}|\mathbf{\Omega}_h])^\perp$. Obviously, these residuals don't contribute to $L(\mathbf{x}_i)$, which measures the norms of data vector projections onto $(range\ \mathbf{\Psi})^\perp \bigcap range\ \mathbf{\Omega}$ (see (5) above). This reduces the probability of anomaly detection by MSD in cases of anomalies having residual norms below η. Note, that the value of η is determined by statistics of the maximum-norm noise realization that has a narrow distribution [2] centered around a value that is not insignificant. This value is a function of σ and the noise subspace rank.

Another disadvantage of using MSD in conjunction with a subspace determination by MOCA, is that the anomaly subspace basis $\mathbf{\Omega}$ found by MOCA is composed of vectors that were directly selected from the data, whereas the background subspace basis satisfies $range\ \mathbf{\Psi} \subset null\ \mathbf{\Omega}^T$. Therefore, the estimated anomaly subspace, as well as the background subspaces are deflected by noise. Since in hyperspectral images the background subspace and anomaly subspace are far from being orthogonal, even small deviations of the anomaly and background subspace estimations may cause background vectors to have a strong contribution to $L(\mathbf{x}_i)$ of (5), which rapidly increases false alarm rate by MSD. This observation is experimentally substantiated in section IV.

D. Proposed algorithm outline and paper organization

The proposed algorithm, denoted here as Anomaly Extraction and Discrimination Algorithm (AXDA), is based on using the background and anomaly subspace estimates by MOCA and is designed to cope with the above MSD drawbacks. The key-point of the proposed algorithm is that it iteratively modifies both $\mathbf{\Omega}$ and $\mathbf{\Psi}$. The modification is performed by removing columns from the matrix $\mathbf{\Omega}$, one at a time, and updating the matrix $\mathbf{\Psi}$ to match the modified $\mathbf{\Omega}$. This significantly reduces the effect of noise on the anomaly detection process. AXDA uses the $\ell_{2,\infty}$-optimality criterion of MOCA to extract all anomaly

pixels belonging to the same anomaly endmember, where anomaly endmembers correspond to columns of $\mathbf{\Omega}_h$. Thus, AXDA combined with MOCA, allows determination of the number of anomalies and the extraction of all pixels belonging to the same type in an unsupervised way. It still applies Extreme Value Theory (EVT) [22] to model the $\ell_{2,\infty}$-norm to construct a sharp, robust and adaptive anomaly detector, which doesn't rely on any prior knowledge about the dimensionality or statistical model of the background, without the need for tuning a one-sided hypothesis threshold, and without any prior knowledge about the number of anomaly classes and/or anomaly endmembers.

This paper is organized as follows: Section II provides a short overview of MOCA [2], which is designed to estimate a signal-subspace defined to include *anomalies*. In section III, we describe the proposed Anomaly Extraction and Discrimination Algorithm (AXDA), which employs signal-subspace and rank estimation results obtained by MOCA for detection and discrimination of anomalies. In section IV, we demonstrate results of applying AXDA to real hyperspectral images. We also show there a comparison of AXDA vs. GMRX [8] and MSD algorithms in terms of Receiver Operating Curves (ROC) obtained by applying the algorithms on 5 hyperspectral images. Finally, in section V, we conclude this work.

II. Maximum Orthogonal Complement Algorithm (MOCA)

In this section we provide a short overview of MOCA, proposed in [2], which is designed to estimate the signal-subspace that is defined to include *anomaly vectors*.

Thus, according to [2], given the signal-subspace rank, k, the signal-subspace estimation $\hat{\mathcal{S}}_k$ that preserves anomaly vectors is optimal in the following sense:

$$\begin{aligned} \hat{\mathcal{S}}_k &= \underset{\mathcal{L}}{\operatorname{argmin}} \ \|\mathcal{P}_{\mathcal{L}^\perp}\mathbf{X}\|^2_{\ell_{2,\infty}} \\ &\quad \text{s.t.} \qquad \operatorname{rank} \mathcal{L} = k, \end{aligned} \tag{6}$$

where $\mathcal{P}_{\mathcal{L}^\perp}$ denotes an orthogonal projection onto $\mathcal{L}^\perp$, $\mathbf{X}$ is a matrix of observed data vectors ordered as its columns, and $\|\mathbf{H}\|_{\ell_{2,\infty}}$ corresponds to the norm of $\mathbf{H}$, defined as the *maximum* ℓ_2 norm of $\mathbf{H}$ columns.

In order to make the minimization of (6) computationally plausible, a suboptimal solution is proposed in [2], according to which the sought basis of $\hat{\mathcal{S}}_k$, denoted by $\mathbf{\Phi}$, is constrained to the following form:

$$\mathbf{\Phi} = [\mathbf{\Psi}_{k-h}|\mathbf{\Omega}_h]\,, \tag{7}$$

where $\mathbf{\Omega}_h$ is a full-rank matrix composed of h columns selected from $\mathbf{X}$, and $\mathbf{\Psi}_{k-h}$ is a matrix with $k-h$ orthogonal columns, obtained via SVD of $\mathcal{P}_{\mathbf{\Omega}_h^\perp}\mathbf{X}$. The main idea of this approach is to collect anomaly

vectors into $\mathbf{\Omega}_h$ in order to directly represent the anomaly vectors subspace, whereas the background vectors subspace is found by applying SVD in the null-space of $\mathbf{\Omega}_h^T$.

The signal-subspace rank estimator is based on examining the maximal data residual norms $\|\mathcal{P}_{\hat{\mathcal{S}}_k^\perp}\mathbf{X}\|_{\ell_{2,\infty}}^2$, for an increasing sequence of rank values k. It is assumed that for some k, $r_q \leq k \leq r$, the signal-subspace $\hat{\mathcal{S}}_k$ is close to the subspace of background vectors, which is of rank r_q. This assumption is plausible due to the SVD-part of the signal subspace determination that is designed to represent well the background vectors subspace. As a result, the background vector residuals in the complementary subspace $\hat{\mathcal{S}}_k^\perp$ are governed by the noise contribution, whereas the anomaly vector residuals in $\hat{\mathcal{S}}_k^\perp$ may still include significant signal contributions. Thus, for $k \geq r_q$, the set of all data-vector residual norms can hypothetically be divided into two subsets as follows:

$$\begin{aligned} \Gamma_k &\triangleq \{\text{squared norms of background vector residuals}\} \\ \Delta_k &\triangleq \{\text{squared norms of the remaining data-vectors}\}. \end{aligned} \tag{8}$$

Once the value of the maximum data-residual squared-norm

$$\eta_k = \max_{j=1,..,N} \|\mathcal{P}_{\hat{\mathcal{S}}_k^\perp}\mathbf{x}_j\|^2 \tag{9}$$

becomes available (with N denoting the number of pixels), the following two hypotheses are formulated:

$$H_0 \quad : \quad \eta_k \text{ belongs to } \Gamma_k, \tag{10}$$

$$H_1 \quad : \quad \eta_k \text{ belongs to } \Delta_k. \tag{11}$$

The rank estimator $\hat{r}$ is set to be equal to the minimal value of k for which the following condition is satisfied:

$$p(H_0|\eta_k) \geq p(H_1|\eta_k), \tag{12}$$

which means that the optimal rank is reached when there is a higher likelihood that the maximum data-residual squared norm η_k is governed by the noise statistics (i.e., it doesn't include significant signal contributions). The details of how to calculate $p(H_0|\eta_k)$ and $p(H_1|\eta_k)$ are found in [2].

