Hyperspectral Image Processing for Automatic Target Detection Applications

Dimitris Manolakis, David Marden, and Gary A. Shaw

■ This article presents an overview of the theoretical and practical issues associated with the development, analysis, and application of detection algorithms to exploit hyperspectral imaging data. We focus on techniques that exploit spectral information exclusively to make decisions regarding the type of each pixel—target or nontarget—on a pixel-by-pixel basis in an image. First we describe the fundamental structure of the hyperspectral data and explain how these data influence the signal models used for the development and theoretical analysis of detection algorithms. Next we discuss the approach used to derive detection algorithms, the performance metrics necessary for the evaluation of these algorithms, and a taxonomy that presents the various algorithms in a systematic manner. We derive the basic algorithms in each family, explain how they work, and provide results for their theoretical performance. We conclude with empirical results that use hyperspectral imaging data from the HYDICE and Hyperion sensors to illustrate the operation and performance of various detectors.

MANY MILITARY AND CIVILIAN applications involve the detection of an object or activity such as a military vehicle or vehicle tracks. Hyperspectral imaging sensors, such as those illustrated in Figure 1, provide image data containing both spatial and spectral information, and this information can be used to address such detection tasks. The basic idea for hyperspectral imaging stems from the fact that, for any given material, the amount of radiation that is reflected, absorbed, or emitted—i.e., the radiance—varies with wavelength. Hyperspectral imaging sensors measure the radiance of the materials within each pixel area at a very large number of contiguous spectral wavelength bands.

A hyperspectral remote sensing system has four basic parts: the radiation (or illuminating) source, the atmospheric path, the imaged surface, and the sensor.

In a passive remote sensing system the primary source of illumination is the sun. The distribution of the sun's emitted energy, as a function of wavelength throughout the electromagnetic spectrum, is known as the solar spectrum. The solar energy propagates through the atmosphere, and its intensity and spectral distributions are modified by the atmosphere. The energy then interacts with the imaged surface materials and is reflected, transmitted, and/or absorbed by these materials. The reflected/emitted energy then passes back through the atmosphere, where it is subjected to additional intensity modifications and spectral changes. Finally, the energy reaches the sensor, where it is measured and converted into digital form for further processing and exploitation.

The signal of interest, that is, the information-bearing signal, is the reflectance spectrum defined by
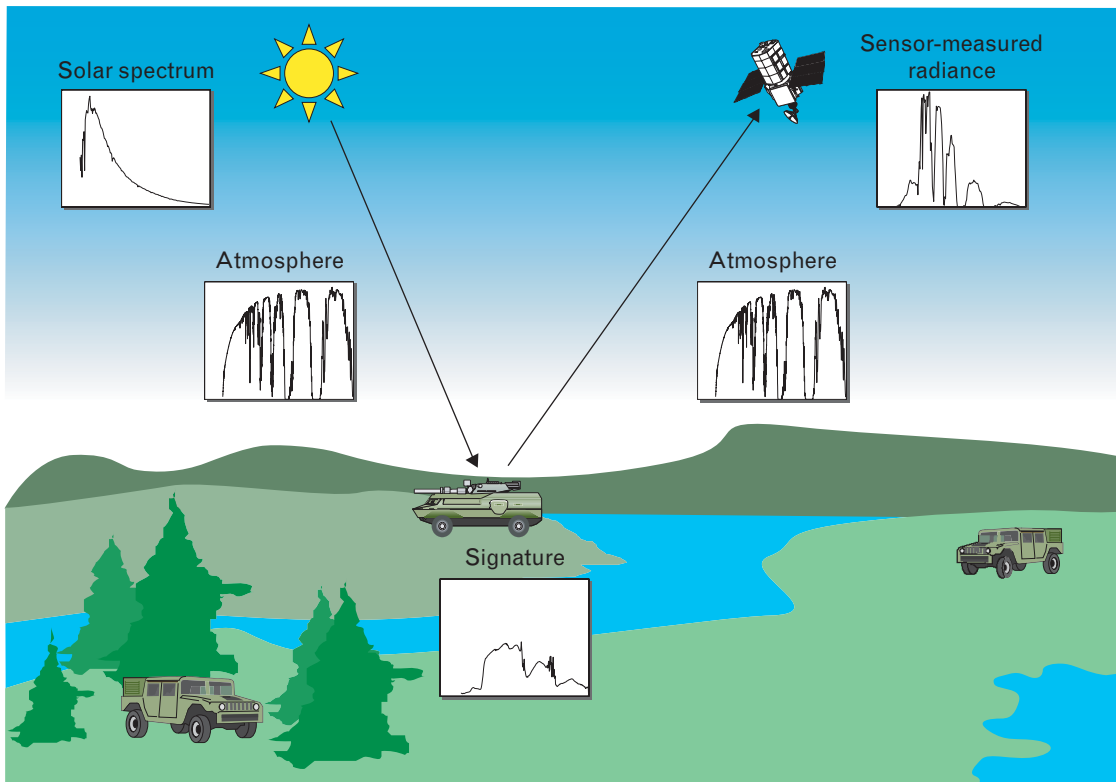
FIGURE 1. Hyperspectral imaging sensors measure the spectral radiance information in a scene to detect target objects, vehicles, and camouflage in open areas, shadows, and tree lines. Imaging sensors on satellites or aircraft gather this spectral information, which is a combination of sunlight (the most common form of illumination in the visible and near infrared), atmospheric attenuation, and object spectral signature. The sensor detects the energy reflected by surface materials and measures the intensity of the energy in different parts of the spectrum. This information is then processed to form a hyperspectral data set. Each pixel in this data set contains a high-resolution spectrum, which is used to identify the materials present in the pixel by an analysis of reflectance or emissivity.

$$\text{reflectance spectrum}(\lambda) = \frac{\text{reflected radiation at band}(\lambda)}{\text{incident radiation at band}(\lambda)}.$$

The reflectance spectrum, or spectral signature, shows the fraction of incident energy, typically sunlight, that is reflected by a material as a function of the wavelength λ of the energy.

To understand hyperspectral imaging data exploitation, it is important to realize how the presence of the atmosphere and the nature of the solar spectrum affect the relationship between the observed radiance spectra and the associated reflectance spectra. Basically, the observed spectrum would have the same shape as the reflectance spectrum if the solar spectrum were flat and the atmosphere had the same

transmittance at all wavelengths. In reality, the measured radiance spectrum is the solar spectrum modified by the transmittance function of the atmosphere and the reflectance spectrum of the imaged ground resolution cell (see Figure 1). The atmosphere restricts where we can look spectrally because it selectively absorbs radiation at different wavelengths, due to the presence of oxygen and water vapor. The signal-to-noise ratio at these absorption bands is very low; as a result, any useful information about the reflectance spectrum is lost. For all practical purposes these spectral bands are ignored during subsequent hyperspectral data analysis.

Any effort to measure the spectral properties of a material through the atmosphere must consider the absorption and scattering of the atmosphere, the subtle effects of illumination, and the spectral re-
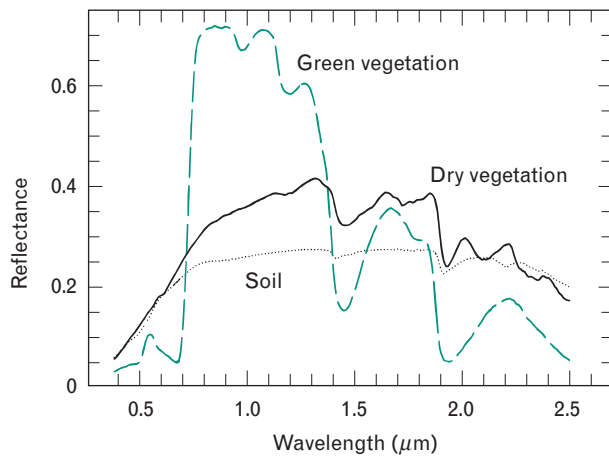
FIGURE 2. Different materials produce different electromagnetic radiation spectra, as shown in this plot of reflectance for soil, dry vegetation, and green vegetation. The spectral information contained in a hyperspectral image pixel can therefore indicate the various materials present in a scene.

sponse of the sensor. The recovery of the reflectance spectrum of each pixel from the observed radiance spectrum is facilitated with the use of sophisticated atmospheric compensation codes.

The resulting reflectance representation, termed the spectral signature, if sufficiently characterized, can be used to identify specific materials in a scene. Figure 2, for example, illustrates the different reflectance spectra for naturally occurring materials such as soil, green vegetation, and dry vegetation.

The hyperspectral sensors discussed in this article measure radiation in the region of the electromagnetic spectrum dominated by solar illumination (0.4 μm to 2.5 μm), where they acquire a complete reflectance spectrum. Unfortunately, such sensors are ineffective at night because of the absence of solar illumination. The problem can be avoided by extending the range of hyperspectral sensors into the thermal infrared region (8 μm to 14 μm), where materials emit more radiation than they reflect from the sun. This extension allows sensors to operate effectively during both day and night.

Hyperspectral imaging sensors are basically advanced digital color cameras with fine spectral resolution at given wavelengths of illumination. Instead of measuring three primary colors—red, green, and blue—these sensors measure the radiation reflected by each pixel at a large number of visible or invisible

frequency (or wavelength) bands. Such an instrument is called an imaging spectrometer. Spectrometers are instruments that divide the impinging electromagnetic radiation into a number of finite bands and measure the energy in each band. (In this sense, the human eye can be considered a crude imaging spectrometer.) Depending upon the design of the spectrometer, the bands may or may not be contiguous and of equal width. Therefore, the measured discrete spectrum is usually not a uniformly sampled version of the continuous one.

The optical system divides the imaged ground area into a number of contiguous pixels. The ground area covered by each pixel (called the ground resolution cell) is determined by the instantaneous field of view (IFOV) of the sensor system. The IFOV is a function of the optics of the sensor, the size of the detector elements, and the altitude. The rectangular arrangement of the pixels builds an image, which shows spatial variation, whereas the spectral measurements at each pixel provide the integrated spectrum of the materials within the pixel.

Scene information is contained in the observed radiance as a function of continuous space, wavelength, and time variables. However, all practical sensors have limited spatial, spectral, radiometric, and temporal resolution, which results in finite-resolution recordings of the scene radiance. The spatial resolution of the sensor determines the size of the smallest object that can be seen on the surface of the earth by the sensor as a distinct object separate from its surroundings. Spatial resolution also relates to how well a sensor can record spatial detail. The spectral resolution is determined by the width $\Delta\lambda$ of the spectral bands used to measure the radiance at different wavelengths λ. The radiometric resolution is determined by the number of bits used to describe the radiance value measured by the sensor at each spectral band. Finally, the temporal resolution is related to how often the sensor revisits a scene to obtain a new set of data.

Spectral and radiometric resolution determine how faithfully the recorded spectrum resembles the actual continuous spectrum of a pixel. The position, number, and width of spectral bands depend upon the sensor used to gather the information. Simple optical sensors may have one or two wide bands, multispec-

Table 1. Comparison of Spatial Processing and Spectral Processing in Remote Sensing

Spatial Processing	Spectral Processing
Information is embedded in the spatial arrangement of pixels in every spectral band (two-dimensional image)	Each pixel has an associated spectrum that can be used to identify the materials in the corresponding ground-resolution cell
Image processing exploits geometrical shape information	Processing can be done one pixel at a time
Very high spatial resolution required to identify objects by shape (many pixels on the target)	No need for high spatial resolution (one pixel on the target)
High spatial resolution requires large apertures and leads to low signal-to-noise ratio	Spectral resolution more important than spatial resolution
Data volume grows with the square of the spatial resolution	Data volume increases linearly with the number of spectral bands
Limited success in developing fully automated spatial-feature exploitation algorithms	Fully automated algorithms for spectral-feature exploitation have been successfully developed for selected applications

tral sensors a few wide bands, and hyperspectral sensors a few hundred narrow bands. Table 1 summarizes some of the distinctions among remote sensing systems that exploit spatial information and spectral information.

As noted in the introductory article to this special issue of the *Journal*, entitled "Spectral Imaging for Remote Sensing," by Gary A. Shaw and Hsiao-hua K. Burke, multispectral sensing and, later, hyperspectral sensing were developed for remote sensing of the natural environment, including such applications as mineral exploration, characterization of ground cover, and characterization of crop health. In these applications, morphological (or shape) information (which is an important prerequisite for remote sensing of man-made objects) is of minimal utility in the detection process, since the various natural materials of interest do not have predefined shapes. For example, an area of diseased crops will have an unpredictable size and shape within a larger field of healthy crops. Similarly, mineral deposits will have unpredictable shapes and may even be intimately mixed with other types of materials.

Another important aspect of traditional applications for hyperspectral imaging is that the applications are oriented toward classifying or grouping similar pixels, when there are typically many instances of each class of pixel, and when occasional errors in pixel classification are not significant, since interpretation of the scene is based on the clustering of the majority of pixels. For example, in an image of crops, if one in every one thousand pixels is misclassified, the isolated errors will not change the overall understanding of whether the crops are healthy or diseased.

In comparison, any target detection application seeks to identify a relatively small number of objects with fixed shape or spectrum in a scene. The processing methods developed for environmental classification are not applicable to target detection for two reasons. First, the number of targets in a scene is typically too small to support estimation of statistical properties of the target class from the scene. Second, depending upon the spatial resolution of the sensor, targets of interest may not be clearly resolved, and hence they appear in only a few pixels or even a single pixel. The isolated character of the targets makes confirmation by means of clustering of like samples problematic. Target detection is still possible, however, in these situations. For example, hyperspectral sensing could be used to increase the search rate and probability of detection for search and rescue operations at sea. A life jacket, which may be unresolved spatially,

can still be detected and identified on the basis of its extreme color contrast with the ocean background.

In practice, hyperspectral sensors represent a deliberate trade-off in which spatial resolution is degraded in favor of improved spectral resolution. By implication, hyperspectral imaging is therefore best matched to applications in which spectral information is more reliable or measurable than morphology or shape information. For example, traditional analysis of passive imagery of military vehicles depends strongly upon being able to determine the length, width, and distinguishing features of a vehicle. If the vehicles are partially hidden under foliage or camouflaged nets, the morphological information is unreliable and alternate means of detection and identification are needed. Since hyperspectral imaging does not rely on shape information, it is less impacted by attempts at concealment and deception.

Historically, interpretation of passive imagery has generally depended upon either human analysis of visual information in a live video or a still image, or else upon machine measurement and analysis of parameters that distinguish a man-made object (i.e., a target) from the natural surroundings. As described in the next section, the high dimensionality of hyperspectral imagery precludes full exploitation of the information through simple visual analysis of the data. In fact, visual representation of the high-dimensional data is itself a challenge, and the essence of hyperspectral detection algorithms is the extraction of information of interest from the high-dimensional data.

Data Representation and Spectral Information Exploitation

In many applications, the basic automated processing problem is to identify pixels whose spectra have a specified spectral shape. This problem raises the following fundamental issues: (1) does a spectrum uniquely specify a material, (2) are the spectra of any material the same when observed at different times, (3) how is the spectrum of a ground material related to the spectrum observed by the sensor, and (4) how should we compare two spectra to decide if they are the same? Answering these questions is a challenging undertaking, involving resources from the areas of spectroscopy, remote sensing, high-dimensional Euclidean geometry, statistics, and signal processing. In this section we make an initial attempt to address these issues by introducing the topics of vector representation of spectra, spectral variability, mixed-pixel spectra, and correlation among spectral bands.

Data Cubes and Spectral Vectors

As a result of spatial and spectral sampling, airborne hyperspectral imaging (HSI) sensors produce a three-dimensional (3D) data structure (with spatial-spatial-spectral components), referred to as a data cube. Figure 3 shows an example of such a data cube. If we extract all pixels in the same spatial location and plot their spectral values as a function of wavelength, the result is the average spectrum of all the materials in the corresponding ground resolution cell. In contrast,
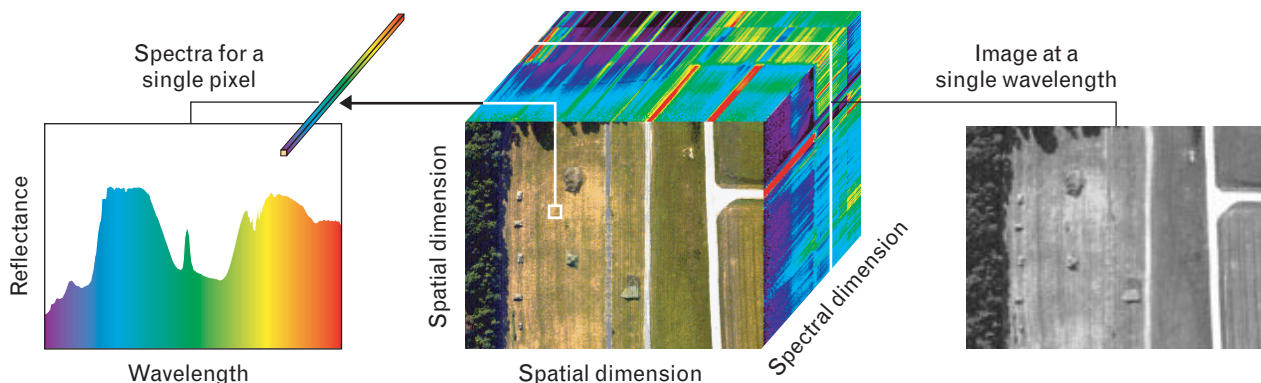

FIGURE 3. Basic data-cube structure (center) in hyperspectral imaging, illustrating the simultaneous spatial and spectral character of the data. The data cube can be visualized as a set of spectra (left), each for a single pixel, or as a stack of images (right), each for a single spectral channel.

the values of all pixels in the same spectral band, plotted in spatial coordinates, result in a grayscale image depicting the spatial distribution of the reflectance of the scene in the corresponding spectral wavelength.

The observed spectral radiance data, or derived apparent surface reflectance data, can be viewed as a scattering of points in an K-dimensional Euclidean space, denoted by $\mathfrak{R}^K$, where K 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 $\mathbf{x} = [x_1, x_2, \ldots, x_K]^T$, where T denotes matrix transposition. The tip of this vector corresponds to an K-dimensional point whose Cartesian coordinates x_i are the radiance or reflectance values at each spectral band. Since each component $x_i \geq 0$, spectral vectors lie inside the positive cone of $\mathfrak{R}^K$. Notice that changes in the level of illumination can change the length of a spectral vector but not its orientation, which is related to the shape of the spectrum. If each material is characterized by a unique deterministic spectrum, which can serve as a spectral fingerprint, we can measure the similarity between two spectra $\mathbf{x}$ and $\mathbf{y}$ by using the Euclidean distance measure

$$\left\| \mathbf{x} - \mathbf{y} \right\| = \sqrt{\sum_{k=1}^{K} \left(x_k - y_k \right)^2}$$

or the spectral angle map (SAM) measure

$$\angle(\mathbf{x}, \mathbf{y}) = \arccos \frac{\mathbf{x}^T \mathbf{y}}{\left\| \mathbf{x} \right\| \left\| \mathbf{y} \right\|},$$

where $\mathbf{x}^T \mathbf{y}$ is the dot product of vectors $\mathbf{x}$ and $\mathbf{y}$. These metrics are widely used in hyperspectral imaging data exploitation, even if there are no optimality properties associated with them.

Exploiting Spectral Correlation

When targets are spatially resolved and have known or expected shapes, the spectral information available in a hyperspectral image may serve to confirm the nature of the target and perhaps provide additional information not available from morphological analysis alone. When targets are too small to be resolved spatially, however, or when they are partially obscured or

of unknown shape (e.g., an oil slick), detection must be performed solely on the basis of the spectral information. The discrimination and detection of different materials in a scene exclusively through the use of spectral information is sometimes termed *nonliteral exploitation*, in reference to the fact that this discrimination and detection does not rely on literal interpretation of an image by means of morphological analysis and inference.

If every material had a unique fixed spectrum and we could accurately measure its value at a fine grid of wavelengths, then we could discriminate between different materials—at least in theory—by comparing their spectra. In this sense, a unique fixed spectrum could be used as a spectral signature of the material. For example, the three types of materials in Figure 2 could be easily discriminated by measuring the spectra in a narrow band around the 1.0-μm wavelength.

