6-2007

Hyperspectral Imagery Target Detection Using Improved Anomaly Detection and Signature Matching Methods

Timothy E. Smetek

HYPERSPECTRAL IMAGERY TARGET DETECTION USING IMPROVED ANOMALY DETECTION AND SIGNATURE MATCHING METHODS

DISSERTATION

Timothy E. Smetek, Major, USAF

AFIT/DS/ENS/07-07

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/DS/ENS/07-07

HYPERSPECTRAL IMAGERY TARGET DETECTION USING IMPROVED ANOMALY DETECTION AND SIGNATURE MATCHING METHODS

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Timothy E. Smetek, BS, MS

Major, USAF

June 2007

AFIT/DS/ENS/07-07

HYPERSPECTRAL IMAGERY TARGET DETECTION

USING IMPROVED ANOMALY DETECTION AND

SIGNATURE MATCHING METHODS

Timothy E. Smetek, BS, MS
Major, USAF

Approved:

	Date
/signed/	
Kenneth W. Bauer Jr. (Chairman)	
/signed/	
David R. Jacques (Dean's Representative)	
/signed/	
Robert T. Brigantic (Member)	
/signed/	
John O. Miller (Member)	
/signed/	
Mark E. Oxley (Member)	

Accepted:

/signed/

Marlin U. Thomas — Date
Dean, Graduate School of Engineering and Management

AFIT/DS/ENS/07-07

Abstract

This research extends the field of hyperspectral target detection by developing autonomous anomaly detection and signature matching methodologies that reduce false alarms relative to existing benchmark detectors, and are practical for use in an operational environment. The proposed anomaly detection methodology adapts multivariate outlier detection algorithms for use with hyperspectral datasets containing tens of thousands of non-homogeneous, high-dimensional spectral signatures. In so doing, the limitations of existing, non-robust, anomaly detectors are identified, an autonomous clustering methodology is developed to divide an image into homogeneous background materials, and competing multivariate outlier detection methods are evaluated for their ability to uncover hyperspectral anomalies. To arrive at a final detection algorithm, robust parameter design methods are employed to determine parameter settings that achieve good detection performance over a range of hyperspectral images and targets, thereby removing the burden of these decisions from the user. The final anomaly detection algorithm is tested against existing local and global anomaly detectors, and is shown to achieve superior detection accuracy when applied to a diverse set of hyperspectral images.

The proposed signature matching methodology employs image-based atmospheric correction techniques in an automated process to transform a target reflectance signature library into a set of image signatures. This set of signatures is combined with an existing linear filter to form a target detector that is shown to perform as well or better relative to detectors that rely on complicated, information-intensive, atmospheric correction schemes. The performance of the proposed methodology is assessed using a range of target materials in both woodland and desert hyperspectral scenes.

AFIT/DS/ENS/07-07

Acknowledgments

First and foremost, I thank my Heavenly Father for blessing me with the strength and energy to complete this research. Without His divine grace, I am quite certain I would not have made it this far. Second, I thank my wife and children for making just as many sacrifices as I did, if not more, in achieving this goal. Pursuing a doctorate degree is no trivial matter in and of itself. Doing so with a family of seven children makes for a tremendously difficult challenge that can only be overcome through teamwork, understanding, and patience—thank you for all of your support.

Next, I would like to express my gratitude to my advisor, Ken Bauer, who gave me the freedom to pursue my research interests as I saw fit—research is much more enjoyable when you don't feel pressured to adhere to someone else's agenda or personal interests. Similarly, I thank my research committee, J.O. Miller, Mark Oxley, Robert Brigantic, and David Jacques, for their time and efforts in helping this dissertation come to fruition. I am also indebted to Mike Eismann of the Air Force Research Laboratory for his insights and ideas that were invaluable in completing the signature matching portion of this research—our conversations were few, but very worthwhile.

Finally, my thanks goes out to my former supervisors, colleagues, and friends that encouraged me to pursue a doctorate degree. Without their confidence in my abilities, I probably would not have embarked on this journey.

Table of Contents

List of Figures

List of Tables

Hyperspectral Imagery Target Detection

Using Improved Anomaly Detection And Signature Matching Methods

I. Introduction

Locating a small number of unique objects scattered across a relatively large geographic area is a common problem faced by many different professions. From an environmental perspective, these objects may be harmful vegetation species that need to be eradicated from croplands or unique minerals that indicate favorable mining locations. For those entrusted with search-and-rescue missions, the object of interest may be a downed-aircraft or adventurers lost in the wilderness. In the field of law-enforcement, it may be necessary to detect illegal border-crossings or the cultivation of illegal crops associated with the drug trade. Finally, the problem faced by the Department of Defense may be detection of tanks, aircraft, surface-to-air missile launchers, command bunkers, and other objects of military significance scattered across a battlefield.

For all of these target detection endeavors, hyperspectral imagery collected from remote sensing platforms provides a powerful means for detecting the targets of interest. Specifically, the unique spectral signatures of objects recorded in hyperspectral imagery can be used to discriminate the target from background materials. Many target detection algorithms have been proposed in the literature that use this spectral information in one of two ways: either the signatures are screened for anomalous spectra that may indicate the presence of the desired target; or, the signatures from the image are compared to known target signatures to determine if there is a match. It is the purpose of this research to

expand upon both of these target detection strategies with the goal of increasing detection accuracy, and thereby increasing the practicality of hyperspectral sensors.

Problem Definition

A common element of current hyperspectral anomaly detection methods is the use of classical mean vector and covariance matrix estimates for groups of signatures in the hyperspectral image. Often, these estimates are used to compute the Mahalanobis Squared Distance (MSD) for a signature, **x**, given as

$$MSD(\mathbf{x}) = (\mathbf{x} - \hat{\boldsymbol{\mu}})^T \mathbf{S}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}})^T \tag{1.1}$$

where

$\hat{\boldsymbol{\mu}}$ = the mean vector estimate, and
$\mathbf{S}$ = the covariance matrix estimate.

A threshold is then applied to the MSDs of all the image signatures, with those lying above the threshold then being associated with anomalous signatures. A serious problem with this approach to anomaly detection is that the inclusion of anomalous signatures in the mean vector and covariance estimates can significantly distort the estimates. These distortions can then lead to inaccurate MSD calculations and the potential for missed targets and increased false alarms. Though this problem is real, it is often ignored in practice.

In regards to signature matching methods, the primary obstacle to successful target detection is the requirement to transform the signatures measured by the sensor into the same type of signatures contained in reference target libraries. Specifically, an object's signature collected by the hyperspectral sensor typically measures the energy radiance—measured in units of watts/meter^2/steradian—originating from the direction of

the object and reaching the sensor. The reference target signature, on the other hand, typically reports the percentage of incident energy that the target material reflects back into the atmosphere. These signatures are referred to as reflectance signatures and are unitless. Based on these differences, it is evident that we cannot directly compare signatures from the hyperspectral sensor to the reflectance signatures for the target. In some cases, the sensor may not be calibrated to measure radiance, but rather records measurements that are only proportional to the actual radiance, further complicating the problem.

The standard procedure for reconciling sensor signatures with reflectance signatures is referred to as atmospheric calibration. In this process, knowledge of sensor location, object location, Sun location, airborne particle concentrations, water vapor concentrations, and atmospheric temperature and density profiles is used to determine the reflectance signature of the object in view of the sensor based on the amount of energy that reached the sensor. This calibration process is by no means precise, and is further complicated by the fact that only a portion, if any, of the aforementioned information may be available for the image being analyzed. In practice, this conversion between radiance and reflectance signatures is a formidable barrier to widespread use of hyperspectral data and remains an open area of research.

Looking beyond inaccurate statistical estimates and atmospheric calibration issues, an additional limitation that is common to both anomaly detection and signature matching is the high level of technical expertise required to employ these types of algorithms. In general, these methods employ sophisticated statistical or mathematical techniques that are well-beyond the education level of the users that can most benefit

from their use. As a result, hyperspectral sensors remain an exotic technology that only benefits those with the educational background to understand the intricacies of how the sensor and its associated algorithms work. This problem is summarized best by Dr. David Landgrebe, one of the pioneers of remote sensing, in his view of the future of hyperspectral algorithms:

> …what is needed is an analysis process that is robust in the sense that it would work effectively for data of a wide variety of scenes and conditions, and can be used effectively by users rather than only by producers of the technology. The algorithms do not need to be simple, but they must be simple to apply and robust against the variety of user problems. (Landgrebe, 2005)

In other words, progress needs to be made in the development of algorithms that are not only accurate, but also accessible to a range of operational users that may wish to employ them.

Research Objectives

In light of the problems just defined, the focus of this research is to achieve the following objectives:

1) *Adapt multivariate outlier detection methods for use as hyperspectral anomaly detectors*. Multivariate outlier detection has been an active area of research for over three decades, yet the algorithms developed in this field have not been used for hyperspectral anomaly detection, a clear exercise in finding outliers in multivariate data. By pursuing this objective, we show that using multivariate outlier detection to find anomalies results in more accurate mean vector and covariance estimates for use in computing MSDs, and thereby

improves detection performance relative to existing benchmark anomaly detectors.

2) *Automate the anomaly detectors developed through Objective 1 so that they can be applied to a range of images with minimal user input or intervention*. To achieve this objective we identify methods for automatically clustering hyperspectral datasets as a preprocessing step to outlier detection, and use Taguchi robust parameter design methods to configure our anomaly detection methods to achieve consistent performance for a variety of detection scenarios.

3) *Develop a signature matching methodology that can effectively detect a range of target materials when little or no information is available to perform an atmospheric calibration on the image*. To satisfy this objective, we build upon the invariant signature matching method originally proposed by Healey and Slater (1999) that eliminates the need for traditional atmospheric calibration. Rather than use the MODTRAN4 atmospheric model to generate target signature subspaces, however, we develop a method that uses only in-scene information to estimate possible target image signatures based on the target's reflectance signature. We then use these estimated target signatures in the Target-Constrained Interference-Minimized Filter (TCIMF) of Ren and Chang (2000) to detect targets of interest.

4) *Automate the signature matching method developed through Objective 3 to minimize required input parameters and user intervention*. To meet this research goal, we designed the proposed signature matching algorithm so the

user only needs to specify the target reflectance signatures as well as generic background reflectance signatures that are likely to exist in the image scene. We show through experimental tests that these inputs are sufficient for a range of targets, but that detection accuracy can be improved if the user is able to obtain more accurate knowledge of the materials in the scene.

By fulfilling these research objectives, we make strides towards a larger objective illustrated by the target detection framework shown in Figure 1. In this framework, we ultimately hope to fuse the output from our proposed anomaly detection and signature matching algorithms to achieve a better detection accuracy than either of the two detectors used individually. Moreover, this framework is intended to operate on an arbitrary hyperspectral image regardless of the measurement units of the sensor, and to minimize the technical expertise and intervention necessary to employ the framework. Future research efforts beyond those conducted in this dissertation are required to complete the fusion component of the detection framework and achieve the final end-state.

Dissertation Outline

In Chapter 2 of this dissertation, we provide a more detailed overview of hyperspectral concepts and attempt to define the basic terminology that pertains to this research. Though we strive to address the key ideas behind hyperspectral image analysis, we omit many details for the sake of brevity. For a more comprehensive explanation of hyperspectral sensors and analytic methods, we suggest texts by Schott (1997), Landgrebe (2003), Richards and Jia (1999), and Chang (2003), as well as the overview given in Landgrebe (2002).

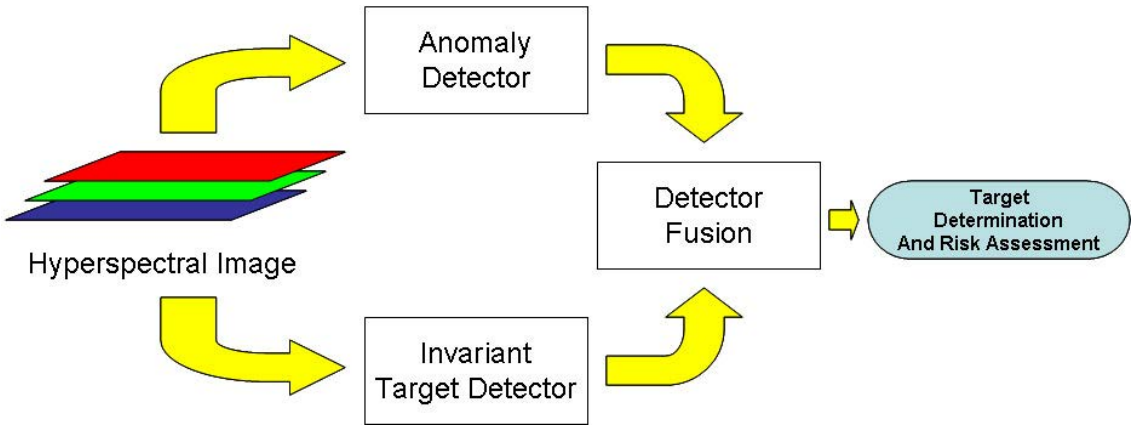


Figure 1. Proposed Target Detection Framework

In Chapter 3, we discuss the hyperspectral anomaly detection problem in more detail and provide a comprehensive literature review of the anomaly detection methods that have been proposed over the last two decades. Based on this review, it becomes evident that the field of anomaly detection has largely ignored multivariate outlier detection concepts, thus providing motivation for our line of research.

Following the anomaly detection review, Chapter 4 summarizes the body of research concerned with invariant target-subspace detection. The review begins with an overview of Healey and Slater's (1999) seminal paper in this area, and proceeds with a description of the various methods that stem from this original work. This review provides the necessary background to understand the context and objectives of our proposed signature matching algorithm presented in Chapter 6.

The proposed anomaly detection method, which we refer to as the AutoDet methodology, is developed in Chapter 5. This chapter begins with a detailed literature review of existing multivariate outlier detection methods and some of the key concepts fundamental to this research area. We then demonstrate the potential problems presented

by outliers in hyperspectral datasets, followed by an experimental evaluation of candidate multivariate outlier methods that are capable of handling large, high-dimensional datasets. The chapter moves on to address the suitability of the *k*-means clustering algorithm as a preprocessor for multivariate outlier detectors, a critical component in the adaptation of outlier detectors to hyperspectral anomaly detection. This discussion is followed by experimental testing of the methods to automatically select the value of *k* for use in the *k*-means algorithm, thereby removing this burden from the user. Finally, the chapter concludes with a Taguchi robust parameter design experiment to optimize the settings of the AutoDet detector, followed by a comparison test demonstrating the superiority of AutoDet to benchmark anomaly detection methods.

Chapter 6 follows the development of AutoDet in Chapter 5 with the construction and evaluation of our proposed signature matching algorithm, which we name AutoMatch. In this chapter we develop the AutoMatch detection methodology through an actual target detection example, and then demonstrate through experimental tests that the method performs as well or better than a benchmark method using more sophisticated atmospheric calibration methods. We conclude Chapter 6 with a discussion of the limitations of AutoMatch and suggestions for further research.

In the final chapter of this dissertation, we summarize the major contributions of our research effort as well as the areas for future research discussed in Chapters 5 and 6. For those readers not interested in reading this dissertation in its entirety, we offer several suggestions. For those knowledgeable of hyperspectral concepts, Chapter 2 can be omitted. For those only interested in the AutoDet anomaly detection methodology,

Chapters 4 and 6 are not necessary. Likewise, those seeking information on AutoMatch may bypass Chapters 3 and 5 with minimal impact.

II. Hyperspectral Data Concepts

In simplest of terms, electromagnetic energy, consisting of Gamma-rays, X-rays, visible light, microwaves, radio waves, etc., can be viewed as energy waves passing through space. These energy waves can be characterized by either their wavelength or their frequency, where the wavelength is the distance between wave peaks, and frequency is the number of waves that pass a fixed point in space per unit time. In remote sensing, wavelength, usually measured in micrometers (μm), is the most common method for characterizing electromagnetic (EM) energy. Figure 2 shows the types of EM energy as they are oriented across the EM spectrum. Notice that the region of the EM spectrum associated with visible light—the energy we can detect with our eyes—comprises a relatively small portion of the overall spectrum. This fact is the foundation upon which hyperspectral imagery is built. We will return to this concept momentarily.

If a sensor capable of detecting EM energy is pointed at an object, the energy detected by the sensor results from two primary mechanisms: either the energy is emitted from the object itself, or it is energy emitted from some other object and reflected by the object at which the sensor is pointed. If an airborne or space HSI sensor is pointed at the Earth's surface, the energy it detects is primarily due to reflected solar energy with a minor contribution from energy emitted from the Earth itself. Figure 3 depicts the relative intensity of these energy sources as a function of wavelength. Again, notice that the region of visible light, 0.4 to 0.7 μm, is only a small portion of the spectrum of energy emitted from the Sun and Earth. Thus, returning to the basic premise of hyperspectral imagery discussed earlier, there is more information about the energy reflected or emitted from an object than can be gained simply from studying visible light.

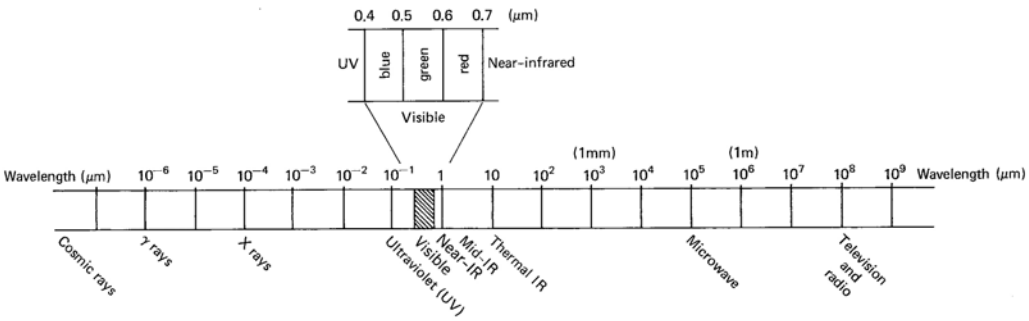


Figure 2. The Electromagnetic Spectrum (Lillesand and Kiefer, 2000)

To further motivate the study of non-visible light reflected and emitted from an object, different materials reflect and emit non-visible EM energy in unique ways, just as visible light is reflected to produce an object's color. In other words, if we measured the EM energy from an object across different wavelengths, the resulting signature will be unique to the object's composition. Thus, rocks will have different signatures than trees, trees will have different signatures than grass, grass will have different signatures from metal surfaces, and so on. The purpose, then, of a hyperspectral sensor is to collect the EM signatures from the Earth's surface so that they can be used to identify materials of interest, whether they be tanks, downed aircraft, land-cover types, or minerals and ores.

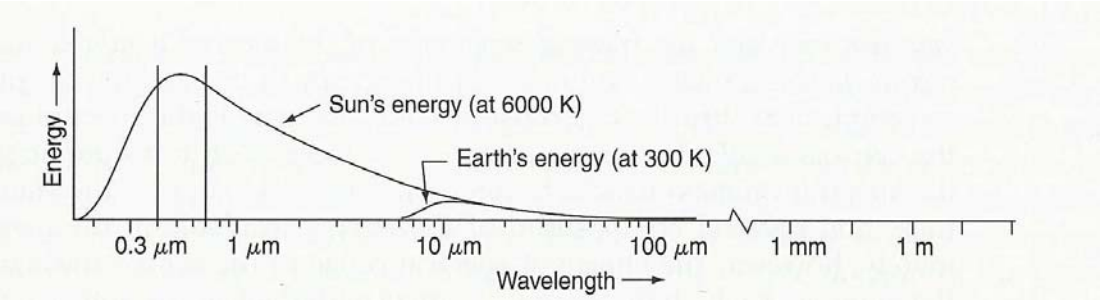


Figure 3. Energy Sources (Lillesand and Kiefer,2000)

Before moving-on to discuss the representation of hyperspectral data, it is important to understand the nature of the energy an airborne or satellite hyperspectral sensor actually receives. As shown in Figure 4, solar energy collected by the sensor arrives via three primary paths. The first path is traveled by energy actually reflected from the Earth's surface. This reflected energy comprises the spectral signature of the material that reflected it, and is the energy the sensor is designed to collect. The second path is taken by energy referred to as path radiance, which is actually solar energy that is reflected by the atmosphere and hence possesses its own unique signature that is characteristic of prevailing atmospheric conditions. This second source of energy is undesirable and acts to distort the energy signatures from the Earth's surface. The third energy path is taken by energy known as skylight. This energy bounces off atmospheric molecules before being reflected by the Earth's surface. The primary impact of skylight is to increase the total illumination of a surface material relative to illumination from solar energy alone. A complicating factor with skylight is that it is typically not uniform throughout an image scene. The total combined energy received by the sensor is called radiance energy, and the intensity of this radiance is what the sensor is measuring. If atmospheric correction is performed on the radiance data, the effects of path radiance,

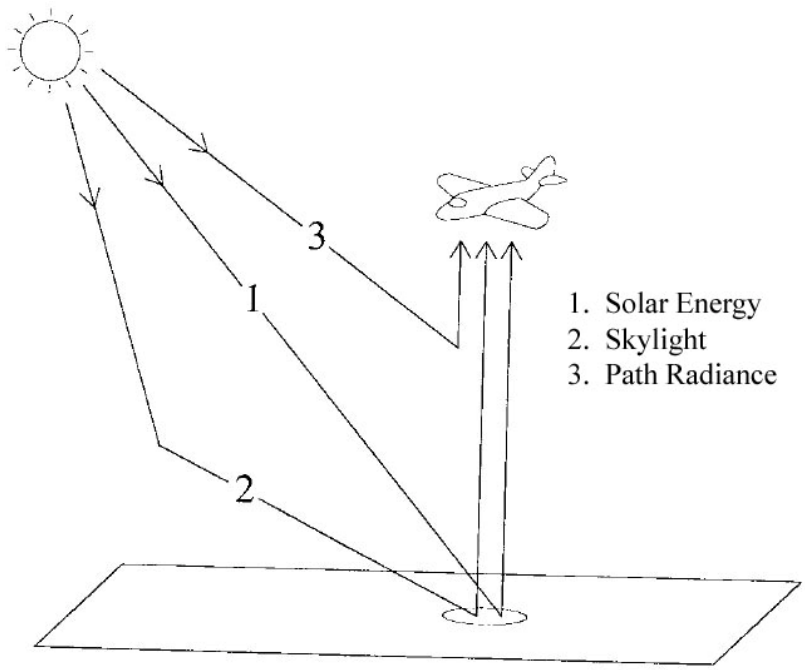


Figure 4. Sources of Sensor Detected Energy (Healey and Slater, 1999)

skylight and other distorting factors can be removed. The signatures resulting from atmospheric correction are referred to as reflectance signatures, and represent the relative amount of energy hitting the Earth's surface that is actually reflected. In other words, reflectance measurements relate the relative fraction of incident energy reflected by a material in the different wavelength bands recorded by the sensor.

In most instances, signature libraries provide material signatures in terms of reflectance. Since an actual sensor measures radiance, a direct comparison cannot be made between library signatures and raw HSI radiance data. As mentioned previously, either radiance must be converted to reflectance, or vice versa. To convert a laboratory reflectance signature into a radiance signature, an atmospheric model must be assumed. A basic model, as discussed by Healey and Slater (1999), is as follows. First, assume the viewing geometry of a hyperspectral scene as shown in Figure 5. The normal vector to a

pixel located at coordinate (x, y) on the Earth's surface is given by **n** and the elevation of the pixel is z_g. Using a polar coordinate system with polar angle θ and azimuthal angle φ, the airborne sensor is located at elevation z_v and direction (θ_v, φ_v). The sun is located at direction (θ_0, φ_0). With these variables defined, the spectral radiance at wavelength λ collected by the sensor pointed at a pixel at (x, y) is given by:

$$\begin{aligned} L(x,y,\lambda) &= T_u\left(z_g, z_v, \theta_v, \phi_v\right) R\left(x,y,\lambda\right) \\ &\cdot\left[KT_d\left(z_g,\theta_0,\phi_0,\lambda\right)E_0\left(\lambda\right)\cos\theta_0 + \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi/2} E_s\left(\theta,\phi,\lambda\right)\cos\theta\sin\theta d\theta d\phi\right] \\ &+P\left(z_g, z_v, \theta_v, \phi_v, \lambda\right), \end{aligned} \tag{2.1}$$

where

$T_u\left(z_g, z_v, \theta_v, \phi_v\right)$ = the upward atmospheric transmittance,
$R\left(x,y,\lambda\right)$ = the spectral reflectance of the material at wavelength λ,
K = a binary constant used to account for occlusion of the pixel,
$T_d\left(z_g,\theta_0,\phi_0,\lambda\right)$ = the downward atmospheric transmittance,
$E_0\left(\lambda\right)$ = the extraterrestrial solar radiance,
$E_s\left(\theta,\phi,\lambda\right)$ = the scattered sky radiance due to skylight, and
$P\left(z_g, z_v, \theta_v, \phi_v, \lambda\right)$ = the path-scattered radiance.

From (2.1) it is evident that if the reflectance signature for a material is known, as well as the atmospheric and viewing geometry parameters, then the radiance signature for the material can be determined. Typically, atmospheric models such as MODTRAN 4 are used to estimate the values of T_u, T_d, E_0, E_s and P for the conditions present at the time an image was obtained. As discussed by Richards and Jia (1999), less rigorous correction methods can also be used such as signature normalization and dark subtraction which are only dependent on the statistics of the image. For these methods to produce useful results, however, atmospheric effects should be uniform over the entire image, an assumption that is image and sensor dependent.

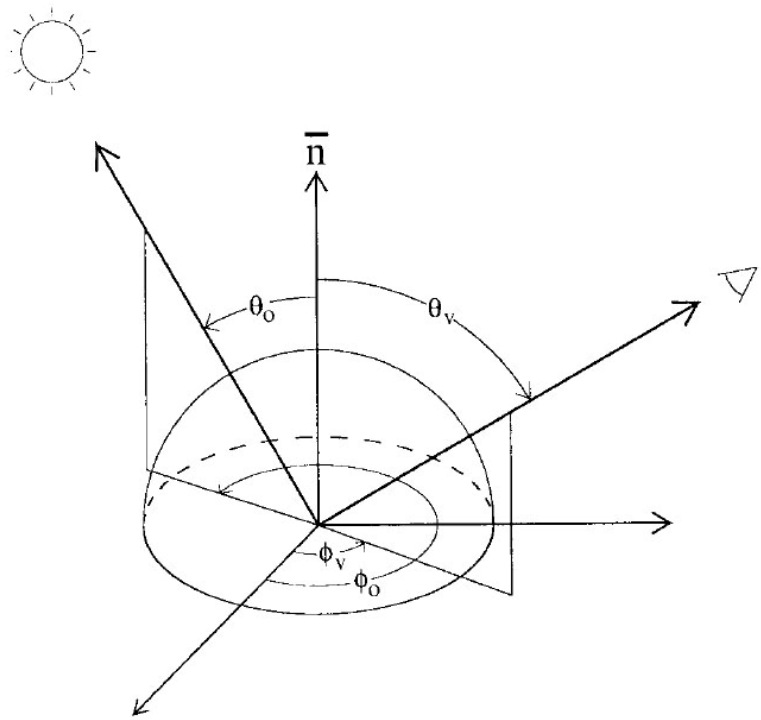


Figure 5. Geometry of a Hyperspectral Image

Because atmospheric conditions are either not known at all or with limited accuracy, and because atmospheric modeling itself is only an approximation of true atmospheric effects, conversion of reflectance to radiance will result in approximate radiance signatures. The resulting approximation error is the primary motivation for developing anomaly detection algorithms that do not require atmospheric correction. This error also motivates target detection methods, such as Healey and Slater's invariant subspace detector, that are less dependent on explicit atmospheric correction for which the precise atmospheric conditions and viewing geometry of an image must be known.

The actual physics and hardware used to collect hyperspectral imagery are beyond the scope of this dissertation; however, the hyperspectral image itself is relatively straight forward to conceptualize. To begin with, the geographic area being imaged is divided

into a raster grid, with each grid cell, or pixel, corresponding to a small, rectangular sub-region of the main image. The dimension of a pixel's edge specifies the spatial resolution of the image, which, for HSI sensors, ranges from fractions to tens of meters. For each pixel, the sensor collects the amount of energy radiated from the pixel's geographic location. As mentioned previously, this energy consists not only of reflected and emitted energy from the pixel region itself, but also from energy scattered by the atmosphere and other nearby regions.

To form the pixel's energy signature, the portion of the EM spectrum for which the sensor is designed is divided into contiguous bands of wavelengths. For hyperspectral sensors, the number of bands ranges from tens to hundreds of bands, as compared to only three to approximately 20 bands for multispectral sensors. There is no generally accepted number of bands that distinguishes multispectral from hyperspectral images. The number of bands and the wavelength interval they encompass define the sensor's spectral resolution.

For each pixel and for each band, the sensor records the aggregate amount of energy received across all wavelengths in the band. Conceptually, the energy information is stored in a three-dimensional array with the first two dimensions identifying a pixel's row and column location in the raster grid, and the third dimension specifying the spectral band. Thus, element (i, j, n) of the data array—also referred to as the data cube—contains the amount of energy detected in wavelength band n for the pixel located in row i, column j of the raster grid. If the image contains N bands, the pixel signature is simply the Nx1 column vector with the nth element corresponding to the energy detected in band n.

To demonstrate the preceding hyperspectral terminology, Figure 6 shows a hyperspectral image of a region of Fort A.P. Hill taken by the COMPASS sensor on 21 July, 2004, at an altitude of approximately 6700 ft. The image contains 255 bands ranging from 0.416 μm to 2.402 μm, which encompasses the visible to mid-IR portion of the EM spectrum. The range of wavelengths for each band is approximately 0.008 μm. The image shown in Figure 6 is a true color image formed by assigning bands 30, 18, and 6 to the red, green, and blue display color guns, respectively. These bands lie within the portion of the EM spectrum associated with red, green, and blue light.

Figure 6. True-Color Image of Fort A.P. Hill Region

As mentioned previously, the purpose of hyperspectral imagery is to collect information on how materials reflect energy, not only in the visible spectrum, but in other

portions of the EM spectrum, as well. To illustrate this point, Figure 7 shows grayscale images of bands 11, 76, and 204, which correspond to visible, near IR, and mid-IR portions of the spectrum, respectively. In these images, the brighter a pixel the higher the intensity of energy collected over the band's corresponding range of wavelengths. By comparing the intensity of the road pixels, for example, in the three images, it is clear that the road surface reflects energy differently as a function of the energy's wavelength. Notice that other materials in the image—trees, grass, the man-made objects in the image—also exhibit different intensities across the three image bands.

Figure 7. Bands 11, 76, and 204 of A.P. Hill Image

To better capture the variation of a pixel's intensity across the hyperspectral bands, the pixel vector is formed, as described earlier. Plotting the elements of the pixel vector as a function of image band gives a visual depiction of the pixel signature. Figure 8 plots the signatures of four pixels containing four different materials—trees, grass, dirt, and road. It is evident from these plots that considerable differences exist in the signatures of different materials. It is these differences that make hyperspectral imagery

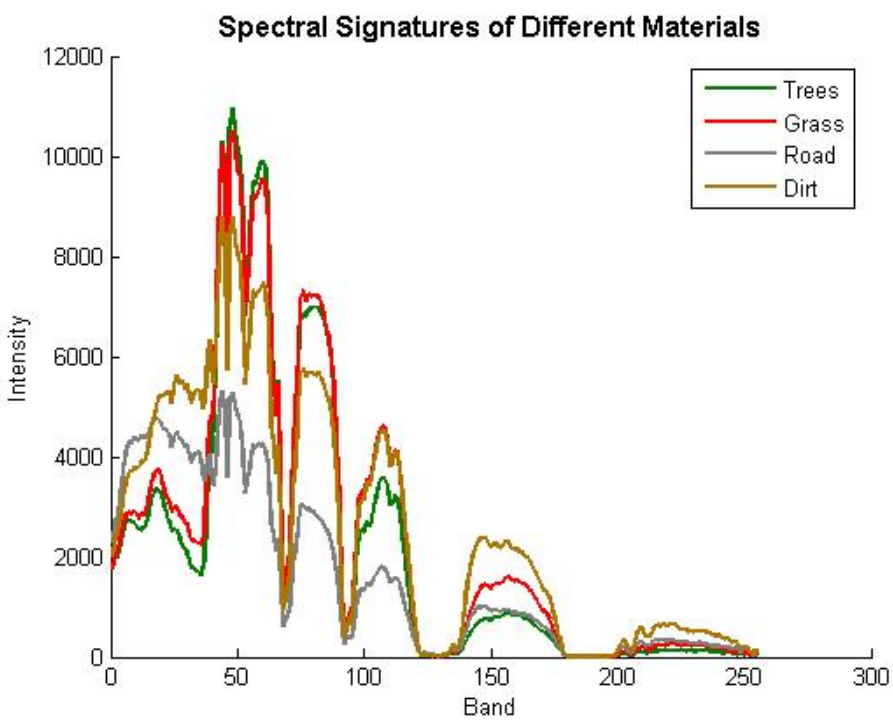


Figure 8. Example of Different Material Spectra

useful in identifying and differentiating materials from one another. Notice, however, that though each material signature is unique somewhere along the spectrum, there are some portions of the spectrum in which the signatures of different materials are quite similar. For example, notice that the signature of the tree pixel coincides closely with the

grass signature in several portions of the spectrum. This condition demonstrates the value in obtaining a wide range of spectral bands in order to adequately separate material classes.