To summarize, the flowchart of MOCA is presented in Fig. 1, according to [2]. The algorithm begins with an initial guess for the signal-subspace rank, such as $k = 1$. At each rank-value iteration, the signal-subspace basis $\mathbf{\Phi}_k = [\mathbf{\Psi}_{k-h}|\mathbf{\Omega}_h]$ for the conjectured rank k, is obtained via a Min-Max SVD

(MX-SVD) algorithm (introduced in [2]) producing $\mathbf{\Omega}_h$ - a matrix composed of h linearly independent columns selected from $\mathbf{X}$, and $\mathbf{\Psi}_{k-h}$ - a matrix with $k-h$ orthogonal columns obtained via SVD of $\mathcal{P}_{\mathbf{\Omega}^\perp}\mathbf{X}$. Then, the data maximum residual-norm is calculated in the $(range\ \mathbf{\Phi}_k)^\perp$. This norm is tested in order to decide if it belongs to the noise hypothesis (this decision is performed by evaluating inequality (12)). If the noise hypothesis passes, the algorithm is terminated, and the estimated signal-subspace and rank equals to the range of the last obtained $\mathbf{\Phi}_k$, and to the last value of k, respectively. Otherwise, the conjectured rank value k is incremented and a new iteration is carried out.

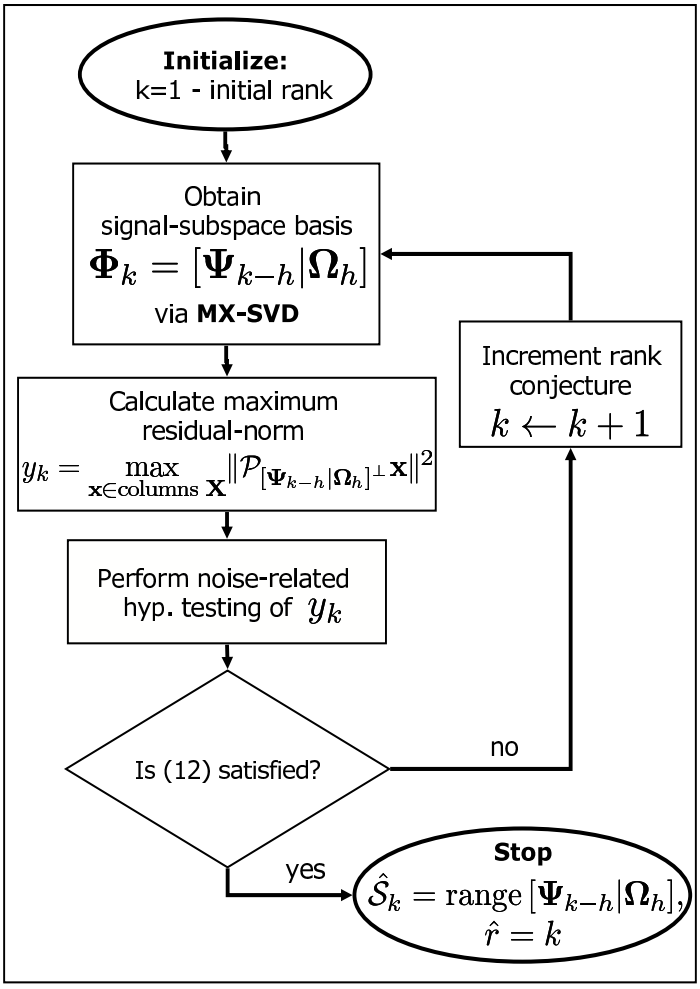

Fig. 1. **Maximum Orthogonal Complement Algorithm (MOCA) flowchart (after [2]).**

III. Anomaly Extraction and Discrimination Algorithm (AXDA)

In this section we propose an Anomaly Extraction and Discrimination Algorithm (AXDA) that employs signal-subspace and rank estimation results obtained by the above described MOCA.

Let's recall that the signal-subspace basis $\mathbf{\Phi}$ produced by MOCA admits the following form:

$$\mathbf{\Phi}_{\hat{r}} = [\mathbf{\Psi}_{\hat{r}-h} | \mathbf{\Omega}_h] \,, \tag{13}$$

where $\hat{r}$ is the estimated signal-subspace rank, the sub-matrix $\mathbf{\Omega}_h$ consists of h linearly independent columns selected from the data matrix $\mathbf{X}$ and the sub-matrix $\mathbf{\Psi}_{\hat{r}-h}$ consists of $\hat{r}-h$ principal components of $\mathcal{P}_{\mathbf{\Omega}^\perp}\mathbf{X}$.

As it was already noted above, the matrix $\mathbf{\Omega}_h$ represents the anomaly vectors subspace. First of all, given the matrix $\mathbf{\Omega}_h$, one can mark anomaly vectors by locating indices of $\mathbf{\Omega}_h$ columns in the original data matrix $\mathbf{X}$. However, this straightforward method does not enable us to find all the pixels in the data that belong to the anomaly subspace, since not all anomalies are guaranteed to be within the columns of $\mathbf{\Omega}_h$. For example, in a case where there are number of vehicles having the same anomalous reflected spectrum, only one pixel representing all vehicle pixels would be collected by MOCA into $\mathbf{\Omega}_h$. Therefore, neither all vehicles, nor all vehicle pixels would be marked by this straightforward method. It was experimentally observed that simple approaches such as looking for data vectors lying close enough to $\mathbf{\Omega}_h$ columns (or, alternatively, to the subspace spanned by $\mathbf{\Omega}_h$ columns) are of a low practical value due to the need for a threshold and due to a high false-alarm rate caused by background interference.

In order to detect and discriminate all anomaly pixels, we propose a new algorithm that extracts all anomalies in the data and associates them with the $\mathbf{\Omega}_h$ columns found by MOCA. As stated earlier, the proposed algorithm is denoted as Anomaly Extraction and Discrimination Algorithm (AXDA). For the sake of clarity, we first present a concise outline of AXDA in Fig. 2.

A. Concise outline of AXDA

The main idea of the algorithm is to iteratively reduce the anomaly vector subspace-rank by dropping columns of $\mathbf{\Omega}_h$, producing submatrices $\{\mathbf{\Omega}_j\}_{j=0}^{h-1}$. Since for a given rank $\hat{r}$, the matrix $[\mathbf{\Psi}_{\hat{r}-h}|\mathbf{\Omega}_h]$ minimizes the $\ell_{2,\infty}$ of data residuals in the $(range\ [\mathbf{\Psi}_{\hat{r}-h}|\mathbf{\Omega}_h])^\perp$ (as noted above), dropping columns from $\mathbf{\Omega}_h$ increases the $\ell_{2,\infty}$-norm of data residuals. Obviously, this change in residual norms occurs in pixels that are well-represented by the dropped column, including the residual norm of the dropped column itself. Therefore, this operation reveals anomaly vectors in the data that belong to the dropped column by increasing their residual norms. The increased residual norms are compared to the $\ell_{2,\infty}$-norm of data residuals from the previous iteration, which are determined by the test in (12) as stemming from noise. If the increased norms exceed the $\ell_{2,\infty}$-norm of data residuals from the previous iteration, the corresponding pixels are marked as belonging to the dropped column and are depleted from data.

