Invited Paper

Is There a Best Hyperspectral Detection Algorithm?

D. Manolakis, R. Lockwood[a], T. Cooley[b] and J. Jacobson[c]

MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420
[a]Space Vehicles Directorate Air Force Research Laboratory
29 Randolph Road, Hanscom AFB, MA 01731-3010
[b]Space Vehicles Directorate Air Force Research Laboratory
Kirtland AFB, 2251 Maxwell Ave, Kirtland AFB, NM, 87117
[c]National Air and Space Intelligence Center, Wright-Patterson AFB, OH

ABSTRACT

A large number of hyperspectral detection algorithms have been developed and used over the last two decades. Some algorithms are based on highly sophisticated mathematical models and methods; others are derived using intuition and simple geometrical concepts. The purpose of this paper is threefold. First, we discuss the key issues involved in the design and evaluation of detection algorithms for hyperspectral imaging data. Second, we present a critical review of existing detection algorithms for practical hyperspectral imaging applications. Finally, we argue that the "apparent" superiority of sophisticated algorithms with simulated data or in laboratory conditions, does not necessarily translate to superiority in real-world applications.

Keywords: Hyperspectral imaging, target detection, statistical modeling, background characterization.

1. INTRODUCTION

A large number of hyperspectral detection algorithms have been developed and used over the last two decades. A partial list includes the classical matched filter, RX anomaly detector, orthogonal subspace projector, adaptive cosine estimator, finite target matched filter, mixture-tuned matched filter, subspace detectors, kernel matched subspace detectors, and joint subspace detectors. In addition, different methods for dimensionality reduction, background clutter modeling, endmember selection, and the choice between radiance versus reflectance domain processing multiply the number of detection algorithms yet further. New algorithms, new variants of existing algorithms, and new implementations of existing methods appear all the time. Furthermore, a large number of papers have been published in attempts to establish the relative superiority of these detectors. The main argument of this paper is that if we take into account important aspects of real hyperspectral imaging problems, proper use of simple detectors, like the matched filter and the adaptive cosine estimator, may provide acceptable performance for practically relevant applications. More specifically this paper has the following objectives. (a) Discuss how the at-sensor radiance physics-based signal model and the low-rank properties of background covariance matrix lead to a parsimonious taxonomy of most widely used detection algorithms. (b) Explain how the limited amount of background data with respect to the high dimensionality of the feature space limit the performance of detection algorithms that are optimum on theoretical grounds. (c) Argue that any small performance gains attained by more sophisticated detectors are irrelevant in practical applications because of the limitations and the uncertainties about many aspects of the situation in which the detector will be deployed. (d) Draw distinction between detectors which require substantial input of expertise and detectors which can be applied automatically with little external input of expertise. (e) Try to answer the question: Is there a best hyperspectral detection algorithm?

2. THE FUNDAMENTAL DETECTION PROBLEM IN HYPERSPECTRAL IMAGING

To detect a target based on its spatial properties (for example, size, shape, texture), the target must me large relative to the pixel size. In this case, some of the pixels are fully occupied by one or more target materials (full pixel or resolved targets). If the spectrum of these target pixels is distinct from the surrounding background (that is, non-target) pixels, the target can be detected spectrally as well as spatially. If a target occupies only part of a pixel (subpixel or unresolved target), it cannot

Correspondence to D.Manolakis. E-mail: dmanolakis@ll.mit.edu, Telephone: 781-981-0524, Fax: 781-981-7271

Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV,
edited by Sylvia S. Shen, Paul E. Lewis, Proc. of SPIE Vol. 7334, 733402 · © 2009 SPIE
CCC code: 0277-786X/09/$18 · doi: 10.1117/12.816917

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms

be detected using spatial techniques. However, under certain conditions, it is possible to detect subpixel targets spectrally. The spectrum of a subpixel target is mixed with the spectrum (or spectra) of the background, which results in a combined spectrum. The detectability of a full pixel target depends on the spectral contrast between target and background. The detectability of subpixel targets is also dependent on the portion of the pixel occupied by the target (fill factor), because it determines the amount of background interference on the observed target spectrum. Additional factors affecting spectral detectability include, environmental conditions, sensor noise, and statistical variability of target and background spectra.

In this paper, we focus on detection algorithms that exploit spectral information, only. Target detection algorithms[1–4] for hyperspectral imaging data, can be grouped into two types:

- **Spectral Anomaly Detection Algorithms** that do *not* require knowledge of the spectral signatures of the targets of interest. Every pixel, whose spectrum does not fit a model of the local or global background, is declared to be a target. Anomaly detectors cannot distinguish among legitimate spectral anomalies generated by manmade objects, natural objects, or targets-of-interest. Clearly, atmospheric compensation is not a prerequisite for the application of anomaly detection algorithms.
- **Spectral Matching Detection Algorithms** that require spectral information about the targets of interest, and they try to identify pixels whose spectrum exhibits a high degree of correlation ("matches") to the expected signature. The spectrum of the target of interest (reference spectrum) can be obtained from a spectral library or from an identified in-scene target pixel.

There are two key observations regarding spectral target detection algorithms. First, for any target detection algorithm to be successful, the target spectrum should be distinguishable from background spectra. Second, spectral matching algorithms exhibit better detection performance than anomaly detectors, because they look for specific spectrally defined targets. Clearly, anomaly detectors are useful for target cueing applications and may perform satisfactorily for isolated manmade targets embedded in natural homogeneous backgrounds.

Another application of hyperspectral imaging is change detection, that is, the use of HSI data obtained over multiple missions over extended time periods to detect environmental or man-made changes. Although there is some commonality between algorithms for target and change detection, there are also significant differences and challenges.[5,6] These papers provide more information about change detection algorithms and their applications.

3. PHYSICS-BASED MODEL FOR AT-SENSOR RADIANCE SIGNAL

We start with a simple discussion of the radiation components in the solar spectral region, that is, the wavelength spectrum from $0.35 - 2.5\mu m$. The total radiation signal reaching the sensor consists of three components:

1. The radiation L_1 reflected from the pixel of interest, that is, the direct and diffused solar radiation incident on the pixel surface and reflected directly into the sensor (reflected radiance).
2. The radiation L_2 reflected from the surface surrounding the pixel of interest and scattered by the air volume into the sensor (adjacency radiance).
3. The path radiance L_3, that is, the photons scattered into the sensor's field-of-view, without ground contact.

Only the reflected radiance component L_1 contains information about the pixel of interest. The task of atmospheric correction is the calculation and removal of components L_2 and L_3, and the retrieval of the pixel reflectance from component L_1. To proceed we need a model for the total at-sensor radiance $L = L_{\text{reflected}} + L_{\text{adjacency}} + L_{\text{path}}$.

A standard physics-based model for the at-sensor spectral radiance signal $L(\lambda_k)$ at a sensor pixel is given by[7]

$$L(\lambda_k) = \frac{A(\lambda_k)}{1 - \rho_e(\lambda_k)S(\lambda_k)}\rho(\lambda_k) + \frac{B(\lambda_k)\rho_e(\lambda_k)}{1 - \rho_e(\lambda_k)S(\lambda_k)} + L_a(\lambda_k) \tag{1}$$

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms

where

$$\begin{aligned}
\lambda_k &= \text{is the central wavelength of the } k\text{th spectral channel,}\\
\rho(\lambda_k) &= \text{is the pixel surface reflectance,}\\
\rho_e(\lambda_k) &= \text{is an average surface reflectance for the surrounding region,}\\
S(\lambda_k) &= \text{is the spherical albedo of the atmosphere,}\\
L_a(\lambda_k) &= \text{is the radiance scattered by the atmosphere without reaching the surface,}\\
A(\lambda_k), B(\lambda_k) &= \text{are coefficients that depend on atmospheric and geometric conditions,}\\
&\quad\ \text{but not on the surface.}
\end{aligned}$$

The first term in (1) corresponds to radiance (from both sky-shine and direct solar illumination) that is reflected from the ground pixel surface and travels directly into the sensor, while the second term corresponds to radiance from the surface surrounding the pixel that is scattered by the atmosphere into the sensor. The spatially averaged reflectance $\rho_e(\lambda_k)$ is used to account for "adjacency effects", that is, radiance contributions that, because of atmospheric scattering, originate from parts of the surface not in the direct line of sight between sensor and imaged pixel. To ignore the adjacency effect we set $\rho_e(\lambda_k) = \rho(\lambda_k)$; however, this correction can result in significant reflectance errors at short wavelengths, especially under hazy conditions and when strong contrasts occur among the materials in the scene.

