Research Article

Hyperspectral Anomaly Detection: Comparative Evaluation in Scenes with Diverse Complexity

Dirk Borghys,[1] Ingebjørg Kåsen,[2] Véronique Achard,[3] and Christiaan Perneel[4]

[1] Department CISS, Royal Military Academy, 2007 Brussels, Belgium
[2] Land and Air Systems Division, Norwegian Defence Research Establishment (FFI), 2007 Kjeller, Norway
[3] Theoretical and Applied Optics Department, French Aerospace Laboratory (ONERA), FR-31055 Toulouse Cedex 4, France
[4] Department of Mathematics, Royal Military Academy, Brussels, Belgium

Correspondence should be addressed to Dirk Borghys, dirk.borghys@gmail.com

Anomaly detection (AD) in hyperspectral data has received a lot of attention for various applications. The aim of anomaly detection is to detect pixels in the hyperspectral data cube whose spectra differ significantly from the background spectra. Many anomaly detectors have been proposed in the literature. They differ in the way the background is characterized and in the method used for determining the difference between the current pixel and the background. The most well-known anomaly detector is the RX detector that calculates the Mahalanobis distance between the pixel under test (PUT) and the background. Global RX characterizes the background of the complete scene by a single multivariate normal probability density function. In many cases, this model is not appropriate for describing the background. For that reason a variety of other anomaly detection methods have been developed. This paper examines three classes of anomaly detectors: subspace methods, local methods, and segmentation-based methods. Representative examples of each class are chosen and applied on a set of hyperspectral data with diverse complexity. The results are evaluated and compared.

1. Introduction

Many types of anomaly detectors have been proposed in literature [1, 2]. The most frequently used anomaly detector is the (spectral only version of the) Reed-Xiaoli (RX) detector [3] that is often used as a benchmark to which other methods are compared. The RX detector characterizes the background by its spectral mean vector μ_B and covariance matrix Σ_B. The actual detector calculates the Mahalanobis distance between the pixel under test r and the background as follows:

$$D_{\mathrm{RX}} = (r - \mu_B)^T \Sigma_B^{-1} (r - \mu_B). \qquad (1)$$

The global RX detector characterizes the background of the complete scene by a single multivariate normal probability density function (pdf). In many scenes, this model is not adequate. For that reason, several variations of the global RX detector have been proposed in literature [1, 2, 4–12]. They can be sub-divided into three classes: subspace methods, local methods, and segmentation-based methods.

In complex scenes the latter category was shown to be very effective and several segmentation-based anomaly detectors (SBAD), not necessarily based on RX, have recently been proposed [13–20]. The aim of the current paper is to compare the results obtained by different types of anomaly detectors in scenes characterized by different types of background. In particular, two rural scenes with subpixel anomalies, a rural scene with some of the targets in shadow, and an urban scene were considered. Representative examples of each of the three previously mentioned classes of anomaly detectors were included in the comparison. In previous work [21], we noted the importance of data reduction and preprocessing on anomaly detection results. The current paper therefore also presents a comparison of results obtained by the different detectors after applying different preprocessing methods.

The evaluation of the detection results is mainly based on receiver-operating characteristic (ROC) curves. For spatially fully resolved targets, the false alarm rate at first detection was also considered. For the two scenes with extended targets,

besides an objective evaluation, a more subjective evaluation is also presented. The rest of the paper is organized as follows. Section 2 presents the used datasets; in Section 3 the examined anomaly detection methods are briefly presented; Section 4 presents the different preprocessing methods that have been applied to the data. The last two sections of the paper present the results and the conclusions. The appendix presents a brief exploratory data analysis that mainly aims at verifying to what extent the different used datacubes comply with the assumption of global or local unimodal multivariate normality.

2. Overview of the Dataset

The analysis was performed on a set of hypercubes of scenes with various backgrounds and representative of three scenarios as follows:

(i) a rural environment with subpixel targets (CAM and OSLO1),

(ii) a rural environment with some of the targets in shadow (BJO),

(iii) an urban environment (OSLO2).

Table 1 presents an overview of the used dataset. The first two datacubes are real hyperspectral images in which a matrix of anomalies was inserted artificially. Figure 1 shows RGB composites of these images on which the targets have been superimposed.

The results shown in this paper were obtained with 10% mixing ratio subpixel anomalies for the CAM scene. For OSLO1, the mixing ratio was varied from 100% to 10%. The inserted anomalies are spectra of a green paint (CAM) and a green fabric (OSLO1).

The BJO image (Figure 2(a)) was acquired over a natural scene with an agricultural region and a small forest near the village of Bjoerkelangen in Norway. The figure shows the target locations with light blue colored rectangles representing the target sizes superimposed on the RGB composite of the scene. Fourteen targets composed of different types of material and with different colors were laid in the scene during the image acquisition. Targets T3–T7 were in shadow. T3 was in deep shadow between the trees, and the four others were in the shadow at the edge of the forest. Table 2 presents the dimensions and material types of the different targets.

The OSLO2 scene (Figure 2(b)) is part of the center of Oslo. In this scene, four targets (T1–T4) were laid out. Their respective dimensions in the image are T1: 5×10, T2: 5×9, T3: 2×6, and T4: 6×7 pixels. Targets T2 and T3 are pieces of green fabric and the other two of a blue plastic. T1–T3 were laid out on the grass in a park, and T4 was put on an asphalt background in the shadow from a building.

The CAM image was rectified and atmospherically corrected. The images BJO, OSLO1, and OSLO2 were not rectified before processing and all processing on these scenes was applied to radiance data, that is, without applying any atmospheric correction.

3. Anomaly Detection Methods

Besides global RX, representative examples of three categories of anomaly detectors are examined in the paper. Figure 3 presents an overview of the selected detectors in the three classes. As can be seen from the figure, many of the investigated methods are RX-based, but for the subspace detection methods and in particular for the SBAD methods, some anomaly detectors that are not related to RX have also been included. The different detectors are briefly described below.

3.1. Subspace Methods. The subspace methods are global and have in common that they apply principal component analysis (PCA) or singular value decomposition (SVD) to the datacube. The first PCA/SVD bands are supposed to represent the background and they are eliminated in different ways by the various subspace methods. Subspace anomaly detectors are thus global anomaly detectors applied on a spectral subset (subspace). For all of the subspace methods, the only parameter is the number of PCA or SVD bands (n_b) that is considered to represent the background. If this number is set too high, targets will disappear in the background, if it is too low, too many false alarms will remain. Automatically determining an optimal value for the dimension of the background subspace remains a current research topic.

3.1.1. Subspace RX (SSRX). In SSRX, the global RX is applied on a limited number of PCA bands. The first PCs are discarded in SSRX.

3.1.2. RX after Orthogonal Subspace Projection (OSPRX). In OSPRX, the first PCA/SVD components define the background subspace and the data are projected onto the orthogonal subspace before applying the RX detector [2, 22].

In the current paper, the SVD of the global spectral covariance matrix Σ is used. Because Σ is positive definite, the SVD is equivalent to the following eigenvector/eigenvalue decomposition:

$$\Sigma = U \Lambda U^T, \tag{2}$$

where U is the matrix of eigenvectors of the decomposition and Λ the diagonal matrix with decreasing eigenvalues. The projection operator P_{SVD} is defined as a function of the first n_b eigenvectors (columns of U), corresponding to the highest eigenvalues, $W = U(1 \ldots n, 1 \ldots n_b)$ as follows:

$$D_{\text{OSPRX}}(r) = r^T (I - P_{\text{SVD}}) r = r^T \left(I - W W^T \right) r, \tag{3}$$

with I the $n \times n$ identity matrix. n is the number of channels in the datacube, and n_b the number of channels used to model the background subspace ($1 \leq n_b < n$).

3.1.3. RX after "Partialling Out" the Clutter Subspace (PORX). In this method, the effect of the clutter in a pixel is removed (partialled out) component-wise by predicting each of its

TABLE 1: Overview of the used dataset.

Name	Site	Sensor name	Number of bands n	Spectral range (in μm)	Spatial resolution (in m)	Image size (in pixels)	Number of targets	Total target size (in pixels)
CAM	Camargue (Fr)	Hymap	126	0.44–2.45	4	150×100	45	45
OSLO1	Oslo (No)	HySpex	80	0.410–0.984	0.25	286×287	81	81
BJO	Bjoerkelangen (No)	HySpex	80	0.410–0.984	0.25	700×1600	14	574
OSLO2	Oslo (No)	HySpex	80	0.410–0.984	0.25	700×1600	4	45

TABLE 2: Material types and sizes (in pixels) of the different targets in the BJO image.

