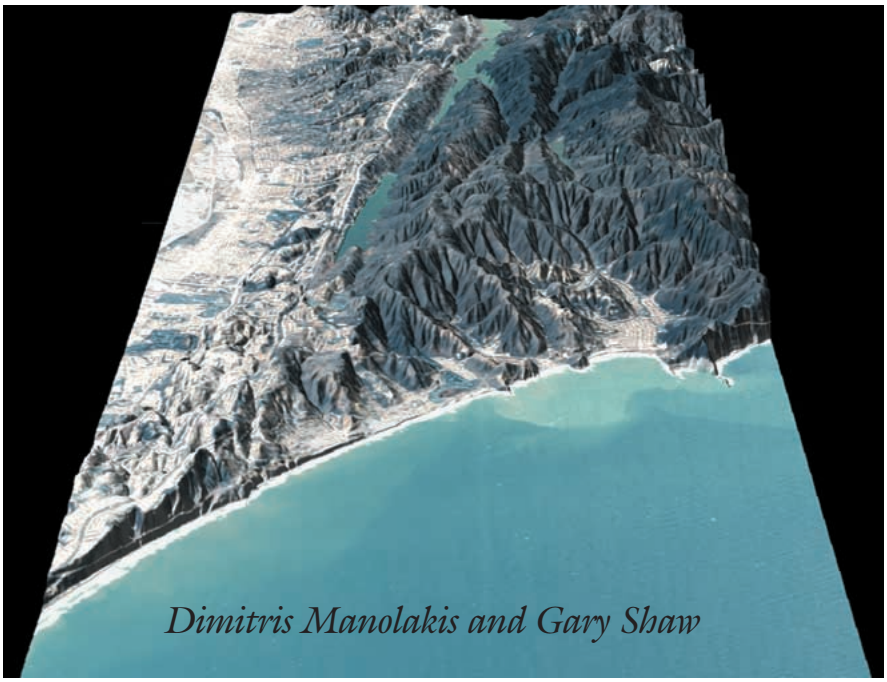
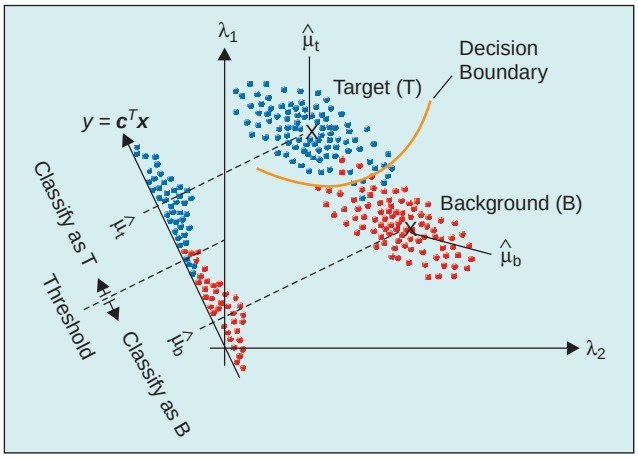


Detection Algorithms for Hyperspectral Imaging Applications

Hyperspectral remote sensing exploits the fact that all materials reflect, absorb, and emit electromagnetic energy, at specific wavelengths, in distinctive patterns related to their molecular composition. Hyperspectral imaging (HSI) sensors in the reflective region of the spectrum (sometimes referred to as imaging spectrometers) acquire digital images in many contiguous and very narrow (nominally about 0.010 μm wide) spectral bands that typically span the visible, near-infrared, and mid-infrared portions of the spectrum (0.4-2.5 μm). This enables the construction of an essentially continuous radiance spectrum for every pixel in the scene. Thus, HSI data exploitation makes possible the remote identification of ground materials-of-interest based on their spectral signatures.

Hyperspectral target detection algorithms can be developed using statistical, physical, or heuristic approaches. Irrespective of the design approach, however, performance evaluation of any detector should be done using statistical criteria. In this article, we focus on the design and evaluation of detection algorithms using statistical hypothesis testing techniques. Statistical inference techniques for normal distributions have been widely studied and used because they are mathematically tractable and provide good performance in many practical situations. We therefore focus on detection algorithms that assume multivariate normal distribution models for HSI data. Although HSI data frequently violate the assumption of normality used to derive the detectors, these algorithms work well in many practical applications, their operation and performance can be understood theoretically, and they can provide the ground for the development of algorithms for nonnormally distributed HSI data. The theoretical development and analysis of detection algorithms

1053-5888/02/$17.00©2002IEEE



▲ *1. Illustration of two-class classification using two spectral bands.*

for arbitrary nonnormal distributions, although highly desirable, it is not mathematically tractable.

This article is organized as follows. First we introduce key concepts and issues including the effects of atmospheric propagation upon the data, spectral variability, mixed pixels, and the distinction between classification and detection algorithms. Detection algorithms for full pixel targets are developed in the following section, using the likelihood ratio approach. Subpixel target detection, which is more challenging due to background interference, is pursued next using both statistical and subspace models for the description of spectral variability. Finally, in the concluding section, we provide some results which illustrate the performance of some detection algorithms using real HSI data. Furthermore, we illustrate the potential deviation of HSI data from normality and point to some distributions that may serve in the development of algorithms with better or more robust performance.

Introduction

As a result of their fine spectral resolution, HSI sensors provide a significant amount of information about the physical and chemical composition of the materials occupying the pixel surface, as well as the characteristics of the atmosphere between the sensor and the surface during the data collection.

Atmospheric Compensation

Due to the effects of the illumination source and the atmosphere, the "raw" radiance spectra obtained by an HSI sensor cannot be directly compared to either laboratory spectra or "raw" spectra collected at other times or places.

To overcome this obstacle, we work with the reflectance spectrum, which indicates the portion of incident energy which is reflected as a function of wavelength. Hence, the properties of the illuminating source and the effects of the propagating atmosphere are removed, and the shape of the reflectance curve is characteristic of the materials in the observed pixel. Once the data have been corrected for the effects of the atmospheric absorption and scattering, the resulting reflectance spectrum for every pixel, can be compared to spectra of known materials available in "spectral libraries."

Spectral Variability and Mixed Pixels

The basic task underlying many HSI applications is to identify different materials based on their reflectance spectrum. In this respect, the concept of a spectral signature, which uniquely characterizes any given material, is highly attractive and widely used. However, spectra observed in the natural world do not exhibit a deterministic signature. The spectra observed from samples of the same material will never be identical, even in laboratory experiments, due to variations in the material surface. The amount of variability is more profound in remote sensing applications due to variations in atmospheric conditions, sensor noise, material composition, location, surrounding materials, and other factors. To make matters worse, totally different material types can have very similar spectra. An additional source of spectral variability are calibration and illumination variations which are not currently handled by atmospheric correction codes.

Despite these difficulties, practical experience has shown that many materials of interest can be identified on the basis of their spectral characteristics. However, the ambiguity introduced by inherent variability of spectral signatures has important implications into the exploitation of HSI data for both civilian and military applications. Despite the intrinsic spectral variability and the occasional lack of identifiability, the concept of spectral signature is widely used in remote sensing spectroscopy. In this article, we assume that different materials are spectrally separable and focus on the problems introduced by the inherent variability of spectral signatures.

Another significant complication arises from the interplay between the spatial resolution of the sensor and the spatial variability present in the ground scene. The sensor integrates the radiance from all materials within the ground surface "seen" by the sensor as an image pixel. Therefore, depending on the spatial resolution of the sensor and the spatial distribution of surface materials within each ground pixel, the result is a HSI data set comprised of "pure" and "mixed" pixels. Mixed pixels present an additional challenge to HSI data exploitation because their spectral signatures do not correspond to any single well-defined material.

Dealing with spectral signature variability and spectral compositions in mixed pixels are among the most challenging problems in HSI data exploitation, both theoretically and practically.

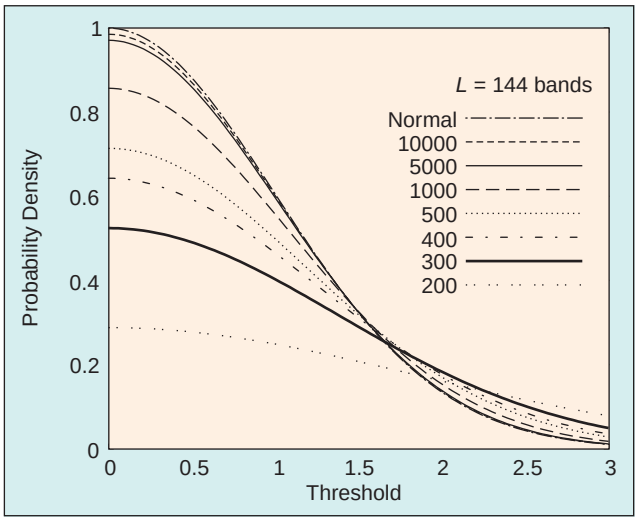
Classification and Target Detection

There are two major applications that rely upon the ability to separate materials based on their spectral signatures: classification and target detection.

The main objective of background classification is to automatically assign all pixels in an HSI data cube into land cover classes or themes, which has led to the term thematic mapping. The user has the task to up-front determine the number and type of classes as well as to quantitatively characterize these classes using spectral libraries or training data and ground truth information. Practical experience has shown that the design of a good classifier requires a sufficient amount of training data for each background class. Clearly, for background classification, the natural criterion of performance is the minimization of the probability of missclassification errors.

In target detection applications, the main objective is to search the pixels of an HSI data cube for the presence of a specific material (target). Conceptually, at least at a theoretical level, target detection can be viewed as a binary hypothesis testing problem, where each pixel is assigned a target or nontarget label. However, there are some fundamental practical differences that have a great impact upon the development and evaluation of practical algorithms for detection versus classification applications. In surveillance applications, the size of the objects (targets) we are searching for constitutes a very small fraction of the total search area. Therefore, the target class will be either empty or sparsely populated. On the other hand, the general "no-target" class includes almost all pixels in the cube and is the union of the different specific background classes. We shall use the term "background" to refer to all nontarget pixels of a scene. Usually, targets are man-made objects with spectra that differ from the spectra of natural background pixels.

