LA-UR- 11-06767

Approved for public release;
distribution is unlimited.

Title:	Evaluating and Improving Local Hyperspectral Anomaly Detectors
Author(s):	Leonardo R. Bachega, Purdue University James Theiler, Los Alamos National Laboratory Charles A. Bouman, Purdue University
Intended for:	Proceedings of AIPR (Applied Imagery Pattern Recognition) Meeting date: 11-13 October 2011 Meeting location: Washington DC, USA Paper due date: 30 November 2011

Form 836 (7/06)

Evaluating and Improving Local Hyperspectral Anomaly Detectors

Leonardo R. Bachega
School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN 47907
lbachega@purdue.edu

James Theiler
Space and Remote Sensing Group
Los Alamos National Laboratory
Los Alamos, NM 87545
jt@lanl.gov

Charles A. Bouman
School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN 47907
bouman@purdue.edu

***Abstract*—This paper addresses two issues related to the detection of hyperspectral anomalies. The first issue is the evaluation of anomaly detector performance even when labeled data is not available. The second issue is the estimation of the covariance structure of the data in local detection methods, such as the RX detector, when the number of available training pixels n is not much larger than (and may even be smaller than) the data dimensionality p.**

Our first contribution is to formulate and employ a mean-log-volume approach for evaluating local anomaly detectors. Traditionally, the evaluation of a detector's accuracy has been problematic. Anomalies are loosely defined as pixels that are unusual with respect to the other pixels in a local or global context. This loose definition makes it easy to develop anomaly detection algorithms – and many have been proposed – but more difficult to evaluate or compare them. Our mean-log-volume approach allows for an effective evaluation of a detector's accuracy without requiring labeled testing data or an overly-specific definition of an anomaly.

The second contribution is to investigate the use of the Sparse Matrix Transform (SMT) to model the local covariance structure of hyperspectral images. The SMT has been previously shown to provide full rank estimates of large covariance matrices even in the $n < p$ scenario. Traditionally, the number of training pixels needed for good estimates of the covariance needs to be at least as large as the data dimensionality (and preferably it should be several times larger). Therefore, when one deploys the RX detector in a sliding window, the choices to select small window sizes are limited because of the $n > p$ restriction associated to the covariance estimation. Our results suggest that RX-style detectors using the SMT covariance estimates perform favorably compared to other methods even (indeed, especially) in the regime of very small window sizes.

I. Introduction

Anomaly detection promises the impossible: it is target detection without knowing anything about the target. In the context of hyperspectral imagery, the anomalous pixels are those that are unusual with respect to the other pixels in a local or global context. A number of anomaly detectors have been developed for hyperspectral datasets, many of which are surveyed by Stein *et. al.* [1], and more recently by Matteoli *et. al.* [2]

Local detectors form an important class of algorithms. They work using a statistical model of the background pixels in the local neighborhood of the pixel under test. In general, only the pixels within a sliding window are used to estimate properties of the local context. To the extent that the background statistical properties are non-stationary across the image, this local statistical characterization has the potential to improve the detection accuracy. One problem with these local methods is that the number of training samples (pixels), n, needed for a good estimate of the covariance must be at least as large as the data dimensionality (number of spectral bands), p, and preferably should be several times larger than p. [3] This $n \gg p$ requirement rules out small window sizes. The potential increase in detection accuracy due to the local characterization of the background (in a small window) is compromised by the lack of adequate training samples needed to estimate the covariance.

Another way to address the covariance estimation problem is to use the Sparse Matrix Transform (SMT). The SMT provides full rank estimates of large covariance matrices even when the number of training samples n is smaller than the data dimensionality p. [4] We have recently shown that the SMT improves the accuracy of "global" anomaly detectors. [5] In this paper, we suggest that RX-style detectors using the SMT covariance estimates perform favorably compared to other methods, even in the regime of very small window sizes.

The rest of this paper is organized as follows: Section II formulates the anomaly detection task and reviews the most commonly used covariance estimation methods used in anomaly detection; Section III describes the SMT covariance estimation and how the SMT estimates yield

highly accurate detectors even when small window sizes are used; Section IV introduces the mean-log-volume as a measure of detection accuracy and show how it can be used to select the window size that maximizes the detection accuracy; Section V presents our main experimental results. Finally, Section VI presents the main conclusions.