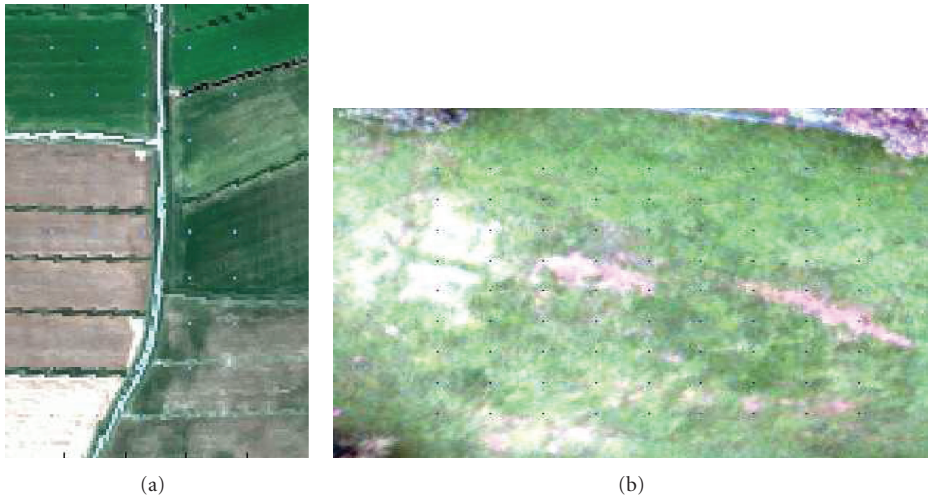| T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14
Green car	Red car	Cloths	Cloths	Painted boards Paint 1	Paint 2	Paint 3	Cloths	Cloths	Painted boards Paint 1	Paint 2	Paint 3	Paint 4	People
8×22	5×11	4×3	2×3	5×7	4×4	3×3	2×7	3×4	8×10	5×7	4×7	5×8	7×16

(a)   (b)

FIGURE 1: RGB color composite of the CAM (a) and OSLO1 (b) datacubes with targets superimposed, respectively, in cyan and black.

(a)   (b)

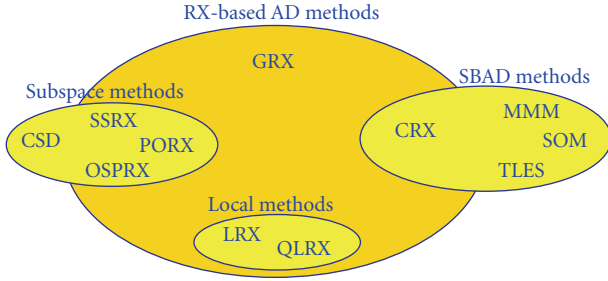FIGURE 2: RGB composite of the BJO (a) and OSLO2 (b) datacubes with target locations indicated.

FIGURE 3: Schematic overview of the examined anomaly detection methods.

FIGURE 4: Sliding triple window used in the local AD methods.

spectral components as a linear combination of its high-variance principal components. The detector applies the RX detector on the residual. Details of the method can be found in [10].

3.1.4. Complimentary Subspace Detector (CSD). The CSD is not an RX-based method. In the CSD, the highest variance principal components are again used to define the background subspace and the other PCs, to define the target subspace (the complimentary subspace) [7]. The PUT is then projected on the two subspaces and the anomaly detector is the difference of the projection onto the target subspace and the background subspace as follows:

$$D_{\text{CSD}}(r) = r^T P_t r - r^T P_b r, \quad (4)$$

where

$$\begin{aligned} P_b &= U(1\ldots n, 1\ldots n_b)U^T(1\ldots n, 1\ldots n_b), \\ P_t &= U(1\ldots n, (n_b+1)\ldots n)U^T(1\ldots n, (n_b+1)\ldots n). \end{aligned} \quad (5)$$

3.2. Local Methods. In the local anomaly detection methods, the statistics of the background are estimated locally in a window around the PUT. A double sliding window is used: a guard window and an outer window are defined, and the background statistics are determined using the pixels between the two (see Figure 4). Sometimes a triple window is used where the covariance matrix of the background is estimated in a larger window than the average local background spectrum.

3.2.1. Local RX (LRX). In LRX, the covariance matrix Σ_B and mean spectrum μ_B of the background are estimated locally in a triple window around the PUT. In the used implementation, the size of the guard window is a parameter from which the size of the two other windows is determined as a function of the number of bands in the image as follow:

$$\begin{aligned} W_\mu &= \min_{k\,\text{odd}}\left(k^2 - W_G^2 \geq \sqrt{10n}\right), \\ W_\Sigma &= \min_{k\,\text{odd}}(k^2 - W_G^2 \geq 10n). \end{aligned} \quad (6)$$

3.2.2. Quasi-local RX (QLRX). Quasi-local RX (QLRX) [9] offers a compromise between the global and local RX

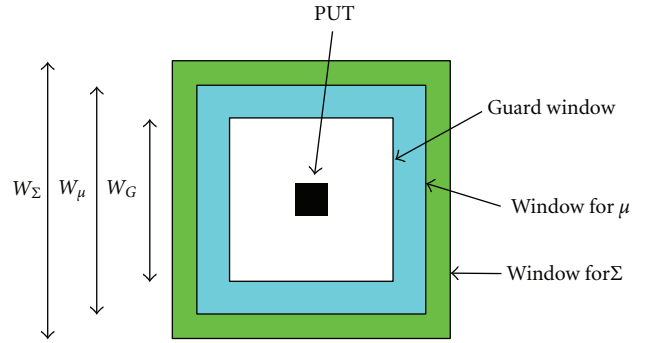approach. In QLRX, the global covariance matrix Σ is decomposed using eigenvector/eigenvalue decomposition (2). The eigenvectors are kept in the RX, but the eigenvalues are replaced by the maximum of the local variance and the global eigenvalue as

$$\lambda_{\text{QL}}^i = \left[\max\left(\lambda_{\text{loc}}^i, \lambda_{\text{glob}}^i\right)\right], \quad (7)$$

where i is an index denoting the pixel in the image. This means that the score of the detector will be lower at locations of the image with high variance (e.g., edges) than in more homogeneous areas. Spectral statistical standardization (see Section 4.4) is applied as a preprocessing step. The local variance is determined in a double sliding window.

3.3. Segmentation-Based Methods. In complex scenes, the hypothesis of a single multivariate normal distribution of background spectra is usually not verified, not even locally. For that reason several segmentation-based anomaly detectors have been proposed in literature. In this paper, four of these methods have been included in the analysis.

3.3.1. Class-Conditional RX (CRX). In CRX, the image is first segmented, the covariance matrix and mean within each class i (i.e., Σ_i and μ_i) are determined. The Mahalanobis distance between the PUT and each of the classes is calculated. The final result is the minimum of these distances:

$$D_{\text{CRX}} = \min_i\left[(r - \mu_i)^T \Sigma_i^{-1}(r - \mu_i)\right]. \quad (8)$$

In the current paper, K-means clustering is used and the parameters of the method are the minimum number of pixels allowed in each class and the maximum number of classes used in the clustering. The number of classes follows from these parameters.

3.3.2. Method Based on Multivariate Normal Mixture Models (MMM) [13]. A Stochastic Expectation Maximization (SEM) algorithm [23] is used for fitting a multivariate normal mixture model to the image for describing the background. The anomaly detector detects pixels that have a low probability according to the fitted model.

The parameters of the method are the maximum number of mixture components and the termination threshold for

the iterative parameters estimation method. The idea behind the mixture model is to use the mixture components (the multivariate normal pdfs) as base pdfs in a row expansion of the true pdf, which can then, in principle, have any shape.

3.3.3. Two-Level Endmember Selection Method (TLES).

The principle of the TLES method [19] is as follows: a small scanning window (50 × 50 pixels) runs over the image and at each position of the window the principal background spectra are determined using a segmentation method based on endmember selection. Endmembers that correspond to at least a given percentage (MP) of the image tile are stored. At the end of the process, an endmember selection is again applied on the stored endmembers and linear unmixing is applied on the image. Anomalies correspond to pixels with a large residue after unmixing. In [19], N-FINDR was used as the endmember selection method. In the current paper, the minimum volume simplex analysis algorithm (MVSA) [24] was used because it was found to give better results. Parameters of the method are the threshold MP and the number of endmembers kept in the two stages of the algorithm.

3.3.4. Method Based on Self-Organizing Maps (SOM).

A trained SOM is considered as a representation of the background classes in the scene. Anomalies are determined by computing the spectral distances of the pixels from the SOM units [16, 17]. The SOM was applied on the first PCA components and run using a square map consisting of NsxNs hexagonal cells. The SOM was optimized sequentially. The parameters of the method are Ns and the number of PCA bands used.

4. Preprocessing Methods

Before applying the actual anomaly detectors, some preprocessing methods were applied to the data. Three different types of preprocessing were applied.

The first type is data dimension reduction, which has two objectives. The first objective is to describe the background better and to obtain more reliable statistical estimation, especially when applying local methods where the number of samples to compute statistics from is low. Moreover, reducing data dimension allows to reduce the size of the windows for the local methods. This reinforces the local aspect of the method and reduces the risk that nearby targets overlap with the window used to compute background statistics. The second objective is to project the data on axes where the anomalies are enhanced, that is, the most separated from the background pixels. In this paper, we focus on two different methods, spectral binning which fulfills the first objective, and kurtosis-based dimension reduction, that attempts to fulfill both.

The second type of preprocessing aims to account for the effects of shadow. In this paper, a simplified approach consisting in square root transforming the data is used (Section 4.5).

Finally, some AD methods need some specific preprocessing that is described in Sections 4.3 and 4.4.

