IMPROVED HYPERSPECTRAL ANOMALY DETECTION IN HEAVY-TAILED BACKGROUNDS

Steven M. Adler-Golden

Spectral Sciences, Inc., 4 Fourth Avenue, Burlington, MA 01803-3304

ABSTRACT

A new metric for anomaly detection in hyperspectral imagery is developed to account for anisotropic heavy tails in covariance-whitened data. The anisotropy, consisting of a variation in tail heaviness with principal component number, commonly occurs when the number of linearly independent components representing the data to within the noise level is less than the number of data dimensions. The detection metric is generated by representing the probability density function of the data with an empirical anisotropic super-Gaussian model for the probability density function. Its performance exceeds that of the RX and Subspace RX methods in examples from CAP ARCHER and HyMap imagery.

Index Terms— Hyperspectral, anomaly, detection, RX

1. INTRODUCTION

Anomaly detection seeks to detect objects of interest that are distinguished simply by being different from the rest of the dataset. Here little if anything is known *a priori* about the probability density function (PDF) for anomalies, denoted f_a, whereas the PDF for the non-anomalous (or "background") data, denoted f_b, can be reasonably well characterized. In hyperspectral imagery (HSI) the multidimensional background PDF is generally derived from second-order spectral statistics gathered from the full image or from a smaller region that may exclude the pixel under evaluation, as in "real-time" and other spatially adaptive methods [1]. The detection metric is derived from the form of the PDF. The best-known anomaly detection method for HSI, the RX algorithm [2], relies on the Mahalanobis distance (MD) metric. MD is the amplitude of the data vector after mean subtraction and application of a linear whitening transform that converts the dataset to one having unit diagonal covariance:

$$\mathrm{MD} = \left(\Sigma_i \, x'^2_i\right)^{1/2} \qquad \textbf{(1)}$$

Here x'_i is the value of the whitened data in the ith coordinate, which is a principal component (PC) eigenvector of the dataset. Approximating the whitened PDF as a multidimensional Gaussian, it is monotonic in MD. Upon further assuming that the anomaly PDF is a constant (since its variation is unknown), thresholding MD amounts to thresholding the likelihood ratio for detection,

$$f_a(x')/f_b(x') \sim 1/f_b(x') \qquad \textbf{(2)}$$

More sophisticated treatments recognize the limitations of the Gaussian form of f_b [3]. Actual whitened data PDFs typically have heavy (or "long" or "fat") tails, in which extreme data values are more numerous than what a Gaussian distribution predicts. One model for this behavior generalizes the multivariate Gaussian distribution to an elliptically contoured (EC) distribution [4]. However, this provides no benefit for anomaly detection, because the EC $f_b(x')$ expression remains monotonic in MD, so the only difference is in the threshold value of the likelihood ratio that corresponds to a particular false alarm rate [3].

In this paper a simple, readily constructed anomaly metric is presented that provides improved detection results when the tail-heaviness of the whitened data is highly anisotropic, i.e., dependent on the direction in PC coordinate space. This behavior commonly occurs when the number of linearly independent components needed to represent the data to within the noise level is less than the number of data dimensions (i.e., spectral channels). The leading (low-numbered) PCs capture the largest variations in the dataset; they tend to show non-Gaussian statistics, characterized by heavy tails. In contrast, the trailing (high-numbered) PCs tend to be noise-dominated and observe Gaussian statistics, with much shorter tails. For example, this behavior is observed in a 400x400-pixel portion of a 52-channel (0.5 to 1.1 μm) hyperspectral image of suburban Indiana taken from a Civil Air Patrol ARCHER aircraft [5]. Fig. 1 (top) shows the data density along two PC components, one low-numbered (the 11th) and the other high-numbered (the 33rd). The distribution is close to symmetric along each PC, and, by definition, the two variances are identical. However, the half-width is narrower and the tail is heavier along the low-numbered PC. A proper anomaly detector, which flags data in the thin regions of the distribution, should bias detection away from such heavy-tailed PCs.