Depletion of such pixels makes the $\ell_{2,\infty}$-norm of data residuals in the current iteration to pass again the noise hypothesis in (12). All operations in this paragraph are performed in block (2) of Fig. 2.

There are two indices, j and s that keep track of anomaly subspace and *total* signal subspace dimensionality, respectively, at each iteration. The index j, which is initialized as $j = h$, denotes the anomaly subspace rank throughout the AXDA iterations. It is decremented by one at each iteration. The index s (initialized as $s = \hat{r}$), denotes the total signal-subspace rank throughout the AXDA iterations. The initialization is depicted in block (1) of Fig. 2. Since the depletion of anomaly vectors is supposed to decrease the anomaly subspace dimensionality in the data by one (see block (3)), one expects the total signal-subspace rank s to decrease by one as well. However, this is not always the case. For example, in cases where the dropped anomaly vector is highly correlated with the background subspace, dropping it from $\mathbf{\Omega}_h$, impairs the ability of $[\mathbf{\Psi}_{\hat{r}-h}|\mathbf{\Omega}_{h-1}]$ (here $s = \hat{r}-1$) to represent well the background subspace. In order to sequentially deplete anomaly vectors at each iteration, one needs to maintain the $\ell_{2,\infty}$-norm of data residuals to be low enough to admit the noise hypothesis in (12) at each iteration (see block (5)). Therefore, in this example, we need to increase the background dimensionality by one, which is performed by retaining s unchanged. Therefore, the decision to decrement the total signal-subspace s rank (see block (7)) is taken only if the reduced-rank subspace meets the maximum-norm noise residual hypothesis (see block (6)).

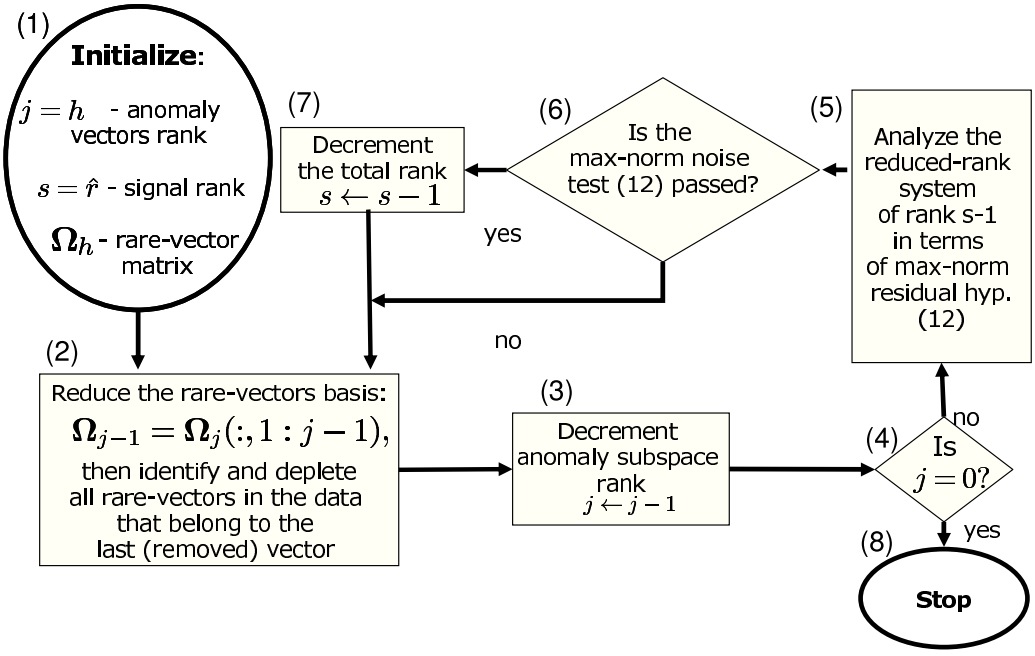

Fig. 2. **A concise outline of Anomaly Extraction and Discrimination Algorithm (AXDA).** The notation in block (2) is MATLAB® notation.

B. Detailed description of AXDA

At this point, we are ready to describe the AXDA algorithm in detail as shown in Fig. 3, where we mainly introduce details of block (2) in Fig. 2. The numbering of the following items correspond to the block numbers in Fig. 3.

1) **Initialization**

   The AXDA algorithm starts by initializing $j = h$, the number of anomaly vectors in $\mathbf{\Omega}_h$ and $s = \hat{r}$, the determined rank of signal subspace $range\ [\mathbf{\Psi}_{s-j}|\mathbf{\Omega}_j]$, and the maximum-residual norm denoted by η_s (see (9)), all as obtained from MOCA.

2) **Reduction of anomaly subspace basis**

   It is important to note that initially, the number $j = h$ of anomaly vectors in the partition $\mathbf{\Phi}_{\hat{r}} = [\mathbf{\Psi}_{\hat{r}-h}|\mathbf{\Omega}_h]$, obtained by MOCA, is optimal for the given signal-subspace rank, i.e., a decrease of h for a given $\hat{r}$ would result in an increased maximum-residual norm and, possibly, of other residual norms of data-vectors.

   We intentionally alter this optimality by dropping the last column of $\mathbf{\Omega}_j$, providing $\mathbf{\Omega}_{j-1}$. This operation is designed to detect anomalies related to the last column of $\mathbf{\Omega}_j$.

3) **Calculation of a new background vector representation basis $\mathbf{\Psi}_{s-j+1}$, corresponding to the new anomaly subspace basis $\mathbf{\Omega}_{j-1}$**

   In order to retain the total signal-subspace rank s, a new background vector representation basis $\mathbf{\Psi}_{s-j+1}$ is calculated by applying SVD on $\mathcal{P}_{\mathbf{\Omega}_{j-1}^{\perp}}\mathbf{X}$ that matches the reduced-rank matrix $\mathbf{\Omega}_{j-1}$.

4) **Calculation of data residual-norms in the obtained residual-subspace**

$$r_i = \|\mathcal{P}_{[\mathbf{\Psi}_{s-j+1}|\mathbf{\Omega}_{j-1}]^{\perp}}\mathbf{x}_i\|^2 \tag{14}$$

5) **Detection of anomaly vectors belonging to the dropped column j**

   In this block we identify indices of all anomaly vector residuals that exceeded the noise level η_s, which is equal to the maximum residual-norm initially obtained from MOCA.

6) **Decision about the next operation, based on previous block results**

   In this block we decide about the next operation based on whether anomaly vectors were found in the previous block. If there are such indices, then we perform the inner loop, in which we deplete the found anomaly vectors, recalculate $\mathbf{\Psi}_{s-j+1}$, and try to detect more anomaly vectors. Otherwise, the depletion of anomaly vectors in this iteration is completed and other operations of current iteration are performed.

7) **Association of found anomaly vectors to j-th column of $\mathbf{\Omega}_j$**

This block belongs to the inner loop of anomaly vectors depletion. We associate all data vectors indices (found in the block (5)) to the dropped column j and store them. Therefore, the corresponding anomaly vectors are denoted as *j-associated anomaly vectors*.