4.1. Dimension Reduction by Spectral Binning (SB).

As noted in [25, 26], dimension reduction can improve hyperspectral anomaly detection performance substantially. We have applied a dimension reduction method based on spectral binning, similar to the method applied in [26]. The binning consists in averaging over groups of neighboring bands, down to a spectral resolution of about 30 nm. The binning tends to improve the signal-to-noise ratio by reducing the relative contribution of photon noise. When this can be done while preserving the relevant spectral features, the result is improved detection performance.

4.2. Kurtosis-Based Dimension Reduction (KDR).

As anomaly detection aims to search for outliers, a projection that enhances outliers applied as a preprocessing can improve detection performance. It has been shown that kurtosis is very sensitive to outliers. In [27] data are projected on the (first) eigenvectors of the kurtosis matrix K:

$$K = \Sigma^{-1} \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)^t \Sigma^{-1} (X_i - \mu)(X_i - \mu)(X_i - \mu)^t, \quad (9)$$

where X_i is the ith element of X, the matrix of observations (the spectrum of the ith pixel), and μ and Σ are the spectral mean and covariance matrix of the datacube. N is the total number of pixels in the image. This method is mainly useful if the data are unimodally distributed, that is, in scenes characterized by a relatively homogeneous background (in this work, the CAM and OSLO1 images). Usually, only the first 3 to 5 kurtosis components are kept for further processing. This is the case for GRX, LRX and QLRX, MMM, and SOM. For the subspace methods and TLES, all kurtosis components were used.

4.3. Spectral Whitening.

If the eigenvalues and eigenvectors of the covariance matrix of the complete image are, respectively, Λ and U, and μ is the average spectral vector of the image, then the spectral whitening of the pixel r is given by [6, 7]

$$r_W = U\Lambda^{-1}U^T(r - \mu). \quad (10)$$

After spectral whitening, a Gaussian distributed variable becomes spherically symmetric and this is sometimes beneficial for detection [6]. Whether whitening is beneficial for the anomaly detector depends on the AD method and the datacube. For CSD, spectral whitening is always applied. The other subspace methods were applied with and without whitening and the best result obtained is reported in this paper.

4.4. Spectral Statistical Standardization.

The spectral statistical standardization converts each spectral band to have a zero mean and a standard deviation of one. This is necessary for the QLRX in order to make sure that the global eigenvalues and the local variances can be interchanged in the algorithm.

4.5. Square Root Transform. Detection performance is generally degraded in shadow, due, among other things, to low signal-to-noise ratio and distortions of the spectral signatures caused by secondary illumination, the adjacency effect, and path-scattered skylight. In addition, the large dynamic range of the data from scenes containing both sunlit and shadowed areas makes the data modeling task more difficult. In order to improve detection performance in shadow, different strategies can be applied: de-shadowing and illumination suppression for estimation of sunlit-like radiance in shadowed areas [28–31], sun/shadow segmentation and application of adapted modeling in the respective areas [32], or transformation of the data to account for the effects of shadow. We have square root transformed the data $r_{\text{sqrt}} = \sqrt{r}$. This reduces the dynamic range of the data, and, perhaps more importantly, makes the noise signal level independent [33], with benefits for data modeling through a suppression of the influence of noisy low-level signals. At low signal levels for homogeneous backgrounds, the dominating source of variation in the signal is Poisson distributed counting (photon) noise, and square root transforming the data yields approximate normality [34]. Square root transforming the data of course also affects the distribution in the, vast majority of, cases where scene clutter is the main source of signal variation, but in unpredictable ways for complex backgrounds.

5. Evaluation Method

Experimental ROC (receiver operating characteristic) curves, showing the detection rate (DR) versus the false alarm rate (FAR), are used to evaluate the results obtained with the various detectors. For the images with resolved targets, a pixel-based ROC curve is calculated for each target, whereas for the images with subpixel targets, an ROC curve is calculated based on all the targets in the image. DR is plotted versus the logarithm of the FAR (the resulting curve is referred to as a logROC), and the area under the logROC curve (the logAUC) is calculated and used as the measure of performance. The reason for using a logarithmic FAR scale is that it ensures equal weight across the range of FAR values.

For extended targets (in BJO and OSLO2), ROC curves give the detection performance for each pixel of the target with respect to the false alarm rate. For defense and security applications, it is also of interest to assess the performance at the first detection of a target. In this paper, for extended targets, we therefore also determined the false alarm rate at the first detection for each of the targets.

Besides these objective evaluation metrics, it is also interesting to look at the type of false alarms that the various detectors produce in the different scenes. Therefore for the "best" detectors, a detection image is shown corresponding to the threshold for which at least one pixel of the most difficult target is detected. This subjective result is shown for scenes with extended targets (BJO and OSLO2).

6. Results and Discussion

6.1. Implementation Issue: Parameter Selection. The different examined AD methods depend on different parameters.

For some of the applied methods, the parameters were set according to experience, whereas for others, where we lack this experience and where no consensus exists in literature for setting the parameters, we chose the optimal parameter setting through an optimization process in order to make the comparison between the detectors fair.

For the local methods, the parameters are the dimensions of the guard window and the outer window(s). The guard window should be set to be larger than the largest target of interest expected to be present in the scene, and size of the outer window(s) is derived from it as explained in Section 3.2.1. For the two datasets with subpixel anomalies, the guard window was set to 1, and for the two other datasets, it was set to 15.

GRX has no parameters.

For subspace methods, the only parameter is the dimensionality attributed to the background subspace. Several methods have been proposed for estimating this "signal subspace," mainly for unmixing purposes [35–39]. The two latter focus on finding the signal subspace dimension in the presence of "rare signals." They are thus likely to add the signal components containing the target to the signal subspace and are therefore less relevant to the choice of the signal subspace in subspace anomaly detection. The remaining methods [35–37] give different results and none of the signal subspace dimensions estimated by these methods correlate in a consistent way with the optimal number of bands in the subspace detectors for the different scenes. A consistent way for identifying the proper dimensionality to use in modeling the background clutter for subspace anomaly detection has yet to be found, as mentioned in the conclusion of [10]. Because the aim of the paper is to compare the different algorithms in the different scenes, the dimensionality parameter is optimized for each detector/scene combination. The complete range of possible background dimensionality (1 to $n-1$) was explored and the results shown are the best results obtained by that algorithm.

Each of the SBAD methods has its own set of parameters. For CRX, two parameters are set: a maximum number of classes and a minimum number of pixels per class. This last parameter can be used to reduce the risk that anomalies form their own classes. Then the maximum number of classes can be set higher than the actual number of background classes. The minimum number of pixels in each of the classes is set to a low percentage (for instance 0.5%) of the total number of pixels. For MMM, the parameters are similar to those of CRX. For SOM and TLES, the parameters, described in Section 3.3, were varied in a reasonable range and the results shown are the best obtained for the examined range of parameters.

6.2. Results for Subpixel Detection in a Rural Environment

6.2.1. Results for CAM. Table 3 shows the logAUC results for the different detectors obtained in the CAM dataset with a 10% mixing ratio. Results are shown without prior data reduction and with two types of data reduction: spectral binning and kurtosis-based data reduction. We see from the

TABLE 3: LogAUC results for CAM scene with a mixing ratio of 10% for different types of data reduction.

AD method	Data reduction method		
	No data reduction	Spectral binning	Kurtosis data reduction
GRX	0.595	0.637	*0.732*
SSRX	0.742	0.778	**1.000**
OSPRX	*0.931*	0.754	0.893
PORX	0.568	0.205	*0.957*
CSD	*0.868*	0.826	0.763
LRX	0.616	0.950	**1.000**
QLRX	0.720	0.828	**1.000**
CRX	0.683	0.830	**1.000**
TLES	0.128	0.268	*0.698*
SOM	0.116	*0.529*	0.120
MMM		**1.000**	**1.000**

TABLE 4: LogAUC results for OSLO1 scene with a mixing ratio of 33% for different types of data reduction.

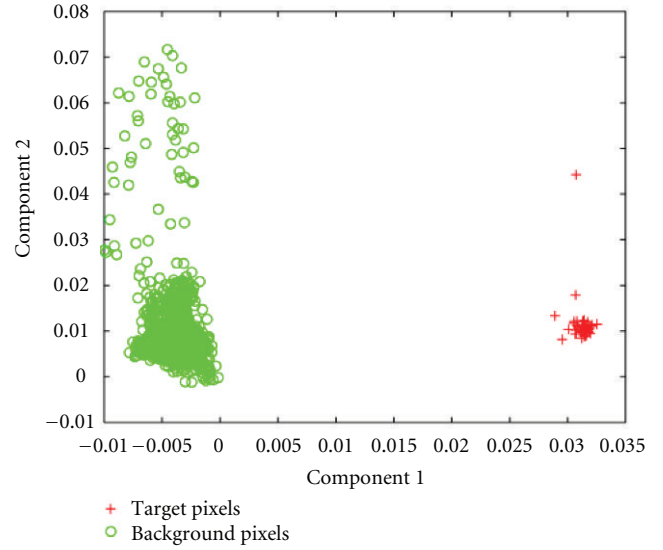| AD	No data reduction	Spectral binning	Kurtosis data reduction
GRX	0.163	0.357	*0.390*
SSRX	0.292	*0.486*	0.394
OSPRX	0.383	*0.486*	0.391
PORX	0.322	*0.493*	0.391
CSD	0.246	0.391	*0.396*
LRX	0.821	0.983	**0.995**
QLRX	0.294	0.395	*0.473*
CRX	0.163	0.357	*0.449*
TLES	0.108	*0.466*	0.356
SOM	0.232	0.385	*0.463*
MMM	0.128	0.544	*0.597*

FIGURE 5: Projection of background (green) and target (red) pixels on the first two kurtosis component axes for CAM10.

table that all SBAD methods, local methods, and GRX benefit from dimension reduction, more from kurtosis dimension reduction (KDR) than from spectral binning (SB) (with exception for SOM, which performs much better after SB than after KDR). For the subspace methods, the results are more diverse: OSPRX and CSD perform best on data with no dimension reduction, whereas for SSRX and PORX a large improvement in performance is observed after KDR.