Atmospheric quantities, $A(\lambda_k)$ and $B(\lambda_k)$, are the product of the top-of-atmosphere solar irradiance, total sun-to-ground transmittance, and cosine of the solar zenith angle. Furthermore, $A(\lambda_k)$ also includes the direct transmission from observer to sun while $B(\lambda_k)$, includes the medium-embedded diffuse transmittance. Thus, the $A(\lambda_k)$ and $B(\lambda_k)$ terms are surface independent coefficients that vary only with atmospheric and geometric conditions.

With the release of MODTRAN™5 the atmospheric contributions to at-sensor radiance (A, B, L_a and S) in (1) can be calculated with a single MODTRAN run using the new DISORT option for atmospheric correction. This option exploits the reciprocity principle of radiation transport theory. Equation (1) shows that the conversion from pixel reflectance to at-sensor radiance is a linear function, dependant upon the spectrum of the pixel of interest, the spatially averaged background spectrum to account for the adjacency effect and the atmospheric conditions, defined by the visibility (or aerosol content), water column density, scene and sensor geometry. Thus the transformation from pixel reflectance to at-sensor-radiance can be expressed using a simple linear equation given by

$$L(\lambda_k) = c(\lambda_k)\rho(\lambda_k) + d(\lambda_k) \tag{2}$$

where

$$c(\lambda_k) = \frac{A(\lambda_k)}{1 - \rho_e(\lambda_k)S(\lambda_k)} \quad \text{and} \quad d(\lambda_k) = \frac{B(\lambda_k)\rho_e(\lambda_k)}{1 - \rho_e(\lambda_k)S(\lambda_k)} + L_a(\lambda_k) \tag{3}$$

We note that $c(\lambda_k)$ and $d(\lambda_k)$ depend only on the atmospheric quantities $A(\lambda_k)$, $B(\lambda_k)$, and $S(\lambda_k)$, which are calculated using MODTRAN, and the spatially averaged reflectance $\rho_e(\lambda_k)$.

4. SPECTRAL VARIABILITY MODELS

The observed spectral radiance data, or derived apparent surface reflectance data, can be viewed as scattering of points in an $p-$dimensional Euclidean space, where p is the number of spectral bands. Each spectral band is assigned to one axis of the space, all axes being mutually orthogonal. Therefore, the spectrum of each pixel can be viewed as a vector defined by

$$\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_p \end{bmatrix}^T \tag{4}$$

where $x_k = L(\lambda_k)$ (radiance) or $x_k = \rho(\lambda_k)$ (reflectance). The tip of this vector corresponds to a point, whose Cartesian coordinates are the values at each spectral band Spectra without variability correspond to a single fixed point, whereas the tip of vector corresponding to spectra with variability can be anywhere within a certain volume of the spectral space. Depending on how we specify this space, there are two widely used ways to describe spectral variability.

The *statistical approach* provides a probability distribution model for the description of the spectral variability. Usually, first- and second-order moments (mean vector and covariance matrix) are employed under a multivariate normal distribution assumption. Clearly, variability is related to the spread of the distribution, and the highest variability is obtained for a uniform distribution over the data space.

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms

The *geometric approach* restricts the spectrum vector to vary in an M-dimensional subspace of the data space $(M < p)$. The observed spectrum is described by a linear combination of vectors $\boldsymbol{s}_k$, which define the variability subspace, can be (a) endmembers determined from spectral libraries or the data, or (b) vectors obtained with statistical techniques (for example, the eigenvectors of the data correlation matrix). Clearly, the variability increases as M increases from one to p.

4.1 Probabilistic Models

Since HSI data cubes are often obtained from scenes with multiple land covers, the underlying theoretical distributions are multimodal. When such situations arise, we subdivide the scene into classes such as each class can be characterized by a unimodal distribution $f_m(\boldsymbol{x})$. This yields a finite (probability) mixture model of the form

$$f(\boldsymbol{x}) = \sum_{m=1}^{M} \pi_m f_m(\boldsymbol{x}) \tag{5}$$

where M is the number of classes and $\pi_k \geq 0$ $(\sum_k \pi_k = 1)$ is the proportion of the k-th class. For the purposes of this paper we determine the classes by interactive specification of land cover areas. Class specification with automated clustering algorithms using spectral similarity metrics (spectrally homogeneous classes) and by fitting normal mixture models (statically homogeneous classes) are discussed in.[8]

Multivariate Normal Distribution If a p-dimensional vector $\boldsymbol{x}$ comes from a multivariate normal (or Gaussian) distribution, its probability density function is given by

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right] \tag{6}$$

where $\boldsymbol{\mu}$ is the $p \times 1$ mean vector with entries $\mu_i = E(x_i)$, $\boldsymbol{\Sigma}$ is the $p \times p$ covariance matrix with entries $\sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$, and $|\ |$ denotes matrix determinant. The symbol E denotes the mathematical expectation operator which provides averaging over the ensemble of all possible realizations. In compact matrix notation we have

$$\boldsymbol{\mu} = E(\boldsymbol{x}) \tag{7}$$

$$\boldsymbol{\Sigma} = E[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T] \tag{8}$$

We often use the shorthand notation $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the normal distribution in (6). If the matrix $\boldsymbol{\Sigma}$ is singular, then with probability one the distribution of $\boldsymbol{x}$ is confined to a subspace of the p-dimensional data space. The quadratic form

$$\Delta^2(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}), \tag{9}$$

is known as the (squared) Mahalanobis distance*. The density function (6) has a constant value on the locus of points on an ellipsoid defined by $\Delta^2(\boldsymbol{x}) = C$. The random variable (9) is distributed according to a (central) chi-square distribution with p degrees-of-freedom

$$\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Rightarrow \Delta^2(\boldsymbol{x}) \sim \chi_p^2 \tag{10}$$

The normal distribution becomes very sparse for large p. Indeed, using (10) we can show that $\Pr(\Delta^2 \leq 3.84) = 0.95$ for $p = 1$, but for $p = 10$ the probability is nearly 0.05.

Multivariate t-Distribution Consider a random vector $\boldsymbol{y} \sim N_p(\boldsymbol{0}, \boldsymbol{R})$ with standardized components $y_k \sim N(0,1)$. Then, the covariance matrix $\boldsymbol{R}$ of $\boldsymbol{y}$ has unit diagonal elements. Suppose now that we randomly modulate the components $y_1, y_2, \ldots, y_p$ with a statistically independent random variable $s \sim \chi_v^2$ and add a constant vector $\boldsymbol{\mu}$ as follows

$$\boldsymbol{x} = \boldsymbol{y}\left(\frac{s}{\nu}\right)^{-1/2} + \boldsymbol{\mu} = \frac{\boldsymbol{y}}{\sqrt{s/\nu}} + \boldsymbol{\mu} \tag{11}$$

*Strictly speaking Δ is the Mahalanobis distance and Δ^2 is the squared Mahalanobis distance; however, for simplicity, the term Mahalanobis distance is used in both cases. The exact meaning should be clear from the context.

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms

The resulting random vector $\boldsymbol{x}$ has a multivariate t-distribution, denoted by $\boldsymbol{x} \sim t_p(\boldsymbol{\mu}, \boldsymbol{R}, \nu)$, with probability density function[?]

$$f(\boldsymbol{x}) = \frac{\Sigma\left(\frac{p+\nu}{2}\right)}{(\pi\nu)^{\frac{p}{2}}\Sigma\left(\frac{\nu}{2}\right)\sqrt{|\boldsymbol{R}|}}\left[1+\frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{R}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]^{-\frac{p+\nu}{2}} \tag{12}$$

where $\Sigma(\ )$ is the Gamma function. The number of degrees of freedom ν controls the tails of the distribution: $\nu = 1$ leads to the multivariate Cauchy distribution (heaviest tails), whereas as $\nu \to \infty$ the t-distribution approaches the multivariate normal distribution (lightest tails) with covariance matrix $\boldsymbol{R}$.