The situation is very different in practice, however. Figure 4 shows the range of variation of spectra, from 37 pixels for a vehicle to 8232 pixels for a tree-covered area. Simple inspection of this plot leads to some crucial observations. If we observe a spectrum residing in the red region of the vehicle class where it overlaps with the green spectrum of the tree class, it is almost
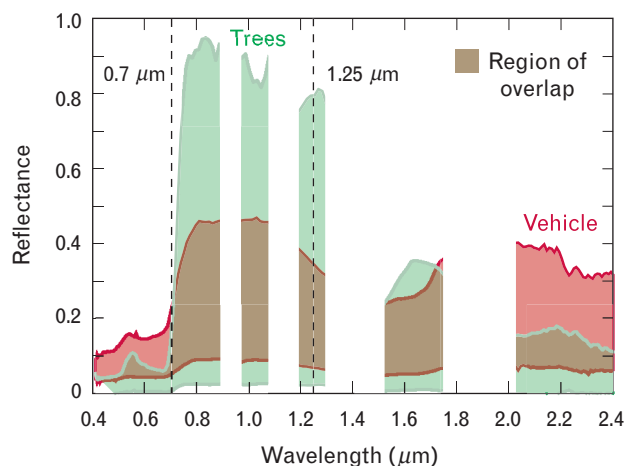

FIGURE 4. Illustration of spectra variability for the *tree* class and the *vehicle* class, where the green and red shaded areas represent the range of observed spectra for trees and vehicles, respectively. At wavelengths where the vehicle spectrum (in red) overlaps the tree spectrum (in green), such as at 0.7 μm and 1.25 μm, it is almost impossible to discriminate between the two classes by using observations from a single spectral band.
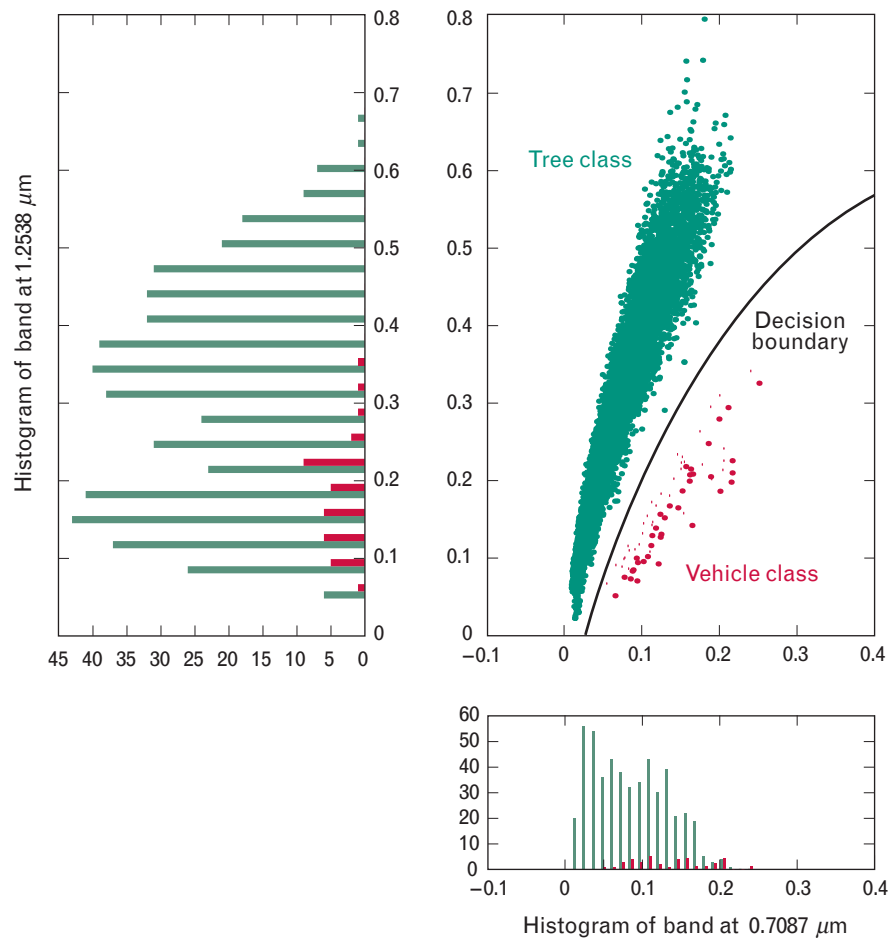
FIGURE 5. Simultaneous exploitation of the spectral bands at 0.7 μm and 1.25 μm in the spectra in Figure 4 makes discrimination possible because of a clearly defined decision boundary between the *tree* class and the *vehicle* class.

impossible to say with certainty if that spectrum corresponds to a vehicle or tree pixel. In actual systems, where we obtain measurements only in certain spectral bands, we would expect discrimination to be even more difficult.

The key question now is this—can we discriminate between vehicle and tree pixels based on spectral observations alone? To answer this question, we select two narrow spectral bands, centered on 0.7 μm and 1.25 μm. Figure 4 shows that it is not possible, using any one band alone, to discriminate between the two classes because they overlap completely. Fortunately, this observation is not the final answer. If we plot the two-band data as points in a two-dimensional space with the reflectance values at the two bands as coordinates, we obtain the scatter plot shown in Figure 5.

Surprisingly, it is now possible to separate the two classes by using a nonlinear decision boundary, despite the complete overlap in each individual band.

The fundamental difference between Figures 4 and 5 is that Figure 5 allows for the exploitation of the extra information in the simultaneous use of two bands. This is the essence of multiple spectral-band data exploitation in remote sensing spectroscopy.

Spectral Variability Models

Unfortunately, a theoretically perfect fixed spectrum for any given material doesn't exist. The spectra observed from samples of the same material are never identical, even in laboratory experiments, because of variations in the material surface. The amount of variability is even more profound in remote sensing ap-

plications because of variations in atmospheric conditions, sensor noise, material composition, location, surrounding materials, and other factors. As a result, measured spectra corresponding to pixels with the same surface type exhibit an inherent spectral variability that prevents the characterization of homogeneous surface materials by unique spectral signatures, as shown in Figure 6(a).

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 that are "seen" by the sensor as a single image pixel, as shown in Figure 6(b). Therefore, depending on the spatial resolution of the sensor and the distribution of surface materials within each ground resolution cell, the result is a hyperspectral data cube comprised of "pure" and "mixed" pixels, where a pure pixel contains a single surface material and a mixed pixel contains multiple materials. Spectral variability and mixed-pixel interference are the main obstacles that need to be addressed and overcome by HSI data-exploitation algorithms.

In the spectral band space, spectra without variability (known as deterministic spectra) correspond to a single fixed point, whereas the tips of vectors corresponding to spectra with variability (known as random spectra) produce a cloud of points. Deterministic spectra are used in spectral libraries as idealized prototype spectral signatures associated with different materials. However, most spectra appearing in real applications are random, and their statistical variability is better described by using probabilistic models.

We next present three families of mathematical models that can be used to characterize the spectral variability of the pixels comprising a hyperspectral data cube. To understand the basic ideas for the different models, we use as an example two spectral bands from a Hyperspectral Digital Imagery Collection Experiment (HYDICE) cube.

Probability Density Models

Figure 7 shows a picture of the ground area imaged by the HYDICE sensor and a scatter plot of the reflectance values for two spectral bands. The scatter plot can help determine areas on the ground that have similar spectral reflectance characteristics. The result shown in the scatter plot is a set of spectral classes that may or may not correspond to distinct well-defined ground-cover classes. This distinction between spectral classes is illustrated in Figure 8, which uses color coding to show several spectral classes in the scatter plot and the corresponding ground-cover areas in the image (these results were obtained by using the *k*-means clustering algorithm with an angular distance metric) [1].
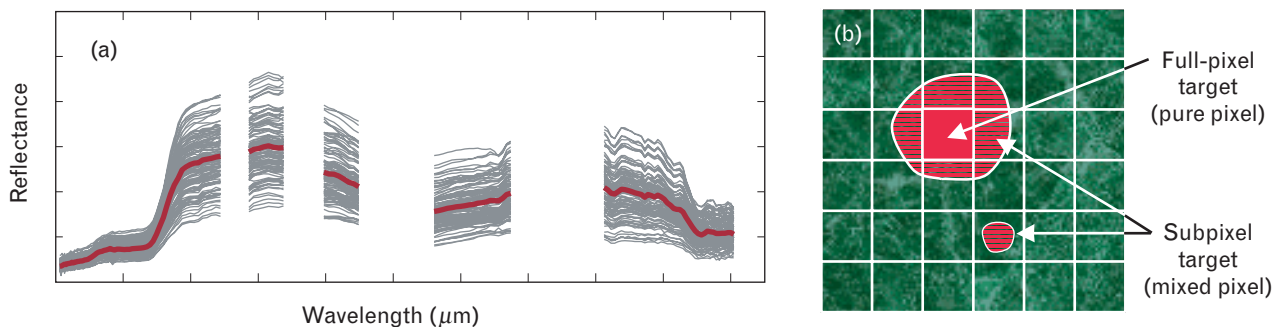

FIGURE 6. Illustration of spectral variability and mixed-pixel interference. (a) Measured spectra corresponding to pixels with the same surface type exhibit an inherent spectral variability that prevents the characterization of homogeneous surface materials with unique spectral signatures. The gaps in the spectra correspond to wavelengths near water absorption bands where the data has been discarded because of low signal-to-noise ratio. (b) Radiance from all materials within a ground resolution cell is seen by the sensor as a single image pixel. Therefore, depending on the spatial resolution of the sensor and the spatial distribution of surface materials within each ground resolution cell, the result is a hyperspectral data cube comprised of pure and mixed pixels, where a pure pixel contains a single surface material and a mixed pixel contains multiple materials. Spectral variability and mixed-pixel interference are the main obstacles that need to be addressed and overcome by hyperspectral data-exploitation algorithms.

A reasonable probabilistic model for such spectral data is provided by the following probability mixture density:

$$f(\mathbf{x}) = \sum_{k=1}^{M} \pi_k f(\mathbf{x}; \mathbf{\theta}_k),$$

where we assume that each observed spectrum $\mathbf{x}$ is generated by first selecting a spectrally homogeneous class k with probability π_k, where $\pi_k \geq 0$ and

$$\sum_{k=1}^{M} \pi_k = 1,$$

and then selecting a spectrum $\mathbf{x}$ according to the conditional probability law $f(\mathbf{x}; \mathbf{\theta}_k)$ specified by the parameter vectors $\mathbf{\theta}_k$.

The component densities $f(\mathbf{x}; \mathbf{\theta}_k)$ are usually chosen as multivariate normal distributions defined by

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{K/2}|\mathbf{\Gamma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{\mu})^T \mathbf{\Gamma}^{-1}(\mathbf{x}-\mathbf{\mu})},$$

where $\mathbf{\mu} \equiv \mathrm{E}\{\mathbf{x}\}$ is the mean vector,

$$\mathbf{\Gamma} \equiv \mathrm{E}\left\{(\mathbf{x} - \mathbf{\mu})(\mathbf{x} - \mathbf{\mu})^T\right\}$$
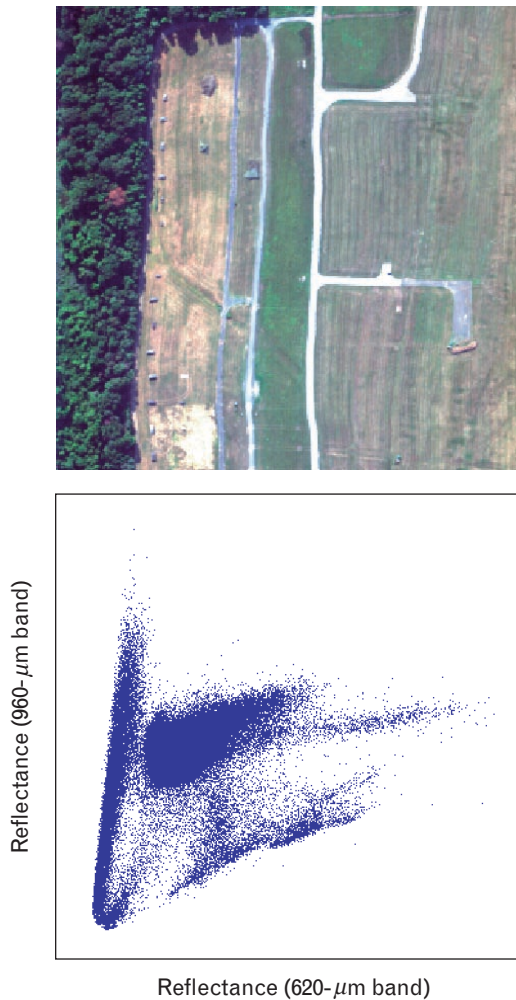

FIGURE 7. Ground image from the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor (top), and a two-dimensional scatter diagram (bottom) illustrating joint reflectance statistics for two spectral bands in the scene.
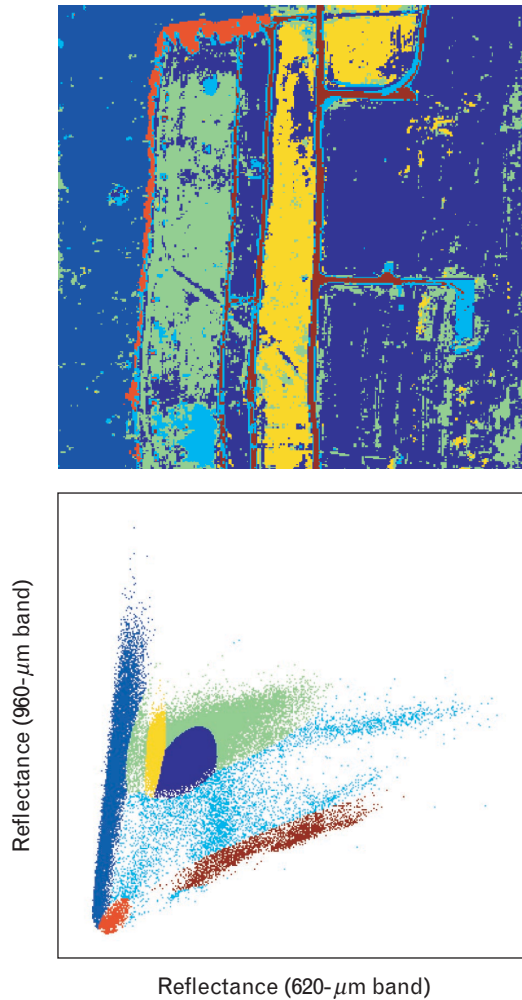

FIGURE 8. Color-coded example of the variety of spectral classes obtained by clustering in the full-dimensional spectral space (top), but illustrated in the scatter plot (bottom) with only two spectral bands.

is the covariance matrix, and $|\Gamma|$ represents the determinant of matrix Γ. We use the notation $\mathbf{x} \sim N(\mu, \Gamma)$ to denote a multivariate normal distribution with mean vector μ and covariance matrix Γ. The quadratic expression in the exponent of the multivariate normal distribution,

$$\Delta^2 = (\mathbf{x} - \mu)^T \Gamma^{-1} (\mathbf{x} - \mu),$$

is a widely used statistical distance measure, known as the *Mahalanobis distance* [2]. This quantity plays a major role in the design and evaluation of target detection algorithms.

Subspace Models

The point specified by the tip of a spectrum vector can be anywhere in the K-dimensional band space. The direction of this vector is specified by the shape of the spectrum and the length of this vector is specified by the amplitude of the spectrum. Therefore, changes in the illumination of a pixel alter the length of its spectral vector but not its direction. In this sense, we can say that the illumination-induced variability is confined into a one-dimensional subspace of the band space.

In general, a *subspace model* restricts the spectrum vector to vary in an M-dimensional subspace of the band space, where $M < K$. Clearly, the variability increases as M increases from one to K. The observed subspace spectrum is described by

$$\mathbf{x} = \sum_{k=1}^{M} a_k \mathbf{s}_k = \mathbf{Sa}. \qquad (1)$$

The vectors $\mathbf{s}_k$ (assumed linearly independent), or equivalently the matrix $\mathbf{S}$ (assumed full rank), provide a basis for the spectral variability subspace. We can introduce randomness to the subspace model in Equation 1 either by assuming that a_k are random variables or by adding a random error vector $\mathbf{w}$ to the right-hand side:

$$\mathbf{x} = \sum_{k=1}^{M} a_k \mathbf{s}_k + \mathbf{w} = \mathbf{Sa} + \mathbf{w}.$$

Subspace variability models can be obtained by using either physical models and the Moderate Resolution Transmittance (MODTRAN) code [3] or statistical approaches such as principal-component analysis [4].

Linear Spectral Mixing Models

The most widely used spectral mixing model is the linear mixing model [5], which assumes that the observed reflectance spectrum, for a given pixel, is generated by a linear combination of a small number of unique constituent deterministic spectral signatures known as *endmembers*. The mathematical equation and the constraints defining the model are

$$\mathbf{x} = \sum_{k=1}^{M} a_k \mathbf{s}_k + \mathbf{w} = \mathbf{Sa} + \mathbf{w}$$

$$\sum_{k=1}^{M} a_k = 1 \qquad \text{(additivity constraint)}$$

$$a_k \geq 0 \qquad \text{(positivity constraint)},$$

where $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_M$, are the M endmember spectra, assumed linearly independent, $a_1, a_2, \ldots, a_M$, are the corresponding abundances (cover material fractions), and $\mathbf{w}$ is an additive-noise vector. Endmembers may be obtained from spectral libraries, in-scene spectra, or geometrical techniques [5].

A linear mixture model based on three or four endmembers has a simple geometrical interpretation as a triangle or tetrahedron whose vertices are the tips of the endmember vectors. In general, spectra satisfying the linear mixture model with both sets of constraints are confined in an K-dimensional simplex studied by the mathematical theory of convex sets [6]. If there are K bands in the mixture model there can be $K + 1$ or fewer endmembers. If $M = K + 1$ there is a unique solution for the fractional abundances; however, we usually assume that $M \leq K$ to allow the use of least-squares techniques for abundance estimation.

Figure 9 uses three-band spectral vectors to illustrate the basic ideas of the probability density, subspace, and linear mixture models for the characterization of spectral variability.

The linear mixture model holds when the materials in the field of view are optically separated so there is no multiple scattering between components. The
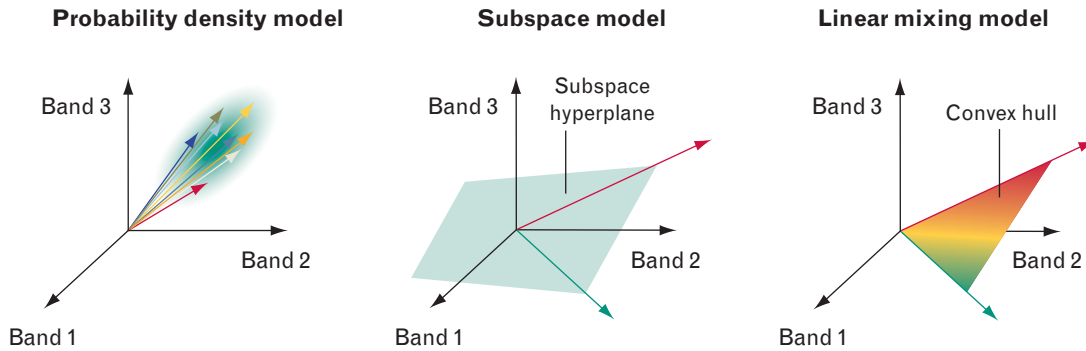
Probability density model   **Subspace model**   **Linear mixing model**

FIGURE 9. Illustration of models for the description of spectral variability. Probability density models provide the probability for a spectrum to be at a certain region of the spectral space. Subspace models specify the linear vector subspace region of the spectral space in which spectral vectors are allowed to reside. Linear mixing models, with the aid of positivity and additivity constraints, permit spectra only in a convex hull of the spectral space.

combined spectrum is simply the sum of the fractional area times the spectrum of each component. However, when different materials are in close contact or close proximity in a scattering surface, such as the mineral grains in a soil or rock, the resulting spectrum—depending on the optical properties of each component—is a highly nonlinear combination of the endmember spectra. In such cases, a nonlinear mixing model is needed to describe the spectrum of mixed pixels.

More elaborate models can be obtained if the endmember spectra are randomly and independently drawn from multivariate normal distributions. Such stochastic mixing models have been developed elsewhere [7], but their applicability is limited because the estimation of their parameters leads to highly nonlinear optimization problems.