The sparseness of the target class implies that there are not sufficient data to train a statistical classifier or statistically evaluate the performance of a target detector. On the other hand, the heavy population of the background class, in conjunction with the emptiness of the target class, allows the use of the "unclassified" HSI cube to statistically characterize the background. In detection applications, where the target probability is very small, minimization of the error probability is not a good criterion of performance, because it can be minimized by classifying every pixel as background. For this reason, we typically seek to maximize the probability of detection while keeping the probability of false alarm under a certain predefined value (Neyman-Pearson criterion [1]).



▲ *2. Probability density function of $p_y(y)$ for various values of the number N_b of the background pixels. The threshold is measured from the ideal value y_o in units of $(\mathbf{s}^T\boldsymbol{\Gamma}_b^{-1}\mathbf{s})^{-1/2}$.*

The basic task underlying many hyperspectral imaging applications is to identify different materials based on their reflectance spectrum.

The amount of a priori information about the spectral signature of the target depends on the requirements of the specific application. A priori information about spectral signatures is available in libraries as reflectance spectra. Therefore, to look for targets with known spectral signatures, the data must first be converted into reflectance, a procedure that may lead to spectral distortions since it generally depends upon assumptions and measurements about atmospheric conditions. If we have no a priori information about the target or we wish to work with radiance, the most reasonable approach is to look for pixels whose spectral content is "significantly" different from the spectral content of the local background. The detection of pixels whose spectral signatures differ from those of the background is known in hyperspectral literature as anomaly detection and is discussed in a companion article.

Design and Evaluation of Target Detectors

Most HSI data processing techniques start with the idea that an observed spectrum can be considered as a vector in a multidimensional space, where the number of dimensions equals the number of spectral bands, L. Taking into consideration spectral variability and receiver noise, the observations provided by the sensor can be modeled, for the purpose of theoretical analysis, as random vectors with specific probability distributions. Given an observed spectrum, $\boldsymbol{x}$, the likelihood ratio (LR) is given by the ratio of the conditional probability density functions

$$\Lambda(\boldsymbol{x}) \triangleq \frac{p(\boldsymbol{x}|\,\text{signal present})}{p(\boldsymbol{x}|\,\text{signal absent})}. \tag{1}$$

If $\Lambda(\boldsymbol{x})$ is larger than the threshold η, the "signal present" hypothesis is accepted. Basically, the LR test accepts as "true" the most "likely" hypothesis.

A practical question of paramount importance to a detection algorithm user is where to set the threshold to keep the number of detection errors (target misses and false alarms) small. Indeed, there is always a compromise between choosing a low threshold to increase the proba-

Hyperspectral remote sensing exploits the fact that all materials reflect, absorb, and emit electromagnetic energy, at specific wavelengths, in distinctive patterns related to their molecular composition.

bility of (target) detection P_D and a high threshold to keep the probability of false alarm P_{FA} low. For any given detector, the trade-off between P_D and P_{FA} is described by the receiver operating characteristic (ROC) curves, which plot $P_D(\eta)$ versus $P_{FA}(\eta)$ as a function of threshold $-\infty < \eta < \infty$. Clearly, any systematic procedure to determine ROC curves or the threshold requires specifying the distribution of the observed spectra $\boldsymbol{x}$ under each of the two hypotheses.

In most practical situations, the conditional probability densities in (1) depend on some unknown target and background parameters (composite hypotheses). Therefore, the ROC curves of any detector depend on the unknown parameters. In this case, it is almost impossible to find a detector whose ROC curves remain an upper bound for the whole range of the unknown parameters (uniformly most powerful (UMP) detector).

An intuitively appealing and widely used approach, in the case of unknown density parameters, is to replace the unknown parameters in the LR (1) with their maximum likelihood estimates. In general, there are no optimality properties associated with the resulting generalized LR (GLR), $\Lambda_G(\boldsymbol{x})$. In practice, however, the GLR leads to detectors that seem to work well in several applications.

Practical target detection systems should function automatically, that is, without operator intervention. This requires an automatic strategy to set a "proper" detection threshold. A high false alarm rate wastes processing and reporting resources and may result to system overloading. Therefore it is critical to keep the false alarm rate constant at a desirable level by using a constant false alarm rate (CFAR) processor. The task of a CFAR algorithm is to provide detection thresholds that are relatively immune to noise and background variation and allow target detection with a constant false alarm rate.

Framework for Detection Algorithm Taxonomy

The key factors that determine the taxonomy of hyperspectral target detection algorithms are: the type of models used for spectral (target or background) variability, the composition of the pixel under test (pure or mixed), and the model used to describe mixed pixels.

There are two widely used ways to describe spectral variability. The geometric approach restricts the spectrum vector to vary in an M-dimensional subspace of the data space ($M < L$). The observed spectrum is described by

$$\boldsymbol{x} = \sum_{k=1}^{M} \alpha_k \boldsymbol{s}_k = \boldsymbol{S}\boldsymbol{\alpha}. \tag{2}$$

The vectors $\boldsymbol{s}_k$ or equivalently the matrix $\boldsymbol{S}$, which define the variability subspace, can be a) unique spectral signatures 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 L. The statistical approach requires 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.

For full pixel targets, there is no significant interaction between target and background other than secondary illumination, shading, etc. Hence, the spectrum, $\boldsymbol{x}$, observed by the sensor is produced either by the target spectrum $\boldsymbol{s}_t$ or the background spectrum $\boldsymbol{s}_b$. In both cases, the observed spectrum is corrupted by additive sensor noise $\boldsymbol{w}$.

For subpixel targets, both the spectrum of the target and the spectrum or spectra of the background contribute to the observed mixed pixel spectrum. There are two widely used models for modeling subpixel targets.

The most widely used spectral mixing model is the linear mixing model [1] (LMM), which assumes that the observed pixel spectrum is generated by a linear combination of a small number of unique constituent deterministic spectral signatures known as "endmembers." The mathematical formulation of the LMM is given by

$$\boldsymbol{x} = \sum_{k=1}^{M} \alpha_k \boldsymbol{s}_k + \boldsymbol{w} = \boldsymbol{S}\boldsymbol{\alpha} + \boldsymbol{w} \tag{3}$$

where $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_M$, are the M endmember spectra, which must be linearly independent, $\alpha_1, \alpha_2, \ldots, \alpha_M$ are the corresponding abundances, and $\boldsymbol{w}$ is an additive noise vector. Endmembers may be obtained from spectral libraries, in-scene spectra, or using geometrical techniques. We point out that the enforcement of positivity ($\alpha_k \geq 0$) and additivity ($\alpha_1 + \alpha_2 + \cdots + \alpha_M = 1$) constraints makes the LMM a replacement model.

If the endmember spectra are randomly and independently drawn from multivariate normal distributions, we have the stochastic mixing model [3], [4].

The choice of a pixel composition assumption (pure or mixed pixel), the selection of a model to account for spectral variability (subspace or probability distribution), and the selection of a mixing procedure leads to different types of target detection algorithms. The detection prob-

lem is typically formulated as a binary hypothesis test with two competing hypotheses: background only (H_0) or target and background (H_1). Since the two hypotheses contain unknown parameters (for example, covariance matrix of the background) that have to be estimated from the data, the detector has to be adaptive, and it is usually designed using the generalized likelihood ratio test (GLRT) approach [1].

Most detection algorithms for full pixel and subpixel targets have been obtained by describing spectral variability using the multivariate normal distribution or a subspace model. Mixed pixels are usually modeled using the LMM. A target detection algorithm based on the stochastic mixing model, known as finite target matched filter, is discussed in [3] and [4].

Finally, we note that in several practical applications, we do not have adequate a priori information about the desired target. In such cases, it is possible to design algorithms that look for spectra which deviate from the local background (anomaly detection). The type of the statistical model used for the background leads to different anomaly detection algorithms. Use of a multivariate normal distribution model leads to the RX algorithm [5], [6] which is often used for anomaly detection. Recently, a new algorithm [7] has been developed, which fuses the local statistics used by the RX algorithm and the clustering statistics obtained using stochastic expectation maximization (SEM) to improve detection performance. More details about anomaly detection algorithms can be found in [8].

Detection Algorithms for Full-Pixel Targets

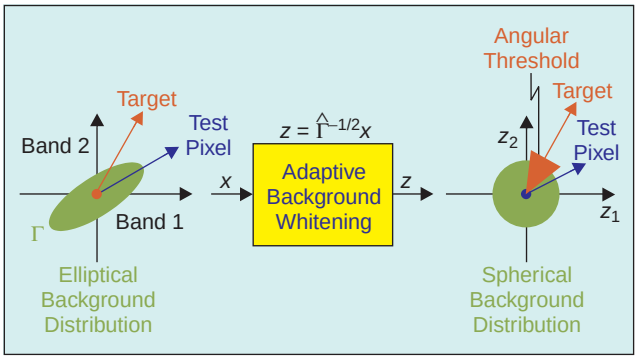
In this section, we assume that target pixels are completely filled by the material-of-interest, that is, we focus attention on full-pixel or resolved targets. In this case, the detection process is complicated by the spectral variabilities of the target and background classes. We can think of the totality of target spectra as constituting a target class and those from the background as being the background class. Let R be the entire L-dimensional space in which the point of L-band spectrum $\mathbf{x}$ falls. To make a decision, we should divide the region R into two regions, R_t and R_b, by some optimum method. A pixel is assigned to the target class if its spectrum $\mathbf{x}$ falls in region R_t or to the background class if $\mathbf{x}$ falls in R_t. We shall pictorially illustrate the various concepts and algorithms using a hypothetical sensor with two spectral bands. Due to the geometrical framework, however, the results and their interpretation hold for spectra produced by HSI sensors with a much larger number of bands. This process is illustrated in Fig. 1 for $L=2$ bands. Clearly, meaningful decision making is possible if the observed target spectra differ to some extent from the observed background spectra. Usually, the two classes overlap and even the best detector will result in misclassification errors. In general, the decision boundary will be a curve corresponding to a nonlinear detector. We can also make a decision by processing the spectrum vector $\boldsymbol{x}$ by a system which calculates a scalar $y=D(\boldsymbol{x})$ and then comparing y to a scalar threshold. Usually, the function $D(\boldsymbol{x})$ is obtained using the LR or GLR approaches. This system, which can be linear or nonlinear, is known as a two-class classifier, discriminant function, statistic, filter, or detector. We shall interchangeably use the terms filter or detector since they are widely used in the engineering literature. We discuss two approaches. First, we shall use the LR to obtain detectors without any structural constraints, that is, detectors with arbitrary decision surfaces in R^L. Second, we focus on the design of detectors with hyperplane decision surfaces. These linear detectors project the data onto a line specified by their coefficient vector with the objective of increasing class separation.