If $\nu > 1$, $\boldsymbol{\mu}$ is the mean of $\boldsymbol{x}$, and if $\nu > 2$, $\nu\boldsymbol{R}/(\nu-2)$ is its covariance matrix, that is, we have

$$E(\boldsymbol{x}) = \boldsymbol{\mu}, \quad \nu > 1 \tag{13}$$

$$\mathrm{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma} = \frac{\nu}{\nu-2}\boldsymbol{R}, \quad \nu > 2 \tag{14}$$

The Mahalanobis distance from $\boldsymbol{x}$ to the center $\boldsymbol{\mu}$, with respect to $\boldsymbol{R}$, follows a central F-distribution with p and ν degrees of freedom

$$\frac{1}{p}\delta^2(\boldsymbol{x}) = \frac{1}{p}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{R}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \sim F_{p,\nu} \tag{15}$$

Because $f(\boldsymbol{x})$ in (12) depends on $\boldsymbol{x}$ only through $\delta^2(\boldsymbol{x})$, the probability density is the same for all $\boldsymbol{x}$ that have the same Mahalanobis distance from $\boldsymbol{\mu}$, and thus the distribution is ellipsoidally symmetric about $\boldsymbol{\mu}$. The distribution of the Mahalanobis distance with respect to $\boldsymbol{\Sigma}$ can be obtained from (14) and (15)

$$\frac{1}{p}\frac{\nu}{\nu-2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \sim F_{p,\nu} \tag{16}$$

One of the advantages of the multivariate t-distribution is the presence of the second degree of freedom ν which adjusts the shape of the distributional tail. When $\nu = 1$ the multivariate t-distribution is equal to the multivariate Cauchy distribution and has very heavy tails. The Cauchy distribution is a member of the family of stable distributions, which have no second-order moments. As $\nu \to \infty$ the distribution tends towards the multivariate normal distribution and has lighter tails.

Elliptically Contoured Distributions The multivariate normal distribution and the multivariate t-distribution belong to the class of elliptically contoured distributions. These distributions have elliptical contours of equal probability and a "bell-shaped surface" with lighter or heavier tails than the normal distribution.

A random vector $\boldsymbol{x}$ has an elliptical distribution with location vector $\boldsymbol{\mu}$ and scale positive definite matrix $\boldsymbol{C}$, if its density takes the form[9]

$$f(\boldsymbol{x}) = c_p|\boldsymbol{C}|^{-1/2}g\left[(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right] \tag{17}$$

where the function g is typically known as the density generator. For a vector $\boldsymbol{x}$ distributed according to (17), we use the notation $\mathrm{ECD}_p(\boldsymbol{\mu}, \boldsymbol{C}; g)$ or simply $\mathrm{ECD}_p(\boldsymbol{\mu}, \boldsymbol{C})$. When $\boldsymbol{\mu} = \boldsymbol{0}$ and $\boldsymbol{C} = \boldsymbol{I}$, we obtain the spherical family of densities. For the multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we have $g(u) = \exp(-u/2)$, $c_p = (2\pi)^{-p/2}$, and $\boldsymbol{C} = \boldsymbol{\Sigma}$. This class of symmetric distributions includes the normal, Student t, Cauchy, and logistic distributions, among others.[10]

4.2 Physics-Based Models: The Linear Mixture Model

The basic premises of linear mixture modeling are that within a given scene : (a) the surface is dominated by a small number of materials with relatively constant spectra (end members), (b) most of the spectral variability within the scene results from varying proportions of the end members, and (c) the mixing relationship is linear if the end members are arranged in spatially distinct patterns.[11]

In the linear mixing model (LMM), the spectrum of a mixed pixel is represented[?] as a linear combination of component spectra (end members). The weight of each end member spectrum (abundance) is proportional to the fraction of the pixel area covered by the end member. If there are p spectral bands, the spectrum of the pixel and the spectra of the end members can be represented by p-dimensional vectors. Therefore, the general equation for mixing by area is given by

$$\mathbf{x} = \sum_{k=1}^{M} a_k\mathbf{s}_k + \mathbf{w} \triangleq \mathbf{S}\mathbf{a} + \mathbf{w} \tag{18}$$

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms

where $\boldsymbol{S} \triangleq [\boldsymbol{s}_1 \boldsymbol{s}_2 \ldots \boldsymbol{s}_M]$ and $\mathbf{a} \triangleq [a_1 a_2 \ldots a_M]^T$, $\mathbf{x}$ is the spectrum of the mixed pixel, $\mathbf{s}_k$ are the spectra of the end members, a_k are their abundances, M is the number of the end members, and $\mathbf{w}$ is an p-dimensional error vector accounting for lack-of-fit and noise effects. Physical considerations dictate the following constraints $a_k \geq 0$ (non-negativity constraint) and $\sum_{k=1}^{M} a_k = 1$ (additivity constraint), which can be enforced, if necessary, to guarantee meaningful parameter values. If we wish to allow for unmodeled end members, we can replace the "=" sign after the summation with the "$\leq$" sign. It should be emphasized that although the positivity constraint is quite meaningful, it is not always enforced because it significantly complicates the solution of detection and unmixing problems.

Fitting a linear mixing model involves two steps: (a) end member identification and (b) abundance estimation. Although there are algorithms where the two steps are interwoven, the objectives of this paper are better served by keeping the two steps distinct. If we know the end members $\mathbf{s}_k$, unmixing can be viewed either as a linear estimation problem or as a linear model fitting problem. Furthermore, in detection and classification applications, we do not need to determine explicitly estimates of $\mathbf{a}$; we can use a measure of the quality of the estimator or a measure of the model "goodness-of-fit" to make decisions.

5. DETECTOR DESIGN STRATEGIES

It is well known from statistical detection theory that decisions based upon the likelihood ratio test are optimum for a wide range of performance criteria. Let $f_0(\boldsymbol{x}|H_0)$ be the conditional probability density function of observing the spectrum $\boldsymbol{x}$ under the H_0 (target absent) hypothesis and $f_1(\boldsymbol{x}|H_1)$ be the probability of observing $\boldsymbol{x}$ under the H_1 (target present) hypothesis. The likelihood ratio test (LRT) is defined by

$$\Lambda(\boldsymbol{x}) \triangleq \frac{f_1(\boldsymbol{x}|\text{target present})}{f_0(\boldsymbol{x}|\text{target absent})} = \frac{f_1(\boldsymbol{x}|H_1)}{f_0(\boldsymbol{x}|H_0)} \tag{19}$$

If $\Lambda(\boldsymbol{x})$ exceeds a certain threshold η then the target present hypothesis is accepted as true, while if $\Lambda(\boldsymbol{x})$ is less than this threshold, the target absent hypothesis is accepted. Depending on whether we want to use the maximum a-posteriori criterion of optimality, a Bayes cost function, or the Neyman-Pearson criterion, the threshold will change, but the detector will remain the same. In surveillance systems we use the Neyman-Pearson (NP) criterion,[12] which maximizes the probability of detection P_{D}while keeping the probability of false alarm P_{FA}under a certain predetermined value.

To determine the likelihood ratio $\Lambda(\boldsymbol{x})$ it is necessary to know the conditional densities $f_0(\boldsymbol{x}|H_0)$ and $f_1(\boldsymbol{x}|H_1)$. However, in many practical cases these distributions have some unknown parameters arranged as a vector $\boldsymbol{\theta}$. If the unknown parameters are nonrandom, we determine the maximum likelihood estimate of $\boldsymbol{\theta}$ under each hypothesis and use these estimates to form the generalized likelihood ratio test (GLRT)

$$\Lambda_G(\boldsymbol{x}) \triangleq \frac{f_1(\boldsymbol{x}|\hat{\boldsymbol{\theta}}_1, H_1)}{f_0(\boldsymbol{x}|\hat{\boldsymbol{\theta}}_0, H_0)}, \quad \hat{\boldsymbol{\theta}}_i = \max_{\boldsymbol{\theta}} f_i(\boldsymbol{x}|\boldsymbol{\theta}, H_i) \tag{20}$$

If the unknown parameters are modeled as random variables with a known a-priori density function $f_\theta(\boldsymbol{\theta})$, we can use the Bayes likelihood ratio test, which is defined by

$$\Lambda_B(\boldsymbol{x}) \triangleq \frac{\int f_1(\boldsymbol{x}|\boldsymbol{\theta}, H_1) f_\theta(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int f_0(\boldsymbol{x}|\boldsymbol{\theta}, H_0) f_\theta(\boldsymbol{\theta}) d\boldsymbol{\theta}} \tag{21}$$

We note that integration over $\boldsymbol{\theta}$ results essentially to a likelihood ratio test. However, the use of this test is prohibited when the a-priori density $f_\theta(\boldsymbol{\theta})$ is complicated or unknown. For this reason we focus on the LRT and the GLRT approaches.

Detection performance is commonly measured in terms of Receiver Operating Characteristic (ROC) curves. The ROC is a plot showing the probability of detection (given by $P_D = \Pr[\Lambda(\boldsymbol{x}|H_1) > \eta]$) versus the probability of false alarm (given by $P_{FA} = \Pr[\Lambda(\boldsymbol{x}|H_0) > \eta]$), as a function of threshold, for specific distributions $f_0(\boldsymbol{x}|H_0)$ and $f_1(\boldsymbol{x}|H_1)$.