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.

Design, Evaluation, and Taxonomy of Target Detectors

In theory, the design and evaluation of detection algorithms is facilitated by assuming some meaningful probability distributions for the target and non-target spectra. The key features of the adopted probabilistic signal models and the used criterion of optimality determine the nature of the obtained detectors and their performance.

Theoretical Design and Performance Evaluation

The mathematical framework for the design and evaluation of detection algorithms is provided by the area of statistics known as binary hypothesis testing. There are several approaches for the systematic design of detection algorithms. However, it is well known that detectors based on the likelihood ratio (LR) test have certain advantages. First, in some sense LR tests minimize the risk associated with incorrect decisions. Second, the LR test leads to detectors that are optimum for a wide range of performance criteria, including the maximization of separation between target and background spectra.

For the purpose of theoretical analysis, spectra are treated as random vectors with specific probability distributions. Given an observed spectrum $\mathbf{x}$ we want to choose between two competing hypotheses:

$$H_0: \quad \text{target absent}$$
$$H_1: \quad \text{target present}.$$

If $p(\mathbf{x}|H_0)$ and $p(\mathbf{x}|H_1)$ are the conditional probability density functions of $\mathbf{x}$ under the two hypotheses, the LR is given by

$$\Lambda(\mathbf{x}) \equiv \frac{p(\mathbf{x}\,|\,\text{target present})}{p(\mathbf{x}\,|\,\text{target absent})}. \qquad (2)$$

If $\Lambda(\mathbf{x})$ exceeds a certain threshold η, then the *target present* hypothesis is selected as true. Basically, the LR test accepts as true the most likely hypothesis. This
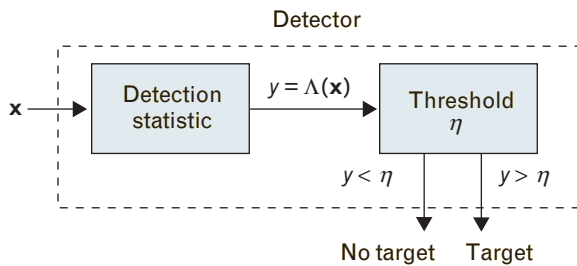
Detector

FIGURE 10. Decomposition of optimum-likelihood-ratio detector. Every detector maps the spectrum **x** of the test pixel (a multidimensional vector) into a scalar detection statistic $y = \Lambda(\mathbf{x})$. Then the detection statistic is compared to a threshold η to decide whether a target is present.

suggests the detector structure shown in Figure 10. The LR test $y = \Lambda(\mathbf{x})$, or any monotonic function of it, provides the information that is used to decide whether a target is present. For this reason, we often refer to this function as a *detection statistic*.

A practical question of paramount importance to a system designer is where to set the threshold to keep the number of detection errors (target misses and false alarms) small and the number of correct detections high. Indeed, there is always a compromise between choosing a low threshold to increase the probability 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 all possible values of the threshold η. Therefore, ROC curves provide the means to evaluate detector performance or compare detectors independently of threshold selection. Figure 11 illustrates the trade-offs in P_D and P_{FA} values for a typical ROC curve. Clearly, any systematic procedure to determine ROC curves or the threshold requires specifying the distribution of the observed spectra **x** under each of the two hypotheses.

If the conditional densities in Equation 2 are completely known (in which case the hypotheses are called *simple* hypotheses), we can choose the threshold to make the detector specified by the LR an optimum detector according to several criteria. The Bayes criterion, used widely in classification applications,

chooses the threshold in a way that leads to the minimization of the probability of detection errors (both misses and false alarms). In detection applications, where the probability of occurrence of a target is very small, minimization of the error probability is not a good criterion of performance, because the probability of error can be minimized by classifying every pixel as background. For this reason, we typically seek to maximize the target probability of detection while keeping the probability of false alarm under a certain predefined value. This approach is the celebrated Neyman-Pearson criterion, which chooses the threshold to obtain the highest possible P_D while keeping $P_{FA} \leq \alpha$. Hence the ROC curve of the optimum Neyman-Pearson detector provides an upper bound for the ROC of any other detector.

In almost all practical situations, the conditional probability densities in Equation 2 depend on some unknown target and background parameters. Therefore, the ROC curves depend on unknown parameters and it is almost impossible to find a detector
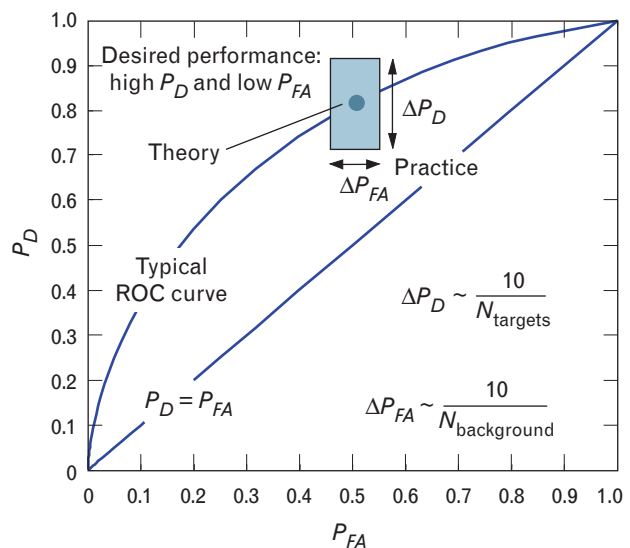

FIGURE 11. The standard performance metric for detection applications is the receiver operating characteristic (ROC) curve, which shows the probability of detection P_D versus the probability of false alarm P_{FA} as the detection threshold takes all possible values. The determination of ROC curves in practice requires the estimation of P_D and P_{FA}. The accuracy of estimating probability as a ratio ($P_A = N_A / N$, where N_A is the number of occurrences of event A in N trials) depends on the size of N. Roughly speaking, the accuracy ΔP of P is in the range of $10/N$.

whose ROC curves remain an upper bound for the whole range of the unknown parameters. A detector with this property is known as a uniformly most powerful (UMP) detector.

An intuitively appealing and widely used approach to handle such situations is to replace the unknown parameters in the LR shown in Equation 2 with their maximum-likelihood estimates. The estimated LR is called the generalized likelihood ratio (GLR), denoted by $\Lambda_G(\mathbf{x})$. The resulting detectors are known as adaptive, or estimate-and-plug, detectors. In general, there are no optimality properties associated with the GLR test. In practice, however, the GLR test leads to adaptive detectors that seem to work well in several applications.

A practical problem in adaptive detection is that the number of pixels in the target class is very small, making the estimation of target density parameters extremely difficult. This limitation is what sets apart detection from classification algorithms in practice. Figure 12 illustrates these differences between theoretical detection and classification problems and their practical versions.

Constant False-Alarm-Rate Detectors

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 in system overloading. Therefore, it is critical to keep the false-alarm rate constant at a desirable level by us-

ing 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. The design of CFAR processors for adaptive detectors is a challenging problem. Figure 13 illustrates the key components and processing steps in a practical CFAR detector. The key problem is obtaining a good model for the background detection statistic.

The desirable performance characteristics of target detection algorithms, in addition to efficient software and effective hardware implementation, are (1) high probability of detection, (2) low probability of false alarm, (3) robustness to deviations from the assumed theoretical model, and (4) CFAR operation under the assumed statistical model. In practice, it is also important to evaluate performance under different tactical scenarios; seasonal changes; targets at various levels of camouflage, concealment, and deception; and different airborne conditions.

A key to quantifying the utility and applicability of hyperspectral sensing to various target detection problems is determining the detection and false-alarm-rate performance of the detection process. Although hyperspectral sensors have been collecting data for many years, three difficulties arise when attempting to derive detector ROC curves: (1) the number of pixels in a hypercube is typically on the order of 10^5, which limits empirical methods of P_{FA} estimation to less than 10^{-4} per cube; (2) the number of targets of a particular type or class in a scene is usu-
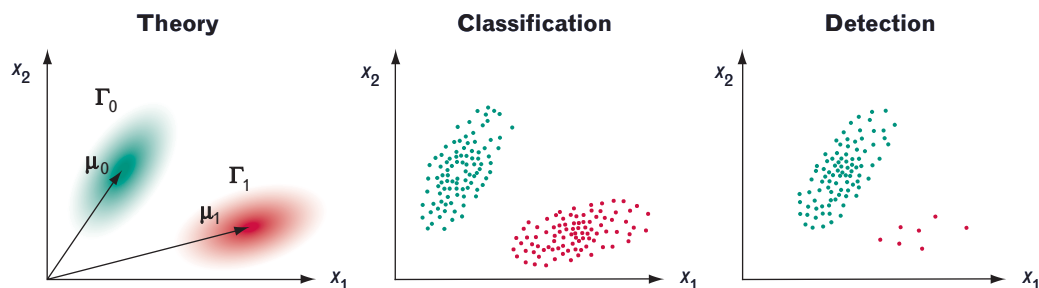

FIGURE 12. In theory, when probability densities for the target and background classes are assumed known, classification and detection problems can both be solved by using the same hypothesis-testing techniques. In practice, however, we need to estimate the statistics of each class from data. This estimation presents some unique challenges for detection applications because the target class is sparsely populated.
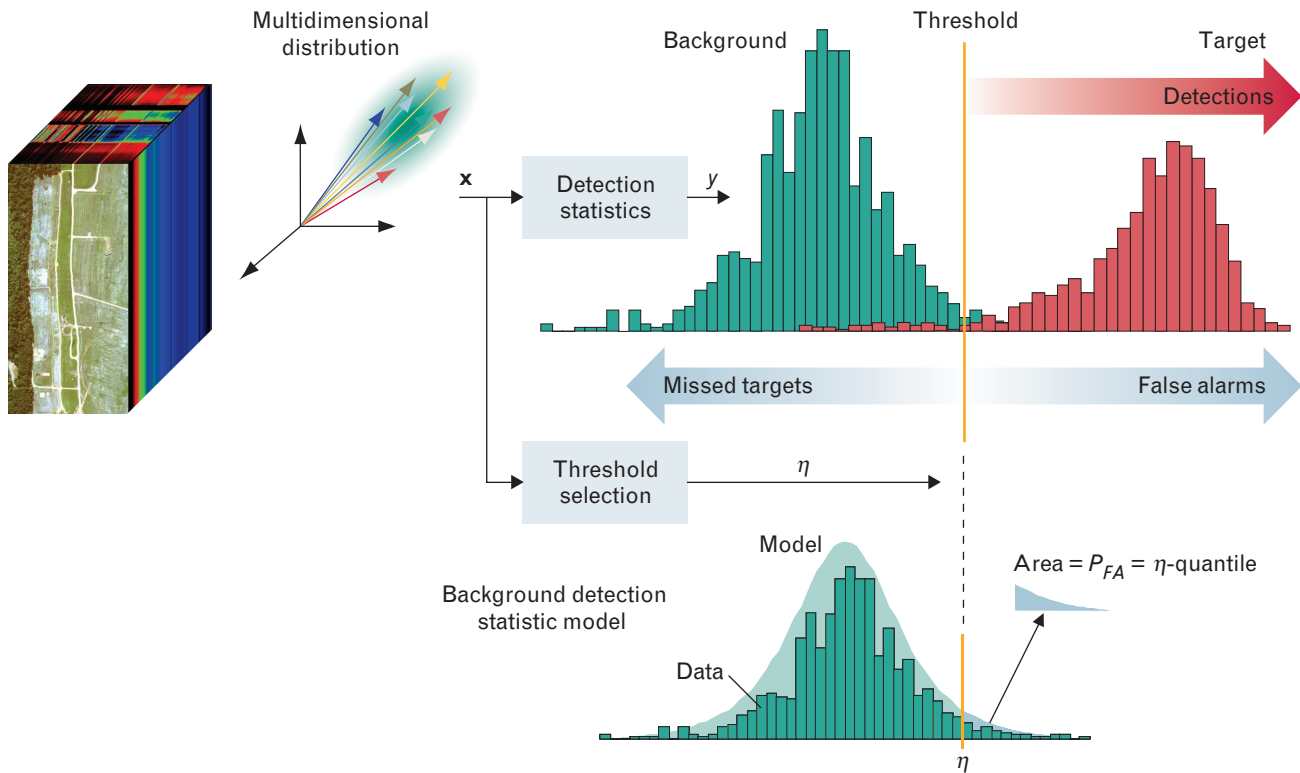
FIGURE 13. A practical constant false-alarm rate (CFAR) adaptive detector maps the spectrum **x** of the test pixel (which is a multidimensional vector) into a scalar detection statistic $y = D(\mathbf{x})$. Then the detection statistics are compared to a threshold η, provided by a CFAR threshold-selection processor, using a background detection statistic model to decide whether a target is present or not present in the pixel data.

ally small, which limits the accuracy that can be achieved in estimating the detection performance; and (3) ground truth for the entire scene is usually limited, which makes confirmation of false alarms difficult (for example, a particular false alarm could be due to variation in the background or it could be due to a spatially unresolved man-made object in the pixel). Regarding the first two of these difficulties, it is well known that as a rule of thumb, the minimum number of pixels N used to estimate a probability P should be $10/P$ or more preferably $100/P$ [8].

Detector Design Road Map

The signal models used to characterize the observed spectra under the target-present and target-absent hypotheses have the most influence upon the design of LR test (optimum) and GLR test (adaptive) detectors. Furthermore, the ability of the models to capture the essential aspects of the data has a direct effect on the performance of the obtained detectors.

For full-pixel targets, there is no significant interaction between target and background other than secondary illumination, shading, and other subtle illumination factors. Hence the spectrum observed by the sensor is produced by either the target spectrum or the background spectrum. In both cases, the observed spectrum is corrupted by additive sensor noise. However, we assume the sensor noise is insignificant or can be accounted for by the target and background distributions.

For subpixel targets, the spectrum observed by the sensor results from a linear or nonlinear mixing of the target and background spectra. The most important consideration is that the background spectrum adds an interference component that makes detection more difficult. The problem of sensor noise is always present as well, and can be incorporated in the target and background signals or modeled as a separate source.

In most 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 non-target class includes almost all pixels in the data cube and is the union of the different specific background classes. We use the term *background* to refer to all non-target pixels of a scene. Usually, targets are man-made objects with spectra that differ from the spectra of natural background pixels.

The detection problem 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, the covariance matrix of the background) that have to be estimated from the data, the detector has to be adaptive, and it is usually designed by using the generalized-likelihood-ratio test (GLRT) approach. As we explain next, however, some special aspects of the detection problem make the estimation of the required model parameters quite challenging.

The sparseness of the target class implies that there are not sufficient data to estimate the parameters of a statistical target model 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 us to use the "unclassified" hyperspectral data cube to statistically characterize the background.

Two issues arise as a result of these observations: (1) how do we specify the parameters of a reasonable target model from *a priori* information, and (2) how do we estimate the parameters of the background model from the observed data? The key question is whether the parameters of the background model remain constant within the region of support used for the estimation.

In the following section we present detection algorithms for full-pixel targets; in this case detection performance is mainly determined by the variability of target and background spectra. Detection algorithms for subpixel targets are treated in a later section. In the subpixel case, detection performance is affected by background interference as well as by the variability of target and background spectra.

Detectors for Full-Pixel Targets

In this section we assume that the pixels for the target and background classes do not contain mixed spectra, and each class can be described by a multivariate distribution with a single set of parameters across the area of interest. First, we present detection algorithms for target and background classes with known statistics. Then we consider the case in which only the background statistics are available. In both cases, we discuss how to estimate the required statistics from available data by using adaptive detectors.

Targets and Backgrounds with Known Statistics

Since statistical decision procedures, based on normal probability models, are simple and often lead to good performance, we model the target and background spectra as random vectors with multivariate normal distributions.

Consider the detection problem specified by the following hypotheses:

$$H_0: \mathbf{x} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Gamma}_0) \quad \text{(target absent)}$$
$$H_1: \mathbf{x} \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Gamma}_1) \quad \text{(target present)}, \tag{3}$$

where the target and background classes follow multivariate normal distributions with different mean vectors and different covariance matrices. For deterministic targets, $\boldsymbol{\Gamma}_1 = 0$ and $\boldsymbol{\mu}_1 = \mathbf{s}$, which implies $\mathbf{x} = \mathbf{s}$ with probability of one. 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 in Equation 2 leads to the following quadratic detector:

$$y = D(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_0)^T \boldsymbol{\Gamma}_0^{-1}(\mathbf{x} - \boldsymbol{\mu}_0)$$
$$- (\mathbf{x} - \boldsymbol{\mu}_1)^T \boldsymbol{\Gamma}_1^{-1}(\mathbf{x} - \boldsymbol{\mu}_1). \tag{4}$$

This detector makes a decision by comparing the Mahalanobis distance of the observed spectrum from the centers of the two target and background classes, as shown in Figure 14. The required threshold η is determined from the formula

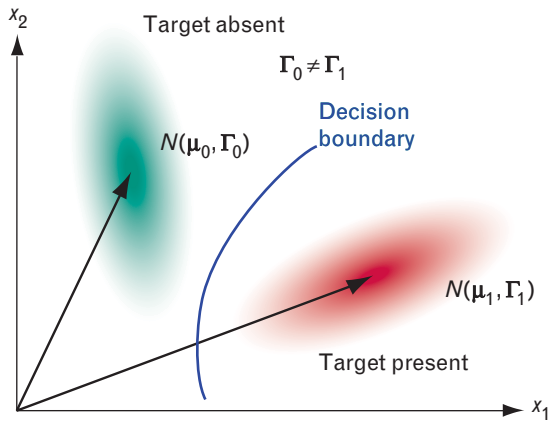$$P_{FA} = \int_{\eta}^{\infty} p(y \,|\, H_0) \, dy = \alpha, \tag{5}$$

FIGURE 14. Illustration of Neyman-Pearson quadratic detector for Gaussian distributions with unequal means and covariances. The decision boundary splits the spectral observation space into two disjoint regions. A detection is declared when spectrum **x** falls in the target-present region.

where α is the desired probability of false alarm. Because of the quadratic form of the Mahalanobis distance, the distribution of the detection statistics y is not normal and therefore difficult to determine.

However, if the target and background classes have the same covariance matrix, that is, $\Gamma_1 = \Gamma_0 \equiv \Gamma$, the quadratic terms in Equation 4 disappear, and the likelihood-ratio detector takes the form of a linear processor

$$y = D(\mathbf{x}) = \mathbf{c}_{MF}^T \mathbf{x},$$

where

$$\mathbf{c}_{MF} = \kappa \Gamma^{-1}(\mu_1 - \mu_0), \tag{6}$$

and where κ is a normalization constant. This detector, which is known as *Fisher's linear discriminant* in the statistical literature [9], is widely used in pattern recognition applications. We use the term *matched filter* (MF), which is typically used in the communications and signal processing disciplines [8, 10]. Figure 15 provides a block diagram representation of a matched-filter–based detector.

A further simplification occurs when the observed spectra have uncorrelated components with equal variances; that is, $\Gamma = \sigma^2 I$. Then the Mahalanobis distance in Equation 4 is reduced to the Euclidean distance. The matched filter becomes a correlation detector that compares the correlations of the input

spectrum with the means of the target and background spectra.

The matched filter in Equation 6 can be also derived by maximizing the following cost function:

$$J(\mathbf{c}) \equiv \frac{\left[E\{y|H_1\} - E\{y|H_0\}\right]^2}{\mathrm{var}\{y|H_0\}}$$

$$= \frac{\left[\mathbf{c}^T(\mu_1 - \mu_0)\right]^2}{\mathbf{c}^T \Gamma \mathbf{c}}, \tag{7}$$

which measures the distance between the means of two normal distributions in units of the common variance. The maximum J_{max}, which is obtained by substituting Equation 6 into Equation 7, is found to be

$$J_{max} = \Delta^2 \equiv (\mu_1 - \mu_0)^T \Gamma^{-1}(\mu_1 - \mu_0), \tag{8}$$

which is the Mahalanobis squared distance between the means of the target and background distributions. The matched filter in Equation 6 with $\kappa = 1/\Delta^2$ minimizes the output variance $\mathbf{c}^T \Gamma \mathbf{c}$, subject to the linear constraint $\mathbf{c}^T \mu_1 = 1$. In the array processing literature this matched filter is known as the minimum-variance distortionless-response beamformer [11].