Likelihood Ratio Detectors

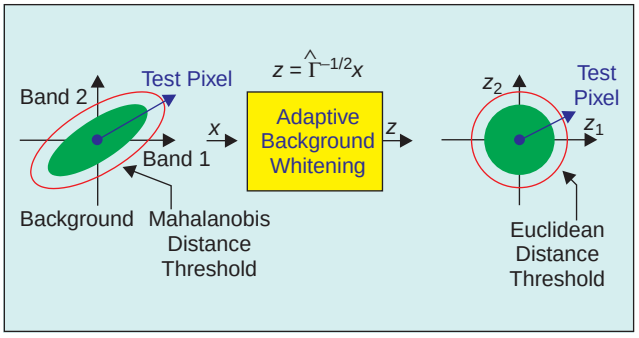
Since statistical decision procedures, based on normal probability models, are simple and often lead to good performance, we shall model the target and background spectra as multivariate normal vectors. A random vector $\boldsymbol{x}$ follows a multivariate normal distribution with mean vector $\boldsymbol{\mu} \triangleq E\{\boldsymbol{x}\}$ and covariance matrix $\boldsymbol{\Gamma} \triangleq E\{(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T\}$, denoted by $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Gamma})$, if its probability density function is given by

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{L/2}|\boldsymbol{\Gamma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Gamma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \tag{4}$$

where $|\boldsymbol{\Gamma}|$ represents the determinant of matrix $\boldsymbol{\Gamma}$.



▲ *3. Illustration of GLRT detectors.*

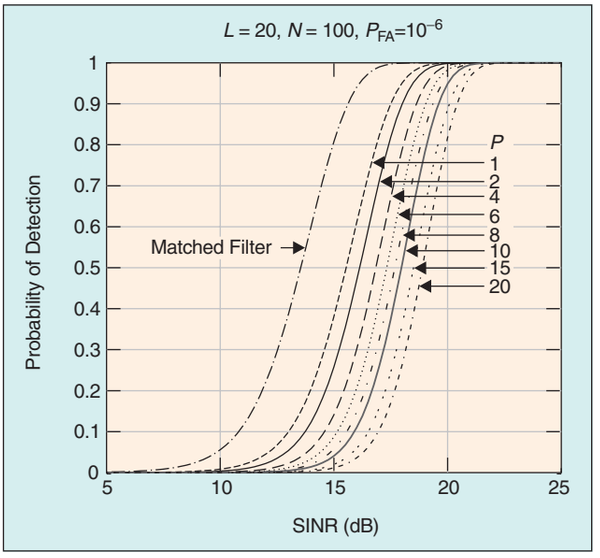


▲ *4. Illustration of unknown target or anomaly detection.*

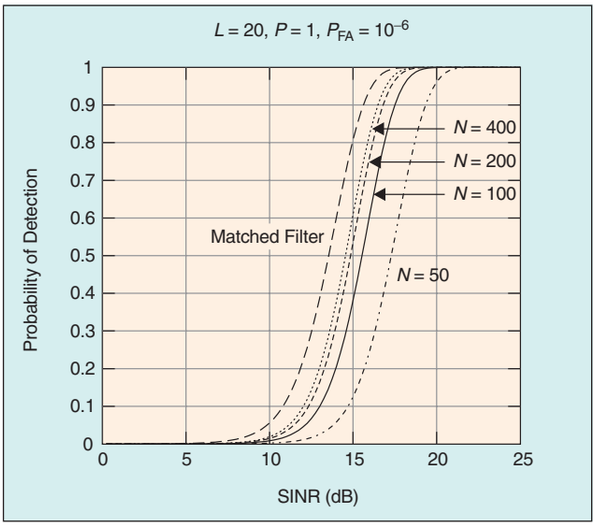
Consider the detection problem specified by the following hypotheses:

$$H_0:\ \boldsymbol{x} \sim N(\boldsymbol{\mu}_b, \boldsymbol{\Gamma}_b) \quad \text{Target absent}$$
$$H_1:\ \boldsymbol{x} \sim N(\boldsymbol{\mu}_t, \boldsymbol{\Gamma}_t) \quad \text{Target present} \tag{5}$$

where the target and background classes follow multivariate normal distributions with different mean vectors and covariance matrices. Since the probability densities are completely specified under each hypothesis, we can design a Neyman-Pearson detector. Indeed, computing the natural logarithm of the LR (1) leads to the quadratic detector



▲ *5. Probability of detection as a function of SINR illustrating the effect of target subspace dimensionality for Kelly's GLRT detector.*



▲ *6. Probability of detection as a function of SINR illustrating the effect of the number of training pixels for Kelly's GLRT detector.*

$$y = D(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_b)^T \boldsymbol{\Gamma}_b^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_b) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_t)^T \boldsymbol{\Gamma}_t^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_t) \tag{6}$$

which compares the Mahalanobis distances of the observed spectrum from the centers of the two classes. The required threshold η is determined from

$$P_{FA} = \int_{\eta}^{\infty} p(y|H_0)\,dy = \alpha \tag{7}$$

where α is the desired probability of false alarm. As a result of the quadratic mapping, the distribution of the random variable y (detector output) is not normal, which makes the performance evaluation of the detector difficult.

If the target and background classes have the same covariance matrix, that is, $\boldsymbol{\Gamma}_t = \boldsymbol{\Gamma}_b \triangleq \boldsymbol{\Gamma}$, the quadratic terms in (5) disappear, and the likelihood ratio detector (5) becomes

$$y = D(\boldsymbol{x}) = (\boldsymbol{\mu}_t - \boldsymbol{\mu}_b)^T \boldsymbol{\Gamma}^{-1} \boldsymbol{x}. \tag{8}$$

In this, highly unlikely case for HSI data, we have a linear detector

$$y = \boldsymbol{c}^T \boldsymbol{x} = \sum_{k=1}^{L} c_k x_k \tag{9}$$

which is specified by the coefficient vector

$$\boldsymbol{c} = \boldsymbol{\Gamma}^{-1}(\boldsymbol{\mu}_t - \boldsymbol{\mu}_b). \tag{10}$$

The output y is now normally distributed because it is a linear combination of normal random variables. This result simplifies the evaluation of the detector and the computation of detection thresholds using (7).

This detector, which is known as Fisher's linear discriminant [15], is widely used in pattern recognition applications. We shall use the term matched filter (MF), which is more widely used in the communications and signal processing literature. There, the matched filter (10) is usually derived by maximizing the cost function

$$J(\boldsymbol{c}) \triangleq \frac{[E\{y|H_1\} - E\{y|H_0\}]^2}{\text{var}\{y|H_0\}} = \frac{[\boldsymbol{c}^T(\boldsymbol{\mu}_t - \boldsymbol{\mu}_b)]^2}{\boldsymbol{c}^T \boldsymbol{\Gamma} \boldsymbol{c}} \tag{11}$$

which measures the distance between the means of two normal distributions in units of the common variance. The maximum, which is obtained by substituting (10) into (11), is

$$J_{\max} = \Delta^2 \triangleq (\boldsymbol{\mu}_t - \boldsymbol{\mu}_b)^T \boldsymbol{\Gamma}^{-1}(\boldsymbol{\mu}_t - \boldsymbol{\mu}_b) \tag{12}$$

which is the Mahalanobis squared distance [10] between the means of the target and background distributions.

The performance of the matched filter is determined by its direction, which specifies the line where the input

data vectors are projected to obtain the detection statistics. Therefore, any filter of the form

$$\boldsymbol{c}_{\mathrm{MF}} = \kappa \boldsymbol{\Gamma}^{-1}(\boldsymbol{\mu}_{\mathrm{t}} - \boldsymbol{\mu}_{\mathrm{b}}) \tag{13}$$

where κ is a normalization constant, will provide the same performance. It can be shown that the filter (13) with $\kappa = 1/\Delta^2$ minimizes the output variance $\boldsymbol{c}^T\boldsymbol{\Gamma}\boldsymbol{c}$ subject to the linear constraint $\boldsymbol{c}^T\boldsymbol{\mu}_{\mathrm{t}} = 1$. In the context of array processing, this processor is known as minimum variance beamformer [11]. A similar detector, termed the constrained energy minimization (CEM) algorithm [12], can be obtained by minimizing the total energy of the filter output under the same constraint.