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms

5.1 Detection Algorithms for Spectra Modeled by Gaussian Distributions

Full pixel or resolved targets can be modeled reasonably well by a multivariate Gaussian distribution. If we also assume a Gaussian distribution for the background clutter, the competing hypotheses are

$$H_0: \quad \boldsymbol{x} \sim N_p(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) \tag{22}$$

$$H_1: \quad \boldsymbol{x} \sim N_p(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t) \tag{23}$$

The mean vectors and covariance matrices of the target and background distributions are assumed to be known. We next determine likelihood ratio detectors for some cases of special interest in practical applications.

Quadratic Detector ($\boldsymbol{\Sigma}_t \neq \boldsymbol{\Sigma}_b$) In the general case, where $\boldsymbol{\Sigma}_t \neq \boldsymbol{\Sigma}_b$, the likelihood ratio is given by

$$\Lambda(\boldsymbol{x}) = \frac{|\boldsymbol{\Sigma}_b|^{1/2} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_t)^T \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_t)\right]}{|\boldsymbol{\Sigma}_t|^{1/2} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_b)^T \boldsymbol{\Sigma}_b^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_b)\right]} \tag{24}$$

Taking the logarithm and ignoring constant terms yields the following detection statistics

$$y = D(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{\mu}_b)^T \boldsymbol{\Sigma}_b^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_b) - (\boldsymbol{x}-\boldsymbol{\mu}_t)^T \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_t) \tag{25}$$

which compares the Mahalanobis distance of the spectrum under test from the means of background and target distributions. The detection statistics is a quadratic function of the observations; hence the name quadratic detector.

The variable $y = D(\boldsymbol{x})$ is a random variable whose probability density depends on which hypothesis is true. If the two densities $f(y|H_0)$ and $f(y|H_1)$ are known, then the probabilities of detection and false alarm are given by

$$P_D(\eta) = \int_\eta^\infty f(y|H_1)dy \quad \text{and} \quad P_{FA}(\eta) = \int_\eta^\infty f(y|H_0)dy \tag{26}$$

The computation of these integrals is quite complicated;[13] therefore, ROC curves for the quadratic detector can be only evaluated by Monte-Carlo simulation.

Target Covariance Shrinkage ($\boldsymbol{\Sigma}_t = \gamma^2 \boldsymbol{\Sigma}_b$) Since most targets are man made objects their spectral variability should be "smaller" than the variability of natural backgrounds. In this case, we can assume that $\boldsymbol{\Sigma}_t = \gamma^2 \boldsymbol{\Sigma}_b$, where $0 < \gamma^2 \leq 1$. In this case, the detection statistic (25), is simplified to

$$y = D(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{\mu}_b)^T \boldsymbol{\Sigma}_b^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_b) - \gamma^{-2}(\boldsymbol{x}-\boldsymbol{\mu}_t)^T \boldsymbol{\Sigma}_b^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_t) \tag{27}$$

$$= \|\tilde{\boldsymbol{x}}_b\|^2 - \gamma^{-2}\|\tilde{\boldsymbol{x}}_t\|^2 \tag{28}$$

where we have used the background whitening transformation

$$\tilde{\boldsymbol{x}}_t \triangleq \boldsymbol{\Sigma}_b^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu}_t), \quad \tilde{\boldsymbol{x}}_b \triangleq \boldsymbol{\Sigma}_b^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu}_b) \tag{29}$$

Matched Filter Detector ($\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_b$) If the target covariance matrix is equal to the background covariance matrix, that is, $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_b$, the likelihood ratio detector becomes a linear function of the observations, that is,

$$y = D(\boldsymbol{x}) = \kappa(\boldsymbol{\mu}_t - \boldsymbol{\mu}_b)^T \boldsymbol{\Sigma}_b^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_b) \tag{30}$$

where κ is a normalization constant. The choice of κ does not affect detection performance; in practice we set $\kappa = 1/\Delta_b^2$ because $y = D(\boldsymbol{\mu}_t) = 1$. This detection statistics can be written as

$$y = \boldsymbol{h}^T(\boldsymbol{x}-\boldsymbol{\mu}_b) \tag{31}$$

where

$$\boldsymbol{h} \triangleq \kappa \boldsymbol{\Sigma}_b^{-1}(\boldsymbol{\mu}_t - \boldsymbol{\mu}_b) \tag{32}$$

The linear processor (31) is known as the matched filter or Fisher's linear discriminant.

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms

Adaptive Matched Filter Detector ($\boldsymbol{\mu}_t = a\boldsymbol{s}_0, \boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_b$, unknown a) The two competing hypotheses are

$$H_0: \quad \boldsymbol{x} \sim N_p(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) \tag{33}$$
$$H_1: \quad \boldsymbol{x} \sim N_p(a\boldsymbol{s}_0, \boldsymbol{\Sigma}_b) \tag{34}$$

Since the value of a is unknown, it has to be estimated from the observation $\boldsymbol{x}$. The maximum likelihood estimate of a under the H_1 hypothesis is

$$\hat{a} = \frac{\boldsymbol{s}_0^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x}}{\boldsymbol{s}_0^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{s}_0} \tag{35}$$

Using (35) and the generalized likelihood ratio rule leads we obtain the following detection statistic

$$y = D_{\text{AMF}}(\boldsymbol{x}) = \frac{(\boldsymbol{s}_0^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x})^2}{\boldsymbol{s}_0^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{s}_0} - 2\boldsymbol{\mu}_b^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x} + \boldsymbol{\mu}_b^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{\mu}_b \tag{36}$$

Anomaly Detector ($\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_b$, unknown $\boldsymbol{\mu}_t$) In many situations of practical interest, we do not have sufficient a priori information to specify the statistics of the target class. Therefore, we can use neither the quadratic detector (25) nor the matched filter detector (30). We clearly need a detector which exclusively uses background information. The maximum likelihood estimate of $\boldsymbol{\mu}_t$ under the H_1 hypothesis is given by $\hat{\boldsymbol{\mu}} = \boldsymbol{x}$. Therefore, the generalized likelihood ratio yields the following detection statistic

$$y = D_{\text{MD}}(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{\mu}_b)^T \boldsymbol{\Sigma}_b^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_b) \tag{37}$$

which is the Mahalanobis distance of the observed spectrum $\boldsymbol{x}$ from the center of the background distribution. The detector (37), which is now commonly referred to as the as the Reed-Xiaoli (RX)[14] anomaly detector, is considered as the benchmark anomaly algorithm for hyperspectral data.

5.2 Additive and Replacement Signal Models: Virtues and Pitfalls

Additive Signal Models The signal model most often used for the development of detection algorithms in hyperspectral imaging is given by

$$H_0: \quad \boldsymbol{x} = \boldsymbol{v} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_b) \tag{38}$$
$$H_1: \quad \boldsymbol{x} = a\boldsymbol{s} + \boldsymbol{v} \sim N_p(a\boldsymbol{s}, \boldsymbol{\Sigma}_b), \ a > 0 \tag{39}$$

This model assumes that (a) the scaled target spectrum $a\boldsymbol{s}$ is superimposed on the background spectrum $\boldsymbol{v}$, which accounts for both clutter and sensor noise, (b) the background has the same mean under both hypotheses (hence we can center the data to zero by subtracting $\boldsymbol{\mu}_b$), and (c) target and background have the same covariance matrix. Clearly, none of these assumptions is valid for real hyperspectral imaging data.

If we assume that a is known, the likelihood ratio approach yields the matched filter detector

$$y = a\boldsymbol{s}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x} \tag{40}$$

Since $a > 0$ the detection statistics $a\boldsymbol{s}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x}$ and $\boldsymbol{s}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x}$ have identical performance (same ROC curves). Therefore, we obtain a matched filter detector which uses only the shape of the target signature

$$y = D_{\text{MF}} = \boldsymbol{s}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x} \tag{41}$$

If the sign of a is unknown, we have to use the GLRT approach. The result is the adaptive matched filter

$$y = D_{\text{AMF}}(\boldsymbol{x}) = \frac{(\boldsymbol{s}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x})^2}{\boldsymbol{s}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{s}} \tag{42}$$

which is obtained from (36) by setting $\boldsymbol{\mu}_b = \boldsymbol{0}$. Careful inspection of (33)-(34) and (38)-(39) shows that for $\boldsymbol{\mu}_b \neq \boldsymbol{0}$ we cannot derive the second model from the first by subtracting the mean value of the background; in other words, the two models are not equivalent.

