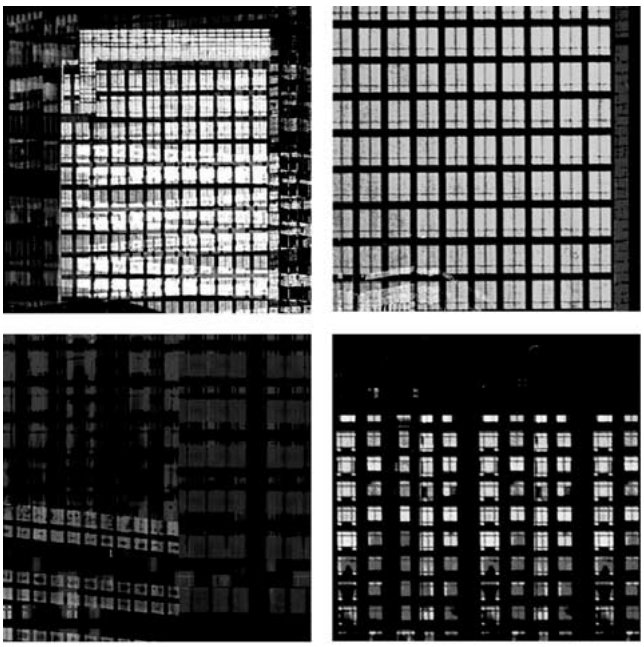
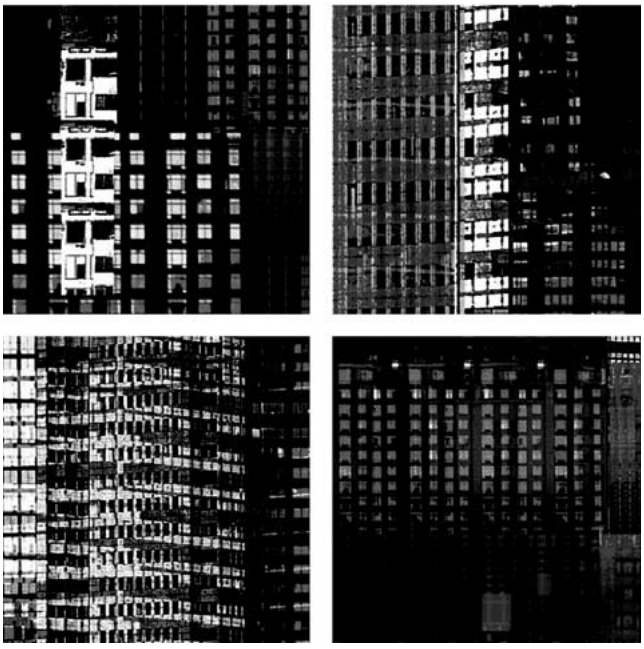
Anomaly Detection from Hyperspectral Imagery



David W. J. Stein, Scott G. Beaven, Lawrence E. Hoff, Edwin M. Winter, Alan P. Schaum, and Alan D. Stocker



Hyperspectral sensors are a new class of optical sensor that collect a spectrum from each point in a scene. They differ from multispectral sensors in that the number of bands is much higher (20 or more) and the spectral bands are contiguous. For remote sensing applications, they are typically deployed on either aircraft or satellites. The data product from a hyperspectral sensor is a three-dimensional array or "cube" of data with the width and length of the array corresponding to spatial dimensions and the spectrum of each point as the third dimension. While this data cube is a convenient way to view the product, at most two of the dimensions can be acquired simultaneously. In the most common configuration, the spectral and one spatial dimension are acquired simultaneously to give a high quality spectral signature for each point in the scene. AVIRIS [1] and many other sensors directed toward terrain classification are flying spot sensors that acquire an image by raster scanning the scene in width while the platform moves to build up the second spatial dimension. More recently, the spatial resolution of hyperspectral sensors has been improved by using a two-dimensional focal plane array that simultaneously acquires a spectrum along a line of points in the scene. The second dimension is then built up by motion of the platform. The sensors TRWIS-III [2], DARKHORSE [3], SEBASS [4], and AHI [5] use this approach. Hyperspectral sensors have a wide range of remote sensing applications including: terrain classification, environmental monitoring, agricultural monitoring, geological exploration, and surveillance.

With the introduction of sensors capable of high spatial and spectral resolution, there has been an increasing interest in using spectral imagery to detect small objects or features of interest. However, detection algorithms that presume a target signature are subject to signal mismatch losses because of the complications of converting sensor data into material spectra. The sensor collects spectral radiance, i.e., radiant flux per unit area per unit solid angle per unit wavelength [6]. Spectral signatures in the visible

to near infrared (VNIR), defined by wavelengths of approximately 0.4 to 1 μm, and short wave infrared (SWIR), defined by wavelengths of approximately 1 to 2.5 μm, are given in terms of reflectance, i.e., the ratio of reflected radiance to incident irradiance [6]. In the long wave infrared (LWIR), covering wavelengths ranging from approximately 7 to 15 μm, the material spectrum is measured in terms of emissivity, the ratio of the emission from the material to that of a blackbody at the same temperature [6]. To apply known-signal detection algorithms, the reflectance or emissivity spectra of the objects of interest must be converted into radiance at the sensor, or the radiance data collected by the sensor must be converted into reflectance or emissivity in a process known as atmospheric calibration [6]. If calibration panels are not available in the scene, then atmospheric compensation of hyperspectral imagery is based on models of illumination and atmospheric scattering and absorption as a function of wavelength, and for LWIR, temperature estimation. Atmospheric transmittance and illumination depend on many factors such as vertical temperature profile, water vapor concentration, concentration of mixed gasses, the concentration and types of aerosols, solar angle, cloud cover, shadowing, and viewing geometry [6]-[9]. Errors in estimates of environmental and sensor parameters may lead to significant errors in the estimate of reflectance spectra [9], which in turn leads to signal processing mismatch losses. Local variations in illumination and temperature further complicate the conversion of radiance to reflectivity or emissivity, respectively [6].

As an alternative to converting sensor data into reflectivity or emissivity, matched target detection algorithms may be implemented by converting spectral signatures into radiance if the prevailing atmospheric, illumination, and sensor parameters are known. Furthermore, if the parameters are known within a range of values, then subspace target models may be developed such that for a large set of conditions the target in radiance at the output of the sensor is well represented by the sub- space, i.e., the angle between the target and the sub- space is small [10]. Algorithms have been developed to detect targets contained within a subspace [10]-[12]. If the target subspaces are designed to be invariant to likely variation in significant parameters, then atmospheric compensation is not required for target detection purposes. However, detection performance generally diminishes as the dimension of the signature subspace increases if SNR is constant.

Target matching approaches are further complicated by the large number of possible objects of interest and uncertainty as to the reflectance/emission spectra of these objects. For example, the surface of an object of interest may consist of several materials, and the spectra may be affected by weathering or other unknown factors. One may be interested in a large number of possible objects each with several signatures. Thus, the multiplicity of possible spectra associated with the objects of interest and

Hyperspectral sensors differ from multispectral sensors in that the number of bands is much higher and the spectral bands are contiguous.

the complications of atmospheric compensation have lead to the development and application of anomaly detectors that seek to distinguish observations of unusual materials from typical background materials without reference to target signatures or target subspaces. In addition, change detectors are used to identify changes within a scene that occur over time without presupposing target signatures.

Anomalies are defined with reference to a model of the background. Background models are developed adaptively using reference data from either a local neighborhood of the test pixel or a large section of the image. Local and global spectral anomalies are defined as observations that deviate in some way from the neighboring clutter background or the image-wide clutter background, respectively. Both approaches have their merits. The local spectral anomaly detector is susceptible to false alarms that are isolated spectral anomalies. For example, consider a scene containing isolated trees on a grass plain. Each separate tree may be detected as a local spectral anomaly even if the image contains a separate region with many pixels of trees. The global spectral anomaly detection algorithms are not susceptible to this type of clutter generated false alarm. However, the global anomaly detector will not find an isolated target in the open if the signature is similar to that of previously classified background material.