Adaptive Matched Filters

The matched filter detector (13) requires the mean vector and the common covariance matrix of the target and background distributions. Furthermore, the resulting detector is optimum (in the Bayes or Neyman-Pearson sense) only when the target and background classes follow multivariate normal distributions with the same covariance matrix, an unlikely situation for real-world HSI data. In practical applications, these quantities are unavailable and have to be estimated from the available data. Under the assumption of low-probability targets, we can use the available data $\boldsymbol{x}(n)$, $n = 1,2,\ldots,N$, to determine the maximum likelihood estimates

$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}(n) \simeq \hat{\boldsymbol{\mu}}_{\mathrm{b}} \tag{14}$$

$$\hat{\boldsymbol{\Gamma}} = \frac{1}{N}\sum_{n=1}^{N}\left[\boldsymbol{x}(n) - \hat{\boldsymbol{\mu}}\right]\left[\boldsymbol{x}(n) - \hat{\boldsymbol{\mu}}\right]^T \simeq \hat{\boldsymbol{\Gamma}}_{\mathrm{b}} \tag{15}$$

of the mean vector and covariance matrix of the background. Unfortunately, there is usually not sufficient training data to determine the mean and covariance of the target. Typically, we use a target spectral signature $\boldsymbol{s}_{\mathrm{t}}$ from a library or the mean of a small number of known target pixels observed under the same conditions. The resulting adaptive matched filter (AMF) is given by

$$y = \frac{\boldsymbol{s}^T\hat{\boldsymbol{\Gamma}}_{\mathrm{b}}^{-1}\boldsymbol{x}}{\boldsymbol{s}^T\hat{\boldsymbol{\Gamma}}_{\mathrm{b}}^{-1}\boldsymbol{s}} \tag{16}$$

where usually the data cube mean is removed from the target and test pixel spectra.

If we know the "true" covariance matrix $\boldsymbol{\Gamma}_{\mathrm{b}}$, the output y under the "target absent" hypothesis, is distributed as $y \sim N(y_{\mathrm{o}}, (\boldsymbol{s}^T\boldsymbol{\Gamma}_{\mathrm{b}}^{-1}\boldsymbol{s})^{-1})$, where $y_{\mathrm{o}} = E\{y\}$. When the required means and covariances are estimated from the data, the resulting estimates are random quantities. If we treat them as constant, we can determine the class-conditional distribution of the detector output as in the known statistics case. However, the correct approach

The basic task underlying many HSI applications is to identify different materials based on their reflectance spectrum.

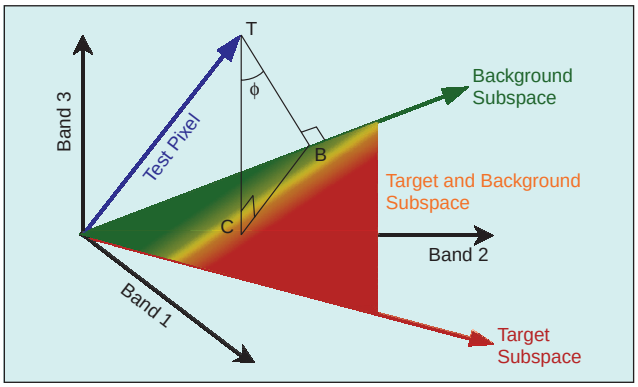
is to treat the estimated means and covariances as random and determine the unconditional distribution of $y = D(\boldsymbol{x})$. Unfortunately, the derivation of unconditional distributions is a very difficult problem, even under the most simplified assumptions. An extremely complicated expression for Fisher's linear discriminant has been obtained by Sitgreaves [13].

The distribution $p_y(y)$, when we use the sample covariance matrix $\hat{\boldsymbol{\Gamma}}_{\mathrm{b}}$, has been determined by Richmond in [14]. It depends upon the number L of spectral bands and the number N_{b} of pixels used to estimate the covariance matrix of the background. The background pixels are assumed to come from a multivariate normal distribution $\boldsymbol{x} \sim N(\boldsymbol{\mu}_{\mathrm{b}}, \boldsymbol{\Gamma}_{\mathrm{b}})$. Fig. 2 shows $p_y(y)$ for $L = 144$ spectral bands for various values of N_{b}. We see that, as the number of pixels used to estimate the covariance matrix of the background increases, the distribution approaches the normal distribution curve (see [15] for more details). This is expected because as $N_{\mathrm{b}} \to \infty$, the sample covariance matrix $\hat{\boldsymbol{\Gamma}}_{\mathrm{b}} \to \boldsymbol{\Gamma}_{\mathrm{b}}$, the true covariance matrix of the background.

A simple algorithm for target detection is the spectral angle mapper (SAM) [60] given by

$$D_{\mathrm{SAM}}(\boldsymbol{x}) \triangleq \frac{\boldsymbol{s}^T\boldsymbol{x}}{(\boldsymbol{s}^T\boldsymbol{s})^{1/2}(\boldsymbol{x}^T\boldsymbol{x})^{1/2}}. \tag{17}$$

Clearly, SAM is the cosine of the angle between the test and target spectra and is always between zero and one because all spectra vectors have positive components. We note that SAM is usually defined in the remote sensing community as the angle between two vectors, instead of the cosine of the angle. SAM will provide adequate performance only for full pixel targets having well separated distributions with small dispersions. There are not any



▲ *7. Illustration of adaptive matched subspace detectors.*

optimality properties associated with the SAM algorithm, even for normally distributed classes.

Detection Algorithms for Subpixel Targets

By definition, subpixel targets occupy only part of the pixel area. The remaining part is filled with one or more materials, which will be collectively referred to as background. In this section, we discuss the detection of compact and isolated subpixel size objects characterized by a known spectral signature with or without variability. As a result of this area mixing, the observed spectral signature can be modeled reasonably well as a linear combination of the target and background spectra. Furthermore, there is always an additive amount of noise from various sources (sensor, atmosphere, etc).

The choice of the mathematical model used to describe the variability of target and background spectra (subspace versus statistical) leads to different families of subpixel target detection algorithms. The variability of the target spectral signature is always described using a subspace model $\boldsymbol{S}\boldsymbol{a}$. If the columns of $\boldsymbol{S}$ are endmembers, the vector $\boldsymbol{a}$ provides their abundances and should satisfy the constraints of the linear mixing model. Otherwise, $\boldsymbol{a}$ simply determines the position of a target in the column space of $\boldsymbol{S}$. The variability of the background can be described using either a subspace model (structured background) or a statistical distribution (unstructured background).

Therefore, the type of the background model drives the development of subpixel detection algorithms.

Unstructured Background Models

Unstructured background models assume that the additive noise has been included in the background $\boldsymbol{b}$, which in turn is modeled by a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Gamma}_b$, that is $\boldsymbol{b} \sim N(\boldsymbol{0}, \boldsymbol{\Gamma})$ (for simplicity, we drop the subscript $_b$ from this point forward). The competing hypotheses are

$$\begin{aligned} H_0&: \ \boldsymbol{x} = \boldsymbol{b}, && \text{Target absent} \\ H_1&: \ \boldsymbol{x} = \boldsymbol{S}\boldsymbol{a} + \boldsymbol{b}, && \text{Target present.} \end{aligned} \tag{18}$$

Hence, $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{\Gamma})$ under H_0 and $\boldsymbol{x} \sim N(\boldsymbol{S}\boldsymbol{a}, \boldsymbol{\Gamma})$ under H_1. In addition, we assume that we have access to a set of training background pixels $\boldsymbol{x}(n)$, $n = 1,2,\ldots,N$, which are independently and identically distributed (IID). The test pixel $\boldsymbol{x}$ and the training pixels are assumed statistically independent L-dimensional random vectors, where L is the number of spectral bands. Since $\boldsymbol{x}(n) \sim N(\boldsymbol{0}, \boldsymbol{\Gamma})$, we can use these pixels to obtain the maximum likelihood estimate of the covariance matrix. Since HSI data have a non-zero mean, we usually remove the estimated mean to comply with this model.

Using the generalized likelihood ratio approach, Kelly [17], [18] obtained the following detector:

$$D_K(\boldsymbol{x}) = \frac{\boldsymbol{x}^T \hat{\boldsymbol{\Gamma}}^{-1} \boldsymbol{S} (\boldsymbol{S}^T \hat{\boldsymbol{\Gamma}}^{-1} \boldsymbol{S})^{-1} \boldsymbol{S}^T \hat{\boldsymbol{\Gamma}}^{-1} \boldsymbol{x}}{N + \boldsymbol{x}^T \hat{\boldsymbol{\Gamma}}^{-1} \boldsymbol{x}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_A \tag{19}$$

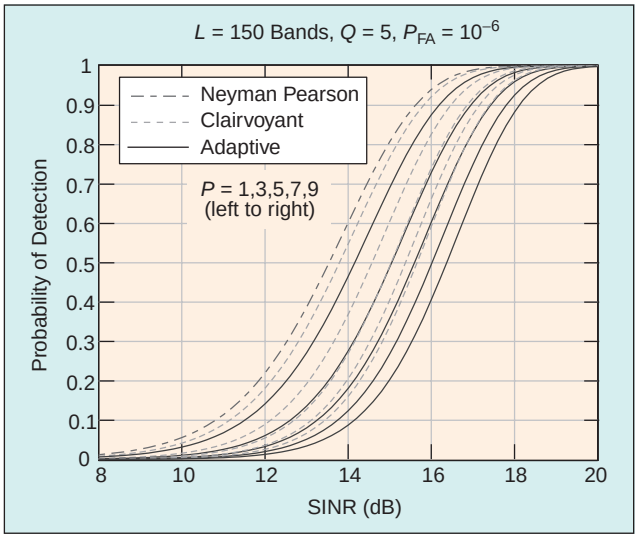
where $\hat{\boldsymbol{\Gamma}}$ is the MLE of the covariance matrix (15). Although there is no optimality test associated with the GLR approach [1], it leads to the design of useful, practical detectors. The threshold parameter η_K determines both the probability of detection, P_D, and the probability of false alarm, P_{FA}.

The matrix $\boldsymbol{S}$ contains the available a priori variability information about the target. This information decreases as we increase the number of columns P (dimensionality of the target subspace) of $\boldsymbol{S}$ and becomes minimum when $P = L$. In this case, we simply know that we are looking for deterministic targets that lie in the data subspace. Since the matrix $\boldsymbol{S}$ has full rank and therefore is invertible, (19) leads to the following detector:

$$D_A(\boldsymbol{x}) = \boldsymbol{x}^T \hat{\boldsymbol{\Gamma}}^{-1} \boldsymbol{x} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_A \tag{20}$$

which was derived by Kelly [19] using the approach discussed here and by Reed and Yu [5] using a multivariate analysis of variance formulation. Basically, $D_A(\boldsymbol{x})$ estimates the Mahalanobis distance of the test pixel from the mean of the background, which is zero for demeaned data. Algorithm (20), which has the CFAR property, is used extensively for anomaly detection in multispectra and hyperspectral data [8].