8) **Depletion of found j-associated anomaly vectors from input data and recalculation of $\mathbf{\Psi}_{s-j+1}$**

Since $(range\ \mathbf{\Omega}_{j-1})^{\perp}$ contains j-associated anomaly vector contributions, the subspace corresponding to $\mathbf{\Psi}_{s-j+1}$ (obtained earlier via SVD in Block (3)) is expected to be diverted in a way that aims to reduce these contributions along with the background vector residual-norms. As a result, not all anomaly vectors corresponding to the dropped j-th column of $\mathbf{\Omega}_j$ may be detected via tresholding their corresponding norms by η_s in Block (5). In order to remedy this problem, we deplete the j-associated anomaly vectors in the data (detected in Block (5)) and perform operations of blocks (3) - (5) again in order to obtain a more precise estimation of the background vectors subspace $\mathbf{\Psi}_{s-j+1}$, which is not diverted by the j-associated anomaly vectors found in Block (5).

9) **Decrementing of anomaly subspace rank**

Once all j-associated anomaly vectors are depleted, the rank of the anomaly subspace can be reduced by one. However, this does not necessarily mean that the total signal representation rank should also drop by one. This can be explained as follows: As it was already noted earlier, the background and anomaly subspaces in the hyperspectral images are not orthogonal. Therefore, if one reduces the rank of $[\mathbf{\Psi}_{s-j}|\mathbf{\Omega}_j]$ by removing a column from $\mathbf{\Omega}_j$, one might transfer a significant amount of background contribution to the complementary subspace $(range\ [\mathbf{\Psi}_{s-j}|\mathbf{\Omega}_{j-1}])^{\perp}$, which means that the reduced-rank subspace basis $[\mathbf{\Psi}_{s-j}|\mathbf{\Omega}_{j-1}]$ might not represent well the signal-subspace of the data after the j-associated anomaly-vectors depletion.

Therefore, the decrementing of anomaly subspace rank j does not necessarily entails decrementing the total signal-basis rank s. Thus, to decide if the total signal-basis rank s should be also decremented, we again employ, in the next blocks, the maximum-norm hypothesis testing (12). Due to the algorithm construction (see blocks (4),(5),(6)), it is guaranteed that at the input to this block, the subspace $(range\ [\mathbf{\Psi}_{s-j+1}|\mathbf{\Omega}_{j-1}])^{\perp}$ doesn't contain signal contributions. It is left to determine if the same holds true for the subspace $(range\ [\mathbf{\Psi}_{s-j}|\mathbf{\Omega}_{j-1}])^{\perp}$, which corresponds to the reduced total signal rank $s-1$.

We start by setting $j \leftarrow j-1$. In the next blocks we perform steps necessary for deciding if to decrement also the total signal-subspace rank s.

10) **Termination condition block**

If the anomaly subspace rank has reached 0, then terminate. Otherwise, continue.

11) **Calculation of $\mathbf{\Psi}_{s-1-j}$ corresponding to a reduced-rank signal-subspace**

In order to decide if decrementing j should also entail the decrementing of the total signal-subspace s, one has to obtain the reduced-rank subspace $[\mathbf{\Psi}_{s-1-j}|\mathbf{\Omega}_j]$ and test the corresponding data residuals. Therefore, in this block, we calculate $\mathbf{\Psi}_{s-1-j}$ by:

$$\mathbf{\Psi}_{s-1-j} = \underset{s-1-j}{\text{SVD}}\, \mathcal{P}_{\mathbf{\Omega}_j^{\perp}} \mathbf{X}. \tag{15}$$

12) **Calculation of maximum data residual-norm η_{s-1} in the obtained residual-subspace**

$$\eta_{s-1} = \max_{\mathbf{x} \in \text{cols } \mathbf{X}} \|\mathcal{P}_{[\mathbf{\Psi}_{s-1-j}|\mathbf{\Omega}_j]^{\perp}} \mathbf{x}\|^2 \tag{16}$$

13) **Performing noise-related hypothesis testing of η_{s-1}**

In this block we assess if η_{s-1} contains signal-contribution. For this purpose we apply the test of equation (12).

14) **Decision if to reduce the total signal-subspace rank s**

If η_{s-1} meets the noise-hypothesis, meaning that the subspace $(range\ \ [\mathbf{\Psi}_{s-1-j}|\mathbf{\Omega}_j])^{\perp}$ doesn't contain signal contributions (i.e., the basis $[\mathbf{\Psi}_{s-1-j}|\mathbf{\Omega}_j]$ represents well the signal-subspace), then s should be decremented. Otherwise, leave s intact and continue to a new iteration.

15) **Decrementing the total signal-subspace rank s**

$$s \leftarrow s - 1, \tag{17}$$

and continue to a new iteration at block (2).

Comments

1) Once the new value of s is determined, we approach a nominal state (at block (2)), where the anomaly vectors matrix rank is decremented by 1, and the signal-subspace basis $[\mathbf{\Psi}_{s-j}|\mathbf{\Omega}_j]$ (with the updated values of j and s) is "MOCA-optimal" with respect to the modified data-matrix $\mathbf{X}$. In order to extract other anomaly vectors, corresponding to the rest of $\mathbf{\Omega}_j$ columns, until the complete depletion of all anomaly vectors, steps 2 - 15 are repeated. The iterations stop when there are no more columns in the anomaly-basis matrix $\mathbf{\Omega}_j$, i.e., $j = 0$.
2) It is important to note that at the end of the AXDA procedure, the signal-subspace basis is composed solely of $\mathbf{\Psi}_s$ ($s \leq \hat{r}$), which constitutes the MOCA-optimal basis of the background vectors.

So the AXDA algorithm equips us also with a anomaly-free (in other words “robust”) estimated background-subspace and rank.