Figure 5 shows a scatter plot of the CAM data with 10% mixing ratio subpixel targets in the 2D space defined by the first two kurtosis components. The figure shows that the targets are very well separated from the background after KDR transform in this scene. This clear separation is not observed on any of the PCA components or on spectral bands before or after spectral binning.

The best results for this scene are indeed obtained with SSRX, LRX, QLRX, CRX, and MMM after KDR. These detectors all achieve a logAUC of 1.0, which means that all targets have been detected with a false alarm rate that is smaller than 1/image size, that is, FAR $< 6.6 * 10^{-5}$. Since the

results are saturated, we cannot properly distinguish between the methods.

6.2.2. Results for OSLO1. Table 4 shows the logAUC results for the different detectors obtained in the OSLO1 dataset with a 33% mixing ratio. Results are shown without prior data reduction and with spectral binning and kurtosis-based data reduction.

LRX clearly outperforms all other detectors. The assumption of local normality is very well met in this image (cf. Table 6 in the Appendix). LRX is also, in contrast to all other methods, able to model the background without influence of targets. For subpixel targets this is particularly true, since the guard window will always contain the whole target. As regards data reduction, for SBAD and local methods and GRX, the results are the same here as for CAM10: the methods all benefit from dimension reduction, and more from KDR than from SB (with exception for TLES which benefits more from SB than from KDR). In this dataset, we also observe an improvement in performance with dimension reduction for the subspace methods, but for these methods SB is generally more beneficial than KDR.

For OSLO1, the behavior of each detector as a function of the mixing ratio was also investigated. Figure 6 shows the logAUC results of the different detectors versus the mixing ratio for the OSLO1 datacube. In the experiment, the mixing ratio was varied from 100% (full pixel anomaly) to 10%. For creating the figure, the data reduction method that gives the best results was selected for each of the detectors. Results of global RX-based methods and CSD are shown as solid lines, LRX and QLRX results as dashed lines, and results of segmentation-based methods as dot-dashed lines. LRX clearly gives the overall best results, followed by SSRX for larger target portions and MMM for smaller. The result of
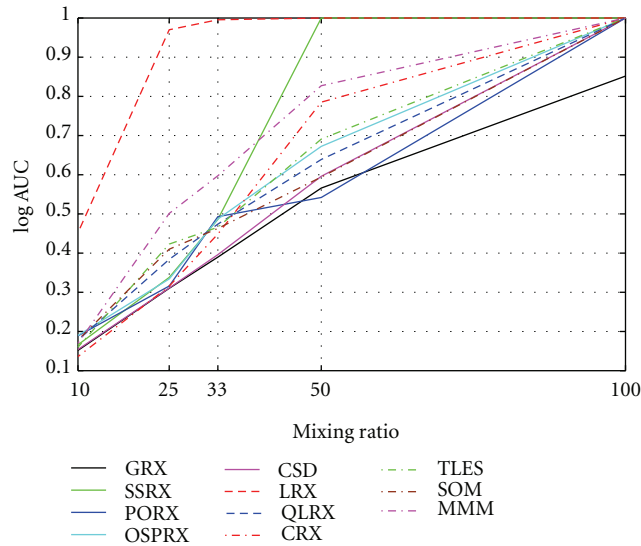
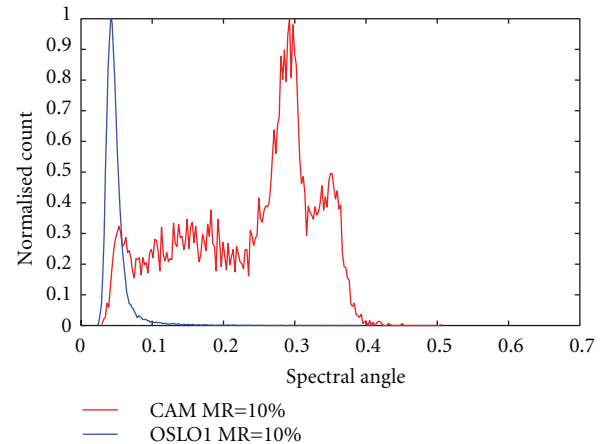FIGURE 6: LogAUC versus mixing ratio for the different detectors in OSLO1.

FIGURE 7: Normalized histograms of the spectral angle of the background pixels with respect to the average target spectrum in the CAM and OSLO1 scenes, at 10% mixing ratio.

SSRX for the 50% mixing ratio is very deviant, suggesting perhaps that subspace fit is somewhat random.

Contrary to the CAM scene, in OSLO1 the performance of the detectors at 10% mixing ratio is very low. The targets are more difficult to detect in the OSLO1 scene than in the CAM scene although the OSLO1 scene has a more homogeneous background and conforms well to the multivariate Gaussian assumption (see discussion of Table 5 in the Appendix). The OSLO1 scene is more difficult than the CAM scene because both the spectral angle and the Euclidean distance between the targets and the different background spectra are much smaller in the former. Figure 7 illustrates this point by means of the normalized histogram of the spectral angle of all background pixels with respect to the average target spectrum for both scenes at a mixing ratio of 10%. The figure shows that the spectral angle is indeed larger in the CAM scene than in the OSLO1 scene. The spread of the spectral angle in the CAM scene also illustrates the heterogeneity of the background.

6.3. Results for a Rural Environment with Some of the Targets in Shadow (BJO). Figures 8 to 10 show a graphical representation of the logAUC (a) and the logarithm of the false alarm rate at first detection (logFARAt1stDet) (b) for each of the detectors and for each target for the BJO scene. The colors represent the value of the respective performance metric. The color map is such that red corresponds to the best performance. The three different figures represent results after different types of preprocessing: Figure 8 shows results obtained without any preprocessing, Figure 9 results after spectral binning, and Figure 10 results after spectral binning and square root transform.

From the figures it is immediately clear that the targets in shadow (T3–T7) are more difficult to detect than the others. T3, hidden in the forest is the most difficult to detect. T2 (a red car) is the most easily detectable target.

We observe a slight improvement of performance from spectral binning for all detectors except TLES and SOM, which have their detection performance for targets in the sun substantially degraded by spectral binning. The reason why we—as opposed to what we saw in the previous datasets—only observe a slight improvement in performance from spectral binning, could be that the targets in sun are so easy to detect that we detect them anyhow, whereas the targets in shadow need some form of compensation for shadow in order to be detectable. The levels of the performance values indicate that this could be the case. For LRX, the size of the windows used to calculate the background statistics depends on the number of bands, and the results hence are not comparable across different numbers of bands.

Square root transforming the data generally improves the detection performance for targets in shadow substantially. It also improves the performance for targets in the sun for some detectors, but for others, mainly the subspace detectors and GRX, it reduces the performance for some sunlit targets—notably targets that are intensity anomalies (T8, T10, and T11), and that hence have their degree of anomaly reduced when the data are square root transformed. The best results on this dataset are obtained with MMM, LRX, and CRX. MMM gives the overall best results, whereas LRX gives the best results for the targets in shadow: for targets T5 and T6, LRX gives significantly better results than MMM. LRX gives also slightly better results than MMM for T10 and T11, which are painted with the same paints as, respectively, T5 and T6. The logAUC and logFARAt1stDet results are globally inter-consistent, but they do show supplementary information. The results are consistent with the complexity of the scene and with the compliance with the multinormal distribution assumption locally shown in Table 6 of the Appendix.