In this article we develop anomaly detectors, i.e., detectors that do not presuppose a signature model of one or more dimensions, for three clutter models: the local normal model, the global normal mixture model, and the global linear mixture model. The local normal model treats the neighborhood of a pixel as having a normal probability distribution. The normal mixture model considers the observation from each pixel as arising from one of several possible classes such that each class has a normal probability distribution. The linear mixture model considers each observation to be a linear combination of fixed spectra, known as endmembers, that are, or may be, associated with materials in the scene, and the coefficients, interpreted as fractional abundance, are constrained to be nonnegative and sum to one. We show how the generalized likelihood ratio test (GLRT) may be used to derive anomaly detectors for the local normal and global normal mixture models. The anomaly detector applied with the linear mixture approach proceeds by identifying target like endmembers based on properties of the histogram of the abundance estimates and employing a matched filter

in the space of abundance estimates. To overcome the limitations of the individual models, we develop a joint decision logic, based on a maximum entropy probability model and the GLRT, that utilizes multiple decision statistics, and we apply this approach using the detection statistics derived from the three clutter models. Examples demonstrate that the joint decision logic can improve detection performance in comparison with the individual anomaly detectors. We also describe the application of linear prediction filters to repeated images of the same area to detect changes that occur within the scene over time.

Anomaly Detection Algorithms and the Generalized Likelihood Ratio Test

The GLRT is the foundation for the anomaly detection algorithms applied to the local normal, normal mixture, and joint exploitation models. We describe the GLRT for unknown signals and, more generally, for signals that belong to a known subspace. We give asymptotic distributions, as the number of reference samples approaches infinity, of the statistics assuming Gaussian noise, and we use these distributions to demonstrate that the performance of the anomaly detectors increases with SNR and diminishes with increasing dimension of the signal subspace. These results are used to compare the performance of two methods of implementing anomaly detectors: a bank of low-dimensional subspace detectors or a single high-dimensional subspace detector. The GLRT and the probability density functions (pdfs) that account for limited sample sizes are developed in more detail in [12].

Thresholding the likelihood ratio provides the hypothesis test that satisfies various optimality criteria including: maximum probability of detection for a given probability of false alarm, minimum expected cost, and minimization of maximal expected cost [13]. The pdf of the data is often assumed to have a parametric form, and parameters, θ, of the distribution are estimated from reference data. Alternatively, the GLRT approach estimates parameters from test and reference data. One postulates a reference data set $\{\nu_j \in C^K | 1 \le j \le N\}$ consisting of N (we allow $N = 0$, which implies that the reference data set is empty) independent identically distributed (iid) samples of dimension K having pdf $p_0(\cdot, \theta_0)$ and a set of test vectors, $\{x_\ell \in C^K | 1 \le \ell \le M\}$, that is to be classified as arising from either pdf $p_1(\cdot, \theta_1)$ (H_1) or $p_0(\cdot, \theta_0)$ (H_0). The GLRT is

$$G(x) = \frac{\max_{\theta_1}(p_1(\{x_\ell\}, \theta_1) p_0(\{\nu_j | 1 \le j \le N\}, \theta_1))}{\max_{\theta_0}(p_0(\{x_\ell\}, \theta_0) p_0(\{\nu_j | 1 \le j \le N\}, \theta_0))} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \quad (1)$$

Kelly [14] and Reed and Yu [15] developed GLRTs for multidimensional image data assuming that the spectrum of the received signal and the covariance of the background are unknown. We denote the normal distribution of mean μ and covariance Γ by $N(\mu, \Gamma)$; $x \overset{d}{=} p$ symbolizes that the random vector x has pdf p, and $x_N \overset{d}{\to} p$ indicates that the pdf of x_N converges to p in measure as $N \to \infty$. The data are modeled as

$$\begin{aligned} H_0&: \; x \overset{d}{=} N(\mu, \Gamma_x) \\ H_1&: \; x \overset{d}{=} N(s, \Gamma_x), \end{aligned} \quad (2)$$

such that s and Γ_x are unknown. We assume a single pixel target; the multipixel target version is a weighted sum of the single-pixel test described here [15]. Let x be the observation vector under test, and let $\{\nu_j \in C^K | 1 \le j \le N\}$ be a set of reference data. The GLRT is then

$$RX(x) = (x-\mu)^T \left(\frac{N}{N+1} \hat{\Gamma}_x + \frac{1}{N+1} (x-\mu)(x-\mu)^T \right)^{-1} \times (x-\mu) \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \quad (3)$$

where $\hat{\Gamma}_x = \frac{1}{N} \sum_{j=1}^{N} (\nu_j - \mu)(\nu_j - \mu)^T$ is the sample covariance matrix of the reference data and $(.)^T$ denotes the conjugate transpose for complex valued data that, when applied to real data, reduces to the ordinary transpose. As $N \to \infty$, RX converges to

$$RX_L(x) = (x-\mu)^T \left(\hat{\Gamma}_x \right)^{-1} (x-\mu) \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \quad (4)$$

RX is a single pixel form of the Reed Xiaoli algorithm that is often approximated by (4). It is monotonically related to minus the log of the probability of the observation x under the null hypothesis. It arises in many other contexts including [16]-[18], and it may be interpreted as a multivariate energy detector.

Asymptotic forms, as $N \to \infty$, of the probability distributions of *RX*, under the H_0 and H_1 hypotheses are given in terms of χ^2 and noncentral χ^2 densities, respectively [19]. Let $\chi^2_n(\cdot)$ denote the χ^2 density on n degrees of freedom (dof), and let $\chi^2_n(\cdot, \lambda)$ denote the noncentral χ^2 density on n dof having noncentrality parameter λ. Let $\delta = 1, 2$ if the data are real or complex valued, respectively. As $N \to \infty$,

$$H_0 : RX \overset{d}{\to} \chi^2_{\delta K}(\cdot) \quad (5)$$

and

$$H_1 : \; RX \overset{d}{\to} \chi^2_{\delta K}(\cdot; (s-\mu)^T \Gamma_x^{-1} (s-\mu)). \quad (6)$$

Note that the distribution of the test statistic under H_0 is independent of the unknown parameters, and thus the test statistic has the constant false alarm rate (CFAR) property.

The GLRT has been extended to situations in which the target signature is unknown and assumed to be in a subspace [12]. Scharf and Friedlander [20] develop the GLRT for the following model:

$$H_0: x = W\phi + n$$
$$H_1: x = V\theta + W\phi + n \tag{7}$$

where W is a matrix such that the columns of W span an interference subspace, V is a matrix such that the columns of V span a signal subspace, and n is additive noise such that $n \overset{d}{=} N(0, \sigma^2\Gamma)$. W, V, and Γ are assumed known, $D = \dim(V)$, θ and ϕ are assumed unknown, and σ^2 may be known or unknown. General procedures have not been developed for simultaneously estimating W and Γ. If either 1) ϕ is locally constant or 2) the data may be segmented into regions such that ϕ is essentially constant on each region, the term $W\phi$ may be absorbed into the noise which is then modeled by $n \overset{d}{=} N(\mu, \Gamma)$, where the parameters μ and Γ are estimated locally or for each segment. With $W = 0$, Γ may be estimated from background reference data, and if $\Gamma = I_{K\times K}$, a basis for W may be estimated as the eigenvectors of a background data correlation matrix having eigenvalues greater than σ^2.