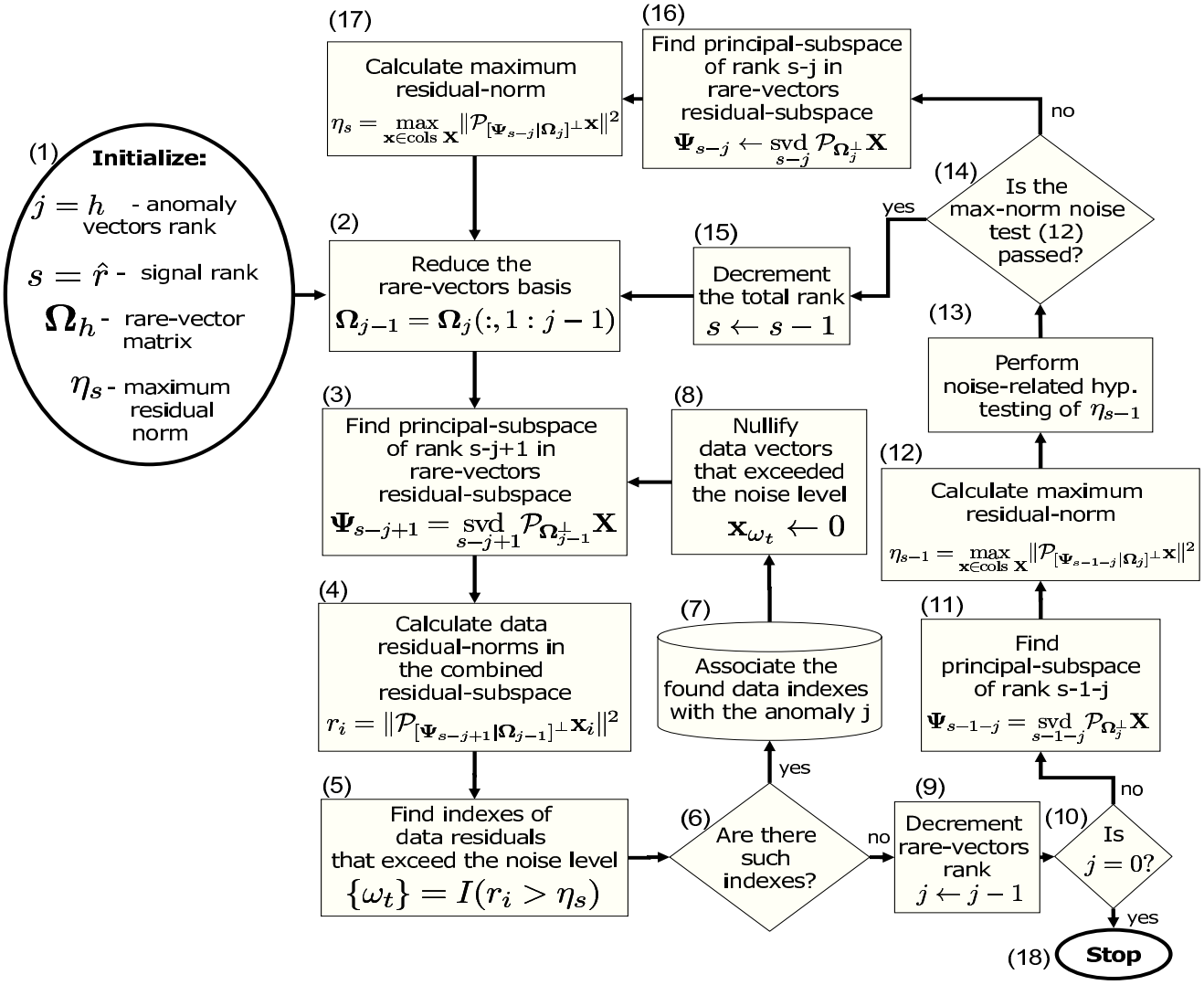

Fig. 3. **Detailed description of Anomaly Extraction and Discrimination Algorithm (AXDA).** The notation in block (2) is MATLAB® notation.

IV. Experiments with Real Hyperspectral Data

In this section we evaluate performance of the MOCA algorithm followed by AXDA postprocessing applying them on real hyperspectral data. For an analysis of the effect of noise on MOCA, using self designed synthetic data experiments with different signal to noise ratios, the reader is referred to [2].

To demonstrate the results, the proposed approach 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 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. 4 one can see results of anomaly detection and discrimination. Shown are images containing the 30th-band of 4 different hyperspectral cubes with different terrain types. The 5th and 6th images are not shown here just because of space limitations and convenience of placing an even number of images in the figure. The left 4 images contain ground-truth anomalies (marked in white and encircled by red ellipses), which were manually identified using side information collected from high resolution RGB images of the corresponding scenes. In Fig. 5, we show one of RGB images used for identifying the ground-truth anomalies. The right 4 images contain anomalies (marked in color) detected by AXDA, overlayed on the white ground-truth pixels. All anomaly pixels of the same type are marked by the same color. There are no missed anomalies in the presented 4 images. The corresponding dimensionality results obtained by MOCA and AXDA are separately summarized in Table. I, where $\hat{r}$ is the signal subspace rank determined by MOCA, h is the anomaly dimensionality, s is the dimensionality of anomaly-free background. Note, that according to the discussion in step 9 of the AXDA algorithm presented in the section III, it is possible that $s \geq \hat{r} - h$. Thus, AXDA allows discrimination of anomalies according to corresponding anomaly endmembers (constituent materials spectra) found by MOCA, though the accuracy of this discrimination is not evaluated in this paper and is under investigation.

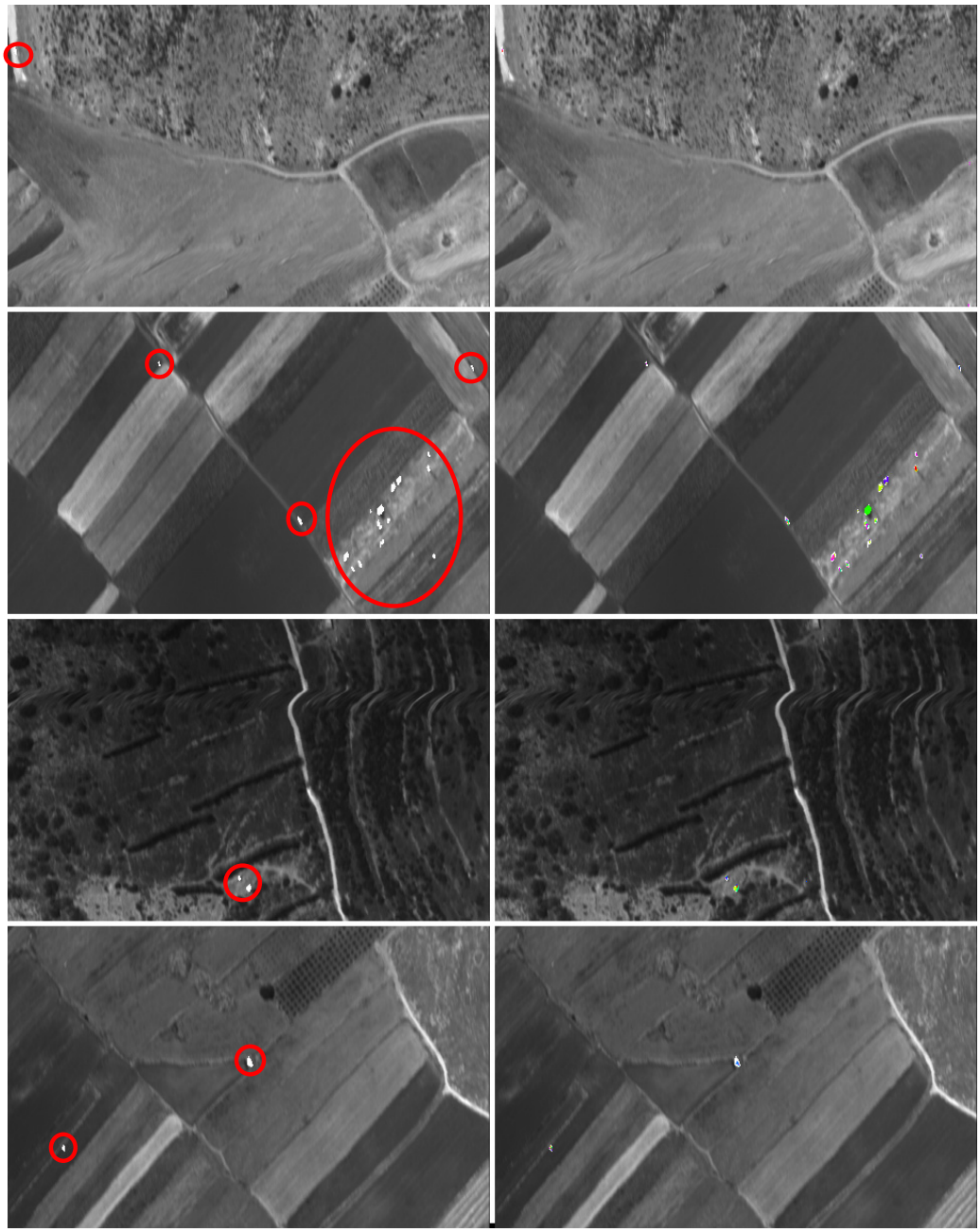

Fig. 4. **AXDA results at the *nominal* operating point.** The left 4 images contain manually identified ground-truth anomalies (marked in white and encircled by red ellipses). The right 4 images contain anomalies (marked in color) detected by AXDA, overlayed on the white ground-truth pixels. There are no missed anomalies in the presented 4 images. All anomaly pixels of the same type are marked by the same color.