II. Hyperspectral Anomaly Detection

Hyperspectral anomaly detection consists in finding pixel regions (objects) in the hyperspectral image with pixels that differ substantially from the background, *i.e.,* the pixels in the regions surrounding these objects.

In general, there is no precise definition of what constitutes an anomaly. A common way of defining anomalies is to say that *anomalies are not concentrated.* [6] Here we assume that anomalous samples are drawn from a broad, uniform distribution with a much larger support than the distribution of typical (*i.e.,* not anomalous) samples. This assumption allows us to describe anomaly detection in terms of a binary classification problem.

A. Anomaly Detection as Binary Classification

Let x be a p-dimensional random vector. We want to classify x as *typical* if it is drawn from a multivariate Gaussian distribution $\mathcal{N}(\mu, R)$, or as *anomalous* if it is drawn from a uniform distribution $\mathcal{U}(x) = c$, where c is some constant. Formally, we have the following hypotheses:

$$\begin{aligned} \mathcal{H}_0 &: \quad \mathrm{x} \sim \mathcal{N}(\mu, R) \\ \mathcal{H}_1 &: \quad \mathrm{x} \sim \mathcal{U}, \end{aligned} \tag{1}$$

where $\mathcal{H}_0$ and $\mathcal{H}_1$ are referred as the *null* and *alternative* hypotheses respectively. According to the *Neyman-Pearson* lemma [7], optimal classifier has the form of a log-likelihood ratio test

$$l(\mathrm{x}) = \log\left\{\frac{p(\mathrm{x};\mathcal{H}_1)}{p(\mathrm{x};\mathcal{H}_0)}\right\} \gtrless l_0, \tag{2}$$

that maximizes the probability of detection, $p(\mathcal{H}_1;\mathcal{H}_1)$ for a fixed probability of false alarm, $p(\mathcal{H}_1;\mathcal{H}_0)$, which is controlled by the threshold l_0.

The log-likelihood ratio test in (2) can be written as

$$\begin{aligned} l(\mathrm{x}) &= \log\left\{\frac{p(\mathrm{x};\mathcal{H}_1)}{p(\mathrm{x};\mathcal{H}_0)}\right\} = \log c - \log p(\mathrm{x};\mathcal{H}_0) \\ &= \log c + \frac{p}{2}\log 2\pi + \frac{1}{2}\log|R| \\ &\quad + \frac{1}{2}(\mathrm{x}-\mu)^t R^{-1}(\mathrm{x}-\mu) \gtrless l_0 \end{aligned} \tag{3}$$

We can incorporate the constant terms in (3) together with l_0 into a new threshold, η, such that the significance test in (3) is equivalent to the test

$$D_R(\mathrm{x}) = \sqrt{(\mathrm{x}-\mu)^t R^{-1}(\mathrm{x}-\mu)} \gtrless \eta. \tag{4}$$

The statistic $D_R(\mathrm{x})$ is interpreted as the Mahalanobis distance between the sample x and the mean μ of the background distribution. If such distance exceeds the threshold η, we label x as an *anomaly*.

In practice, one does not know the true parameters μ and R of the background pixel distribution $\mathcal{N}(\mu, R)$. In order to compute the statistic $D_R(\mathrm{x})$ in (4) , the practitioner needs first to compute good estimates $\hat{\mu}$ and $\hat{R}$ of μ and R respectively, from the samples (pixels) available.

B. Sliding Window-based Detection

The RX detection algorithm [8], [9] uses a sliding window centered at the pixel x, as illustrated in Fig. 1. The window pixels are used to compute the covariance estimate $\hat{R}$ of the background. As argued in [2] the pixels closest to x within the *Guard window* are left out of the estimation to avoid contaminating the estimate with potentially anomalous pixels. The dimension of the guard window is chosen according to the expected maximum size of an anomalous object. An interesting variation of the RX detector (not investigated here) uses a third window around x, larger than the guard window but smaller than the outer window, to estimate the mean μ. [2] The motivation is that a good estimate of the mean requires fewer pixels than a good estimate of the covariance.

The pixels within the outer window are used as the training pixels in the estimation of the covariance matrix R. The choice of the window size is a compromise between two factors: (i) The window should be small enough that it covers a homogeneous region of the background, therefore, being accurately modeled by the multivariate Gaussian $\mathcal{N}(\mu, R)$; (ii) The window should be large enough that the number of pixels within the outer window is enough to produce reliable estimates of the covariance R. At least $p+1$ pixels are required for non-singular sample covariance estimates.