A key assumption in the derivation of (19) was that the covariance matrix of the background is the same under the two hypotheses. However, for subpixel targets the amount of background covered area is different under the two hypotheses. Therefore, it is more appropriate to use the following hypotheses:



▲ *8. Probability of detection as a function of SINR, illustrating the effect of target subspace dimensionality, for structured background detectors.*

$$H_0:\ \boldsymbol{x}=\boldsymbol{b}, \quad \text{Target absent}$$
$$H_1:\ \boldsymbol{x}=\boldsymbol{S}\boldsymbol{a}+\sigma\boldsymbol{b}, \quad \text{Target present} \tag{21}$$

which implies that $\boldsymbol{x}\sim N(\boldsymbol{0},\boldsymbol{\Gamma})$ under H_0 and $\boldsymbol{x}\sim N(\boldsymbol{S}\boldsymbol{a},\sigma^2\boldsymbol{\Gamma})$ under H_1. 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 GLR approach leads to the following adaptive coherence/cosine estimator (ACE) detector [20], [21]

$$D_{\text{ACE}}(\boldsymbol{x})=\frac{\boldsymbol{x}^T\hat{\boldsymbol{\Gamma}}^{-1}\boldsymbol{S}(\boldsymbol{S}^T\hat{\boldsymbol{\Gamma}}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\hat{\boldsymbol{\Gamma}}^{-1}\boldsymbol{x}}{\boldsymbol{x}^T\hat{\boldsymbol{\Gamma}}^{-1}\boldsymbol{x}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{\text{ACE}} \tag{22}$$

which can be obtained from (19) by removing the number N of background training pixels from the denominator.

If we use the adaptive whitening transformation $\tilde{\boldsymbol{x}}\triangleq\hat{\boldsymbol{\Gamma}}^{-1/2}\boldsymbol{x}$, where $\hat{\boldsymbol{\Gamma}}\triangleq\hat{\boldsymbol{\Gamma}}^{1/2}\hat{\boldsymbol{\Gamma}}^{1/2}$ is the square-root decomposition of the estimated covariance matrix, the ACE can be expressed as

$$D_{\text{ACE}}(\boldsymbol{x})=\frac{\tilde{\boldsymbol{x}}^T\tilde{\boldsymbol{S}}(\tilde{\boldsymbol{S}}^T\tilde{\boldsymbol{S}})^{-1}\tilde{\boldsymbol{S}}^T\tilde{\boldsymbol{x}}}{\tilde{\boldsymbol{x}}^T\tilde{\boldsymbol{x}}}=\frac{\tilde{\boldsymbol{x}}^T\boldsymbol{P}_{\tilde{S}}\tilde{\boldsymbol{x}}}{\tilde{\boldsymbol{x}}^T\tilde{\boldsymbol{x}}} \tag{23}$$

where $\tilde{\boldsymbol{S}}\triangleq\hat{\boldsymbol{\Gamma}}^{-1/2}\boldsymbol{S}$ and $\boldsymbol{P}_{\tilde{S}}\triangleq\tilde{\boldsymbol{S}}(\tilde{\boldsymbol{S}}^T\tilde{\boldsymbol{S}})^{-1}\tilde{\boldsymbol{S}}^T$ is the orthogonal projection operator onto the column space of $\tilde{\boldsymbol{S}}$. Since $\boldsymbol{P}_{\tilde{S}}^2=\boldsymbol{P}_{\tilde{S}}$, we can write (22) as

$$D_{\text{ACE}}(\boldsymbol{x})=\frac{\left\|\boldsymbol{P}_{\tilde{S}}\tilde{\boldsymbol{x}}\right\|^2}{\left\|\tilde{\boldsymbol{x}}\right\|^2}=\cos^2\theta \tag{24}$$

which shows that $D_{\text{ACE}}(\boldsymbol{x})$ is equal to the cosine of the angle between the test pixel and the target subspace into the whitened coordinate space. This is illustrated in Fig. 3.

Using the whitening transformation, the anomaly detector (19) can be expressed as $D_{\text{A}}(\boldsymbol{x})=\tilde{\boldsymbol{x}}^T\tilde{\boldsymbol{x}}$ which is the Euclidean distance of the test pixel from the background mean in the whitened space (see Fig. 4). We note that, in the absence of a target direction, the detector uses the distance from the center of the background distribution.

For targets without variability, we have $P=1$ and the target subspace $\boldsymbol{S}$ is specified by the direction of a single vector $\boldsymbol{s}$. Then, the formulas for the previous GLR detectors are simplified to

$$D(\boldsymbol{x})=\frac{(\boldsymbol{s}^T\hat{\boldsymbol{\Gamma}}^{-1}\boldsymbol{x})^2}{(\boldsymbol{s}^T\hat{\boldsymbol{\Gamma}}^{-1}\boldsymbol{s})(\psi_1+\psi_2\boldsymbol{x}^T\hat{\boldsymbol{\Gamma}}^{-1}\boldsymbol{x})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{25}$$

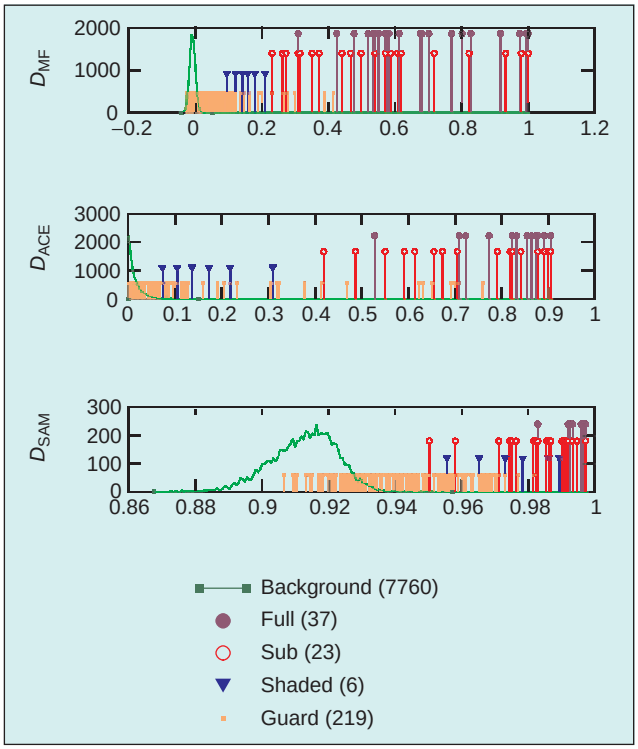
where $(\psi_1=N,\psi_2=1)$ for the Kelly detector and $(\psi_1=0,\psi_2=1)$ for the ACE algorithms. Kelly's algorithm was derived for real-valued signals and was applied to multispectral target detection in [6]. The one-dimensional version of ACE has been derived in [22], [23]. Finally, we note that, if $(\psi_1=N,\psi_2=0)$, we obtain the adaptive matched filter (AMF) detector [24], [25].

Determining the distribution of the various GLRT detectors is an elaborate process [17], [19], [5], [26], [21]. It turns out that the distribution of the different detector outputs involves a noncentral F-distribution. The noncentrality parameter is the theoretical signal-to-interference plus noise ratio (SINR)

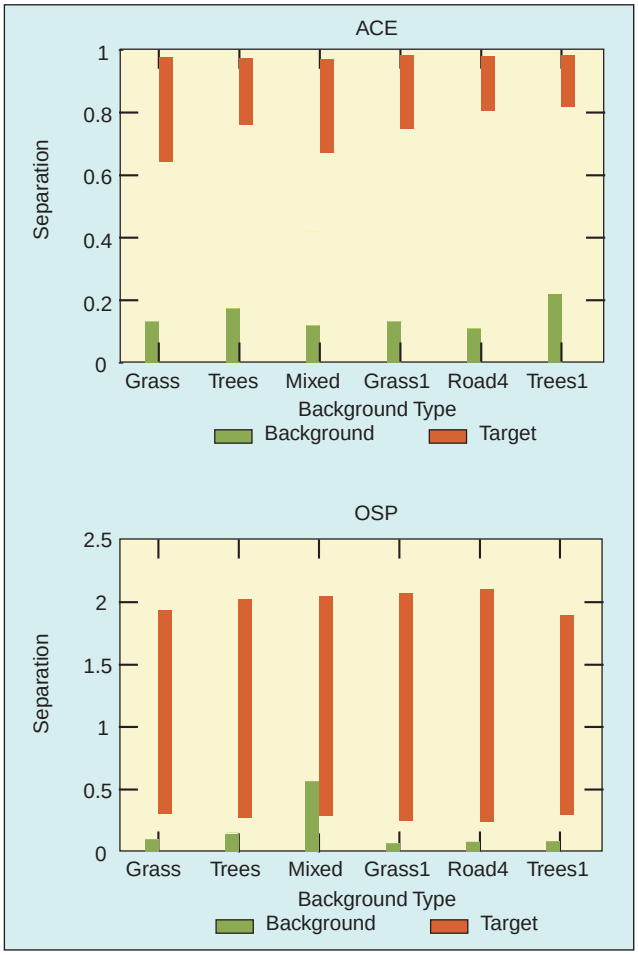
$$\text{SINR}_{\text{o}}=(\boldsymbol{S}\boldsymbol{a})^T\boldsymbol{\Gamma}^{-1}(\boldsymbol{S}\boldsymbol{a}). \tag{26}$$

The performance of all GLRT detectors depends only on the dimensional integer numbers L, P, N and the opti-

▲ 9. HYDICE Forest Radiance I data.