Although the choice of the normalization factor does not affect the performance of the matched filter, in practice we use the formula

$$y = D(\mathbf{x}) = \frac{(\mathbf{x} - \mu_0)^T \Gamma^{-1}(\mu_1 - \mu_0)}{(\mu_1 - \mu_0)^T \Gamma^{-1}(\mu_1 - \mu_0)} \tag{9}$$

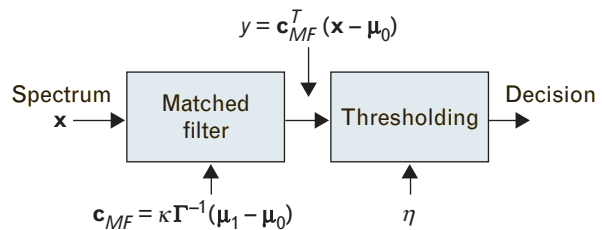because it takes the value $y = 1$ when $\mathbf{x} = \mu_1$.

FIGURE 15. Block diagram of an optimum matched filter detector. The first part is a linear filter that computes the detection statistic for each pixel. The second part compares the detection statistic to a predefined threshold to decide whether a target is present or absent.
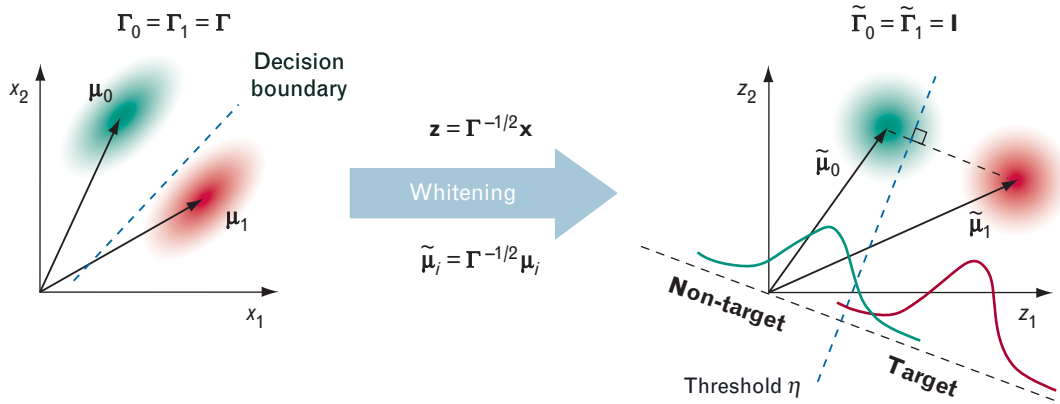
FIGURE 16. Optimum Neyman-Pearson detector for Gaussian distributions with equal covariance matrices. The original spectral observation space is shown at the left, and the whitened space with spherical covariance, created by the whitening transformation $\mathbf{z} = \mathbf{\Gamma}^{-1/2}\mathbf{x}$, is shown at the right.

A detection algorithm, which is often confused with the matched filter given by Equation 6, is the constrained energy minimization (CEM) algorithm [12]. It is obtained by minimizing the total energy of the output of the linear filter

$$y(n) = \mathbf{c}^T \mathbf{x}(n),$$

i.e.,

$$E = \frac{1}{N}\sum_{n=1}^{N} y^2(n)$$

$$= \mathbf{c}^T \left[\frac{1}{N}\sum_{n=1}^{N} \mathbf{x}(n)\mathbf{x}^T(n)\right]\mathbf{c} = \mathbf{c}^T \hat{\mathbf{R}}\mathbf{c},$$

subject to the constraint $\mathbf{c}^T\boldsymbol{\mu}_1 = 1$. The matrix $\hat{\mathbf{R}}$ is the estimated correlation (*not* covariance) matrix of the cube. The CEM filter is given by

$$\mathbf{c}_{\text{CEM}} = \frac{\hat{\mathbf{R}}^{-1}\boldsymbol{\mu}_1}{\boldsymbol{\mu}_1^T\hat{\mathbf{R}}^{-1}\boldsymbol{\mu}_1},$$

and the minimum energy is given by

$$E_{\min} = \frac{1}{\boldsymbol{\mu}_1^T\hat{\mathbf{R}}^{-1}\boldsymbol{\mu}_1}.$$

The CEM algorithm algorithm becomes the matched filter if we remove the mean of the cube from all the data pixels and the reference signature $\boldsymbol{\mu}_1$.

The matched-filter output is the projection of the test spectrum $\mathbf{x}$ along the direction of the parameter vector $\mathbf{c}_{\text{MF}}$. We wish to determine the direction that provides the best separability between the two classes. The direction of the optimum matched filter for a spherical target and background distribution $\mathbf{\Gamma} = \sigma^2 \mathbf{I}$ is along the direction of $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0$. For elliptical distributions this direction is modified by the transformation $\mathbf{\Gamma}^{-1}$. If we use the square-root decomposition $\mathbf{\Gamma} = \mathbf{\Gamma}^{1/2}\mathbf{\Gamma}^{1/2}$, the output of the matched filter is

$$y = \kappa\left[\mathbf{\Gamma}^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\right]^T\left[\mathbf{\Gamma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu}_0)\right].$$

The transformation $\mathbf{z} = \mathbf{\Gamma}^{-1/2}\mathbf{x}$, known as *whitening* or *spherizing*, creates a random vector with spherical distribution. The output of the matched filter is the projection of the whitened spectrum $\mathbf{z}$ along the direction of $\tilde{\boldsymbol{\mu}}_1 - \tilde{\boldsymbol{\mu}}_0$, where $\tilde{\boldsymbol{\mu}}_i = \mathbf{\Gamma}^{-1/2}\boldsymbol{\mu}_i$, $i = 0, 1$. Figure 16 illustrates the operation of the matched filter in the original spectral space and the whitened space. The term "matched" signifies that the detector evaluates the amount of correlation (i.e., matching) between the background-centered reference target signature $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0$ and the test-pixel spectrum $\mathbf{x} - \boldsymbol{\mu}_0$ in the whitened space.

Under the assumption that $\mathbf{x}$ is normal, the output $y = D(\mathbf{x})$ of the matched filter in Equation 9 is 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 Equation 5. Indeed, it can be easily shown that
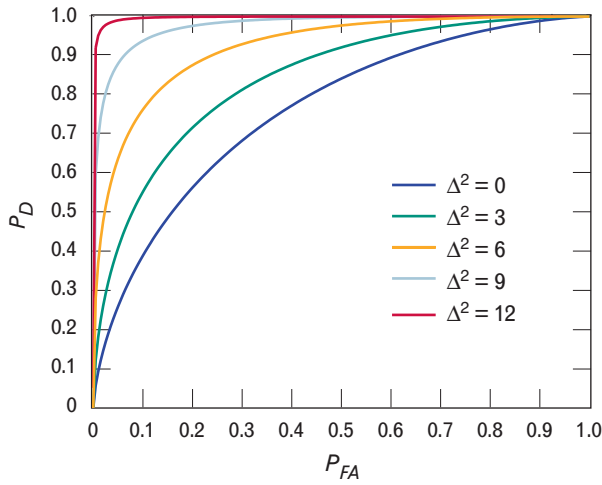
FIGURE 17. ROC curves for the optimum matched filter with the Mahalanobis distance Δ^2 as a parameter.

FIGURE 18. ROC curves for the optimum matched filter with P_{FA} as a parameter.

$$y = D_{MF}(\mathbf{x}) \sim \begin{cases} N(0, \Delta^2), & \text{under } H_0 \\ N(\Delta^2, \Delta^2), & \text{under } H_1 . \end{cases}$$

Therefore, to the degree that the normal model is accurate, the performance of the matched filter is completely determined by the Mahalanobis distance between the target and background distributions.

The probability of detection and the probability of false alarm are related by

$$P_D = Q[Q^{-1}(P_{FA}) - \Delta],$$

where

$$Q(x) = \frac{1}{2\pi} \int_x^\infty \exp\left(-\frac{1}{2} t^2\right) dt$$

is the complementary cumulative distribution function [2, 8].

Figure 17 shows P_D versus P_{FA} ROC curves with Δ^2 as a parameter, whereas Figure 18 shows P_D versus $10\log_{10}(\Delta^2)$ curves with P_{FA} as a parameter.

Anomaly Detection

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 nor the matched filter. We clearly need a statistic exclusively based on $\boldsymbol{\mu}_0$ and
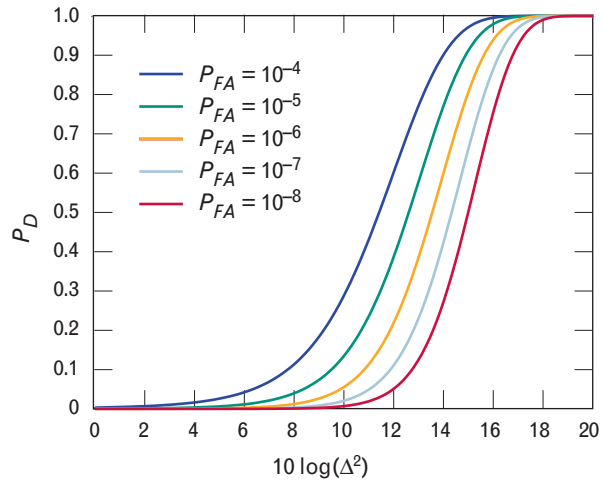
$\boldsymbol{\Gamma}_0$. We consider only the case $\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma}$, which leads to a mathematically tractable problem. The likelihood ratio test for the hypotheses in Equation 3, when only $\boldsymbol{\mu}_0$ and $\boldsymbol{\Gamma}_0$ are known, is given by

$$y = D(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_0)^T \boldsymbol{\Gamma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_0), \quad (10)$$

which is the Mahalanobis distance of the test pixel spectrum from the mean of the background class [2]. This detector, illustrated in Figure 19, is known in the remote sensing literature as the *anomaly detector*. We notice that, even if the two covariance matrices are assumed equal as in the matched-filter case, the lack of knowledge about the target statistics makes the anomaly detector nonlinear. Figure 20 shows a block diagram representation of the anomaly detector given by Equation 10. The distribution of the detection statistic is

$$y = D(\mathbf{x}) \sim \begin{cases} \chi_K^2(0), & \text{under } H_0 \\ \chi_K^2(\Delta^2), & \text{under } H_1 , \end{cases}$$

where $\chi_K^2(\beta)$ is a noncentral chi-squared distribution with K degrees of freedom and noncentrality parameter β. We notice that the performance of the anomaly detector depends on the Mahalanobis distance given by Equation 8 between the target and background distributions. The threshold η is determined by substituting $\chi_K^2(0)$ into Equation 5.

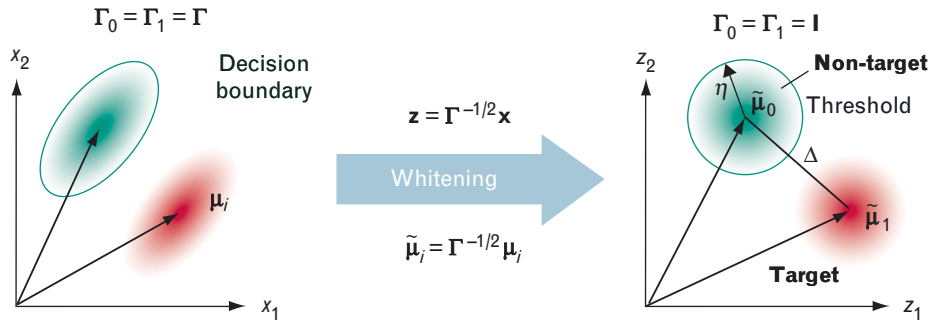Figure 21 shows P_D versus P_{FA} ROC curves with

FIGURE 19. Operation of a CFAR detector for targets with unknown mean and covariance (anomaly detection). This detector is optimum, in the Neyman-Pearson sense, if the unknown target covariance is equal to the covariance of the background class. The threshold, which depends on the orientation in the band space, becomes orientation-independent in the whitened space.

Δ^2 as a parameter for $K = 144$ bands. Figure 22 shows P_D versus P_{FA} ROC curves with $\Delta^2 = 9$ and the number of bands K as a parameter. Figure 23 shows P_D versus $10\log_{10}(\Delta^2)$ curves with P_{FA} as a parameter.

Adaptive Detectors

In practice the background statistics are unknown and they have to be estimated from the data. If the detectors operate in a sparse target environment, we can estimate the mean and covariance of the background by using all pixels within an area of interest. The size of the area is chosen large enough to assure the invertibility of the covariance matrix and small enough to assure spectral homogeneity (stationarity). This estimate-and-plug approach has two consequences: (1) the obtained adaptive detectors lose their Neyman-Pearson optimality, and (2) the detection statistics do not have a normal distribution. Clearly, as the quality of the estimated covariance matrix improves, the performance of the adaptive detectors should approach that of optimum detectors. It has been shown that, if the distribution of the back-

ground and target classes is indeed normal and the covariance matrix is estimated by using $N \geq 3K$ pixels, the loss in detection performance is about 3 dB [13].

Detection Algorithms for Subpixel Targets

By definition, subpixel targets occupy only part of the pixel area. The remaining part of the pixel is filled with one or more materials, which are 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 material mixing, the observed spectral signature can be modeled reasonably well as a linear combination of the
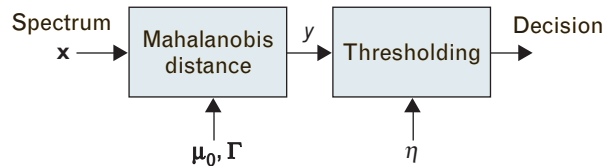

FIGURE 20. Block diagram of optimum anomaly detector for Gaussian distributions with the same covariance matrix and different mean vectors.
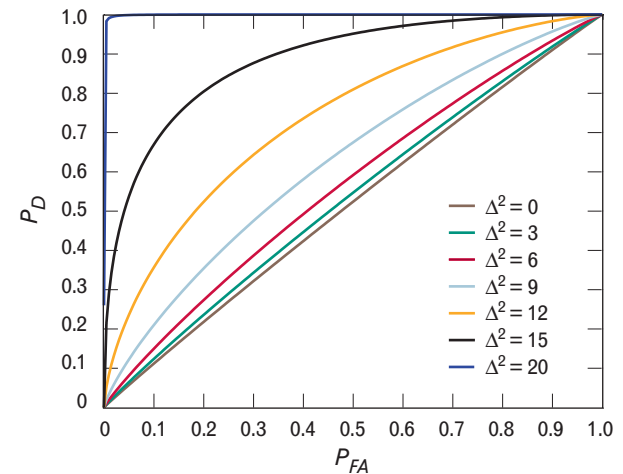

FIGURE 21. ROC curves for the optimum anomaly detector for $K = 144$ bands with the Mahalanobis distance Δ^2 as a parameter.
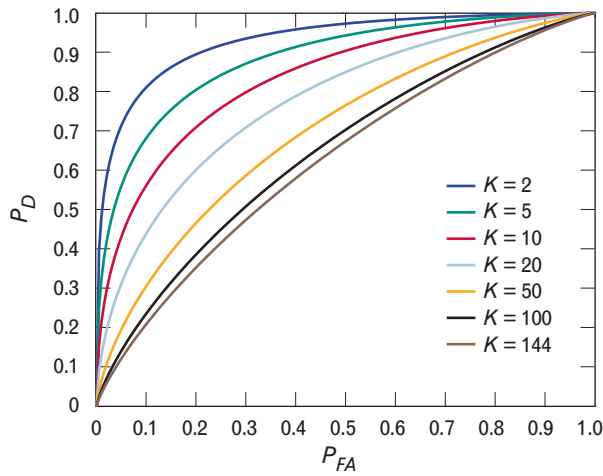
FIGURE 22. ROC curves for the optimum anomaly detector for $\Delta^2 = 9$ with the number of bands K as a parameter.
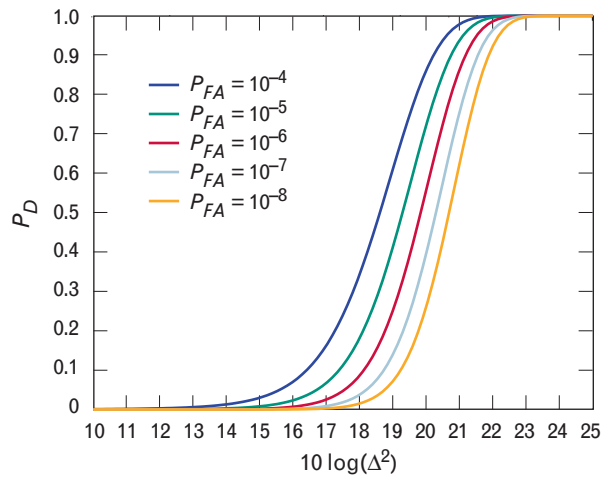

FIGURE 23. ROC curves for the optimum anomaly detector for $K = 144$ bands with P_{FA} as a parameter.

target and background spectra, as shown in Figure 24. Furthermore, there is always an additive amount of noise from various sources (primarily the sensor and the atmosphere).

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 by using a subspace model $\mathbf{Sa}$. If the

FIGURE 24. Linear mixing increases class variability because all mixed-pixel spectra always lie on the line that connects the component spectra. This makes the discrimination more difficult between spectra including both target and background pixels.

columns of $\mathbf{S}$ are endmembers, the vector $\mathbf{a}$ provides their abundances and should satisfy the constraints of the linear mixing model. Otherwise, the vector $\mathbf{a}$ simply determines the position of a target in the column space of $\mathbf{S}$.

The variability of the background can be described by using either a subspace model (structured background) or a statistical distribution (unstructured background). Therefore, the choice of background model leads to two different classes of subpixel target detection algorithms.

Unstructured Background Models

In unstructured background models such as those shown in Figure 25, we assume for convenience that the additive noise has been included in the background $\mathbf{v}$, which in turn is modeled by a multivariate normal distribution with mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Gamma}_0$; that is, $\mathbf{v} \sim N(\boldsymbol{\mu}, \boldsymbol{\Gamma})$. (For simplicity, we drop the subscript 0 from this point forward, and we remove the mean $\boldsymbol{\mu}_0$ from the observations $\mathbf{x}$.) The competing hypotheses are

$$H_0: \quad \mathbf{x} = \mathbf{v} \qquad \text{target absent}$$
$$H_1: \quad \mathbf{x} = \mathbf{Sa} + \mathbf{v} \quad \text{target present}.$$

Hence, $\mathbf{x} \sim N(\mathbf{0}, \boldsymbol{\Gamma})$ under H_0 and $\mathbf{x} \sim N(\mathbf{Sa}, \boldsymbol{\Gamma})$ under H_1. In addition, we assume that we have access to a set of training background pixels $\mathbf{x}(n)$, $1 \leq n \leq N$,
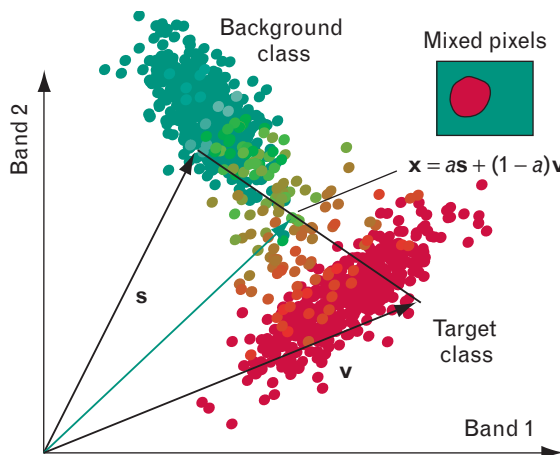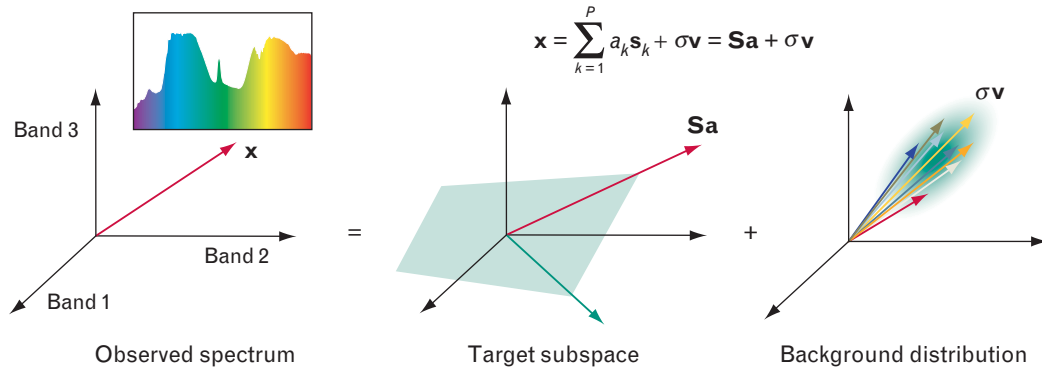
FIGURE 25. Illustration of signal model for subspace targets in homogeneous normal background.

which are independently and identically distributed. The test pixel **x** and the training pixels are assumed statistically independent. It is important to note that, since HSI data have a non-zero mean, we must remove the estimated mean from the data cube and the target subspace vectors to comply with this model.