The GLRT has also been derived for subspace signal models such that the background covariance matrix is unknown. Yu et al. derive the GLRT for multiband data in which the target signature is confined to a subset of the bands [21]. Kraut et al. consider the more general case of a subspace signature model [22]. They derive the GLRT assuming

$$H_0: \ x = n_0$$
$$H_1: \ x = V\theta + n_1 \tag{8}$$

where V is as above, $n_0 \sim N(0, \Gamma)$, $n_1 \sim N(0, \sigma^2\Gamma)$, Γ is unknown, and σ^2 may be either known or unknown. If σ^2 is known, then without loss of generality, one may assume that $\sigma^2 = 1$. The case of unknown σ^2 results in a GLRT such that the probability distribution under the null hypothesis is invariant to scalar variation of the test data. The detection statistic for the case rank $(V) > 1$ and $\sigma^2 = 1$ is described. As above, let $\hat{\Gamma}$ be the sample covariance of the reference data. $\hat{\Gamma}$ is positive definite with probability 1 if $N > K$ and Hermitian, and therefore, it has Cholesky factorization $\hat{\Gamma} = H^T H$. Define $\hat{\Gamma}^{-1/2} = H^{-1}$, $\hat{\Phi} = \hat{\Gamma}^{-1/2}(V)$, and $P_{\hat{\Phi}} = \hat{\Phi}(\hat{\Phi}^T\hat{\Phi})^{-1}\hat{\Phi}^T$, i.e., $P_{\hat{\Phi}}$ is orthogonal projection (in the whitened space) onto $\widetilde{\Phi}$, the subspace spanned by the columns of $\hat{\Phi}$. The GLRT for this case is [22]

$$\mathrm{KSM}(x) = \frac{\left(\hat{\Gamma}^{-1/2}x\right)^T P_{\hat{\Phi}}\left(\hat{\Gamma}^{-1/2}x\right)}{1 + \frac{1}{N}\left(x^T\hat{\Gamma}^{-1}x\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{9}$$

As $N \to \infty$,

$$H_0: \ \mathrm{KSM} \overset{d}{\to} \chi^2_{\delta D}(\cdot) \tag{10}$$

and

$$H_1: \ \mathrm{KSM} \overset{d}{\to} \chi'^2_{\delta D}(\cdot; (V\theta)^T\Gamma^{-1}(V\theta)). \tag{11}$$

The distributions of these GLRT for finite N are derived in [22]. KSM is a generalization of RX, and computationally the difference lies in the projection operator in (9).

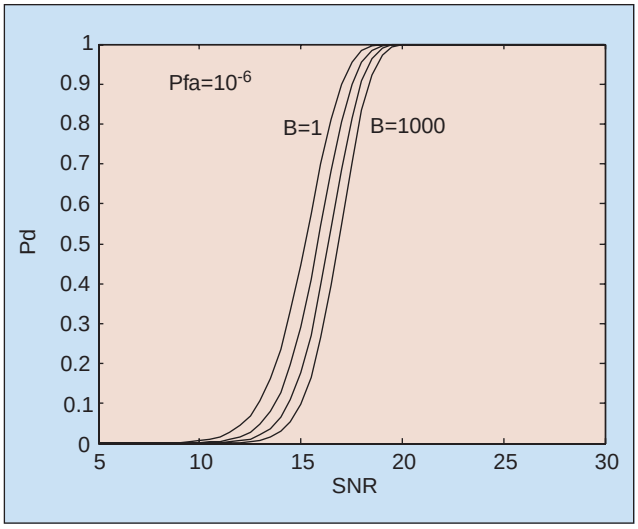
We compare the predicted performance of two methods of implementing anomaly detection algorithms. For each possible target signature of interest, one might develop a subspace that encapsulates the uncertainty as to illumination and atmospheric transmission [10] and utilize a bank of subspace detectors. Alternatively, one might develop a single subspace that encapsulates all uncertainty as to the target model. Figs. 1 and 2 illustrate the performance of these approaches by comparing probability of detection (Pd) at a fixed probability of false alarm (Pfa). The Pd of a bank of filters at a given Pfa is easily calculated from the Pd at each Pfa of the individual filters and the assumptions: 1) the filters have independent output under the null hypothesis and 2) if a target corresponding to the target model of one of the filters is present, then that filter will have maximal output. Further analysis is required if the filters do not have independent output under the null hypothesis. Fig. 1 shows the asymptotic performance of a bank of subspace detectors for which the dimension of the signal subspace, D, is 6 as a function of SNR and the number of subspaces, B. Fig. 2 shows the asymptotic performance of RX as a function of SNR and K, the dimension of the observation space. Fig. 2 also illustrates the performance of KSM as a function of SNR and D, the dimension of the signal subspace, by letting $K = D$ in the figure. Fig. 2 illustrates the degradation, at a fixed SNR, in predicted performance as the dimension of the observation space or the signal subspace increases. Comparing Figs. 1 and 2 shows that a bank of 1000 subspace detectors, such that each subspace has dimension 6, has better predicted performance than a 100 dimensional RX or KSM detector, assuming, as noted above, independent filters. In general, the filters will not be independent under H_0 unless the subspaces satisfy appropriate orthogonality conditions. For example, the bank of 1000 filters, each using a subspace of dimension 6, requires an SNR of 17.5 to achieve a predicted Pd of 0.5, while the RX detector applied to an observation space of dimension 100 or the KSM detector applied to a signal subspace of dimension 100 requires an SNR of 19 dB to have a predicted Pd of 0.5. However, the bank of subspace detectors has a much greater computational load as the set of target subspaces must be accurately determined, and the quadratic form is computed at each pixel for each subspace detector while the RX and KSM algorithms are implemented with a single quadratic form.

Furthermore, the bank of subspace detectors may suffer from mismatch losses that will degrade performance.

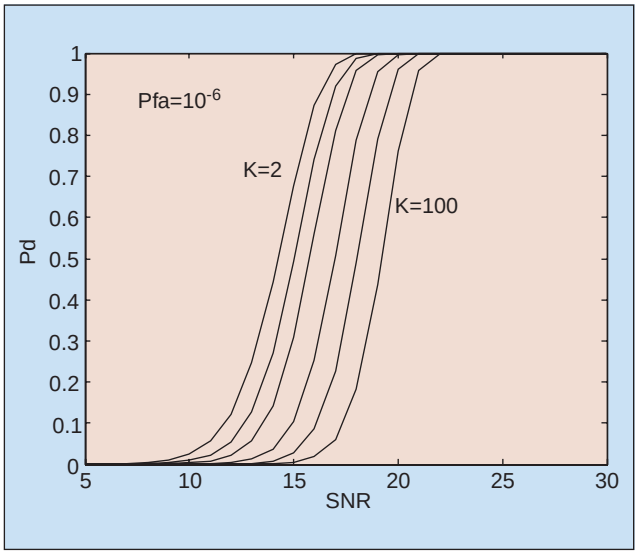
In section four anomaly detectors with non-Gaussian pdfs are utilized. Assume the following model:

$$H_0: \; x \sim p_0(\cdot;\theta_0)$$
$$H_1: \; (x-s) \sim p_0(\cdot;\theta_0), \tag{12}$$

where θ_0 is a set of parameters of the pdf p_0. If θ_0 is known or estimated from a large sample of reference data and s is unknown, then the GLRT is well approximated by:



▲ *1. Limiting performance of KSM for D=6 and B=1,10,100,1000 where D is the dimension of the subspace, and B is the number of filters in the filter bank.*



▲ *2. Limiting performance of RX for K=2,5,10,25,50,100, where K is the dimension of the observation space, and also the limiting performance of the KSM algorithm if K is replaced by D, the dimension of the signal subspace. Note that the greater the uncertainty in the signal, as manifested in a signal subspace of larger dimension, the greater the SNR required for equivalent performance.*

$$G(x) \equiv -\log(p_0(x;\theta_0)) \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{13}$$

Note that (13) generalizes the RX algorithm (4).

Models of Hyperspectral Imagery and Anomaly Detection Algorithms

Anomaly detection algorithms are used to distinguish observations from the background when target models are not available or are unreliable. In this section we outline three models: the local Gaussian model, the global Gaussian mixture model, and the global linear mixture model. These models have been widely used in the community and provide complementary interpretations of the data. For each model we describe an associated anomaly detection algorithm. The anomaly detection algorithms for the local and global normal mixture models are applications of the GLRT.