The benefit of obtaining a large number of bands is further illustrated by Figure 9 in which the variation in spectral signature for a specific material—in this case, grass—is plotted. To generate this plot, a group of pixels containing grass were identified, and for each band, the mean, maximum, minimum, and standard deviation across all pixels were

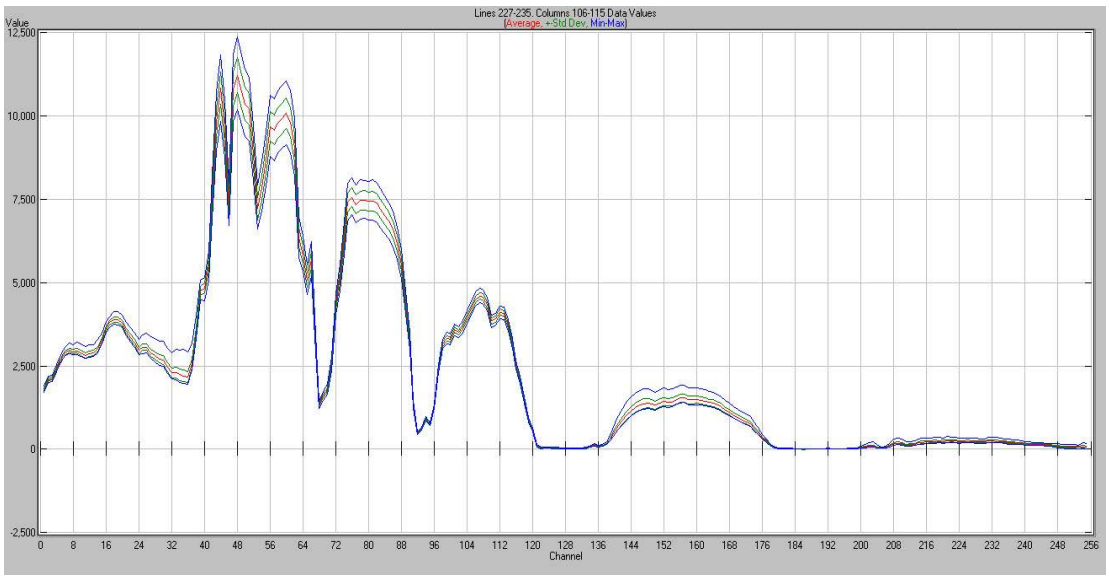


Figure 9. Example of Material Spectra Variation

computed. As shown in this plot, the spectral signature for a material is by no means deterministic. Rather, the signature can be expected to vary across pixels that contain the same material. The variability can result from variations in atmospheric conditions, the presence of other materials that are not visible at the spatial resolution of the image, sensor noise, and other factors. A result of this variability is that samples of two different

materials can conceivably have near-identical signatures, at least over portions of the EM spectrum, as seen in the grass and tree pixels in Figure 8.

Inspection of Figure 9 indicates that the variability of the radiance data within each band is not likely to be constant across bands. This condition would suggest a complex covariance or correlation structure across spectral bands in hyperspectral data. To show the truth in this assertion, Figure 10 provides a visual representation of the correlation matrix of a subset of the Fort A.P. Hill image, as shown in Figure 11. Figure 10 is a 255x255 pixel image with each pixel corresponding to an element of the correlation matrix. The color of each pixel indicates the approximate value of the respective correlation coefficient.

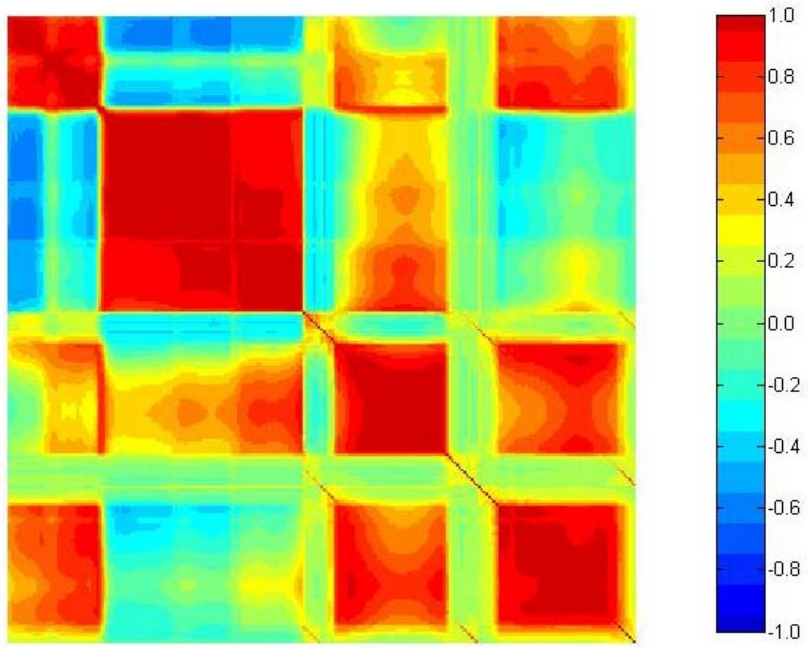


Figure 10. Correlation Matrix Example

From Figure 10, several conclusion can be made concerning the correlation structure of hyperspectral data. First, the red squares along the diagonal of the figure indicate strong positive correlation between bands that lie close together in the EM spectrum. This correlation suggests the use of data reduction techniques—such as principal component analysis (PCA), discriminant analysis, and the projection pursuit method proposed by Jimenez and Landgrebe (1999)—to reduce the dimensionality of the data. These methods are, in fact, commonly used in hyperspectral analysis when limited

Figure 11. Subset Image of Fort A.P. Hill

sample sizes are available in order to improve covariance and correlation estimates, and hence, classification accuracy.

A second implication of the correlation structure is that the entire matrix should be estimated rather than using simplifying assumptions such as band independence or constant variance across bands. Because the entire matrix must be estimated,

classification methods that rely on a covariance or correlation estimate, such as MLE classification, some types of cluster analysis, and Mahalanobis' Distance-based methods, can become computationally expensive.

A third conclusion that is evident in Figure 10, is that not all image bands provide useful information. In particular, the striations of near-zero correlation running through the matrix correspond to atmospheric absorption bands in which the energy at these wavelengths is almost entirely absorbed by the atmosphere. Consequently, the sensor detects primarily random noise at these wavelengths. This phenomenon is confirmed by the spectral plots in Figures 8 and 9 where it is seen that the energy intensity is near zero from bands 122 to 132, and from 179 to 199. These bands correspond to the strongest absorption wavelengths of 1.36 to 1.44 μm and 1.8 to 1.96 μm, respectively. Because these bands provide little information, they can be removed from the data set with little effect (though in some instances, these bands may provide a means of estimating noise covariance.)

The preceding discussion has provided a basic overview of hyperspectral imagery concepts so as to introduce the necessary terminology for subsequent chapters. For a more in-depth treatment of remote sensing, the reader is referred to Richards and Jia (1999). Landgrebe (2002) provides a more thorough description of hyperspectral imagery and associated classification issues. Jimenez and Landgrebe (1998) discuss the implications of high-dimensional hyperspectral data on MLE classification. Finally, for examples of covariance estimation methods for high-dimensional data with limited samples, see Hoffbeck and Landgrebe (1996), Tadjudin and Landgrebe (1999), and Jackson and Landgrebe (2002).

III. Overview of Existing Anomaly Detection Methods

One of the key components of the proposed target detection framework discussed in Chapter One is the anomaly detector. In this chapter, the general hyperspectral anomaly detection problem is discussed followed by a review of the related literature. The chapter concludes with a discussion of potential extensions that may improve the performance of anomaly detection methods.

The Anomaly Detection Problem

Hyperspectral anomaly detection is essentially a pattern classification problem in which each pixel vector is classified as either being anomalous to the image scene or as being a member of the scene's predominant materials—often referred to as the image background. As a point of departure for defining this classification problem, we first present the constant risk Bayes classifier as it pertains to hyperspectral data. For this classifier, we assume that a pixel vector, $\mathbf{x}$, is composed of one of C background materials, ω_i, $i = 1,\ldots,C$, and that $P(\omega_i)$ is the prior probability that an arbitrary pixel in the scene contains background material ω_i. From Bayes' formula, the discriminant function for this classifier is

$$g_i(\mathbf{x}) = P(\omega_i \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega_i)P(\omega_i)}{\sum_{j=1}^{C} p(\mathbf{x} \mid \omega_j)P(\omega_j)} \tag{3.1}$$

where

$p(\mathbf{x} \mid \omega_i) =$ is the density function of class ω_i,
$P(\omega_i \mid \mathbf{x}) =$ the posterior probability that a given vector, $\mathbf{x}$, is a member of class ω_i.

The decision rule for this classifier is to conclude that a pixel vector, $\mathbf{x}$, is composed of material i if $g_i(\mathbf{x}) > g_j(\mathbf{x})$ for all $j \neq i$.

In the context of anomaly detection, it is implied that there is a set, A, that contains the indices of materials considered anomalous. Theoretically, there is also a set, B, that contains the indices of background materials. Thus, to classify a pixel as an anomaly, it must be true that

$$\min_{i \in A} g_i(\mathbf{x}) > \max_{j \in B} g_j(\mathbf{x}). \tag{3.2}$$

Unfortunately, in the case of anomaly detection, the set A is not known, which obviously implies that the priors and densities are not known for the anomalous materials. Therefore, Bayes classification cannot be used directly. Depending on what is known of set B, however, the discriminant functions of (3.2) may still be of use in detecting anomalies. For instance, if $g_j(\mathbf{x})$ is sufficiently small for all j in B, we may conclude that the pixel does not fit well in any background classes, and therefore is anomalous. This idea of using background material classes to detect anomalies can be used to define three general types of anomaly detection problems: detection when background classes are known; detection when the background classes are unknown, but can be estimated; and detection when the background is unknown but is assumed to contain one class. These problem types are further defined in the following paragraphs.

Problem Type I: Background Classes Known

In this type of anomaly detection problem, it is assumed that the materials that comprise the majority of the image scene are known and can be defined in terms of class probability density functions and prior probabilities. Detecting anomalies in this case is a matter of computing the discriminant functions of (3.2) for a given pixel and declaring it

an anomaly if all $g_j(\mathbf{x})$ for the background classes is sufficiently small relative to the distributions of the background classes. Though this problem type is most appealing from a theoretical view, it is practically the least common problem since the background materials are rarely known with any confidence. However, if anomaly detection is to be performed on an image with a relatively uniform, simple background—a desert scene, for example—formulating the problem in this manner may be realistic and present good detection results.

Problem Type II: Background Classes Estimated

Though knowledge of the exact background classes may not be available, it may be possible to estimate the number of classes and their respective densities using the image scene. For example, cluster analysis can be used to segment the image pixels into similar groups. The densities can then be estimated and the detection problem solved as a Type I problem. An example of this approach is given by Carlotto (2005). A limitation of the Type II formulation is the number of background materials must be assumed either explicitly or implicitly via a threshold criterion. This assumption will directly impact the false alarms that occur during anomaly detection.

Problem Type III: Single Background Class Assumed

This problem type is similar to Type II in that there is no prior knowledge of the number or characteristics of the background classes. However, in this formulation it is assumed that all the background classes can be treated as one class, ω_b, with detection then proceeding as in the Type I problem. That is to say, a pixel is considered an anomaly if $g_b(\mathbf{x})$ is sufficiently small. To make this one-class assumption valid, the background density function to which a pixel is compared is estimated from a relatively

small window of pixels surrounding the pixel of interest. It is assumed that this window will contain a relatively homogenous set of pixels that can adequately represent a background class. The benchmark RX anomaly detection method proposed by Reed and Yu (1990) relies on this premise. Another approach for strengthening the one-class assumption is to use the entire image, or perhaps a larger window, to estimate a single, often multi-mode, density function for the background class. This approach has the advantage of providing more pixels for covariance estimation, but requires more sophisticated density estimation methods.

General Remarks and Challenges

Before reviewing the various anomaly detection methods found in the literature, several remarks should be made concerning the field of anomaly detection. First, the three types of anomaly detection problems defined in the preceding paragraphs are by no means exhaustive, but they do summarize the primary views of anomaly detection found in the literature. Second, problem Type II and III are the predominant formulations, especially in a military context, since ground truth data that conclusively identifies background materials is seldom available. Third, all of the problem types are essentially outlier detection problems in multi-dimensional space. From this viewpoint, several challenges to anomaly detection are evident.

First, to ensure the accuracy of an anomaly detector based on parametric classification, the background classes must be accurately identified and modeled, especially in the tails of any assumed density functions. For problem types II and III, this challenge can be formidable since, by definition, little is known about the background materials. A second challenge arises from noise and artifacts commonly found in

hyperspectral imagery. Specifically, a malfunctioning sensor or random noise can generate pixel signatures that are indeed anomalies, but are uninteresting and contribute to the false alarm rate. To complicate matters, image smoothing methods that are commonly used to remove noise and artifacts can also remove true anomalies, thereby confounding attempts at detection. A third challenge, associated primarily with window-based methods, is selection of a window size. If the window is too large, it may contain a large enough number of target pixels such that they no longer appear as anomalies. If the window is too small, there may be insufficient pixels to estimate the covariance matrix inverse that is common to many anomaly detection methods.

A final challenge of anomaly detection is determining a meaningful method for comparing the performance of different detectors. The root of the problem is anomalies, by definition, occur with a very low probability in an image scene. Therefore, simply computing the overall classification accuracy of a detector can be misleading since classifying every image pixel as background will result in a very high overall accuracy. For example, if 100 of the 28800 pixels in the scene of Figure 9 were true anomalies, classifying all the scene pixels as background gives an overall accuracy of 99.7%. This challenge is further complicated by a limited number of hyperspectral data sets that contain true anomalies verified by ground truth. Current methods used to evaluate anomaly detectors either produce a version of an operating characteristic curve that plots true positive fractions against the number of false alarms per square kilometer, or simply visually compare output images showing the location of anomalies found by candidate detectors.

Summary

In this section, basic concepts of hyperspectral imagery were introduced in order to provide an understanding of what the data actually represents, as well as insights to its basic characteristics. Also, the anomaly detection problem was more formally defined and subdivided into three basic problem types that are differentiated by the level of knowledge of the background materials in the hyperspectral scene. Finally, some of the more significant challenges of hyperspectral anomaly detection were discussed. In the following section, an outline of recent anomaly detection methods found in the literature is presented.

Literature Review

A review of the technical literature for anomaly detection methods indicates that research in this field can be divided into two general categories: local anomaly detection and global anomaly detection. Local detectors are characterized by the use of a processing window centered on a pixel of interest. The pixels in the window are used to estimate background material statistics. The pixel of interest is compared to this background model and a determination is made whether or not the pixel is an anomaly. The window is then centered around another image pixel and the process repeated. Global detectors, on the other hand, attempt to compare each pixel to a statistical model representative of the entire image, rather than just the neighborhood of pixels in the immediate vicinity of the pixel of interest. The following sections outline different local and global anomaly detectors proposed for hyperspectral imagery.

Local Anomaly Detectors

The RX Detector

The benchmark local anomaly detector to which other detection methods are compared is the Reed-Xiaoli (RX) detector proposed by Reed and Yu (1990). This detector was originally developed for multispectral imagery, but has proven effective for hyperspectral imagery, as well. As summarized by Stein, Beaven, Hoff, Winter, Schaum, and Stocker (2002), the RX detector is derived using a generalized likelihood ratio test (GLRT). To form the likelihood ratio, it is first assumed that a processing window contains a set of N background pixels, $\mathbf{v}_j$, $j = 1,2,\ldots,N$, with probability density function $p_0(\cdot,\theta_h)$, where θ_h, h=0,1, are unknown parameters that must be estimated for the null (h=0) and alternative (h=1) hypotheses. It is also assumed that there is a set of M pixels, $\mathbf{x}_t$, $t = 1,2,\ldots,M$, to be tested, where $p_1(\cdot,\theta_h)$ is the pdf of the test pixels. When the test set contains a single pixel, the likelihood ratio becomes

$$G(\mathbf{x}) = \frac{\max\limits_{\theta_1}\left\{p_1\left(\mathbf{x}_t,\theta_1\right)p_0\left(\left\{\mathbf{v}_j \mid 1 \le j \le N\right\},\theta_1\right)\right\}}{\max\limits_{\theta_0}\left\{p_1\left(\mathbf{x}_t,\theta_0\right)p_0\left(\left\{\mathbf{v}_j \mid 1 \le j \le N\right\},\theta_0\right)\right\}}. \tag{3.3}$$

If $G(\mathbf{x})$ is less than some specified threshold, then the null hypothesis that the pixel, $\mathbf{x}$, comes from a different distribution than the background pixels is supported. Under the assumption that the window pixel vectors have a normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, Reed and Yu show that the GLRT of (3.3) reduced to the following:

$$RX(\mathbf{x}) = \left(\mathbf{x}-\hat{\mu}\right)^T\left(\frac{N}{N+1}C + \frac{1}{N+1}\left(\mathbf{x}-\hat{\mu}\right)\left(\mathbf{x}-\hat{\mu}\right)^T\right)^{-1}\left(\mathbf{x}-\hat{\mu}\right) \tag{3.4}$$

where

$\hat{\mu}$ = the estimate of the window mean, and
C = the estimate of the window covariance matrix.

As the number of window pixels, N, gets large, (3.4) converges to the following:

$$RX_L(\mathbf{x}) = \left(\mathbf{x} - \hat{\mu}\right)^T C^{-1}\left(\mathbf{x} - \hat{\mu}\right) \tag{3.5}$$

which is simply the Mahalanobis squared distance between the pixel vector, **x**, and the mean of the processing window pixels. Equation (3.5) is the most commonly used form of the RX detector. Reed and Yu go on to show that the RX statistic under the null hypothesis has a Chi-Square distribution with L degrees of freedom, where L is the dimensionality of **x**. Because the distribution of the statistic under the null hypothesis is independent of the estimated parameters, the statistic has the constant false alarm rate (CFAR) property.

To apply the RX detector, a processing window—typically a square window—is centered on an image pixel and either (3.4) or (3.5) is computed. Using the resulting value, either the null hypothesis can be formally tested using the Chi-Square distribution and a desired confidence level, or it can be used to create a grey-scale image that visually depicts the pixels that produced a high RX value. A combination of the two methods can also be used to produce an image showing all pixels for which the null hypothesis is not rejected.

The primary limitation of the RX detector is the subjectivity associated with choosing the processing window size. In order to estimate the inverse covariance matrix, the window must contain at least as many pixels as the number of dimensions of the image, otherwise, the covariance matrix will be singular and the inverse undefined. Depending on the amount of clutter in the image, however, ensuring the window exceeds the dimensionality may result in a window that is too large to detect anomalies of interest, rather than isolated natural materials such as trees, rocks, etc. In cases where smaller

window sizes are desired, band selection or data reduction can be performed on the original data to reduce the dimensionality and allow a smaller processing window. Beyond ensuring that the window contains sufficient pixels for accurate covariance estimation, little guidance exists as to the best choice of window size.

In addition to window size specification, the RX detector also suffers from difficulties in detecting relatively large anomalies. In fact, this problem is related to the window size problem in that to detect large anomalies, a large window needs to be used to ensure the anomaly pixels do not dominate the statistics. If the window is too large, however, it may contain sufficient clutter to inhibit anomaly detection. This dilemma is a fundamental problem for all local detection methods that employ a single processing window.

RX-Related Methods

The basic RX detector has been extended in a number of different ways to account for its limitations. Chang and Chiang (2002) present variations of the RX detector that are useful for real-time anomaly detection in which each pixel is classified as its data is received by the sensor. Chang and Chiang also present an automatic threshold method for determining the value of the output statistic beyond which a pixel should be classified as an anomaly. Chang and Chiang's versions of the RX detector are derived from the insight that the RX detector in (3.5) is essentially a matched filter operating on the pixel of interest, **x**. That is, (3.5) is of the form

$$M_{\mathbf{d}}(\mathbf{x}) = \kappa \mathbf{d}^T (\mathbf{x} - \boldsymbol{\mu}) \tag{3.6}$$

where

$\mathbf{d}$ = the matched signal, and
κ = a scaling constant.

For the RX detector of (3.5), $\mathbf{d}^T = (\mathbf{x} - \boldsymbol{\mu})^T\mathbf{C}^{-1}$ and $\kappa = 1$. By viewing the RX detector as a matched filter, Chang and Chiang present the following three RX-like detectors:

Normalized RX Detector:

$$NRX(\mathbf{x}) = \left(\frac{\mathbf{x}-\boldsymbol{\mu}}{\|\mathbf{x}-\boldsymbol{\mu}\|}\right)^T \mathbf{C}^{-1}\left(\frac{\mathbf{x}-\boldsymbol{\mu}}{\|\mathbf{x}-\boldsymbol{\mu}\|}\right) \tag{3.7}$$

Modified RX Detector:

$$MRX(\mathbf{x}) = \left(\frac{\mathbf{x}-\boldsymbol{\mu}}{\|\mathbf{x}-\boldsymbol{\mu}\|}\right)^T \mathbf{C}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right) \tag{3.8}$$

RX Detector with Background Subtraction:

$$BRX(\mathbf{x}) = \left(\mathbf{x}-\mathbf{1}\right)^T \mathbf{C}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right) \tag{3.9}$$

Chang and Chiang assert that the detectors of (3.7)-(3.9) are only useful for detecting anomalies whose spectral signature are well-characterized by second-order statistics alone, and that materials whose signatures are characterized only by first-order statistics may go undetected. To counter this problem, the correlation matrix is used in the three detectors rather than the covariance estimate. This substitution gives the following three detectors:

Correlation-Based Normalized RX Detector:

$$CNRX(\mathbf{x}) = \left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right)^T \mathbf{R}^{-1}\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) \tag{3.10}$$

Correlation-Based Modified RX Detector:

$$CMRX(\mathbf{x}) = \left(\|\mathbf{x}\|\right)^{-1} \mathbf{x}^T\mathbf{R}^{-1} \tag{3.11}$$

Correlation-Based RX Detector with Background Subtraction:

$$CBRX(\mathbf{x}) = \left(\mathbf{x}-\mathbf{1}\right)^T \mathbf{R}^{-1}\mathbf{x} \tag{3.12}$$

Experiments conducted by Chang and Chiang using (3.10)-(1.12) show, based on visual inspection, improvement over (3.7)-(3.9) in detecting anomalies in HYDICE and AVIRIS hyperspectral sensor data sets. However, the true benefit in using the correlation-based detectors is their ability to be used in a real-time processing environment. Chang and Chiang contend that (3.7)-(3.9) cannot be executed as each pixel is acquired because the entire image is needed to compute the covariance estimate. However, by using methods presented by Chang and Chiang (2001) to compute the sample correlation matrix, (3.10)-(3.12) can be processed in real-time as pixels are acquired; this processing can even be conducted using a parallel architecture. An extension to (3.10)-(3.12) is given by Hsueh and Chang (2004) in which strong anomalies already detected during the real-time processing are removed from data to prevent the anomaly pixels from dominating and confounding future computations.

The automatic threshold method proposed by Chang and Chiang to identify anomalies using the detector output is relatively straightforward. First the desired metric from (3.7)-(3.9) is computed for every image pixel. A histogram is then constructed for these metric values. Using the histogram as an empirical distribution, any pixels with values exceeding a specified confidence level are considered anomalies. Though this method is more objective than visual inspection, the selection of the confidence level is still somewhat subjective.

According to Riley, Newsom, and Andrews (2004), the success of the basic RX detector is dependent on a high signal-to-noise ratio for the hyperspectral data. If several of the bands are particularly noisy, outliers in these bands can trigger false alarms with the RX detector. To overcome this problem, Riley et al propose modifying the RX

detector by using an estimate of the noise covariance matrix in place of the window covariance matrix in (3.5). This detector, referred to as the Weighted Euclidean Distance detector, has the property of computing the distance between the pixel of interest and the window's mean vector while giving more weight to the vector components corresponding to bands with low noise.

To estimate the noise covariance matrix, Riley et al. assume that a typical pixel signature within the processing window should be a relatively smooth function of wavelength. That is to say, a pixel vector such as the ones shown in Figure 8 should not have any sharp spikes. Thus, for each pixel vector in the window, the vector is first divided by the norm of the window mean vector to remove any gross fluctuations due to atmospheric attenuation or illumination. Then, the difference is computed between each vector component and an interpolation of the component before and after it in the vector. These differences are assumed to be caused by noise, and the variance of the differences is therefore assumed to provide an estimate of the noise variance. As an alternative to using the noise covariance matrix in (3.5), Riley et al. also propose adding a multiple of the noise covariance matrix to the window covariance matrix.

Dealing with Non-Gaussian Data

A key assumption for the success of the RX detector is that the distribution of the window pixels is a single Gaussian. This assumption permits mathematically tractable analysis of the generalized likelihood ratio given by (3.3), and hence, derivation of the RX detector. Hyperspectral data, however, is typically not Gaussian which means that the RX detector may produce high false alarm rates when non-linear distributions exist. Kwon and Nasrabadi (2005) propose a variant of the RX method that first uses radial-

basis kernel functions to implicitly transform the original data into a higher dimensional feature space where Kwon and Nasrabadi contend the Gaussian assumption is valid. The use of kernel functions allows the RX metric to be computed in the transformed space without explicitly transforming the data or having to compute the higher-dimensional covariance matrix. The kernel-based RX detector is compared to the RX detector using three different data sets. Operating curves for these tests indicate that the kernel-based RX method is significantly better at detecting anomalies, particularly at low false alarm rates.

Adapting RX for Detecting Large Anomalies

An RX-based method that attempts to overcome the RX detector's limited ability to detect large anomalies that span multiple pixels is given by Gaucel, Guillaume, and Bourennane (2005). The method is called whitening spatial correlation filtering (WSCF), and first entails applying a whitening transformation to the pixels in the processing window to produce data with zero mean vector and unit covariance. After the transformation, the RX detector becomes

$$I(\mathbf{x}) = \|\tilde{\mathbf{x}}\| \tag{3.13}$$

where

$\tilde{\mathbf{x}}$ = the whitened version of the pixel of interest, $\mathbf{x}$.

Now, if an anomaly is considerably larger than a single pixel, (3.13) will still fail to detect the anomaly because the anomaly pixels may dominate the statistics of the processing window. To counter this affect, Gaucel et al. incorporate a second term to (3.13) to inflate the detector output if the pixel's eight neighbors have the same magnitude and direction. The WSCF detector then becomes

$$I_{WSCF}(\mathbf{x}) = \|\tilde{\mathbf{x}}\| + \alpha \sum_{j \in v(i)} \rho_j \|\tilde{\mathbf{x}}_j\| \tag{3.14}$$

where

$v(i)$ = the set of eight pixel vectors corresponding to the neighbors of $\mathbf{x}$,
$\tilde{\mathbf{x}}_j$ = the whitened version of the jth neighbor of $\mathbf{x}$,
ρ_j = the correlation coefficient between $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}_j$, and
α = a scaling parameter.

Gaucel et al show that (3.14), with an appropriate value of the scaling parameter, performs better than the RX detector on a simulated data set. In their experiments, the authors adjust the scaling parameter until a maximum probability of detection is achieved for known anomalies. Unfortunately, little guidance is provided for selecting the parameter when nothing is known of the size or signature of the anomalies—a more likely scenario in actual applications.

Other Local Detector Methods

To overcome the RX detector's limited ability to detect large anomalies, Kwon, Der and Nasrabadi (2003) propose a dual-window detector that places a smaller processing window inside a larger window. The mean and sample covariance matrices are then computed for each window. The difference matrix between the two covariance matrices is also found and its corresponding eigenvalues computed. These eigenvalues are divided into two groups—one corresponding to negative values and the other to positive values. Kwon et al. assert that a small number of the positive eigenvalues can be used to extract spectrally distinct materials contained in the inner window. In other words, if the materials in the inner window are considerably different from those in the outer window, the difference between the two covariance matrices should have significant structure reflected by the positive eigenvalues and corresponding

eigenvectors. By projecting the difference between the outer and inner window mean vectors onto the space of the eigenvectors of the large, positive eigenvalues, the resulting scalar should have a large absolute value if the materials in the two windows are significantly different. This projection leads to what Kwon et al refer to as the dual-window eigen separation transform (DWEST) detector:

$$D(\mathbf{x}) = \left| \sum_{\mathbf{v}_i} \mathbf{v}_i^T \mathbf{m}_{diff}(\mathbf{x}) \right| \tag{3.15}$$

where

$\mathbf{v}_i$ = the ith eigenvector corresponding to the ith eigenvalue from the set of large, positive eigenvalues

$\mathbf{m}_{diff}(\mathbf{x})$ = the difference between the mean vectors of the outer and inner windows

Where the DWEST detector is designed to locate anomalies larger than a single pixel by using two nested windows, Liu and Chang (2004) extend this approach to find both large and small anomalies. To accomplish this task, Liu and Chang propose a three-window nested detector in which a small window corresponding to small anomalies is nested within a larger middle window corresponding to large anomalies. These two windows are nested within an even larger outer window which is used to model the image background. To determine if an anomaly exists in either the inner or middle window, Liu and Chang compute what Chang (2003) refers to as the orthogonal projection divergence (OPD) between the inner and outer window means, and between the middle and outer window means. The OPD between two vectors, $\mathbf{s}_i$ and $\mathbf{s}_j$, is given by

$$OPD(\mathbf{s}_i, \mathbf{s}_j) = \left(\mathbf{s}_i P_{\mathbf{s}_j}^{\perp} \mathbf{s}_i + \mathbf{s}_j P_{\mathbf{s}_i}^{\perp} \mathbf{s}_j \right)^{1/2} \tag{3.16}$$

where

$$P_{\mathbf{s}_k}^{\perp} = \mathbf{I} - \mathbf{s}_k \left(\mathbf{s}_k^T \mathbf{s}_k \right)^{-1} \mathbf{s}_k^T$$

Each term in the OPD metric is the residual of one of the vectors projected into the orthogonal subspace of the other vector. Hence, a large value for the OPD metric indicates that the two vectors point in different directions from one another. By computing the OPD for the inner and outer window and the OPD for the middle and outer window, a detector can be constructed that outputs the maximum of these two scores. If the detector outputs a significantly large number, the pixel of interest is considered an anomaly. Liu and Chang refer to this method of anomaly detection as nested spatial window-based target detection (NSWTD).

The NSWTD method and the DWEST detector are attempts to take into account the size of the anomaly when classifying a pixel. In other words, these methods utilize spatial information as well as spectral information to find anomalies. A more sophisticated method for using spatial information is given by Schweizer and Moura (2000) and Schweizer and Moura (2001). In this method, a three-dimensional Markov random field (MRFs) is used to capture the spatial correlation between pixels. Schweizer and Moura use inner and outer processing windows that are both further divided into smaller Markov windows. An approximate maximum likelihood method is used to estimate the MRF parameters. The GLRT is used to determine both a single hypothesis version of the anomaly detector and a binary hypothesis version. Execution of the detectors consists of moving the processing windows over each pixel, estimating the MRF parameters, computing an output statistic, and testing the appropriate hypothesis that the inner window is composed of background material only, or background material

and an unknown target material. The most significant contribution of Schweizer and Moura's method is that the number of computations grows linearly with the number of spectral bands. In comparison, the number of computations for the RX detector grows as the square of the number of bands. Schweizer and Moura contend that this characteristic allows more, potentially useful, information to be used for anomaly detection. Improvements in actual detection performance with the GMRF method, as tested by Schweizer and Moura, are only moderate compared to the RX detector.

Hazel (2000) presents a different approach to using GMRFs for anomaly detection in which GMRFs are first used to automatically segment the hyperspectral image into a specified number of classes. This segmentation information is then used in a GMRF anomaly detector to locate anomalies.

Goovaerts, Jacquez, Warner, Crabtree, and Marcus (2004) propose a method for using spatial information in anomaly detection that begins by performing a principal components analysis on the original image data. The results of the PCA are then used to produce *k* principal component (PC) images using the largest 75% of the PCs. For each of the *k* PC images, a filter window containing n pixels is passed over the entire image. The filter used for this task is:

$$m_k(\mathbf{u}) = \sum_{i=1}^{n} \lambda_{ik} z_k(\mathbf{u}_i) \tag{3.17}$$

where

$\mathbf{u}$ = the coordinates of the pixel being processed,
$z_k(\mathbf{u}_i)$ = the intensity value of the *i*th pixel in the window with coordinates $\mathbf{u}_i$,
λ_{ik} = a weight assigned to the *i*th pixel in the window when image k is being processed

The weights must sum to unity and are derived using a method developed by Goovaerts (1992) referred to as factorial kriging. This method determines the weights by taking into account the spatial correlation of the pixels in the processing window.

Once the filter is passed over each image, the residual for each pixel is computed between the original pixel intensity and the filter value. The end result of these computation is *k* images showing these residual values. Each of the residual images are then scanned for anomalies. The scanning is performed by defining an inner and outer processing window and computing what is referred to as the local indicator of spatial autocorrelation (LISA) statistic:

$$LISA(\mathbf{u}) = \bar{r}_k(\mathbf{u})\left(\frac{1}{J}\sum_{i=1}^{J} r_k(\mathbf{u}_i)\right) \tag{3.18}$$

where

$\bar{r}_k(\mathbf{u}) =$ the average residual value in the inner window, and

$r_k(\mathbf{u}_i) =$ the residual value of the *i*th pixel located at coordinate $\mathbf{u}_i$ in the outer window containing J pixels.

For anomaly pixels, the LISA statistic will produce large negative values since anomalies are expected to produce inner and outer window residual means of opposite sign. After performing this scan, *k* new images are obtained showing the LISA value for each pixel.

To determine which LISA scores indicate anomalies, Goovaerts, et al, use a Monte Carlo simulation to estimate the LISA distribution for each of the k images. The simulation entails randomly sampling *J* pixels to form the outer window and then re-computing the LISA scores. This process is repeated many times to form a sufficiently large sample of scores. The resulting sample is used as the empirical distribution of the

LISA statistic. With this distribution, a *p*-value is estimated for each pixel. The process is repeated for each of the *k* LISA images resulting in *k* new images showing the *p*-values of each pixel.