The key assumptions used for the detectors using the covariance matrix of the background are (1) the background is homogeneous and can be modeled by a multivariate normal distribution, (2) the background spectrum interfering with the test-pixel spectrum has the same covariance matrix with the background training pixels, (3) the test and training pixels are independent, and (4) the target and background spectra interact in an additive manner (additive instead of the replacement model), as illustrated in Figure 24. This means that we do not enforce the additivity and positivity constraints discussed earlier in the section on linear spectral mixture models.

Using the GLR approach, E.J. Kelly obtained the following detector [14, 15]:

$$D_\kappa(\mathbf{x}) = \frac{\mathbf{x}^T\hat{\mathbf{\Gamma}}^{-1}\mathbf{S}(\mathbf{S}^T\hat{\mathbf{\Gamma}}^{-1}\mathbf{S})^{-1}\mathbf{S}^T\hat{\mathbf{\Gamma}}^{-1}\mathbf{x}}{N + \mathbf{x}^T\hat{\mathbf{\Gamma}}^{-1}\mathbf{x}} \mathop{\gtrless}_{H_0}^{H_1} \eta_\kappa, \quad (11)$$

where $\hat{\mathbf{\Gamma}}$ is the maximum-likelihood estimate of the covariance matrix

$$\hat{\mathbf{\Gamma}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^T(n).$$

Although there is no optimality criterion associated

with the GLR test approach [8], 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 **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 **S** and becomes minimum when $P = K$. In this case, we simply know that we are looking for deterministic targets that lie in the data subspace. Since the matrix **S** has full rank and therefore is invertible, Equation 11 leads to the following detector:

$$D_A(\mathbf{x}) = \mathbf{x}^T\hat{\mathbf{\Gamma}}^{-1}\mathbf{x} \mathop{\gtrless}_{H_0}^{H_1} \eta_A, \quad (12)$$

which was derived by Kelly [16] using the approach presented here, and by I.S. Reed and X. Yu [17] using a multivariate analysis of variance formulation. Basically, $D_A(\mathbf{x})$ estimates the Mahalanobis distance of the test pixel from the mean of the background, which is zero for demeaned data. The algorithm in Equation 12, which has the CFAR property, is the adaptive version of Equation 10 and it is used extensively for anomaly detection in multispectral and hyperspectral imaging applications [7].

A key assumption in the derivation of Equation 11 was that the covariance matrix of the background is the same under the two hypotheses. For subpixel targets, however, the amount of background covered area is different under the two hypotheses. Therefore, it is more appropriate to use the following hypotheses

H_0: $\mathbf{x} = \mathbf{v}$ target absent

H_1: $\mathbf{x} = \mathbf{Sa} + \sigma\mathbf{v}$ target present ,

implying $\mathbf{x} \sim N(\mathbf{0}, \sigma^2\mathbf{\Gamma})$ under H_0 and $\mathbf{x} \sim N(\mathbf{Sa}, \sigma^2\mathbf{\Gamma})$ under H_1. In other words, the background has the same covariance structure under both hypotheses 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 [18, 19]:

$$D_{ACE}(\mathbf{x}) =$$

$$\frac{\mathbf{x}^T\hat{\mathbf{\Gamma}}^{-1}\mathbf{S}(\mathbf{S}^T\hat{\mathbf{\Gamma}}^{-1}\mathbf{S})^{-1}\mathbf{S}^T\hat{\mathbf{\Gamma}}^{-1}\mathbf{x}}{\mathbf{x}^T\hat{\mathbf{\Gamma}}^{-1}\mathbf{x}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{ACE} , \quad (13)$$

which can be obtained from Equation 11 by removing the additive term N from the denominator.

If we use the adaptive whitening transformation

$$\mathbf{z} \equiv \hat{\mathbf{\Gamma}}^{-1/2}\mathbf{x} ,$$

where $\hat{\mathbf{\Gamma}} = \hat{\mathbf{\Gamma}}^{1/2}\hat{\mathbf{\Gamma}}^{1/2}$ is the square-root decomposition of the estimated covariance matrix, the ACE can be expressed as

$$D_{ACE}(\mathbf{x}) = \frac{\mathbf{z}^T\tilde{\mathbf{S}}(\tilde{\mathbf{S}}^T\tilde{\mathbf{S}})^{-1}\tilde{\mathbf{S}}^T\mathbf{z}}{\mathbf{z}^T\mathbf{z}} = \frac{\mathbf{z}^T\mathbf{P}_{\tilde{\mathbf{S}}}\mathbf{z}}{\mathbf{z}^T\mathbf{z}} ,$$

where $\tilde{\mathbf{S}} \equiv \hat{\mathbf{\Gamma}}^{-1/2}\mathbf{S}$ and $\mathbf{P}_{\tilde{\mathbf{S}}} \equiv \tilde{\mathbf{S}}(\tilde{\mathbf{S}}^T\tilde{\mathbf{S}})^{-1}\tilde{\mathbf{S}}^T$ is the or-

thogonal projection operator onto the column space of matrix $\tilde{\mathbf{S}}$. Since $\mathbf{P}_{\tilde{\mathbf{S}}}^2 = \mathbf{P}_{\tilde{\mathbf{S}}}$, we can write Equation 13 as

$$D_{ACE}(\mathbf{x}) = \frac{\left\|\mathbf{P}_{\tilde{\mathbf{S}}}\mathbf{z}\right\|^2}{\left\|\mathbf{z}\right\|^2} = \cos^2\theta ,$$

which shows that, in the whitened coordinate space, $y = D_{ACE}(\mathbf{x})$ is equal to the cosine square of the angle between the test pixel and the target subspace.

With the whitening transformation, the anomaly detector in Equation 12 can be expressed as

$$D_A(\mathbf{x}) = \mathbf{z}^T\mathbf{z} ,$$

which is the Euclidean distance of the test pixel from the background mean in the whitened space. We note that, in the absence of a target direction, the detector uses the distance from the center of the background distribution.

For targets with amplitude variability, we have $P = 1$ and the target subspace $\mathbf{S}$ is specified by the direction of a single vector $\mathbf{s}$. Then the formulas for the previous GLR detectors are simplified to

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

where $(\psi_1 = N, \psi_2 = 1)$ for the Kelly detector and $(\psi_1 = 0, \psi_2 = 1)$ for the ACE. Kelly's algorithm was derived for real-valued signals and has been applied to multispectral target detection [20]. The one-dimen-
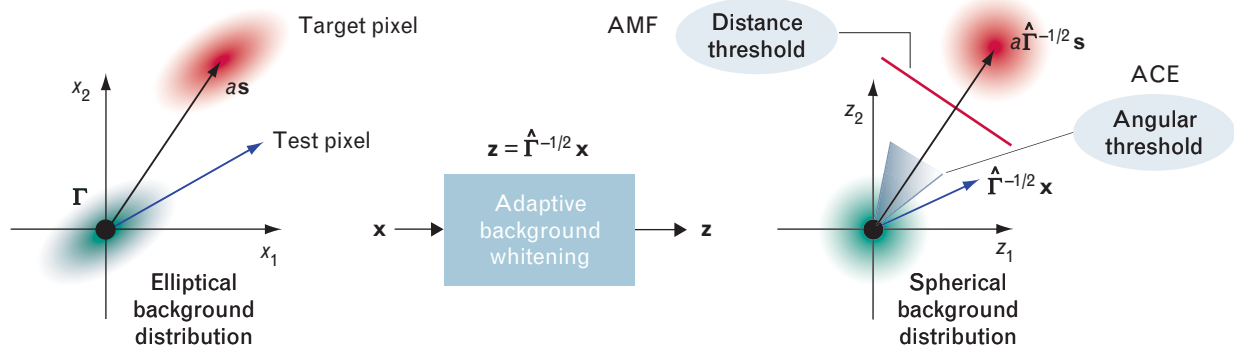

FIGURE 26. Illustration of generalized-likelihood-ratio test (GLRT) detectors. These are the adaptive matched filter (AMF) detector, which uses a distance threshold, and the adaptive coherence/cosine estimator (ACE) detector, which uses an angular threshold. The test-pixel vector and target-pixel vector are shown after the removal of the background mean vector.
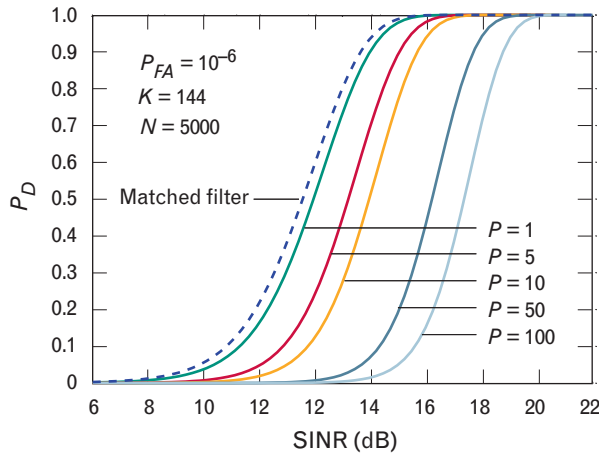
FIGURE 27. ROC curves for probability of detection P_D as a function of SINR, showing the effect of target subspace dimensionality.
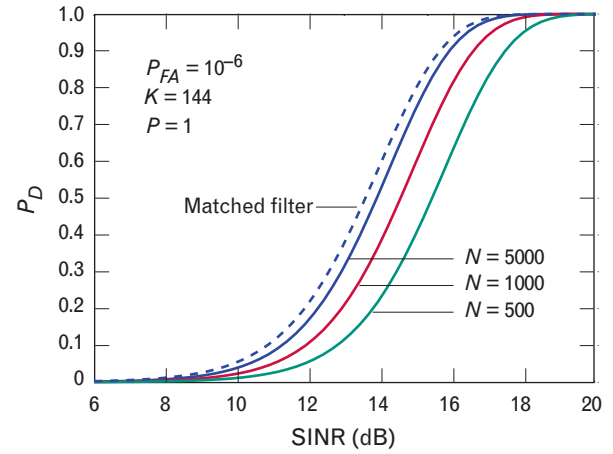

FIGURE 28. ROC curves for probability of detection P_D as a function of SINR, showing the effect of the number of training pixels.

sional version of ACE has been derived elsewhere [21, 22]. Finally, we note that if ($\psi_1 = N$, $\psi_2 = 0$) we obtain the adaptive matched filter (AMF) detector [23, 24].

The length of a spectral vector increases or decreases when the overall illumination increases or decreases, but its angular orientation remains fixed. Equivalently, the shape of the spectrum remains the same, but its amplitude changes proportionally to the illumination. For this reason, angle distances play an important role in HSI data processing. These issues, and the operation of angle-versus-distance-based detectors, are illustrated in Figure 26.

Determining the distribution of the various GLRT detectors is an elaborate process [14, 16, 17, 19, 25]. 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) given by

$$ \text{SINR}_0 = (\mathbf{Sa})^T \boldsymbol{\Gamma}^{-1} (\mathbf{Sa}) . $$

The performance of all GLRT detectors depends only on the dimensional integer numbers K, P, and N, and the optimum SINR_0 parameter. Under the H_0 hypothesis (target absent), $\text{SINR}_0 = 0$, the output distribution becomes central, and the probability of false alarm depends only on the parameters K, P, and N. Therefore, all GLRT detectors discussed in this section have the CFAR property. For a remote sensing system with high signal-to-noise ratio, the detection

performance depends predominately on the fraction of the target occupying the pixel, since this fraction determines the SINR. Figure 27 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 $P = K$, which corresponds to the adaptive anomaly detector in Equation 12. Figure 28 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 [8] (i.e., a known target in noise with known covariance matrix) for comparison.

Structural Background Models

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

$$ \begin{aligned} H_0: \quad & \mathbf{x} = \mathbf{Ba}_{b,0} + \mathbf{w} && \text{target absent} \\ H_1: \quad & \mathbf{x} = \mathbf{Sa} + \mathbf{Ba}_{b,1} + \mathbf{w} && \text{target present}, \end{aligned} \tag{14} $$

where the $K \times P$ matrix $\mathbf{S}$ has to be specified by the user, the $K \times Q$ matrix $\mathbf{B}$ is determined from the data,
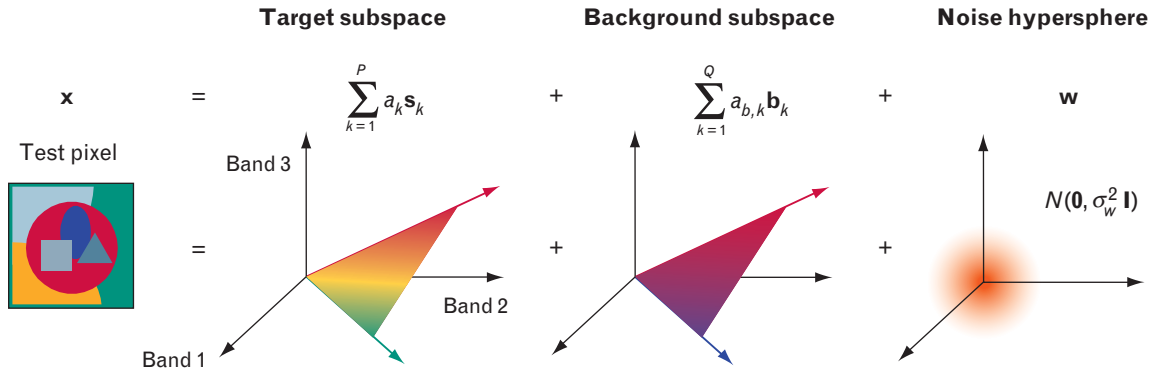
FIGURE 29. Illustration of structured background model when we use linear mixing with positivity and additivity constraints to describe the target and background subspaces (only the part within a convex hull of each subspace is allowed). For unconstrained linear mixing or statistically determined subspaces, the full subspace ranges are permissible.

and $\mathbf{w} = N(\mathbf{0}, \sigma_w^2 \mathbf{I})$. Figure 29 provides a geometrical illustration of the structured background model. Since we assume that the coefficient vectors $\mathbf{a}$, $\mathbf{a}_{b,0}$, and $\mathbf{a}_{b,1}$ are unknown constants, the only source of randomness is the additive noise $\mathbf{w}$ with unknown variance σ_w^2.

If we compute the maximum-likelihood estimates of the unknown quantities and substitute the likelihood ratio, we obtain the following GLRT detector [8, 26, 27]:

$$y = D_{SB}(\mathbf{x}) = \frac{\mathbf{x}^T \mathbf{P}_B^\perp \mathbf{x}}{\mathbf{x}^T \mathbf{P}_{SB}^\perp \mathbf{x}} = \frac{\left\| \mathbf{P}_B^\perp \mathbf{x} \right\|^2}{\left\| \mathbf{P}_{SB}^\perp \mathbf{x} \right\|^2}, \quad (15)$$

where $\mathbf{P}_A^\perp = \mathbf{I} - \mathbf{P}_A$ and $\mathbf{P}_A = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is the projection matrix onto the column space of matrix $\mathbf{A}$. The matrix $\mathbf{SB} = [\mathbf{S}\ \mathbf{B}]$ denotes the matrix obtained by combining the target and background subspaces. We note that $\|\mathbf{P}_B^\perp \mathbf{x}\|$ and $\|\mathbf{P}_{SB}^\perp \mathbf{x}\|$ are the Euclidean distances of the test pixel from the background subspace and the combined target and background subspaces, respectively (see Figure 30).

From the orthogonal triangle TCB we have $\cos\phi = $ TC/TB, which implies that $1/D_{SB}(\mathbf{x}) = \cos^2\phi$. That is, the detection statistic is based on the angle between the perpendiculars drawn from the test pixel to the target and the combined target-background subspaces. Because any monotonic function of $D_{SB}(\mathbf{x})$ can be used as a detection statistic, we often prefer the ratio

$$y = D_{SB}(\mathbf{x}) = \frac{\mathbf{x}^T \mathbf{P}_B^\perp \mathbf{x} - \mathbf{x}^T \mathbf{P}_{SB}^\perp \mathbf{x}}{\mathbf{x}^T \mathbf{P}_{SB}^\perp \mathbf{x}}$$

$$= \frac{(\text{CB})^2}{(\text{CT})^2} = \tan^2\phi \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{SB} \, ,$$

because its probability density function can be conveniently expressed in terms of the F-distribution. In-
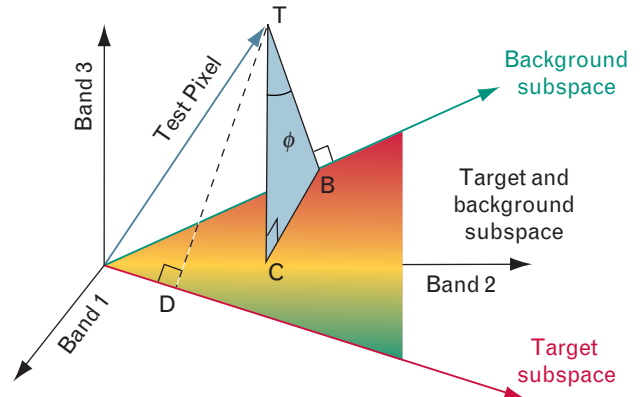

FIGURE 30. Geometrical illustration of GLRT subspace target detector. The length of TC measures the goodness of fit for the full linear model (target plus background) to the test-pixel spectrum $\mathbf{x}$, whereas the length of TB measures the goodness of fit for the reduced model (background). If TC $\cong$ TB, the inclusion of target does not improve the fit, and we decide that the target is not present. The length of TD measures the goodness of fit of the target model to the test-pixel spectrum; however, for subpixel targets, comparing TD to TB cannot be used to decide whether a target is present or not present.

deed, the distribution of the detection statistics is given by

$$D(\mathbf{x}) \sim \frac{P F_{P,K-P-Q}(\text{SINR}_0)}{K - P - Q},$$

where $F_{P,K-P-Q}(\text{SINR}_0)$ is noncentral F-distribution with P numerator degrees of freedom, $K - P - Q$ denominator degrees of freedom, and noncentrality parameter SINR_0:

$$\text{SINR}_0 = \frac{(\mathbf{Sa})^T \mathbf{P}_B^{\perp}(\mathbf{Sa})}{\sigma_w^2},$$

which depends on the unknown variance σ_w^2. Notice that σ_w^2 is not required for the computation of the detection statistics because it cancels out. In the statistical literature $D_{SB}(\mathbf{x})$ is denoted by $F(\mathbf{x})$ and is known as the F test. The threshold η_{SB} is established by the required probability of false alarm

$$P_{FA} = 1 - F_{P,K-P-Q}(0, \eta_{SB}),$$

because $\text{SINR}_o = 0$ under H_0. [With a slight abuse of notation we use $F_{P,K-P-Q}(\text{SINR}_0, \eta_{SB})$ to denote the cumulative distribution function of the noncentral F random variable.] Since the distribution of $D(\mathbf{x})$ under H_0 is known, we can set the threshold η_{SB} to attain CFAR operation. The probability of detection

$$P_D = 1 - F_{P,K-P-Q}(\text{SINR}_0, \eta_{SB})$$

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

It can be shown that, independent of the normality assumption, the detector in Equation 15 maximizes the SINR [8, 10]. Furthermore, it has been proven that "within the class of invariant detectors which have the same invariances as the GLRT, the GLRT is uniformly most powerful (UMP) invariant. This is the strongest statement of optimality one could hope to make for a detector" [26, p. 2156].