Local Normal Model

The Gaussian model leads to the anomaly detectors (3), (4). If the mean and covariance matrix are estimated within a neighborhood of every test pixel, one obtains the RX algorithm [15]. The local Gaussian model may not be valid for hyperspectral data if relatively small regions contain multiple materials. The BHEP (Barringhaus, Henze, Epps, and Pulley) test [23] was used to evaluate the local normal model. The BHEP test statistic compares the empirical characteristic function of the data, transformed to zero mean and identity covariance, with the characteristic function of the normal distribution having mean zero and identity covariance matrix. The test is consistent, invariant under affine transformations, and applicable to any number of samples and data dimensions. The first three moments of the limiting distribution, as the number of samples approaches infinity, are known and can be used to approximate thresholds of the test statistic corresponding to prescribed probabilities of rejecting the null hypothesis when it is true. The test was applied to a hyperspectral scene to evaluate the goodness-of-fit of a local normal model. For these data, at a probability of false rejection of 10^{-5}, the local normal hypothesis was rejected for more than 90% of the pixels. Thus, the local normal model may not capture the complexity of hyperspectral imagery. The presence of multiple materials in close proximity suggests that global mixture distributions will provide more accurate descriptions.

Gaussian Mixture Model

The Gaussian mixture model is one method of characterizing spectral data observed from nonhomogenous, multicomponent scenes. This approach models each datum as an observation of a random vector having one of

several possible multivariate Gaussian distributions. The pdf of the scene is a Gaussian mixture distribution:

$$g(x) = \sum_{c=1}^{C} \pi_c N(x|\mu_c, \Gamma_c); \quad \pi_c \geq 0; \quad \sum_{c=1}^{C} \pi_c = 1, \tag{14}$$

where π_c is the probability of class c.

It is applied as a global model since the parameters are estimated over relatively large regions. In our work, the parameters of the Gaussian mixture model are estimated using the stochastic expectation maximization (SEM) approach [24], which is a variation on the expectation maximization (EM) algorithm [25]. For each iteration SEM randomly assigns observations to current classes, using the conditional distribution of the classes given the observations, and updates the class statistics as the class sample statistics. The iterations cease when a convergence criterion based on mean likelihood or the change in the parameter values is satisfied.

The Gaussian mixture model may also be used for scene classification. Pixel classification, i.e., segmentation, is achieved by assigning individual pixels to a class using a maximum a posteriori (MAP) classifier [26]. The MAP classifier identifies the most likely class for a given observation according to Bayes Law.

Detection algorithms may be derived from the Gaussian mixture clutter model. Anomaly detection may be achieved by applying the GLRT (13) to the model (14). Alternatively, data may be classified according to the MAP classifier, and a class conditional RX algorithm, (4), may be used to identify anomalies. This approach is denoted GMRX.

Linear Mixture Model

Hyperspectral imagery may also contain pixels that are composed of several materials, and the linear mixture model has been developed to analyze such data [27-32]. This approach models the observation x_i as

$$x_i = \sum_{m=1}^{M} a_{mi} s_m + \varepsilon; \quad a_{mi} \geq 0; \quad \sum_{m=1}^{M} a_{mi} = 1, \tag{15}$$

where $s_m \in R^K$ is the spectrum (reflective or emissive) of a pure pixel of material m, $a_{mi} \in R$ is the fractional abundance of material m at pixel i, and ε is additive noise. The observations are modeled as convex combinations of constituent spectra, s_m, known as endmembers. The constraints imposed on the abundance values represent the facts that a pixel cannot contain more than a full pixel nor a negative quantity of a substance, that materials replace each other rather than combine via superposition, and that the set of endmembers should include all material types present in the scene. The linear mixture model is a global model in that the set of endmembers is assumed to be valid over relatively large regions. Fig. 3 illustrates observations expressed as a convex combination of three endmembers. The endmember identified as a shade point consists of sensor noise and path radiance.

Several different procedures have been developed to automatically find the endmembers, s_m, from hyperspectral data. If M endmembers are to be found, the data are typically transformed to a subspace of dimension M-1. The endmembers have been obtained as the vertices of the minimal volume simplex, or an approximation thereof, that encloses a subset of the transformed data. The ORASIS algorithm, the first demonstrated real-time autonomous endmember selection method, uses this approach [29]. Constructing the minimal volume enclosing simplex is computationally intensive, however the algorithm does not require pure pixels in order to uniquely identify endmember spectra [30]. Alternatively, one may define the endmembers to be the vertices of the maximal volume simplex such that the vertices consist of observation vectors. This approach underlies the NFINDR algorithm [31]. For this method to produce endmembers, there must be at least one pure pixel of each material [30]. Alternatively, the iterative error analysis (IEA) approach finds the endmembers by performing a series of constrained unmixings and choosing the endmembers to be those observations that minimize the error in the unmixed image [32].

Given the endmembers, the abundance estimates are obtained as the solution of a constrained least squares problem. If the inequality constraints are utilized, the abundance estimates are the solution of a quadratic programming problem. To reduce the computational requirements, certain authors have neglected the inequality constraints, and the abundance coefficients are then found as the solution to a system of linear equations. The computational burden is then dominated by the computation of a matrix vector product at each pixel.

NFINDR has been coupled with a two-step procedure, the stochastic target detector (STD), for the autonomous detection of global spectral anomalies. This process first identifies certain of the endmembers as target-like and then applies a matched filter for each of the target-like endmembers in the abundance space. Targets are assumed to be rare, and furthermore, the abundance of a target is assumed to be near one on a pixel that contains the target and near zero on a pixel that does not. Therefore the histogram of the abundance values of a target endmember will have a high peak to RMS value, and only a small percentage of abundance values, approximately the percent of pixels containing the target, will be above a noise threshold. Using these criteria, the indices of the abundance vector $a_i = (a_{1i}, \ldots, a_{Mi})$ are segregated into sets $\mathrm{I}_x = \{\mathrm{j}_1, \ldots, j_s\}$ and $\mathrm{I}_y = \{\mathrm{k}_1, \ldots, k_t\}$ consisting of the indices of the clutter components and the target-like components, respectively. Let $x_i = (a_{j_1 i}, \ldots, a_{j_s i})$ and $y_i = (a_{k_1 i}, \ldots, a_{k_t i})$ be the clutter and target like components of the abundance vector a_i, respectively. For each target-like component, $1 \leq \tau \leq t$, the following detection problem presents itself, dropping the index i:

$$H_0:\ x = c_x + m_x \text{ and } y = c_y + m_y$$
$$H_1(\tau):\ x = c_x + m_x^{\tau} \text{ and } y = c_y + m_y^{\tau}, \tag{16}$$

where: m_x^{τ} and m_y^{τ} are the mean of x and y, respectively, when a target corresponding to component τ is present in a pixel; m_x and m_y are the mean of x and y under the null hypothesis; the additive components, c_x and c_y, are the fluctuations of the clutter-like and target-like components. Both the null and alternative hypotheses have a mean value added to a zero mean random fluctuation. Note that the mean in the alternative hypotheses replaces the clutter mean in the null hypotheses. These hypotheses are modeling the fact that the target obscures the background and does not add to it. The target and clutter are assumed to experience similar fluctuations in the scene. We assume that

$$m_x^{\tau} = (0,\ldots,0), \text{ and } m_y^{\tau} = (0,\ldots,0,\overbrace{1}^{\text{1 in component } \tau},0,\ldots,0).$$

Under these assumptions the matched filter for each target type may be written as [21], [33]:

$$\lambda_{\tau}(x,y) = (m_y^{\tau} - m_y)^T (\Gamma_{yy} - \Gamma_{yx}\Gamma_{xx}^{-1}\Gamma_{xy})^{-1} \times (y - \Gamma_{yx}\Gamma_{xx}^{-1}(x - m_x)) \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{17}$$

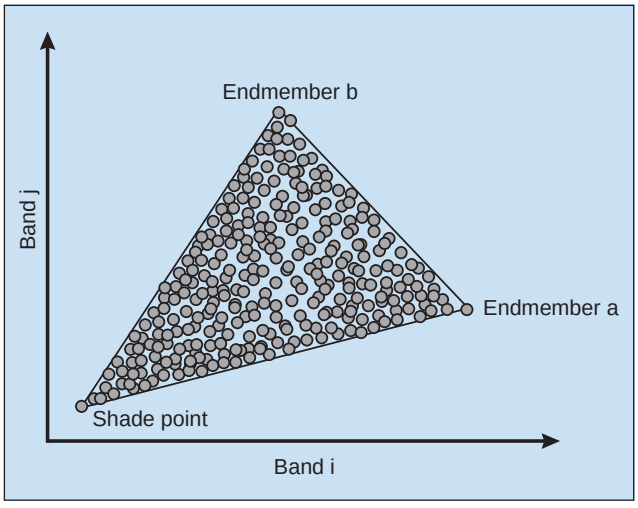
where

$$\Gamma = \begin{bmatrix} \Gamma_{xx} & \Gamma_{xy} \\ \Gamma_{yx} & \Gamma_{yy} \end{bmatrix}$$

and

$$\Gamma_{ab} = E\left((a - \mu_a)(b - \mu_b)^T\right).$$

The matched filter is the likelihood ratio assuming Gaussian distributed noise, and it is the linear filter having maximal SNR for distributions having second-order moments. The linear transform $H: x \to m_y + \Gamma_{yx}\Gamma_{xx}(x - m_x)$ is the linear prediction filter of y given x having minimal mean square error, and the covariance matrix of $y - Hx$ is $\Gamma_{yy} - \Gamma_{yx}\Gamma_{xx}^{-1}\Gamma_{yy}$. A bank of matched filters of the form (17) is used to detect the presence of any of the targets $1 \le \tau \le t$.