C. Covariance Estimation Methods

In this section, we discuss some of the methods used to estimate the covariance matrix R.

1) Sample Covariance: Let $X = [\mathrm{x}_1, \cdots, \mathrm{x}_n]$ be the set of n i.i.d. p-dimensional Gaussian random vectors drawn from $\mathcal{N}(0, R)$. The sample covariance S is given by

$$S = \frac{1}{n} X X^t.$$

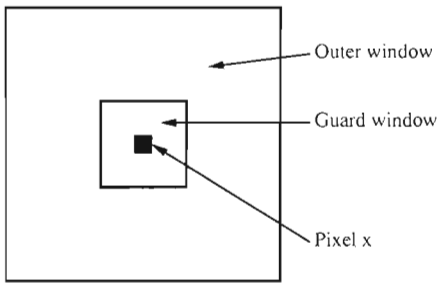

Fig. 1. Square sliding window used in the RX detection algorithm. The pixels in the outer window are used to compute the covariance estimate $\hat{R}$ of the background surrounding the pixel x. The pixels within the inner window (referred as the *guard window* are not used in the covariance computation to avoid that potential anomalous pixels contaminate the estimate $\hat{R}$.

which is the unconstrained maximum likelihood estimate of R. [7]

When $n < p$, the sample covariance S is singular, with rank n and *overfits* the data. As argued in [3], [2], in the case of hyperspectral data, it is usually desirable to have $n \geq 10p$ so that S is a reliable estimate of R. But even when n is small and S is by itself unreliable, the sample covariance is still useful as a starting point for the regularized shrinkage estimates reviewed below as well as the SMT introduced in Section III.

2) Diagonal: Because it is the inverse of R that is used in (4) , it is important that the estimate of R be full-rank. A simple way to obtain a full-rank estimate of R with a small number of samples n (especially when $n < p$) is to treat all the p dimensions as uncorrelated and simply estimate the variances for each of the p coordinates. This results in the estimator

$$D = diag(S),$$

which is generally of full-rank and can be well estimated even with small n. However, D tends to *underfit* the the data since the assumptions that the coordinates are uncorrelated is typically unrealistic.

3) Shrinkage: The shrinkage estimation is a very popular method of regularizing estimates of large covariance matrices. [10], [11], [12] It is based on the combination of the sample covariance matrix S that *overfits* the data with another estimator T (called the shrinkage target) that *underfits* the data:

$$\hat{R} = (1 - \alpha)S + \alpha T, \tag{5}$$

where $\alpha \in [0, 1]$. The choice of the value α that maximizes the likelihood of the estimate $\hat{R}$ is typically done through a cross-validation procedure.

The most common variation of the shrinkage method [10], [11] uses $\sigma^2 I$ as the shrinkage target, where σ^2 is the average variance across all the p dimensions and I is the $p \times p$ identity matrix. The covariance estimator is given by

$$\hat{R} = (1 - \alpha)S + \alpha\sigma^2 I. \tag{6}$$

A variation of (5) proposed by Hoffbeck and Landgrebe [12] uses $D = diag(S)$ as the shrinkage target, resulting in the following shrinkage estimator

$$\hat{R} = (1 - \alpha)S + \alpha D. \tag{7}$$

The authors in [12] also propose a computationally efficient leave-one-out cross-validation (LOOC) scheme to estimate α in (7) .

4) Quasilocal Covariance: This method proposed by Caefer *et. al.* [13] considers the eigen-decomposition of the covariance matrix $R = E\Lambda E^t$, and makes the observation that the eigenvalues in the matrix Λ are more likely to change across different image locations while the eigenvectors in E remain mostly pointed to the same directions across the entire image.