For the final step of the process, Goovaerts et al propose either computing the average *p*-value for each pixel across the *k* images (S1 metric), or computing the average absolute deviation of the *p*-value from 0.5 for each pixel across the *k* images (S2 metric). A threshold for these final statistics must be specified in order to determine which pixels are anomalies. No guidance is provided for determining this threshold.

Goovaerts et al test their proposed method using two hyperspectral data sets that image regions of Yellowstone National Park. Experiments study the effects of: the number of PC images used; the final output metric employed; the signal-to-noise ratio of the images; and the internal window size. In general, the ROC curve analysis from these experiments show that the method performs best using more PCs and the S2 metric. The method performs well when detecting anomalies in a relatively uniform background, but produces considerable false alarms in more complex images. The authors claim the method is robust to variations in signal-to-noise ratio; however, this conclusion appears somewhat subjective based on the presented results.

The poor performance of the Goovaerts et al method given a highly cluttered image is characteristic of most local detectors. Rosario (2004) attributes this problem to the fact that local detectors tend to reduce the complex background to a set of statistics that misrepresent the background. To overcome this problem, Rosario uses a three-window, logistic regression-based anomaly detector. All three windows are centered on the pixel to be tested. The inner window is hypothesized to contain anomalous material,

the middle window represents background material, and the third outer window also contains background material but is used to achieve a better estimate of the background variability. Using these windows, two sets of metrics are computed. The first set contains the spectral angle between each pixel vector in the outer window and the middle window mean pixel value. The second set contains the spectral angle between each pixel vector in the outer window and the inner mean pixel value. The two sets of metrics are assumed to have probability density functions (pdf) denoted by $g_0(\mathbf{x})$ and $g_1(\mathbf{x})$, respectively. If the inner window indeed contains anomalous pixels, it is assumed that $g_1(\mathbf{x})$ is an exponential distortion of $g_0(\mathbf{x})$:

$$\frac{g_1(\mathbf{x})}{g_0(\mathbf{x})} = \exp\left(\alpha + \beta\mathbf{x}\right) \tag{3.19}$$

where α and β are parameters to be estimated.

If, under the null hypothesis, the pixel vector being tested is not an anomaly, then the estimate for β in (3.19) should be equal to zero. Rosario outlines a procedure for estimating β, as well as the variance of the background pixels using an estimate of $g_0(\mathbf{x})$. A test statistic for the null hypothesis is also given that Rosario asserts is Chi-Square distributed. Hence, anomaly detection consists of computing the test statistic for each pixel vector and comparing it to a critical value from the Chi-Square distribution. Experimental tests of this detector on a single hyperspectral data set show drastically better ROC curve performance relative to the RX, DWEST, and two other common anomaly detectors.

Global Anomaly Detectors

The common element of the local anomaly detection methods is some form of moving processing window that is used to characterize the background materials in the immediate vicinity of the test pixel. Global detectors take a different approach by attempting to characterize the different background materials contained in the entire image and then determining if any pixels are not well-defined by these materials. In general, global methods can be divided into two groups: mixture model-based methods and distribution-based methods. These two types of detectors are discussed in the following sections.

Mixture Model Methods

A common view of hyperspectral imagery is that the image scene contains M distinct background materials, or endmembers, each of which has a characteristic spectral signature given by the vector $\mathbf{s}_m$. Each pixel vector in the image is then assumed to be a linear mixture of these pure signatures as well as additive noise. This linear mixture model is given by

$$\mathbf{x} = \sum_{m=1}^{M} \alpha_m \mathbf{s}_m + \mathbf{n} \tag{3.20}$$

where

α_m = the mixing, or abundance, fractions, and
$\mathbf{n}$ = additive noise.

With this model in mind, it is reasonable to assume that if the predominant endmembers in the image can be identified and then used to fit (3.20) to each pixel in the image, any pixels with a poor model fit are likely to be anomalies. This premise is the point of departure for mixture-based methods, with the primary differences between methods

being the manner in which endmembers are identified and the manner in which the abundance fractions are used.

Grossman, Bowles, Haas, Antoniades, Grunes, Palmadesso, Gillis, Tsang, Baumback, Daniel, Fisher, and Triandaf (1998) present a mixture detector system called the Optical Real-time Adaptive Spectral Identification System (ORASIS). The ORASIS searches the image for pixels, or exemplars, that span the feature space to within a user-specified tolerance—details are not provided as to how this search is accomplished. A PCA is then performed on the exemplars to determine the fundamental subspace of the image. The dominant PCs are then used to transform the exemplars into the reduced subspace. These transformed exemplars become the endmembers. For each endmember, a filter is constructed to detect the respective material, and these filters are each passed over the image to produce an abundance map for each endmember. A histogram-based method is then used to screen the abundance maps for those that best represent target materials—again, no details are provided on how this screening is accomplished. A final image is then produced that shows which pixels contained significant amounts of any of the target materials. A spatial filter corresponding to the hypothesized target shape is then passed over the image to further eliminate false alarms. The ORASIS detector is tested against a single hyperspectral data set. Grossman et al. conclude, based on visual inspection, that the ORASIS detector is effective in detecting anomalies, though no comparisons are made to other detection algorithms.

A method similar to ORASIS is the NFINDR algorithm with stochastic target detector (STD) discussed by Stein, et al. NFINDR, first presented by Winter (1999), is a method for extracting endmembers from a hyperspectral image. The premise behind

NFINDR is that for an image with *N* pure endmembers, the simplex formed with the endmembers as vertices will produce the largest volume in *N*-1 space. Further, every other pixel vector in the image will be a linear combination of the simplex endmembers. Thus, given a user-specified value for *N*, NFINDR looks for *N* pixels in the image that produce the largest volume in (*N*-1)-space. This task is accomplished by starting the search with a set of *N* random pixel vectors selected from the scene. Starting with the first pixel in the image, each pixel is substituted for one of the *N* pixels in the set and the volume of the simplex re-computed. If the volume increases, the substituted pixel is left in the set, the vector it replaced is discarded, and the next pixel in the image is evaluated. The process is repeated until the simplex volume fails to increase beyond a specified threshold. The final set of pixels represent the pure endmembers for the image. The endmembers are then used to perform a least squares fit for every pixel in the image using (3.20), and *N* abundance maps are then produced corresponding to the *N* endmembers. The abundance maps are simply images whose pixel values indicate the relative amount of the respective material contained in the pixel.

As outlined by Stein et al., the STD portion of the method consists of using the abundance maps to identify target-like endmembers. It is hypothesized that target materials will be relatively rare, producing abundance maps with relatively few intense pixels. It is also assumed that actual targets will have an abundance value near 1.0 in the target endmember abundance maps. Using these assumptions, histogram analysis is used to screen the abundance maps for those that represent target endmembers, and pixels with abundance values close to unity in these maps are marked as anomalies. Winter (2004) applies the NFINDR/STD method to finding surface mines. Results of this experiment

show that the method performs exceptionally well when targets are larger than the spatial resolution of the image, but tends to produce an increased number of false alarms when a pixel is spatially larger than the targets.

An alternative method to finding image endmembers is the Greedy Monte Carlo (GMC) linear unmixing method proposed by Clare, Bernhardt, Oxford, Murphy, Godfree, and Wilkenson (2003). In the GMC method, a sample of pixel vectors is randomly chosen from the image. Individually, each pixel vector in the sample is used to fit (3.20) to all the image pixels. The sample vector that produces the minimum sum of absolute residuals over the entire image is selected as the first basis vector. A second sample is then selected, and the best vector—combined with the first basis vector—that provides the best fit to the image pixel is selected for the second basis vector. This process continues until the sum of absolute residuals fails to decrease. Since the first few basis vectors selected may not fit the data very well, Clare et al. propose, once convergence has been achieved, removing the first basis vector and finding a new vector to take its place. The same procedure is conducted for the second basis vector, and so on, until convergence is again achieved. Once a basis is selected, the corresponding vectors are used to fit (3.20) to every pixel in the image. The pixels with the largest residuals are considered anomalies. Tests conducted with a single data set indicate that the GMC detector has better ROC curve performance than the RX detector for extremely low to mid-range false alarm rates.

A similar approach to the GMC detector is the iterative error analysis (IEA) approach proposed by Neville, Staenz, Szeredi, Lefebvre, and Hauff (1999) which also iteratively finds basis vectors to minimize the overall error of the unmixed image.

Details of how the basis selection is performed for IEA were not available at the time of this writing.

Distribution-Based Methods

Where global mixture-based anomaly detectors attempt to unmix each pixel vector according to (3.20) and identify those pixels containing high concentrations of target-like endmembers, distribution-based methods are concerned with finding probability distributions that globally model the data, and then identifying anomalous pixels that are outliers for these distributions. Stein, Beaven, Hoff, Winter, Schaum, and Stocker (2002) propose that the global distribution is a mixture of Gaussian distributions defined by the following pdf:

$$g(\mathbf{x}) = \sum_{c=1}^{C} \pi_c N\left(x \mid \mu_c, \Sigma_c\right);\ \pi_c \geq 0;\ \sum_{c=1}^{C} \pi_c = 1 \tag{3.21}$$

where

C = the number of material classes in the image,
π_c = the probability of class c,
μ_c = the mean vector of class c, and
Σ_c = the covariance matrix of class c.

Anomaly detection, according to Stein et al., becomes a matter of estimating the parameters for each class, segmenting the image using a maximum a posteriori (MAP) classifier, and designating as anomalies those pixels that do not fit well in their assigned class. Stein et al. propose using a stochastic expectation maximization (SEM) method proposed by Moon (1993) for parameter estimation, though no guidance is given for determining the number of classes contained in the image. A related anomaly detection method that also tests pixel fit relative to mixtures of the *C* classes is the stochastic mixing model described by Schaum and Stocker (1997).

An alternative approach for determining the image classes is given by Catterall (2004) who proposes using a version of *k*-means clustering to group the image pixels into *k* classes, where *k* is presumably specified by the user. Caterall goes on to model each class with a Multivariate Normal Inverse Gaussian (MNIG) distribution. This distribution is recommended because it is better able to fit unimodal, heavy-tailed distributions that are characteristic of hyperspectral data. An expectation-maximization method introduced by Øigård and Hanssen (2002) is used to estimate the MNIG parameters. Once a distribution is fit to each image class, the negative log-likelihood of each pixel belonging to its respective class distribution is computed. High-values of the negative log-likelihood function indicate a pixel is an anomaly. The threshold value of the function at which anomalies are declared is user-specified. A simple comparison of the MNIG detector to the RX detector using a single hyperspectral image visually indicate better detection capability with the MNIG detector.

Carlotto (2005) proposes a cluster-based anomaly detector (CBAD) that is similar to the MNIG detector. The primary difference between the two methods is that Carlotto simply computes the Mahalanobis distance between each pixel and its assigned cluster. In other words, Carlotto assumes a Gaussian distribution of the clusters rather than a heavy-tailed distribution, as suggested by Stein et al. By making this assumption, the CBAD method is computationally less demanding than the MNIG detector, though the accuracy may not be as good.

A limitation of the methods suggested by Carlotto and Caterall is that the number of clusters to use is subjective. To get around this problem, Chang (2003) discusses a projection-base method originally proposed by Chiang, Chang, and Ginsberg (2001) that

looks for global outliers without explicitly modeling each material class. The method begins by whitening the hyperspectral data so that the projection procedure is translation invariant. A variant of projection pursuit—originally introduced by Friedman and Tukey (1974)—is then used to find a projection vector, **a**, that projects the image data into a single dimension in which a projection index (PI) is maximized. Proposed PI's are skewness or kurtosis of the projected data since these metrics are generally indicators that outliers are present in otherwise Gaussian data. A genetic algorithm is used to search for the best value of **a** that maximizes the PI. In theory, the resulting projected image data should have high skewness or kurtosis which is assumed to be caused by anomalous materials. To find the threshold value of the projected data that represents an anomaly, a histogram is constructed. It is then assumed that a zero value in the histogram represents a separation between background pixels and anomaly pixels. Hence, the first zero value found in the histogram is used as the threshold value above which a pixel is considered an anomaly.

Under the assumption that other projections may also reveal anomalies, a zero vector—the mean of the whitened data—is placed in the columns of the whitened data matrix corresponding to the anomaly pixels identified with the first projection. A second projection is then determined and anomalies identified in the same manner as with the first projection. This process continues until the PI converges to zero, indicating that no additional outliers exist. The anomalies found from each stage of the method are combined to form a final binary image indicating the location of the anomalies. Chang shows tests of the method on an image containing known anomaly targets. The method

locates all the anomalies after three projections with few false alarms. There is no guidance as to how many projections to use for an arbitrary image.

Achard, Landrevie, and Fort (2004) further investigate the method of Chiang et al. by using a modified method for finding the projection vectors. Rather than initiate the genetic algorithm search with an arbitrary set of projection vectors, the eigenvectors from a PCA and minimum noise fraction (MNF) analysis are first used to project the original image data. The eigenvectors producing the largest PI are then included in the initial generation of the genetic algorithm search. Achard et al. test their method against two hyperspectral data sets containing known anomalies and compare it to projecting with the PCA eigenvectors only and to projecting with each of the original pixels in the image. For each type of projection, the first six projections are used for anomaly detection with skewness and kurtosis used for the PI. Results of the tests showed that using the PCA eigenvectors to initialize the genetic algorithm search with kurtosis as the PI is more successful than the other tested methods at locating anomalies with few false alarms. However, Acard et al. also fail to provide any objective method for the number of projections to use in finding anomalies. This limitation can be problematic since at some point in the detection procedure projections will contain more and more false alarms.

Literature Review Summary

In the preceding pages, the significant anomaly detection methods found in the literature were presented. These methods are classified as either local detectors that locate anomalies relative to the pixels in a local neighborhood, or as global detectors that attempt to characterize the global distribution of the image pixel vectors and find outliers relative to this distribution. The local detectors are further categorized as being similar to

the benchmark RX detector or as employing a significantly different approach. The global detectors are further described as those based on the linear-mixture model of (3.20) or as distribution-based methods. As mentioned earlier, all of these methods follow the underlying theme of finding outliers relative to some assumed statistical model. Additionally, it is evident that none of these methods use multivariate outlier detection methods to accurately estimate detector statistics. It will become evident in Chapter 5 that this is a serious omission that may degrade the detection accuracy of the methods outlined in this chapter.

IV. Overview of Invariant Target Detection Methods

Introduction

A significant challenge of hyperspectral data classification is accurate comparison of pixel signatures to known material signatures collected in laboratory conditions. The primary cause of this challenge is the pixel vector elements that comprise the pixel signature typically report the energy radiance detected by the sensor at the respective band wavelengths, whereas library signatures for a material typically report the percent of incident energy reflected by the material at different wavelengths. Therefore, a conversion must be made from either radiance signatures to reflectance signatures, or vice versa. Conversion from radiance to reflectance entails removing the effects of skylight, viewing geometry, path radiance, and atmospheric conditions from sensor radiance measurements to obtain an estimate of the reflectance that would be obtained in the laboratory. Conversion from reflectance to radiance entails combining these effects with the laboratory reflectance signature to obtain an estimate of the sensor radiance reading. In either case, atmospheric and viewing geometry parameters must be known or estimated for a hyperspectral image for the conversion to be made. These parameters may not be available to the scientist attempting to classify a hyperspectral image, and if they are, the atmospheric and illumination models required for the conversion may not provide sufficient accuracy to enable an accurate and usable conversion.

As outlined in Richards and Jia (1999), a common method for circumventing the conversion problem is to identify training pixels for each class within the image itself and use these pixels to define the probability distributions for maximum likelihood classification. This method eliminates the need for any conversion, but it shifts the

burden to identifying a sufficient number of training pixels to adequately define the class probability distributions. This task is further complicated by the requirement for ground truth data to verify the class membership of training pixels. These problems can generally be overcome for land-cover classification studies of geographic areas in which ground truth data is attainable, but in the case of target detection studies, these problems present a more formidable challenge.

A significant difference between target detection and land-cover classification studies is the target material of interest often does not appear in the image with very high frequency; therefore, manually defining a sufficient number of target material training pixels may not be possible. Also, military target detection may further be complicated by an inability to gather ground truth data for the hyperspectral scene. These factors severely limit the practicality of defining training sets, and argue for some form of signature matching detection that does not require atmospheric correction of the hyperspectral data.

In the following sections, a signature matching method proposed by Healey and Slater (1999) that is invariant to the atmospheric conditions and viewing geometry associated with a hyperspectral image is discussed. This method is referred to as invariant subspace target detection. Existing extensions to Healey and Slater's method are also presented. An understanding of these invariant subspace target detection methods provides the context for the target detection methodology given in Chapter 6.

The Original Method and Extensions

The invariant subspace target detection method, originally proposed by Healey and Slater (1999) for the purpose of detecting pure-pixel targets, attempts to define the subspace of radiance signatures that a target material's reflectance signature is mapped into as a function of atmospheric conditions and viewing geometry. With this subspace defined, any pixel vector from an arbitrary hyperspectral radiance image that lies within the subspace is designated a target. To define the target material's radiance subspace, Healey and Slater begin with the radiance model for a pixel located at coordinate (*x*,*y*) given in Equation 2.1 and restated here:

$$\begin{aligned} L(x,y,\lambda) = {} & T_u\left(z_g, z_v, \theta_v, \phi_v\right) R\left(x,y,\lambda\right) \\ & \cdot\left[K T_d\left(z_g, \theta_0, \phi_0, \lambda\right) E_0\left(\lambda\right)\cos\theta_0 + \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi/2} E_s\left(\theta,\phi,\lambda\right)\cos\theta\sin\theta d\theta d\phi \right] \\ & + P\left(z_g, z_v, \theta_v, \phi_v, \lambda\right) \end{aligned} \tag{4.1}$$

where

$T_u\left(z_g, z_v, \theta_v, \phi_v\right)$ = the upward atmospheric transmittance,
$R\left(x,y,\lambda\right)$ = the spectral reflectance of the material at wavelength λ,
K = a binary constant used to account for occlusion of the pixel,
$T_d\left(z_g, \theta_0, \phi_0, \lambda\right)$ = the downward atmospheric transmittance,
$E_0\left(\lambda\right)$ = the extraterrestrial solar radiance,
$E_s\left(\theta,\phi,\lambda\right)$ = the scattered sky radiance due to skylight,
$P\left(z_g, z_v, \theta_v, \phi_v, \lambda\right)$ = the path-scattered radiance,
z_g = the altitude of the pixel,
z_v = the altitude of the sensor,
θ_v = the polar angle of the sensor relative to the pixel,
ϕ_v = the azimuthal angle of the sensor relative to the pixel,
θ_0 = the polar angle of the sun relative to the pixel,
ϕ_0 = the azimuthal angle of the sun relative to the pixel, and
λ = the wavelength of interest.

If the laboratory reflectance signature for a target material is known, then it is theoretically possible to use (4.1) to determine the radiance of the material detected by the sensor for a given set of viewing geometry and atmospheric parameters. By systematically using different combinations of these parameters, (4.1) can be used to generate a sample set of the possible radiance spectra that can be produced from the reflectance signature. To use (4.1) for this purpose, explicit functions for T_u, T_d, E_0, E_s, and P must be known; however, closed forms for these functions may be difficult to obtain. As a surrogate for (4.1), Healey and Slater use the MODTRAN 3.5 atmospheric modeling program to generate a set of radiance spectra for different combinations of solar zenith angle, atmospheric gas profiles, aerosol profiles, and sensor altitudes.

Once the sample set of radiance spectra are obtained for the target material's reflectance spectra, the subspace in which they lie can be estimated. To accomplish this task, Healey and Slater recommend using spectral value decomposition (SVD) to find an orthonormal set of basis vectors—the eigenvectors of the sample radiance spectra—that can be used to define the radiance subspace. To determine the dimensionality of the subspace, and hence the number of basis vectors to use, it is assumed that each of the sample radiance vectors can be approximated by a linear combination of the basis vectors. That is to say,

$$L_i \approx \sum_{j=1}^{N} a_{ij}\mathbf{m}_j \qquad 1 \le i \le C \tag{4.2}$$

where

L_i = the ith radiance spectrum from the sample set,
C = the total number of sample radiance spectra generated,
N = the assumed dimensionality of the radiance subspace,
a_{ij} = the weighting coefficient for the jth basis vector, and
$\mathbf{m}_j$ = the jth basis vector.

In this model, it is assumed that the N basis vectors used to define the subspace are the N eigenvectors from the SVD corresponding to the N largest eigenvalues.

If all the eigenvectors are used for the model of (4.2), then there will be no residual error between the L_i and their respective linear combination of the basis vectors, where the error for the ith radiance spectra is given by:

$$E_i(N) = \left\| L_i - \sum_{j=1}^{N} a_{ij}\mathbf{m}_j \right\|^2 . \tag{4.3}$$

However, using all the eigenvectors may produce a relatively large subspace that may overlap the radiance subspaces of other materials. Thus, finding the appropriate number of basis vectors to use to define the radiance subspace reduces to finding the minimum value of N that produces a sufficiently low total squared error given by:

$$E_T(N) = \sum_{i=1}^{C} E_i(N). \tag{4.4}$$

By minimizing the number of basis vectors, it is hoped that sufficient separability will result between the target material subspace and other material subspaces. Experimental tests conducted by Healey and Slater using 498 material spectra from the USGS spectral library indicate that using N=9 basis vectors produces a sufficiently low error to adequately model an arbitrary material's radiance subspace. Thus, selecting the nine largest eigenvectors from the SVD of the sample set of radiance spectra is expected to

provide a good estimate of the radiance subspace for a target material's reflectance spectra.

Once the subspace basis vectors are determined, they can be used to detect the presence of the target material in a hyperspectral radiance image. To perform this detection, the pixel vectors in the image are first normalized since the size of the approximation error given by (4.3) depends on scalings of the spectral vector. For an arbitrary pixel vector, L, the normalized vector is simply,

$$\hat{L} = \frac{L}{\|L\|} \tag{4.5}$$

It is then assumed that the normalized vector is a linear combination of the subspace basis vectors. In other words,

$$\hat{L} = \sum_{j=1}^{N} \alpha_j \mathbf{m}_j + \mathbf{n}, \tag{4.6}$$

where $\mathbf{n}$ is an error term. Under the assumption that the error terms are independent with constant variance, the maximum likelihood estimates for the α_j terms are given by:

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{M}^T \mathbf{M}\right)^{-1} \mathbf{M}^T \hat{L}, \tag{4.7}$$

where the columns of $\mathbf{M}$ are the basis vectors. Also due to the independent and identically distributed (iid) error assumption, the estimates of (4.7) minimize the errors, which are given by

$$r = \left\| \hat{L} - \sum_{j=1}^{N} \hat{\alpha}_j \mathbf{m}_j \right\|. \tag{4.8}$$

These error terms can be thresholded in order to determine if a pixel is comprised of the target material. Thus, invariant subspace target detection consists of cycling through all the image pixels and using (4.7) to estimate the bases multipliers for each pixel. The

error terms for each pixel are then computed using (4.8), and any errors that have a sufficiently low probability of belonging to a Gaussian distribution with zero mean and unit variance are designated as targets. It is important to note that once the basis vectors have been determined for a target material, they can be used to detect targets in any number of arbitrary hyperspectral radiance images. Thus, the potentially time consuming task of generating a representative sample of radiance spectra using the MODTRAN 3.5 model needs to be performed only once in order to identify the subspace basis vectors. This feature makes invariant subspace target detection an attractive option for the target detector component of the proposed target detection framework.

Extension to Sub-pixel Targets

Healey and Slater's original invariant subspace target detector was designed for detecting pure-pixel targets. Thai and Healey (1999) adapt the original methodology for use in detecting sub-pixel targets. This extension is made by first assuming that an arbitrary pixel in a hyperspectral radiance image can be modeled in the following manner:

$$\mathbf{y} = \mathbf{B}\boldsymbol{\varphi} + \mathbf{T}\boldsymbol{\theta} + \mathbf{n} = \mathbf{A}\begin{bmatrix}\boldsymbol{\varphi}\\ \boldsymbol{\theta}\end{bmatrix} + \mathbf{n} \tag{4.9}$$

where

$\mathbf{B}$ = a matrix whose columns form the basis of the image background,
$\mathbf{T}$ = a matrix whose columns form the basis of the target material radiance subspace,
$\boldsymbol{\varphi}$ = a vector of multipliers for $\mathbf{B}$,
$\boldsymbol{\theta}$ = a vector of multipliers for $\mathbf{T}$, and
$\mathbf{n}$ = a vector of error terms.

The basis vectors that comprise $\mathbf{T}$ are found using the same procedure used for the original invariant subspace target detector. In order to find the basis vectors that

comprise **B**, the hyperspectral image is arranged as the matrix, **Y**, in which each row corresponds to a pixel vector from the image. To ensure the background subspace defined by **B** has as little overlap with **T** as possible, any pixel vectors that are fit well by **T** are removed from **Y**. A singular value decomposition is then performed on **Y** to produce an orthonormal set of vectors that can be used to define the subspace of the image background.

The question then arises as to how many basis vectors to include in **B** so as to adequately define the background subspace. Thai and Healey assert that enough vectors should be include to account for a minimum amount of the total variance, but not so many vectors that the background and target subspaces overlap significantly. Also, any background basis vectors that are used should not project well into the target subspace, again to prevent overlap of the two subspaces. Basis selection thus proceeds by adding eigenvectors to **B** until the variance accounted for by the vectors is between a specified upper and lower threshold. Also, the magnitude, δ_j, of the projection onto **T** of the *j*th vector added to **B** must be below a specified threshold. That is to say,

$$\delta_j = \left\| \mathbf{T}^T \mathbf{u}_j \right\| \leq \eta \tag{4.10}$$

where

$\mathbf{u}_j$ = the *j*th basis vector, and
η = the specified threshold

The threshold value used in (4.10) should be arbitrarily close to zero since a vector that is orthogonal to **T** will have a δ_j-value of zero.

Under the assumption that the **n**-terms from (4.9) are i.i.d. with a Gaussian distribution, then the multipliers in (4.9) can be estimated as

$$\begin{bmatrix} \hat{\boldsymbol{\varphi}} \\ \hat{\boldsymbol{\theta}} \end{bmatrix} = \mathbf{A}^{+}\mathbf{y} \tag{4.11}$$

where

$$\mathbf{A}^{+} = \text{the pseudo-inverse of } \mathbf{A}.$$

The i.i.d. assumption of the error terms further leads to the formation of a generalized likelihood ratio that can be used to test if the target material is present in the pixel, **y**. This ratio is given by

$$\tilde{\Lambda}(\mathbf{y})^{2/l} = \frac{\mathbf{y}^{T}\left(\mathbf{I}-\mathbf{B}\mathbf{B}^{\mathrm{T}}\right)\mathbf{y}}{\mathbf{y}^{T}\left(\mathbf{I}-\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\right)\mathbf{y}} \propto \frac{p\left(\mathbf{y} \mid \text{Target and Background } (\mathrm{H}_1)\right)}{p\left(\mathbf{y} \mid \text{Background Only } (\mathrm{H}_0)\right)} \tag{4.12}$$

where

$\mathbf{I}$ = the identity matrix, and
$\mathbf{Q}$ = the Gram-Schmidt transformation of $\mathbf{A}$.

Implementation of Thai and Healey's sub-pixel target detector consists of first determining the basis, **T**, for the desired target material. As mentioned previously, this basis only needs to be determined once for a given material and can then be used for any number of arbitrary hyperspectral radiance images. Once **T** is defined, the hyperspectral image is divided into sub-regions and the background subspace basis, **B**, is then determined for each sub-region. For each pixel in the image, the ratio of (4.12) is computed using the appropriate **B**-matrix for the pixel's sub-region. Any pixels with a sufficiently high ratio are initially designated as containing the target material. However, as indicated by Thai and Healey, it is possible that a pixel with a high ratio contains a target spectrum component, $\mathbf{T\theta}$, of (4.9) that contains negative elements. Thus, the initial set of designated target pixels are screened using the following validity check:

$$\hat{\mathbf{t}} = \mathbf{T}\hat{\boldsymbol{\theta}} \geq \mathbf{0}. \tag{4.13}$$

Any designated pixels that satisfy (4.13) are designated as actual target pixels.

In order to better determine the dimensionality of the background subspace basis, **B**, Thai and Healey (2002) provide a modification to the sub-pixel detection method. In particular, the method for selecting the number of basis vectors for the columns of the **B**-matrix is reduced to finding a value of *i* that maximizes the following ratio:

$$K_i = \frac{\tilde{\Lambda}_i\left(\bar{\mathbf{y}}_t\right)}{\tilde{\Lambda}_i\left(\bar{\mathbf{y}}_b\right)}, \qquad M \le i \le N \tag{4.14}$$

where

$\bar{\mathbf{y}}_t$ = the mean spectral vector over the set of vectors in **T**,
$\bar{\mathbf{y}}_b$ = the mean spectral vector over the region of the image used to define the background,
$\tilde{\Lambda}_i\left(\mathbf{y}\right)$ = the generalized likelihood ratio obtained for **y** using a background model with *i* basis vectors,
M = the minimum number of vectors required to ensure a minimum threshold value of variance is accounted for by the basis vectors, and
N = the maximum number of vectors required to ensure a maximum threshold value of variance is not exceeded by the basis vectors.

The motivation for using this method to select the number of vectors in **B** can be understood by considering an image pixel that contains both target and background material. Such a pixel can be modeled by (4.9). If we also assume that the target and background subspaces will overlap if too many basis vectors are used for **B**, then it can be shown that (4.14) will increase as basis vectors are added to **B**. However, as the overlap between subspaces increases, then the ratio will begin to decrease. This decrease occurs because the subspace defined by **B** will eventually contain enough of the target subspace to allow a target vector to be written as a linear combination of the vectors in **B**. This condition will result in low values of the likelihood ratio, even if the target material

is present, since the vectors in **T** are no longer required to model the pixel. Finding the value of *i* that maximizes (4.14) helps prevent this phenomenon from occurring, thus ensuring high values of the ratio in (4.12) when the target material is present in the pixel.

Another modification of the invariant sub-pixel target detector is presented by Zhang and Gu (2004) who assert that using singular value decomposition or PCA to derive the target and background subspace bases is imprecise if the subspaces contain nonlinearities. As an alternative method for defining the subspaces, Zhang and Gu use kernel PCA with a radial basis kernel to specify the subspace as a linear combination of the kernel matrix elements. In this new method, the **T** and **B** matrices in Thai and Healey's formulation of (4.9)-(4.12) are replaced by kernel subspace matrices $\mathbf{T}_k$ and $\mathbf{B}_k$. After making this substitution, Zhang and Gu's method proceeds in the same manner as the original invariant sub-pixel target detection method.

Zhang and Gu's assertion that SVD is not necessarily the best method for selecting the basis vectors is verified by research conducted by Bajorski, Ientilucci, and Schott (2004) and Bajorski and Ientilucci (2004). In the former study, Bojorski et al. test three different basis selection methods for determining the background matrix, **B**. The first method is the SVD approach used by Thai and Healey. The second method is the Pixel Purity Index (PPI) method described by Boardman, Kruse, and Green (1995) that projects all the image pixel vectors in thousands of different random directions in the spectral space. Pixels that repeatedly receive very low or very high projection magnitudes are identified as endmember materials that can be used as basis vectors. The final selection method is referred to as the Maximum Distance (MaxD) method and was originally proposed by Lee (2003). MaxD forms the basis by searching for pixel vectors

that form the corners of a simplex in the spectral space. Details of this method are provided in Bajorski et al. The comparison tests of the three methods indicate that the MaxD method produces the best detection results for an AVIRIS image of a complex urban scene, while the SVD method is best for a less complex HYDICE image. No rationale is given for this results, but it is evident that the scene complexity may influence the performance of invariant sub-pixel target detection.

In all the versions of the sub-pixel detection method discussed up to this point, designation of a pixel vector as a target is based on the generalized likelihood ratio (GLR) given in (4.12). However, derivation of (4.12) assumes that the residuals in the model given by (4.9) are i.i.d. Gaussian random variables. According to Bajorski, Ientilucci, and Schott (2004) and Manolakis, Siracusa, and Shaw (2001), this assumption is usually not valid for hyperspectral data. To account for this problem, Liu and Healey (2004) propose that the class conditional density functions used in the GLR be estimated using non-parametric methods. Specifically, Liu and Healey calculate the magnitudes of the residuals for all the image pixels under the null hypothesis that no target material is contained in the pixel, as well as the magnitudes of the residuals for all the image pixels under the alternative hypothesis that the pixels contain both target and background materials. For each of these sets of magnitudes, a histogram is constructed to serve as the empirical density functions required for the GLR of (4.12).

Other Extensions to the Original Method

In the derivation of the original invariant subspace target detection method, Healey and Slater assume that the surface normal of an irradiated pixel emanates from the center of the Earth. That is to say, it is assumed that the pixel is flat. To account for

pixels that may actually have an aspect component—that lie on a hill, or slope—Slater and Healey (1999) develop a function that generates new radiance spectra given a radiance spectra for a flat pixel. With this function, the output radiance spectra from MOTDRAN 3.5 using a combination of atmospheric and viewing geometry parameters can be used to generate additional radiance spectra corresponding to different surface orientations.

Just as the original invariant subspace target detector did not consider surface orientation of the pixel, it also did not consider the zenith angle of the sensor. That is to say, the original method only considered the viewing geometry parameters of solar zenith angle and sensor altitude, but assumed that the sensor was located at the nadir of the pixel. Suen, Healey, and Slater (2001) extend the original method by incorporating the sensor zenith angle—θ_v in Figure 5—as a parameter to be varied in generating the sample set of radiance vectors. This modification to the original methodology increases the total number of sample vectors that must be generated with MODTRAN, but leaves the remainder of the detection process intact.