▲ *3. Data are depicted as a convex combination of a shade point and two other endmembers.*

Multiple Algorithm Fusion

Three complementary statistical models of hyperspectral data and associated detection algorithms that do not presume a target signature or subspace of signatures have been presented. Evaluation of the resulting detection algorithms on hyperspectral imagery provides examples of each algorithm outperforming the other two. Since many of the conditions and parameters that govern performance are unknown or poorly characterized in an operational setting, it can be very difficult to dynamically select the optimal detection algorithm. To obtain more consistent anomaly detection performance from a single algorithm, one might investigate more general statistical models [34], [35]. Alternatively, we construct a multiple-algorithm-decision or fusion criterion that combines multiple discrimination features. In a conventional fusion approach, simple logical operations (e.g., AND, OR) would be applied to the individual detector outputs at the post-threshold or object-declaration stage. These ad-hoc rules tend to produce highly variable performance results from one case to another (including occasional losses in sensitivity), and it is notoriously difficult to specify the "best" fusion logic in advance. In the approach described here, a modeled joint distribution of the detection statistics is constructed, and the multiple algorithm decision criterion is obtained by applying a threshold test to the resulting joint pdf (13). As described below, the joint distribution is obtained by modeling the marginal distribution of the detection statistics and their correlation.

Marginal Distributions

Gamma mixture distributions were used to model the output of the RX, GMRX, and STD algorithms. The gamma density with shape parameter V and scale parameter a is given by $p_{\gamma}(x|a,\nu) = (x/a)^{\nu-1}\exp(-x/a)(a\Gamma(\nu))^{-1}; x \ge 0$. An M component gamma mixture distribution is defined by

$$\psi(x;a,\nu,w) = \sum_{j=1}^{M} w_j p_{\gamma}(x|a_j,\nu_j); \quad w_j \ge 0, \text{and} \sum_{j=1}^{M} w_j = 1. \tag{18}$$

As described above, if the background has a normal pdf, then the output of a quadratic detector has a central or noncentral χ^2 distribution (5), (6). The noncentral χ^2 function may be expressed as a mixture of gamma distributions [36]. Furthermore, parameter estimation error distorts the output statistics of quadratic detectors from χ^2 to noncentral χ^2 and thus to gamma-mixture. The output of quadratic detectors acting on normal mixture

data may be modeled by a mixture of noncentral χ^2 distributions, which in turn may be modeled as a mixture of gamma distributions. Thus, the output of the GMRX and RX algorithms should be well modeled using gamma mixture pdfs, and this has been verified empirically. We have also found empirically that the positive valued side of the output of the NFINDR-STD algorithm is well modeled using gamma mixture distributions. Fig. 4 illustrates the fit of a three-term gamma mixture model complementary cumulative distribution function (CCDF) to the empirical CCDF of the output of the GMRX detector. For these data, the RX and STD output were similarly well described by a three-term gamma mixture model.

Modeling the Joint Distribution

Given a set of marginal distributions, we construct the maximum entropy joint distribution such that the marginals of the joint distribution are equal to the given marginals and such that the normal-score correlation of the joint distribution is equal to the empirical normal-score correlation [37]. Let $x=(x_1,\ldots,x_K)\in R^K$ be a random vector such that x_k has marginal distribution $x_k=\Psi_k$ with CDF Ψ_k. Let Φ be the zero mean unit variance normal CDF. Then

$$z_k=\Phi^{-1}\circ\Psi_k(x_k)\overset{d}{=}N(0,1), \tag{19}$$

and $C=\mathrm{cov}(z_1,\ldots,z_K)$ is the normal score correlation of $x=(x_1,\ldots,x_K)$. Define

$$T(x_1,\ldots,x_K)=(\Phi^{-1}(\Psi(x_1)),\cdots,\Phi^{-1}(\Psi(x_K))) \tag{20}$$

and let

$$p(z)=\frac{1}{(2\pi)^{k/2}\|C\|^{1/2}}\exp\left(-\frac{1}{2}z^TC^{-1}z\right). \tag{21}$$

Define

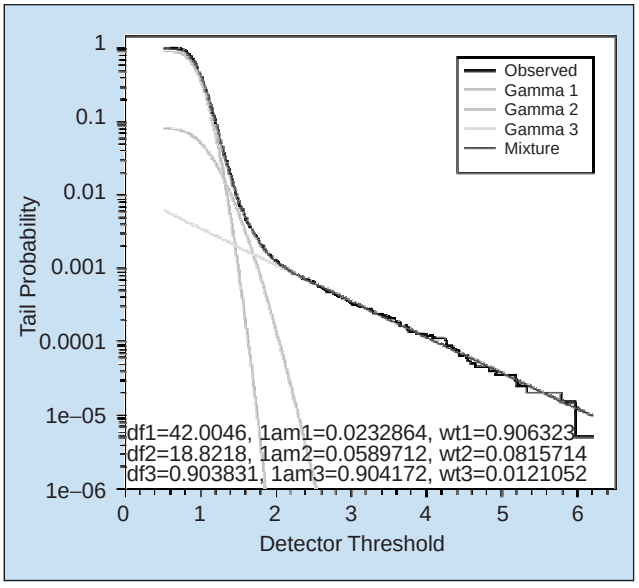
$$\begin{aligned} f(x)&=p(T(x))\|\nabla T(x)\| \\ &=\|C\|^{-1}\exp\left(-\frac{1}{2}T(x)^T(C-I)^{-1}T(x)\right)\prod_{k=1}^{K}\psi_k(x_k), \end{aligned} \tag{22}$$

where $\|M\|$ denotes the determinant of the matrix M and $\nabla T(x)$ is the Jacobian of T at x; f is the pull back of the normal pdf p via the transform T. Then f is the distribution having maximum entropy among distributions with marginal distributions ψ_k, and normal score correlation C [37]. We use this construction to model the joint distribution of the (RX,GMRX,STD) output using the gamma mixture marginals described above. Joint decision curves are defined as contours of the joint density function (13). This defines the multiple algorithm fusion (MAF) approach. Alternatively, one may transform the data so that the marginals have exponential densities, refit the marginals with gamma densities [38], and apply the above construction to obtain the joint density.

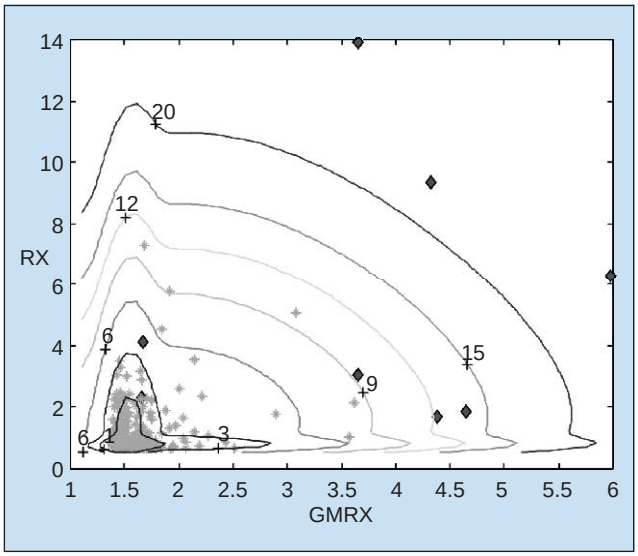
Fig. 5 illustrates the joint decision boundaries, i.e., the contours of the joint pdf obtained from the above construction applied to the RX and GMRX algorithms. The asterisks and diamonds are GMRX-RX values of background pixels and test objects, respectively. Evidently, the number of false alarms incurred in order to place all targets in the correct class is significantly diminished if the joint decision contours, rather than thresholds applied to the RX or GMRX values alone, are used to segregate the background region from the target region.