The observation above suggests that one can obtain a global estimate of the eigenvector matrix E using all the pixels in the image, and then can adjust the eigenvalues in Λ locally by computing the variances independently in each direction using only pixels that are within the sliding window. Since the number of pixels in the entire image, we typically have $n \gg p$, and so the sample covariance S will provide a full-rank global estimate and its eigenvectors, $\hat{E}_{global}$ can be used as the estimates of E across all positions of the sliding window. Finally, the estimate of the matrix Λ is estimated locally at each position of the sliding window, by computing variances in each of the global eigenvector directions. This approach results in the *quasilocal* estimator of covariance:

$$\hat{R} = \hat{E}_{global}\hat{\Lambda}_{local}\hat{E}^t_{global}.$$

III. The Sparse Matrix Transform (SMT)

The Sparse Matrix Transform (SMT) [4], [5] can be used to provide full-rank estimates of the covariance matrix R used in the detection framework in Section II. The method decomposes the true covariance R into the product $R = E\Lambda E^t$, where E is the orthonormal matrix containing the eigenvectors of R and Λ is a diagonal matrix containing the eigenvalues of R. The SMT then provides the estimates $\hat{E}$ and $\hat{\Lambda}$ with the diagonal elements of $\hat{\Lambda}$ being strictly positive.

A. SMT Covariance Estimation

Given a training set with n independent p-dimensional i.i.d random vectors drawn from the multivariate Gaussian $\mathcal{N}(0, R)$, and organized into the data matrix $X = [\mathrm{x}_1, \cdots, \mathrm{x}_n]$. The Gaussian likelihood of observing the data X is given by

$$l(X; R) = \frac{|R|^{-n/2}}{(2\pi)^{np/2}} \exp\left\{-\frac{1}{2} trace(R^{-1}S)\right\} \quad (8)$$

where $S = \frac{1}{n}XX^t$ is the sample covariance, a sufficient statistic for the likelihood of the data X. The joint maximization of (8) with respect to E and Λ results in the maximum likelihood (ML) estimates

$$\hat{E} = \arg\min_{E \in \Omega_K} \left\{\left|\mathrm{diag}(E^t S E)\right|\right\} \quad (9)$$

$$\hat{\Lambda} = \mathrm{diag}(\hat{E}^t S \hat{E}) \ , \quad (10)$$

where Ω_K is the set of allowed orthonormal transforms.

If $n > p$, and the set Ω_K includes all orthonormal transforms, then the solution to (9) and (10) is given by the sample covariance; *i.e*, $\hat{E}\hat{\Lambda}\hat{E}^t = S$. However, as discussed in Section II, when $n < p$, the sample covariance, S overfits the data and is a poor estimate of the true covariance R.

In order to regularize the covariance estimate, we impose the constraint that Ω_K be the set of sparse matrix transforms (SMT) or order K. More specifically, we will assume that the eigen-transformation has the form

$$E_K = \prod_{k=1}^{K} G_k = G_1 \cdots G_K \in \Omega_K \ , \quad (11)$$

for a model order K. Each G_k is a *Givens rotation* [4] over some (i_k, j_k) coordinate pair by an angle θ_k,

$$G_k = I + \Theta(i_k, j_k, \theta_k),$$

where

$$[\Theta]_{ij} = \begin{cases} \cos(\theta_k) - 1 & \text{if } i = j = i_k \text{ or } i = j = j_k \\ \sin(\theta_k) & \text{if } i = i_k \text{ and } j = j_k \\ -\sin(\theta_k) & \text{if } i = j_k \text{ and } j = i_k \\ 0 & \text{otherwise} \end{cases} , \quad (12)$$

and K is the model order parameter.

The optimization of (9) is non-convex, so we use a greedy optimization approach to design each rotation, G_k, in sequence to minimize the cost [4]: Let $S_{k-1} = G_{k-1}^t S_{k-2} G_{k-1}$. At the kth step of the greedy optimization, we select the pair of coordinates (i_k, j_k) such that

$$(i_k, j_k) = \arg_{i,j} \max \left(\frac{(S_{k-1})_{ij}^2}{(S_{k-1})_{ii}(S_{k-1})_{jj}} \right) ,$$

i.e, the most correlated pair of coordinates, and choose the angle

$$\theta_k = \frac{1}{2} \tan^{-1} \left(\frac{-2(S_{k-1})_{i_k j_k}}{(S_{k-1})_{i_k i_k} - (S_{k-1})_{j_k j_k}} \right)$$

that completely decorrelates the i_k and j_k dimensions. This greedy optimization procedure can be done fast if a graphical constraint can be imposed to the data. [14]