In some target detection studies, the target material signature may be derived from the image itself rather than from laboratory measurements. This scenario may unfold when anomaly detection is used to identify potential targets in an image and it is desired to find these same targets types in other parts of the image or in different images all-together. Slater and Healey (2001) adapt their original invariant subspace methodology to operate in these scenarios. The significant component of this new methodology is a functional relationship that relates the radiance signature of the target identified in the original image to potential variants of the signature that may occur under different

atmospheric conditions in another image. This functional relationship can be used in conjunction with MODTRAN 4 to generate a sample set of radiance spectra for the target. With this sample set, the original invariant subspace detection method can be applied in the usual way. Again, the primary use for this extension is to identify and track interesting target materials found in one image across other images taken at different times under different atmospheric and viewing geometry conditions.

Another variant of target detection arises when the target material of interest is actually a mixture of materials. This scenario may arise in a military context when the target of interest is painted in a multi-color camouflage scheme and the color pattern has a smaller dimension than an image pixel. Suen and Healey (2001) present an invariant detection method for detecting these types of targets. The method begins by assuming that under a set of atmospheric and viewing geometry conditions specified by index j, a pixel signature, $\mathbf{p}_{j,\beta}$, for a pixel containing the target mixture is a linear mixture of the N different materials:

$$\mathbf{p}_{j,\beta} = \sum_{k=1}^{N} \beta_k \mathbf{s}_{kj} + \mathbf{n} \tag{4.15}$$

where

β_k = the fraction of the pixel occupied by the kth material (these coefficients must be non-negative and sum to unity),
$\mathbf{s}_{kj}$ = the radiance spectra of the kth material under a set of atmospheric and viewing geometry conditions identified by j, and
$\mathbf{n}$ = a Guassian error term.

With this assumption, Suen and Healey use MODTRAN 4 to generate a set of L radiance signatures for each mixture material, where each of the signatures corresponds to a different combination, j, of atmospheric and viewing geometry parameters. Starting with

an arbitrary pixel in the image, the β_k coefficients in (4.15) are estimated to minimize the error between the actual pixel signature and the fitted value from the mixture model. These coefficients are estimated using quadratic programming for each of the L conditions.

After the coefficients have been estimated for the pixel, the resulting L sets of coefficients are searched to find the set, j, that produced the smallest residual error. Because the residuals are assumed to be Gaussian noise with zero mean and constant variance, the magnitude of the residual for condition set j can be thresholded to determine if the pixel contains a mixture of the N target materials. This process is repeated for each pixel in the image, though the L radiance spectra for each of the target materials do not need to be regenerated.

The invariant subspace target detection methods discussed up to this point have been concerned with detecting spectral signatures that are similar to a target spectra. As the spatial resolution of hyperspectral imagery improves, however, it is also possible to detect target materials by matching the texture of a target material to the texture of a pixel's surrounding region. Shi and Healey (2005) exploit this idea by using the concept of multiband correlation functions and invariant subspace methods to detect targets based on the target's texture. Element (m, n) of the multiband correlation matrix between bands L_i and L_j for the pixel located at coordinate (x, y) is given by:

$$C_{ij}(m,n) = E\left\{\left[L_i(x,y) - \bar{L}_i\right]\left[L_j(x+m,y+n) - \bar{L}_j\right]\right\} \tag{4.16}$$

where

$\bar{L}_i$ = the mean vector for band i, and
$\bar{L}_j$ = the mean vector for band j.

Computing these matrices for a subset, *W*, of all band pair combinations in the hyperspectral image provides a means for discriminating different target textures. However, the mutliband correlation matrices for a specific texture are sensitive to atmospheric and viewing geometry parameters. To account for this problem, Shi and Healey use the DIRSIG synthetic image generation model of Schott, Brown, Raqueno, Gross, and Robinson (2002) to generate a set of images corresponding to different environmental conditions. From this set of images, a representative set of multiband correlation matrices can be computed for each target texture. Shi and Healey then present a method for determining the subspace defined by these matrices. Target texture detection is then performed by computing the distance between the vector representation of a pixel's *W* multiband correlation matrices and the subspace of the target texture vector. Sufficiently small values of this distance indicate that the pixel texture matches the target texture. A statistical test for significance of the distance metric does not exist.

Summary

The preceding discussion provided an overview of invariant subspace target detection methods that rely on the MODTRAN4 radiative transfer model to generate sets of target radiance signatures that are representative of a target in an image scene. These methods serve as a background to the target methodology proposed in Chapter 6 that replaces the MODTRAN4-based signatures with a set of target signatures derived from in-scene information, thereby making this target detection approach more accessible to a wide range of image analysts.

V. Improved Anomaly Detection Using Multivariate Outlier Methods

Detecting anomalies in hyperspectral imagery is essentially a multivariate outlier detection problem in which it assumed that the background materials in the image constitute a set of homogeneous populations that are potentially contaminated by anomalous pixel vectors. When anomaly detection is viewed in this light, it would seem natural that the numerous outlier detection methods and principles proposed in the statistical literature are applicable to finding hyperspectral anomalies. However, as indicated by the literature review in Chapter III, classical outlier detection methods have yet to find their way into the field of hyperspectral analysis. Moreover, the negative effects that outliers impose on classical statistical methods are seldom addressed by current anomaly detection methods, if they are even acknowledged at all. This omission is particularly troublesome since many of the anomaly detection methods rely on the Mahalanobis distance and other covariance matrix-based metrics which are extremely sensitive to the presence of even a small number of outlying observations. It is the subject of this chapter to investigate this problem further and to propose a methodology for using outlier detection methods to find anomalies in hyperspectral data. It is shown that such a method is capable of finding anomalies at low false alarm rates relative to the benchmark RX detector and a cluster-based anomaly detector. Particular emphasis is placed on developing an anomaly detector that can be applied with minimal input from the user.

The remainder of the chapter proceeds by first discussing the basic problems imposed by outliers and surveying the existing technical literature on multivariate outlier detection methods. With this background in-hand, the significance of outliers in the

hyperspectral context is demonstrated using simulated hyperspectral data. Experimental tests are then presented that indicate the BACON algorithm of Billor, Hadi, and Velleman (2000) and the FAST-MCD algorithm of Rousseeuw and van Driessen (1999) are amenable to hyperspectral anomaly detection. To use the BACON and FAST-MCD algorithms in an autonomous fashion it is necessary to apply them to homogenous data. To this end, we demonstrate that the *k*-means clustering algorithm is a reasonable method for clustering hyperspectral image data into homogeneous groups, and we also evaluate different methods for automatically determining the value for *k*. We then combine the BACON and FAST-MCD methods with the *k*-means algorithm to produce an autonomous anomaly detector, and use Taguchi robust parameter design methods to produce a final algorithm that consistently produces high detection accuracy across a range of hyperspectral images. Finally, it is shown that the robustly configured algorithm, referred to as AutoDet, is superior to two benchmark anomaly detectors when applied to a range of actual hyperspectral images.

Key Outlier Detection Concepts

The challenge of dealing with outliers in statistical data has persisted for centuries. As described by Barnett and Lewis (1994), Daniel Bernoulli wrote the following statement in 1777 concerning his analysis of astronomical observations:

> I see no way of drawing a dividing line between those that are to be utterly rejected and those that are to be wholly retained; it may even happen that the rejected observation is the one that would have supplied the best correction to the others. (Bernoulli and Allen, 1961).

In 1852, seventy-five years after Bernoulli's apparent frustration with the handling of outlying observations, the first journal article pertaining to outliers was written by Benjamin Pierce. In his article, Pierce draws the following conclusion:

> In almost every true series of observations, some are found, which differ so much from the others as to indicate some abnormal source of error not contemplated in the theoretical discussions, and the introduction of which into the investigations can only serve…to perplex and mislead the inquirer. (from Barnett and Lewis, 1994).

From these comments by Bernoulli and Pierce, it is evident that the presence of outliers and their ability to mislead a well-intentioned scientist have been recognized since the outset of formal scientific analysis. Over the years, much theoretical work has been conducted to formalize the outlier problem and to offer statistical methods that are either robust to their presence or can be used to ascertain their existence. For a thorough discussion of these developments, the reader is directed to the text by Barnett and Lewis (1994) or the journal article by Beckman and Cook (1983). From the material presented by these authors, we can distill several key concepts that point to the relevance of outlier methods to the problem of finding anomalies in hyperspectral data. Specifically, the concepts of breakdown point, masking, and swamping are discussed in the following paragraphs in order to set the stage for using multivariate outlier detection methods for anomaly detection.

Estimator Breakdown Point

Barnett and Lewis attribute the concept of a breakdown point to Hodges (1967) and Hampel (1968) (1971) who used it to describe the resistance of robust estimation methods to the presence the outliers. In simple terms, the breakdown point of an estimator is the fraction of arbitrary contaminating observations that can be present in a sample before the value of the estimator can become arbitrarily large. In other words, the

breakdown point specifies the fraction of outliers in a sample that can theoretically cause the estimator to produce values that are meaningless since they cannot be bounded in any way.

For location and covariance estimators—the two estimators that are most germane to outlier detection—Lopuhaa and Rousseeuw (1991) give more formal definitions of the breakdown point. For a location estimator, $\mathbf{t}_n$, at a collection of observations, $\mathbf{X}$, the breakdown point, $\varepsilon^*(\mathbf{t}_n, \mathbf{X})$, is defined as:

$$\varepsilon^*\left(\mathbf{t}_n, \mathbf{X}\right) = \min_{1 \le m \le n}\left\{\frac{m}{n} : \sup_{\mathbf{Y}_m}\left\|\mathbf{t}_n\left(\mathbf{X}\right) - \mathbf{t}_n\left(\mathbf{Y}_m\right)\right\| = \infty\right\} \tag{5.1}$$

where

$\mathbf{Y}_m =$ a collection of n observations corrupted by replacing m observations from $\mathbf{X}$ with arbitrary values.

From (5.1), it can be seen that the breakdown point for a location estimator is the smallest fraction of a sample that can be corrupted by outliers before the distance between the true sample mean and the corrupted sample mean can become arbitrarily large.

The formal definition of the breakdown point for the covariance estimator, $\mathbf{C}_n$, is given by Lopuhaa and Rousseeuw to be:

$$\varepsilon^*\left(\mathbf{C}_n, \mathbf{X}\right) = \min_{1 \le m \le n}\left\{\frac{m}{n} : \sup_{\mathbf{Y}_m} D\left(\mathbf{C}_n\left(\mathbf{X}\right), \mathbf{C}_n\left(\mathbf{Y}_m\right)\right) = \infty\right\} \tag{5.2}$$

where

$$D\left(\mathbf{A}, \mathbf{B}\right) = \max\left\{\left|\lambda_1\left(\mathbf{A}\right) - \lambda_1\left(\mathbf{B}\right)\right|, \left|\lambda_p\left(\mathbf{A}\right)^{-1} - \lambda_p\left(\mathbf{B}\right)^{-1}\right|\right\}, \text{ and}$$

$\lambda_i\left(\mathbf{A}\right) =$ the ith ordered eigenvalue of $\mathbf{A}$.

In words, (5.2) states that the breakdown point for a covariance estimator is the smallest fraction of a sample that can be corrupted by outliers before the difference between the largest eigenvalues of the true covariance estimate and that of the corrupted covariance estimate becomes arbitrarily large, or the difference between the smallest eigenvalues of the two estimates is arbitrarily close to zero.

In the context of estimating the mean vector and covariance matrix for a sample of data, it is advantageous to use estimators with as high a breakdown point as possible, with the theoretical limit being 50%, as explained by Rousseeuw and Leroy (1987). Unfortunately, the breakdown points for the classical mean and covariance estimators are only $1/N$, where N is the sample size (Donoho and Huber, 1983). In other words, the classical mean and covariance estimators can potentially produce unbounded estimates, in the sense of (5.1) and (5.2), with as little as one contaminating observation present in the sample. Extending this idea further, any metric that uses the classical mean and covariance estimate is also prone to breakdown with only a single outlier. Since the Mahalanobis distance is such a metric, any method that relies on this distance should not be trusted if outliers are suspected to be in the sample. All known variants of the RX anomaly detector fall in this category of suspicion. In the multivariate analysis world, the generally accepted remedy for this problem is to obtain robust estimates of the mean and covariance. These estimates can then be used in Mahalanobis distance-based methods to detect the presence of outliers. Methods that follow this prescription are outlined later in this chapter.

The Masking Effect

In addition to estimator breakdown, the phenomenon of outlier masking also argues for the use of outlier detection methods for detecting hyperspectral anomalies. Masking refers to the condition of very strong outliers distorting non-robust mean and covariance estimates to such a degree that weaker outliers appear ordinary in terms of their Mahalanobis distances. The degree of masking is measured in terms of an increase in Type II error, or false negatives, since observations that are truly outlying are classified as part of the uncontaminated population of data.

To formalize the concept of masking, Becker and Gather (1999) developed the outlier detector masking breakdown point that specifies the smallest fraction of outliers in a sample that can induce the masking affect. Becker and Gather prove that the masking breakdown point for an outlier detector that uses a mean and covariance estimator is bounded by the breakdown points of these two estimators. Further, if the two estimators have the same breakdown point, then the masking breakdown point of the detector is equal to the estimator breakdown point. An immediate conclusion that can be drawn from these findings is that non-robust Mahalanobis distance-based outlier detectors can be affected by masking in the presence of a single outlying observation. Since hyperspectral anomaly detectors are essentially multivariate outlier detectors, this conclusion is also relevant to finding hyperspectral anomalies. Therefore, multivariate outlier detection methods that are resistant to masking should also be considered for finding hyperspectral anomalies.

The Swamping Effect

A further reason for employing multivariate outlier detection methods for anomaly detection is to combat the swamping effect. Where masking refers to the increase of Type II error due to the presence of outliers, swamping refers to the increase in Type I error caused by outliers. As explained by Hadi (1992) in the context of Mahalanobis distance-based outlier detection methods:

> …not all observations with large [Mahalanobis distance] values are necessarily outliers. For example, a small cluster of outliers will attract [the mean vector] and will inflate [the covariance estimate] in its direction and away from some other observations which belong to the pattern suggested by the majority of observations, thus yielding large [Mahalanobis distance] values for these observations.

In other words, swamping occurs when outliers sufficiently distort the mean vector and covariance estimate so that good observations are incorrectly classified as outlying.

With the concept of swamping in mind, it can be argued that relatively poor receiver operating characteristic curve performance of Mahalanobis distance-based anomaly detectors is due in-part to swamping of the detector. In particular, if large anomalies are present in the processing window of a RX-type detector, the anomalous pixels may distort the mean vector and covariance matrix to the extent that false alarms occur. To ensure against this source of false alarms, multivariate outliers detection methods should be employed that use robust estimation methods for the mean vector and covariance matrix. Following this strategy helps ensure that the false alarm rate for an anomaly detector is inline with the accepted alpha-level for the detector.

Desirable Detector Properties

Based on the foregoing discussion, multivariate outlier detection methods with high breakdown point and resistance to the masking and swamping effects are generally

desirable for actual applications. In deciding if a detector's breakdown point is sufficient for a given detection scenario, the number of anticipated outliers should be considered. For a problem with N observations in p dimensions, if less than $N/(p+1)$ outliers are expected in the dataset, detectors with a breakdown point of $1/(p+1)$ may work perfectly well, alleviating the need to use higher-breakdown methods that may be more computationally complex. If it is impractical or impossible to judge the fraction of outliers contained in a sample, high breakdown methods provide a more conservative and reliable option for detecting the outliers.

To determine if a detector is resistant to the masking affect, the result of Becker and Gather (1999) can be applied. Specifically, if a detector has constituent estimators with known breakdown points, then the masking breakdown point will be no less than the smallest of the estimator breakdown points. Hence, the breakdown point of the detector can be used as a guide for assessing its resistance to masking. There is no similar result that formally explains a detector's resistance to swamping. For Mahalanobis distance-based detectors, however, it would seem intuitive that a detector's resistance to swamping is linked to its ability to accurately estimate the mean vector and covariance matrix for the good observations. If the mean vector and covariance matrix are accurately estimated, then it is less likely that the swamping effect will cause good observations to be labeled as outliers. Therefore, the breakdown point of the detector should also provide an indicator for the detector's resistance to swamping—high-breakdown detectors should not experience swamping unless the fraction of outliers exceeds the detector's breakdown point.

Another desirable property of an outlier detection method is affine equivariance. A detector is affine equivariant if the results it produces do not depend on translations, rotations, or changes of scale of the original data. A detector possesses the affine equivariance property if the estimators used in the detector are themselves affine equivariant. For Mahalanobis distance-based detectors, the detector is affine equivariant if the location and covariance estimators used to compute the distance are affine equivariant. Referencing Rousseeuw and Leroy (1987), a location estimator, *T*, is affine equivariant if and only if

$$T\left(\mathbf{AX}+\mathbf{b}\right)=\mathbf{A}T\left(\mathbf{X}\right)+\mathbf{b} \tag{5.3}$$

where

$\mathbf{X}$ = the p-dimensional data matrix,
$\mathbf{A}$ = a non-singular $p\times p$ matrix, and
$\mathbf{b}$ = a p-dimensional vector.

A covariance estimator is affine equivariant if and only if

$$C\left(\mathbf{AX}+\mathbf{b}\right)=\mathbf{A}^{T}C\left(\mathbf{X}\right)\mathbf{A}. \tag{5.4}$$

Again, the primary benefit of using affine equivarient methods to find outliers is the data can be translated, rotated or scaled without affecting the detection results. In the context of hyperspectral imagery, this property is particularly important since data transformation techniques such as principal component analysis are often applied to the data to reduce the dimensionality. Additionally, hyperspectral data is often scaled to reduce the effects of uncalibrated data, the atmosphere, and varying dynamic ranges between bands.

Examples of affine equivariant detection methods identified by Rousseeuw and Leroy are convex peeling (Barnett, 1976, Bebbington, 1978), ellipsoidal peeling (Helbling, 1983, Titterington, 1978), classical Mahalanobis distance methods, iterative

deletion, iterative trimming (Gnanadesikan and Kettenring, 1972), and depth trimming. Some of these methods, as well as additional outlier detection techniques, are discussed further in the following section.

Multivariate Outlier Detection Literature

The preceding section introduced the concepts of breakdown point, masking, and swamping to explain the impact outliers can have in foiling efforts to reveal them. With these ideas in mind, we now discuss the various methods that have been proposed over the years to detect outliers. These methods can be divided into two general groups: robust Mahalanobis distance-based methods, and non-traditional methods. The robust distance methods use some form of robust estimation to obtain mean vector and covariance estimates for the data. The Mahalanobis distance is then computed for each observation using these robust estimates, and observations whose distances exceed a critical value—generally from the Chi-square distribution if the data is multivariate normal—are labeled as outliers. For the non-traditional methods, the Mahalanobis distance is either not used for detection, or it is not used in a robust form. Rather, some alternative statistic is exploited that is presumably better at revealing outliers or that is computationally easier to compute than distances based on robust mean and covariance estimates. These groups will be discussed individually in the following sections.

Robust Distance Methods

Of all the multivariate outlier detection methods found in the literature, robust distance-based methods are the most numerous. In order to better explain how these methods have evolved over the last two decades, they are presented in roughly chronological order in the following paragraphs. Also, in an effort to keep the focus of

this discussion on existing outlier detection methods, much of the theoretical work in the related area of robust estimation is not addressed here. For the reader who is interested in the theory underpinning some of these robust distance methods, articles by Maronna (1976), Tyler (1988), Lopuhaa (1989), Lopuhaa and Rousseeuw (1991), Lopuhaa (1992), Butler, Davis and Jhun (1993), Rocke (1996), and Becker and Gather (1999) are suggested.

M-Estimation Method

One of the earliest robust distance methods was proposed by Campbell (1980) who suggested using *M*-estimators to obtain robust mean vector and covariance matrix estimates. *M*-estimators were originally proposed by Maronna (1976) as an affine equivariant method for obtaining robust mean vector and covariance matrices for possible use in linear discrimination, principal component analysis, and outlier detection. The M-estimates of a location vector, **t**, and a scatter matrix, **V**, are defined as the solution to the following system of equations:

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} u_1\left[\left\{\left(\mathbf{x}_i-\mathbf{t}\right)^T \mathbf{V}^{-1}\left(\mathbf{x}_i-\mathbf{t}\right)\right\}^{\frac{1}{2}}\right]\left(\mathbf{x}_i-\mathbf{t}\right)=\mathbf{0}, \\ &\frac{1}{n}\sum_{i=1}^{n} u_2\left[\left(\mathbf{x}_i-\mathbf{t}\right)^T \mathbf{V}^{-1}\left(\mathbf{x}_i-\mathbf{t}\right)\right]\left(\mathbf{x}_i-\mathbf{t}\right)\left(\mathbf{x}_i-\mathbf{t}\right)^T=\mathbf{V}, \end{aligned} \tag{5.5}$$

where the functions u_1 and u_2 are functions of the Mahalanobis distance that must satisfy certain assumptions. In general, these functions serve as weighting functions that minimize the impact outlying observations have on the mean and covariance estimates. Different forms of the weighting functions have been proposed in the literature.

To find a solution for (5.5), iterative methods are typically employed; however, there is no guarantee that the global optimum can be found. As determined by Maronna,

a weakness of these estimators is a breakdown point of only $1/(p+1)$, which can be problematic if operating in high-dimensional space.

MVE and MCD Methods

As a high-breakdown point alternative to the M-estimation method, Rousseeuw (1983) proposes the minimum volume ellipsoid (MVE) and minimum covariance determinant (MCD) as methods for estimating the location and scatter of the data. The MVE method searches for the minimal volume ellipsoid that encompasses at least h of the observations, with h taken as $[n/2]+1$. The mean vector estimate is the center of the ellipsoid, and the covariance is the ellipsoid itself multiplied by a correction factor to achieve consistency with a multivariate normal distribution. In a similar manner, the MCD looks for the sub-sample of h observations whose covariance matrix has the smallest determinant. The mean vector is then taken as the mean of the h observations, and the covariance estimate is the covariance of the h observations multiplied by a consistency factor. Upon obtaining the MVE or MCD estimates, they are then used to compute the Mahalanobis distance of all the observations to detect outliers. The advantage of the MVE and MCD is their high breakdown point of 50%, which makes them very useful for highly contaminated data. A disadvantage of these estimators is the combinatorial optimization problem that must be solved to find their exact solutions. In practice, search heuristics are employed to find approximate solutions.

A practical means for searching for an approximate MVE solution is proposed by Rousseeuw and Leroy (1987) and again by Rousseeuw and van Zomeren (1990). This method—referred to as the resampling method—entails drawing m sub-samples of size $p+1$ from the original data, where m is chosen to ensure a high probability that at least

one sub-sample will be free of outliers. For each sub-sample, the covariance matrix is computed and either inflated or deflated to include *h* of the observations from the original sample. The volumes of each of the *m* resulting ellipsoids are then approximated, and the one with the minimum volume is used to form the MVE estimate. To improve the efficiency of the MVE estimate, Rousseeuw and Leroy go on to recommend a *reweighting* step in which the mean vector and covariance matrix are recomputed using only the observations whose Mahalanobis squared distance relative to the MVE mean vector and covariance matrix fall below a suitable quantile of a Chi-Square distribution with *p* degrees of freedom. This reweighting step is also recommended by Rousseeuw and van Zomeran (1990), while Lopuhaa and Rousseeuw (1991) show that it preserves the breakdown point of the MVE.

Stahel-Donoho Estimator Method

In addition to suggesting the MVE and MCD estimators for use in robust distance outlier detectors, Rousseeuw and Leroy (1987) also allude to using Stahel-Donoho estimators in the robust distance computation. These estimators, proposed independently by Stahel (1981) and Donoho (1982), compute the mean vector and covariance matrix by assigning decreasing weight to observations that are outlying relative to some projection of the data to univariate space. Specifically, *outlyingness* of an observation $\mathbf{x}_i$ is defined to be:

$$u_i = \sup_{\|\mathbf{v}\|=1} \frac{\left|\mathbf{v}^T\mathbf{x}_i - \underset{j}{med}\left(\mathbf{v}^T\mathbf{x}_j\right)\right|}{\underset{k}{med}\left|\mathbf{v}^T\mathbf{x}_k - \underset{j}{med}\left(\mathbf{v}^T\mathbf{x}_j\right)\right|} \tag{5.6}$$

where

$\mathbf{v}$ = a *p*-dimensional projection vector.

Upon determining the u_i for all observations, the mean vector and covariance matrix are estimated as:

$$T(\mathbf{X}) = \frac{\sum_{i=1}^{n} w(u_i)\mathbf{x}_i}{\sum_{i=1}^{n} w(u_i)} \tag{5.7}$$

and

$$V(\mathbf{X}) = \frac{\sum_{i=1}^{n} w(u_i)(\mathbf{x}_i - T(\mathbf{X}))(\mathbf{x}_i - T(\mathbf{X}))^T}{\sum_{i=1}^{n} w(u_i)}, \tag{5.8}$$

where $w(u_i)$ is a positive, decreasing weighting function.

The Stahel-Donoho estimator is an attractive robust estimator because it has a high breakdown point which asymptotically approaches 50%, as shown by Donoho (1982). However, as explained by Rousseeuw and Leroy, the primary difficulty with these estimators is the computation of the outlyingness values. Apparently, no satisfactory method has been proposed to find these values, thereby preventing these estimators from experiencing any practical use for outlier detection. However, Gasko and Donoho (1982) propose a method that uses these estimators to identify leverage points in multiple regression data.

Hadi's Forward Search Method

Returning to the MVE-based outlier detection method proposed by Rousseeuw and Leroy (1987) and Rousseeuw and van Zomeren (1990), Hadi (1992) identifies several limitations with the approach. First, the user must decide upon the number of sub-samples to use in the resampling scheme. This choice is not obvious since it depends on the presumably unknown fraction of outliers that exist in the data. A second limitation

is that the covariance matrices for the sub-samples are estimated using only p+1 observations which could lead to singularities or highly inaccurate estimates. The final problem highlighted by Hadi is that several of the sub-samples may have covariance determinants close to zero, leaving the user with the task of choosing which sub-sample to use to form the MVE estimate. Since these sub-samples may have considerably different covariance structures, their resulting MVE estimates will likely be different. Thus, choosing the correct sub-sample is not obvious.

To correct for the limitations of the original MVE resampling method, Hadi proposes an MVE-based, non-affine equivariant outlier detector that begins by computing the vector of coordinate-wise medians for the original data. The median vector is then used to estimate the covariance matrix for the data. These location and covariance estimates are then used to compute robust Mahalanobis distances for the observations. The $[(n+p+1)/2]$ observations with the smallest distances are identified and used to form classical mean vector and covariance estimates and a new set of distances for all the observations. From this latest set of distances, the p+1 observations with the smallest distances are selected to form what is referred to as the basic subset. This basic subset is analogous to a sub-sample in the MVE resampling method with two notable differences. First, the basic subset is composed of observations closest to the centroid of the sample as determined by the robust, coordinate-wise median Mahalanobis distances. Second, there is only one basic subset in Hadi's method as opposed to potentially hundreds of sub-samples in the resampling MVE method. This considerable reduction in the number of subsets makes Hadi's method less computationally complex and faster to execute.

Once the basic subset is formed, it is then used to estimate a new mean vector, covariance matrix, and Mahalanobis distances. The distances are sorted and used to create a new basic subset that is one additional observation larger in size then the previous subset. This process continues until the basic subset contains $h=[(n+p+1)/2]$ observations—this value of h is chosen in order to be consistent with the method of Rousseeuw and Zomeren (1990). When the final basic subset is obtained, its mean vector and covariance matrix are estimated. A small-scale correction factor is then applied to the covariance matrix and Mahalanobis distances are computed for all observations. Since the distribution of the resulting distances are not known without knowing the distribution of the original data, Hadi suggests graphically inspecting the distances for outlying observations. If the original data can be assumed Gaussian, then the squared distances can be compared to a suitable quantile of the Chi-Square distribution with p degrees of freedom. Minor modifications to the stopping criteria, covariance correction factor, and initial basic subset formation for Hadi's method are given by Hadi (1994).

Atkinson's Forward Search Method

Sharing the same concerns with the MVE resampling method as Hadi, Atkinson (1993) proposed an affine equivariant forward search algorithm similar in nature to Hadi's method. Atkinson's forward search method begins by randomly selecting a subset of $m=p+1$ observations and using this subset to estimate a mean vector and covariance matrix. The covariance matrix is inflated or deflated to include h of the original observations, and the volume of the resulting matrix is recorded. The adjusted covariance matrix is then used to compute the Mahalanobis squared distances for all observations and the $m+1$ observations with the smallest distances are used to repeat the

process, while any observations whose squared distances exceed a critical Chi-Square threshold are identified as potential outliers. When $m=n$, the entire process is repeated with a new random subset of $m=p+1$ observations. After executing the algorithm through the desired number of random starting subsets, the adjusted covariance matrix that gave the smallest volume over all trials can be used for the final robust mean and covariance estimates and subsequent outlier detection. However, Atkinson does not recommend identifying outliers in this manner. Rather, he uses a graphical method known as stalactite plots to analyze which observations consistently emerged as outliers in each stage of the algorithm. Examples of Atkinson's method are given by Atkinson (1994).

Hawkins' Feasible Solution Algorithm

Motivated by the need to use efficient starting solutions for M-estimation and other iterative robust estimators, Hawkins (1994) proposed the Feasible Solution Algorithm (FSA) for obtaining approximations to Rousseeuw's MCD estimator. Hawkins also suggests that the MCD estimate resulting from the FSA can be used to detect outliers using the usual robust distance scheme. The FSA begins by first assuming that there are at most h outliers in the data. A random sample of $(n-h)$ observations is then selected from the original sample of n observations, with the remaining h observations trimmed from the data. The randomly selected observations are used to form an initial mean vector and covariance estimate along with the respective covariance determinant. Next, for each possible pair of observations with one observation coming from the randomly selected subset and the other from the trimmed subset, an updating formula provided by Hawkins is used to determine the reduction in covariance determinant if the pair of observations is interchanged between subsets. The pair of

observations that produces the greatest reduction in the covariance determinant are then swapped and the process repeated until no swaps can be identified that reduce the determinant value. The subset of *n-h* observations that results after no further improvements can be made is referred to as a feasible solution. The entire process is then repeated to find additional feasible solutions. The final MCD estimate is obtained from the feasible solution that produced the smallest covariance determinant.

Hawkins claims that the MCD estimate resulting from the FSA satisfies necessary conditions for global optimality, but not the sufficient conditions. In other words, a global solution will also be an FSA solution, but FSA solutions are not always global solutions. To improve the chances of finding the global solution with the FSA, Hawkins suggests increasing the number of random starts of the algorithm. For very small problems with $n<50$ and $p<=6$, simulation studies conducted by Hawkins indicate that 100 random starts ensure a 0.99 probability that the final FSA solution is the global MCD. No guidance is provided for data sets of larger magnitude.

Because the FSA requires evaluation of all pairs of observations from the two subsets at each iteration, the algorithm does not scale to large data sets very well. To account for this problem, Hawkins and Olive (1999) modify the original FSA to significantly reduce the number of pairs that must be evaluated at each iteration, thus making the algorithm more conducive to large data sets.

Compound Estimation Method

The robust distance outlier detection methods discussed to this point follow one of three strategies: 1) use of what Rocke and Woodruff (1996) refer to as smooth estimators, such as M-estimators or Stahel-Donoho estimators; 2) use of combinatorial estimators

such as the MVE or MCD; and 3) use of forward search methods as proposed by Hadi and Atkinson. In an effort to unify these strategies under one outlier detection method, Rocke and Woodruff propose a compound estimation outlier detection method that culminates the research of Rocke and Woodruff (1993), Woodruff and Rocke (1993), Woodruff and Rocke (1994), and Rocke (1996). The high-breakdown point, affine equivariant detector is composed of two phases. The objective of Phase I is to obtain a robust estimate of the data set's location and shape. This estimate is achieved by first using Hawkins' FSA to obtain an approximate MCD estimate of the location and shape. The MCD estimate is then used for the starting point of Atkinson's forward search method as opposed to the mean vector and covariance matrix of a random subset of p+1 points originally suggested by Atkinson. The non-outlying points identified by Atkinson's method are used to compute the starting mean vector and covariance matrix estimates for a modified, high-breakdown point M-estimation method proposed by Rocke (1996). The rationale for obtaining the final estimates in this manner is that the forward search method achieves better results given a good starting point, while M-estimation is also more likely to find the globally optimal solution if the initial estimate is close to this solution.

An additional feature of the Phase I process is a partitioning scheme designed to counter the fact that MCD computations grow exponentially with the sample size. Rather than attempt to apply the compound MCD, forward selection, and M-estimation method to the entire data set, the original data is randomly partitioned into a user-specified number of subsets. Robust estimates are then obtained for each subset and the covariance estimate with minimum determinant is used for Phase II. By partitioning the problem in

this manner, computations will grow linearly with the sample size, and also allow the methodology to be implemented using parallel processing.

Phase II of the compound estimation method involves computing the Mahalanobis squared distances for all the observations using the robust estimates from Phase I, scaling these squared distances so that they are consistent with distances obtained from multivariate normal data, and comparing the scaled distances to a suitable threshold from a Chi-squared distribution with p-degrees of freedom. The scaling method proposed by Rocke and Woodruff is somewhat unique and demands further explanation since it addresses a problem common to robust estimation methods. To begin, the shape estimate produced by methods such as the MCD or M-estimation gives an unbiased estimate whose expectation, according to Grubel and Rocke (1990), is some multiple of the true covariance matrix for elliptically symmetric distributions. Thus, Mahalanobis distances derived from the these shape estimates are some multiple of the true distances. If these distances are not scaled to be consistent with the underlying distribution of the data, unacceptable Type I or Type II errors may result. To account for this problem, Woodruff and Rocke suggest standardizing the squared distances to the h/n quantile of a Chi-squared distribution with p degrees of freedom, where $h=[(n+p+1)/2]$.