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms

Replacement Signal Models When an opaque target occupies part of the pixel it covers or "replaces" part of the background. If we denote by $\boldsymbol{s}$ the spectrum of the target, by $\boldsymbol{v}$ the spectrum of the background, and by α the fraction of the pixel area filled by the target, the observed pixel spectrum is

$$\boldsymbol{x} = \alpha \boldsymbol{s} + (1-\alpha)\boldsymbol{v}, \quad 0 \leq \alpha \leq 1 \tag{43}$$

which is known as the *replacement signal model* for sub-pixel targets. We assume that there is additive sensor noise which, for simplicity, has been incorporated into the background and target spectra. There are two interesting special cases

$$\boldsymbol{x} \approx \alpha \boldsymbol{s} + \boldsymbol{v}, \quad \text{if } 0 < a \ll 1 \tag{44}$$

$$\boldsymbol{x} \approx \boldsymbol{v}, \quad \text{if } a \approx 1 \tag{45}$$

Therefore, the additive target model (39) is a good approximation for targets with very small or very large fill factors. For targets with intermediate fill factor values it is quite challenging to explain the physical meaning of the additive model.

Since the fill factor a in the replacement model is unknown, we can derive a detection algorithm using the GLRT approach. The observation $\boldsymbol{x}$ under the H_1 hypothesis has a normal distribution with mean and covariance given by

$$\boldsymbol{\mu}(a) = a\boldsymbol{\mu}_t + (1-a)\boldsymbol{\mu}_b \tag{46}$$

$$\boldsymbol{\Sigma}(a) = a^2\boldsymbol{\Sigma}_t + (1-a)^2\boldsymbol{\Sigma}_b \tag{47}$$

The generalized likelihood ratio, which is a nonlinear function of a, is given by

$$\Lambda_G(\boldsymbol{x}) = \frac{\max_a N_p(\boldsymbol{\mu}(a), \boldsymbol{\Sigma}(a))}{N_p(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)} \tag{48}$$

Maximization of the numerator requires nonlinear optimization techniques. This algorithm, which was introduced by Schaum and Stocker,[15] is known as Finite Target Matched Filter (FTMF). The maximization in the numerator of (48) has to be done numerically; a closed form solution is available for the special case $\boldsymbol{\Sigma}_t = \gamma^2\boldsymbol{\Sigma}_b$. The success of this algorithm depends on how accurately we know the covariance matrix of the target class. Clearly, this is a difficult requirement to satisfy in practical applications.

Another attempt to exploit the replacement model for improved detection of subpixel targets by Boardman[16,17] led to the Mixture Tuned Matched Filter (MTMF), which has been implemented in the ENVI software package. The MTMF uses the output of the standard matched filter and an infeasibility index, which is related to the target fill-factor. The MTMF uses two thresholds: one for the output of the matched filter and one for the infeasibility index.

5.3 Detection Algorithms for Additive Subspace Models of Spectral Variability

We saw that the optimum detector (59) for targets and backgrounds modeled by normal distributions with different covariance matrices is a quadratic function of the observations. The quadratic detector is reduced to the linear matched filter (31) if $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_b$ or if the observations are described by the additive signal model (39). The additive signal model assumes a *fixed* target with known shape and unknown amplitude. A mathematically tractable approach to introduce additional target variability is to assume that target "lies" at a known *subspace* of the spectral observation space. The rank q_t of the subspace reflects the degree of uncertainty about the target. Signals with known shape, that is known direction and unknown amplitude, have rank one. The uncertainty about the target is maximum for $q_t = p$.

Subspace Targets on Normally Distributed Backgrounds For subspace targets, the additive model (39) takes the form

$$\begin{array}{lll} H_0: & \boldsymbol{x} = \boldsymbol{v}, & \boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b) \\ H_1: & \boldsymbol{x} = \mathbf{S}\mathbf{a} + \boldsymbol{v}, & \boldsymbol{x} \sim N(\mathbf{S}\mathbf{a}, \boldsymbol{\Sigma}_b) \end{array} \tag{49}$$

We assume that the background covariance matrix $\boldsymbol{\Sigma}_w$ is known. This assumption simplifies the derivation of several detection algorithms and their performance. The approach is reasonable when we use a large number of pixels to estimate

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms

the covariance matrix; if the covariance matrix is assumed unknown, we obtain a family of adaptive[18] detection algorithms. The maximum likelihood estimator of $\boldsymbol{a}$ is given by

$$\hat{\boldsymbol{a}} = (\boldsymbol{S}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{S})^{-1} \boldsymbol{S}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x} \tag{50}$$

Using the generalized likelihood ratio approach, we obtain the following detector[19–22]

$$y = D(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{S} (\boldsymbol{S}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{S})^{-1} \boldsymbol{S}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x} \tag{51}$$

Although there is no optimality test associated with the GLRT approach,[12] it leads to the design of useful practical detectors.

If $q_t = p$ and the matrix $\boldsymbol{S}$ has full rank (that is, $\boldsymbol{S}$ is invertible), equation (51) leads to the following detector[23]

$$y = D_{\mathrm{MD}}(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{\mathrm{A}} \tag{52}$$

which is identical to the RX anomaly detection algorithm for full pixel targets (37).

A key assumption in the derivation of (51) was that the covariance matrix of the background is the same under the two hypotheses. However, for sub-pixel targets the pixel area covered by background is different under the two hypotheses. Therefore, it is more reasonable to assume that the covariance under the H_1 hypothesis is given by $\beta^2 \boldsymbol{\Sigma}_b$. In other words, the background has the same covariance structure but different variance. This variance is directly related to the fill factor of the target, that is, the percentage of the pixel area occupied by the target object. The competing hypotheses are

$$\begin{array}{lll} H_0: & \boldsymbol{x} = \boldsymbol{v}, & \boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_b) \\ H_1: & \boldsymbol{x} = \mathbf{Sa} + \beta \boldsymbol{v}, & \boldsymbol{x} \sim N(\mathbf{Sa}, \beta^2 \boldsymbol{\Sigma}_b) \end{array} \tag{53}$$

The GLRT approach leads to the following Adaptive Coherence/Cosine Estimator (ACE) detector[18,24]

$$y = D_{\mathrm{ACE}}(\boldsymbol{x}) = \frac{\boldsymbol{x}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{S} (\boldsymbol{S}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{S})^{-1} \boldsymbol{S}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{\mathrm{ACE}} \tag{54}$$

For $q_t = 1$, that is, for targets with *known* shape, the target subspace $\boldsymbol{S}$ is specified by the direction of a single vector $\mathbf{s}$

$$y = D(\boldsymbol{x}) = \frac{(\boldsymbol{s}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x})^2}{(\boldsymbol{s}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{s})(\boldsymbol{x}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{x})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{55}$$

The ROC curves for the subspace target detectors depend monotonically on the signat-to-clutter ratio

$$\mathrm{SCR_o} = (\boldsymbol{Sa})^T \boldsymbol{\Sigma}_b^{-1} (\mathbf{Sa}) \tag{56}$$

The performance of all GLRT detectors depends only on the dimensional integer numbers p, q, and the optimum $\mathrm{SCR_o}$ parameter.

Subspace Targets on Subspace Backgrounds If we model the background by a subspace model $\boldsymbol{v} = \boldsymbol{B}\boldsymbol{a}_b + \boldsymbol{w}$, where $\boldsymbol{B}$ is a $p \times q_b$ matrix and $\boldsymbol{w}$ a $p \times 1$ random vector, we have the following hypotheses

$$\begin{array}{lll} H_0: & \boldsymbol{x} = \boldsymbol{B}\boldsymbol{a}_{\mathrm{b},0} + \boldsymbol{w} & \text{(Target absent)} \\ H_1: & \boldsymbol{x} = \boldsymbol{S}\boldsymbol{a} + \boldsymbol{B}\boldsymbol{a}_{\mathrm{b},1} + \mathbf{w} & \text{(Target present)} \end{array} \tag{57}$$

For the derivation of the detection algorithm we assume that $\boldsymbol{S}$ and $\boldsymbol{B}$ are known. If we assume that $\boldsymbol{w} \sim N_p(\boldsymbol{0}, \sigma_w^2 \boldsymbol{I})$, with known σ_w^2, the optimum detector is obtained using the likelihood ratio. The result is[25]

$$y = D_{\mathrm{SB}}(\boldsymbol{x}) = \boldsymbol{x}^T (\boldsymbol{P}_S^{\perp} - \boldsymbol{P}_B^{\perp}) \boldsymbol{x} / \sigma_w^2 \tag{58}$$