Detection Performance

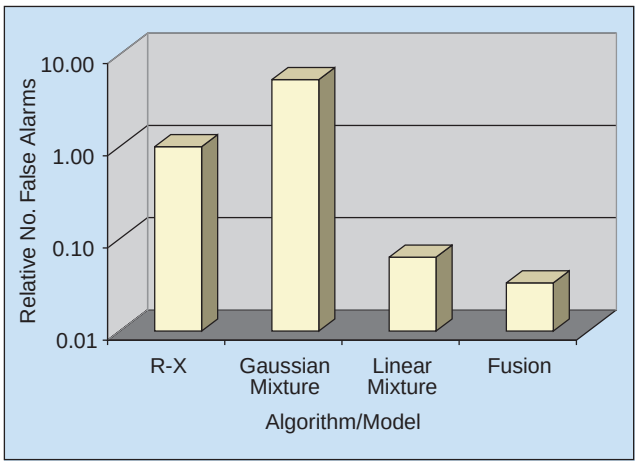
The algorithms were applied to a hyperspectral image in order to compare performance. The relative performance of RX, based on the local normal model, GMRX, based on the Gaussian mixture model, NFINDR-STD, based on the linear mixture model, and MAF is illustrated in Figs. 6 and 7. Fig. 6 shows, for a fixed Pd, the ratio of the number of false alarms incurred using each of the algorithms to the number incurred using the RX algorithm. For this example, GMRX produces more false alarms and NFINDR-STD produces fewer false alarms than RX, while MAF yields fewer false alarms then either of the individual algorithms. In Fig. 8, the algorithms are compared on the basis of relative probability of detection at a fixed number of false alarms using MAF as the standard. For this example, the algorithms are ranked in order of increasing number of targets detected as follows: RX, GMRX, NFINDR-STD, and MAF. These results demonstrate improved detection performance from jointly exploiting anomaly detection algorithms based on the three background clutter models.



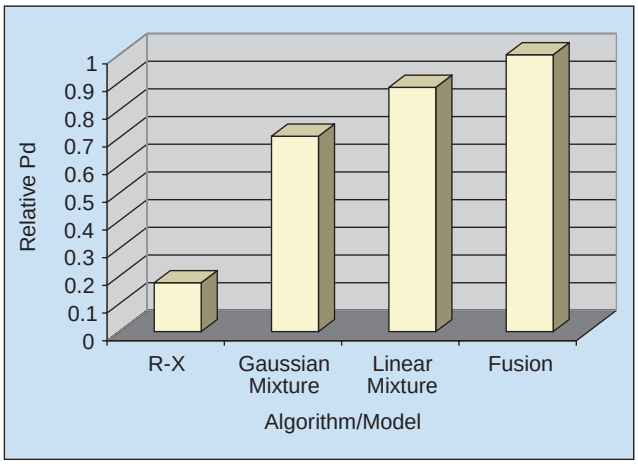
▲ *4. Comparison of the empirical CCDF of the GMRX output and the CCDF of the three-term gamma mixture fit to hyperspectral data.*



▲ *5. Contours of the maximum entropy joint distribution of the RX and GMRX statistics.*



▲ *6. False alarm performance of anomaly detectors based on different stochastic background models at a fixed probability of detection relative to the RX algorithm.*



▲ *7. Detection performance of anomaly detectors at a fixed probability of false alarm relative to fusion of multiple anomaly detectors .*

Hyperspectral Change Detection

If an object of interest is inserted into a scene, moves within a scene, or undergoes a change of its reflectivity or emissivity during a time interval, then change detection techniques may be used to ascertain this event. Geographically registered images from the beginning and end of this time interval are assumed to be available. Due to variations in atmospheric conditions, background illumination, temperature, and possibly sensor response, simple subtraction of observed radiances from two perfectly registered images typically produces a cluttered residual, in which it is impossible to distinguish natural changes from the removal or appearance of small objects. However, at each wavelength, the radiance measurements of the same material viewed under two sets of conditions are linearly related [6]. Jensen applied this relationship to multispectral Landsat imagery to identify changes in urban land use [39]. He separately estimates the linear regression coefficients for each wavelength. This approach is generalized in the Chronochrome algorithm [40]. It employs a multivariate linear regression filter to predict the radiance of an image of a scene from an earlier radiance image of the scene. Changes occurring at any pixel in the image are indicated by thresholding an RX detection statistic applied to the residual consisting of the difference between the image and the predicted image.

Spectrally dependent changes in path radiance, atmospheric attenuation, scene illumination, and sensor gain and offset that accumulate over time all modify sensed radiances according to an affine transformation [6]. Let x_i, be a hyperspectral vector collected at time i. One expects that $x_2 = O + Gx_1$, where the offset O models the change in path radiance and the gain G models the change in illumination, temperature, sensor gain, and atmospheric transmission from time t_1 to time t_2. All of these physical effects are consistent with the matrix G being diagonal. However, in practice nondiagonal matrices provide noticeably better models, which may be due to imperfect pixel-to-pixel registration or spatial variations in temperature (LWIR) or illumination (VNIR-SWIR).

A multivariate regression analysis of hyperspectral data collected at several times can often be used to estimate the model parameters O and G. This first step in the Chronochrome method produces a least mean squared estimate $\hat{x}_2$ of any pixel value at time 2 given by

$$\hat{x}_2 = E(x_2) + \Gamma_{21}\Gamma_{11}^{-1}(x_1 - E(x_1)) \tag{23}$$

with

$$\Gamma_{ij} = E\Big((x_i - E(x_i))(x_j - E(x_j))^T\Big),$$

where the parameters $E(x_1)$, $E(x_2)$, Γ_{11}, and Γ_{21} are estimated from the data. Equation (23) describes a regression hyperplane.

The second step in the Chronochrome algorithm consists of finding those pixels for which the estimated signal

is improbable. This is fundamentally the problem addressed by the "RX" algorithm for detecting multivariate anomalies, which prescribes that the magnitude of $\varepsilon = x_2 - \hat{x}_2$ be compared to a threshold after it is whitened with its own covariance structure. That is, one looks for large values of

$$C(\varepsilon) = \varepsilon^T \Gamma_\varepsilon^{-1} \varepsilon. \tag{24}$$

where

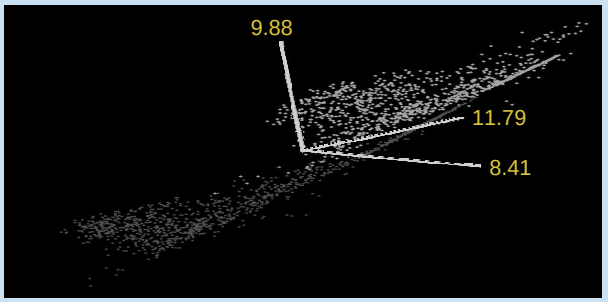
$$\Gamma_\varepsilon = E(\varepsilon\varepsilon^T) = \Gamma_2 - \Gamma_{21}\Gamma_1^{-1}\Gamma_{12}.$$

The transformation (23) fails to account for local variation in temperature or illumination affecting longwave and reflectance data, respectively, as the regression coefficients are computed globally. For example, in the 8-12 μm region, radiance is driven by temperature, which is unpredictable over periods of only a few hours. The emissivity of most background constituents are high, typically above 0.9, implying that there is a rather well defined single spectral direction corresponding to temperature. This direction is approximated as the eigenvector corresponding to the largest eigenvalue of the matrix Γ_ε, and it is suppressed by the RX statistic.