When $P = 1$, it can be shown [27] that the amplitude of the target signature in Equation 15 can be estimated by the formula

$$\hat{a} = \frac{\mathbf{s}^T \mathbf{P}_B^{\perp} \mathbf{x}}{\mathbf{s}^T \mathbf{P}_B^{\perp} \mathbf{s}} \sim N[a, \sigma_w^2(\mathbf{s}^T \mathbf{P}_B^{\perp} \mathbf{s})^{-1}]. \qquad (16)$$

Detection algorithms based on estimating $\hat{a}$ do not have the CFAR property, because σ_w^2 is unknown, and they are not optimum in the Neyman-Pearson sense. We notice that the detection algorithm known as the orthogonal subspace projector (OSP) [28] is just the numerator of Equation 16. The GLRT detector in Equation 15 and the OSP detector use the same signal model given in Equation 14. The GLRT detector maximizes the SINR for any noise distribution and is CFAR in the case of Gaussian noise. The OSP has none of these properties.
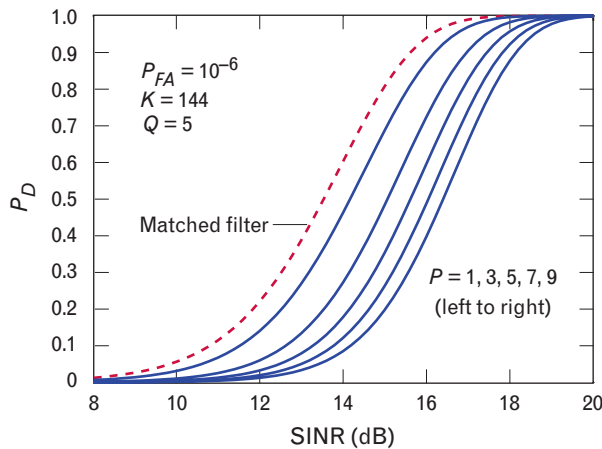

FIGURE 31. ROC curves for probability of detection P_D as a function of signal-to-interference-plus-noise ratio (SINR), illustrating the effect of target subspace dimensionality P.
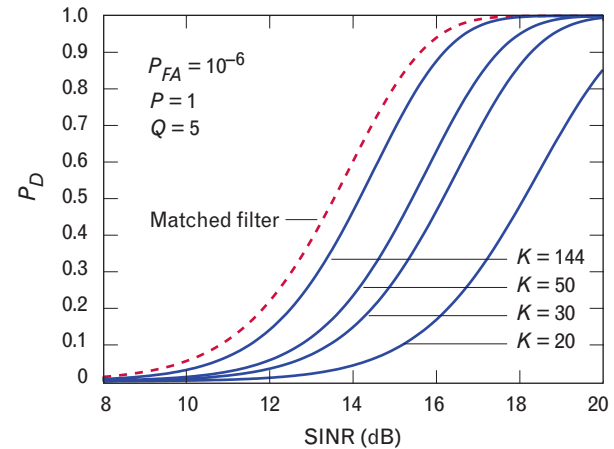

FIGURE 32. ROC curves for probability of detection P_D as a function of SINR, illustrating the effect of the number K of spectral bands.

To gain additional insight into the performance of the adaptive GLRT subspace detector, we consider a background subspace with $Q = 5$, $K = 144$ bands, and the probability of false alarm fixed at $P_{FA} = 10^{-6}$. The dimension of the target subspace is varied from $P = 1$ to $P = 9$ in steps of 2. Figure 31 shows plots of the probability of detection as a function of the SINR for fixed values of P_{FA}, Q, P, and K. 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. We see that for a given probability of detection, there is a loss in SINR due to the need to estimate the target abundance and the noise variance.

Figure 32 shows a similar family of curves parameterized with respect to the number of spectral bands K, with all other parameters held fixed. As we intu-itively expect, performance improves as the number of bands increases (that is, as we supply more information into the detector). Clearly, the improvement is more dramatic when we start adding more bands and becomes less significant after a certain point. The performance of the optimum-matched-filter detector does not depend on the number of bands. More details and experimental results regarding the performance of detection algorithms based on subspace background models can be found elsewhere [27].

As is evident from the preceding discussion, a variety of hyperspectral imaging detection algorithms have been reported in the literature, many with a heritage in radar detection algorithms. Since the notation used by the various authors is not consistent, and since the connection to earlier algorithm work in radar is not always noted, we have created a taxonomy of detection algorithms, employing a common nota-

Table 2. Taxonomy of Hyperspectral Imaging Target Detection Algorithms

		Signal Model	Assumptions	Detector $y = D(\mathbf{x})$	Name	Comments
Full pixel targets		$H_0: \mathbf{x} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Gamma}_0)$ $H_1: \mathbf{x} \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Gamma}_1)$	Known $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \boldsymbol{\Gamma}_0, \boldsymbol{\Gamma}_1$	$(\mathbf{x} - \boldsymbol{\mu}_0)^T \boldsymbol{\Gamma}_0^{-1}(\mathbf{x} - \boldsymbol{\mu}_0) -$ $(\mathbf{x} - \boldsymbol{\mu}_1)^T \boldsymbol{\Gamma}_1^{-1}(\mathbf{x} - \boldsymbol{\mu}_1)$	Bayes or Neyman-Pearson quadratic classifiers	$\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_0 \Rightarrow$ $y = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \boldsymbol{\Gamma}^{-1} \mathbf{x}$ Plug-in detectors: estimate $\boldsymbol{\mu}_1, \boldsymbol{\mu}_0, \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_0$ from training data
			Training data for $\boldsymbol{\mu}_0, \boldsymbol{\Gamma}_0$ only	$(\mathbf{x} - \hat{\boldsymbol{\mu}})^T \hat{\boldsymbol{\Gamma}}^{-1}(\mathbf{x} - \hat{\boldsymbol{\mu}})$	Mahalanobis distance	Anomaly detection RX algorithm CFAR
Target Resolution	Subpixel targets	$H_0: \mathbf{x} = \mathbf{v}$ $H_1: \mathbf{x} = \mathbf{Sa} + \sigma\mathbf{v}$ $\mathbf{S}$ known $(K \times P)$ matrix $H_0: \mathbf{x} \sim N(0, \ )$ $H_1: \mathbf{x} \sim N(\mathbf{Sa}, \sigma^2\ )$	Training data under $H_0: \mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(N)$ $\hat{\boldsymbol{\mu}} = \dfrac{1}{N}\sum_{n=1}^{N} \mathbf{x}(n)$ $\hat{\boldsymbol{\Gamma}} = \dfrac{1}{N}\sum_{n=1}^{N}[\mathbf{x}(n) - \hat{\boldsymbol{\mu}}][\mathbf{x}(n) - \hat{\boldsymbol{\mu}}]^T$ $\mathbf{x} \leftarrow (\mathbf{x} - \hat{\boldsymbol{\mu}}); \quad \mathbf{s}_k \leftarrow (\mathbf{s}_k - \hat{\boldsymbol{\mu}})$	$\dfrac{\mathbf{x}^T\hat{\boldsymbol{\Gamma}}^{-1}\mathbf{S}(\mathbf{S}^T\hat{\boldsymbol{\Gamma}}^{-1}\mathbf{S})^{-1}\mathbf{S}^T\hat{\boldsymbol{\Gamma}}^{-1}\mathbf{x}}{\psi_1 + \psi_2\mathbf{x}^T\hat{\boldsymbol{\Gamma}}^{-1}\mathbf{x}}$	Name ψ_1 ψ_2 σ Kelly N 1 1 AMF 1 0 1 ACE 0 1 ? ? = unknown AMF = adaptive matched filter ACE = adaptive coherence/cosine estimator	CFAR $\text{SINR}_o = (\mathbf{Sa})^T \boldsymbol{\Gamma}^{-1}(\mathbf{Sa})$ $y = \kappa\, \mathbf{s}^T\hat{\boldsymbol{\Gamma}}^{-1}\mathbf{x}$ (matched filter) Low-rank covariance matrix approximation: $\hat{\boldsymbol{\Gamma}} \approx \sum_{k=1}^{Q} \lambda_k \mathbf{u}_k \mathbf{u}_k^T$ $\Rightarrow \hat{\boldsymbol{\Gamma}}^{-1} \approx \sum_{k=1}^{Q} \dfrac{1}{\lambda_k} \mathbf{u}_k \mathbf{u}_k^T$
			$P = 1 \Rightarrow \mathbf{S} \to \mathbf{s}$	$\dfrac{\mathbf{s}^T\hat{\boldsymbol{\Gamma}}^{-1}\mathbf{x}}{(\mathbf{s}^T\hat{\boldsymbol{\Gamma}}^{-1}\mathbf{s})^{1/2}(\psi_1 + \psi_2\mathbf{x}^T\hat{\boldsymbol{\Gamma}}^{-1}\mathbf{x})^{1/2}}$		
		$H_0: \mathbf{x} = \mathbf{Ba}_{b,0} + \mathbf{w}$ $H_1: \mathbf{x} = \mathbf{Sa} + \mathbf{Ba}_{b,1} + \mathbf{w}$	$\mathbf{w} \sim N(0, \sigma_w^2\mathbf{I})$ $\mathbf{P}_A^{\perp} = \mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$	$\dfrac{\mathbf{x}^T(\mathbf{P}_B^{\perp} - \mathbf{P}_{SB}^{\perp})\mathbf{x}}{\mathbf{x}^T\mathbf{P}_{SB}^{\perp}\mathbf{x}}$	GLRT detector	CFAR $\text{SINR}_o = \dfrac{\left\|\mathbf{P}_B^{\perp}\mathbf{Sa}\right\|^2}{\sigma_w^2}$
				$\mathbf{s}\mathbf{P}_B^{\perp}\mathbf{x}$	Orthogonal subspace projector (OSP)	Non-CFAR

tion, by using the following criteria: (1) the spectral composition of the observed pixels (full pixels or mixed pixels); (2) the mathematical approach used to model the spectral variability of target and background spectra (subspace or statistical); and (3) the available *a priori* information about target or background classes (signature-based detection versus anomaly detection).

Table 2 lists this taxonomy of the most fundamental target detection algorithms discussed in this article. We focus on algorithms derived with well-established statistical models and procedures. A number of additional algorithms reported in the literature can be obtained from this table through mathematical approximations or simplifications. For example, low-rank approximations or diagonal loading of the estimated background covariance matrix lead to algorithms that may provide improved performance under certain practical circumstances. The use of un-supervised classification techniques to derive more homogeneous background segments followed by subsequent application of the algorithms in Table 2 may or may not lead to improved detection performance.

Practical Performance Evaluation

The successful implementation of automatic hyperspectral target detection algorithms requires contending with many practical details and challenges that arise because of departures of the data from the theoretical assumptions made in deriving the various algorithms. For example, the scene reflectances may not conform to the assumed probability distribution and, even if they do, atmospheric effects, sensor errors, and sampling artifacts can corrupt the data.

Table 3 provides a list of desirable features for target detection algorithms, as well as several performance metrics that need to be quantified when we evaluate these algorithms for practical applications. In this section, we touch upon some of these issues to provide a flavor of the difficulties encountered in the practical performance assessment of HSI target detection algorithms.

Hyperspectral Imaging Data and Ground Truth

Airborne hyperspectral imagery data collected by the HYDICE [29] sensor at the U.S. Army Aberdeen Proving Grounds, Maryland, on 24 August 1995 are used to illustrate various issues and algorithms discussed in this article. HYDICE collects calibrated (post-processed) spectral radiance data in 210 wavelengths spanning 0.4 to 2.5 μm in nominally 10-nm-wide bands. The image in Figure 33 is a color representation of this data. We also use data from the NASA Hyperion sensor to illustrate the detection of

---

Table 3. Detection Algorithm Performance Metrics and Assessment

Desirable Features	*Empirical Evaluation*

High P_D and low P_{FA}	Detection performance
	    target-background separation
	    detected target pixels at a given P_{FA}
	    P_{FA} to detect all target pixels

Robustness to model mismatches—similar performance under similar situations	

	Sensitivity to
	    externally supplied parameters
	    covariance corruption by target pixels
	    covariance regularization
	    dimensionality reduction

Ease of use—minimal user interaction	
Automated threshold selection	
Reasonable robustness to supplied and estimated parameters	
	Performance under different tactical scenarios; seasonal changes; targets at various levels of camouflage, concealment, and deception; and different airborne conditions
Acceptable computational complexity	

Classes selected for statistical analysis

Class name	Selection technique	Sample size
Grass	Spatially adjacent	7760
Trees	Spatially adjacent	8232
Mixed	Spatially adjacent	7590
Blocks 1–8	Spatially adjacent	25,000

FIGURE 33. Color image representing the hyperspectral data from the HYDICE sensor data cube, showing the division of the data cube into rectangular blocks to reduce spatial inhomogeneity. These data are used to illustrate the empirical performance of different target detection algorithms.

subpixel targets. The Hyperion sensor covers the range from 0.4 to 2.5 μm, with 220 wavebands and a spatial resolution of thirty meters.

A quantitative evaluation of the performance of HSI detectors using real data is complicated. We would like to have calibrated data free of sensor arti-

facts, since these artifacts degrade detector performance. Furthermore, we would like to have comprehensive *a priori* knowledge of all the materials in the scene in order to score the performance of the various detectors. This comprehensive *a priori* knowledge, or *ground truth*, involves information such as precise location, size, and orientation of the materials in the scene. For spatially resolved target materials, some pixels will be entirely filled by the material, some may be partially filled, and some may be in shadow. Therefore, for the purpose of assessing the performance of automated detection algorithms, it is useful to carefully review all the data and manually label the pixels in the vicinity of a target as full, mixed, shadow, and guard pixels, to distinguish the performance differences among different detectors. This approach involves developing target masks, as illustrated in Figure 34, which are created by a laborious manual review of the imagery. Figure 35 shows the reflectance spectra of the four types of pixels specified by the target mask in Figure 34. The mean value of the full-pixel spectra is shown by a thick black line in all four plots.

The data were analyzed in units of calibrated spectral radiance for the characterization of joint statistics and surface reflectance for the modeling of target detection statistics. While it is known that data artifacts exist in the calibrated spectral radiance because of an
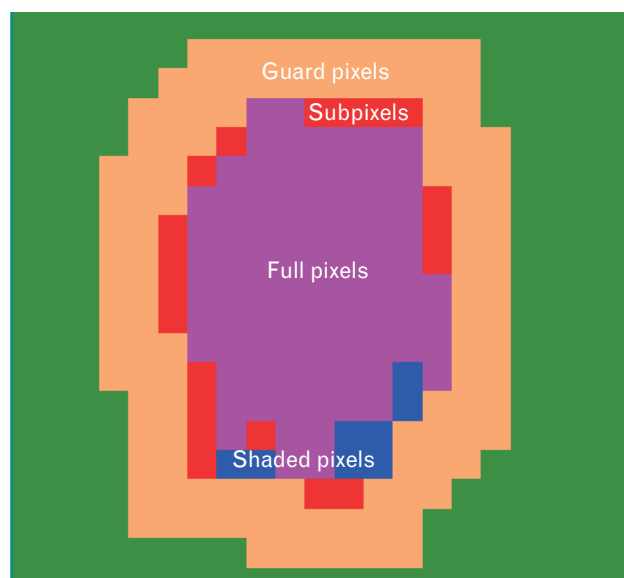

FIGURE 34. Example of a target mask, illustrating the various types of pixels identified in the canonical data sets.
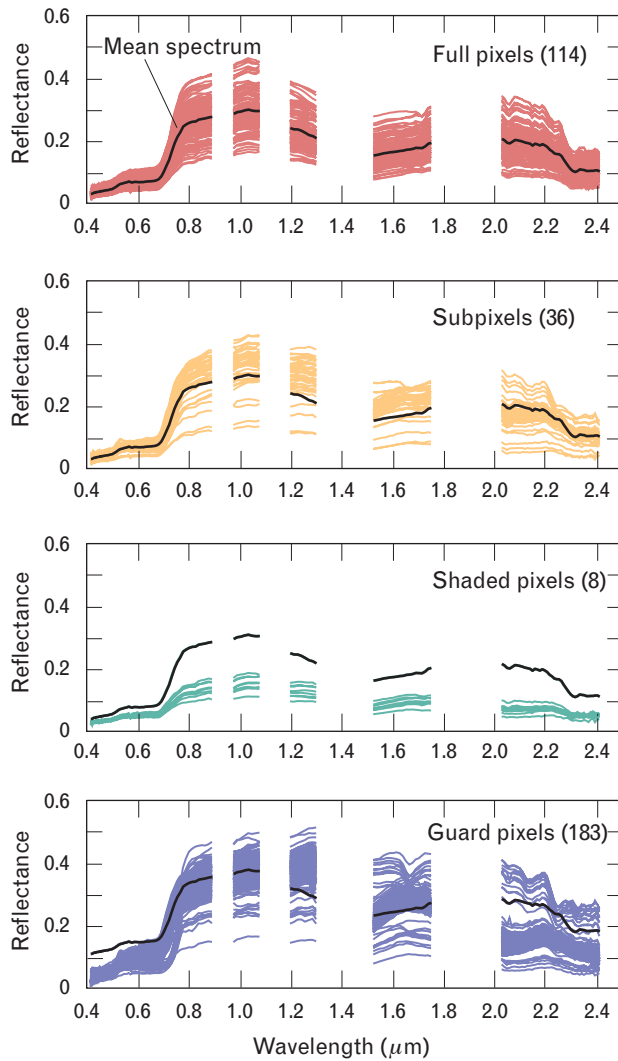
FIGURE 35. Reflectance spectra of the four types of pixels specified by a target mask such as the one shown in Figure 33. The thick line shown in all plots is the mean spectrum of the full-target pixels.

interpolation process used to replace dead pixels, no special screening for these anomalies was performed. For the multivariate analysis examples, only 144 of the 210 channels were used to avoid regions of low signal-to-noise ratio (water-vapor bands).

Empirical Investigation of Detection Performance

We now present some typical results regarding the performance of different detection algorithms, using hyperspectral imaging data obtained by the HYDICE and Hyperion sensors for both full-pixel and subpixel targets.

As we have already mentioned, the preeminent performance metric for target detection applications is an ROC curve. However, the relatively modest size of the available data cubes (about 2×10^5 pixels) and the very small number of target pixels (usually fewer than 10^2 pixels) make the use of reliable ROC curves questionable. To avoid this problem, we used the plots illustrated in Figure 36. The top plot shows the probability of exceedance as a function of the detector output (detection statistics). The probability of exceedance $P_y(\eta)$ is defined as the probability that the value of a random variable y with probability density function $f_y(y)$ exceeds a threshold η, such that

$$ P_y(\eta) = \int_\eta^\infty f_y(y)\, dy = 1 - \int_\infty^\eta f_y(y)\, dy = 1 - F_y(\eta), $$

where $F_y(\eta)$ is the cumulative distribution function.