If σ_w^2 is unknown we can use the generalized likelihood ratio. This leads to the following algorithm[12,25]

$$y = D_{\mathrm{GSB}}(\boldsymbol{x}) = \frac{\boldsymbol{x}^T (\boldsymbol{P}_B^{\perp} - \boldsymbol{P}_{SB}^{\perp}) \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{P}_{SB}^{\perp} \boldsymbol{x}} = \frac{\boldsymbol{x}^T \boldsymbol{P}_B^{\perp} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{P}_{SB}^{\perp} \boldsymbol{x}} - 1 \tag{59}$$

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms

where $\boldsymbol{P}_A^{\perp} = \boldsymbol{I} - \boldsymbol{P}_A$ and $\boldsymbol{P}_A = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ is the projection matrix onto the column space of matrix A. The subscript SB denotes projection operators using the combined matrix $[\boldsymbol{S}\,\boldsymbol{B}]$. For full pixel targets the denominator in () should be replaced by $\boldsymbol{x}^T\boldsymbol{P}_B^{\perp}\boldsymbol{x}$. The ROC curves are monotonically related to the following signal-to-clutter ratio

$$\mathrm{SCR_o} = (\mathbf{Sa})^T\mathbf{P}_B^{\perp}(\mathbf{Sa})/\sigma_w^2 \tag{60}$$

which depends on the length of the target projection onto the "background-free" subspace and the "everywhere-present" noise variance σ_w^2. In the statistical literature $D_{\mathrm{SB}}(\boldsymbol{x})$ is denoted by $F(\boldsymbol{x})$ and is known as the F-test. When $q_b = 1$, it can be shown that the amplitude of the target signature in (57) can be estimated by the formula?

$$\hat{a} = \boldsymbol{s}^T\boldsymbol{P}_B^{\perp}\boldsymbol{x}/\boldsymbol{s}^T\boldsymbol{P}_B^{\perp}\boldsymbol{s} \tag{61}$$

The numerator of (61), that is, $y = \boldsymbol{s}^T\boldsymbol{P}_B^{\perp}\boldsymbol{x}$, is known as the orthogonal subspace projector (OSP)[26] detection algorithm.

Another subspace detector[27] uses the following hypotheses

$$\begin{array}{ll} H_0: & \boldsymbol{x} = \boldsymbol{B}\boldsymbol{a}_b + \boldsymbol{w}, \quad \boldsymbol{w} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_w) \\ H_1: & \boldsymbol{x} = \boldsymbol{S}\boldsymbol{a} + \boldsymbol{w} \end{array} \tag{62}$$

with known $\boldsymbol{S}$, $\boldsymbol{B}$, and $\boldsymbol{\Sigma}_w$. Since $\boldsymbol{\Sigma}_w$ is known, whitening reduces (62) to (57) with $\sigma_w^2 = 1$. This yields the detector $y = \tilde{\boldsymbol{x}}^T(\boldsymbol{P}_{\tilde{S}}^{\perp} - \boldsymbol{P}_{\tilde{B}}^{\perp})\tilde{\boldsymbol{x}}$ in the whitened space or the detector

$$y = \boldsymbol{x}^T\boldsymbol{\Sigma}_w^{-1}\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{\Sigma}_w^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{\Sigma}_w^{-1}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{\Sigma}_w^{-1}\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{\Sigma}_w^{-1}\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{\Sigma}_w^{-1}\boldsymbol{x} \tag{63}$$

Careful inspection of (62) suggests that the target subspace replaces the background subspace, but a part of the background described by $\boldsymbol{w} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_w)$ is the same under both hypotheses. The major practical obstacle with this approach is obtaining $\boldsymbol{B}$ and $\boldsymbol{\Sigma}_w$. It has been suggested[28] that $\boldsymbol{\Sigma}_w = \boldsymbol{I}$; in this case (63) is reduced to (58).

5.4 Beyond Gaussianity

It is well known that hyperspectral backgrounds are non-Gaussian. Therefore, it is important to (a) understand what are the implications on the performance of the detection algorithms derived under the Gaussian assumption and (b) to develop detection algorithms for alternative more accurate models of hyperspectral backgrounds.

Elliptically Contoured Distributions It has been shown[29] that the AMF, Kelly GLRT,[19] and ACE[18] detection algorithm structures are *not* a product of Gaussianity, but rather a product of elliptical symmetry which the Gaussian distributions happen to possess. Indeed, it turns out that a large class of ECDs lead to the same detector structures as the normal distribution. This suggests that the AMF, Kelly-GLRT, and ACE detectors are robust to deviations from normality and to backgrounds with heavy-tails. A detector for an ECD with exponential tails (obtained by changing the exponent of the normal distribution from two to one) has been developed[4] and applied to real hyperspectral data.

Mixtures of Gaussian Distributions Clearly, a mixture of normal or elliptically contoured distributions provide a better model for hyperspectral backgrounds. Thus, we could use this model in conjunction with the GLRT to obtain a better detection algorithm. Unfortunately, there are two major obstacles in this approach: (a) estimating the parameters of the model is computationally expensive, and (b) using the same amount of data to estimate a much larger number of parameters results to less accurate estimates (bias-variance tradeoff). Therefore, it is not guaranteed that this more sophisticated model will lead to improved detection performance. Another approach starts by segmenting the data into multiple classes, using a clustering algorithm. Then we find the closest cluster to each test pixel and apply a detector designed using the mean and covariance of this cluster. Although intuitively appealing, this approach does not perform as well as expected.

Bayesian Detectors Use of the Bayesian LRT (52) with various types of a-priori distributions[30] results to several detection algorithms. Although this approach may provide some additional insight to existing detection algorithms, the resulting algorithms cannot be easily applied to practical applications. A nice discussion of several extensions to classical detection algorithms is given in[31] and the provided references.

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms

6. THEORY VERSUS PRACTICE: IMPLICATIONS OF REAL-WORLD CONSIDERATIONS

Most hyperspectral detection algorithms have been known and studied extensively by the signal processing, radar, and pattern recognition communities for a long time. The hyperspectral imaging community has adopted and adapted these algorithms to the special characteristics of spectral signals. However, some efforts have been undertaken to extend some algorithms by taking into account the target replacement signal model or the convex-space constrained linear mixing model. The large number of hyperspectral detection algorithms can be reduced to a handful of fundamental algorithms if we take into account the target and background information that is available in practical applications. In theory we assume that target and background information (means, covariances, or subspaces) are known (optimum detectors); in practice, these quantities are estimated from available data (Estimate-and-Plug detectors). Given target and background measurements we can easily compute means, covariances, and subspaces; however, the quality of the estimates depends on the size and nature of the available data. Although Gaussian or ECD mixture models provide a better fit to background data, their use is prohibited because they are computationally expensive, the quality of the estimated parameters is questionable (same data are used to estimate much more parameters), and the computation of the likelihood ratio is quite involved. We next outline how these considerations and some additional practical limitations affect the selection and use of detection algorithms for practical applications.

6.1 Benchmark Detection Algorithms

In most practical hyperspectral detection applications, targets and backgrounds are described by means, covariances, and subspaces. Table 1 shows a taxonomy of detection algorithms based on the use of covariance or subspace models and on how the subspace models were obtained; the table is by no means exhaustive.

If we set aside how the target or background subspaces are determined, how we estimate the covariance matrix, and we bypass special cases of the quadratic detector, we end up with a few fundamental ("benchmark") algorithms, which are summarized in Table 2. The theoretical performance of these algorithms has been extensively studied for normal and elliptically contoured distribution.

Table 1. Detection algorithm taxonomy based on the available target and background information (covariance matrix or subspace).

	TARGET			
BACKGROUND	No Info	Data $\Rightarrow \boldsymbol{\Sigma}_t$ $\boldsymbol{\Sigma}_t = \gamma^2 \boldsymbol{\Sigma}_b$	Data $\Rightarrow \boldsymbol{S}$	MODTRAN $\Rightarrow \boldsymbol{S}$ Radiance Domain
$\boldsymbol{\Sigma}_b$	RX anomaly detector	Quadratic detector, MTMF,[16] FTMF,[15] CTM,[4] NMF,[4]	MF, AMF, ACE[2]	MF, AMF, ACE[2]
$\boldsymbol{\Sigma}_b = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T \Rightarrow \boldsymbol{B}$	SubSpace RX[28] $\tilde{\boldsymbol{\Sigma}}_b = \sum_{k=d}^{p} \lambda_k \boldsymbol{q}_k \boldsymbol{q}_k^T$		Subspace GLRT,[25] JSD[4] (mean and $\boldsymbol{Q}$) Use data from other collects	Subspace GLRT, Healy[32] (SVD) MSD[4] (mean and $\boldsymbol{Q}$)
Endmembers $\Rightarrow \boldsymbol{B}$			See[33]	

6.2 Reflectance and Radiance Domain Processing

The information-bearing signal in HSI applications, is the reflectance spectrum of each pixel in the data cube of the imaged area. The reflectance spectrum is a function of the physical properties and the structure of materials within the imaged pixel; in this sense, can be thought as a "spectral fingerprint" (signature) for these materials. However, the data cubes obtained by the HSI sensors consist of the radiance spectrum for each pixel. In principle, the radiance spectrum is

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms

Table 2. Benchmark hyperspectral detection algorithms.