Finally, for an SMT of order K, we have the estimates

$$\hat{E}_K = G_1 \cdots G_K \quad (13)$$

$$\hat{\Lambda}_K = \mathrm{diag}(\hat{E}_K^t S \hat{E}_K) \ , \quad (14)$$

with the covariance estimate given by

$$\hat{R}_{SMT} = \hat{E}_K \hat{\Lambda}_K \hat{E}_K^t . \quad (15)$$

B. SMT Model Order

The model order parameter K can be estimated using cross-validation [4], [14], a Wishart Criterion [5], or the minimum description length (MDL) approach derived in [5]. We used the MDL criterion for the experiments in this paper. According to the MDL criterion, we select the smallest value of K such that the following inequality is satisfied:

$$\max_{ij} \left(\frac{[S_K]_{ij}^2}{[S_K]_{ii}[S_K]_{jj}} \right) \leq 1 - \exp\left(\frac{-\log n - 5 \log p}{n} \right) \ ,$$

where $S_K = \hat{E}_K^t S \hat{E}_K$.

It is often useful to express the order of the SMT as $K = rp$, where r is the average number of rotations per coordinate, being typically very small ($r < 5$) for several previously studied datasets.

C. Shrinkage SMT

The SMT covariance estimate in (15) can be used as a shrinkage target, alternative to the ones described in Section II-C3, resulting in the following Shrinkage-SMT estimate:

$$\hat{R} = (1 - \alpha) S + \alpha \hat{R}_{SMT} \ .$$

IV. Ellipsoid Mean Log-Volume

In this section, we develop the *Ellipsoid Mean Log-Volume*, a novel metric to evaluate the accuracy of anomaly detection algorithms that make detection decisions based on a Mahalanobis statistic such as D_R in (4) . Different versions of these detectors use different techniques to estimate the covariance yielding different detection accuracies depending on how well the covariance estimate $\hat{R}$ approximates the true background covariance R.

Traditionally, receiver operating characteristics (ROC) curves have been widely used to evaluate anomaly detectors. The ROC approach requires both samples labeled as *typical* and samples labeled as *anomalous* in order to estimate the both the *probability of detection* and the *probability of false alarm* used in the ROC analysis. Unfortunately, anomalies are rare events and it is often difficult to have enough data labeled as *anomalous* in order to estimate the probability of detection required in the ROC analysis.

The approach developed here seeks to characterize how well the estimates of the background model (*i.e.,* $\hat{\mu}$ and $\hat{R}$) fit the training (typical) pixel data, overcoming the limitation of the ROC analysis described above. More specifically, we evaluate the volume of the hyperellipsoid within the region

$$(\mathrm{x} - \hat{\mu})^t \hat{R}^{-1} (\mathrm{x} - \hat{\mu}) \leq \eta^2, \tag{16}$$

where η controls the probability of false alarm, as described previously. Such a volume is evaluated by the following expression:

$$V(R, \eta) = \frac{\pi^{p/2} \sqrt{|R|}}{\Gamma(1 + p/2)} \eta^p. \tag{17}$$

Smaller values of $V(R, \eta)$ indicate smaller probabilities that an anomalous data point would fall within the hyperellipsoid region of (16) . Based on this observation, the core idea in our approach is to use the value of $V(R, \eta)$ as a proxy for the probability of missed detection. Therefore, for a fixed probability of false alarm, smaller values of $V(R, \eta)$ indicate more accurate detection. Because the direct computation of $V(R, \eta)$ tends to be numerically unstable, often leading to numerical overflow for large values of p, in practice we work with $\log V(R, \eta)$ as our measure of accuracy.

This approach has been used before in global anomaly detection [5], [15], [16], but we are extending it here to local sliding window-based anomaly detection. These detectors produce a different local estimate of the background covariance at each location of the sliding window across the image. We suggest measuring detection accuracy in terms of the expected log-volume of the hyperellipsoid, $E[\log V(\hat{R}, \eta)]$ across the whole hyperspectral image, where each different estimate $\hat{R}$ is computed for each position of the sliding window using local training data pixels.

V. Experiments

All experiments in this section were performed using the*Blindrad* hyperspectral dataset, [17] a HyMap image of Cook City, MT of 800×280 pixels, each with 126 hyperspectral bands. Fig. 2 displays a RGB rendering of this dataset.