TABLE I

No. image	$\hat{r}$	h	s
1	10	2	10
2	15	9	11
3	10	5	8
4	16	8	12
5	15	7	13
6	11	4	10

As it was noted above, in Fig. 5, we show one of the RGB images used for identifying the ground-truth anomalies. The scene under consideration is shown in a high-resolution (2672×4000) color RGB-image. The ground-truth anomalies are encircled by red ellipses. As it can be seen, the detected anomalies correspond to vehicles and small agriculture facilities, which occupy a few pixel segments.

In Fig. 6, we compare between GMRX [8], MSD [1] and the proposed AXDA in terms of Receiver Operation Characteristic (ROC) curves. For the purpose of ROC curves generation, 6 hyperspectral images were used, in which the total number of anomaly segments count is 25.

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 labelling. 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.

In order to obtain multiple operating points for AXDA, an additional parameter should be introduced to the proposed algorithm. A reasonable place for such a parameter is in the noise hypothesis relation in (12). However, due to special characteristics of maximum-norm noise distribution, which is very narrow - almost deterministic (see [2]), any factor introduced to this relation would result in almost the same decision. Thus, AXDA has naturally a single(nominal) operating point dictated by the noise statistical properties.

Yet, for the sake of comparison, we've introduced a rather compelling parameter γ to the equation of block (5) in Fig. 3, which now reads as:

$$\{\omega_t\} = I(r_i > \gamma\eta_s). \tag{18}$$

In words, the noise-related threshold value η_s (measured in a previous iteration) is multiplied by the factor γ in order to produce a new threshold value. The lower the factor γ is, the more data vectors

Fig. 5. **High resolution RGB image of the analyzed scene,** used as a ground-truth indication for AXDA results verification. The ground-truth anomalies are encircled by red ellipses.

will be treated as anomaly-vectors and be associated to the dropped column j of $\mathbf{\Omega}_j$. In our simulation, we have used 30 values of γ, which were uniformly sampled from $[0.8, 1.2]$. The position of nominal operating point of AXDA (for $\gamma = 1$) is pointed out by a red arrow. As can be seen from the figure, the nominal operating point provides a high detection rate (24 detected anomalies) with a significantly low false alarm rate (6 false alarm segments).

The GMRX 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 GMRX, too small clusters, and hence unreliable, were eliminated. In Fig. 7 one can see results of the GMRX algorithm, applied to the same 4 hyperspectral cubes as AXDA, with a GLR parameter producing the same false alarm rate as AXDA at the nominal operating point (which equals to 6 false alarm segments). As in Fig. 4, The left 4 images contain manually identified ground-truth anomalies (marked in white and encircled by red ellipses), whereas the right 4 images contain anomalies (marked in red), detected by GMRX, overlayed

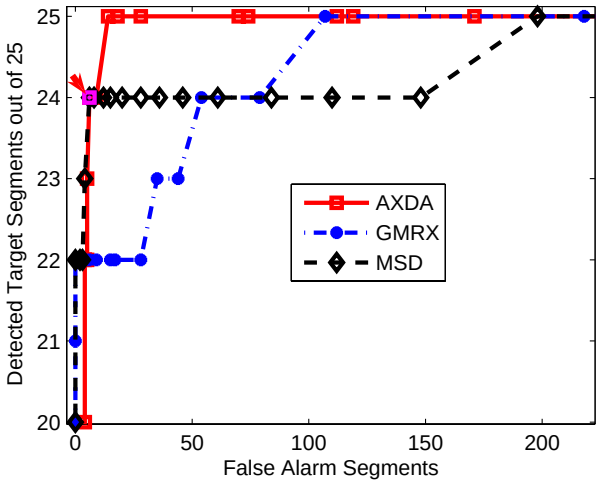

Fig. 6. **ROC curves corresponding to GMRX, MSD and AXDA.** The nominal operating point of AXDA is marked in magenta color and is pointed out by the arrow. This point corresponds to 24 detected anomalies and 6 false alarm segments.

on the white ground-truth pixels. The missed targets are encircled by cyan ellipses.

The MSD algorithm was provided an anomaly-free estimation of the background basis $\mathbf{\Psi}_s$ estimated by AXDA, which uses the anomaly subspace basis $\mathbf{\Omega}_h$ provided by MOCA, since MOCA and AXDA combined are unique in their ability to perform an unsupervised determination of both anomaly and background subspaces and their ranks.

Fig. 6 clearly shows that for the examined images AXDA has a better performance than GMRX and MSD, in most of the range of the tested parameters. It is also important to note, that in contrast to MSD and GMRX, AXDA allows an unsupervised determination of the nominal operating point, determined by maximum-norm noise statistical properties. Moreover, AXDA has an ability to discriminate between different types of anomalies.