Because the Mahalanobis distances are only asymptotically Chi-square distributed, Woodruff and Rocke suggest further modification to the distances beyond the standardization just mentioned. Specifically, they suggest conducting a simulation study to determine the $1-\alpha_1$ quantile of distances obtained from normal samples of size n in dimension p and computed with mean vectors and covariance matrices from the Phase I procedure. Any observations whose Mahalanobis distance falls under this quantile is

then used to compute a new mean vector and covariance estimate, thus producing a covariance estimate for the 1-α_1 fraction of the original sample. According to Rocke and Woodruff, under the assumption of multivariate normality, this covariance estimate is a multiple of the true covariance matrix with the multiple given to be:

$$k\left(p,\alpha_1\right)=\frac{F_{\chi^2_{p+2}}\left(\chi^2_{p;1-\alpha_1}\right)}{1-\alpha_1} \tag{5.9}$$

The covariance matrix of the 1-α_1 fraction is multiplied by $k(p, \alpha_1)$ and the resulting covariance matrix used to recomputed Mahalanobis squared distances for all the observations. These final squared distances can then be compared to the desired quantile of a Chi-squared distribution with p degree of freedom in order to detect outliers.

Smallest Half-Volume Method

Rocke and Woodruff's compound estimation method represents a combination of two somewhat theoretical approaches to detecting outliers. The main drawback to MCD and M-estimation strategy for robust distance detection is their large computational burden that limits their utility relative to large-scale problems. As a less-formal, intuitive alternative for outlier detection on large datasets, Egan and Morgan (1998) propose the smallest half-volume (SHV) method. The basic premise behind the SHV method is that good observations in a dataset will tend to cluster closely together in Euclidean space. To identify a cluster of good data, the method begins by mean-centering and standardizing each column of the data matrix using the respective column mean and standard deviation. This process is referred to as auto-scaling. Using the auto-scaled data, an nxn distance matrix is formed in which element d_{ij} is the Euclidean distance from observation i to observation j. Thus, each column of the distance matrix records how close observation j

is to all other observations. With this idea in mind, each column of the distance matrix is sorted in ascending order. For each sorted column, the sum of the first $n/2$ distances is computed. The column with the smallest sum is identified, and the $n/2$ observations used in computing this column's sum are labeled as good data. The good data are then used to form a robust mean vector and covariance matrix, and to re-perform the auto-scaling procedure. To detect outliers, the mean vector and covariance estimates are used as robust inputs to the classic Mahalanobis distance detector.

A primary advantage of the SHV method is it does not require any matrix inversions, thus reducing the computational complexity relative to other robust distance methods. The SHV also obtains its solution in one pass of the method, as opposed to searching from many starting points in the manner of the MVE resampling method. The weaknesses of the SHV method are concerned primarily with evaluating the final Mahalanobis distances for indications of outliers. Because the final covariance matrix is estimated from the $n/2$ observations closest to the centroid, it will likely be a multiple of the true covariance estimate, as suggested by Rocke and Woodruff (1996). Therefore, the Mahalanobis distances should be scaled before conducting an formal significance tests. Unfortunately, Egan and Morgan provide no guidance in this area nor do they suggest an alternative method for analyzing the distances, though their simulation tests indicate the SHV method is effective at uncovering outlying observations.

Resampling by Half-Means Method

In the same article in which the SHV method is proposed, Egan and Morgan (1998) also develop the Resampling by Half-Means (RHM) method for detecting outliers. This method makes use of the auto-scaling concept to create samples of robust distances

for the observations. The RHM method begins by randomly selecting $n/2$ observations from the dataset without replacement. Each of the selected observations is used to form a row of the matrix $\mathbf{X}(i)$, where i denotes the iteration of the method. The mean and standard deviation are computed for each column of $\mathbf{X}(i)$. These estimates are then used to auto-scale the original data matrix. The magnitude of each row of the auto-scaled matrix is computed, which is equivalent to computing the distance of each auto-scaled observation to the centroid of the data. The distances for the n observations are saved in the vector $\mathbf{l}(i)$ which, in turn, constitutes the ith column of a matrix $\mathbf{L}$. This process is repeated for iteration $i+1$ until the desired number of iterations is achieved.

After the last iteration is complete, each column of L is sorted in ascending order. For each of the sorted columns, the observations corresponding to the largest 5% of the distances are identified. Outliers are identified as those observations whose distances appear in the upper 5% of distances an unusually large number of times. Unfortunately, no guidance is provided as to what how many appearances is indicative of an outlier; thus, this method ultimately relies on subjective judgment by the analyst to label observations as outlying.

Bivariate Boxplot Method

An informal method for detecting outliers in univariate data is to construct a boxplot that visually depicts the location, spread, and skewness of the data. Zani, Riani, and Corbellini (1998) develop a method for building a bivariate boxplot and suggest how it may be used to mind multivariate outliers. To build the bivariate boxplot for a pair of variables, the inner region for the plot—analogous to the univariate boxplot's interquartile region—is determined through the use of convex hull peeling originally

proposed by Bebbington (1978). Convex hull peeling entails identifying the observations on the convex hull of the bivariate data cloud, trimming these observations from the dataset, and repeating the process until only a desired percentage of the original observations remain. For the purpose of the bivariate boxplot, Zani et al. suggest trimming the data until 50% of the observations remain. These observations define the inner region for the boxplot. To ensure a smooth ellipse that visually depicts this inner region, Zani et al. use the method of B-splines (Ammeraal, 1992) to fit a curve to the convex hull of the inner region. The centroid for the boxplot is computed as the arithmetic mean of the observations contained in the inner region.

For a univariate boxplot, the outer region for the plot is defined by the interval $[x_{0.25}-1.5IR, x_{0.75}+1.5IR]$, where $x_{0.25}$ and $x_{0.75}$ are the 0.25 and 0.75 quantiles of the data, respectively, and $IR=x_{0.75}-x_{0.25}$. For normally distributed data, such a region is expected to include 99.3% of the observations. To construct an analogous region for the bivariate boxplot, Zani et al. suggest multiplying the inner region ellipsoid by the factor l=1.58, which, for normally distributed data, omits approximately 1% of the data.

To detect multivariate oultiers, Zani et al. recommend constructing a bivariate boxplot for every pair of variables. Any observation that is outside the 90% convex hull in any of the plots is removed from the data set. The remaining observations are then used as the starting point for the forward search method of Hadi (Hadi, 1992, Hadi, 1994) or Atkinson (1993). The authors claim that using bivariate boxplots in this manner make the forward search more computationally efficient, presumably because the initial basic subset for the search should contain considerably more than p+1 points.

Partitioning Method

Upon experimental testing of their compound estimator method, Rocke and Woodruff (1996) concluded that the method has difficulty detecting outliers in highly-contaminated datasets in which the outlier fraction was above 35%. In addition to this limitation, Kosinski (1999) also gives an example dataset in which the MCD-derived half-sample is not necessarily outlier-free. Claiming these weaknesses as a reason for caution when using MCD-based detection methods, Kosinski (1999) proposes an alternative detection method that searches for the partition of data that separates good observations from bad observations. Kosinski's partitioning method is essentially a repeated application of the forward search method of Hadi or Atkinson against multiple random starting subsets of size p+1. The number of starting subsets is selected to ensure a minimum probability that at least one of the subsets will be free of outliers; derivation of this number is provided by the author.

Once the forward search is applied to all the starting subsets, it is hoped that at least one α-partition of the data has been obtained, where an α-partition is defined by the following four characteristics:

1) The "good" part of the partition contains the majority of the data, where the majority is defined to be the quantity h=[(n+p+1)/2].
2) All the Mahalanobis squared distances for the bad observations are significant at the specified α-level of a Chi-Squared distribution with p degrees of freedom.
3) All the Mahalanobis distances for the bad observations are larger than those for the good observations.

4) The Mahalanobis squared distances for the good observations are not significant at the α-level of a Chi-Squared distribution with p degrees of freedom.

If only one α-partition results from the forward search process, then outlying observations are those contained in the bad partition. If multiple α-partitions result, Kosinski offers a procedure to ultimately arrive at only one partition by iteratively identifying and removing the strongest outliers.

Simulation tests conducted by Kosinski indicate that the partitioning method is less susceptible to masking and swamping effects than Rocke and Woodruff's compound estimation method, particularly when the outlier fraction is above 25%. These tests were run at different proximities of the outliers to the good data, and for p=2 and p=5. Whether or not these results are scalable to larger problems, or if the partitioning method is computationally feasible for larger problems is not clear.

FAST-MCD Method

When the MVE and MCD estimation methods were originally proposed by Rousseeuw (1983), the MVE received initial attention for outlier detection because it was computationally less expensive to find an approximate MVE solution. However, Butler, Davies, and Jhun (1993) showed that the MCD has better statistical efficiency than the MVE since the MCD is asymptotically normal. Additionally, Davies (1992) showed that the MVE has a lower convergence rate than the MCD. According to Rousseeuw and van Driessen (1999), these theoretical findings, combined with the need for accurate estimators for use in outlier detection schemes, the MCD began to gain favor over the MVE as the preferred robust estimator for outlier detection. The main drawback to using

the MCD, however, was the high computational complexity involved with searching the space of half-samples of a dataset to find the one whose covariance matrix had the minimum determinant. To address this problem, Rousseeuw and van Driessen proposed the FAST-MCD outlier detection method that uses a key theoretical finding in conjunction with a partitioning method to rapidly search for an approximate MCD solution. The primary theorem proved by Rousseeuw and van Driessen states that if one starts with a half-sample of data, orders the entire data set based on Mahalanobis distances derived from the half-sample's mean vector and covariance matrix, and selects a new half-sample from the observations with smallest distances, the covariance determinant of the new half-sample will be less than or equal to the old half-sample covariance determinant with equality occurring only when the mean vector and covariance matrices for the old and new half-samples are equal. By repeatedly applying this theorem to a dataset—a process referred to as a C-step—it is possible to converge to at least a local optimal MCD solution. A further finding based on experimental results indicates that if the starting half-sample is capable of converging to a good solution, the covariance determinant will begin to rapidly converge after only two C-steps.

To effectively use this tendency of rapid convergence, Rousseeuw and van Driessen develop a partitioning, or nesting, scheme in which the original data set is randomly partitioned into smaller subsets. A small number of C-steps are then performed on each subset. The covariance matrices are then taken from the subsets whose C-steps gave the smallest covariance determinants, and these covariance matrices are then used as starting points for C-steps on the entire data set. For each covariance matrix, C-steps are performed until convergence of the covariance determinant. The covariance matrix

corresponding to the smallest determinant is used for the final MCD estimate. For very large data sets, this nesting scheme is altered to first operate on random samples from the original dataset. As the method progresses, the entire data set is gradually introduced into the C-step scheme.

To obtain consistency of the MCD covariance estimate when the data is multivariate normal, the final covariance matrix is multiplied by the following scaling factor:

$$\frac{\underset{i}{med}\, d^2_{(\mathbf{T},\mathbf{S})}(i)}{\chi^2_{p,0.5}} \tag{5.10}$$

where

$d^2_{(\mathbf{T},\mathbf{S})}(i) =$ the Mahalanobis squared distance of the *i*th observation relative to the MCD mean, **T**, and covariance matrix, **S**, and

$\chi^2_{p,0.5} =$ the 0.5 quantile of a Chi-squared distribution with p degrees of freedom.

After scaling the covariance matrix using (5.10), the authors also recommend computing a one-step reweighted estimate in the same manner as Rousseeuw and van Zomeren's method discussed earlier. The final mean vector and covariance matrix obtained from the one-step reweighted estimate are then use in the classical Mahalanobis distance outlier detection scheme to identify outliers. Testing performed by Rousseeuw and van Driessen indicate the FAST-MCD method is capable of handling problems with 50000 observations in 30 dimensions. This algorithm was implemented in S-Plus 4.5 and SAS/IML 7 as a robust estimation option.

The BACON Method

The desire to find an outlier detection method that is applicable to very large datasets is echoed by Billor, Hadi, and Velleman (2000). However, where the FAST-MCD method attempts to use nesting and C-steps to search for an optimal solution, Billor et al. make two observations concerning robust distance computation as a guide to developing the Blocked Adaptive Computationally Efficient Outlier Nominator (BACON). The first observation is that the added computational complexity of trying to find optimal robust estimators may not be justified by significantly better outlier detection. The second observation is that insisting upon a completely affine equivariant method may add substantial computational complexity to an algorithm without a proportional improvement in the detection of outliers. Using these two observations, Billor et al. develop BACON as a method that "abandons" optimality conditions in favor of a very fast outlier detection strategy that can be run in a non-robust, affine equivariant mode with breakdown point of 20%, or in a robust, near-affine equivariant mode with a breakdown point of 40%.

The BACON method is derived from the forward search method of Hadi (1992) and Hadi (1994), and begins its search for outliers in much the same manner by selecting an initial basic subset of good observations. The manner in which the initial basic subset is chosen depends on whether the user wishes to have a lower breakdown point method that is affine equivariant, or a high-breakdown point method that is not completely affine equivariant. In the former case, the initial basic subset contains the p+1 observations with the smallest Mahalanobis distances relative to the mean vector and covariance matrix for the entire dataset. In the latter case, the basic subset is formed from the p+1

observations with smallest distances relative to the component-wise median of the observations and the covariance matrix derived from this median vector. Using the component-wise median makes the BACON method more robust to outliers at the expense of affine equivariance since the median estimator is not affine equivariant.

Once the initial basic subset is selected, its mean vector and covariance matrix are estimated and used to compute Mahalanobis distances for all observations. Once these distances are obtained, they can be compared to the square root of an appropriate quantile from the Chi-Squared distribution with p degrees of freedom. However, since the covariance matrix used to compute the distances is estimated from a small sample of points, the threshold value must first be multiplied by a correction factor given by:

$$c_{npr} = c_{np} + c_{hr} \tag{5.11}$$

where

$$\begin{aligned} c_{hr} &= \max\left\{0, (h-r)/(h+r)\right\}, \\ h &= \left[(n+p+1)/2\right], \\ r &= \text{the size of the current basic subset, and} \\ c_{np} &= 1 + \frac{p+1}{n-p} + \frac{1}{n-h-p}. \end{aligned}$$

The c_{np} term in (5.11) is the same as the correction factor introduced by Hadi (1994). This factor was apparently derived using simulation studies, though no further details are available on how this study was conducted.

Upon comparing the Mahalanobis distances to the corrected threshold value, any observations that fall below the threshold are added to the basic subset. A new mean vector and covariance matrix are computed for the basic subset, new Mahalanobis distances are obtained, and any observations falling under the threshold value are added

to the basic subset. This process continues until the basic subset does not change between iterations. Once the iterations stop, any observations not contained in the basic subset are considered outliers. This iteration process is significantly different than Hadi's forward search method in that multiple observations can be added to the basic subset at each iteration whereas the original method only allowed the basic subset to grow by one observation per iteration. This modification employed by BACON makes the algorithm much faster than the original method by sacrificing any attempt to find an approximate MVE solution which required the basic subset to grow at a much slower rate.

Experimental tests with the BACON method show that it is less computationally expensive than Hadi's forward search and that the number of iterations required by the method remains relatively constant as the sample size increases. Test also indicate that BACON has a good null-behavior and that the breakdown points for the non-robust start and robust start are approximately 20% and 40%, respectively.

The MCD-EHD Method

A common method for detecting multiple outliers in univariate data is to identify the most outlying observation, delete it from the data, find the second-most outlying observation, delete it, and so on, until no other observations can be considered outlying. These methods are referred to iterative deletion method. A multivariate, robust distance version of iterative deletion is given by Viljoen and Venter (2002) who refer to the method as Minimum Covariance Determinant-Extreme Hotelling Deviate (MCD-EHD). The MCD-EHD method is based on Wilks' (1963) test for a single multivariate outlier, and Caroni and Prescott's (1992) sequential application of Wilks' test to find multiple

outliers. These two precursor methods will be summarized first before discussing the MCD-EHD method further.

Wilks (1963) proposed a relatively straight-forward process for detecting a single outlier in a multivariate data set assuming a mean-shifted normal model for the data and the outlier. The mean-shifted model simply means that the main population is normally distributed and that the population of the outlier is also normally distributed with the same covariance as the good data but with mean vector shifted by some constant. Wilks' method begins by computing the following ratio for the *j*th observation in the dataset, $j=1,\ldots,n$:

$$W_j = \frac{\left|\mathbf{A}^{(j)}\right|}{\left|\mathbf{A}\right|} \tag{5.12}$$

where

$\mathbf{A}$ = the covariance matrix of the entire dataset, and
$\mathbf{A}^{(j)}$ = the covariance matrix of the dataset with observation j removed.

With these W_j computed, the smallest of these values, D_1, is tested against a critical value from the distribution of the W_j (Wilks provides tables of critical values for the W_j distribution obtained from simulation studies.) If D_1 exceeds the critical value, the corresponding observation is labeled an outlier. In essence, Wilks' method is attempting to find the observation that has the largest impact on the size of the covariance matrix—as determined by the covariance determinant—and then determine if this observation is indeed outlying.

To extend Wilks' method to find multiple outliers, Caroni and Prescott (1992) suggest iterative applications of the original procedure. This iterative approach proceeds

by computing the non-robust Mahalanobis distances for each of the observations in the data set and identifying the observation with the largest distance. This observation is deleted from the dataset, new distances are computed, and the observation with the new largest distance is removed. The process continues until the number of observations removed reaches a user-specified upper bound on the number of outliers present in the data. At each iteration of the method, the maximum Mahalanobis distance is recorded. When the deletion phase of method is complete, the recorded distances form a set $\{T_r\}$, $r=l+1,\ldots,n$, where r is the number of observations used in a given iteration, and l is the user-specified lower bound on the number of good observations in the original dataset.

To determine the number of outliers, the T_r values are tested against a corresponding sequence of critical values $\{c_r\}$, beginning with T_{l+1}. For the first T_r that exceeds its critical value, the number of good observations is determined to be r-1 and the number of outliers is determined to be n-r. The outlying observations correspond to those removed during the first n-r iterations of the deletion process.

Viljoen and Venter demonstrate that the primary limitation of Caroni and Prescott's method is its susceptibility to masking and swamping due to the non-robust Mahalanobis distances used in the deletion phase of the method. Viljoen and Venter offer the MCD-EHD as a solution to this problem. Rather than use the non-robust Mahalonobis distances in the deletion phase of the Caroni-Prescott method, the authors recommend using robust distances computed from FAST-MCD estimates of the mean vector and covariance matrix. Since this strategy can be computationally expensive due to computing FAST-MCD estimates at each iteration, the Viljoen and Venter offer the alternative of computing the FAST-MCD estimate only once assuming that the dataset

has a lower bound of l good observations. Using this simplification, the T_r values become the n-l largest orders statistics of the Mahanalobis distances computed using the single FAST-MCD estimate.

In addition to using robust distances, the MCD-EHD method further modifies the Caroni-Prescott method by using a different derivation for the sequence of critical values $\{c_r\}$. For further details on these critical values, refer the original article by Viljoen and Venter (2002).

Closest Distance to Center Method

With the exception of the FAST-MCD and BACON algorithms, the robust distance outlier detection methods discussed so far are somewhat theoretical in nature and may not necessarily scale to large problems of practical interest. To address such problems, several methods can be found in the technical literature that are somewhat less formal in nature, but can handle large data sets in relatively high dimension. One such robust distance-based method is the Closest Distance to Center (CDC) method proposed by Chiang, Pell, and Seasholtz (2003). The CDC method proceeds by scaling the data so that each of the p attributes has zero mean and unit variance using the auto-scaling procedure discussed earlier. A mean vector is then computed for the scaled data. For each auto-scaled observation, the distance is computed from the observation to the mean vector. This distance is computed in one of two ways: the Euclidean distance can be used (CDC_2); or, the maximum norm distance can be computed as the maximum component-wise distance from the mean vector (CDC_m). Regardless if the CDC_2 or the CDC_m method is used, the next step of the method is to identify the $n/2$ observations with the smallest distances. The mean vector and covariance matrix are then computed for these

observations and used as a starting point for the Ellipsoidal Multivariate Trimming (MVT) method of Walczak and Massart (1995). The MVT method continues by computing the Mahalanobis distances for the entire dataset using the CDC_2 or CDC_m mean vector and covariance matrix. The $n/2$ observations with the smallest Mahalanobis distances are identified and used to estimate a new mean vector and covariance matrix. This process repeats until the covariance estimate stabilizes. The final mean vector and covariance matrix can be used to compute robust Mahalanobis distances. The authors seem to recommend graphical analysis of the Mahalanobis distances to identify outliers.

A limitation of the CDC method is the use of auto-scaling in the first step of the method. For a data matrix arranged with each observation as a row in the matrix, auto-scaling the matrix entails subtracting the respective column mean, m_j, from each element of the matrix, and then dividing the differences by the respective column standard deviation, s_j. Because the component means and standard deviations are not robust to outliers, the auto-scaling method may confuse the outlier search. As alternatives to auto-scaling, Chiang et al. offer several alternatives. The *robust* alternative was originally suggested by Huber (1989) and requires m_j be replaced by the column-wise median in the auto-scaling method, and that s_j be replaced s_{MAD} given by:

$$s_{MAD} = 1.482 \underset{i}{median} \left\{ \left| x_i - x_{median} \right| \right\} \tag{5.13}$$

where

x_i = the ith element of column j, and
x_{median} = the median of all elements in column j.

The *modified* version of auto-scaling proposed by Chiang et al. involves replacing the column mean and standard deviation by the mean and standard deviation of the $n/2$

column elements closest to the column median. A final alternative based on the S_n and Q_n estimators of Rousseeuw and Croux (1993) is suggested for data that may be have an asymmetrical distribution.

Experimental tests of the CDC method and the different scaling alternatives indicates that using CDC or the scaling alternatives improves outlier detection, while employing the scaling alternative as a preprocessing step for CDC provides the best detection performance of the detectors under evaluation.

Robust Clustering Detector

A common element of all the robust-distance methods discussed so far, as well as most multivariate outlier detection methods, in general, is the assumption that the majority of observations come from a single, good population. In many applications, however, a dataset may be comprised of observations from several good populations. For hyperspectral imagery, this latter condition is usually the case with the scene's background materials—grass, road, trees, water, dirt, etc.—each forming a good population sample in which we want to find outliers. Applying single-population outlier detection methods to these types of multi-population datasets will undoubtedly produce misleading results.

To address this complication, Hardin and Rocke (2004) propose an outlier detection method that performs a robust cluster analysis on the dataset to divide the observations into similar clusters, and then applies an MCD-based robust distance outlier detector to each cluster to find outlying observations. More specifically, Hardin and Rocke's method begins with the user specifying the number of clusters, k, to use in the analysis. A robust clustering method developed by Woodruff and Reiners (2004) is then

used to perform the cluster analysis in order to avoid the negative effects outliers can have on classical clustering algorithms. The clustering method formulates the cluster analysis problem as a zero-one non-linear integer program that attempts to optimally assign each observation to one of the k clusters in order to minimize the sum of the cluster covariance determinant matrices. To account for the presence of outliers, the integer program formulation includes an additional cluster that is used to contain any observations that do not fit well in the any of the k clusters. An approximate integer program solution is found heuristically using simulated annealing.

Once the cluster analysis is complete, a mean vector and covariance matrix are computed for each cluster, the Mahalanobis squared distances are computed from every observation to each cluster center, and each observation is assigned to the cluster to which it is closest. For the first cluster, $h_1=[(n_1+p+1)/2]$ of the observations with the smallest Mahalanobis distances are selected, where n_1 is the total number of observations assigned to cluster one. This *half-sample* of observations is used to estimate a new mean vector, covariance matrix, and Mahalanobis distances for the cluster. These operations are repeated on the cluster one data until the covariance matrix stabilizes. The final half-sample of observations is used to form the cluster's MCD estimate. This process is then repeated for the remaining clusters.

When the MCD half-samples, mean vectors, and covariance matrices have been determined for each cluster, the Mahalanobis squared distances are again computed from each observation to each cluster center. To determine if any of the distances indicate outliers, Hardin and Rocke suggest a significance test based on the distribution of Mahalanobis squared distances when the distance are computed from MCD mean vector

and covariance estimates. This distance distribution was originally derived by Hardin and Rocke (2005) and is given as:

$$\frac{c_j\left(m_j - p + 1\right)}{pm_j} d^2_{S_j^*}\left(X_i, \bar{X}_j^*\right) \sim F_{p,m_j-p+1} \tag{5.14}$$

where

$d^2_{S_j^*}\left(X_i, \bar{X}_j^*\right) =$ the Mahalanobis squared distance of observation X_i relative to cluster j's MCD mean vector, $\bar{X}_j^*$, and MCD covariance matrix, S_j^*,

$$c_j = \frac{P\left(\chi^2_{p+2} < \chi^2_{p,h_j/n_j}\right)}{h_j / n_j},$$

$h_j =$ the half-sample size of cluster j,

$n_j =$ the number of observations in cluster j, and

$m_j =$ either a small-sample or asymptotic correction factor.

The correction factor, m_j, was originally derived by Croux and Haesbroeck (1999), and has both an asymptotic and small-sample form. The small-sample approximation for m_j is given by:

$$m_{pred} = m_{asy} e^{\left(0.725 - 0.00663p - 0.078\ln(n)\right)} \tag{5.15}$$

where

$m_{asy} =$ the asymptotic value of the factor.

The computations required for m_{asy} can be found in Hardin and Rocke (2004). Tests performed by Hardin and Rocke indicate that using this distributional fit for the

Mahalanobis squared distances as opposed to a Chi-Squared distribution results in significantly smaller Type I and Type II errors.

Non-Traditional Methods

A common limitation with all robust distance-based outlier detection methods is the requirement to find a subset of outlier-free data from which robust estimates of the mean vector and covariance matrix can be obtained. Unfortunately, there is no existing method that can find an outlier-free subset with 100% certainty. In other words, there is always a chance that the "outlier-free" sample contains some outliers. Should this condition exist, the ability of the respective detection method to find outliers may be impaired. Though, empirical tests indicate that all of the robust-distance methods discussed in the previous section seem to have some level of resistance to this problem, researchers have proposed alternative methods that attempt to avoid robust Mahalanobis distances altogether. Because these methods comprise a small minority of existing multivariate outlier detectors, we shall refer to them as *non-traditional* methods. In the following paragraphs, the significant non-traditional outlier detection methods found in the technical literature are outlined. As in the previous section, these methods are discussed in chronological order to illustrate how these methods have evolved over time.

Principal Component Methods

One of the earliest distance-free methods for detecting multiple outliers in multivariate data is described by Gnanadesikan and Kettenring (1972) and is originally attributed to Rao (1964). This method makes the assumption that the dataset falls in the linear subspace defined by the first p-q principal components of the sample covariance matrix. Under this assumption, it is argued that outliers will have a large deviation from

this sub-space as measured by the sum of the magnitudes of their projections onto the last q eigenvectors. More specifically, outliers in a pxn dataset, $\mathbf{Y}$, are observations, $\mathbf{y}_j$, with large values of:

$$d_j^2 = \sum_{i=p-q+1}^{p} \left[\mathbf{l}_i^T \left(\mathbf{y}_j - \overline{\mathbf{y}} \right) \right] \tag{5.16}$$

where

$\mathbf{l}_i$ = the eigenvector corresponding to the ith smallest eigenvalue of the covariance matrix, and

$\overline{\mathbf{y}}$ = the mean vector of $\mathbf{Y}$.

Gnanadesikan and Kettenring suggest analyzing the d_j^2 values through the use of a gamma probability plot where the shape parameter is estimated using a method proposed by Wilk and Gnanadesikan (1964).

In addition to Rao's method, Gnanadesikan and Kettenring suggest other informal uses of the principal component scores for detecting outliers:

i) Construct scatter plots of bivariate and trivariate sets of component scores corresponding to the smallest eigenvalues to detect unusual groupings of the data.

ii) Construct normal probability plots for each of the last few sets of principal component scores. Because the principal component transformation is linear, the resulting scores may be more normally distributed. Hence, outliers may be easier to detect in these probability plots.

iii) Construct plots of the component scores associated with each of smallest eigenvalues against distances computed using the scores for

the largest eigenvalues. An example of this procedure would entail plotting, for each observation, the last component score against the magnitude of the vector formed by the first three component scores. Such a plot may reveal abnormalities in the data in the same manner that residual abnormalities are identified using residual plots of linear regression data.

iv) To determine which observation, $\mathbf{y}_j$, has the biggest impact on the orientation and scale of the sample covariance matrix, compute the following metric for each observation:

$$t_j^2 = \sum_i \lambda_i \left[\mathbf{l}_i^T \left(\mathbf{y}_j - \overline{\mathbf{y}} \right) \right]^2 = \left(\mathbf{y}_j - \overline{\mathbf{y}} \right)^T \mathbf{S} \left(\mathbf{y}_j - \overline{\mathbf{y}} \right) \tag{5.17}$$

where

λ_i = the ith largest eigenvalue of $\mathbf{S}$,
$\mathbf{S}$ = the sample covariance matrix, and
$\mathbf{l}_i$ = the eigenvector of $\mathbf{S}$ corresponding to λ_i.

Observations with large values of t_j^2 can be considered outliers.

Unfortunately, a limitation of these methods is they are devoid of any formal tests of significance, relying upon the analyst to subjectively determine how an outlier should manifest itself.

<u>Mahalanobis Distance Decomposition Method</u>

As an alternative to computing robust Mahalanobis distances to detect outliers, Kim (2000) derives two decompositions of the Mahalanobis distance and uses scatter plots of the component terms to uncover outlying observations. Thus, rather than use the Mahalanobis distances themselves to find outliers, Kim suggests analyzing the

constituent parts of the Mahalanobis distances for an observation to determine how the distance was achieved. Kim provides no guidance on the distribution of the components of the Mahalanobis distance, thus requiring subjective analysis of the suggested scatter plots to identify outliers.

Projection Pursuit Detection

In order to avoid the masking and swamping effects associated with the classical Mahalanobis distance detector as well as the computational complexities of robust-distance detectors, Pan, Fung, and Fang (2000) propose a detector that uses univariate projections of the original data and univariate outlier detection to identify multivariate outliers. The method begins by projecting the original data onto a vector located on the *p*-dimensional unit hypersphere. For each projected observation, the following metric is computed:

$$J_n\left(\mathbf{x},\mathbf{a}\right)=\sqrt{n}\left(V\left(\mathbf{x},\mathbf{a},F_n\right)-V\left(\mathbf{x},\mathbf{a},F\right)\right) \tag{5.18}$$

where

$\mathbf{a} =$ the projection vector on the unit hypersphere,
$F_n =$ the empirical distribution function of the data,
$F =$ the distribution function of the data, and
$V\left(\square\right) =$ the Hampel univariate outlier identifier.

The Hampel identifier was described in Hampel, Ronchetti, Rousseeuw, and Stahei (1986), and is defined as:

$$V(\mathbf{x},\mathbf{a},F)=\frac{\left(\mathbf{a}^T\mathbf{x}-med\left(F^{\mathbf{a}}\right)\right)}{\underset{\mathbf{x}}{med}\left(\left|\mathbf{a}^T\mathbf{x}-med\left(F^{\mathbf{a}}\right)\right|\right)} \tag{5.19}$$

where

$$F^{\mathbf{a}} = \text{is the distribution of } \mathbf{a}^T\mathbf{x}, \text{ and}$$
$$F = \text{the distribution of } \mathbf{x}.$$

The process of projecting the data onto different vectors is repeated s times to form a sequence $J_s(\mathbf{x})=(J(\mathbf{x},\mathbf{a}_1), J(\mathbf{x},\mathbf{a}_2),\ldots, J(\mathbf{x},\mathbf{a}_s))$ for each observation. As shown by Cui and Ting (1994), the sequence ($J(\mathbf{x},\mathbf{a})$: $\mathbf{a}$ is on the unit hypersphere) is a stochastic process that converges weakly to a Guassian process with continuous sample path. The sample paths have zero mean and a covariance function $R_{\mathbf{x}}(\mathbf{a},\mathbf{b})$ that is derived by Pan et al., where $\mathbf{a}$ and $\mathbf{b}$ are two vectors on the unit hypersphere. That is to say,

$$\mathbf{J}_s(\mathbf{x}) = \left(J(\mathbf{x},\mathbf{a}_1), J(\mathbf{x},\mathbf{a}_2),\ldots, J(\mathbf{x},\mathbf{a}_s)\right)^T \sim N\left(\mathbf{0},\mathbf{R}_s(\mathbf{x})\right), \tag{5.20}$$

where $\mathbf{R}_s(\mathbf{x})$ is generated using the covariance function $R_{\mathbf{x}}(\mathbf{a},\mathbf{b})$.

After $\mathbf{J}_s(\mathbf{x})$ is generated for each observation in the original dataset, the Mahalanobis squared distances can be computed in the s-dimensional space for each observation using the observation's respective covariance matrix, $\mathbf{R}_s(\mathbf{x})$. These squared distances can be compared to a quantile of the Chi-Square distribution with s degrees of freedom to identify outliers, as proved by Pan et al.