In order to give an idea of the type of false alarms produced by the three best detectors, in Figure 11 the results of the three best detectors (MMM, LRX, and CRX after spectral binning and sqrt transform) are superimposed on a grayscale image of the BJO scene. The shown results are
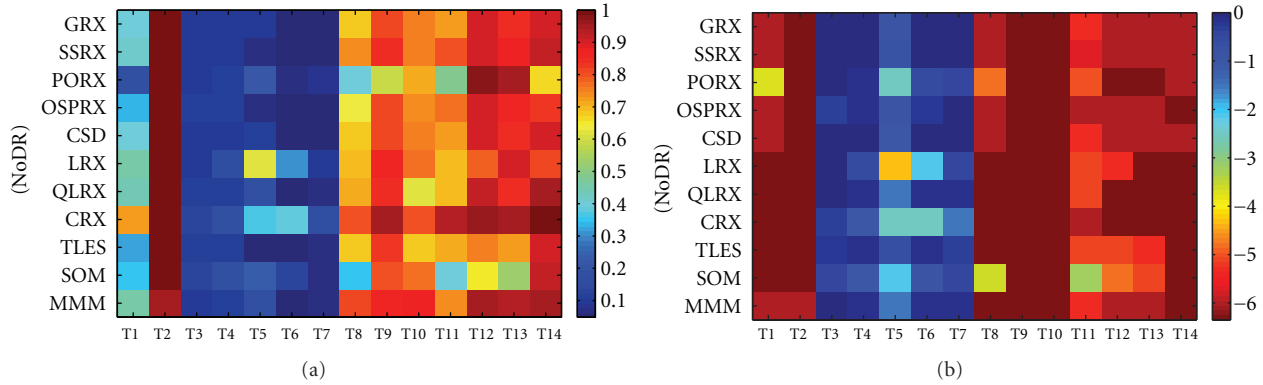
Figure 8: logAUC and logFARAt1stDet results per target without any data reduction for BJO.
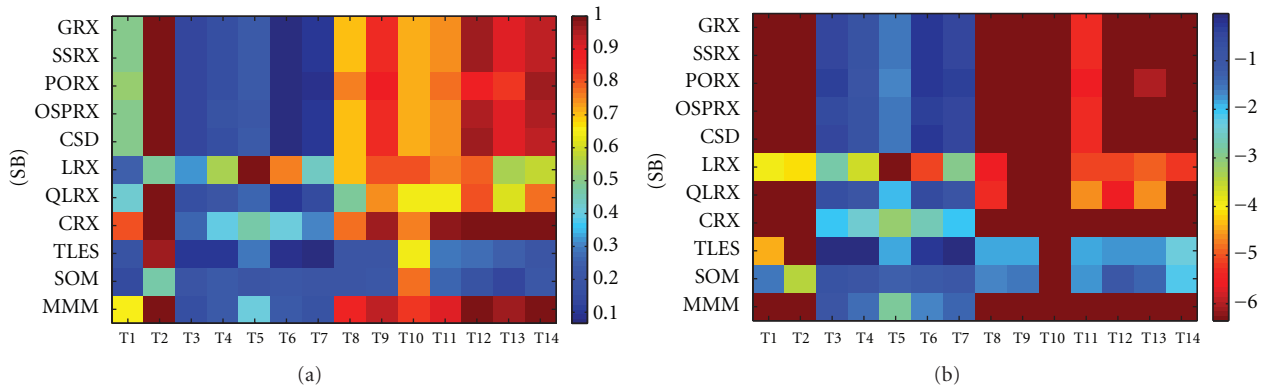

Figure 9: logAUC and logFARAt1stDet results per target after spectral binning for BJO.

thresholded detection results with the threshold set to the lowest first detection level for the true targets (i.e., the threshold for which at least one pixel of the most difficult target is detected). The figure shows that the false alarms produced by MMM mainly consist of isolated pixels in the forest and also some more extended false alarms at the top right of the image. LRX produces some small false alarms in the forest while CRX detects part of the stream as well as some detections in the forest. Most of the false alarms are detected by only one detector. On the other hand, for each target, except T7, an overlap in the detected zone for the three detectors is seen. The detectors are thus complimentary and fusing their results may be of interest. Likely causes of the complementarity of the results are that LRX are able to account more correctly for local illumination, whereas MMM/CRX are able to model locally heterogeneous background (forest) more precisely. The results for CRX indicate that too few classes are used: a large background structure like the stream is poorly modelled.

6.4. Results for the Urban Scene (OSLO2).

Figure 12(a) shows the logAUC results for the OSLO2 scene. Figure 12(b) shows the logFARAt1stDet results. For creating the figure, the data reduction method that gives the best results was selected for each of the detectors. The best data reduction method was spectral binning for most detectors, but for OSPRX and

QLRX the best results were obtained without data reduction, and for LRX, KDR gave the best results. The superiority of spectral binning over KDR for most detectors is to be expected because of the complexity of the scene (cf. Table 5 in the Appendix), and similarly the good performance of KDR for LRX—it can be attributed to local normality, see Table 6 in the Appendix. The mixed results for subspace detectors are consistent with the results for CAM and OSLO1. The result for QLRX, on the other hand, is not, but we should probably not read too much into this, since the detector more or less fails to detect the targets.

It can be seen that the values of logAUC are much lower than for the other datacubes. The maximum value obtained here is around 0.5. This is due to the complexity of the scene: the targets inserted into the scene are not the only anomalies. In an urban environment many objects can present an anomalous spectrum, for example, cars, special roof materials, and so forth. The comparison therefore only shows how well the different anomaly detection methods cope with urban "clutter."

From the two figures it can be seen that MMM gives the overall best results, followed by TLES and CRX. Of the two latter CRX gives the best results according to the logAUC metric and TLES according to the logFARAt1stDet metric. As could be expected the SBAD methods thus obtain the best results in the urban scene.
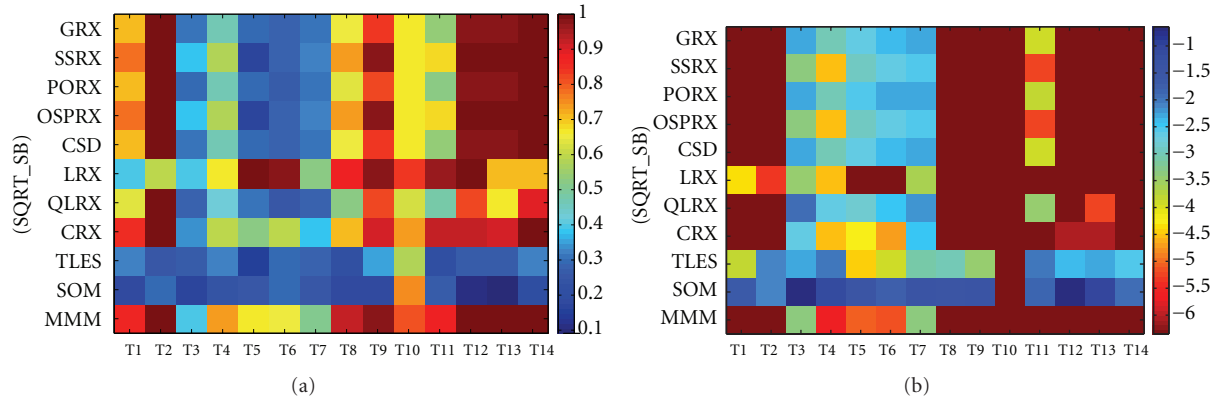
(a)

(b)

FIGURE 10: logAUC and logFARAt1stDet results per target after SQRT transform and spectral binning for BJO.
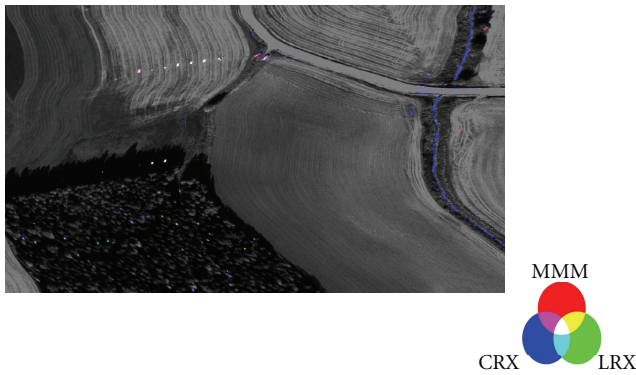

FIGURE 11: Color composite of the detection results for MMM, LRX, and CRX at lowest 1st detection threshold, superimposed on B/W image of BJO (R: MMM results, G: LRX results, B: CRX results).

GRX and all subspace methods perform quite bad, except for OSPRX that gives quite good results for targets T1, T2, and T4. Contrary to what we have seen in the other datasets, LRX does not perform particularly well in this dataset—despite good compliance with the multivariate normal distribution assumption locally near the targets, cf. Table 6 in the Appendix. The reason for this is probably that the remaining (non-target) parts of the scene are not well described by a local multivariate normal distribution, and hence we get lots of false alarms. This assumption is partly verified by comparing with the results of OSLO1. The target material of targets T2 and T3 in OSLO2 is the same as the material of the OSLO1 targets, and the local background is very similar (grass at two different places in Oslo), so a difference in performance between the images ought to be due to a difference in the number of false alarms.

As mentioned above, an urban scene presents many objects that may have an anomalous spectrum and that thus will be considered a false alarm in the above evaluation. It is therefore of interest to give an idea of these false alarms for each of the detectors. Figure 13 presents the results of the three best detectors (MMM, CRX, and TLES), superimposed on a grayscale image of the OSLO2 scene. The results shown

are thresholded detection results with the threshold set to the lowest first detection level for the true targets.

The figure shows that target 1 is the most easy, and that it was detected completely by the three detectors. T2 and T3 have been completely detected by MMM, while the two other detectors (at the selected thresholds) detect only a part of the interior. On the contrary, T4 has been completely detected by TLES and CRX, while MMM only detects a part of its interior. The figure also shows that many of the "false alarms" are quite different for the three detectors. TLES detects parts of the vegetation. CRX detects a set of small objects next to the building on the lower left in the image. Some cars and small structures on roofs have been detected by a combination of detectors. This subjective evaluation shows that the three detectors that perform best according to the "objective" evaluation are quite complimentary. Examining further the properties of their results may lead to interesting ideas on fusion of anomaly detectors.