In all experiments, a sliding window like the one described in Fig. 1 moves across the image and, at each position it estimates the covariance R from the samples of the outer window using several covariance estimation methods previously discussed. Such covariance is used to compute D_R in (4) for each pixel within the guard window. The radius η is adjusted globally so that a fraction of the points corresponding to a fixed probability of false alarm is left out of the ellipsoid region. Finally we compute the expected value $E[\log V(\hat{R}, \eta)]$ over all window positions and take that as the measure of anomaly detection performance.

Fig. 3 shows the *coverage plots* with the expected log-volume of ellipsoid *vs.* the probability of false alarm for different window sizes. The hyperspectral bands of the dataset were rotated to the *Quasilocal* coordinate system by the matrix $\hat{E}^t_{global}$ (see Sec. II-C4). These "ROC-like" curves suggest that the regularized methods are more accurate, especially when small window sizes are used. When large window sizes are used, the unregularized sample covariance has its performance similar to the regularized methods.

Fig. 4 compares the performance of several detectors in both the original and the quasilocal coordinate systems at two different fixed false alarm rates. The diagonal covariance estimate performs poorly in the original coordinates (Figs. 4(a) and 4(b)), but remains a competitive method in the quasilocal coordinates (Figs. 4(c) and 4(d)); in fact, the diagonal estimator in quasilocal coordinates is just the quasilocal covariance estimator suggested by Caefer *et. al.* [13]. The Shrinkage-SMT estimates are among the best methods in both spaces, though in the quasilocal space, Shrinkage-Diagonal detectors perform just as well. When the window size used to estimate the covariance matrix grows large, we observe the increase in the expected ellipsoid log-volume; *i.e.,* the degradation of the detection accuracy for all the methods. This degradation is due to the distribution of the background pixels being non-stationary across the image. Therefore, the estimate of the covariance using large windows tends to yield poor estimates. When small window sizes are used, the training pixels are more likely to come from a homogeneous region with Gaussian distribution. Nevertheless, this is a regime where poor estimates of the covariance are due to the limited number of training samples, as observed in the curves for detectors using the sample covariance. On the other hand, the results suggest that the regularized methods perform best with smaller window sizes. Finally, the

Fig. 2. RGB rendering of the 800 × 280 pixel *Blindrad* hyperspectral dataset, captured using a HyMap sensor with 126 channels.