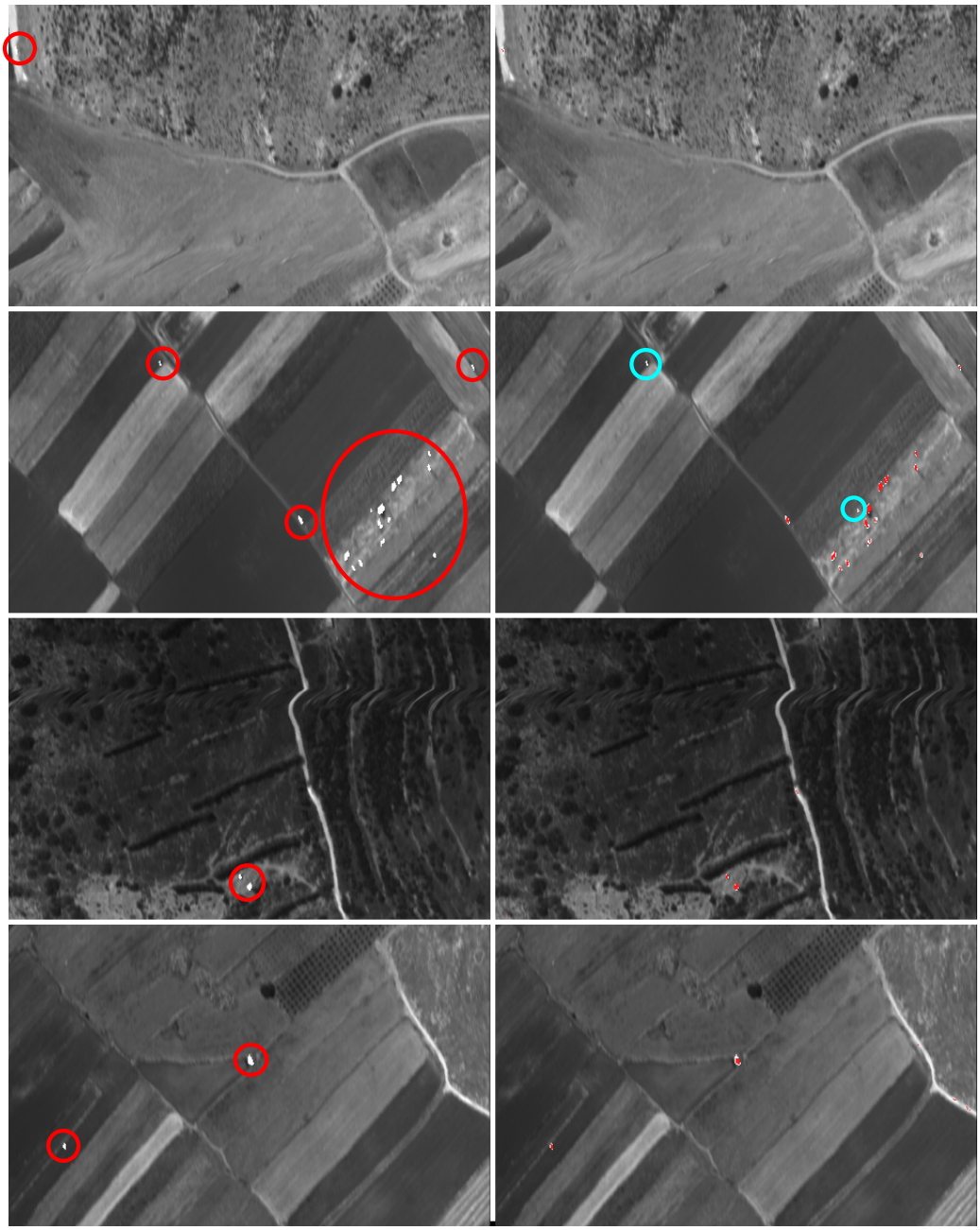

Fig. 7. **GMRX Anomaly Detection Results for GLRT parameter producing the same false alarm rate as AXDA at its nominal operating point.** The left 4 images contain manually identified ground-truth anomalies (marked in white and encircled by red ellipses). The right 4 images contain anomalies (marked in red) detected by GMRX, overlayed on the white ground-truth pixels. Missed anomalies are encircled by cyan ellipses.

V. Summary and Conclusion

In this work we have proposed an algorithm for anomaly detection, discrimination and population estimation of anomalies of the same type, called AXDA. The algorithm is based on a signal-subspace and rank estimation provided by MOCA [2]. By its construction, the signal basis consists of two groups of basis vectors. One group spans the subspace of anomalies. The second group is designed to represent background pixel residuals belonging to the subspace that is complementary to the subspace of the anomalies. The proposed AXDA extracts anomaly pixels by removing an anomaly basis vector from the anomaly vectors group and compensating for its removal by augmenting the background vectors related subspace. This operation causes a violation of the noise hypothesis condition in vectors that are highly correlated with the removed anomaly basis vector. Such vectors are detected, associated with the removed basis vector, and depleted from the data. This way we obtain groups of data vectors associated with each one of the anomaly basis vectors.

In experiments with real hyperspectral image cubes AXDA was shown to have a better performance than GMRX and MSD, in most of the range of the tested parameters. Since the anomaly and background subspaces are unknown in advance, the MSD algorithm was provided the anomaly-free estimation of the background basis $\mathbf{\Psi}_s$ obtained from AXDA and the anomaly subspace obtained from MOCA. This provides MSD subspace-related information that is (at least) as good as AXDA has for the detection of anomalies. It is also important to note, that in contrast to MSD and GMRX, AXDA is equipped with an unsupervised determination of the nominal operating point. AXDA also has a capability to discriminate between different types of anomalies, though the accuracy of this discrimination, as well as the accuracy of population estimation of anomalies of the same type, are not evaluated in this paper and are under investigation. Moreover, AXDA allows also an anomaly-free (robust) estimation of the background-subspace and rank.

It turns out now that MOCA in combination with AXDA provide means to meet a wide range of signal-subspace estimation scenarios:

1) *Estimation of a signal-subspace that includes anomaly-vectors.*
2) *Detection of anomaly-vectors and determination of their subspace.*
3) *Providing a natural (nominal) operating point for anomaly detection.*
4) *Estimation of the pure (free of outliers) background-subspace.*

APPENDIX I
NOISE VARIANCE ESTIMATION PROCEDURE

AXDA and MOCA strongly rely on the assumption of additive white Gaussian noise of known variance. The correct specification of the noise variance is of paramount importance since it determines the signal subspace rank (see (12)) and, as a result, affects the detection/false-alarm rates. In this appendix we describe a technique used for the estimation of noise variance in each hyperspectral channel. The estimated noise variance is then used for a band-wise normalization of the noise variance to 1.

It was observed in experiments with real hyperspectral data that *overestimation* of the noise variance by about a half an order of magnitude has little influence on AXDA performance. Although using an overestimated value of the noise variance (causing a poorer representation of the background) would result in the underestimation of signal subspace rank, the false alarm rate remains mostly unchanged. This happens since background misrepresentations are tested by rule (12) (used in steps 9 and 13 of AXDA), which depends on the noise variance as well. Thus, the overestimated noise variance raises the "effective threshold value", which naturally leaves the background misrepresentations undetected. Thus, an overestimated noise level may just slightly impair detection rate of anomalies that aren't prominent enough.

Using an *underestimated* value of the noise variance is less favorable because of special statistical properties of the maximal norm of noise. As shown in [2], the maximal norm of noise has a narrow distribution, explained by Extreme Value Theory results. Therefore, there is a high likelihood that the maximal norm of noise would obtain an almost deterministic value (see Figs. 4 and 5 in [2]). Thus, if the underestimated noise variance makes the "effective threshold value" implied by (12) lower than the almost deterministic maximal norm of noise, MOCA would never terminate its iterations (or will terminate too late). This would result in a significant signal-subspace rank overestimation, which may cause the background subspace of increased-rank to include the anomalies and to significantly impair the anomaly detection rate. Therefore, the noise variance estimation technique proposed below prefers *noise variance overestimation*.

In CCD-based hyperspectral systems, the noise is a combination of dark current noise, photon (shot) noise and fixed pattern noise (FPN) [19]. The FPN is due to different sensor responsivities, which is estimated and compensated out by calibrating the sensor. It turns out that even at mild light intensities, the *photon noise* may be dominant [19]. The photon noise problem arises from the statistical nature of photon production. The probability distribution for n photons in an observation window is known to be

Poisson:

$$p(n) = \frac{M^n e^{-M}}{n!}, \tag{19}$$

where M is the average number of photons within the given observation window. For the linear part of the CCD response function, the image intensity I is linearly proportional to n, i.e.,

$$I = gn, \tag{20}$$

for some proportionality coefficient g. Since the Poisson distribution approaches a normal distribution for large M, the photon noise in I can be modelled as having a zero-mean normal distribution with std σ_e satisfying:

$$\sigma_e(I) = \sqrt{gH}, \tag{21}$$

where H being the mean of I satisfying $H = gM$, is considered to be the clean signal.