▲ *10. Response of the matched filter, ACE, and SAM detectors to background and different types of target pixels.*



▲ *11. Illustration of target visibility enhancement for different target detection algorithms.*

mum SINR_o parameter. Under the H_0 hypothesis (target absent) $\mathrm{SINR}_o = 0$, the output distribution becomes central , and the probability of false alarm depends only on the parameters L, P, N. Therefore, all GLRT detectors discussed in this section have the CFAR property. For a remote sensing system with high SNR, the detection performance depends on the fraction of the target occupying a pixel because this determines the SINR. Fig. 5 illustrates the effects of target dimensionality on detection performance for Kelly's GLRT detector. We see that as P increases (that is as a priori information about the target decreases), detection performance deteriorates. The worst performance is obtained when $P = L$, which corresponds to the anomaly detector (20). Fig. 6 illustrates performance as a function of the number of training pixels. Clearly, performance improves as N increases, that is as the estimate of the interference covariance matrix becomes more accurate. In both figures, we have included the curve indicating the performance of the optimum matched filter [1] (known target in noise with known covariance matrix) for comparison.

Structured Background Models

When the background variability is modeled using a subspace model, the target detection problem involves choosing between the following competing hypotheses

$$\begin{aligned} H_0&: \ \boldsymbol{x} = \boldsymbol{B}\boldsymbol{a}_b + \boldsymbol{w}, \quad \boldsymbol{w} \sim N(\boldsymbol{0}, \sigma_w^2 \boldsymbol{I}), \quad \text{Target absent} \\ H_1&: \ \boldsymbol{x} = \boldsymbol{S}\boldsymbol{a}_t + \boldsymbol{B}\boldsymbol{a}_b + \boldsymbol{w} = \boldsymbol{Z}\boldsymbol{a} + \boldsymbol{w}, \quad \text{Target present} \end{aligned} \tag{27}$$

where $\boldsymbol{S}\ (L \times P)$ has to be specified by the user, $\boldsymbol{B}\ (L \times Q)$ has to be determined from spectral libraries or the data, $\boldsymbol{Z} \stackrel{\triangle}{=} [\boldsymbol{S}\ \ \boldsymbol{B}]$, and $\boldsymbol{a}^T \stackrel{\triangle}{=} [\boldsymbol{a}_t^T \ \ \boldsymbol{a}_b^T]^T$. Use of the GLR approach leads to the following detection statistics:

$$D_{\mathrm{ASD}}(\boldsymbol{x}) = \frac{\boldsymbol{x}^T (\boldsymbol{P}_B^{\perp} - \boldsymbol{P}_Z^{\perp}) \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{P}_Z^{\perp} \boldsymbol{x}} = \frac{(\mathrm{CB})^2}{(\mathrm{CT})^2} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{\mathrm{ASD}} \tag{28}$$

where $\boldsymbol{P}_B^{\perp}$ and $\boldsymbol{P}_Z^{\perp}$ are projection matrices defined by $\boldsymbol{P}_A^{\perp} = \boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}^T \boldsymbol{A})^{-1} \boldsymbol{A}^T$ with $\boldsymbol{A} = \boldsymbol{B}$ and $\boldsymbol{A} = \boldsymbol{Z}$, respectively. This is illustrated in Fig. 7, where $\boldsymbol{P}_B^{\perp}\boldsymbol{x}$ is the perpendicular TB from the test pixel to the background subspace and $\boldsymbol{P}_Z^{\perp}\boldsymbol{x}$ is the perpendicular TC from the test pixel to the target and background subspace. The detector computes the square of the tangent of angle ϕ; however, $\cos^2 \phi$ can be also used as a detection statistics.

In the statistical literature $D_{\mathrm{ASD}}(\boldsymbol{x})$ is denoted by $F(\boldsymbol{x})$ and is known as the F-test. This decision statistic was first developed in the context of the linear statistical modeling [27]. In signal processing, it is known [1], [28] as the adaptive subspace detector (ASD). The random variable $T_{\mathrm{ASD}}(\boldsymbol{x})$ is distributed as

$$T_{\mathrm{ASD}}(\boldsymbol{x}) \frac{L - P - Q}{P} \sim F_{P, L-P-Q}(\mathrm{SINR}_o) \tag{29}$$

where $F_{P,L-P-Q}(\mathrm{SINR}_o)$ is the noncentral F distribution with P numerator degrees of freedom, $L-P-Q$ denominator degrees of freedom, and the noncentrality parameter: the SINR

$$\mathrm{SINR}_o \triangleq \frac{\left\| \boldsymbol{P}_B^{\perp} \boldsymbol{S} \boldsymbol{a}_t \right\|^2}{\sigma_w^2} \tag{30}$$

which depends on the unknown variance σ_w^2. Notice that σ_w^2 is not required for the computation of the detection statistics. The threshold η_{ASD} is specified by the required probability of false alarm

$$P_{\mathrm{FA}} = 1 - F_{P,L-P-Q}(0, \eta_{\mathrm{ASD}}) \tag{31}$$

because $\mathrm{SINR}_o = 0$ under H_0. Since the distribution of $T_{\mathrm{ASD}}(\boldsymbol{x})$ under H_0 is known, we can set the threshold η_{ASD} to attain CFAR operation. The probability of detection

$$P_D = 1 - F_{P,L-P-Q}(\mathrm{SINR}_o, \eta_{\mathrm{ASD}}) \tag{32}$$

depends on the unknowns $\boldsymbol{a}_t$ and σ_w^2; therefore, it cannot be maximized to obtain an optimum Neyman-Pearson detector.

For targets without variability, that is, when $P = 1$, it can be shown that the ML estimate of the abundance a_t of the target signature vector $\boldsymbol{s}$ is given by

$$\hat{a}_t = \frac{\boldsymbol{s}^T \boldsymbol{P}_B^{\perp} \boldsymbol{x}}{\boldsymbol{s}^T \boldsymbol{P}_B^{\perp} \boldsymbol{s}} \sim N\left(a_t, \sigma_w^2 \left(\boldsymbol{s}^T \boldsymbol{P}_B^{\perp} \boldsymbol{s}\right)^{-1}\right). \tag{33}$$

Clearly, detection algorithms based on $\hat{a}_t$ cannot have the CFAR property because σ_w^2 is unknown. Also, since the distribution under H_1 depends on a_t and σ_w^2, which are both unavailable, the probability of detection cannot be maximized to attain Neyman-Pearson optimality. Clearly, the theoretically predicted Gaussian distribution of the target abundance conflicts with the physical constraint $0 \le \hat{a}_t \le 1$.

A widely used detection algorithm is the orthogonal subspace projector (OSP) [29]

$$T_{\mathrm{OSP}}(\boldsymbol{x}) = \boldsymbol{s}^T \boldsymbol{P}_B^{\perp} \boldsymbol{x}. \tag{34}$$

However, it is better to use the normalized statistic (32) because it can be interpreted as the abundance estimate for the desired target. The resulting algorithm is not CFAR because the abundance of the target and the variance of the sensor noise are unknown. The operation $\boldsymbol{e}_b = \boldsymbol{P}_B^{\perp} \boldsymbol{x}$ removes from $\boldsymbol{x}$ the part which belongs to the background $\boldsymbol{B}$. The decision is based on the correlation $y = \boldsymbol{s}^T \boldsymbol{e}_b$ between the residual "target plus noise" and the target template $\boldsymbol{s}$, which is proportional to the angle between those two vectors.

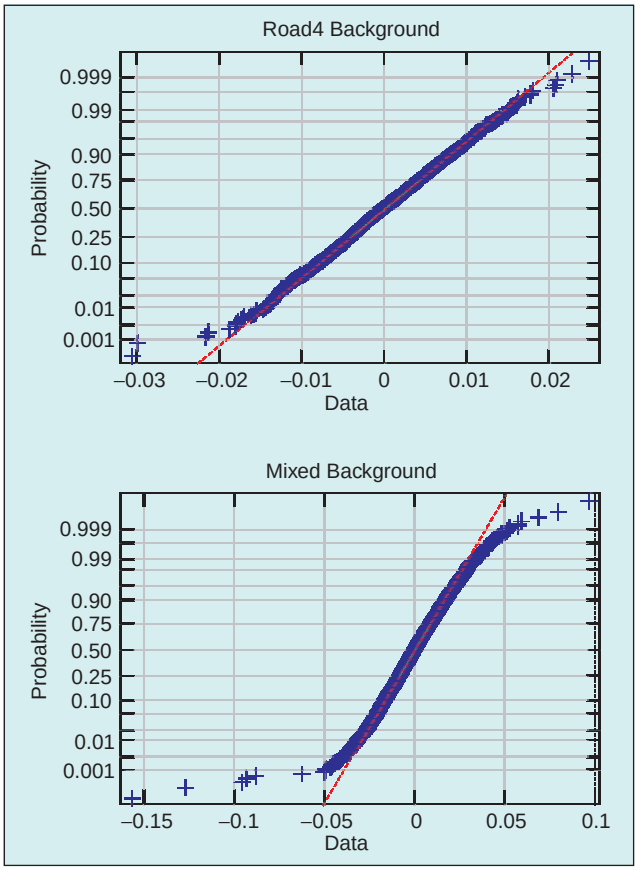
To get some insight into the performance of the GLRT detector we consider a background with $Q = 10$ end members. The number of bands is $L = 150$ and the

Dealing with spectral signature variability and spectral compositions in mixed pixels are among the most challenging problems in HSI data exploitation.

probability of false alarm is fixed at $P_{\mathrm{FA}} = 10^{-5}$. The dimension of the target subspace is varied from $P = 1$ to $P = 10$. Fig. 8 shows plots of the probability of detection as a function of the SINR for fixed values of P_{FA}, Q, P, and L. The family of curves, which is parameterized with respect to P, shows that performance deteriorates as the dimensionality of the target subspace increases (that is as the a priori information about the target decreases) as is expected.

Practical Detector Performance Evaluation

The implementation of hyperspectral target detection algorithms in a real-world environment involves confrontation with many "practical details" and challenges that result from the violation of the theoretical assumptions used for the derivation of the various algorithms.