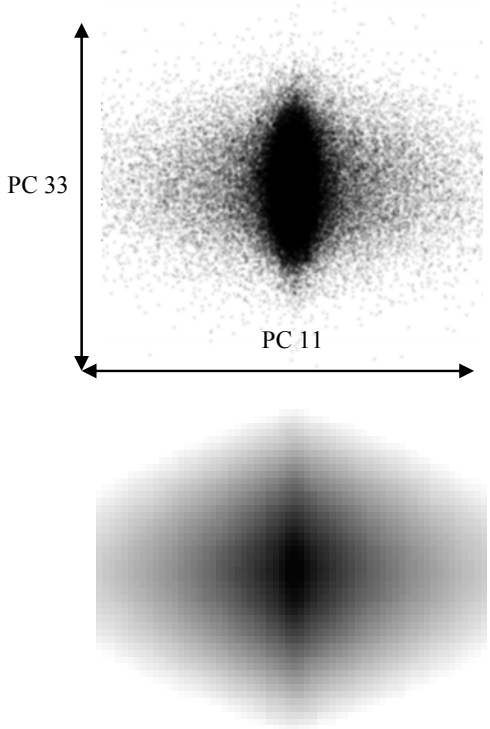

Fig. 1. Top, scatter plot for the whitened ARCHER 400x400-pixel sub-image along the 11th and 33rd PC coordinates (origin is at the center). Bottom, an AS approximation to the data distribution.

An ad hoc remedy to the heavy-tail problem is to remove a leading-PC subspace of the data, i.e., to simply omit some number of leading terms in the Eq. (1) sum [5]. While this method, called Subspace RX (SSRX), improves performance in many cases [6], it remains empirically based. It would be more rigorous and less arbitrary to retain the likelihood ratio test while employing a non-Gaussian, non-EC description of the background PDF. Some methods for characterizing such PDFs have been developed that utilize high-order statistics calculations [7] or Monte Carlo sampling [3], which are computationally intensive. Here we present a method based on an anisotropic super-Gaussian (AS) model for the background PDF, defined to within a normalization constant by

$$f_b(x') \sim \exp\left[-\Sigma_i \, (a_i|x'_i|)^{P_i}\right] \qquad \textbf{(3)}$$

Whereas in the Gaussian PDF the logarithm of the data density along each PC coordinate is linear in the second power of the coordinate, here the power P varies with the PC number and is 2 or less. An AS representation of the ARCHER data scatter plot is shown in Fig. 1 (bottom).

Since $f_b(x')$ is monotonic in the exponent or some arbitrary power thereof, the AS anomaly detection metric is taken as

$$\mathrm{AS} = \left[\Sigma_i \, (a_i|\, x'_i|)^{P_i}\right]^n \qquad \textbf{(4)}$$

The free parameters a_i and P_i may be readily obtained from the whitened data via histogramming and linear regression. Referring to Eq. (1), MD represents the special case of Eq. (3) in which $P_i = 2$, $a_i = 1$ and $n = ½$. In the general case n may be chosen as some average inverse of the P_i, so that the AS output has a similar dynamic range to the MD output; however, its value has no effect on detection performance. A drawback of the Eq. (3) PDF model is that it fits an EC distribution only in the Gaussian special case. Therefore the AS metric may be poorer than MD for anomaly detection with some datasets. The real potential benefit of the AS metric is with datasets in which P_i varies strongly with i, such as the Fig. 1 data.

2. IMPLEMENTATION AND APPLICATION

Parameter Extraction

We denote H_i, a one-dimensional function of coordinate i, as the integral of Eq. (3) over all PC coordinates except the ith. Given a finite data set, each H_i corresponds to a histogram of data frequency versus PC coordinate value. Noting that Eq. (3) is separable into i-dependent factors, it is straightforward to show that

$$\ln(H_i) = b_i - a_i^{P_i}|x'_i|^{P_i} \qquad \textbf{(5)}$$

where the b_i are constants. Appropriate a and P parameters may be extracted directly from the data histograms. A simple procedure for each i is to set a trial P_i value, determine a_i from a linear regression fit of $\ln(H_i)$ using $|x'_i|^{P_i}$ as the independent variable, and repeat the process with various P_i values to determine a best-fitting pair of a_i and P_i. We have used this procedure with trial P values ranging from 0.1 to 2.0 in increments of 0.1. The histogram bins used in the fit are restricted to those containing some minimum number of data points in order to insure adequate statistics, although this may introduce a very small bias towards heavier tails (lower P values).

Typical P values obtained from the ARCHER data with this procedure are shown in Fig. 2 (top). There is a gradual but unmistakable increase in P with PC number, and the highest values, around 1.5 or greater, occur where the eigenvalues level off, a further indication of noise-limited behavior. We also analyzed a 400x400-pixel region of a HyMap hyperspectral data strip taken at Davis, CA, which

is of an agricultural area containing some manmade objects. These data have both a higher signal-to-noise and a more extended spectral range (0.5-2.4 μm) than the ARCHER data. A further increase in signal-to-noise was obtained by re-binning the data from 120 to 60 spectral bands. As shown in Fig. 2 (bottom), here there is much more scatter in the *P* values, and the increase, if any, is much more gradual. The values for the high-numbered PCs do not exceed 1.5 and average around 1.1, indicating that these PCs retain considerable surface clutter information.