To implement this projection pursuit method, several obstacles need to be overcome: 1) the projection vectors need to be determined; 2) the distribution of the projected observations must be known in order to compute the Hampel identifier values; and 3), a sample estimate for $\mathbf{R}_s(\mathbf{x})$ must be computed. To address the first problem, Pan et al. recommend using the quasi-Monte Carlo method TFWW—originally proposed by Tang (1977) and Fang, Wang, and Wong (1992)—to generate a set of vectors uniformly scattered over the p-dimensional unit hypersphere. To overcome the second obstacle, Pan et al. recommend using bootstrap methods to estimate the distribution functions

required by the Hampel identifier. Finally, to estimate the covariance matrix for an observation, Pan et al. provide estimation formulas as well as a regularization method for use when *s*, the number of projection vectors, is very large.

Based on tests with relatively small datasets, Pan et al. demonstrate that their method is effective at detecting outliers while achieving relatively low false alarm rates. No evidence is provided to suggest this method is scaleable to larger problems. In fact, using this method for high-dimensional datasets can be problematic since the number of projection vectors generated by the TFWW method to achieve uniform coverage of the *p*-dimensional unit hypersphere grows non-linearly with *p*. Further discussion of this problem is provided by Fang and Wang (1994).

Juan-Prieto Method

Empirical tests conducted by Juan and Prieto (2001) indicate that the robust distance methods of Rocke and Woodruff (1996), Hawkins (1994), Rousseeuw and van Driessen (1999), and Maronna and Yohai (1995), have difficulty detecting clusters of concentrated outliers, particularly when the clusters are relatively close to the good data. To overcome this perceived weakness of robust distance methods, Juan and Prieto suggest a distance free method based on angles. Specifically, the authors state that the projections of observations on the *p*-dimensional unit hypersphere are uniformly distributed when the observations have an ellipsoidal distribution, as shown by Eaton (1983). Based on this characteristic, Juan and Prieto claim that the angles between the projected observation vectors and an arbitrary reference direction, $\mathbf{u}_0$, have a Beta distribution. The form of the Beta distribution is provided by the authors.

To detect outliers, the original observations are projected onto the unit hypersphere. A reference direction, $\mathbf{u}_0$, is then selected using a method suggested by Juan and Prieto. The angles between the projected observations and $\mathbf{u}_0$ are then computed. The authors then suggest using a quantile-quantile plot of the angles to determine if they follow the beta distribution. Alternatively, the distributional fit of the angles can be assessed by analyzing the spacings between the ordered values of $F(\mathbf{w}_i)$, the theoretical distribution function of the angles evaluated at each angle, $\mathbf{w}_i$. If the angles actually follow the prescribed distribution, the spacings should be uniformly distributed. To test this hypothesis, Juan and Prieto suggest using the distribution of the largest spacing in a uniform sample introduced by David (1981). This distribution function is given by:

$$P\left(\bar{D}_{(n)} \le y\right) = \sum_{0 \le i \le 1/y} (-1)^i \binom{n+1}{i} (1-iy)^n \tag{5.21}$$

where

$\bar{D}_{(n)}$ = the largest spacing between n ordered, uniformly distributed points.

From this distribution function, a critical value for the largest spacing can be computed and all the largest spacing tested for significance. If the test fails, any corresponding observations preceding the largest spacing are considered outliers. To detect multiple outlier clusters, this entire process is repeated until the spacings indicate uniformity of the angles.

Though Juan and Prieto present empirical tests that indicate their angle detector is effective at finding outliers, there is one significant limitation with the method. Specifically, finding the critical value for the largest spacing test using (5.21) can be

computationally challenging and time-consuming when n is large. Complicating the matter further, Juan and Prieto empirically show that the critical value is also dependent on the dimensionality of the original data. To help alleviate these problems, the authors provide an approximation that can be used for higher dimensions based on the critical values for the univariate case. Even with this approximation, however, computation of the critical value can still be challenging in practice.

Chiang-Pell-Seasholtz PCA Method

Where Gnanadesikan and Kettenring (1972) proposed somewhat informal methods for using PCA to find multivariate outliers, Chiang, Pell, and Seasholtz (2003) give a PCA method that includes significance tests for outliers. The method begins by performing a PCA on the original data to arrive at the $p \times a$ matrix, **P**, containing the eigenvectors corresponding to the a largest eigenvalues. For each observation, $\mathbf{x}_i$, the T^2-statistic is computed as:

$$T^2 = \mathbf{x}_i^T \mathbf{P} \Sigma_a^{-2} \mathbf{P}^T \mathbf{x}_i \tag{5.22}$$

where

Σ_a = the matrix containing the first a rows
and columns of the original covariance matrix.

The threshold value for T^2 is given by MacGregor and Kourti (1995) to be:

$$T_\alpha^2 = \frac{a(n-1)(n+1)}{n(n-a)} F_\alpha(a, n-a) \tag{5.23}$$

where

$F_\alpha(n_1, n_2)$ = the (1-α)-quantile of the F-distribution
with n_1 and n_2 degrees of freedom.

In addition to testing if an observation is an outlier using the components for the *a* largest eigenvalues, Chiang et al also suggest testing the observation using the *p*-*a* components for the remaining eigenvalues. To perform this test, the authors recommend using the *Q*-statistic of Jackson and Mudholkar (1979) defined as:

$$Q = \mathbf{e}^T \mathbf{e} \tag{5.24}$$

where

$$\begin{aligned} \mathbf{e} &= \text{the residual vector resulting from fitting} \\ &\quad \text{the subspace of the first } a \text{ eigenvectors to} \\ &\quad \text{the observation } \mathbf{x} \\ &= \left(\mathbf{I} - \mathbf{P}\mathbf{P}^T\right)\mathbf{x}. \end{aligned}$$

The threshold value for the *Q*-statistic is provided by Chiang et al.

If either the T^2- or *Q*-statistics for an observation exceed their respective critical value, the observation is labeled an outlier and removed from the data set. Once all observations are tested, the entire process is repeated using only the non-outlying observations. The algorithm terminates when no additional observations are labeled outliers between iterations, or when the total number of outliers detected reaches *n*/2.

Because the initial PCA is performed on a potentially contaminated dataset, Chiang et al. caution that their detection method may fail since the starting covariance matrix may be significantly distorted by outliers. To guard against this condition, the authors propose using robust PCA algorithms such as those developed by Helge, Liang, and Kvalheim (1995), Li and Chen (1985), Croux and Ruiz-Gazen (1996), and Hubert, Rousseeuw, and Verboven (2002). No guidance is provided on how such an implementation would proceed.

Max-Eigen Difference (MED) Method

Adding to the arsenal of principal component-based outlier detection methods, Gao, Li, and Wang (2005) propose the Max-Eigen Difference (MED) method. The method proceeds by computing the eigenvalues and eigenvectors of the sample covariance matrix of the entire dataset. For each observation, $\mathbf{x}_i$, the eigenvalues and eigenvectors are then computed for the covariance matrix obtained when $\mathbf{x}_i$ is removed from the dataset. Using these eigenvalues and eigenvectors, the following values are computed for each observation:

$$d_i = \left\| \lambda_1^{(i)} v_1^{(i)} - \lambda_1 v_1 \right\| \left(1 - \prod_{j=1}^{p} I_{\left\{ y_{ij}^2 < \lambda_j \right\}} \right) \tag{5.25}$$

where

λ_j = the jth largest eigenvalue for the sample covariance matrix,

v_j = the eigenvector associated with λ_j,

$\lambda_j^{(i)}$ = the jth largest eigenvalue for the sample covariance matrix when the ith observation is removed from the sample,

$v_j^{(i)}$ = the eigenvector associated with $\lambda_j^{(i)}$,

$y_{ij} = \left(\mathbf{x}_i - \bar{\mathbf{X}} \right)^T v_j$, and

$I_{\{\cdot\}}$ = an indicator function.

After computing the d_i values, the MED statistic for each observation is computed as:

$$MED_i = \frac{d_i}{\sum_{j=1}^{n} d_j}. \tag{5.26}$$

Upon decomposing the MED statistic, Gao et al. are able to show that large MED values indicate outlier observations. Specifically, the decomposition illustrates that an

observation with a large MED may indicate: 1) the observation has a first principal component score that is much larger than the other observations; 2) the observation may have relatively large scores on the other component axes; and 3) the observation is not close to the centroid of the data. An observation with large MED may posses any combination of these characteristics.

Based on the properties of the MED, Gao et al. recommend detecting outliers by plotting the MED values against the observation indices. Any observations that appear to have a large MED relative to the other observations are labeled as outliers. This labeling is a subjective decision made by the analyst. The authors provide no formal significance test for the MED statistic. Empirical tests of the MED detector indicate that it is superior to the classical Mahalanobis distance detector, and provides similar results to an MVE-based robust distance detector. In one test, the MED detector identified outliers that were overlooked by the robust distance method.

Summary

The preceding paragraphs provide a survey of the significant multivariate outlier detection methods developed over the last three decades. As presented, these methods can be divided into two basic families: robust distance methods that attempt to apply robust estimation to the classical Mahalanobis distance detector, and non-traditional methods that by-pass robust estimation in favor of some other characteristic of outliers that can be exploited to reveal their presence. In the following sections, the foundation laid by this literature review is used to investigate further the relevance of outlier detection methods to hyperspectral anomaly detection and to develop procedures to apply multivariate outlier detectors to uncover anomalies.

Outlier Impact Experiments

As stated previously, Donoho and Huber (1983) show that the breakdown points for the classical mean and covariance estimate are only $1/N$. Thus, it is theoretically possible for a single, well-placed outlier to distort these estimates to the extent that Mahalanobis distances produced by these estimates are no longer useful in detecting outliers. Though this single-observation breakdown can occur in theory, it is natural to wonder if this phenomenon is of any practical concern to hyperspectral anomaly detection. Relatively limited results found in West et al. (2005) and Farrell and Mersereau (2005) demonstrate that using contaminated covariance matrix estimates can degrade target detector performance. To more comprehensively address this issue, we conduct several experiments to assess the magnitude of masking and swamping when the classical Mahalanobis distance outlier detector is applied to simulated data possessing similar mean vectors, covariance structure, dimensionality, and number of observations as actual hyperspectral data. These tests are discussed in the following sections.

Simulated Gaussian Data Experiments

This first experimental test performed measures the degree of masking and swamping that can occur in controlled multivariate Gaussian data as a function of the number of outliers contained in the data. Multivariate Gaussian data is chosen for these tests because it is relatively straight-forward to generate random variates from this distribution, as well as the fact that many hyperspectral analysis techniques make the Gaussian assumption. The experiment was executed as follows:

1) A sample of N_b= 2000 p-dimensional observations was generated from a multivariate Gaussian distribution with a specified mean vector and covariance matrix. Refer to these observations as the "background" data.
2) A specified number, N_o, of *outlier* observations were randomly selected with replacement from a set of observations with a different mean vector and covariance matrix from the background data. These outliers were added to the background data to form the *contaminated* dataset of size $N= N_b+ N_o$.
3) The mean vector and covariance matrix were estimated for the contaminated dataset and used to compute the Mahalanobis Squared Distances (MSDs) for all observations in the contaminated dataset.
4) The 0.95 quantile for a Chi-Square distribution with p-degrees of freedom was determined and used to threshold the MSDs from Step 3. Observations whose MSDs exceeded the threshold were considered *detected* outliers.
5) The number of true positives was computed and recorded to be the number of outlier observations classified as detected outliers. If the number of true positives is less than the number outliers introduced into the sample at Step 2, masking is occurring.
6) The number of false alarms was computed as the number of background observations classified as detected outliers. If the number of false alarms is greater than αN, where α=0.05, swamping is occurring.
7) Steps 1 through 6 were repeated 50 times using the same background mean vector and covariance matrix in Step 1 and set of outlier observations in Step 2. The mean number of true positives and false alarms was computed for

these 50 iterations, as well as the 95%-confidence intervals for these mean estimates.

8) Steps 1 through 7 were repeated using a different value for N_o. The values of N_o used for the experiment ranged from 0 to 500 in increments of ten.

9) Steps 1 through 8 were repeated using a different mean vector and covariance matrix in Step 1 and a different set of outlier observations in Step 2.

Before presenting the results of this experiment, several comments need to be made. First, the mean vectors, covariance matrices, and outlier observations used in Steps 1 and 2 were determined from sets of pixel vectors obtained from two actual hyperspectral images. The first hyperspectral image is a COMPASS sensor image of Fort A.P. Hill, Virginia. From this image, pixel vectors corresponding to grass, road, dead grass, trees, and shadow were manually selected. The mean vectors for each of these sets of pixel vectors are shown in Figure 12. The error bars in the chart denote one-standard deviation above and below the mean for each band. These five materials were used to form the different background-outlier combinations listed in Table 1. The second image used for this experiment is an AVIRIS image of the National Mall in Washington, D.C. From this image, pixel vectors corresponding to grass, asphalt, gravel, roofing, and water were manually selected. The mean vectors for these materials are shown in Figure 13, and the respective background-outlier combinations are listed in Table 1. It should be noted that the sets of pixel vectors were not mixed between the two images when forming the background-outlier pairs. Additionally, Table 2 lists the number of pixel vectors collected for each material and the number of bands, p, for each image.

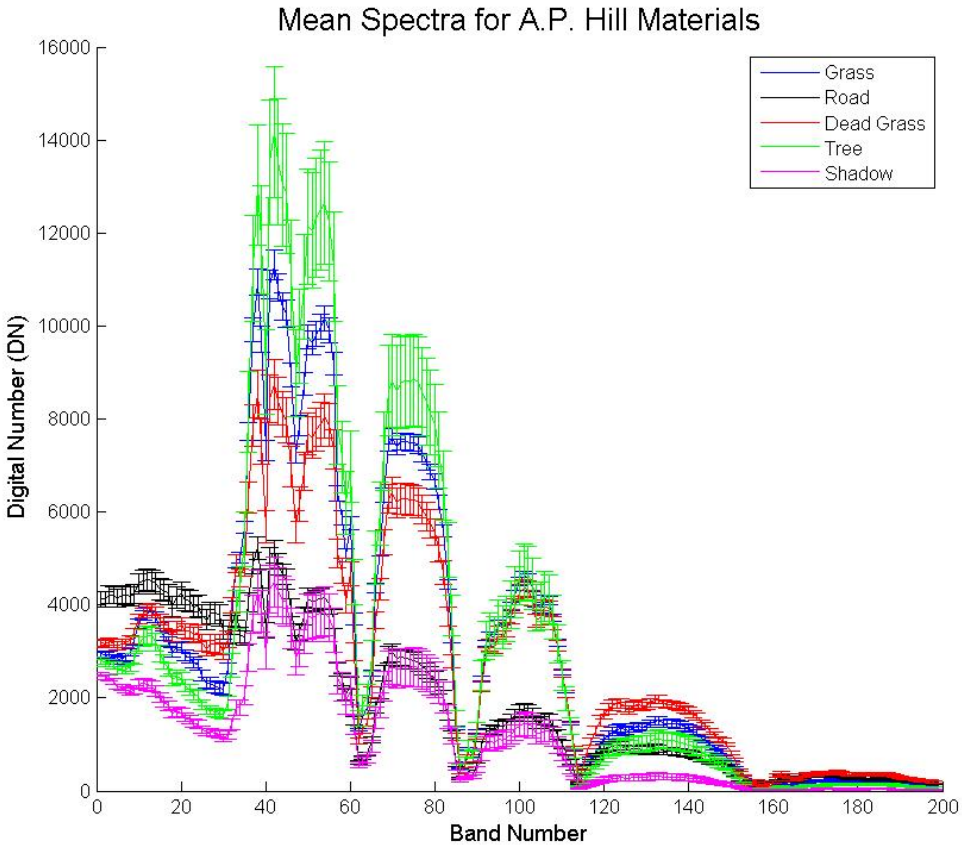


Figure 12. Mean Vectors of Spectra from Fort A.P. Hill Image

Table 1. Background-Outlier Material Combinations for Multivariate Gaussian Experiments

Fort A.P. Hill Combinations		D.C. Mall Combinations	
Background	**Outlier**	**Background**	**Outlier**
Grass	Road	Grass	Asphalt
Grass	Dead Grass	Grass	Gravel
Grass	Trees	Grass	Water
Grass	Shadow	Grass	Roof
Dead Grass	Road	Asphalt	Grass
Dead Grass	Grass	Asphalt	Water
Dead Grass	Trees	Asphalt	Gravel
Dead Grass	Shadow	Asphalt	Roof
Road	Grass	Gravel	Asphalt
Road	Dead Grass	Gravel	Grass
Road	Trees	Gravel	Water
Road	Shadow	Gravel	Roof
		Water	Asphalt
		Water	Grass
		Water	Gravel
		Water	Roof

A second comment concerning this experiment is the number, N_b, of observations used to form the background dataset. A value of 2000 observations was used to ensure a reasonably accurate estimate of the mean vector and covariance matrix in Step 3 of the experiment based on the suggested sample size of at least $10p$ given in Jimenez and Landgrebe (1998). Viewed in another light, 2000 observations is larger than the 1600 observations contained in a somewhat large 40x40 pixel processing window for a local hyperspectral anomaly detector such as the RX detector. Thus, 2000 observations is a realistic sample size that may be encountered in practice.

As a final comment on the experimentation method, it should be noted that in all cases tested, MSDs were also computed using the mean vector and covariance matrix

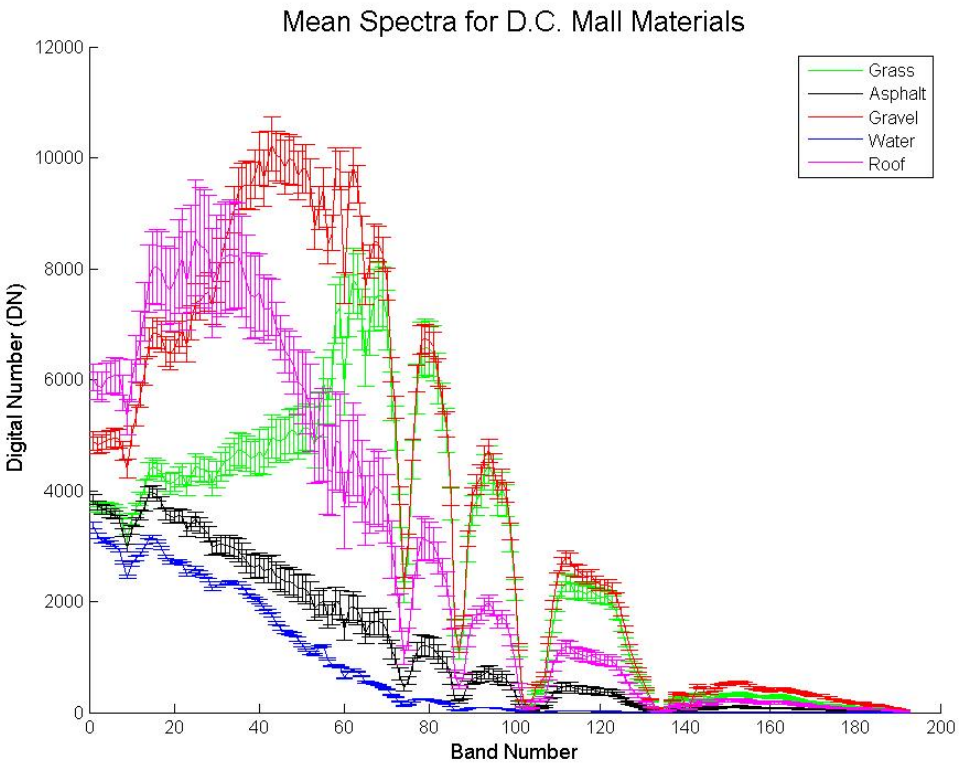


Figure 13. Mean Vectors of Spectra from D.C. Mall Image

estimated from the background data only. These MSDs were also compared to the Chi-Squared threshold to determine the expected number of true positives and false alarms given uncontaminated mean and covariance estimates. These numbers serve as a benchmark for the detection accuracy that can be achieved if the mean and covariance are robustly estimated, presumably using multivariate outlier detection methods.

The results of the multivariate Gaussian data experiments using the Fort A.P. Hill signatures are summarized in Figure 14 and Table 3. Figure 14 depicts the mean number of true positives identified by the non-robust Mahalanobis Squared Distance detector for each level of outliers tested. For reference, the line representing perfect detection performance is included in the graphs. If the masking effect is present for a given

Table 2. Sample Sizes of Spectra Collected from Fort A.P. Hill and D.C. Mall Images
This table lists the number of signatures collected from the Fort A.P. Hill and D.C. Mall images for the respective materials. These signatures are used to compute mean vectors and covariance matrices which are then used to generate new samples of multivariate Gaussian data.

Fort A.P. Hill Image Spectra (*p*=198)		**D.C. Mall Image Spectra (*p*=191)**	
Material	**Number Collected**	**Material**	**Number Collected**
Grass	2300	Grass	1219
Road	1626	Asphalt	1207
Dead Grass	600	Gravel	1227
Trees	1006	Roof	980
Shadow	805	Water	2100

background/outlier combination, the respective curve should deviate from this reference line. Additionally, error bars are used in the graphs to depict the 95% confidence intervals for the mean number of true positives. Table 3 reports the sensitivity of the false alarm rate to the number of outliers present. If our detection efforts are actually affected by swamping, we would expect to see the number of false alarms increase as the number of outliers in the dataset increases. For reference, the last row of Table 3 gives the expected number of false alarms for the dataset (consisting of 2000 observations plus the number of outliers present) using the non-robust MSD detector and a significance level of α=0.05.

From the information reported in Figure 14, two significant conclusions are evident. First, the masking effect does occur in these simulated datasets. Clear examples of masking are seen in the Grass/Road background/outlier combination and in all cases where shadow spectra are used for the outliers. For the Grass/Road case, the fraction of true positives detected falls below 0.75 with only 1.96% of the dataset contaminated by

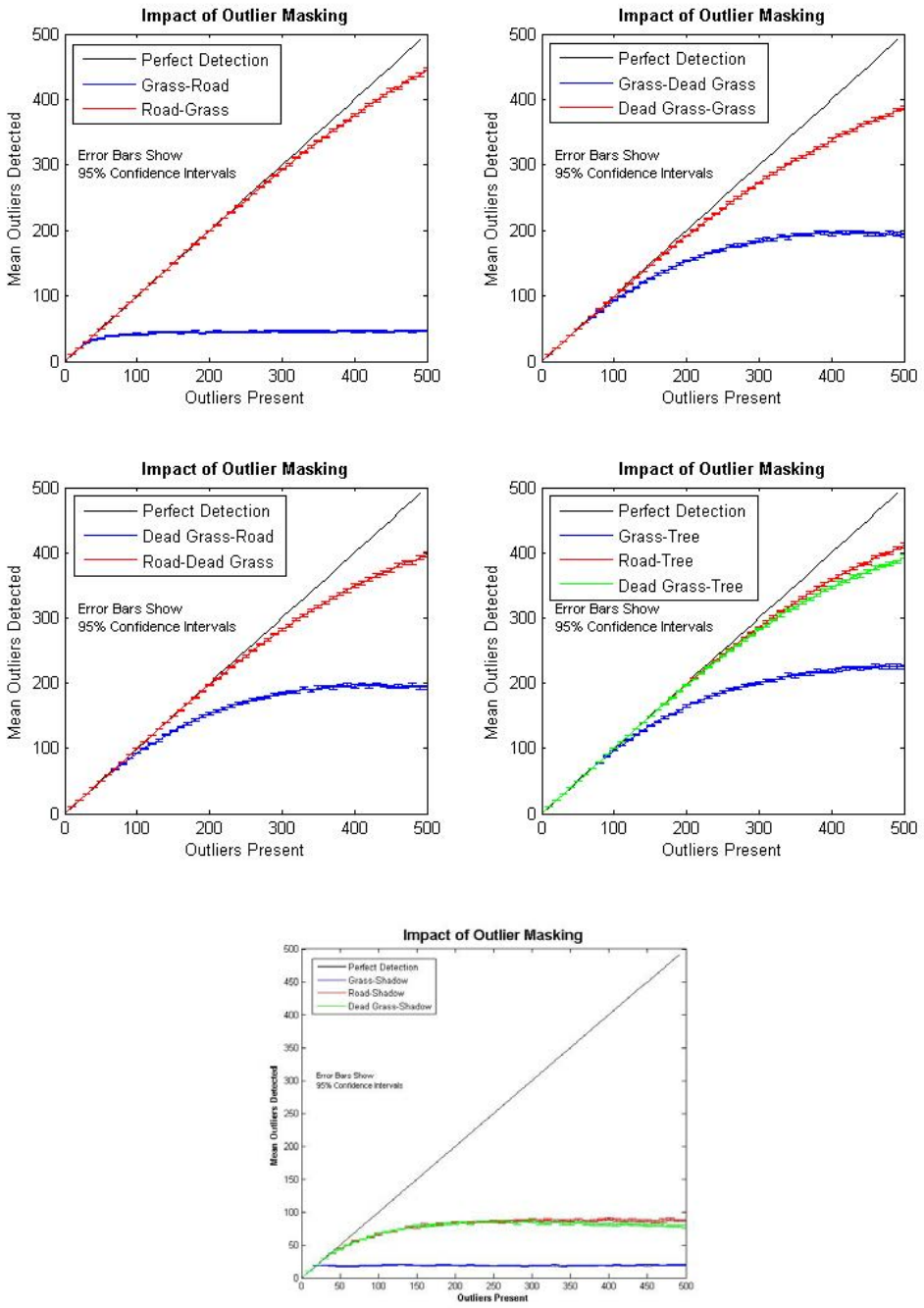


Figure 14. Outliers Detected for Fort A.P. Hill Background-Outlier Combinations

Table 3. Number of False-Alarms for Multivariate Gaussian Experiments using Fort A.P. Hill Data

Background/Outlier	95% Confidence Intervals for Mean Number of False Alarms as a Function of the Number of Outliers Present (Percent Contamination Given in Parentheses)			
	50 (2.4%)	100 (4.8%)	300 (13.0%)	500 (20.0%)
Grass/Road	(67.9, 72.0)	(66.2, 71.0)	(72.3, 76.8)	(95.2, 100.2)
Grass/Dead Grass	(57.2, 61.1)	(46.9, 50.6)	(38.1, 41.7)	(41.8, 45.4)
Grass/Trees	(48.7, 52.6)	(35.8, 39.4)	(25.9, 28.9)	(26.8, 30.1)
Grass/Shadow	(72.7, 76.6)	(77.4, 81.4)	(115.2, 120.7)	(184.6, 190.8)
Dead Grass/Road	(56.7, 60.4)	(48.6, 51.6)	(34.8, 38.6)	(37.4, 40.7)
Dead Grass/Grass	(46.1, 49.7)	(32.6, 35.4)	(13.5, 15.7)	(10.0, 11.9)
Dead Grass/Trees	(34.8, 37.7)	(23.6, 26.1)	(9.6, 11.1)	(6.6, 8.1)
Dead Grass/Shadow	(58.9, 63.3)	(55.3, 58.7)	(60.6, 64.1)	(85.8, 90.8)
Road/Grass	(37.0, 40.0)	(23.8, 25.9)	(7.6, 9.0)	(5.1, 5.9)
Road/Dead Grass	(35.8, 39.3)	(23.4, 26.4)	(10.1, 11.7)	(7.0, 8.8)
Road/Trees	(33.2, 36.0)	(21.5, 23.9)	(7.1, 8.6)	(5.7, 7.1)
Road/Shadow	(56.7, 60.9)	(54.6, 58.5)	(59.4, 63.3)	(82.3, 86.6)
Expected False Alarms for α=0.05	**102.5**	**105**	**115.0**	**125.0**

outliers. When the contamination climbs to 20% (2000 good observations and 500 outliers), the true positive fraction is only 0.10. The second conclusion is that some of the background/outlier combinations seem much more resistant to the masking effect than others, even when reversing the roles of the two materials induces significant masking. For example, using road as the background and grass as the outlier results in an average of approximately 411 true positives when 500 outliers are in the sample as opposed to only 46 when the roles of the materials are reversed. The reason for this curious result warrants further explanation which we will return to momentarily.

The primary conclusion drawn from Table 3 is that the swamping effect does not appear to manifest itself as strongly as theoretically predicted. Only in the case of the grass/shadow combination do we see the number of false alarms exceed the number expected by the choice of significance level. In fact, for virtually all cases, we see the

number of false alarms actually decrease with the contamination level until higher contamination levels are reached. This result is somewhat peculiar and will be further explained at the same time we discuss the counter-intuitive result revealed in Figure 14.

We now turn our attention to the test results produced from the D.C. Mall data. These results are shown in Figure 15 and Table 4, and reveal similar conclusions to those found using the A.P. Hill data. The masking effect is again quite significant for a number of the material combinations, but we also see that reversing the role of the materials changes the degree of masking considerably. In particular, we note that water induces a strong masking effect whenever it is used as the outlier; however, in all cases where water is the background, masking is relatively insignificant for all levels of contamination tested. In Table 4, we see the same decrease in false alarms as before, though in the cases with water acting as the outlier, the onset of swamping occurs with fewer outliers than with any of the A.P. Hill cases.

Based on these experimental results with the two different data sets, it is reasonable to conclude that masking, and to a lesser degree, swamping, can occur in simulated multivariate Gaussian data that is similar to hyperspectral data in terms of mean vectors, covariance matrices, and dimensionality. As stated previously, we confirmed this conclusion by also using the uncontaminated mean vector and covariance matrix estimate in the MSD detector for all 71400 samples tested. Upon using these estimates in the Mahalanobis distance classifier, 100% of the outliers were correctly identified. In other words, in the instances where the non-robust detector failed to find the known outliers, the failure was solely due to inaccurate mean and covariance estimates obtained from the contaminated samples. Thus, it would seem obvious that

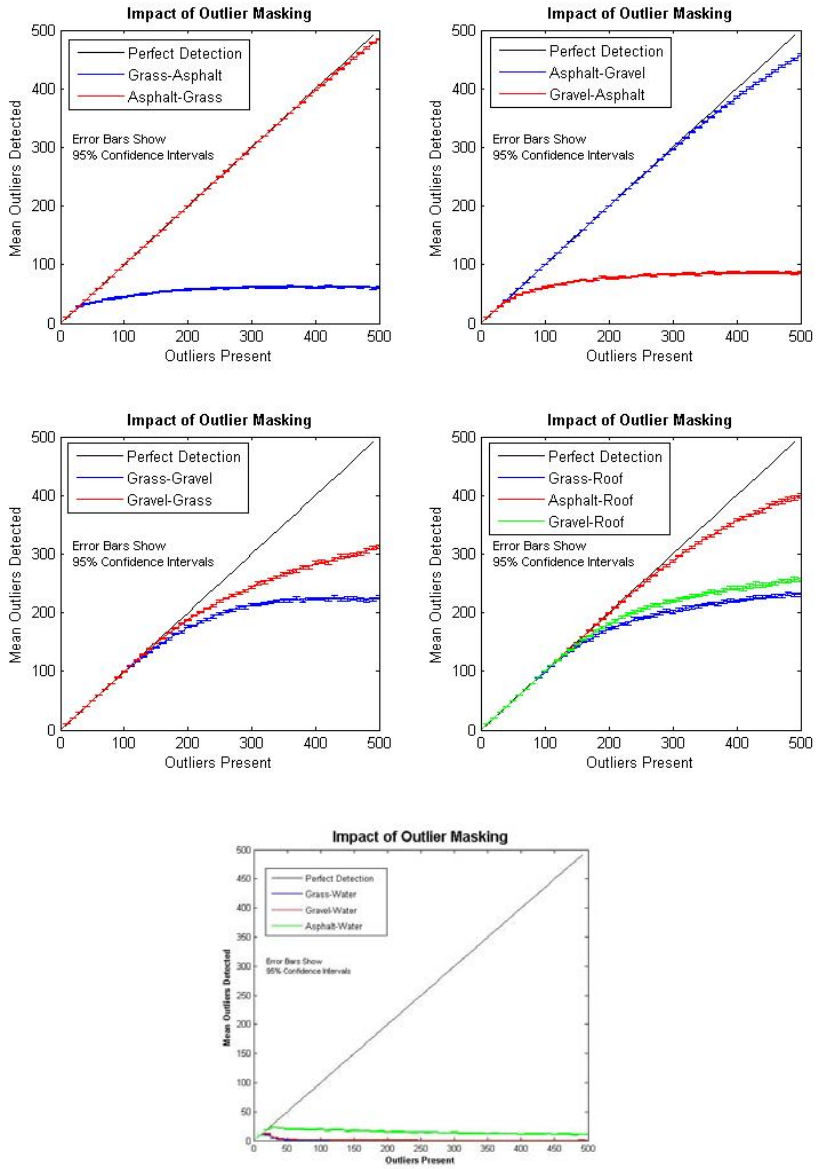


Figure 15. Outliers Detected for D.C. Mall Background-Outlier Combinations

using multivariate outlier detection methods that attempt to avoid these inaccurate estimates can be useful for detecting hyperspectral anomalies.