▲ *12. Evaluation of the matched filter CFAR property for two different backgrounds.*

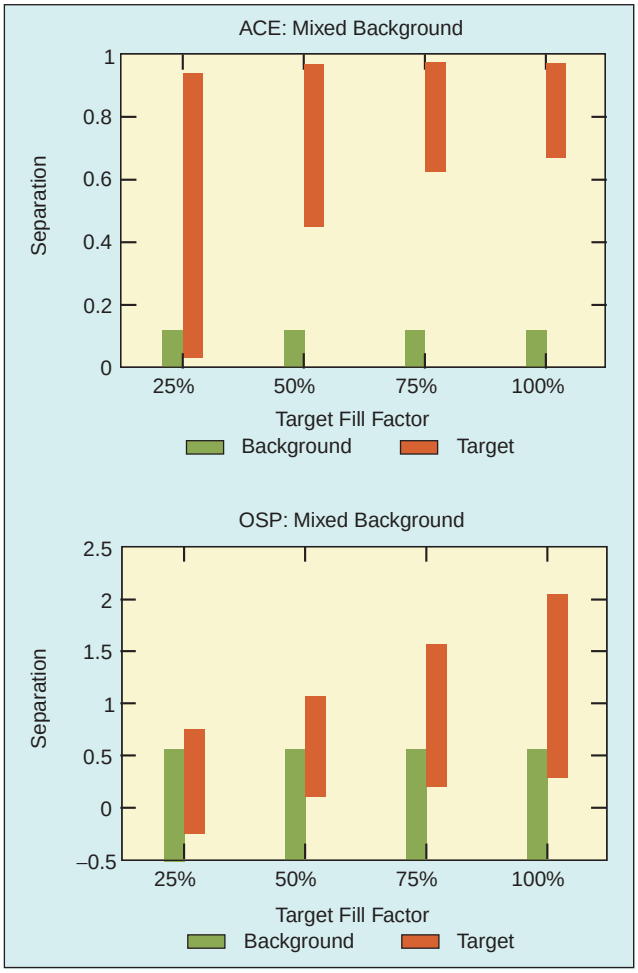
The key assumptions used for the detectors using the covariance matrix of the background are:

- The background is homogeneous and can be modeled by a multivariate normal distribution;
- The background spectrum interfering with the test pixel spectrum has the same covariance matrix with the background training pixels;
- The test and training pixels are independent;
- The target and background spectra interact in an additive manner (additive instead of replacement model).

The key assumptions used for the detectors using the background subspace model are:

- The spectrum of every pixel can be adequately modeled by the LMM;
- The modeling error has uncorrelated components with the same variance;
- The distribution of the modeling error is multivariate normal;
- The background subspace $\boldsymbol{B}$ is perfectly known;
- The target subspace template $\boldsymbol{S}$ is known.

To illustrate various issues regarding the discussed algorithms, we shall use the Forest Radiance I data (see Fig. 9) collected with the HYDICE sensor. We have selected the three areas outlined in Fig. 9 to investigate three different types of background: grass (G), trees (T), and mixed grass-road (GR). The first two scenes are relatively homogeneous, whereas the third scene consists of a nonhomogeneous background including different types of grass and roads. Also considered were classes resulting from a supervised classification process performed to isolate spectrally similar (not necessarily spatially adjacent) pixels. Data from classes selected from this analysis were labelled using a number after the class label, for example "grass2."



▲ *13. Target-background separability as a function of the target fill factor a.*

Regarding targets, we have chosen a multipixel vehicle target. The mean value of the set is used as the target template $\mathbf{s}$ required for the implementation of each detection algorithm.

The performance evaluation of detection algorithms in practice is challenging due to the limitations imposed by the limited amount of target data. As a result the establishment of accurate ROC curves is quite difficult. We shall compare the various algorithms in terms of their ability to operate in CFAR mode and the enhancement of the separation between targets and background they provide. Sensor and environmental noise do not seem to be significant factors. The CFAR property depends on the capability to accurately model the detection statistics $D(\boldsymbol{x})$ of the background pixels for a given algorithm. The enhancement of target visibility will be illustrated visually by plotting the detection statistics for every pixel in the scene.

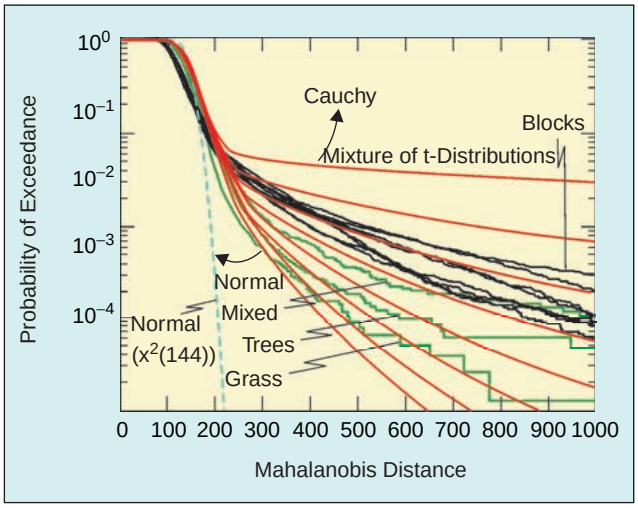
Another complicating factor is that, in the case of real data, we can identify pure or mixed pixels with only a certain level of confidence. Therefore, it is useful to label the pixels in the vicinity of a target as full, mixed, shadow, and guard pixels and distinguish among such pixels when we compare different detectors. Taking into consideration that the number of such pixels much smaller than the number of background pixels, a useful way to represent the output of any target detector is shown in Fig. 10. The key idea is to represent the background response by its histogram and the response of the various "target" pixels by stems indicating the magnitude of the detector's output and the type of the pixel.

To illustrate these points, we investigate the performance of the ACE and OSP detectors, by assessing their capacity to enhance the "visibility" of the desired target and to accurately model the background statistics. Roughly speaking, a target becomes more visible to the detector when the background-target separation increases. For the estimate of the background subspace $\boldsymbol{B}$ for the OSP detector, we used the first $Q=10$ significant eigenvectors of the estimated correlation matrix of the background pixels (the "known" target pixels were excluded). These eigenvectors capture more than 99% of the data cube energy in each case. Since the estimated covariance matrix is low rank, its inverse can be approximated by $\boldsymbol{\Gamma}^{-1} \simeq (\boldsymbol{I} - \boldsymbol{U}_Q \boldsymbol{U}_Q^T)$, where $\boldsymbol{U}_Q$ contains the Q dominant eigenvectors. Since the columns of $\boldsymbol{U}_Q$ are orthonormal, $\boldsymbol{P}_{U_Q}^{\perp} = \boldsymbol{I} - \boldsymbol{U}_Q \boldsymbol{U}_Q^T$ is the orthogonal subspace projector onto the compliment of the back-

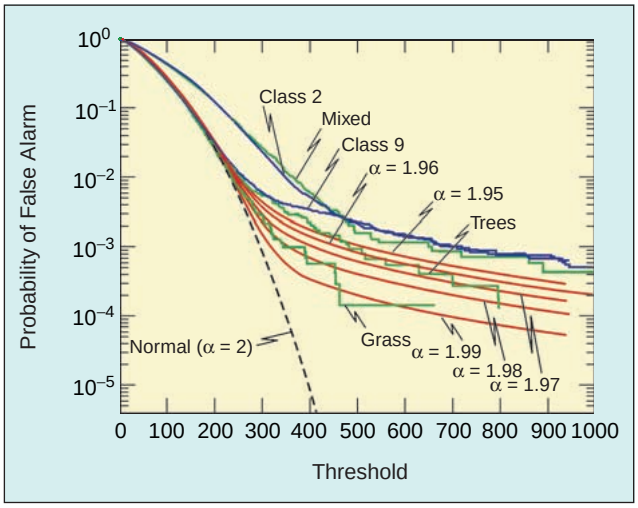
ground subspace spanned by the columns of $\boldsymbol{B}$ or $\boldsymbol{U}_Q$ [30].

The bar charts in Fig. 11 provide the range of the detection statistic of the target and the maximum value of the background detection statistics for various backgrounds. The target-background separation or overlap is the quantity used to evaluate target visibility enhancement. For example, it can be seen that the ACE detector performs better than the OSP algorithm for the six data sets shown.

The expected probability distribution of the detection statistics under the "target absent" hypothesis can be compared to the actual statistics using a quantile-quantile (Q-Q) plot. A Q-Q plot shows the relationship between the quantiles of the expected distribution and the actual data. An agreement between the two is illustrated by a straight line. The Q-Q plots in Fig. 12 illustrate the comparison between the experimental detection statistics to the theoretically predicted ones for the matched filter algorithms. The actual statistics for two different backgrounds is compared to the normal distribution. A straight line shown that the postulated model provides a good fit and therefore can be used to estimate the threshold for CFAR operation.



▲ *14. Modeling the statistics of the Mahalanobis distance.*



▲ *15. Modeling the matched filter output statistics using stable distributions.*

The performance evaluation of detection algorithms in practice is challenging due to the limitations imposed by the limited amount of target data.

The previous results dealt with full-pixel or resolved targets. To evaluate detection performance for subpixel targets, we have simulated subpixel targets using formula (3). Subpixel targets were simulated by adding a randomly chosen target pixel from the target pixel set to each of the background pixels at a constant fraction. The results shown in Fig. 13, show target-background separability as a function of the target fill factor a for the ACE and OSP detectors. Clearly, target visibility improves with the size of the target. A more detailed comparison of a large set of detection algorithms is provided in [31]. It has been shown that taking into consideration target variability using a subspace model can increase detection performance [32].