7. Conclusions

This paper evaluates the performance of anomaly detection methods in scenes with different backgrounds and types of targets: agricultural scenes with subpixel targets, an agricultural scene with some of the targets in shadow, and an urban scene. Three classes of anomaly detectors were considered besides the global RX: subspace methods, local methods, and segmentation-based anomaly detection (SBAD) methods.

For subpixel anomaly detection in scenes of low complexity (rural and non shadow), LRX gives the best results, followed by MMM. From the investigated global RX-based methods SSRX and OSPRX give the best results. For the SBAD and local methods and GRX, detection results are improved by data reduction, and (with minor exceptions) more by kurtosis dimension reduction than by spectral binning. The improvement of results after kurtosis-based data reduction for most of the detectors illustrates the potential of customizing the data reduction method.

For the rural scene with some of the targets in shadow the results show that it is important to account for the effects
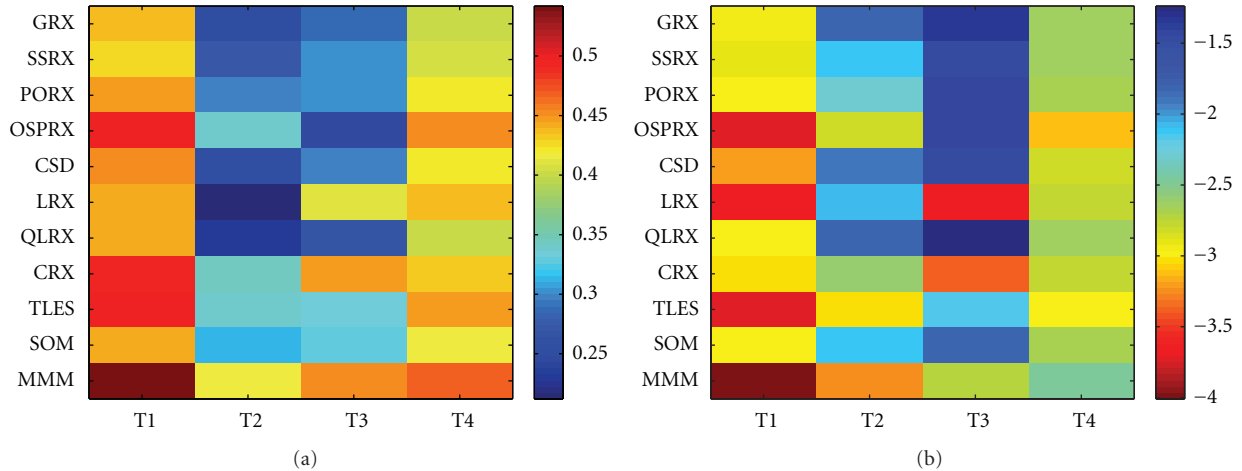
(a)

(b)

FIGURE 12: logAUC and logFARAt1stDet results per target for OSLO2 using the best data reduction method for each detector.
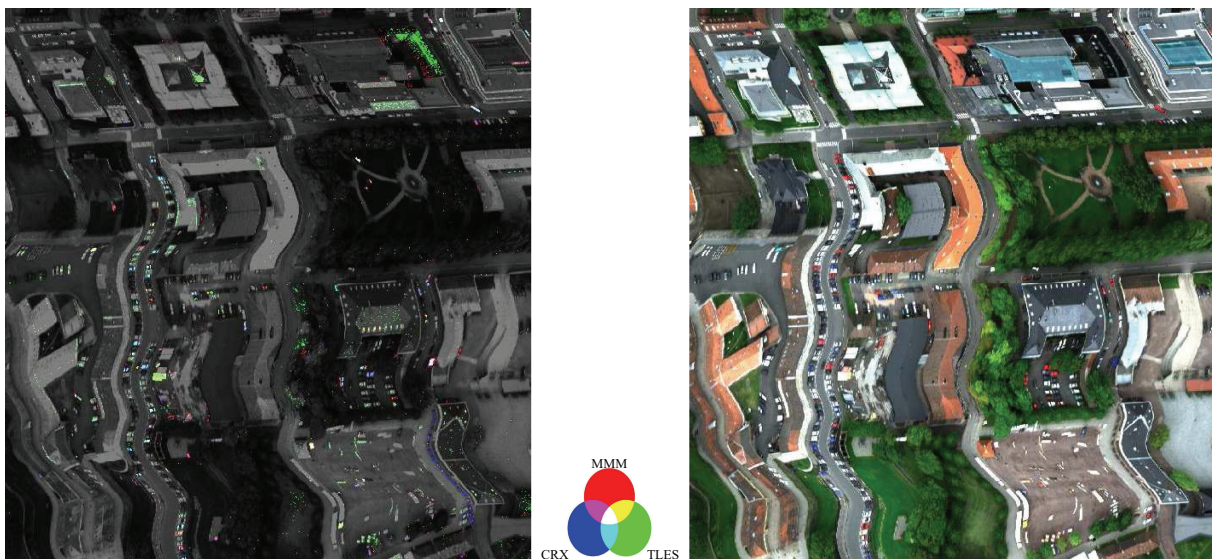

FIGURE 13: Left: color composite of the detection results for MMM, CRX, and TLES at lowest 1st detection threshold, superimposed on B/W image of OSLO2 (R: MMM results, G: CRX results, B: TLES results).

of shadow. In this paper this was done using a simplified approach consisting in square root transforming the data. After this transform MMM gives the best overall results, followed by LRX and CRX. For some targets LRX gives significantly better results than MMM. These three best detectors produce different false alarms while producing a common detection for all but one of the targets. They are thus complementary to each other and fusion of their results should be beneficial.

In the urban environment the SBAD methods perform best. The overall best result for the urban scene is obtained by MMM, TLES, and CRX. Of the globl RX-based methods OSPRX gives the best results in this dataset. Subjective evaluation of the detection results shows that the best performing detectors give complimentary results, and that "false alarms" are mainly due to objects with anomalous spectra in the scene such as cars and parts of buildings.

Further investigation of this complementarity may lead to efficient detector fusion.

Appendix

Exploratory Data Analysis

The main aim of the exploratory data analysis is to investigate how well the different datacubes comply with the assumption of unimodal multivariate normality. If a distribution is multivariate normal, the square of the Mahalanobis distances of its samples follows a χ^2 distribution with degrees of freedom equal to the dimension of the multivariate variable [40]. The compliance can then be investigated visually using a Q-Q plot of the empirical cumulative distribution function (CDF) of the Mahalanobis distance and the CDF of the theoretical χ^2 distribution. Figure 14 shows the Q-Q plots
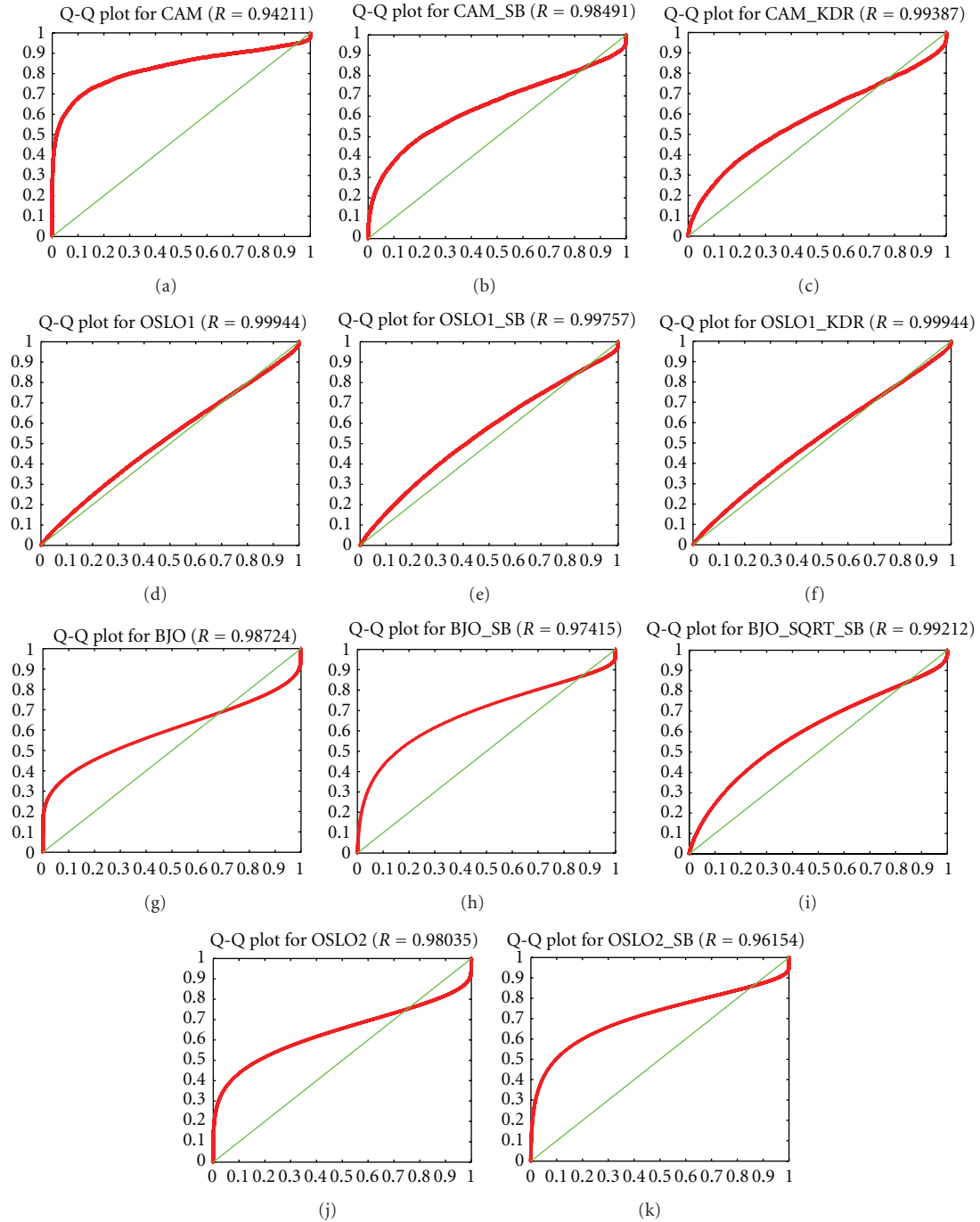
FIGURE 14: Q-Q plots for the different data cubes used in this paper.

for the four scenes and the different used preprocessing methods. Table 5 shows the correlation coefficient between the empirical and theoretical CDF of the Mahalanobis distance as well as the maximal deviation between the two (the Kolmogorov-Smirnov test statistic).