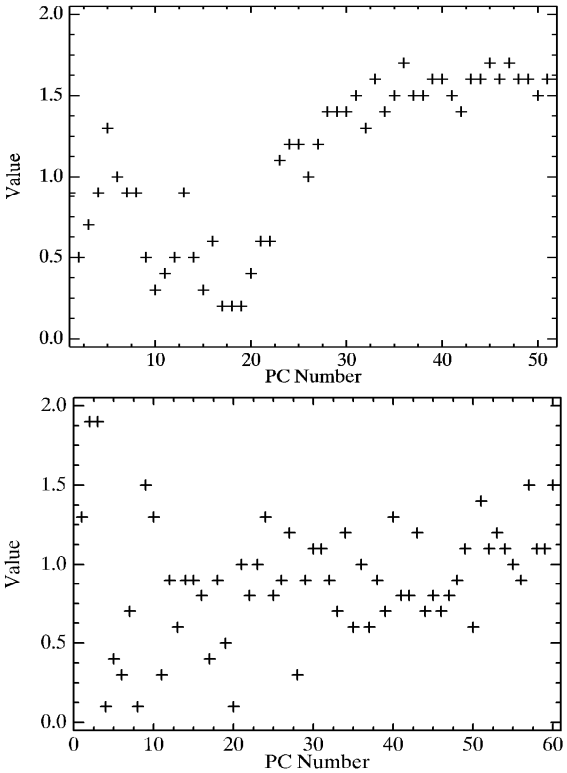

Fig. 2. Values of the PC coordinate power *P* for the ARCHER (top) and HyMap (bottom) image sub-regions.

Performance Characterization

For an initial comparison of the MD (RX) and AS anomaly metrics, we applied them to the detection of three green tarps in the ARCHER sub-image. Receiver-operator characteristic (ROC) curves for detecting the 28 fully-filled tarp pixels are shown in Fig. 3. The new AS metric provides a dramatic improvement in detection performance over the MD (RX) metric, amounting to one to two orders of magnitude reduction in false detections. Also shown in Fig. 3 are the MD and AS metric results with the first four PC terms excluded; the former represents the optimal SSRX method. Interestingly, removal of these terms improves performance with the AS metric as well as the MD metric.

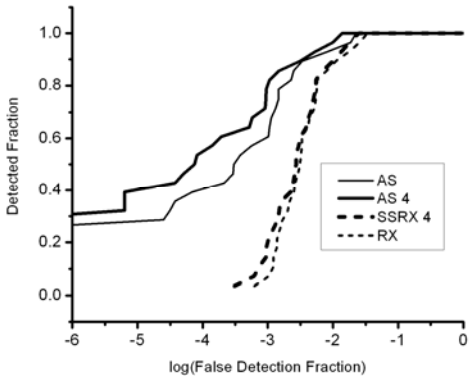

Fig. 3. ROC curves for detection of a tarp target in the ARCHER image; the first four PC terms are excluded in the heavy curves.

We generated additional test cases via a synthetic method [3] that uses data from a large image containing different terrain types or material classes in different regions. The background data are taken as a region dominated by a particular terrain type or subset of material classes. The target data are a spectrally diverse collection of pixels, preferably of a contrasting surface type, taken from outside the background region. The target pixels are "embedded" in the background region and the anomaly detection algorithm exercised to detect the entire target collection. We used this approach to evaluate the anomaly metrics on multiple urban targets from the ARCHER and HyMap scenes. The ARCHER targets consisted of 12 pixels from residential areas outside the 400x400-pixel sub-image. The HyMap targets consisted of 18 pixels from the University of California campus, located several hundred image lines away from the agricultural background area. In both datasets the target pixels were mainly from roofs, roads, parking lot pavement, and vehicles. For simplicity the targets were excluded from the background PDF statistics; this corresponds to detecting a rare target in a large scene, or to a spatially adaptive method of gathering the background statistics. Here the target embedding is virtual, as the ROC curves are constructed directly from the values of the anomaly metric for the target and background pixels.

Results for the ARCHER urban targets are shown in Fig. 4. The AS advantage over the RX metric is smaller here than with the tarp targets in Fig. 3, but still amounts to an order of magnitude or more reduction in false detections. Excluding leading PC terms leads to slightly better performance. Results for the HyMap urban targets are shown in Fig. 5. Again, the AS method performs better than the RX method, although by a much smaller margin than with the ARCHER scene (up to a twofold reduction in false detections). Here exclusion of leading PC terms slightly worsens the performance of both the RX and AS metrics.