When the spectral observation vector $\boldsymbol{x}$ is distributed as $N(\boldsymbol{\mu},\boldsymbol{\Gamma})$, its Mahalanobis distance follows a chi-squared distribution with L degrees of freedom. By removing the mean, we obtain the anomaly detector (19). However, for nonnormal data the distribution of Mahalanobis distance is not chi-squared. Fig. 14 shows the probability of false alarm for the three sets shown in Fig. 9 as well as eight blocks obtained by partitioning this data cube into a four by two matrix. The figure also shows theoretical predictions based on a chi-squared and a mixture of two F-distributions. Evidently, the F-mixture provides a good description for the body and the tails of the underlying distribution. We note that if the data follow an elliptical multivariate t distribution, the Mahalanobis distance follows a univariate F distribution [33]. The multivariate normal and t distributions is a special case of the family of *elliptically contoured* distributions [33] specified by the distribution $f(\boldsymbol{x}) = |\boldsymbol{\Gamma}|^{-1/2}\, g\{(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Gamma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}$. The form of function $g(\ \ )$ leads to distributions with heavier or lighter tails than the normal.

The heavy tails in the univariate distribution of the Mahalanobis distance imply heavy tails in the multivariate distribution of the data. Therefore, heavy tails may appear not only in the quadratic Mahalanobis distance, but in other linear and quadratic statistics employed in several widely used [34], [31] target detection techniques.

The family of symmetric α-stable ($S\alpha S$) distributions provides a good model for data with impulsive behavior. They are characterized by a parameter α (characteristic exponent) that takes values in the range $0 < \alpha \leq 2$. The value

$\alpha = 1$ leads to the Cauchy distribution and the value $\alpha = 2$ to the Gaussian. The stable distributions result from the central limit theorem if we remove the finite variance constraint. The only stable distribution with finite second-order moments is the normal distribution. Since α-stable distributions follow from the central limit theorem they are invariant under linear transformations. Since there is no closed-form expression for their probability density function, SαS random variables are specified [11], [35], [36] by their characteristic function (that is, the Fourier transform of the PDF)

$$\Phi(\xi) = \exp\left(j\mu\xi - |\sigma\xi|^{\alpha} \right) \tag{35}$$

where α is the characteristic exponent, σ is a scale parameter, and μ is a location parameter. The heaviness of the tails increases as α increases from 1 (Cauchy) to 2 (Gaussian). The estimation of the parameters of a stable distribution from data is challenging due to the presence of "spikes." The published compilation [36] provides a comprehensive review of statistical techniques for stable distributions from a practical perspective. The estimation method used in this article [37] is based on the use of the characteristic function.

Fig. 15 shows the probability of false alarm when the matched filter detector (16) is used for different scenes as well as superimposed theoretical curves obtained using the family of SαS distributions for various values of α. It can be seen that the tails of the empirical P_{FA} curves can be modeled by the heavier tails of the stable distribution.

Conclusions

This article provided a tutorial review and state-of-the-art of target detection algorithms for hyperspectral imaging applications. The main obstacles in the development of effective detection algorithms are the inherent variability target and background spectra. The use of adaptive algorithms deals quite effectively with the problem of unknown background; however, the lack of sufficient target data makes the development and estimation of target variability models challenging. Hyperspectral target detection is a multidisciplinary problem that should draw upon different scientific and engineering areas.

Acknowledgments

We wish to express our gratitude to Capt. Frank Garcia, DUSD (S&T), Manager of the Hyperspectral Technology Assessment Program, for his enthusiastic support and to SITAC for providing the calibrated and ground-truthed HYDICE data used in this work. We also thank Sylvia S. Shen for providing exceptionally detailed and helpful comments and David Marden for helping with the experimental results. This work was sponsored by the Department of the Defense Air Force contract F19628-95-C-0002.

Dimitris G. Manolakis received the B.S. in physics and the Ph.D. in electrical engineering from the University of Athens, Greece. He is currently a member of the technical staff at MIT Lincoln Laboratory, Lexington, MA. Previously, he was a Principal Member, Research Staff, at Riverside Research Institute. He has taught at the University of Athens, Northeastern University, Boston College, and Worcester Polytechnic Institute. He is coauthor of the textbooks *Digital Signal Processing: Principles, Algorithms, and Applications* (Prentice-Hall, 1996, 3d ed.) and *Statistical and Adaptive Signal Processing* (McGraw-Hill, 2000). His research interests include digital signal processing, adaptive filtering, remote sensing, array processing, pattern recognition, and radar systems.

Gary Shaw received the B.S. and M.S. degrees in electrical engineering from the University of South Florida, Tampa, and the Ph.D. in electrical engineering from the Georgia Institute of Technology, Atlanta. He is Senior Staff Member at Lincoln Laboratory. Since joining Lincoln Laboratory in 1980, his work has primarily been directed toward algorithms and architectures for high-speed digital signal processing, communication, and remote sensing.

References

[1] S.M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1998.

[2] J.B. Adams, M.O. Smith, and A.R. Gillespie, "Remote geochemical analysis: Elemental and mineralogical composition," in *Imaging spectroscopy: Interpretation Based on Spectral Mixture Analysis*, C.M. Pieters and P.A.J. Englert, Eds. Cambridge, U.K.: Cambridge Univ. Press, 1993, pp. 145-166.

[3] A.D. Stocker and A. Schaum, "Application of stochastic mixing models to hyperspectral detection problems," *SPIE Proc.*, vol. 3071, April 1997.

[4] A. Schaum and A. Stocker, "Spectrally-selective target detection," in *Proc. ISSSR*, 1997.

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

[6] X. Yu and I.S. Reed, "Comparative performance analysis of adaptive multiband detectors," *IEEE Trans. Signal Processing*, vol. 41, pp. 2639-2656, Aug. 1993.

[7] D.W. Stein and S.G. Beaven, "The fusion of quadratic detection statistics applied to hyperspectral imagery," *IRIA-IRIS Proc. 2000 Meeting of the MSS Specialty Group on Camouflage, Concealment, and Deception*, Mar. 2000, pp. 271-280.

[8] D. Stein, S. Beaven, L. Hoff, E. Winter, A. Schaum, and A. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Processing Mag.*, vol. 19, pp. 58-69, Jan. 2002.

[9] R.A. Fisher, "The use of multiple measurement in taxonomic problems," *Ann. Eugenics*, vol. 7, pp. 179-188, 1936.

[10] R.A. Johnson and D.W. Wichem, *Applied Multivariate Statistical Analysis*. Englewood Cilffs, NJ: Prentice Hall, 1998.

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

[12] J.C. Harsanyi, "Detection and classification of subpixel spectral signatures in hyperspectral image sequences," Ph.D. dissertation, University of Maryland, 1993.

[13] R. Sitgreaves, "On the distribution of two random matrices used in classification," *Ann. Math. Statist.*, vol. 23, no. 2, pp. 263-270, June 1952.

[14] C.D. Richmond, "Derived PDF of maximum likelihood signal estimator which employs an estimated noise covariance," *IEEE Trans. Signal Processing*, vol. 44, pp. 305-315, Feb. 1996.

[15] D. Manolakis and G. Shaw, "Directionally constrained or constrained energy minimization adaptive matched filter: Theory and practice," in *Imaging Spectrometry*, M. Descour and S. Shen, Eds., San Diego, CA, Aug. 2001, SPIE.

[16] R.A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*. San Diego, CA: Academic, 1997.

[17] E.J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, pp. 115-127, March 1986.

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

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

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

[21] S. Kraut, L.L. Scharf, and L.T. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Processing*, vol. 49, no. 1, pp. 1-16, Jan. 2001.

[22] E. Conte, M. Lops, and G. Ricci, "Asymptotically optimum radar detection in compound-Gaussian clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 2, pp. 617-625, 1995.

[23] L.L. Scharf and L.T. McWhorter, "Adaptive matched subspace detectors and adaptive coherence," in *Proc. 30th Asilomar Conf. Signals and Systems*, 1996, pp. 1114-117.

[24] F.C. Robey, D.R. Fuhrmann, E.J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208-218, Jan. 1992.

[25] W. Chen and I.S. Reed, "A new CFAR detection test for radar," *Digital Signal Processing*, vol. 1, no. 4, pp. 198-214, 1991.

[26] C.D. Richmond, "Performance of the adaptive sidelobe blanker detection algorithm in homogeneous environments," *IEEE Trans. Signal Processing*, vol. 48, no. 5, pp. 1235-1247, May 2000.

[27] S.R. Searle, *Linear Models*. New York: Wiley, 1971.

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

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

[30] C.H. Gierull, "Statistical analysis of the eigenvector projection method for adaptive spatial filtering of interference," *IEE Proc. Radar Sonar Navigation*, vol. 144, pp. 57-63, 1997.

[31] D. Manolakis, C. Siracusa, D. Marden, and G.Shaw, "Hyperspectral adaptive matched filter detectors: Practical performance comparison," in *Algorithms for Multispectral and Hyperspectral Imagery*, Orlando, FL, April 2001, SPIE.

[32] D. Manolakis, C. Siracusa, and G. Shaw, "Adaptive matched subspace detectors for hyperspectral imaging applications," presented at IEEE Int. Conf. ASSP, Salt Lake City, Utah, May 2001.

[33] R.J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.

[34] D. Manolakis, G. Shaw, and N. Keshava, "Comparative analysis of hyperspectral adaptive matched filter detectors," in *Algorithms for Multi-spectral and Hyperspectral Imagery VI*, S.S. Shen and M.R. Descour, Eds., Orlando, FL, April 2000, vol. 4049, pp. 2-17, SPIE.

[35] G. Samorodnitsky and M.S. Taqqu, *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance*. New York: Chapman and Hall, 1994.

[36] R.J. Adler, R.E. Feldman, and M.S. Taqqu, *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Boston, MA: Birkhauser, 1998.

[37] S. Kogon and D. Williams, "Characteristic function based estimation of stable distribution parameters," in *A Pratical Guide to Heavy Tails: Statistical Techniques and Applications*, R.J. Adler, R.E. Feldman, and M.S. Taqqu, Eds. Boston, MA: Birkhauser, 1998, pp. 311- 335.