Thus, the photon noise variance is not constant and, therefore, doesn't meet the noise stationarity property assumed in MOCA. Nevertheless, in our real-data simulations, we have empirically found that using

$$\sigma_{0.98} = \sqrt{gH_{0.98}} \tag{22}$$

in AXDA, as an estimate of the noise std in each band (with $H_{0.98}$ denoting the 0.98 quantile of image intensities in the band), decreases false alarm rate caused by high image intensity pixels, while allowing a reasonably high anomaly detection rate. The 0.98 quantile corresponds to almost maximum image intensity, ignoring 2% of the most intense image values that may stem from anomalies.

The only thing left is to estimate g. According to (21), g satisfies:

$$g = var\left(\frac{e}{H}\right), \tag{23}$$

where e denotes pixel noise and var denotes variance. Note, that random variables $\{e_i/H_i\}$, where i denotes pixel index, are identically distributed. If one assumes that they are independent, then g can be estimated by:

$$\hat{g} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{e}_i}{\hat{H}_i}\right)^2, \tag{24}$$

where $\hat{e}_i$ is a noise estimation, and $\hat{H}_i$ is a clean image intensity estimation.

The estimation of $\{H_i\}$ can be obtained via a 2D linear prediction as follows:

$$\hat{H}_i = \sum_{j \in \mathcal{K}} a_j I_{i,j}, \tag{25}$$

where $\mathcal{K}$ denotes the set of a 2D neighborhood indices, $I_{i,j}$ denotes the image intensity at position j in the neighborhood of a pixel i, and $\{a_j\}$ denote the linear prediction coefficients, obtained via least squares over the *whole image*. In our simulations we used a 5×5 neighborhood.

The estimation of $\{e_i\}$ is then given by

$$\hat{e}_i = I_i - \hat{H}_i. \tag{26}$$

Unfortunately, the estimation of $\{H_i\}$ given in (25) is inaccurate in non-smooth image regions such as edges and/or anomaly pixels. Therefore, the estimates $\hat{e}_i$ and $\hat{H}_i$ from these regions should not be accounted in (24) for the estimation of g. In order to filter out the undesired contributions of $\hat{e}_i$ and $\hat{H}_i$, we estimate $\sqrt{g}$ using median absolute deviation (MAD), proposed in [27], for a robust estimation of the standard deviation of e/H as follows:

$$\sqrt{\hat{g}} = MAD\left(\hat{e}/\hat{H}\right) = \underset{i=1,\ldots,N}{median} \left|\hat{e}_i/\hat{H}_i\right|, \tag{27}$$

where N is the total number of hyperspectral pixels.

Using the estimated values of g and H_i and substituting to (22), we obtain an estimate of the effective noise std $\sigma_{0.98}$ in each band and normalize the noise to unity variance in each band of the hyperspectral cube for further processing.

ACKNOWLEDGMENT

The authors would like to thank Mr. Eyal Madar for his helpful and constructive comments, and to the anonymous reviewers for their very useful comments and advice.

REFERENCES

[1] L. L. Sharf and B. Friedlander "Matched Subspace Detectors," *IEEE Trans. Signal Proc., vol. 42, no. 8, pp. 2446 – 2157, August 1994.*

[2] O. Kuybeda, D. Malah and M. Barzohar "Rank Estimation and Redundancy Reduction of High-Dimensional Noisy Signals with Preservation of Rare Vectors," *To Appear (Dec. 2007), IEEE Trans. Signal Proc.*

[3] 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. 1, pp. 1760 - 1770, Oct. 1990.*

[4] 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.*

[5] 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*

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

[7] 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*.

[8] 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.*

[9] P. Stoica and Y. Selen. "Model-order selection: a review of information criterion rules", *Signal Processing Magazine, IEEE, vol. 21, Issue 4, July 2004, pp. 36-47.*

[10] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-33, pp. 387-392, Apr. 1985.*

[11] M. Tipping and C. Bishop "Probabilistic principal component analysis", *Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol. 61, No. 3, pp. 611 - 622, 1999*

[12] C-I Chang and Q. Du "Estimation of Number of Spectrally Distinct Signal Sources in Hyperspectral Imagery", *IEEE Trans. on Geoscience and Remote Sensing, Vol. 42, No. 3, pp. 608 - 619, March 2004*

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

[14] B. Thai and G. Healey "Invariant subpixel material detection in hyperspectral imagery", *IEEE Trans. on Geoscience and Remote Sens., vol. 40, no. 3, pp. 599-608, March 2002*

[15] Q. Du and and C.-I Chang "A signal-decomposed and interference-annihilated approach to hyperspectral target detection,", *IEEE Trans. on Geoscience and Remote Sens. vol. 42, no. 4, pp. 892-906, April 2004.*

[16] 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.*

[17] M. J. Carlotto "A Cluster-Based Approach for Detecting Man-Made Objects and Changes in Imagery" *IEEE Trans. on Geoscience and Remote Sensing, vol. 43, no. 2, pp. 374-386 Feb. 2005.*

[18] L.L. Scharf *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis,* Addison-Welsey Publishing Company, 1993.

[19] H. Faraji and W. J. MacLean "Noise Removal in Digital Images" *IEEE Trans. Image Proc., vol. 15, no. 9, pp. 2676-2685 Sept. 2006.*

[20] P.J. Huber, *Robust Statistics,* Wiley: New York 1981.

[21] M. E. Winter and E. M. Winter, "Comparison of approaches for determinig end-members in hyperspectral data," *Aerospace Conference Proceedings, IEEE,* vol. 3, pp. 305 – 313, March 2000.

[22] M. Leadbetter, *Extremes and related properties of random sequences and processes,* Springer Series in Statistics, 1982.

[23] S.I. Resnick, *Extreme Values, Regular Variation, and Point Process,* Springer Verlag, New York., 1987.

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

[25] Gumbel, E.J., *Statistics of Extremes.,* Columbia University Press., 1958.

[26] C.I. Chang and Q. Du “Interference and noise-adjusted principal components analysis.” *IEEE Trans. on Geoscience and Remote Sensing, Vol 37, No. 5, pp. 2387 – 2396, 1999*

[27] D.L. Donoho, “De-noising by soft-thresholding”, *IEEE Trans. Inform. Theory, vol. 41, no. 3, pp. 613-627, May 1995.*

[28] D. Manolakis, C. Siracusa, and G. Shaw “Hyperspectral Subpixel Target Detection Using the Linear Mixing Model”, *IEEE Trans. Geoscience and Remote Sensing, vol. 39, no. 7, pp. 1232-1249, July 2001.*

[29] J. Dias and J. Nascimento “Estimation of Signal Subspace on Hyperspectral Data”, *in Proc. of SPIE, vol. 5982, pp. 191-198, Bruges, Belgium, September 2005.*

[30] W. X. Zheng “A Least-Squares Based Method for Autoregressive Signals in the Presence of Noise” *IEEE Trans. on circ. and syst.II: analog and digital signal processing, vol. 46, no. 1, pp. 81-84 Jan. 1999.*