As can be expected the OSLO1 scene, consisting of a very homogeneous background, conforms best to the assumption of global multivariate normality. Both the original data

and the data after kurtosis data reduction have a high correlation coefficient between the two CDFs and a low value for the KS-statistic. None of the other datasets comply with the global normality assumption. For the CAM scene the multivariate normality improves by preprocessing and the best normality is achieved after kurtosis data reduction when all kurtosis components are considered. When only the first five components are considered, the global normality
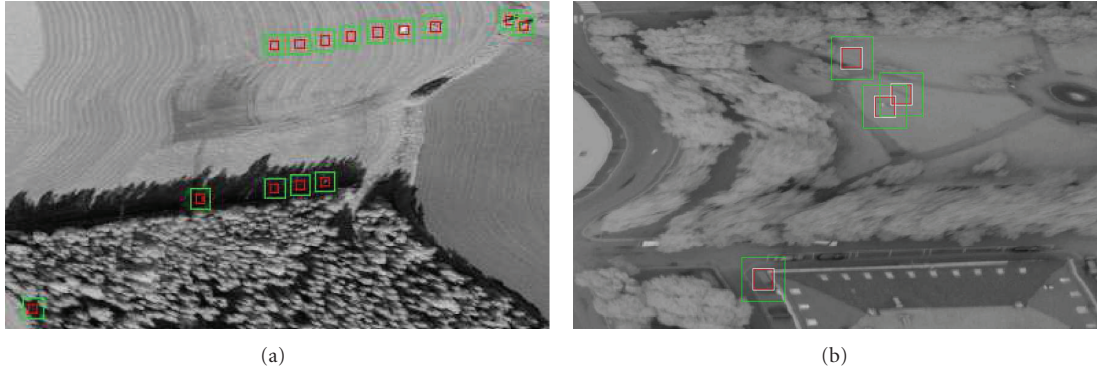
(a)   (b)

FIGURE 15: Guard window (red) and outer window used for calculating local Σ centered at each target location for BJO (a) and OSLO2 (b).

TABLE 5: Correlation coefficient and Kolmogorov-Smirnov test statistic between empirical and theoretical CDF of the Mahalanobis distance and the condition number of global Σ.

Name	Preprocessing	Correlation	KS statistic	Condition number of Σ
CAM	None	0.937	0.579	$1.11e + 10$
	SB	0.984	0.295	$7.23e + 08$
	KDR	0.993	0.178	$2.91e + 08$
	KDR (5 kcp)	0.892	0.540	$2.95e + 00$
OSLO1	None	0.999	0.049	$9.11e + 03$
	SB	0.997	0.093	$8.32e + 03$
	KDR	0.999	0.049	$6.73e + 00$
	KDR (5 kcp)	0.973	0.251	$4.91e + 00$
BJO	None	0.986	0.278	$1.63e + 05$
	SB	0.973	0.343	$1.03e + 05$
	SQRT + SB	0.992	0.188	$1.06e + 05$
OSLO2	None	0.979	0.336	$3.38e + 04$
	SB	0.961	0.409	$2.27e + 04$

assumption is not met. For BJO the combination of the square root transform and spectral binning improves the normality of the data.

Several of the investigated anomaly detection methods (the global RX-based methods and LRX) rely on the estimation and inversion of the spectral covariance matrix Σ. It is known that the sample covariance matrix in many cases needs to be regularized before inversion [41, 42]. The regularization makes the problem of finding the inverse mathematically stable, but if the initial matrix does not have a stable inverse, that is, is well-conditioned, the obtained inverse might not lead to a good detection result for the detector. In Table 5 the condition numbers of the covariance matrices of the complete images are also given. The condition number is the ratio between the highest and lowest singular value and it provides an indication of the accuracy of the results of matrix inversion. In CAM and OSLO1 the various preprocessing methods appear to decrease the condition number. This decrease is particularly significant for KDR in the

OSLO1 scene. For KDR, contrary to the normality assumption, considering only the first five components reduces the condition number substantially.

The local anomaly detection methods estimate the characteristics of the background in a local window around the current pixel. For LRX the data in that local window are supposed to follow a multivariate normal distribution. In order to assess the validity of this assumption, the normality was checked in the neighborhood of each target in the different scenes. The neighborhood is defined in the same way as for the actual LRX detector. Table 6 shows the same estimators of normality as well as the condition number as in Table 5 but based on a local background estimation. As there are different targets in each of the scenes, the average and standard deviation over all targets is given. One can notice that for the BJO and OSLO2 scene the local normality assumption is much better met than the global one. For OSLO1 and BJO the condition number of local Σ is also better than that of the global Σ. The large standard deviation of the condition numbers are due to some targets for which local covariance matrix has a very high condition number. For CAM, this is the case for three targets and the median of the local condition number in that scene is 5.3×10^9 (no preprocessing). For the scenes with a limited number of targets this is explored in more detail in Table 7. The table shows the values obtained for the three estimators in the local window around each of the targets. Figure 15 shows the guard window (red) and the outer window (green) used for estimating local Σ superimposed on a grayscale representation of the two scenes. The green window is the outer window used when no data reduction is applied. For BJO and OSLO2, where the guard window is 15×15 and the number of bands $n = 80$, this means that the outer window has a size of 33×33. In OSLO2, LRX is applied after KDR and only 5 kurtosis components are kept. This results in an outer window of 17×17 represented by the white squares in Figure 15(b).

From Table 7 it appears that the normality assumption in the BJO scene is best obeyed for the local neighborhood of targets T8–T13. T3 has the most heterogeneous background, as can be also be seen in Figure 15(a). T1 and T2 deviate from the normality assumption because of target contamination: the outer windows overlap part of the adjacent target. It is

TABLE 6: Correlation coefficient and Kolmogorov-Smirnov test statistic between empirical and theoretical CDF of the Mahalanobis distance around each of the targets and the condition number of local Σ.

Name	Preprocessing	Correlation mean ± std	KS statistic mean ± std	Condition number of local Σ mean ± std
CAM	None	0.95 ± 0.03	0.48 ± 0.17	$1.15e + 10 \pm 2.3e + 10$
	SB	0.98 ± 0.01	0.28 ± 0.11	$3.97e + 09 \pm 8.4e + 09$
	KDR	0.992 ± 0.004	0.15 ± 0.04	$8.25e + 08 \pm 2.6e + 08$
	KDR (5 kcp)	0.991 ± 0.011	0.128 ± 0.07	$6.08e + 01 \pm 1.04e + 02$
OSLO1	None	0.999 ± 0.0003	0.04 ± 0.009	$6.33e + 03 \pm 5.6e + 03$
	SB	0.998 ± 0.0008	0.068 ± 0.01	$3.02e + 03 \pm 2.5e + 03$
	KDR	0.9995 ± 0.0003	0.04 ± 0.009	$3.90e + 01 \pm 3.3e + 01$
	KDR (5 kcp)	0.993 ± 0.003	0.11 ± 0.03	$1.20e + 01 \pm 2.0e + 01$
BJO	None	0.994 ± 0.02	0.076 ± 0.10	$7.34e + 04 \pm 1.4e + 05$
	SB	0.996 ± 0.006	0.089 ± 0.06	$5.38e + 04 \pm 8.6e + 04$
	SQRT + SB	0.996 ± 0.006	0.089 ± 0.06	$5.38e + 04 \pm 8.6e + 04$
OSLO2	None	0.996 ± 0.006	0.097 ± 0.06	$3.27e + 04 \pm 1.0e + 04$
	SB	0.996 ± 0.005	0.102 ± 0.05	$2.08e + 04 \pm 1.8e + 04$

TABLE 7: Correlation coefficient and Kolmogorov-Smirnov test statistic between empirical and theoretical CDF of the Mahalanobis distance calculated in a local window around each target and the condition number of local Σ for BJO and OSLO2, both without preprocessing.