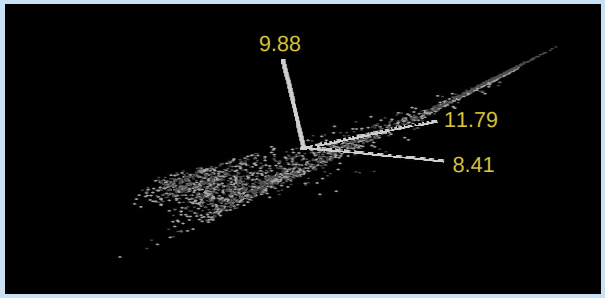
Fig. 8 illustrates the application of Chronochrome to data from the hyperspectral thermal sensor SEBASS [4]. The first plot shows projections of spectral signals collected on two consecutive days. Each scatter plot manifests the well-defined edges and triangular structures consistent with a linear mixture model. The second plot overlays predictions of the signals using (24) atop the actual data, and it shows large apparent errors. The final plot reveals that the error ε in fact tends to be concentrated in one direction (temperature). The off-diagonal outliers shown in the third plot all correspond to pixels on vehicles that had been present in only one scene. These anomalous signals are sufficiently distinct from those associated with temperature that RX applied to ε enables target detection at high SNR.

Summary and Conclusions

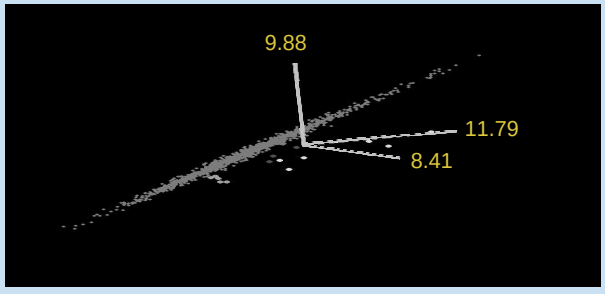
Anomaly detectors for hyperspectral data, i.e., detectors that do not require target signature models, have been developed based on fundamental detection theoretic principles, including the GLRT and approximations thereof. In this article, we have outlined the underlying theory for the application of anomaly detection to systems with inherently high dimensionality, and we have demonstrated that the performance improves with SNR and diminishes with increasing dimension. SNR depends on the characteristics of the targets, clutter, environment, and sensor. The differences between the anomaly detection approaches are tied to the underlying clutter models that are imposed. Here we have given an overview of three principle types of background clutter models: locally Gaussian, Gaussian-mixture, and linear mixture models. Additionally, a method for jointly exploiting multiple anomaly detectors through fusion of their outputs is described. Sample relative performance of the detectors and their joint performance support the use of multiple-model anomaly detection in the analysis of hyperspectral data. The Chronochrome algorithm extends the idea of anomaly detection to the detection of changes within a scene occurring over time without employing signature models. Future work on anomaly detection should include the development of dimensionality reduction transforms that preserve SNR for the targets of interest so that the improved performance provided by reduced dimensionality may be realized. Figs. 1 and 2 suggest that improved anomaly detection could result from using a bank of subspace detectors rather than a global anomaly detector, and methods of defining this collection of subspaces should be explored. Further improvements in anomaly detection may result from using improved clutter models as suggested by the work on stochastic mixture models that unifies the linear mixture and Gaussian mixture ap-



(a)



(b)



(c)

▲ *8. (a) Day 1 scatter (red), day 2 scatter (green). (b) Chronochrome prediction (green) of day 1 data (red). (c) Error scatter concentrated in temperature direction, with real changes displaced from diagonal.*

proaches [34], [35]. Furthermore, the requirements for image registration in the application of Chronochrome need to be established.

Acknowledgments

The authors would like to thank the Defense Advanced Research Projects Agency and the Office of Naval Research for supporting this work and the reviewers for their helpful suggestions.

David W.J. Stein received the B.A. in mathematics-economics from Reed College in 1979 and the Ph.D. in mathematics from Brandeis University in 1986. He is presently a Scientist at SPAWAR Systems Center, San Diego, where he has worked on modeling and signal processing issues for SONAR, RADAR, and spectral sensors. He has published over 30 journal and conference articles. He holds five patents in the area of signal processing.

Scott G. Beaven is a Corporate Senior Staff member at Space Computer Corporation, where he works on hyperspectral and multisensor signal processing. He has worked in the area of remote sensing since 1990. From 1990-1995 he was an ONR Doctoral Fellow at the University of Kansas Radar Systems and Remote Sensing Laboratory, where he worked on microwave sensing of the polar oceans for ONR and NASA. From 1995-2000 he was with the Space and Naval Warfare Systems Center (SPAWAR) in San Diego, where he was involved in the analysis and processing of hyperspectral sensor imagery for DARPA and ONR-sponsored programs. He is a member of IEEE and has published over 30 conference and journal articles.

Lawrence E. Hoff received the B.S.E.E. degree from Iowa State University, Ames, the M.S.E.E. degree from the University of Michigan, Ann Arbor, and the Ph.D. degree in information and computer science from the University of California, San Diego. He has specialized in target detection and classification, signal processing, and information theory. Since 1989 his work has been focused on target detection and recognition of targets in hyperspectral infrared and in UHF synthetic aperture radar data. He currently works as an independent contractor for Hoff Engineering.

Edwin M. Winter is Research Director for sensor systems at Technical Research Associates, Inc. He received his Ph.D. in geophysics from the University of California, Los Angeles, in 1972 and has been active in remote sensing for 29 years. His recent technical emphasis has been the analysis of data from hyperspectral sensors. He has presented papers at SPIE, IEEE, ERIM, EUROPTO, IRIS and other technical meetings and has published papers in the Physical Review, Remote Sensing of the Environment, Applied Optics as well as numerous conference proceedings.

Alan P. Schaum received a Ph.D. in theoretical physics in 1978 from the Johns Hopkins University, where he had applied relativistic quantum field theory to the study of elementary particle bound states. Until 1983 he worked at Bell Laboratories, applying information and queuing theories to network planning problems. For the past 18 years at the Naval Research Laboratory, he has developed signal and image processing methods in support of intelligence operations. He has published over 70 technical papers and is first author on more than 50 of these. His current research concentrates on hyperspectral detection theory and the modeling of terrestrial background signatures and their dynamics.

Alan D. Stocker has been Vice President of Space Computer Corporation, Los Angeles, since 1987. He received his M.S.E.E. from Stanford University in 1981 and previously worked on the technical staff of Hughes Aircraft Corporation and MARK Resources, Inc. in the area of radar signal processing. For the past nine years his primary technical focus has been the development and implementation of data processing algorithms for multispectral and hyperspectral imaging sensors. He has authored over 30 papers on various aspects of hyperspectral data analysis and processing, including sensor correction and calibration, spectral signature phenomenology, target detection and recognition, and real-time systems.

References

[1] R.O. Green, T.G. Chrien, P.J. Nielson, C.M. Sarture, B.T. Eng, C. Chovit, A.T. Murray, M.L. Eastwood, and H.I. Novack, "Airborne visible/infrared imaging spectrometer (AVIRIS): Recent improvements to the sensor and data facility," in *SPIE, Imaging Spectrometry of the Terrestrial Environment*, vol. 1937, G. Vane, Ed., Sept. 1993, pp. 180-190.

[2] S.R. Sandor-Leahy, D. Beiso, M.A. Figueroa, M.A. Folkman, D.A. Gleichauf, T.R. Hedman, P.J. Jarecke, and S. Thordarson, "TRWIS III hyperspectral imager: Instrument performance and remote sensing applications," in *SPIE Imaging Spectrometry IV*, vol. 3438, M.R. Descour and S.S. Shen, Eds., July 1998, pp. 13-22.

[3] C. Stellman, G. Hazel, F. Bucholtz, T. Michalowicz, A. Stocker, and W. Schaaf, "Real-time hyperspectral detection and cuing," *Opt. Eng.*, vol. 39, no. 7, pp. 1928-1935, July 2000.