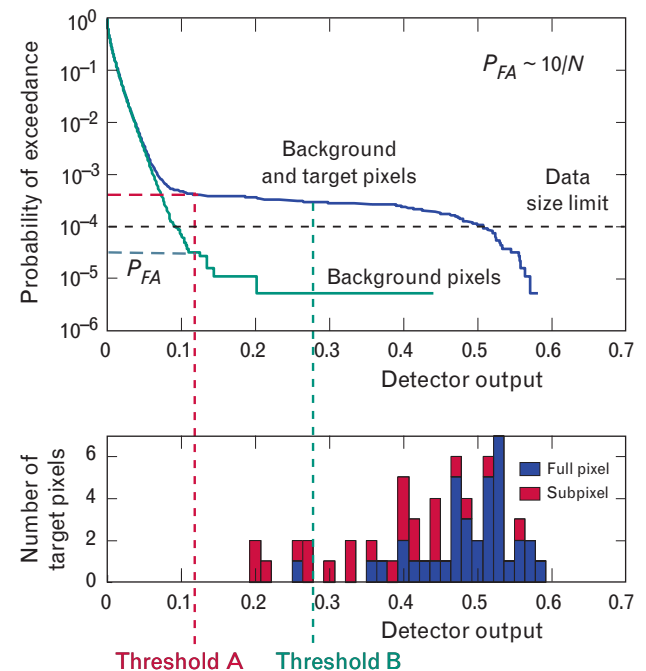Since ground truth is available, we can exclude the

FIGURE 36. Example of plots used to evaluate detection performance. The top plot shows the probability of exceedance as a function of threshold. This information can be used to select a threshold for a desired false-alarm rate. The histogram in the lower plot shows the number of target pixels as a function of detector statistics, for both full-pixel and subpixel targets.
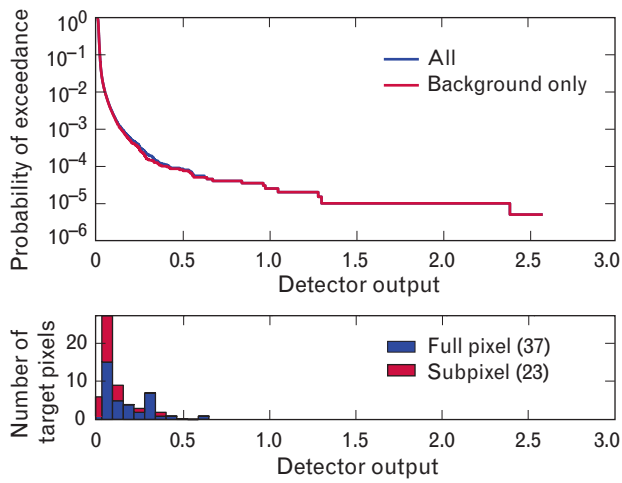
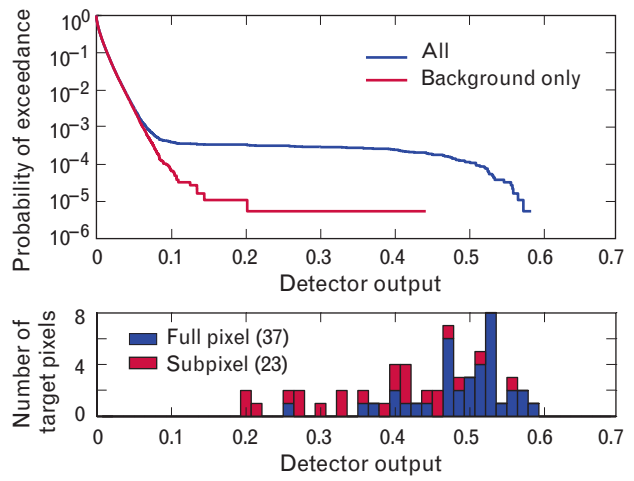**FIGURE 37.** Performance of Mahalanobis distance-based anomaly detector for target A.

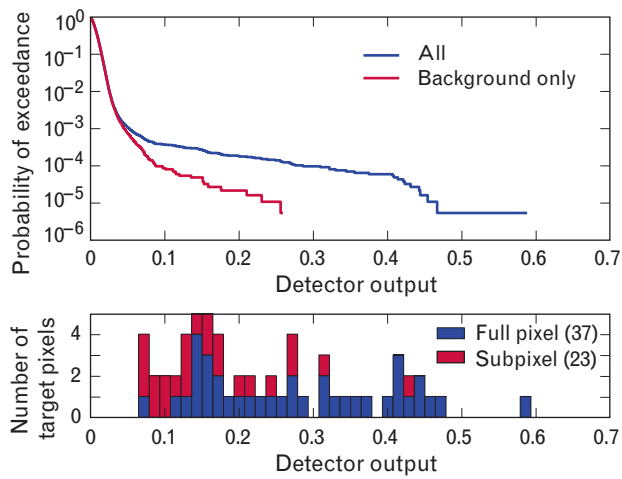**FIGURE 38.** Performance of the ACE detector for target A.
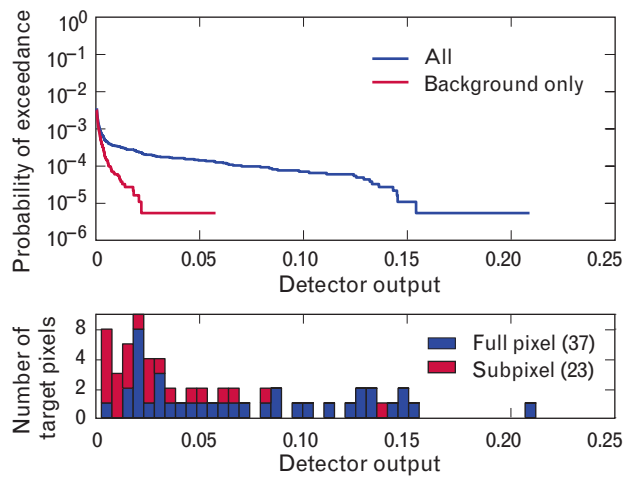
FIGURE 39. Performance of the AMF detector for target A.

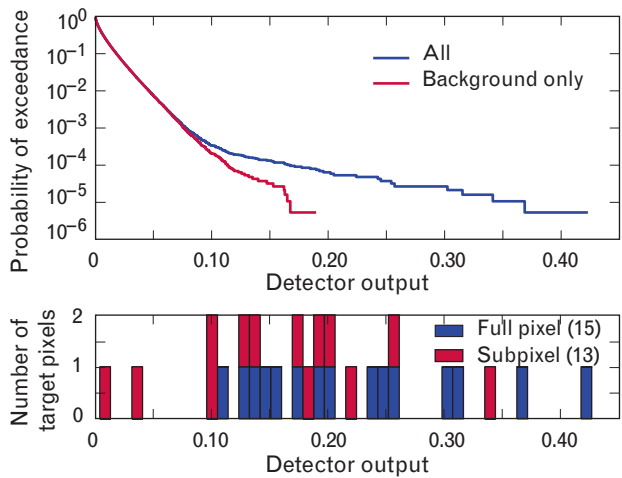**FIGURE 40.** Performance of the GLRT detector for target A.

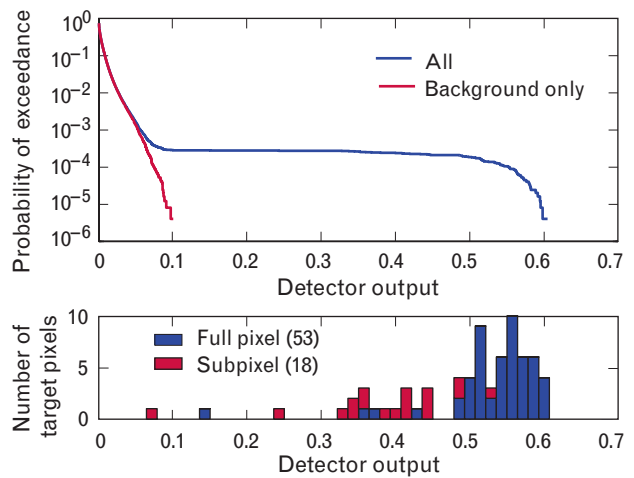**FIGURE 41.** Performance of the ACE detector for target B.

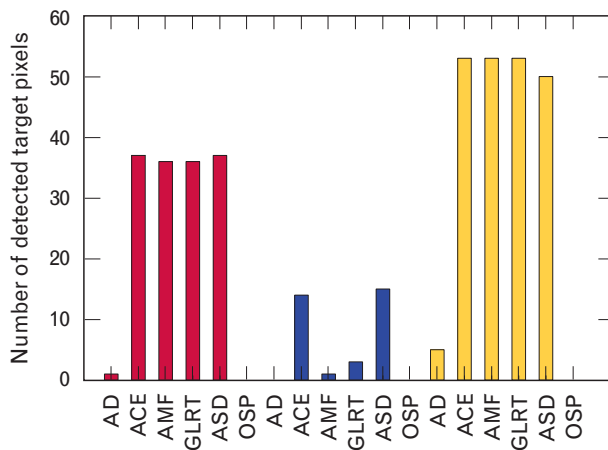**FIGURE 42.** Performance of the ACE detector for target C.

FIGURE 43. Summary of detection performance of the various detectors for three different target signatures. Each bar shows the number of detected targets for a 10^{-4} probability of false alarm.
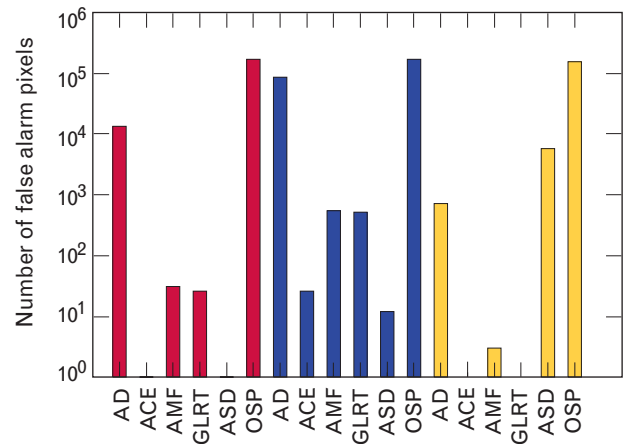

FIGURE 44. Summary of detection performance of the various detectors for three different target signatures. Each bar shows the probability of false alarm for a threshold that assures the detection of all full-pixel targets.

target pixels. The exceedance curve with targets excluded shows the probability of false alarm as a function of threshold and can be used to select a threshold for a desired false-alarm rate. Inspection of exceedance curves with and without target pixels included shows that even a small number of targets can change the tail of the distribution significantly. This reinforces the argument that estimates of P_{FA} are inaccurate when $P_{FA} < 10/N$, where N is the number of available pixels. This limit is more profound when we deal with the probability of detection, where the number of target pixels may be as low as one pixel. Since stating probabilities could be misleading in such cases, we plot a histogram, such as the lower plot in Figure 36, which shows the number of target pixels as a function of the detector output, including both full pixel and subpixel targets.

From Figure 36 we can conclude that the performance of a detector improves as the body of the target histogram moves toward and beyond the tail of the background probability of exceedance curve. In this sense, careful inspection of Figures 37 through 40 indicates the following: (1) the three detectors (ACE, AMF, GLRT), searching for a target A with known spectral signature, outperform the Mahalanobis distance-based anomaly detector (AD), which does not use the spectral signature of target A; (2) the ACE algorithm performs better than the AMF and GLRT

algorithms because its target histogram is mostly located outside the range of the background probability of exceedance curve; (3) the performance for subpixel targets is inferior to that for full-pixel targets. Indeed, the response for full-pixel targets typically exceeds that for subpixel targets in all detectors.

Figures 41 and 42 illustrate the performance of the ACE detector for two additional targets, labeled B and C, in the same Forest Radiance I background. As Figure 41 shows, target B is more difficult to detect than target A (Figure 38) and target C (Figure 42).

To provide a synoptic performance comparison of the main algorithms discussed in this article, Figure 43 shows the number of correctly detected target pixels for three spectrally different targets. The threshold in each case was chosen to attain a false-alarm rate of 10^{-4}. Similarly, Figure 44 shows the resulting false-alarm rate when the threshold is chosen to assure the detection of all target pixels in each case. These figures indicate that the ACE algorithm typically provides better performance than any other algorithm. Although this performance may not be significantly better than other algorithms, it is consistent across different targets and backgrounds.

Another point of practical significance, which becomes apparent from Figures 38 through 42, is the difficulty of selecting a threshold to attain a given probability of false alarm when $P_{FA} < 10/N$. In order

to set the CFAR threshold in a real-time system, the data must conform to some underlying statistical model, and the parameters of the model must be inferred from the data. We have seen that the presence of target pixels (even a very small number) lifts the tail of the exceedance curve significantly. Clearly, estimating the CFAR threshold using data containing targets will lead to the choice of an unnecessarily high threshold. There are remedies to this problem, such as methods that exclude outliers in estimating CFAR threshold. A more important consideration, however, is the accuracy of the underlying statistical model.

Figure 45 shows the actual background statistics and the theoretically predicted statistics for the ACE detector algorithm, assuming that the hyperspectral data follow a multivariate normal distribution. We see that there is a significant discrepancy at the tail of the distributions, where the model tends to underestimate the false-alarm rates for thresholds above 0.1. Therefore, we need models that can provide a more accurate description of the background statistics at very low probabilities of false alarm. This topic is addressed in the next section.

The reader should be warned that the results presented are indicative of the typical performance we
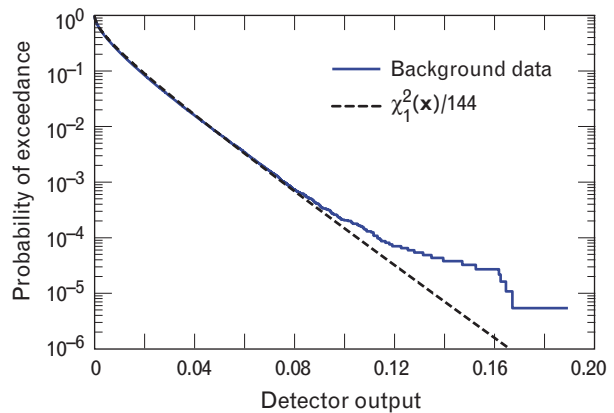

FIGURE 45. Experimental and theoretically expected background probabilities of exceedances for the ACE detector.

have seen in several experiments. It is possible that fine-tuning an algorithm may provide better performance in certain situations; however, it is not guaranteed that this improvement will hold in general.

Evaluation of detection performance for subpixel targets is even more challenging because it is costly to deploy many targets. Furthermore, detection is more difficult if the target fills only a small fraction of the pixel—say, less than fifty percent. Consider, for example, the Hyperion hyperspectral sensor, which has
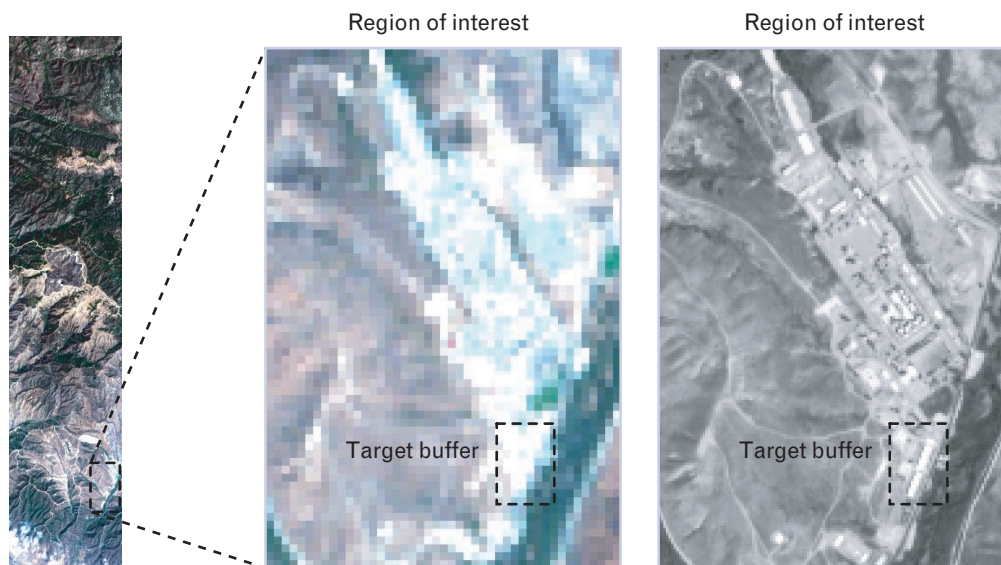

FIGURE 46. Camp Pendleton, California, test area, imaged by the Hyperion sensor (left). The two detailed views of the target site are a color image (center), indicative of the spatial resolution of the Hyperion data sensor, and a high-resolution panchromatic image (right). The desired target material of interest is a rooftop in the scene, as shown in the target buffer area.

a spatial resolution of 30 × 30 meters. An automobile, which covers an area less than ten square meters, would be very difficult to detect with such a sensor, since the pixel fill factor is only 11%. The article by John P. Kerekes and Jerrold E. Baum in this issue, "Hyperspectral Imaging System Modeling," examines the sensitivity to fill factor in more detail.

Figure 46 shows an area from the Camp Pendleton, California, test site, imaged with the NASA Hyperion sensor. The target-site area, shown in a color image and in a high-resolution panchromatic image, includes a rooftop that constitutes the desired target. The ground truth regarding the location of the target was derived from Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor data for the same site, as shown in Figure 47. We define a 10 × 6 pixel buffer in the Hyperion data that is certain to include the target pixels. The spectral signature for the rooftop, required by the three detectors, was obtained from the AVIRIS data cube.

To compare the three detectors (ACE, AMF, and GLRT), we compute the detection statistics for the
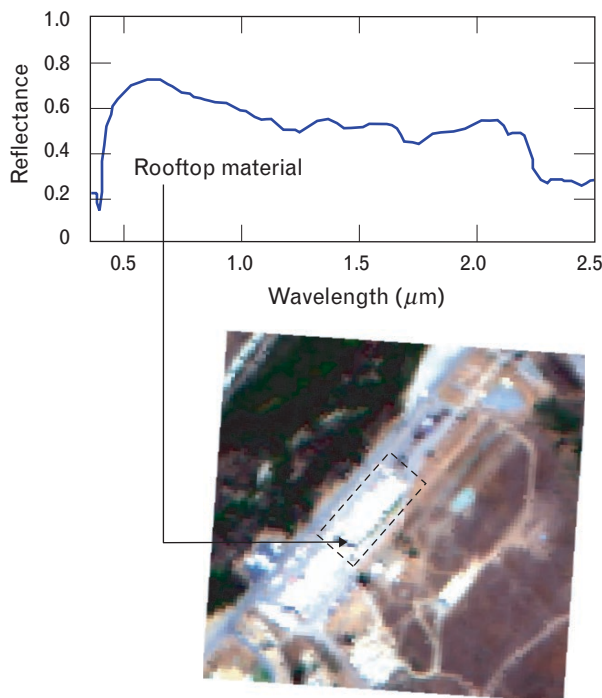

FIGURE 47. Ground truth for the location of the rooftop target shown in Figure 45, obtained by the AVIRIS sensor. Its reference spectral signature, which corresponds to rooftop material, was obtained from a spectral library.

data cube and determine the maximum value within the target buffer. Then we set the threshold to this maximum value and determine the number of false alarms, as shown in Figure 48 for the ACE detector. Figure 49 shows the detection statistics in the target buffer for the ACE, AMF, and GLRT algorithms. All three algorithms can detect the target (circled pixel) with fewer than forty false alarms over the entire image. The ACE algorithm has the smallest number of false alarms, but the numbers for all algorithms change even with a small change of the threshold.

Background Modeling with Elliptically Contoured Distributions

To date, analytic derivation of hyperspectral target detectors has been based on signal models involving multivariate normal distributions. In many practical cases, however, the actual response of a detector to the background pixels differs from the theoretically predicted distribution for Gaussian backgrounds. In fact, the empirical distribution usually has heavier tails compared to the theoretical distribution (see Figure 50), and these tails strongly influence the observed false-alarm rate of the detector. Clearly, the detection problem cannot be addressed in an optimum manner until the departures of the background distribution from normality are understood. The focus of this section is to introduce a non-Gaussian model for the background statistics and discuss the implications for detector design.