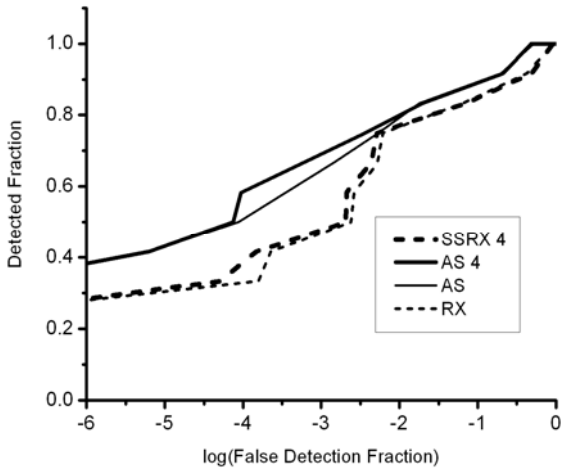

Fig. 4. ROC curves for detecting twelve urban pixels "embedded" in the Fig. 1 background image using various methods.

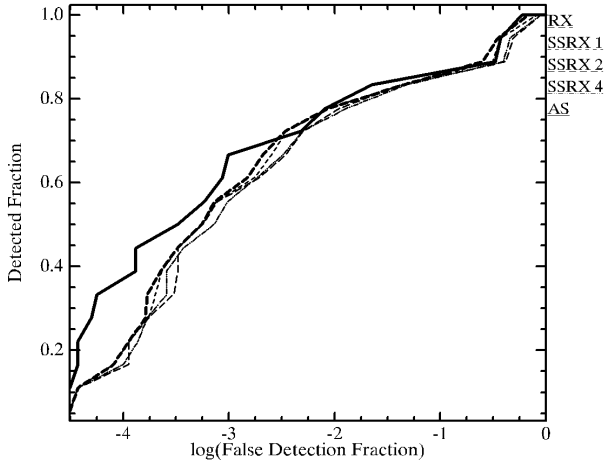

Fig. 5. ROC curves for detecting eighteen urban pixels "embedded" in the HyMap image of Davis, CA using various methods. Heavy solid line = AS, heavy dashed line = RX (MD), thin dashed lines = SSRX with 1, 2 or 4 PC's omitted.

3. CONCLUSIONS

Detection of anomalies in hyperspectral and other multidimensional datasets via the likelihood ratio test relies on the ability to model the background probability density function (PDF). Using an anisotropic super-Gaussian (AS) model for the PDF, a new anomaly detection metric that is monotonic in this PDF is found to outperform both the RX and subspace RX (SSRX) [5] methods in our test cases. There is undoubtedly room for further improvement. For simplicity, the present work was conducted using large-area PDF statistics. Better performance may result from locally adaptive schemes for defining the background [8], as well as by including a target PDF, derived from general knowledge of target spectral properties, in the likelihood ratio for detection [6]. In addition, it would be very desirable to develop an improved, but still easily calculated, PDF model that can describe EC as well as non-EC distributions.

Acknowledgement: The author thanks João M. Romano (US Army ARDEC, Picatinny Arsenal, NJ) for assistance with this paper.

4. REFERENCES

[1] C.-I. Chang, and S.-S. Chiang, "Anomaly Detection and Classification for Hyperspectral Imagery," IEEE Trans. Geosci. Remote Sensing, 40, 1314-1325 (2002).

[2] 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, 38, pp. 1760-1770, (1990).

[3] M. Bernhardt, J. Heather and O. Watkins, "Hyperspectral Anomaly Detection and Un-mixing in Fat-Tailed Clutter", 3rd EMRS DTC Technical Conference, Edinburgh (2006).

[4] D. Marden, and D. Manolakis, "Algorithms for hyperspectral imaging data exploitation using non-Gaussian elliptically contoured distributions", Imaging Spectroscopy VIII, S.S. Shen, ed., vol. 4816, SPIE, Seattle, WA, (2002).

[5] W. Kendall, "The Civil Air Patrol's ARCHER Hyperspectral Detection System," Proc. Specialty Group on Camouflage, Concealment and Deception, Military Sensing Symposium, (2005).

[6] A.P. Schaum, "Hyperspectral anomaly detection beyond RX," Proc. SPIE 6565, Algorithms and Technologies for Multispectral, Hyperspectral and Ultraspectral Imagery XIII, Paper No. 656502 (2007).

[7] H. Ren, Q. Du, J. Wang, C.-I. Chang, J.O. Jensen, and J.L. Jensen, "Automatic Target Recognition for Hyperspectral Imagery Using High-Order Statistics," IEEE Trans. Aerospace Electronic Systems, v. 42, No. 4, (October 2006).

[8] S.R. Soofbaf, H. Fahimnejad, M.J. Valadan Zoej, and B. Mojaradi, "Anomaly Detection Methods for Hyperspectral Imagery," Map World Forum, Hyderabad, India, (January 22-25, 2007).