Name	Formula	Target	Background
Mahalanobis Distance	$\boldsymbol{x}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{x},\ \ \boldsymbol{x} \leftarrow (\boldsymbol{x}-\boldsymbol{\mu}_b)$	Unspecified	$\boldsymbol{\Sigma}_b$
Quadratic Detector	$(\boldsymbol{x}-\boldsymbol{s})^T\boldsymbol{\Sigma}_b^{-1}(\boldsymbol{x}-\boldsymbol{s}) - \boldsymbol{x}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{x},\ \boldsymbol{s}=\boldsymbol{\mu}_t-\boldsymbol{\mu}_b$		
Matched Filter	$\boldsymbol{s}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{x}$	$\boldsymbol{s}$	$\boldsymbol{\mu}_b$,$\boldsymbol{\Sigma}_b$
ACE	$\dfrac{(\boldsymbol{s}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{x})^2}{(\boldsymbol{s}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{s})(\boldsymbol{x}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{x})}$	$\boldsymbol{s}$	$\boldsymbol{\mu}_b$,$\boldsymbol{\Sigma}_b$
Subspace-Target MF	$\boldsymbol{x}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{x}$	$\boldsymbol{S}$	$\boldsymbol{\mu}_b$,$\boldsymbol{\Sigma}_b$
Subspace-Target ACE	$\dfrac{\boldsymbol{x}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{x}}$	$\boldsymbol{S}$	$\boldsymbol{\mu}_b$,$\boldsymbol{\Sigma}_b$
GLRT (Target and background subspaces)	$\dfrac{\boldsymbol{x}^T\boldsymbol{P}_B^{\perp}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{P}_{SB}^{\perp}\boldsymbol{x}}$, Full-pixel $\Rightarrow \boldsymbol{P}_{BS}^{\perp}=\boldsymbol{P}_S^{\perp}$	$\boldsymbol{S}$	$\boldsymbol{B}$

obtained by "modifying" the solar spectrum by the reflectance spectrum of the pixel, the transmittance characteristics of the intervening atmosphere, and the response of the sensor. Besides bearing *different information*, radiance and reflectance spectra are measured in *different units*. Therefore, *all data (that is, target signatures and data cubes) required by a detection algorithm should be either in the "reflectance-domain" or in the "radiance-domain"*. The benchmark detection algorithms in Table 2 can be used in both domains by proper selection of the necessary quantities. The advantages and disadvantages of applying target detection algorithms in each domain is an area of active research and debate.

6.3 Target Modeling and Estimation

The target signature can be obtained from a spectral library. If several spectra, from a given target, are available, we have two choices: (a) Set the target signature $\boldsymbol{s}$ equal to the mean vector of the existing spectra, (b) Use an endmember selection techniques or the SVD to determine a low-dimensional target subspace $\boldsymbol{S}$. If the processing is done in the radiance domain, we can include "environmental" variability by using a forward physics-based model (usually MODTRAN) to generate all possible radiance target spectra and then represent them by a lower dimensional subspace that captures most variability.

6.4 Background Modeling and Estimation

Typically, the background clutter is described by a mean vector and a covariance matrix, which can be estimated form the available data locally, block-wise, or globally. The local estimation is done using a running window centered at the pixel of interest; pixels around the pixel of interest are ignored using a guard mask to avoid spoiling the covariance matrix. The determination of the background subspace $\boldsymbol{B}$ can be done by applying SVD[32] or endmember selection techniques to a set of background spectra from the data cube.

If we wish to take into consideration the non-homogeneity or multi-class nature of the training data, we can use an unsupervised classification algorithm to segment the background into spectrally homogeneous classes and then perform detection per each class. Alternatively, we can use a gaussian mixture model and the associated likelihood ratio detector. However, the extremely large number of unknown parameters cannot be reliably estimated form the available data. Although there is no definitive conclusion regarding this argument, preliminary results and experience from corresponding discrimination problems support the position that using all data to estimate a single covariance provides good performance.

Possible presence of targets in the background estimation data lead to the corruption of background covariance matrix by target spectra. This may lead to significant performance degradation; therefore, it is extremely important that, the estimation of $\boldsymbol{b}$ and $\boldsymbol{\Sigma}$ should be done using a set of "target-free" pixels that accurately characterize the background. Some approaches to attain this objective include: (a)run a detection algorithm, remove a set of pixels that score high, recompute the covariance with the remaining pixels, and "re-run" the detection algorithm, and (b) before computing the covariance, remove the pixels with high projections onto the target subspace.[32]

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms

6.5 Covariance Estimation, Inversion, and Regularization

Another important consideration is the number of pixels required for a reliable and numerically invertible estimate of the background covariance matrix. Certainly, the number of used pixels should exceed the number of bands; however, in practice, it should be much larger (typically, 10 to 100 times the number of bands) to obtain a statistically reliable estimate of the covariance matrix. To estimate and invert the covariance matrix of the background we combine the Dominant Mode Rejection (DMR) approach with regularization. The basic idea underlying all DMR techniques is to estimate only the large eigenvalues and corresponding eigenvectors of the covariance matrix $\mathbf{\Sigma}$. The advantage is that we can obtain better estimates of $\mathbf{\Sigma}$ with fewer spectra. Given the spectral decomposition of the estimated covariance matrix $\mathbf{\Sigma} = \sum_{k=1}^{p} \lambda_k \boldsymbol{q}_k \boldsymbol{q}_k^T$ the DMR-regularized inverse is given by

$$(\mathbf{\Sigma}_{\text{DMR}} + \delta \boldsymbol{I})^{-1} = \frac{1}{\alpha + \delta} \left[\boldsymbol{I} - \sum_{i=1}^{d} \left(\frac{\lambda_i - \alpha}{\lambda_i + \delta} \right) \boldsymbol{q}_i \boldsymbol{q}_i^T \right] \tag{64}$$

where $\alpha = \frac{1}{p-d} \sum_{i=d+1}^{p} \lambda_i$ is the average of the smaller $p-d$ eigenvalues of $\mathbf{\Sigma}$ and the loading factor δ introduces diagonal loading to the dominant modes. If we assume that $\lambda_i \gg \alpha$, for all $1 \leq i \leq d$, we obtain the Principal Component Inversion (PCI) approximation of the inverse covariance matrix

$$\tilde{\mathbf{\Sigma}}_{\text{PCI}}^{-1} = \frac{1}{\alpha} \left(\boldsymbol{I} - \sum_{i=1}^{d} \boldsymbol{q}_i \boldsymbol{q}_i^T \right) = \frac{1}{\alpha} \left(\boldsymbol{I} - \boldsymbol{Q}_1 \boldsymbol{Q}_1^T \right) \tag{65}$$

which provides a link between covariance-based[34] and subspace-based detection algorithms. Equation (64) is mostly used for the inversion of background covariance matrix; however, it can be used to invert a target covariance matrix, if available.

6.6 Detection Algorithm Robustness

The matched-signature type detectors assume a target with a perfectly *known* spectral signature. In practice, the target signature is either imperfectly measured (target mismatch) and/or it exhibits spectral variability. Robust matched filter algorithms[35] use covariance regularization to address the problems resulting from the uncertainty and/or variability of target signatures. Covariance regularization can be also interpreted using a Bayesian[36] framework, which was first introduced in the context of ridge regression. Robustness to other types of mismatch (e.g., covariance) have been not fully investigated.

6.7 Threshold Selection

Finally, an important issue, which is often ignored in theoretical and empirical investigations, is the selection of the detection threshold η. In automated surveillance systems, the threshold should be chosen to assure an operationally useful constant false alarm rate (CFAR). In a user-in-the-loop approach, interactive threshold selection through visual assessment may be useful. However, objective threshold selection requires accurate modeling of the background statistics. A practical approach to CFAR threshold selection[35,37] is based on the theory of extreme value distributions.