[4] J.A. Hackwell, D.W. Warren, R.P. Bongiovi, S.J. Hansel, T.L. Hayhurst, D.J. Mabry, M.G. Sivjee, and J.W. Skinner, "LWIR/MWIR imaging hyperspectral sensor for airborne and ground-based remote sensing," in *SPIE Imaging Spectrometry II*, vol. 2819, M.R. Descour and J.M. Mooney, Eds., Nov. 1996, pp. 102-107.

[5] P.G. Lucey, T.J. Williams, M. Mignard, J. Julian, D. Kobubun, G. Allen, D. Hampton, W. Schaff, M.J. Schlangen, E.M. Winter, W.B. Kendall, A.D. Stocker, K.A. Horton, and A.P. Bowman, "AHI: An airborne long-wave hyperspectral imager," in *SPIE Airborne Reconnaissance XXII*, vol. 3431, W.G. Fishell, A.A. Andraitis, M.S. Fagan, J.D. Greer, and M.C. Norton, Eds., Nov. 1998, pp. 36-43.

[6] J.R. Schott, *Remote Sensing: The Image Chain Approach*. Oxford, U.K.: Oxford Univ. Press, 1997.

[7] F.G. Smith, *Atmospheric Propagation of Radiation*. Bellingham, WA: SPIE Optical Engineering Press, 1993,.

[8] G.P Anderson, A. Berk, P.K. Acharya, M.W. Matthew, L.S. Bernstein, J.H. Chetwynd, H. Dothe, S.M. Adler-Golden, A.J. Ratkowski, G.W. Felde, J.A. Gardner, M.L. Hoke, S.C. Richtsmeier, B. Pukall, J. Mello, and L.S. Jeong, "MODTRAN4: Radiative transfer modeling for remote sensing," in *SPIE, Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, vol. 4049, S. Shen and M. Descour, Eds., Apr. 2000, pp. 176-183.

[9] H. Burke, M.K. Griffin, and J.W. Snow, "Mitigation of atmospheric effects in hyperspectral data analysis," in *SPIE, Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, vol. 4049, S. Shen and M. Descour, Eds., Apr. 2000, pp. 434-443.

[10] G. Healey and D. Slater, "Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 2706-2717, Nov. 1999.

[11] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Mag.*, vol. 19, pp. 29-43, Jan. 2002.

[12] D. Manolakis, G. Shaw, and N. Keshava, "Comparative analysis of hyperspectral adaptive matched filter detectors," in *SPIE, Algorithms for Multispectral, Hyperspectral, and Ultraspectral, Imagery VI,* vol. 4049, S. Shen and M. Descour, Eds., Apr. 2000, pp. 2-17.

[13] H.V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988, pp. 7-43, 55.

[14] E.J. Kelly, "Adaptive detection and parameter estimation for multidimensional signal models," MIT Lincoln Laboratory, Lexington, MA, Tech. Rep. 848, Apr. 1989.

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

[16] L.M. Novak and M.C. Burl, "Optimal speckle reduction in polarimetric SAR imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, pp. 293-305, Mar. 1990.

[17] L.M. Novak, G.J. Owirka, and A.L. Weaver, "Automatic target recognition using enhanced resolution SAR data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, pp. 157-175, Jan. 1999.

[18] R.A. Dillard and G.M. Dillard, "Likelihood-ratio detection of frequency-hopped signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, pp. 543-553, Apr. 1996.

[19] A.D. Whalen, *Detection of Signals in Noise*. New York: Academic, 1971, pp. 109-118.

[20] L.L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Processing*, vol. 42, pp. 2146-2157, Aug. 1994.

[21] X. Yu, L.E. Hoff, I.S. Reed, M. Chen, and L.B. Stotts, "Automatic target detection and recognition in multiband imagery: a unified ml detection and estimation approach," *IEEE Trans. Image Processing*, vol. 6, pp. 143-156, Jan. 1997.

[22] S. Kraut, L.L. Scharf, L.T. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Processing*, vol. 49, pp. 1-16, Jan. 2001.

[23] N. Henze and T. Wagner, "A new approach to the BHEP tests for multivariate normality," *J. Multivariate Analysis*, no. 62, pp. 1-23, 1997.

[24] P. Masson and W. Pieczynski, "SEM algorithm and unsupervised statistical segmentation of satellite images," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 618-633, May 1993.

[25] T.K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Mag.*, vol. 31, pp. 47-60, May 1993.

[26] S.G. Beaven, D. Stein, and L.E. Hoff, "Comparison of Gaussian mixture and linear mixture models for classification of hyperspectral data," in *Proc. IGARSS 2000*, Honolulu, HI, 24-28 July 2000, pp. 1597-1599.

[27] J.W. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *4th Annu. JPL Airborne Geoscience Workshop*, Jet Propulsion Laboratory, Pasadena, CA, JPL Publication 93-26, vol. 1, 1993, pp. 11-14.

[28] J.W. Boardman, "Analysis understanding and visualization of hyperspectral data as convex sets in n-space," in *SPIE, Imaging Spectrometry*, vol. 2480 M.R. Descour, J.M. Mooney, D.L. Perry, L.R. Illing, Eds., 1995, pp. 14-22.

[29] J.M. Grossmann, J. Bowles, D. Haas, J.A. Antoniades, M.R. Grunes, P. Palmadesso, D. Gillis, K.Y. Tsang, M. Baumback, M. Daniel, J. Fisher, and I. Triandaf, "Hyperspectral analysis and target detection system for the adaptive spectral reconnaissance program (ASRP)," in *SPIE, Algorithms for Multispectral and Hyperspectral Imagery IV*, vol. 3372, 13-14 Apr. 1998, Orlando, FL, pp. 2-13.

[30] M.D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 542-552, May 1994.

[31] M.E. Winter, "Fast autonomous spectral endmember determination in hyperspectral data," in *Proc. 13th Int. Conf. Applied Geologic Remote Sensing*, vol. II, Vancouver, BC, Canada, 1999, pp. 337-344.

[32] R.A. Neville, K. Staenz, T. Szeredi, J. Lefebvre, and P. Hauff, "Automatic endmember extraction from hyperspectral data for mineral exploration," in *4th Int. Airborne Remote Sensing Conf. Exhibition/21st Canadian Symposium on Remote Sensing*, Ottawa, Ontario, Canada, 21-24 June 1999, pp. 891-896.

[33] L.E. Hoff, A.M. Chen, X. Yu, and E.M. Winter, "Generalized weighted difference algorithm for weak target detection in multiband imagery," in *SPIE, Signal and Data Processing of Small Targets*, vol. 2561, O.E. Drummand Ed., July 1995, pp. 141-152.

[34] A.D. Stocker and A.P. Schaum, "Application of stochastic mixing models to hyperspectral detection problems," in *SPIE, Algorithms for Multispectral and Hyperspectral Imagery III*, vol. 3071, S.S. Shen and A.E. Iverson, Eds., Aug. 1997, pp. 47-60.

[35] D.W.J. Stein, "Stochastic compositional models applied to subpixel analysis of hyperspectral imagery," in *SPIE, Imaging Spectrometry VII,* vol. 4480, S.S. Shen and M.S. Descour, Eds., July 2001.

[36] S.J. Press, "Linear combinations of noncentral chi-square variates," *Ann. Math Stat.*, no. 37, pp. 480-487, 1966.

[37] M.J.W. Jansen, "Maximum entropy distributions with prescribed marginals and normal score correlations," in *Distributions with Given Marginals and Moment Problems*, V. Benes and J. Stepan, Eds. New York: Kluwer, 1997, pp. 87-92.

[38] P.A.P. Moran, "Statistical inference with bivariate gamma distributions," *Biometrika*, vol. 56, no. 3, pp. 627-634, 1969.

[39] J.R. Jensen, "Urban/suburban land use analysis," in *Manual of Remote Sensing*, R.N. Colwell, Ed., vol. 2. Dobbs Ferry, NY: Sheridan, 1983, pp 1571-1666.

[40] A. Schaum and A. Stocker, "Subclutter target detection using sequences of thermal infrared multispectral imagery," in *SPIE, Algorithms for Multispectral and Hyperspectral Imagery,* vol. 3071, S.S. Shen, M.R. Descour, Eds., 1997, pp. 12-22.