A key characteristic of normal random vectors is the elliptical shape of their equal probability contours. It has been shown [30] that multiplying a normal random vector $\mathbf{z} \sim N(\mathbf{0}, \boldsymbol{\Gamma})$ with a non-negative random variable α produces a random vector

$$\mathbf{x} = \alpha \mathbf{z}, \tag{17}$$

whose contours of equal probability still have elliptical shape. The density of the contours is controlled by the probability density of the random variable α. When Equation 17 is applied to hyperspectral data modeling, the normal vector $\mathbf{z}$ can be used to account for the intrinsic spectral variability, whereas the scalar random variable α can be used to model environmental effects associated with the illuminated pixel area. The conditional probability density of $\mathbf{x}$ given α is
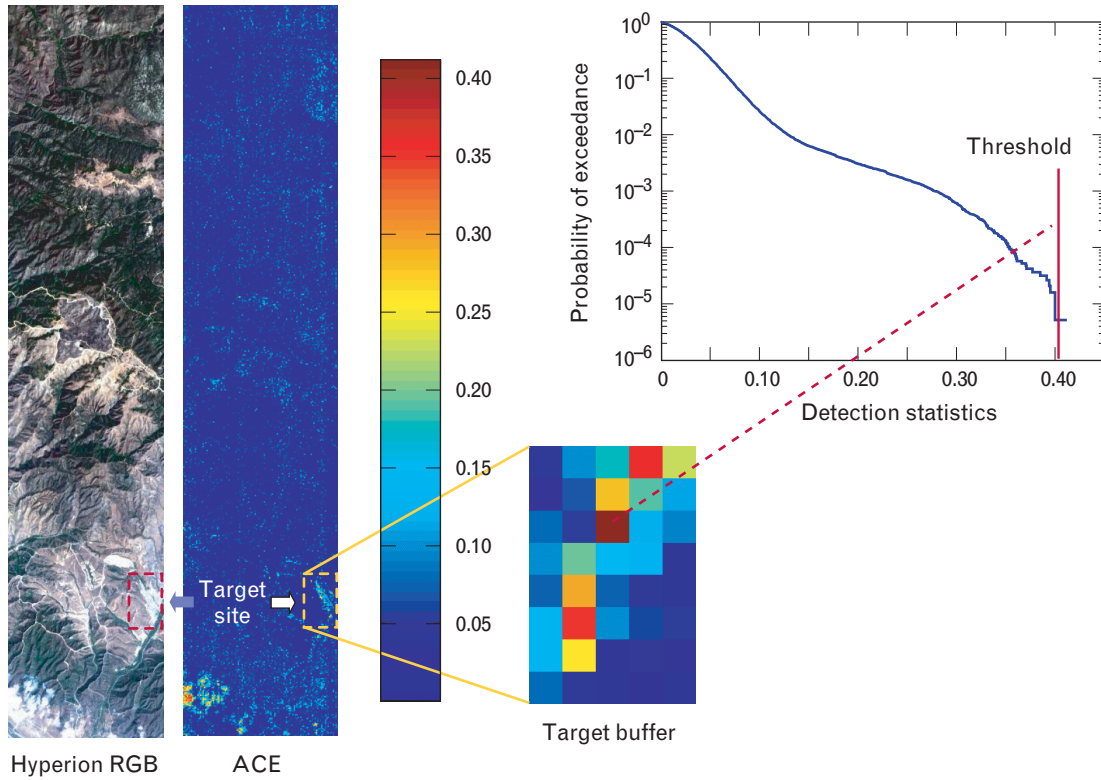
FIGURE 48. Illustration of the threshold-selection process for the ACE algorithm. ACE detection statistics for the Hyperion red-green-blue (RGB) color image are computed, and the maximum value within the target buffer (inside the target site) is used to set the threshold value and to determine the number of false alarms.

$$f_{x|\alpha}(x|\alpha) =$$

$$\frac{1}{(2\pi)^{K/2}\alpha^K|\Gamma|^{1/2}} \exp\left(-\frac{x^T\Gamma^{-1}x}{2\alpha^2}\right). \quad (18)$$

Therefore, the probability density of $\mathbf{x}$ is

$$f(\mathbf{x}) = \int_0^\infty f_{\mathbf{x}|\alpha}(\mathbf{x}|\alpha) f_\alpha(\alpha)\, d\alpha. \quad (19)$$

Using Equation 18 we can express Equation 19 as

$$f(\mathbf{x}) = (2\pi)^{-p/2}|\Gamma|^{-1/2} h_K(d), \quad (20)$$

where $d = \Delta^2$ is the Mahalanobis distance of $\mathbf{x}$ from $\boldsymbol{\mu} = 0$, and

$$h_K(d) = \int_0^\infty \alpha^{-K} \exp\left(-\frac{1}{2}\frac{d}{\alpha^2}\right) f_\alpha(\alpha)\, d\alpha$$

is a positive, monotonically decreasing function for all K. We say that the random vector $\mathbf{x}$ with probability density in Equation 20 has an *elliptically contoured distribution* (ECD) function. This is denoted by using the shorthand notation $EC(\boldsymbol{\mu}, \Gamma, h)$. Note that the normal distribution is a special case of Equation 20 with

$$h_K(d) = \exp\left(-\frac{1}{2}d\right).$$

In addition, the class of ECDs includes the multivariate t, the multivariate Cauchy, and the double exponential distributions. Many of the properties of the multivariate normal distribution extend also to the class of ECDs. Thus all the marginals of an ECD, and the conditional distributions of some variables, given the values of the others, are also ECDs. Given this property, we can use the linear mixing model and elliptically contoured distributed random vectors to create more accurate subpixel spectral models.
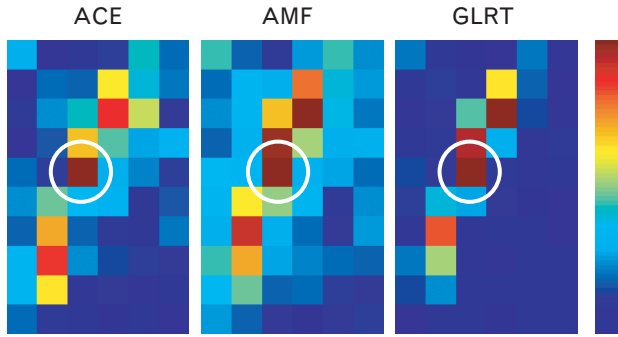
FIGURE 49. Detection statistics in the target buffer area for the ACE, AMF, and GLRT target detection algorithms. All three algorithms can detect the target (circled pixel) with fewer than forty false alarms in the entire cube of about 200,000 pixels.

The random vector $\mathbf{x} \sim EC(\boldsymbol{\mu}, \boldsymbol{\Gamma}, h)$ can be generated by

$$\mathbf{x} = \boldsymbol{\Gamma}^{1/2}(\alpha \mathbf{z}) + \boldsymbol{\mu}, \qquad (21)$$

where $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I})$ and α is a non-negative random variable independent of $\mathbf{z}$. The density of the ECD is uniquely determined by the probability density of α, which is known as the *characteristic* probability density function of $\mathbf{x}$. Note that $f_\alpha(\alpha)$ and $\boldsymbol{\Gamma}$ can be specified independently.

One of the most important properties of random vectors with ECDs is that their joint probability density is uniquely determined by the probability density of the Mahalanobis distance

$$f_d(d) = \frac{1}{2^{K/2} \Gamma(K/2)} d^{(K/2)-1} h_K(d),$$

where $\Gamma(K/2)$ is the Gamma function. As a result, the multivariate probability density identification and estimation problem is reduced to a simpler univariate one. This simplification provides the cornerstone for our investigations in the statistical characterization of hyperspectral background data.

If we know the mean and covariance of a multivariate random sample, we can check for normality by comparing the empirical distribution of the Mahalanobis distance against a chi-squared distribution. However, in practice we have to estimate the mean and covariance from the available data by using

the maximum-likelihood estimates. When the number of pixels N is greater than about $10K$, the estimated Mahalanobis distance, for all practical purposes, follows a chi-squared distribution.

From the multitude of ECDs discussed in statistics literature, the elliptically contoured t-distribution [2] has been shown to provide a good model for many hyperspectral data sets [32]. This distribution is defined by

$$t_K(\mathbf{x}; \boldsymbol{\mu}, \mathbf{C}, \upsilon) = \frac{\Gamma[(K+\upsilon)/2]}{\Gamma(\upsilon/2)(\upsilon\pi)^{K/2}|\mathbf{C}|^{1/2}}$$

$$\cdot \left[1 + \frac{1}{\upsilon}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right]^{-\frac{K+\upsilon}{2}},$$

where $E(\mathbf{x}) = \boldsymbol{\mu}$, $Cov(\mathbf{x}) = \frac{\upsilon}{\upsilon-2}\mathbf{C}$, υ is the number of degrees of freedom, and $\mathbf{C}$ is known as the scale matrix. The Mahalanobis distance is distributed as

$$\frac{1}{K}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \sim F_{K,\upsilon}, \qquad (22)$$

where $F_{K,\upsilon}$ is the F-distribution with K and υ degrees of freedom. The integer υ controls the tails of the distribution: $\upsilon = 1$ leads to the multivariate Cauchy
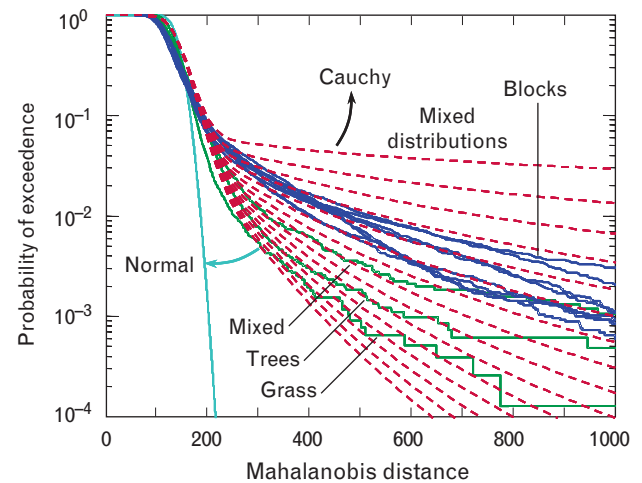

FIGURE 50. Modeling the distribution of the Mahalanobis distance for the HSI data blocks shown in Figure 33. The blue curves correspond to the eight equal-area subimages in Figure 33. The green curves represent the smaller areas in Figure 33 and correspond to trees, grass, and mixed (road and grass) materials. The dotted red curves represent the family of heavy-tailed distributions defined by Equation 22.

distribution (heavier tails), whereas as $v \rightarrow \infty$ the t-distribution approaches the multivariate normal distribution (lighter tails). Random vectors from any EC t distribution can be generated by Equation 21, if the random variable α has a $(\chi_K^2)^{-1/2}$ distribution.

To identify the joint spectral distribution of hyperspectral backgrounds, we compute the probability of exceedance of Mahalanobis distance for various HYDICE data sets. The empirical distributions are compared to the theoretical chi-squared and F-distributions corresponding to the multivariate normal and t-distributions. To reduce the effects of spatial inhomogeneity, we divide the data cube into rectangular blocks, as shown earlier in Figure 33. We also consider three regions, identified by the white boxes in the figure, which define three classes that were selected by their spatial proximity. In the upper right is a *grass* region, in the middle left is a *tree* region, and at the bottom is a *mixed* region of road and grass. These regions define the pixels selected for three of the classes considered. Figure 50 shows the distribution of the Mahalanobis distance for all blocks, plus the three spatially determined classes.

The results in Figure 50, which are representative of several other data sets, indicate that the distribution of HYDICE and AVIRIS data cubes cannot be accurately modeled by the multivariate normal distribution. However, the elliptically contoured multivariate t-distribution provides a promising model. We note that the t-distribution tends to the normal distribution when the number of degrees of freedom increases.

These models can be used to select the threshold for CFAR detectors more accurately, develop detection algorithms that better exploit the statistics of hyperspectral background data, and test the robustness of detectors designed under the normality assumption. Indeed, it has been shown [33] that the structure of the GLRT, AMF, and ACE algorithms remains mathematically unchanged, and they retain the CFAR property for any $h_K(d)$. However, their probability of detection depends on the form of the function $h_K(d)$.

In summary, ECD models are more flexible and provide a better fit to hyperspectral data than the simpler Gaussian distribution. Furthermore, the fact that

the GLRT, AMF, and ACE detectors for ECD and normal distributions have the same structure explains their robustness to deviations from normality. The refinement and application of ECD models in order to enhance performance of hyperspectral CFAR detectors is an important new area of research.

Conclusions

This article presents a tutorial review of the state of the art in target detection algorithms for hyperspectral imaging applications, providing a consistent mathematical notation and framework for algorithms drawn from many disparate sources in the literature. The main obstacles in the development of effective detection algorithms are the inherent variability in target and background spectra. The use of adaptive algorithms deals quite effectively with the problem of unknown backgrounds; the lack of sufficient target data, however, makes the development and estimation of target variability models challenging. Hyperspectral target detection is a multidisciplinary problem that draws results from different scientific and engineering areas. There is a significant signal processing component, however, and expertise from the areas of array processing, radar, and detection theory will provide crucial contributions to future progress.

The empirical performance assessments presented in this article hint at the detection and false-alarm performance achievable with automated matched-filter detection algorithms. However, additional effort in modeling the variability of targets and the non-Gaussian behavior of backgrounds is needed to produce accurate ROC curves for a greater variety of target and background classes.

REFERENCES

1. R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Recognition,* 2nd ed. (Wiley, New York, 2001).

2. R.J. Muirhead, *Aspects of Multivariate Statistical Theory* (Wiley, New York, 1982).

3. G. Healey and D. Slater, "Models and Methods for Automated Material Identification in Hyperspectral Imagery Acquired under Unknown Illumination and Atmospheric Conditions," *IEEE Trans. Geosci. Remote Sens.* **37** (6), 1999, pp. 2706–2717.

4. R.A. Johnson and D.W. Wichern, *Applied Multivariate Statistical Analysis,* 4th ed. (Prentice Hall, Upper Saddle River, N.J., 1998).

5. J.B. Adams, M.O. Smith, and A.R. Gillespie, "Imaging Spectroscopy: Interpretation Based on Spectral Mixture Analysis," chap. 7 in *Remote Geochemical Analysis: Elemental and Mineralogical Composition,* C.M. Pieters and P.A.J. Englert, eds. (Cambridge University Press, Cambridge, U.K.), 1993, pp. 145–166.

6. J.W. Boardman, "Automating Spectral Unmixing of AVIRIS Data Using Convex Geometry Concepts," *4th Ann. JPL Airborne Geoscience Workshop* **1,** *Pasadena, Calif., 25 Oct. 1993,* pp. 11–14.

7. D.W.J. Stein, S.G. Beaven, L.E. Hoff, E.M. Winter, A.P. Schaum, and A.D. Stocker, "Anomaly Detection from Hyperspectral Imagery," *IEEE Signal Process. Mag.* **19** (1), 2002, pp. 58–69.

8. S.M. Kay, *Fundamentals of Statistical Signal Processing* (Prentice Hall, Englewood Cliffs, N.J., 1998).

9. R.A. Fisher, "Multiple Measures in Taxonomic Problems," *Ann. Eugenics* **7,** 1936, pp. 179–188.

10. 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, Boston, 2000).

11. H.L. Van Trees, *Detection, Estimation, and Modulation Theory,* pt. 4: *Optimum Array Processing* (Wiley, New York, 2002).

12. W.H. Farrand and J.C. Harsanyi, "Mapping the Distribution of Mine Tailings in the Coeur d'Alene River Valley, Idaho, through the Use of a Constrained Energy Minimization Technique," *Rem. Sens. Environ.* **59** (1), 1997, pp. 64–76.

13. I.S. Reed, J.D. Mallett, and L.E. Brennan, "Rapid Convergence Rate in Adaptive Arrays," *IEEE Trans. Aerosp. Electron. Syst.* **10** (6), 1974, pp. 853–863.

14. E.J. Kelly, "An Adaptive Detection Algorithm," *IEEE Trans. Aerosp. Electron. Syst.* **22** (1), 1986, pp. 115–127.

15. E.J. Kelly, "Adaptive Detection in Non-Stationary Interference, Part III," Technical Report 761, Lincoln Laboratory (10 June 1986), DTIC# ADA-185622.

16. E.J. Kelly and K.W. Forsythe, "Adaptive Detection and Parameter Estimation for Multidimensional Signal Models," Technical Report 848, Lincoln Laboratory (19 April 1989), DTIC# ADA-208971.

17. I.S. Reed and X. Yu, "Adaptive Multiple-Band CFAR Detection of an Optical Pattern with Unknown Spectral Distribution," *IEEE Trans. Acoust. Speech Signal Process.* **38** (10), 1990, pp. 1760–1770.

18. S. Kraut and L.L. Scharf, "The CFAR Adaptive Subspace Detector Is a Scale-Invariant GLRT," *IEEE Trans. Signal Process.* **47** (9), 1999, pp. 2538–2541.

19. S. Kraut, L.L. Scharf, and L.T. McWhorter, "Adaptive Subspace Detectors," *IEEE Trans. Signal Process.* **49** (1), 2001, pp. 1–16.

20. X. Yu, I.S. Reed, and A.D. Stocker, "Comparative Performance Analysis of Adaptive Multiband Detectors," *IEEE Trans. Signal Process.* **41** (8), 1993, pp. 2639–2656.

21. E. Conte, M. Lops, and G. Ricci, "Asymptotically Optimum Radar Detection in Compound-Gaussian Clutter," *IEEE Trans. Aerosp. Electron. Syst.* **31** (2), 1995, pp. 617–625.

22. L.L. Scharf and L.T. McWhorter, "Adaptive Matched Subspace Detectors and Adaptive Coherence," *Proc. 30th Asilomar Conf. on Signals and Systems,* Pacific Grove, Calif., 3–6 Nov. 1996, pp. 1114–117.

23. F.C. Robey, D.R. Fuhrmann, E.J. Kelly, and R. Nitzberg, "A CFAR Adaptive Matched Filter Detector," *IEEE Trans. Aerosp. Electron. Syst.* **28** (1), 1992, pp. 208–216.

24. W.-S. Chen and I.S. Reed, "A New CFAR Detection Test for Radar," *Dig. Signal Process.* **1** (4), 1991, pp. 198–214.

25. C.D. Richmond, "Performance of the Adaptive Sidelobe Blanker Detection Algorithm in Homogeneous Environments," *IEEE Trans. Signal Process.* **48** (5), 2000, pp. 1235–1247.

26. L.L. Scharf and B. Friedlander, "Matched Subspace Detectors," *IEEE Trans. Signal Process.* **42** (8), 1994, pp. 2146–2157.

27. D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral Subpixel Target Detection Using the Linear Mixing Model," *IEEE Trans. Geosci. Remote Sens.* **39** (7), 2001, pp. 1392–1409.

28. J.C. Harsanyi and C.-I. Chang, "Hyperspectral Image Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Approach," *IEEE Trans. Geosci. Remote Sens.* **32** (4), 1994, pp. 779–785.

29. L.J. Rickard, R. Basedow, E. Zalewski P. Silvergate, and M. Landers, "HYDICE: An Airborne System for Hyperspectral Imaging," *SPIE* **1937,** 1993, pp. 173–179.

30. K. Yao, "A Representation Theorem and Its Application to Spherically-Invariant Random Processes," *IEEE Trans. Inform. Theory* **19** (5), 1973, pp. 600–608.

31. E. Conte and M. Longo, "Characterization of Radar Clutter as a Spherically Invariant Random Process," *IEE Proc. F* **134** (2), 1987, pp. 191–197.

32. D. Manolakis and D. Marden, "Non Gaussian Models for Hyperspectral Algorithm Design and Assessment," *Proc. IEEE IGARSS* **3,** Toronto, 24–28 June 2002, pp. 1664–1666.

33. C.D. Richmond, "Adaptive Array Processing in Non-Gaussian Environments," *Proc. 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing, 24–26 June 1996, Corfu, Greece,* pp. 562–565.

DIMITRIS MANOLAKIS
is a staff member in the Sensor Technology and System Applications group. His research experience and interests include the areas of digital signal processing, adaptive filtering, array processing, pattern recognition, remote sensing, and radar systems. He received a B.S. degree in physics and a Ph.D. in electrical engineering from the University of Athens, Greece. 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, and he is a coauthor of the textbooks *Digital Signal Processing: Principles, Algorithms, and Applications* (Prentice-Hall, 1996, 3rd ed.) and *Statistical and Adaptive Signal Processing* (McGraw-Hill, 2000).

DAVID MARDEN
is a research student in the Sensor Technology and System Applications group. His research interest includes image and signal processing, statistical models, and data compression. He received a B.S. degree in electrical engineering from Boston University, and is in the process of completing a Ph.D. degree in electrical engineering from Northeastern University. He has spent the last few years at Lincoln Laboratory doing graduate research, focusing on statistical models of hyperspectral image data.

GARY A. SHAW
is a senior staff member in the Advanced Space Systems and Concepts group. His current research focus is on collaborative sensing concepts, including energy-efficient techniques for both line-of-sight and non-line-of-sight electro-optical communication and networking. He joined Lincoln Laboratory in 1980 to work on adaptive processing for radar systems, and was a member of the study team that launched a program in the early 1980s to develop concepts for affordable space radar. From 1984 to 1987 he served as test director and assistant leader of the ALCOR and MMW radars at the Kwajalein Reentry Measurements Site. After his Kwajalein tour he served as assistant group leader and then associate group leader in what is now the Sensor Technology and System Applications Group, where he applied his signal processing expertise to radar and passive sensor algorithm development. He received B.S. and M.S. degrees in electrical engineering from the University of South Florida, and a Ph.D. degree in electrical engineering from the Georgia Institute of Technology, where he was both a President's Fellow and a Schlumberger Fellow.