6.8 Detection Performance Evaluation

A ROC curve plots the P_{D} versus the P_{FA} as a function of the threshold η. Theoretical ROC curves can be obtained by making assumptions about the probability density functions under the two hypotheses. In practice, probabilities have to be estimated from real data. Their accuracy depends on the number of available target and background pixels. Typically, a reliable estimate of a probability P requires at least $10/P$ pixels.[38] Since the number of target pixels is often very limited, the estimate of P_{D} is not accurate. This implies that, ROC curves should be used in empirical detection evaluation with extreme care. The lack of widely accessible data from well designed target detection experiments with accurate ground truth makes the experimental evaluation and comparison of detection algorithms extremely difficult. Clearly, this has led to a proliferation of claims regarding the discovery of "better" target detectors.

6.9 Some Operational Considerations

For a fully fair comparison of detection algorithms, we should take the position of an operational user and consider how the use of each detector in a practical application affects its performance and utility. Two very important practical considerations are: good performance on a diverse variety of targets and backgrounds, and operational ease of use by non-experts. Furthermore, robustness to the selection of "free-parameters" and operation without "man-in-the-loop" are also critical for surveillance applications.

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms

7. SUMMARY

This paper is a first attempt to answer the question: Is there a best hyperspectral detection algorithm? To provide a definitive answer, we need well-truthed data sets for a diverse range of targets and backgrounds under various scenarios, target fill-factors, and atmospheric conditions. Statistically significant numbers of target and background pixels are necessary for the construction of reliable ROC curves. Since we may never reach this point, a reasonable approach is to use detection algorithms with good and well-understood theoretical properties. It is our experience that if we take sufficient care to understand the essential properties of matched filter and ACE detectors and apply them properly, we can obtain good performance. Furthermore, it turns out that, any small performance gains attained by more sophisticated detectors are irrelevant in practical applications because of the limitations and the uncertainties about many aspects of the situation in which the detector will be deployed.

REFERENCES

[1] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging application," *IEEE Signal Processing Magazine* , pp. 29–43, January 2002.

[2] D. Manolakis, D. Marden, and G. Shaw, "Target detection algorithms for hyperspectral imaging application," *Lincoln Laboratory Journal* **14**(1), pp. 79–116, 2003.

[3] D. Manolakis, "Taxonomy of detection algorithms for hyperspectral imaging applications," *Optical Engineering* **44**(6), p. 066403, 2005.

[4] A. Schaum, "Hyperspectral detection algorithms: from old ideas to operational concepts to next generation," *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII* **6233**(1), p. 623305, SPIE, 2006.

[5] J. Theiler, "Quantitative comparison of quadratic covariance-based anomalous change detectors," *Appl. Opt.* **47**(28), pp. F12–F26, 2008.

[6] M. T. Eismann, J. Meola, A. D. Stocker, S. G. Beaven, and A. P. Schaum, "Airborne hyperspectral detection of small changes," *Appl. Opt.* **47**(28), pp. F27–F45, 2008.

[7] S. Adler-Golden, Michael, Matthew, L. Bernstein, R. Levine, A. Berk, S. Richtsmeier, P. Acharya, G. Anderson, J. W. Felde, J. Gardner, M. Hoke, L. Jeong, B. Pukall, A. Ratkowski, and H. Hua K. Burke, "Atmospheric correction for short-wave spectral imagery based on MODTRAN4," **3753**, SPIE, (Bellingham, WA), 1999.

[8] J. Cipar, R. Lockwood, T. Cooley, and P. Grigsby, "Background spectral library for Fort A.P. Hill, Virginia," in *Remote Sensing and Modeling of Ecosystems for Sustainability*, D. R. Shaw and W. Gao, eds., pp. 35–46, SPIE, 11 2004.

[9] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York, NY, 3rd ed., 2003.

[10] K. T. Fang, S. Kotz, and K. W. Ng, *Symmetric Multivariate and Related Distributions*, Chapman and Hall, New York, 1990.

[11] J. B. Adams, M. O. Smith, and A. R. Gillespie, "Imaging spectroscopy: Interpretation based on spectral mixture analysis," in *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, C. M. Pieters and P. A. J. Englert, eds., pp. 145–166, Cambridge University Press, (Cambridge), 1993.

[12] S. M. Kay, *Fundamentals of Statistical Signal Processing*, Prentice Hall, New Jersey, 1998.

[13] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Computer Science and Scientific Computing, Academic Press, New York, second ed., 1990.

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

[15] A. Schaum and A. Stocker, "Spectrally-selective target detection," *Proceedings of ISSSR* , 1997.

[16] J. W. Boardman, "Leveraging the high dimensionality of aviris data for improved sub-pixel target un- mixing and rejection of false positives: mixture tuned matched filtering," in *Summaries of the Seventh Annual JPL Airborne Geoscience Workshop*, R. O. Green, ed., **1**, p. 55, JPL Publication 97-21, (Pasadena, California), 1998.

[17] E. J. Ientilucci and J. R. Schott, "Target detection in a structured background environment using an infeasibility metric in an invariant space," *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI* **5806**(1), pp. 491–502, SPIE, 2005.

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms

[18] S. Kraut, L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Transactions on Signal Processing* **49**, pp. 1–16, January 2001.

[19] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. on Aerospace and Electronic Systems* **22**, pp. 115–127, March 1986.

[20] E. J. Kelly, "Adaptive detection in non-stationary interference, part III," Tech. Rep. 761, MIT Lincoln Laboratory, Lexington, MA, 1987.

[21] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. on Aerospace and Electronic Systems* **28**, pp. 208–218, January 1992.

[22] W. Chen and I. S. Reed, "A new CFAR detection test for radar," *Digital Signal Processing* **1**(4), pp. 198–214, 1991.

[23] E. J. Kelly and K. M. Forsythe, "Adaptive detection and parameter estimation for multidimensional signal models," Tech. Rep. 848, MIT Lincoln Laboratory, Lexington, MA, April 1989.

[24] S. Kraut and L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Trans. Signal Processing* **47**, pp. 2538–2541, Sept. 1999.

[25] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*, McGraw-Hill Companies, Inc., Boston, 2000.

[26] J. C. Harsanyi and C. I. Chang, "Detection of low probability subpixel targets in hyperspectral image sequences with unknown backgrounds," *IEEE Trans. Geoscience and Remote Sensing* **32**, pp. 779–785, July 1994.

[27] A. P. Schaum, "Spectral subspace matched filtering," *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VII* **4381**(1), pp. 1–17, SPIE, 2001.

[28] A. P. Schaum, "Hyperspectral anomaly detection beyond RX," *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIII* **6565**(1), p. 656502, SPIE, 2007.

[29] C. D. Richmond, *Adaptive Array Signal Processing and Performance Analysis in Non-Gaussian Envirnoments*. Doctor of philosophy, Massachusetts Institute of Technology, Cambridge, MA, May 1996.

[30] A. P. Schaum and A. D. Stocker, "Joint hyperspectral subspace detection derived from a bayesian likelihood ratio test," *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII* **4725**(1), pp. 225–233, SPIE, 2002.

[31] J. Theiler, B. R. Foy, and A. M. Fraser, "Beyond the adaptive matched filter: nonlinear detectors for weak signals in high-dimensional clutter," *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIII* **6565**(1), p. 656503, SPIE, 2007.

[32] B. Thai and G. Healey, "Invariant subpixel material detection in hyperspectral imagery," *IEEE Trans. on Geosci. and Rem. Sensing* **40**, pp. 599–607, March 2002.

[33] P. Bajorski, E. J. Ientilucci, and J. R. Schott, "Comparison of basis-vector selection methods for target and background subspaces as applied to subpixel target detection," *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X* **5425**(1), pp. 97–108, SPIE, 2004.

[34] T. C. D. Manolakis, R. Lockwood and J. Jacobson, "Hyperspectral detection algorithms: Use covariances or subspaces?," in *SPIE Annual Meeting*, (San Diego, CA), 2009.

[35] D. Manolakis, D. Zhang, M. Rossacci, R. Lockwood, T. Cooley, and J. Jacobson, "Maintaining cfar operation in hyperspectral target detection using extreme value distributions," *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIII* **6565**(1), p. 65651W, SPIE, 2007.

[36] N. Nasrabadi, "Regularized spectral matched filter for target recognition in hyperspectral imagery," *Signal Processing Letters, IEEE* **15**, pp. 317–320, 2008.

[37] E. Ensafi and A. D. Stocker, "An adaptive cfar algorithm for real-time hyperspectral target detection," *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIV* **6966**(1), p. 696605, SPIE, 2008.

[38] J. D. Echard, "Estimation of radar detection and false alarm probabilities," *IEEE Trans. on AES* **27**, pp. 255–260, March 1991.

ACKNOWLEDGMENTS

This work was sponsored by the Department of Defense under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the United States Government.

Downloaded from SPIE Digital Library on 15 Mar 2010 to 18.51.1.125. Terms of Use: http://spiedl.org/terms