The results of the multivariate Gaussian experiments indicated two counter-intuitive results: 1) severe masking that occurs for a background/outlier combination does

Table 4. Number of False Alarms for Multivariate Gaussian Experiments using D.C. Mall Data

Background/Outlier	95% Confidence Intervals for Mean Number of False Alarms as a Function of the Number of Outliers Present (Percent Contamination Given in Parentheses)			
	50 (2.4%)	100 (4.8%)	300 (13.0%)	500 (20.0%)
Grass/Asphalt	(59.3, 63.8)	(57.6, 61.1)	(78.4, 83.3)	(128.7, 135.7)
Grass/Gravel	(35.5, 38.8)	(24.0, 27.1)	(14.4, 16.7)	(16.9, 19.8)
Grass/Water	(87.3, 91.4)	(103.8, 108.4)	(205.0, 209.9)	(371.4, 378.7)
Grass/Roof	(26.8, 29.3)	(16.3, 18.3)	(13.8, 15.5)	(19.6, 22.0)
Asphalt/Grass	(9.2, 11.0)	(2.7, 3.7)	(0.1, 0.4)	(0.0, 0.2)
Asphalt/Water	(72.9, 77.7)	(79.0, 83.6)	(108.4, 113.4)	(169.0, 174.5)
Asphalt/Gravel	(10.7, 12.5)	(3.6, 4.7)	(0.3, 0.7)	(0.1, 0.4)
Asphalt/Roof	(15.3, 17.6)	(6.5, 7.9)	(1.4, 2.3)	(1.5, 2.1)
Gravel/Asphalt	(54.2, 58.3)	(50.4, 54.2)	(57.4, 62.3)	(89.1, 94.3)
Gravel/Grass	(29.2, 32.7)	(18.2, 20.4)	(7.4, 8.9)	(6.5, 7.8)
Gravel/Water	(85.0, 89.0)	(99.9, 104.6)	(192.7, 197.8)	(335.9, 343.4)
Gravel/Roof	(23.0, 25.8)	(13.0, 15.6)	(10.4, 12.5)	(13.6, 16.1)
Water/Asphalt	(19.3, 21.7)	(9.1, 10.6)	(1.1, 1.9)	(0.5, 1.0)
Water/Grass	(3.0, 4.1)	(0.3, 0.7)	(0.0, 0.0)	(0.0, 0.0)
Water/Gravel	(3.2, 4.4)	(0.6, 1.1)	(0.0, 0.1)	(0.0, 0.0)
Water/Roof	(7.2, 8.7)	(1.4, 2.1)	(0.0, 0.1)	(0.0, 0.1)
Expected False Alarms for α=0.05	**102.5**	**105**	**115.0**	**125.0**

not necessarily occur when the roles of the materials are reversed; and 2) the number of false positives may actually decrease with the contamination level, a contradiction to the theoretical swamping effect. To explain these phenomenon, we focus on the Grass/Water combination derived from the D.C. Mall image. To help visualize how outlier contamination is affecting the non-robust MSD detector, we use only band 40 and band 60 for these two materials to compute representative mean vectors and covariance matrices. Next, 2000 multivariate normal random variates are generated using the grass reference estimates. We then successively add 1, 10, 50, 100, 300, and 500 randomly generated multivariate Gaussian water signatures to the grass signatures to create six

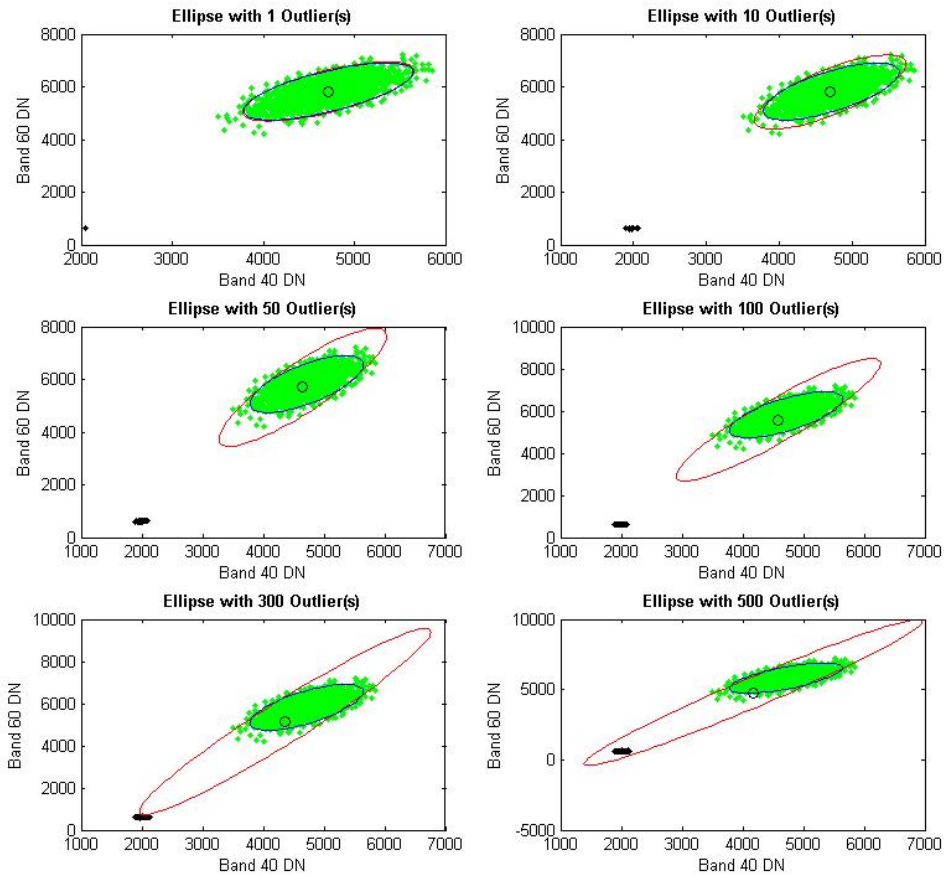


Figure 16. Covariance Ellipse Distortion for High Variance Background Material
This figure shows the impact of outlying observations on the 95% threshold ellipse for the non-robust MSD detector. The green dots represent grass observations and the black dots represent water observations acting as outliers. The blue ellipse shows the 95% threshold derived from the uncontaminated data. The black circle shows the location of the contaminated mean vector.

contaminated datasets. For each dataset, the 95% threshold ellipse is computed to visually depicted how the MSD detector will identify outliers. The datasets and ellipses are plotted in Figure 16. For reference, the ellipse generated using the covariance matrix of the clean data only is plotted in blue in each graph. We repeat this process to produce Figure 17 after first reversing the roles of the grass and water signatures.

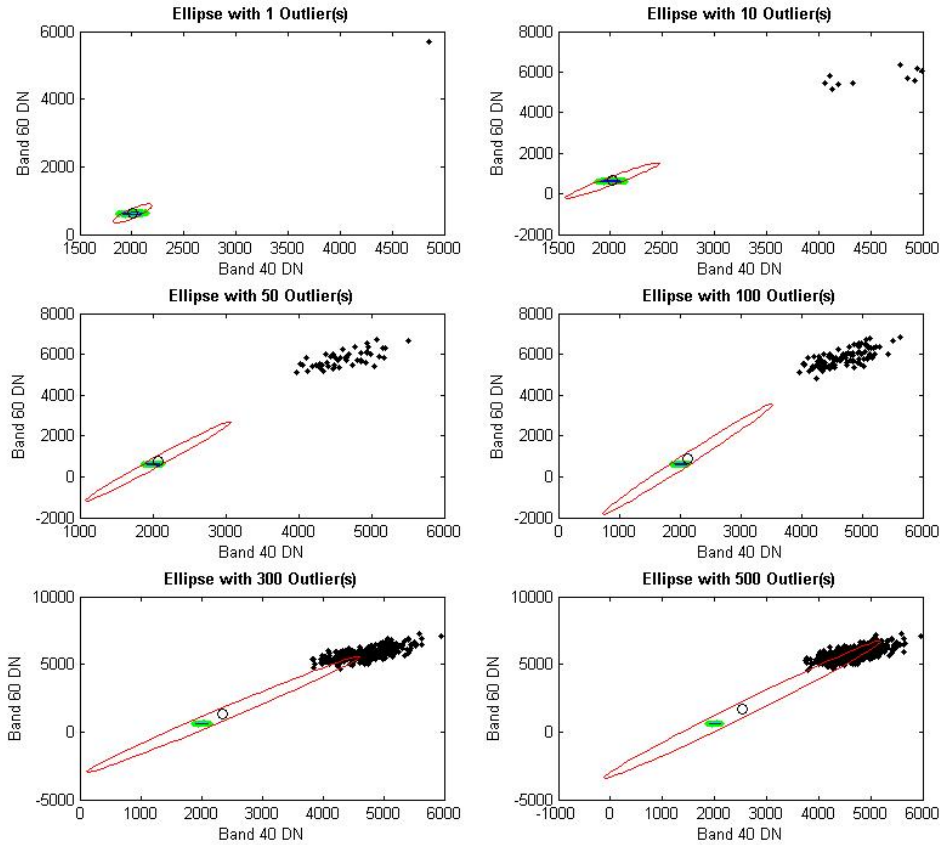


Figure 17. Covariance Ellipse Distortion for Low-Variance Background Material
This figure shows the impact of outlying observations on the 95% threshold ellipse for the non-robust MSD detector. The green dots represent water observations and the black dots represent grass observations acting as outliers. The blue ellipse shows the 95% threshold derived from the uncontaminated data. The black circle shows the location of the contaminated mean vector.

Inspection of Figures 16 and 17 reveal interesting results. First, and most importantly, the contamination level required to significantly distort the covariance ellipse is not very high. For the case when water is the background material, a single outlier (a contamination level of only 0.05%) is enough to significantly rotate the covariance axes. With ten outliers (a contamination level of 0.5%), the length of the

primary covariance axis is significantly distorted. Thus, it is obvious that relatively few outliers can significantly change the covariance structure of a dataset.

So why is the non-robust MSD detector still able to detect outliers for the Water/Grass case? The answer to this question is given by our second result. Specifically, background materials with relatively low variance compared to the variance of the outlier material may still detect outliers under high contamination levels because the distorted ellipse is too *narrow* to envelope all the outlying observations. Conversely, when the background material has relatively large variance compared to the outlier material variance, the distorted ellipse is still relatively *fat* and will eventually encompass the concentrated outliers. This phenomenon is clearly evident in Figures 16 and 17. This result would appear to explain the first peculiarity noted earlier.

Addressing the second peculiarity of the decreasing false alarms rates, we return to Figures 16 and 17 for our third result noticing that, in both background/outlier combinations, the 95% threshold ellipse initially inflates in such a manner that more of the good observations are included in the ellipse. For the case of the concentrated water background, virtually all the water observations are enveloped after only ten outliers are introduced to the dataset. The envelopment is not as severe when grass is the background; however, in this case we note that increasing the number of outliers eventually causes the ellipse to narrow, leading to some observations initially declared good to be classified as anomalies. This results helps explain why the number of false alarms initially decrease with the number of outliers for some background/outlier combinations, and then rise as the contamination level continues to increase.

Principal Axis Rotation Tests

In the preceding section it was demonstrated that, in the bivariate case, outliers can distort covariance ellipses used by the MSD detector to find outliers. However, it would be satisfying to have further evidence that these distortions are also occurring with full-dimensional hyperspectral data. To achieve this goal, an experiment similar to the preceding Gaussian data experiment was conducted as follows:

1) A sample of N_b= 2000 p-dimensional background observations was generated from a multivariate Gaussian distribution with a specified mean vector and covariance matrix.
2) A Principal Components Analysis (PCA) was performed on the original hyperspectral data used to generate the background dataset, and the first normalized component axis identified. Denote this axis as the *reference axis*, $\mathbf{e}_{ref}$.
3) A specified number, N_o, of outlier observations was randomly selected with replacement from a set of observations with a different mean vector and covariance matrix from the background data. These outliers were added to the background data to form the contaminated dataset of size $N= N_b+ N_o$.
4) A PCA was performed on the contaminated dataset and the first component axis identified. Denote this axis as the *distorted axis*, $\mathbf{e}_{dist}$.
5) The angle between the two vectors $\mathbf{e}_{ref}$ and $\mathbf{e}_{dist}$ was computed as

$$\theta_{cont} = \left(\frac{180}{\pi}\right)\cos^{-1}\left(\mathbf{e}_{ref}^{T}\mathbf{e}_{dist}\right) \quad (5.27)$$

6) The angle, θ_{clean} was also computed using the first principal component axis of the uncontaminated data rather than $\mathbf{e}_{dist}$. This angle indicates if any axis

deflection can be expected if robust methods are used to estimate the mean vector and covariance matrix.

7) Steps 1 through 6 were repeated 50 times to obtain mean estimates for the angles, θ_{cont} and θ_{clean}. The 95%-confidence intervals were also computed for these mean estimates.

8) Steps 1 through 7 were repeated for different values of N_o ranging from 0 to 100 in increments of 10.

9) Steps 1 through 8 were repeated for different background/outlier combinations. The same combinations used in the Gaussian Data Experiment were used for this experiment.

To assess the significance of the angles computed in Steps 5 and 6, a threshold angle, θ_0, was computed for each background material using a Monte Carlo simulation. For each material, this simulation entailed generating 800 samples of 2000 multivariate Gaussian observations using the material's mean vector and covariance matrix. For each sample, a PCA was performed and the first principal component axis identified. The angles were then computed between these axes and the first principal component axis of the original hyperspectral data whose mean vector and covariance matrix were used to generate the samples. For a given material, the end-result of this procedure was 800 angles. The 0.95-quantile of these angles served as the desired threshold angle, θ_0, for the material. These angles represent the expected deflection of the first principal component axis of a background material due solely to random sampling from the material's underlying multivariate Gaussian distribution.

Table 5. Principal Component Axis Distortion Results for Fort A.P. Hill Data

Background/ Outlier	θ_0	Number of Outliers					
		0	**10**	**20**	**30**	**50**	**100**
Grass/Road	1.9						
θ_{cont}		(0.7, 0.9)	(13.6, 13.7)	(15.5, 15.5)	(16.2, 16.3)	(16.9, 16.9)	(17.4, 17.4)
θ_{clean}		(0.7, 0.9)	(0.8, 1.1)	(0.8, 1.1)	(0.8, 1.0)	(0.8, 1.0)	(0.8, 1.2)
Grass/D. Grass	1.9						
θ_{cont}		(0.8, 1.1)	(3.9, 4.3)	(5.7, 6.2)	(7.1, 7.4)	(8.5, 8.8)	(10.2, 10.4)
θ_{clean}		(0.9, 1.1)	(0.8, 1.0)	(0.8, 1.1)	(0.9, 1.2)	(0.8, 1.1)	(0.8, 1.0)
Grass/Tree	1.9						
θ_{cont}		(0.6, 0.8)	(4.1, 4.6)	(6.4, 6.9)	(7.7, 8.1)	(8.8, 9.3)	(10.0, 10.4)
θ_{clean}		(0.6, 0.8)	(0.8, 1.1)	(0.8, 1.0)	(0.8, 1.0)	(0.7, 1.0)	(0.8, 1.0)
Grass/Shadow	1.9						
θ_{cont}		(0.8, 1.0)	(20.1, 20.2)	(21.9, 22.1)	(22.6, 22.7)	(23.2, 23.3)	(23.6, 23.7)
θ_{clean}		(0.8, 1.0)	(0.8, 1.1)	(0.8, 1.0)	(0.8, 1.1)	(0.8, 1.1)	(0.9, 1.2)
D. Grass/Road	1.08						
θ_{cont}		(0.5, 0.6)	(7.0, 7.1)	(11.2, 11.3)	(13.6, 13.7)	(16.6, 16.7)	(19.6, 19.6)
θ_{clean}		(0.5, 0.6)	(0.5, 0.6)	(0.5, 0.7)	(0.4, 0.5)	(0.5, 0.6)	(0.5, 0.7)
D. Grass/Grass	1.08						
θ_{cont}		(0.5, 0.6)	(1.5, 1.8)	(3.0, 3.3)	(4.1, 4.4)	(5.9, 6.1)	(8.9, 9.1)
θ_{clean}		(0.5, 0.6)	(0.5, 0.7)	(0.4, 0.6)	(0.5, 0.7)	(0.4, 0.6)	(0.5, 0.6)
D. Grass/Tree	1.08						
θ_{cont}		(0.4, 0.6)	(5.6, 5.8)	(8.5, 8.8)	(10.2, 10.6)	(12.3, 12.6)	(14.3, 14.6)
θ_{clean}		(0.4, 0.6)	(0.5, 0.6)	(0.4, 0.5)	(0.5, 0.6)	(0.4, 0.6)	(0.4, 0.6)
D. Grass/Shadow	1.08						
θ_{cont}		(0.4, 0.6)	(13.5, 13.8)	(19.9, 20.1)	(23.4, 23.6)	(26.9, 27.1)	(29.8, 30.0)
θ_{clean}		(0.4, 0.6)	(0.4, 0.6)	(0.4, 0.5)	(0.5, 0.6)	(0.4, 0.5)	(0.5, 0.6)
Road/Grass	1.4						
θ_{cont}		(0.6, 0.9)	(39.2, 39.5)	(45.3, 45.5)	(47.5. 47.6)	(49.0, 49.1)	(50.3, 50.3)
θ_{clean}		(0.6, 0.9)	(0.5, 0.8)	(0.5, 0.8)	(0.6, 0.8)	(0.5, 0.8)	(0.6, 0.8)
Road/D. Grass	1.4						
θ_{cont}		(0.5, 0.8)	(27.6, 28.2)	(37.2, 37.6)	(41.0, 41.2)	(44.3, 44.5)	(46.7, 46.8)
θ_{clean}		(0.5, 0.8)	(0.5, 0.7)	(0.5, 0.8)	(0.5, 0.7)	(0.6, 0.9)	(0.5, 0.8)
Road/Tree	1.4						
θ_{cont}		(0.6, 0.8)	(47.0, 47.2)	(50.6, 50.8)	(51.6, 51.9)	(52.4, 52.7)	(53.2, 53.4)
θ_{clean}		(0.6, 0.8)	(0.5, 0.7)	(0.6, 0.8)	(0.4, 0.6)	(0.5, 0.7)	(0.6, 0.8)
Road/Shadow	1.4						
θ_{cont}		(0.5, 0.7)	(8.9, 9.3)	(15.0, 15.6)	(19.0, 19.6)	(23.9, 24.5)	(28.7, 29.2)
θ_{clean}		(0.5, 0.7)	(0.5, 0.7)	(0.6, 0.8)	(0.4, 0.6)	(0.6, 0.8)	(0.6, 0.8)

The results of this test are summarized in Tables 5 and 6 for the Fort A.P. Hill- and D.C. Mall-derived data, respectively. The contents of these tables offer several pieces of information. First, the Monte Carlo thresholds for each background material are listed in the θ_0-column of the tables. Next, the 95% confidence intervals for the mean value of θ_{cont} for each test case are listed for representative levels of contamination.

Table 6. Principal Component Axis Distortion Results for D.C. Mall Data

Background/ Outlier	θ_0	Number of Outliers					
		0	10	20	30	50	100
Grass/Asphalt	1.13						
θ_{cont}		(0.5, 0.7)	(3.7, 3.9)	(5.6, 5.8)	(6.6, 6.8)	(7.7, 7.8)	(8.9, 9.0,)
θ_{clean}		(0.5, 0.7)	(0.5, 0.7)	(0.4, 0.6)	(0.5, 0.6)	(0.5, 0.7)	(0.6, 0.7)
Grass/Gravel	1.13						
θ_{cont}		(0.4, 0.6)	(11.8, 12.3)	(21.0, 21.5)	(27.2, 27.6)	(34.2, 34.6)	(40.6, 40.7)
θ_{clean}		(0.4, 0.6)	(0.5, 0.7)	(0.5, 0.7)	(0.5, 0.6)	(0.5, 0.7)	(0.5, 0.6)
Grass/Water	1.13						
θ_{cont}		(0.5, 0.6)	(3.1, 3.3)	(4.4, 4.5)	(5.0, 5.1)	(5.7, 5.8)	(6.3, 6.4)
θ_{clean}		(0.5, 0.6)	(0.5, 0.6)	(0.5, 0.7)	(0.5, 0.6)	(0.5, 0.6)	(0.5, 0.6)
Grass/Roof	1.13						
θ_{cont}		(0.5, 0.6)	(8.6, 9.4)	(29.3, 31.5)	(50.2, 52.4)	(65.1, 66.1)	(71.5, 72.0)
θ_{clean}		(0.5, 0.6)	(0.5, 0.7)	(0.5, 0.7)	(0.5, 0.6)	(0.5, 0.6)	(0.5, 0.6)
Asphalt/Grass	0.7						
θ_{cont}		(0.3, 0.4)	(17.5, 17.8)	(21.0, 21.2)	(22.5, 22.6)	(23.7, 23.8)	(24.6, 24.8)
θ_{clean}		(0.3, 0.4)	(0.3, 0.4)	(0.4, 0.5)	(0.3, 0.4)	(0.4, 0.5)	(0.3, 0.4)
Asphalt/Water	0.7						
θ_{cont}		(0.3, 0.4)	(1.9, 2.0)	(3.5, 3.6)	(4.9, 5.0)	(7.0, 7.1)	(10.3, 10.4)
θ_{clean}		(0.3, 0.4)	(0.3, 0.4)	(0.3, 0.4)	(0.3, 0.4)	(0.3, 0.4)	(0.3, 0.4)
Asphalt/Gravel	0.7						
θ_{cont}		(0.4, 04)	(3.6, 3.7)	(4.0, 4.0)	(4.0, 4.1)	(4.1, 4.2)	(4.2, 4.3)
θ_{clean}		(0.4, 0.4)	(0.4, 0.5)	(0.4, 0.5)	(0.3, 0.4)	(0.4, 0.4)	(0.3, 0.4)
Asphalt/Roof	0.7						
θ_{cont}		(0.3, 0.4)	(20.8, 21.9)	(24.6, 25.8)	(26.2, 27.0)	(28.1, 28.8)	(29.2, 29.6)
θ_{clean}		(0.3, 0.4)	(0.3, 0.4)	(0.3, 0.4)	(0.4, 0.5)	(0.3, 0.4)	(0.3, 0.4)
Gravel/Asphalt	1.7						
θ_{cont}		(0.7, 1.0)	(12.1, 12.3)	(14.2, 14.3)	(15.0, 15.1)	(15.7, 15.8)	(16.3, 16.4)
θ_{clean}		(0.7, 1.0)	(0.7, 0.9)	(0.8, 1.0)	(0.8, 1.0)	(0.7, 1.0)	(0.7, 1.0)
Gravel/Grass	1.7						
θ_{cont}		(0.7, 1.0)	(5.6, 6.0)	(8.0, 8.3)	(9.3, 9.7)	(11.0, 11.3)	(12.5, 12.7,)
θ_{clean}		(0.7, 1.0)	(0.7, 0.9)	(0.8, 1.0)	(0.8, 1.0)	(0.7, 1.0)	(0.7, 1.0)
Gravel/Water	1.7						
θ_{cont}		(0.7, 0.9)	(13.2, 13.4)	(15.0, 15.1)	(15.7, 15.7)	(16.3, 16.3)	(16.7, 16.8)
θ_{clean}		(0.7, 0.9)	(0.7, 1.0)	(0.8, 1.0)	(0.6, 0.9)	(0.7, 0.9)	(0.7, 0.9)
Gravel/Roof	1.7						
θ_{cont}		(0.7, 0.9)	(18.6, 19.8)	(29.7, 30.9)	(35.56, 36.5)	(40.5, 41.2)	(44.7, 45.2)
θ_{clean}		(0.7, 0.9)	(0.7, 1.0)	(0.7, 1.1)	(0.7, 0.9)	(0.7, 0.9)	(0.7, 1.0)
Water/Asphalt	1.6						
θ_{cont}		(1.2, 1.3)	(38.1, 38.6)	(39.9, 40.2)	(40.3, 40.6)	(40.7, 40.9)	(41.1, 41.3)
θ_{clean}		(1.2, 1.3)	(1.1, 1.2)	(1.1, 1.2)	(1.2, 1.3)	(1.1, 1.2)	(1.1, 1.2)
Water/Grass	1.6						
θ_{cont}		(1.1, 1.2)	(59.6, 59.9)	(59.5, 59.7)	(59.6, 59.8)	(59.7, 59.8)	(59.8, 59.9)
θ_{clean}		(1.1, 1.2)	(1.1, 1.2)	(1.1, 1.2)	(1.1, 1.2)	(1.1, 1.3)	(1.1, 1.3)
Water/Gravel	1.6						
θ_{cont}		(1.1, 1.3)	(39.4, 39.7)	(39.5, 39.7)	(39.5, 39.6)	(39.5, 39.6)	(39.5, 39.6)
θ_{clean}		(1.1, 1.3)	(1.2, 1.3)	(1.1, 1.2)	(1.2, 1.3)	(1.1, 1.2)	(1.2, 1.3)
Water/Roof	1.6						
θ_{cont}		(1.1, 1.2)	(23.4, 23.8)	(23.4, 23.7)	(23.4, 23.7)	(23.5, 23.6)	(23.4, 23.6)
θ_{clean}		(1.1, 1.2)	(1.2, 1.3)	(1.2, 1.3)	(1.1, 1.2)	(1.2, 1.3)	(1.1, 1.2)

Finally, 95% confidence intervals are also provided for the mean value of θ_{clean} for each background/outlier combination.

From these tables we clearly see that the orientation of the covariance ellipsoid in the full-dimensional hyperspectral space is indeed changing significantly for all tested background/outlier combinations when only ten outliers are added to the 2000-observation samples of good data. For the Water/Grass case in Table 6, the axis deflection is approximately 60 degrees compared to the respective 95% threshold of 1.6 degrees. We also note that the deflection angles appear to converge to a limiting value as the contamination level increases. This result is reasonable given that the increasing number of outliers are added to the same general location in *p*-dimensional space relative to the good observations.

A final observation that we take from this test is the nature of the deflection angles when the outliers are removed from the sample dataset. In all cases tested, these deflections fail to be significant relative to the Monte Carlo-derived threshold. This result is another clear indication that the distortions evident in the contaminated covariance matrices are due solely to the presence of outliers. Thus, we can again conclude that multivariate outlier detection methods that attempt to account for the affects of outliers in their detection computations are promising alternatives for hyperspectral anomaly detection.

Multivariate-t Data Experiments

In all of the experiments discussed to this point, we have used multivariate normal random generators to create our simulated datasets. However, whether or not hyperspectral data is actually Gaussian in nature is an on-going debate in the

hyperspectral research community. An alternative to the Gaussian model for hyperspectral data is to use heavy-tailed, elliptically contoured distributions, as discussed by Kerekes and Manolakis (2004), Manolakes et al. (2005), and Caterall (2004). Given that many multivariate outlier detection methods assume Gaussian data, particularly when statistical testing is used to detect outliers, we ask the following: can we expect reasonable performance from a detector that assumes Gaussian data when the actual data has a more heavy-tailed distribution?

To resolve this issue, we conducted an experiment using the representative mean vector and covariance matrices from the Fort A.P. Hill grass and road spectra. Using the grass mean vector and covariance estimates, 2000 simulated signatures were randomly generated using a multivariate *t*-distribution. This sample of good data was then augmented with a specified number of outliers generated from a multivariate *t*-distribution using the road mean vector and covariance matrix. Contaminated samples were formed in this manner for contamination levels ranging from 0 to 250 outliers and the degrees of freedom for the multivariate *t*-distribution ranging from 4 to 32 in increments of 4. For each combination of contamination level and degrees of freedom, 100 contaminated samples were generated and the mean number of true positives and false positives averaged over these 100 samples.

For each of the 100 samples at a specific test case, three detection methods were used to find the outliers. Method 1 uses the mean vector and covariance matrix computed from the good sample data to obtain Mahalanobis squared distances for all the observations in the sample. Because these squared distances, δ, come from multivariate *t*-distributed data, they are distributed as

$$\frac{\delta \nu}{p(\nu - 2)} \sim F_{p,\nu} \tag{5.28}$$

where p is the dimensionality of the data, ν is the degrees of freedom, and $F_{a,b}$ is the F-distribution with a and b degrees of freedom (Manolakis et al, 2005). Using this fact, any observations whose scaled squared distances exceed the 0.95-quantile from the F-distribution with p and ν degrees of freedom are identified as outliers. Detecting outliers in this manner simulates how well the MSD detector is expected to perform if the correct mean vector, covariance matrix, and degrees of freedom can be accurately estimated from the data.

Method 2 is similar to the first with one significant difference: the mean vector and covariance matrix are estimated from the contaminated sample. Thus, we still assume that the degrees of freedom for the underlying t-distributed data can be determined for the sample.

In Method 3, we assume that the data is Gaussian and use the BACON outlier detection method discussed in Billor, Hadi, and Velleman (2000) to identify outliers. By using these three different detection methods on each of the simulated samples, we can assess the implications of using robust Gaussian outlier detection methods to detect outliers as opposed to using detection methods that attempt to more accurately model the underlying distribution of data but use non-robust mean vector and covariance matrix estimation methods.

The results of this experiment are reported in Figures 18 and 19. Each graph in Figure 18 corresponds to a different value of ν used to generate the simulated data. The lines in the graphs indicate the mean number of true positives detected by a particular

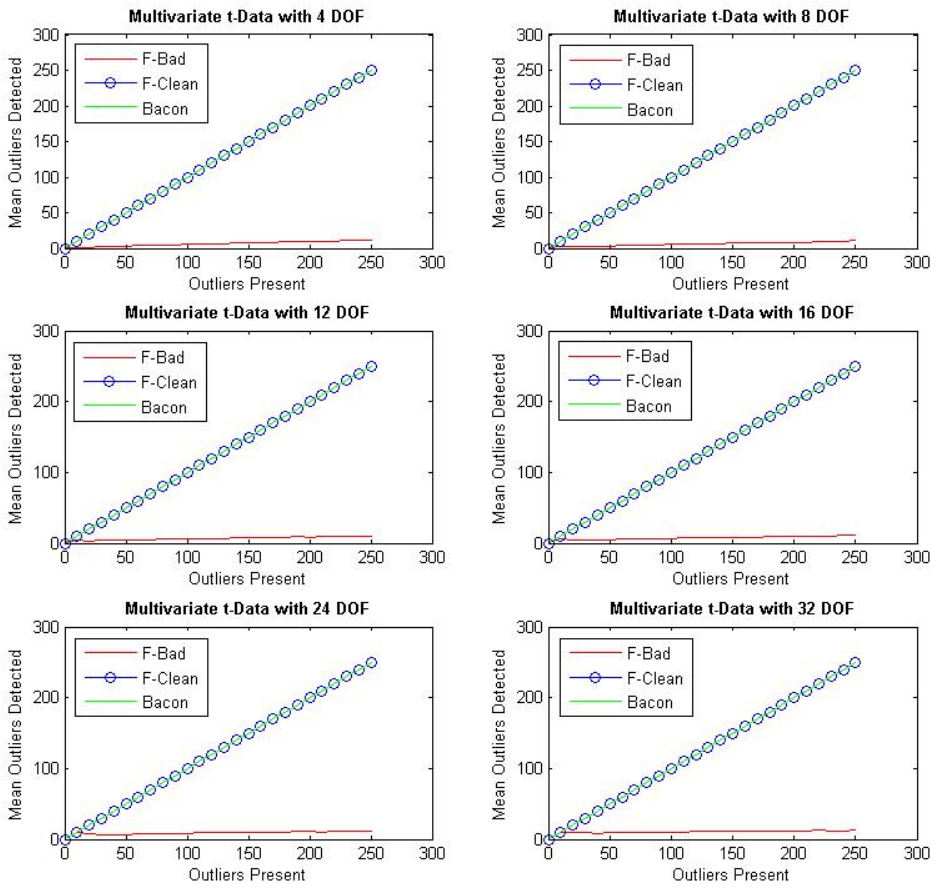


Figure 18. Number of Outliers Detected for Multivariate-*t* Data Tests

detector as the contamination level increases. The *F-Clean* line refers to Method 1, *F-Bad* refers to the Method 2, and *Bacon* line refers to Method 3. To reduce clutter, confidence intervals are not provided on the graphs; however, the maximum standard error encountered over all test cases does not exceed 3.6. Figure 19 is similar to Figure 18, but reports the mean false positives obtained for the experiment.

Inspection of Figure 18 reveals several notable results. First, we see that even if we can determine the proper *F*-distribution for the Mahalanobis distances with perfect

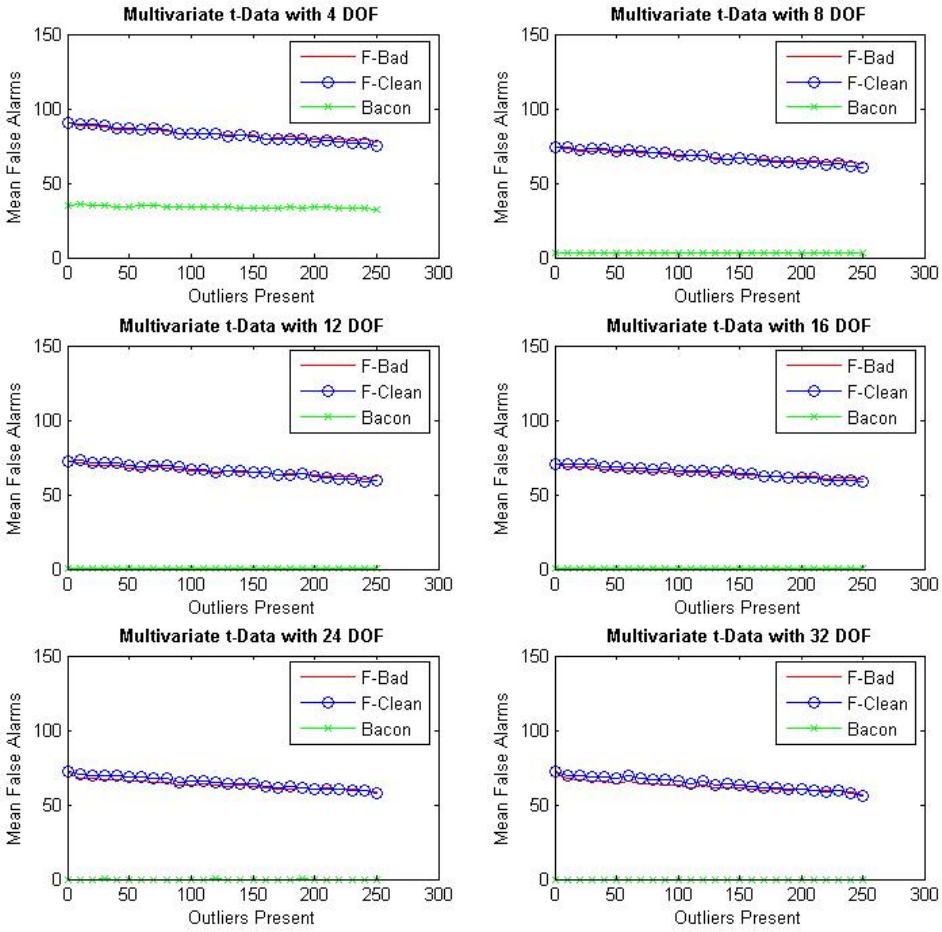


Figure 19. Number of False-Alarms for Multivariate-*t* Data Tests

clarity, using contaminated data to estimate the mean vector and covariance matrix may still result in significant masking, as indicated by the "F-Bad" lines. Second, using accurate mean vector and covariance estimates, along with the correct distribution of the Mahalanobis squared distances, perfect detection is achieved for all cases tested. This result, combined with the first, again underscores the importance of robustly estimating the sample's mean vector and covariance matrix. Finally, we see that if we inaccurately assume a Gaussian model for the data, but use a robust-distance method to find the outliers, perfect detection is also obtained. Further, this result holds even for data with

very heavy tails. The implication of this result is that using multivariate outlier detectors to find hyperspectral anomalies—even when using an inaccurate Gaussian assumption—may be a practical alternative to the difficult task of assuming a non-Gaussian model and attempting to estimate the proper distribution of the Mahalanobis distances.