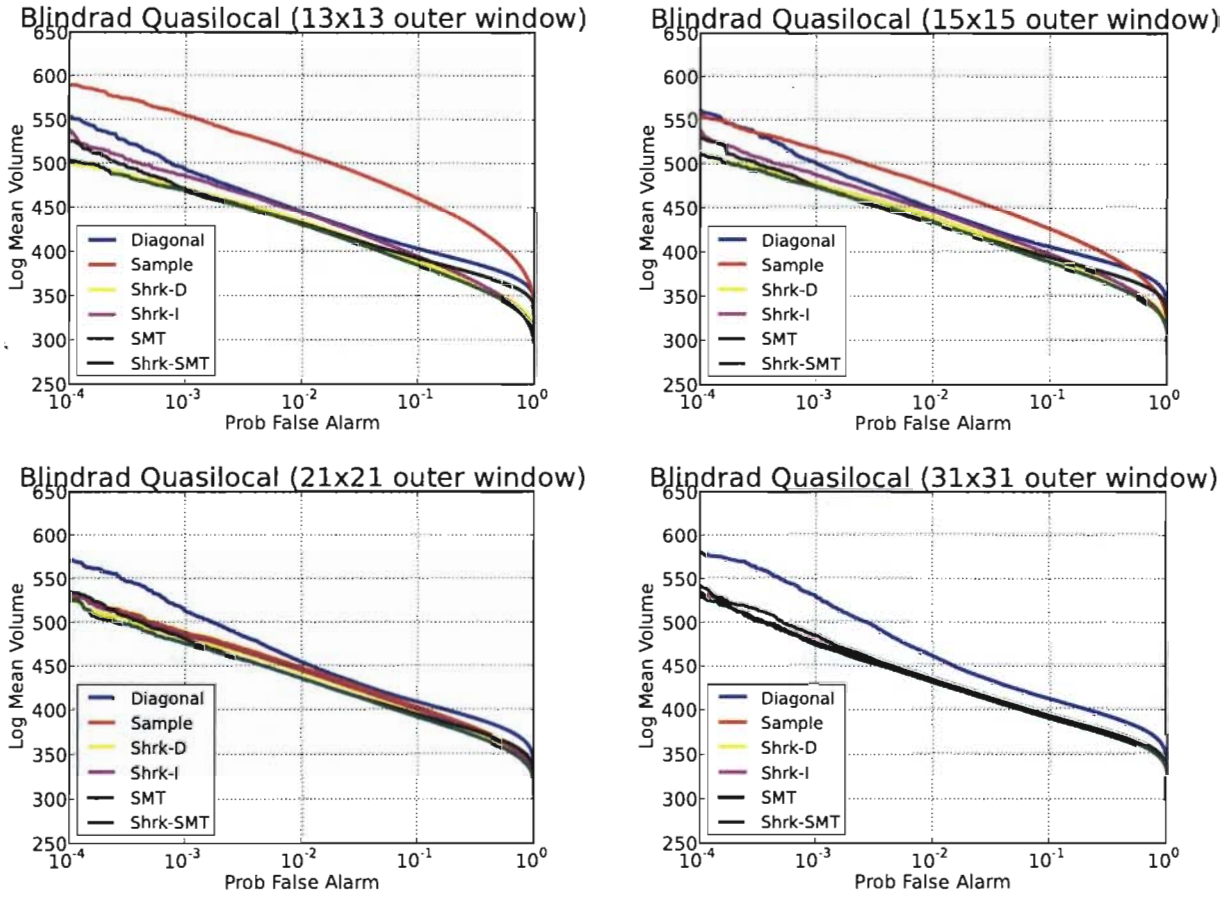

Fig. 3. Coverage plots with the expected ellipsoid log-volume *vs.* probability of false alarm for various outer window sizes.