TGT	Correlation	KS statistic	Condition number of local Σ
		BJO	
T1	0.9976	0.121	$1.746e + 05$
T2	0.9968	0.104	$1.884e + 05$
T3	0.9329	0.420	$5.405e + 05$
T4	0.9991	0.055	$6.775e + 03$
T5	0.9995	0.043	$6.864e + 03$
T6	0.9994	0.052	$8.845e + 03$
T7	0.9992	0.057	$6.654e + 03$
T8	0.9999	0.027	$1.345e + 04$
T9	0.9999	0.022	$1.264e + 04$
T10	0.9998	0.033	$1.144e + 04$
T11	0.9997	0.031	$1.134e + 04$
T12	0.9997	0.032	$1.426e + 04$
T13	0.9999	0.021	$1.123e + 04$
		OSLO2	
T1	0.9995	0.052	$3.780e + 04$
T2	0.9986	0.096	$4.352e + 04$
T3	0.9991	0.052	$3.021e + 04$
T4	0.9874	0.190	$1.935e + 04$

well known that target contamination degrades the results of detectors [42, 43]. For the targets in the shadow at the edge of the forest (T4–T7) the normality assumption is reasonably well met and the condition number of the local covariance matrix is lower than for the other targets. The outer windows for these targets fall entirely in the shadow zone.

Table 7 shows that in OSLO2 the assumption of local normality is best met for targets T1 and T3 while for T4 the assumption is less valid. The corresponding figure shows that T4 has indeed the most heterogeneous local background. There is some target contamination between T2 and T3 when no data reduction is applied, while this is not the case when KDR is applied and only five kurtosis components are used.

Acknowledgments

References

[1] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 58–69, 2002.

[2] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerospace and Electronic Systems Magazine*, vol. 25, no. 7, pp. 5–27, 2010.

[3] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1760–1770, 1990.

[4] C. I. Chang and S. S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 6, pp. 1314–1325, 2002.

[5] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: a nonlinear anomaly detector for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 2, pp. 388–397, 2005.

[6] A. Schaum, "Advanced methods of multivariate anomaly detection," in *Proceedings of the IEEE Aerospace Conference*, pp. 1–7, March 2007.

[7] A. P. Schaum, "Hyperspectral anomaly detection beyond RX," in *Proceedings of the SPIE Algorithms and Technologies for Multispectral, Hyperspectral and Ultraspectral Imagery XII*, vol. 6565, 2007.

[8] E. Lo and L. T. C. John Ingram, "Hyperspectral anomaly detection based on minimum generalized variance method," in *Proceedings of the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIV*, vol. 6966, March 2008.

[9] C.E. Caefer, J. Silverman, O. Orthal, D. Antonelli, Y. Sharoni, and S.R. Rotman, "Improved covariance matrices for point target detection in hyperspectral data," *Optical Engineering*, vol. 47, no. 7, Article ID 076402, 2008.

[10] E. Lo and A. Schaum, "A hyperspectral anomaly detector based on partialling out a clutter subspace," in *Proceedings of the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV*, vol. 7334, April 2009.

[11] E. Lo, "Maximized subspace model for hyperspectral anomaly detection," *Pattern Analysis and Applications*, pp. 1–11, 2011.

[12] E. Lo, "Variable subspace model for hyperspectral anomaly detection," *Pattern Analysis and Applications*, pp. 1–13, 2011.

[13] I. Kåsen, P. E. Goa, and T. Skauli, "Target detection in hyperspectral images based on multi-component statistical models for representation of background clutter," in *Proceedings of the SPIE on Electro-Optical and Infrared Systems: Technology and Applications*, vol. 5612, pp. 258–264, October 2004.

[14] D. Blumberg, E. Ohel, and S. Rotman, "Anomaly detection in noisy multi- and hyperspectral images of urban environments," in *Proceedings of the 3rd GRSS/ISPRS Symposium*, Tempe, Ariz, USA, March 2005.

[15] C. Willis, "Anomaly detection in hyperspectral imagery using statistical mixture models," in *Proceedings of the 2nd EMRS DTC Technical Conference*, Edinburgh, UK, 2005.

[16] O. Duran and M. Petrou, "A time-efficient method for anomaly detection in hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3894–3904, 2007.

[17] O. Duran and M. Petrou, "Spectral unmixing with negative and superunity abundances for subpixel anomaly detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 1, pp. 152–156, 2009.

[18] T. Veracini, S. Matteoli, M. Diani, and G. Corsini, "Fully unsupervised learning of Gaussian mixtures for anomaly detection in hyperspectral imagery," in *Proceedings of the 9th International Conference on Intelligent Systems Design and Applications (ISDA '09)*, pp. 596–601, December 2009.

[19] D. Borghys, E. Truyen, M. Shimoni, and C. Perneel, "Anomaly detection in hyperspectral images of complex scenes," in *Proceedings of the 29th Symposium of the European Association of Remote Sensing Laboratories*, Chania, Greece, June 2009.

[20] Y. Tarabalka, T. V. Haavardsholm, I. Kåsen, and T. Skauli, "Real-time anomaly detection in hyperspectral images using multivariate normal mixture models and GPU processing," *Journal of Real-Time Image Processing*, vol. 4, no. 3, pp. 287–300, 2009.

[21] D. Borghys, I. Kasen, V. Achard, and C. Perneel, "Comparative evaluation of hyperspectral anomaly detectors in different types of background," in *Proceedings of the SPIE Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII*, vol. 8390, Baltimore, April 2012.

[22] C. I. Chang, "Orthogonal subspace projection (OSP) revisited: a comprehensive study and analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 502–518, 2005.

[23] P. Masson and W. Pieczynski, "SEM algorithm and unsupervised statistical segmentation of satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 31, no. 3, pp. 618–633, 1993.

[24] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, pp. III250–III253, July 2008.

[25] S. S. Shen and E. M. Bassett, "Information theory based band selection and utility evaluation for reflective spectral systems," in *Proceedings of the SPIE Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII*, vol. 4725, pp. 18–29, April 2002.

[26] I. Kåsen, A. Rødningsby, T. V. Haavardsholm, and T. Skauli, "Band selection for hyperspectral target detection based on a multinormal mixture anomaly detection algorithm," in *Proceedings of the SPIE Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIV*, vol. 6966, March 2008.

[27] D. Peña, F. J. Prieto, and J. Viladomat, "Eigenvectors of a kurtosis matrix as interesting directions to reveal cluster structure," *Journal of Multivariate Analysis*, vol. 101, no. 9, pp. 1995–2007, 2010.

[28] E. A. Ashton, B. D. Wemett, R. A. Leathers, and T. V. Downes, "A novel method for illumination suppression in hyperspectral images," in *Proceedings of the SPIE Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIV*, vol. 6966, March 2008.

[29] R. Richter and A. Müller, "De-shadowing of satellite/airborne imagery," *International Journal of Remote Sensing*, vol. 26, no. 15, pp. 3137–3148, 2005.

[30] S. M. Adler-Golden, M. W. Matthew, G. P. Anderson, G. W. Felde, and J. A. Gardner, "An algorithm for de-shadowing spectral imagery," in *Proceedings of the 11th JPL Airborne Earth Science Workshop*, pp. 3–4, March 2002.

[31] M. Shimoni, G. Tolt, C. Perneel, and J. Ahlberg, "Detection of vehicles in shadow areas using combined hyperspectral and lidar data," in *Proceedings of the Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 4427–4430, Vancouver, Canada, July 2011.

[32] B. D. Wemett, J. K. Riek, and R. A. Leathers, "Dynamic thresholding for hyperspectral shadow detection using Levenberg-Marquardt minimization on multiple gaussian illumination distributions," in *Proceedings of the SPIE Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV*, vol. 7334, April 2009.

[33] T. Skauli, "Sensor noise informed representation of hyperspectral data, with benefits for image storage and processing," *Optics Express*, vol. 19, no. 14, pp. 13031–13046, 2011.

[34] P. McCullagh and J. Nelder, *Generalized Linear Models*, chapter 6, CRC Press, 1999.

[35] J. Harsanyi, W. Farrand, and C. Chang, "Determining the number and identity of spectral endmembers: an integrated approach using neyman pearson eigenthresholding and iterative constrained rms error minimization," in *Proceedings of the 9th Thematic Conference on Geologic Remote Sensing*, February 1993.

[36] C. I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 608–619, 2004.

[37] J. Bioucas-Diaz and M. Nascimiento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, 2008.

[38] N. Acito, M. Diani, and G. Corsini, "A new algorithm for robust estimation of the signal subspace in hyperspectral images in the presence of rare signal components," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3844–3856, 2009.

[39] N. Acito, "Hyperspectral signal subspace identification in the presence of rare signal components," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 4, pp. 1940–1954, 2010.

[40] B. Manly, Ed., *Multivariate Statistical Methods*, Chapman and Hall, Boca Raton, Fla, USA, 1995.

[41] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, 2004.

[42] S. Matteoli, M. Diani, and G. Corsini, "Different approaches for improved covariance matrix estimation in hyperspectral anomaly detection," in *Proceedings of the Annual Meeting of the Italian National Telecommunications and Information Theory Group (GTTI '09)*, 2009.

[43] J. Theiler and B. R. Foy, "Effect of signal contamination in matched-filter detection of the signal on a cluttered background," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 98–102, 2006.