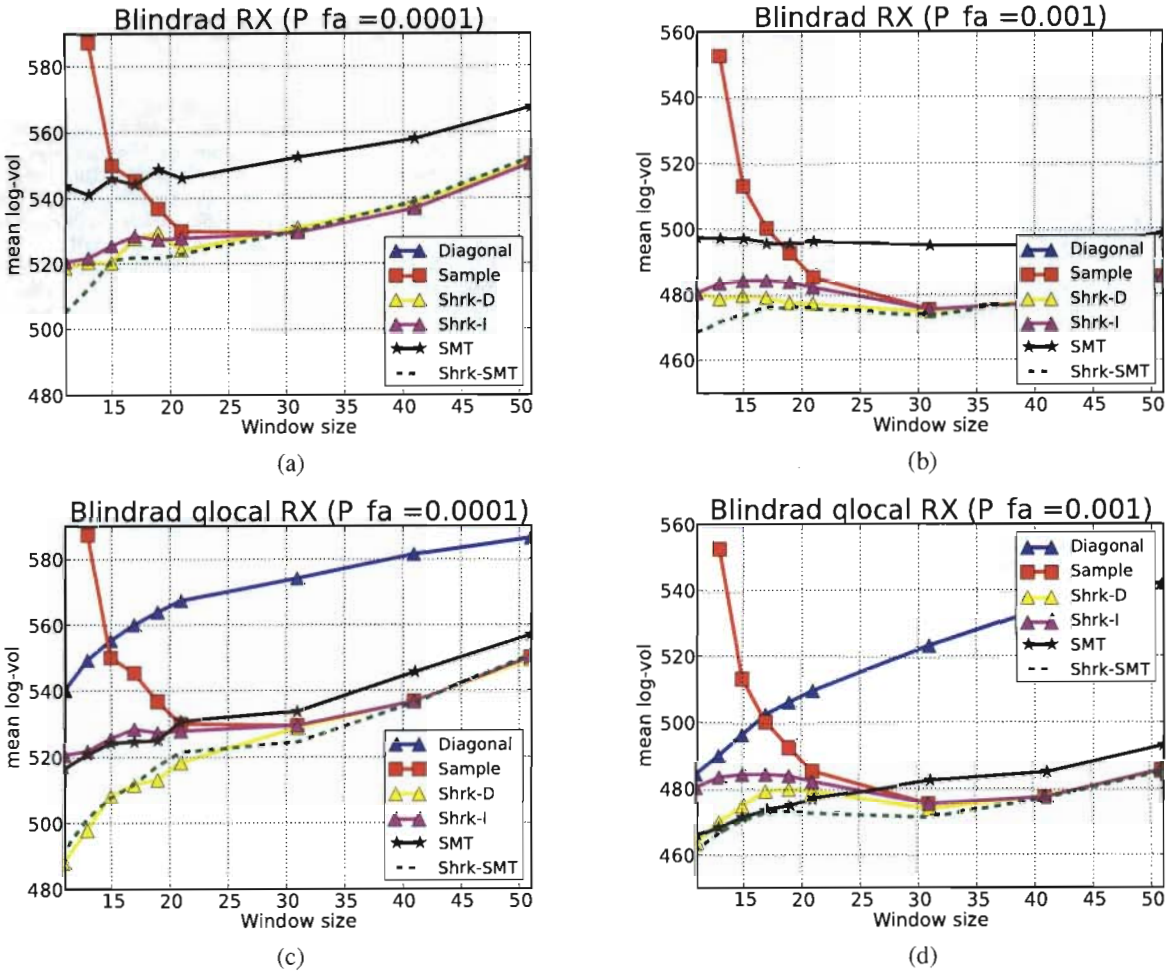

Fig. 4. Expected ellipsoid log-volume *vs.* the dimension of the sliding window fixed probabilities of false alarm in both the original, (a) and (b), and the quasilocal, (c) and (d), coordinate systems.

practitioner can use the curves in Fig. 4 as a criterion to select the window size that produces the most accurate detector for a chosen covariance estimation method.

VI. Conclusions

In this paper we have shown how to use the expected log-volume of ellipsoid to measure local detector accuracy. This measure was used to compare different detectors as well as a to provide a criterion for selecting the optimal size of the sliding window. We have also shown how to use the SMT to produce regularized covariance estimates to be used in detection. While Shrinkage-SMT often produces good results, our results show that Shrinkage-Diagonal performs just as well when combined with the quasilocal method proposed in [13]. In the future, we plan to address how to push the covariance methods to work with even smaller window sizes.

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, pp. 58–69, Jan 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. 5–28, Jul 2010.

[3] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerospace and Electronic Systems*, vol. 10, pp. 853–863, 1974.

[4] G. Cao, L. R. Bachega, and C. A. Bouman, "The sparse matrix transform for covariance estimation and analysis of high dimensional signals," *IEEE Trans. Image Processing*, vol. 20, no. 3, pp. 625–640, 2011.

[5] J. Theiler, G. Cao, L. R. Bachega, and C. A. Bouman, "Sparse matrix transform for hyperspectral image processing," *IEEE J. Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 424–437, 2011.

[6] I. Steinwart, D. Hush, and C. Scovel, "A classification framework for anomaly detection," *J. Machine Learning Research*, vol. 6, pp. 211–232, 2005.

[7] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. New Jersey: Prentice Hall, 1998, vol. II.

[8] J. Y. Chen and I. Reed, "A detection algorithm for optical targets in clutter," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-23, no. 1, pp. 46–59, Jan 1987.

[9] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1760–1770, 1990.

[10] J. H. Friedman, "Regularized discriminant analysis," *J. American Statistical Association*, vol. 84, no. 405, pp. 165–175, 1989.

[11] O. Ledoit and M. Wolf, "A well-conditioned estimator for large dimensional covariance matrices," *J. Multivariate Analysis*, vol. 88, pp. 365 – 411, February 2004.

[12] J. P. Hoffbeck and D. A. Landgrebe, "Covariance matrix estimation and classification with limited training data," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 7, pp. 763–767, 1996.

[13] 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. 7, p. 076402, 2008.

[14] L. R. Bachega, G. Cao, and C. A. Bouman, "Fast signal analysis and decomposition on graphs using the sparse matrix transform," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 5426–5429, 2010.

[15] J. Theiler and D. R. Hush, "Statistics for characterizing data on the periphery," *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 4764–4767, 2010.

[16] J. Theiler, "Ellipsoid-simplex hybrid for hyperspectral anomaly detection," *Proc. IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2011.

[17] D. Snyder, J. Kerekes, I. Fairweather, R. Crabtree, J. Shive, and S. Hager, "Development of a web-based application to evaluate target finding algorithms," *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2, pp. 915–918, 2008.