A further justification for using multivariate outlier detection methods is provided in Figure 19. In these graphs it is seen that the BACON detector produces fewer false alarms over the test cases than the other detectors. In fact, as the degrees of freedom used to generate the multivariate *t*-data increase and the data becomes more Gaussian, the number of false alarms for the BACON detector are close to zero. The primary reason for the BACON detector's low false alarm rates is the use of a Bonferroni significance test to threshold the Mahalanobis distances. This significance test uses a significance level of α/n as opposed to α, as used by the other two detectors. Using the same procedure for the other detectors would likely reduce their false alarm rates in a similar manner, though further testing is required to confirm this assertion.

In the multivariate Gaussian, principal axis rotation, and multivariate-*t* data tests presented in the previous paragraphs we demonstrated the validity of the masking and swamping problems in the context of hyperspectral data, and showed the potential benefit of using multivariate outlier detection methods to avoid these problems. The outcome of these tests clearly demonstrate that masking is a realistic concern for simulated Gaussian and multivariate-*t* data with mean vectors, covariance structures, and dimensionality similar to actual hyperspectral data. It was also shown that fairly high contamination levels must exist in a dataset for swamping to become a significant factor. However, it was also revealed that the apparent absence of masking and swamping does not mean that

the underlying structure of the data is not distorted by outliers. On the contrary, distortions can be expected with contamination levels as low as 0.05%.

Finally, and perhaps most significantly, it was demonstrated that using mean vector and covariance estimates with the influence of outliers removed—the ultimate goal of robust-distance outlier detectors—achieved perfect detection results for the cases tested, even if heavy-tailed distributions are used. In the following section, we will evaluate different multivariate outlier detection methods to identify those methods that are well-suited for detecting anomalies in hyperspectral data.

Evaluation of Multivariate Outlier Detection Methods

As stated previously, the experiments conducted thus far were conducted using both contaminated and uncontaminated datasets to determine the ideal benefit of using robust mean vector and covariance matrix estimates. In the Multivariate Gaussian and Multivariate-*t* data tests, it was found that using robust estimates resulted in perfect outlier detection for all the cases tested. Further, it was demonstrated in the Multivariate-*t* Data experiment that the BACON multivariate outlier detector was effective in finding outliers in simulated hyperspectral datasets, even when the data deviates from BACON's Gaussian assumption. We now build upon these findings by evaluating the ability of several different multivariate outlier detection methods to detect outliers in simulated multivariate Gaussian and multivariate-*t* datasets. The methods that we consider for this evaluation are the BACON detector, the FAST-MCD detector of Rousseeuw and van Dreissen (1999), a modification of the Stahel-Donoho Estimator (SDE) detector originally proposed by Stahel (1981) and Donoho (1982), and the angle-based detector proposed by Juan and Prieto (2001). These methods were chosen for this study because

they represent a range of robust MSD and non-traditional detectors and are computationally conducive to handling large, high-dimensional datasets.

In the following paragraphs, we first summarize the four algorithms used in the evaluation, and then we compare their ability to detect outliers in simulated datasets. It should be noted that complete details of each algorithm are not provided in this dissertation. For a more complete explanation of the algorithms, the reader should consult the original technical articles.

The FAST-MCD Detector

The primary objective of the FAST-MCD detector is to rapidly search for a solution to the following non-linear optimization problem:

$$\begin{aligned} &\min \quad \det(\mathbf{S}) = \det\left(\frac{\sum_{i=1}^{n} t_i \left(\mathbf{x}_i - \sum_{j=1}^{n} t_j \mathbf{x}_j \Big/ \sum_{j=1}^{n} t_j \right) \left(\mathbf{x}_i - \sum_{j=1}^{n} t_j \mathbf{x}_j \Big/ \sum_{j=1}^{n} t_j \right)^T}{\sum_{i=1}^{n} t_i - 1} \right) \\ &\text{s.t.} \\ &\qquad \sum_{i=1}^{n} t_i = h = \left[\frac{(n+p+1)}{2} \right], \; t_i \in \{0,1\}, \end{aligned} \tag{5.29}$$

where $\mathbf{x}_i$ is an observation vector, n is the total number of observations in the dataset, det(•) is the determinant operator, and $(\bullet)^T$ is the transpose operator. The search is conducted by first selecting a user-specified number of random subsets of size h from the original dataset. For each subset, a C-step procedure is performed consisting of the following: 1) the Mahalanobis squared distances are computed for all observations in the dataset using the mean vector and covariance matrix of the subset data; 2) the distances

are sorted; and 3) the *h* observations from the original dataset with smallest squared distances are used to form a new subset. Rousseeuw and van Driessen (1999) prove that repeated applications of the *C*-step procedure to a dataset will produce a new subset of size *h* that has a covariance determinant less than or equal to that of the preceding estimate.

After applying the *C*-step procedure to each random subset until convergence of the respective covariance determinant, the subset that produced the smallest covariance determinant is identified. The mean vector of this subset is used for the robust mean estimate of the original dataset, and the covariance matrix of the subset is used as the robust estimate of the data's shape matrix. This shape matrix is then scaled to be consistent with Gaussian data in the sense that the median of the Mahalanobis squared distances obtained using the scaled covariance matrix is equal to the 0.5-quantile of a Chi-Squared distribution with *p* degrees of freedom. The resulting scaled matrix becomes the robust covariance estimate. These robust estimates are used to compute robust Mahalanobis squared distances for each observation in the dataset. Any observation whose squared distance exceeds an appropriate quantile of the Chi-squared distribution with *p* degrees of freedom is considered an outlier.

To allow the FAST-MCD method to handle very large datasets, Rousseeuw and van Driessen also propose a nesting scheme that initiates the search by selecting a random sample of the original data and forming the initial subsets from this random sample. As the search proceeds, more and more of the original data is included in the search until the final solution is obtained. The FAST-MCD method is implemented in S-Plus 4.5 as the cov.mcd function and in SAS/IML 7 as the MCD function.

The BACON Detector

The BACON detector proposed by Billor et al. (2000) is a robust distance detector designed to rapidly identify outliers in very large datasets. The algorithm is relatively simple to implement with the added advantage that it is very fast relative to the other detectors we consider, even for extremely large datasets.

BACON attempts to find outlying observations by first identifying a basic subset of *clean* observations close to the centroid of the data. The user has the option of using either a robust, non-affine equivariant or a non-robust, affine equivariant method to find this subset. Once determined, the basic subset is used to estimate a mean vector and shape matrix for the dataset. The shape matrix is multiplied by a small-sample correction factor derived by Billor et al. from a Monte Carlo simulation study. Using the mean vector and scaled shape matrix, Mahalanobis squared distances are computed for all observations in the dataset. Any observations whose squared distances are less than an appropriate quantile of the Chi-Squared distribution with p degrees of freedom are then used to form a new basic subset. This process is repeated until the basic subset fails to increase in size between iterations. Any observations not in the basic subset when the algorithm terminates are considered outliers.

The Juan-Prieto Detector

The outlier detector proposed by Juan and Prieto (2001)—hereafter referred to as the Juan-Prieto detector—is a non-traditional outlier detection method that avoids the computation of Mahalanobis distances altogether. Thus, the method offers a good contrast to the other methods we consider. The Juan-Prieto detector is also designed to

locate concentrated outliers, which, intuitively, would seem to match well with the problem of finding targets in a hyperspectral scene.

The underlying statistical theory exploited by the Juan-Prieto detector is that *p*-dimensional Gaussian data projected onto the *p*-dimensional unit hypersphere has a Uniform distribution. Further, the angles between each normalized vector and a reference direction will have a Beta distribution. These properties are also reasonably robust to departures from normality if the data is elliptically symmetrical. With this theory in-mind, the Juan-Prieto detector begins by normalizing all the observation vectors so that they have a magnitude of one, and thus lie on the *p*-dimensional unit hypersphere. A reference direction is then chosen using a non-linear optimization method suggested by Juan and Prieto, and the angles between the reference direction and each normalized vector are computed. To determine if these angles have the prescribed Beta distribution, they are entered as arguments to the inverse of the appropriate Beta distribution function. If the angles indeed have the proper distribution, the outputs to the inverse distribution function should, in turn, have a Uniform distribution. This hypothesis is tested by analyzing the maximum spacing between the ordered inverse function outputs. If the maximum spacing is not consistent with a Uniform distribution, all corresponding observations beyond the maximum spacing in the ordered inverse function outputs are considered outliers.

The Modified Stahel-Donoho Estimator (SDE) Detector

The original SDE detector proposed by Maronna and Yohai (1995) is a robust distance method that arrives at mean vector and covariance matrix estimates using a

robust estimation method originally proposed by Stahel (1981) and Donoho (1982). The SDE mean vector, **T**, and covariance matrix, **S**, are given by:

$$\mathbf{T}(\mathbf{X}) = \frac{\sum_{i=1}^{n} w_i \mathbf{x}_i}{\sum_{i=1}^{n} w_i} \tag{5.30}$$

and

$$\mathbf{S}(\mathbf{X}) = \frac{\sum_{i=1}^{n} w_i \left(\mathbf{x}_i - \mathbf{T}(\mathbf{X})\right)\left(\mathbf{x}_i - \mathbf{T}(\mathbf{X})\right)^T}{\sum_{i=1}^{n} w_i} \tag{5.31}$$

where the w_i are weights whose magnitudes depend on the degree to which the corresponding observation is outlying. Though different weight functions can be employed, Maronna and Yohai demonstrate empirically that the following function provides good statistical efficiency of the estimator:

$$w_i = I\left(r_i \le c\right) + \left(c / r_i\right)^q I\left(r_i > c\right) \tag{5.32}$$

where

$I(\square)$ = the indicator function, and
r_i = the measure of "outlyingness" for observation *i*.

The parameters *c* and *q* in (5.32) are constants that can be derived using Monte Carlo simulation to achieve an acceptable level of bias for the estimator. The r_i metric in (5.32) for an observation vector, $\mathbf{x}_i$, is defined as:

$$r_i = \sup_{\|\mathbf{a}\|=1} \left\{ \frac{\left| \mathbf{a}^T \mathbf{x}_i - \operatorname{med}_j \left(\mathbf{a}^T \mathbf{x}_j \right) \right|}{\operatorname{med}_k \left| \mathbf{a}^T \mathbf{x}_k - \operatorname{med}_j \left(\mathbf{a}^T \mathbf{x}_j \right) \right|} \right\} \tag{5.33}$$

The interpretation of (5.33) is we are looking for some projection vector, **a**, on the *p*-dimensional unit hypersphere that maximizes the standardized distance between the projection of $\mathbf{x}_i$ onto **a** and the centroid of the projected dataset onto **a**. To ensure a robust estimate of r_i, the median of the projected data is used to estimate the centroid, and the median absolute deviation (MAD) is used to estimate the standard deviation. The rationale for using (5.33) to measure *outlyingness* is that for elliptically symmetric data, an outlier in *p*-dimensional space will be an outlier in some univariate projection of the data.

Once the robust estimates of (5.30) and (5.31) are obtained, they can be used to compute robust Mahalanobis distances for all the observations in the dataset. Maronna and Yohai suggest that these distances are *F*-distributed, and provide a suitable critical value for screening them for outliers. Hence, implementing the SDE outlier detector entails: 1) computation of the r_i for each observation; 2) using these values to compute (5.30) and (5.31); 3) using the robust estimates to compute Mahalanobis squared distances for the observations; and 4) using the appropriate critical value to screen the distances for outliers. The practical challenge in using this detector, however, is solving the non-linear optimization problem given by (5.33). Due to the non-differentiable objective function, derivative-free optimization methods must be used to search for a local solution.

Rather than using penalty or barrier function methods to solve (5.33), Maronna and Yohai suggest generating random points on the unit hypersphere that have a Uniform distribution. Each point, or vector, is then substituted into (5.33) to find an approximate solution to the maximization problem. As an alternative to random vector generation, we propose using number theoretic methods (NTM) to generate points that are uniformly scattered—as opposed to uniformly distributed—on the unit hypersphere. We favor this method because NTM point generation requires fewer points to evenly cover the unit hypersphere than random point generation, as explained in Fang and Wang (1994). Thus, given the same number of points generated by the two methods, we can be more confident of an even search of the feasible region with NTM generation than with random generation.

By modifying Maronna and Yohai's SDE detector using NTM point generation, we define the SDE-NTM generator as follows:

1) Generate a set of uniformly scattered points, or vectors, an the p-dimensional unit hypersphere using the TFWW method outlined in Fang and Wang (1994).
2) Use the vectors from Step 1 to find an approximate solution to (5.33) for each observation.
3) Use the r_i's from Step 2 to compute the mean vector and covariance estimates given by (5.30) and (5.31), respectively.
4) Compute the Mahalanobis squared distance for each observation relative to the robust mean and covariance estimates computed in Step 3.
5) Scale the squared distances from Step 4 by the median of the squared distances, and declare as outliers any observation whose scaled squared

distance, d^*, exceeds $F(\alpha/n; p, n\text{-}2p)/F(0.5; p, n\text{-}2p)$, where α is a specified significance level and $F(\bullet; a, b)$ is the F-distribution function with a and b degrees of freedom.

The critical value given in Step 5 is based on empirical simulation studies conducted by Maronna and Yohai that indicate the following:

$$d^*_{(i)} \approx F(i/(n+1); p; n-k) / F(0.5; p; n-k) \tag{5.34}$$

where

$$d^*_{(i)} = \text{the } i\text{th ordered } d^*, \text{ and}$$
$$k \in [p, 2p].$$

Though the SDE-NTM detector offers a more efficient procedure for finding approximate solutions to (5.33) relative to the original SDE detector, the method is still computationally expensive, particularly in high-dimensions. To reduce the number of unnecessary computations, we suggest computing and storing uniformly scattered sets of points for different combinations of dimensionality and numbers of points.

The preceding paragraphs outlined the four multivariate outlier detection methods used in the comparison tests described in the following sections. Again, these methods were selected based on their perceived ability to handle very large datasets, as well as the different detection strategies they employ. By comparing the relative performance of this diverse set of detectors, it is hoped that useful insights may be obtained as to how best multivariate outlier detection may be used to find hyperspectral anomalies.

Algorithm Comparisons

To compare the BACON, FAST-MCD, Juan-Prieto, and SDE-NTM outlier detection methods, we used a test similar to the Multivariate Gaussian Data experiment. Specifically, the test proceeded as follows:

1) Generate 2000 *p*-dimensional observations from a multivariate variate Gaussian distribution with mean vector and covariance matrix derived from the same Fort A.P. Hill or D.C. Mall datasets used in the previous experiments.
2) Generate a specified number of outlier observations from a multivariate Gaussian distribution with mean vector and covariance matrix derived from one of the Fort A.P. Hill or D.C. Mall datasets, and combine these outliers with the background dataset created in Step 1.
3) Apply each of the four outlier detection methods to the contaminated dataset and record the number of true positives and false alarms detected by each method. As a benchmark, apply the classical MSD detector to the dataset as well to determine the detection accuracy if non-robust methods are used.
4) Repeat Steps 1 through 3 30 times and estimate the mean true positives and false alarms detected by each method. Also compute the standard error for each mean estimate.
5) Repeat Steps 1 through 4 for a higher level of contamination. The contamination ranged from 0 to 500 outliers in increments of 50 outliers.
6) Repeat Steps 1 through 5 for a different background-outlier combination. The combinations used in this experiment correspond to those most affected by

Table 7. Background-Outlier Material Combinations used for Multivariate Outlier Detector Comparisons

Fort A.P. Hill Combinations		D.C. Mall Combinations	
Background	**Outlier**	**Background**	**Outlier**
Grass	Road	Grass	Asphalt
Grass	Shadow	Grass	Water
Dead Grass	Shadow	Asphalt	Water
Roac	Shadow	Gravel	Asphalt

masking in the previous experiments. These combinations are listed in Table 7 for the Fort A.P Hill and D.C. Mall images.

7) Repeat Steps 1 through 6 using a multivariate-*t* distribution with twelve degrees of freedom to generate the background and outlier observations in Steps 1 and 2, respectively. This distribution was chosen based on conclusions presented by Manolakis and Mardin (2002), Kerekes and Manolakis (2004), and Manolakis et al (2005).

The outcome of our tests are summarized in Tables 8 through 11. Tables 8 and 9 show the mean true-positives obtained by each detector for the Gaussian and multivariate *t*-distributed data, respectively. Similarly, Tables 10 and 11 show the mean number of false-positives for the two distributions. For each mean value, the standard error is also reported as a measure of detector performance variability. To keep these tables as concise as possible, we have only included results from a subset of the contamination levels tested; however, we feel they are sufficient to show the relative performance of the detectors.

From Tables 8 and 9, several conclusions can be made. First, it is clear that the classical, non-robust Mahalanobis distance detector suffers significantly from masking, as indicated by the low number of true positives across all the material combinations,

Table 8. True Positives for Outlier Detection Method Comparison Tests (Multivariate Gaussian Data)

Background/ Outlier	Number Of Outliers	True Positives by Method									
		Classical		BACON		F.MCD		Juan-Prieto		SDE-NTM	
		Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Grass/ Road (A.P. Hill)	50	46.3	2.1	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	12.1	2.8	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0
	300	2.9	1.5	300.0	0.0	300.0	0.0	286.4	6.8	300.0	0.0
	500	1.5	1.1	500.0	0.0	500.0	0.0	481.0	13.9	500.0	0.0
Grass/ Shadow (A.P. Hill)	50	45.0	2.3	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	14.2	3.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0
	300	4.4	1.8	300.0	0.0	300.0	0.0	283.9	10.5	300.0	0.0
	500	2.2	1.1	500.0	0.0	500.0	0.0	468.6	17.1	500.0	0.0
Dead Grass/ Shadow (A.P. Hill)	50	45.3	2.0	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	15.7	2.7	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0
	300	4.8	1.4	300.0	0.0	300.0	0.0	284.1	10.4	300.0	0.0
	500	2.4	1.8	500.0	0.0	500.0	0.0	471.4	15.7	500.0	0.0
Road/ Shadow (A.P. Hill)	50	49.2	0.9	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	72.9	3.2	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0
	300	68.4	5.3	300.0	0.0	300.0	0.0	7.5	41.1	300.0	0.0
	500	43.4	5.4	500.0	0.0	500.0	0.0	21.9	84.5	500.0	0.0
Grass/ Asphalt (D.C. Mall)	50	11.7	3.1	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	0.0	0.2	100.0	0.0	100.0	0.0	70.0	43.1	100.0	0.0
	300	0.0	0.0	300.0	0.0	300.0	0.0	297.5	2.6	300.0	0.0
	500	0.1	0.3	500.0	0.0	500.0	0.0	491.8	5.3	500.0	0.0
Grass/ Water (D.C. Mall)	50	0.0	0.0	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	0.0	0.0	100.0	0.0	100.0	0.0	86.7	34.6	100.0	0.0
	300	0.0	0.0	300.0	0.0	300.0	0.0	300.0	0.0	300.0	0.0
	500	0.0	0.0	500.0	0.0	500.0	0.0	500.0	0.0	500.0	0.0
Asphalt/ Water (D.C. Mall)	50	27.1	2.7	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	0.0	0.0	100.0	0.0	100.0	0.0	88.9	30.2	100.0	0.0
	300	0.0	0.0	300.0	0.0	300.0	0.0	299.2	1.1	300.0	0.0
	500	0.0	0.0	500.0	0.0	500.0	0.0	496.8	2.8	500.0	0.0
Gravel/ Asphalt (D.C. Mall)	50	20.5	2.6	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	0.0	0.0	100.0	0.0	100.0	0.0	89.2	30.3	100.0	0.0
	300	0.0	0.0	300.0	0.0	300.0	0.0	299.4	0.9	300.0	0.0
	500	0.0	0.0	500.0	0.0	500.0	0.0	497.8	1.8	500.0	0.0

Table 9. True Positives for Outlier Detection Method Comparison Tests (Multivariate-*t* Data)

Background/ Outlier	Outliers Present	True Positives by Method									
		Classical		BACON		F.MCD		Juan-Prieto		SDE-NTM	
		Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Grass/ Road (A.P. Hill)	50	45.6	1.9	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	12.4	2.3	100.0	0.0	100.0	0.0	4.9	18.9	100.0	0.0
	300	5.8	1.6	300.0	0.0	300.0	0.0	286.4	5.8	300.0	0.0
	500	4.1	2.0	500.0	0.0	500.0	0.0	475.2	13.3	500.0	0.0
Grass/ Shadow (A.P. Hill)	50	44.2	1.9	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	15.2	2.8	100.0	0.0	100.0	0.0	3.2	12.5	100.0	0.0
	300	8.7	2.8	300.0	0.0	300.0	0.0	278.5	11.8	300.0	0.0
	500	8.2	2.4	500.0	0.0	500.0	0.0	465.0	23.4	500.0	0.0
Dead Grass/ Shadow (A.P. Hill)	50	44.3	2.5	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	16.8	3.3	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0
	300	8.7	2.4	300.0	0.0	300.0	0.0	282.4	8.7	300.0	0.0
	500	7.6	2.5	500.0	0.0	500.0	0.0	468.6	17.4	500.0	0.0
Road/ Shadow (A.P. Hill)	50	48.4	1.0	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	63.9	4.1	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0
	300	67.6	5.8	300.0	0.0	300.0	0.0	32.2	83.8	300.0	0.0
	500	59.3	4.0	500.0	0.2	500.0	0.0	85.8	174.8	500.0	0.0
Grass/ Asphalt (D.C. Mall)	50	9.7	2.2	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	0.1	0.3	100.0	0.0	100.0	0.0	71.0	43.6	100.0	0.0
	300	0.2	0.6	300.0	0.0	300.0	0.0	295.9	3.2	300.0	0.0
	500	0.2	0.4	500.0	0.0	500.0	0.0	487.4	7.9	500.0	0.0
Grass/ Water (D.C. Mall)	50	0.0	0.0	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	0.0	0.0	100.0	0.0	100.0	0.0	90.0	30.5	100.0	0.0
	300	0.0	0.0	300.0	0.0	300.0	0.0	300.0	0.0	300.0	0.0
	500	0.0	0.0	500.0	0.0	500.0	0.0	500.0	0.2	500.0	0.0
Asphalt/ Water (D.C. Mall)	50	23.4	2.9	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	0.1	0.4	100.0	0.0	100.0	0.0	88.6	30.1	100.0	0.0
	300	0.0	0.0	300.0	0.0	300.0	0.0	298.6	1.1	300.0	0.0
	500	0.0	0.2	500.0	0.0	500.0	0.0	493.4	3.6	500.0	0.0
Gravel/ Asphalt (D.C. Mall)	50	18.8	2.2	50.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
	100	0.0	0.0	100.0	0.0	100.0	0.0	82.2	37.4	100.0	0.0
	300	0.0	0.0	300.0	0.0	300.0	0.0	298.7	1.9	300.0	0.0
	500	0.0	0.0	500.0	0.0	500.0	0.0	495.9	3.1	500.0	0.0

contamination levels, and distributions. Second, we see that the BACON, FAST-MCD, and SDE-NTM detectors successfully identify all outliers in the cases tested. This finding is true for both the Gaussian data and the multivariate *t*-data, which is confirms our previous finding that robust estimation of the mean vector and covariance matrix can improve detection accuracy relative to the classical MSD detector, even if the Gaussian assumption is not valid. The ability of these detectors to successfully find outliers in heavy-tailed distributions is important since it provides an alternative to the challenging task of correctly identifying a specific distribution from the multivariate *t*-distribution family. A third observation from Tables 8 and 9 is the inability of the Juan-Prieto detector to find outliers when the contamination level is relatively low. The likely cause of this limitation is that relatively few outliers are not likely to affect the uniformity of the data when projected onto the unit hypersphere.

Turning to the false positive data reported in Tables 10 and 11, it is seen that when all detectors are applied to the Gaussian data, the number of false positives is close to zero for all levels of contamination and material combinations. The reason for this seemingly ideal false positive rate is the use of a Bonferoni significance level of α/n used to threshold the respective test statistics for the different detectors, where α=0.05 and n is the total number of observations. For all cases tested, the expected number of false alarms for the significance level used is less then one.

In the case of the multivariate-*t* data, the false alarm data is somewhat more interesting. First, we note that the false alarms for the BACON detector remain close to zero. In contrast, the false alarms for the FAST-MCD and SDE-NTM methods are

Table 10. False Positives for Multivariate Outlier Detector Comparisons (Multivariate Gaussian Data)

Background/ Outlier	Outliers Present	False Positives by Method									
		Classical		BACON		F.MCD		Juan-Prieto		SDE-NTM	
		Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Grass/ Road (A.P. Hill)	0	0.0	0.0	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.2
	50	0.1	0.3	0.0	0.0	0.1	0.4	0.0	0.0	0.1	0.3
	100	0.0	0.0	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.2
	300	0.0	0.2	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.0
	500	0.2	0.5	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.0
Grass/ Shadow (A.P. Hill)	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2
	50	0.0	0.0	0.0	0.0	0.2	0.4	0.0	0.0	0.1	0.3
	100	0.0	0.2	0.0	0.0	0.1	0.3	0.0	0.0	0.1	0.3
	300	0.0	0.0	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.0
	500	0.1	0.3	0.0	0.0	0.0	0.2	0.7	1.9	0.0	0.0
Dead Grass/ Shadow (A.P. Hill)	0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0
	50	0.0	0.0	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.2
	100	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2
	300	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	500	0.0	0.2	0.0	0.0	0.1	0.3	1.3	6.6	0.0	0.0
Road/ Shadow (A.P. Hill)	0	0.0	0.2	0.0	0.0	0.1	0.3	0.0	0.0	0.1	0.3
	50	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2
	100	0.0	0.0	0.0	0.0	0.2	0.5	0.0	0.0	0.1	0.3
	300	0.0	0.0	0.0	0.0	0.1	0.3	1.5	4.8	0.0	0.2
	500	0.1	0.3	0.0	0.0	0.2	0.6	8.8	25.9	0.0	0.2
Grass/ Asphalt (D.C. Mall)	0	0.1	0.3	0.0	0.0	0.2	0.4	0.0	0.0	0.1	0.3
	50	0.0	0.0	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.2
	100	0.0	0.0	0.0	0.0	0.1	0.3	2.6	5.7	0.0	0.0
	300	0.2	0.5	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.2
	500	0.2	0.6	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.0
Grass/ Water (D.C. Mall)	0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0
	50	0.0	0.2	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.0
	100	0.0	0.0	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.0
	300	0.3	0.5	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.0
	500	0.6	0.8	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0
Asphalt/ Water (D.C. Mall)	0	0.1	0.3	0.0	0.0	0.2	0.4	0.0	0.0	0.1	0.3
	50	0.0	0.0	0.0	0.0	0.1	0.4	0.0	0.0	0.0	0.0
	100	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2
	300	0.1	0.3	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2
	500	0.1	0.3	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.0
Gravel/ Asphalt (D.C. Mall)	0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2
	50	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2
	100	0.1	0.3	0.0	0.0	0.1	0.3	0.0	0.2	0.1	0.3
	300	0.2	0.4	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.0
	500	0.2	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 11. False Positive for Multivariate Outlier Detector Comparison (Multivariate-*t* Data)

Background/ Outlier	Outliers Present	False Positives by Method									
		Classical		BACON		F.MCD		Juan-Prieto		SDE-NTM	
		Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Grass/ Road (A.P. Hill)	0	14.8	3.5	0.0	0.0	39.4	6.5	0.0	0.0	27.9	4.6
	50	10.6	2.8	0.0	0.0	38.7	7.3	0.0	0.0	25.2	5.3
	100	10.9	3.5	0.1	0.3	37.1	6.7	1.0	3.1	23.9	6.2
	300	11.1	3.0	0.0	0.2	30.4	5.9	0.0	0.0	14.6	3.4
	500	14.1	3.3	0.0	0.0	28.5	5.3	0.0	0.0	11.6	3.9
Grass/ Shadow (A.P. Hill)	0	14.6	3.4	0.0	0.2	40.5	7.2	0.0	0.0	26.8	5.1
	50	10.0	2.9	0.0	0.2	39.8	7.1	0.0	0.0	26.3	5.5
	100	10.6	2.8	0.0	0.0	38.8	6.2	0.9	4.6	24.5	4.8
	300	10.5	2.9	0.1	0.3	30.1	6.3	0.0	0.0	15.8	4.5
	500	10.1	3.5	0.0	0.0	26.5	6.1	0.0	0.0	10.0	3.9
Dead Grass/ Shadow (A.P. Hill)	0	14.4	3.4	0.0	0.0	41.3	8.0	0.0	0.0	28.0	6.3
	50	9.9	3.1	0.0	0.0	38.9	6.9	0.0	0.0	25.8	5.4
	100	10.0	3.2	0.0	0.0	36.5	6.6	0.1	0.4	23.2	6.0
	300	10.0	3.2	0.0	0.2	31.2	6.0	0.0	0.0	16.3	3.8
	500	11.1	3.4	0.0	0.2	26.0	7.1	0.0	0.0	10.8	3.7
Road/ Shadow (A.P. Hill)	0	14.2	3.1	0.0	0.2	41.6	6.0	0.0	0.0	27.9	5.6
	50	6.1	1.9	0.0	0.2	38.8	6.4	0.0	0.0	25.6	5.6
	100	5.5	2.0	0.0	0.0	40.5	5.4	0.0	0.0	25.0	5.0
	300	4.3	1.7	0.0	0.0	31.9	6.3	2.8	7.4	16.5	4.4
	500	3.8	2.5	0.0	0.0	27.6	5.9	21.6	62.6	11.4	3.3
Grass/ Asphalt (D.C. Mall)	0	13.4	2.8	0.1	0.3	39.7	6.9	0.0	0.0	27.3	4.5
	50	11.9	2.9	0.0	0.0	39.4	6.2	0.0	0.0	25.8	4.5
	100	11.0	2.8	0.0	0.0	36.1	6.5	1.1	3.2	23.0	4.8
	300	16.0	3.6	0.0	0.2	33.1	4.6	0.0	0.0	17.0	4.0
	500	19.5	3.4	0.0	0.2	26.1	4.3	0.0	0.0	10.5	3.8
Grass/ Water (D.C. Mall)	0	14.2	3.1	0.0	0.0	37.2	6.0	0.0	0.0	25.7	4.6
	50	10.7	2.9	0.0	0.0	36.6	6.1	0.0	0.0	23.7	4.3
	100	12.9	3.0	0.0	0.0	35.4	5.5	0.0	0.0	22.8	4.8
	300	17.4	2.3	0.0	0.0	31.0	5.3	0.0	0.0	15.9	3.1
	500	25.2	5.2	0.0	0.0	27.3	6.8	0.0	0.0	10.3	4.2
Asphalt/ Water (D.C. Mall)	0	14.0	3.3	0.0	0.2	40.3	5.6	0.0	0.0	26.9	4.8
	50	11.4	2.7	0.0	0.2	38.2	6.2	0.0	0.0	25.6	4.6
	100	13.0	3.1	0.0	0.0	35.5	5.9	0.1	0.5	23.5	4.5
	300	17.5	3.6	0.0	0.2	33.2	4.9	0.0	0.0	17.8	3.8
	500	20.1	4.4	0.1	0.3	26.0	5.1	0.0	0.0	11.3	3.9
Gravel/ Asphalt (D.C. Mall)	0	14.5	3.3	0.1	0.3	39.3	5.6	0.0	0.0	27.2	4.8
	50	10.7	3.0	0.0	0.0	38.3	6.1	0.0	0.0	24.7	5.4
	100	11.8	3.3	0.0	0.0	34.6	6.4	0.1	0.6	22.2	6.0
	300	16.1	3.2	0.0	0.2	32.1	6.8	0.0	0.0	16.5	4.4
	500	21.0	3.5	0.0	0.2	26.7	5.1	0.0	0.0	10.4	3.5

significantly higher. The reason for this difference is that both the FAST-MCD and SDE-NTM methods arrive at a covariance estimate by first estimating the shape matrix of the data by trimming away all the observations far from the center. Because good observations may also be trimmed, the shape matrix underestimates the true variance of the good data. Hence, either the shape matrix or the Mahalanobis distances derived from it must be scaled before testing the distances for outliers. We have found that the scaling process used for both the FAST-MCD and SDE-NTM methods still tend to underestimate the true variance in the data, particularly when compared to the scaling method used by the BACON detector. Hence, it is reasonable to expect more false alarms with the FAST-MCD and SDE-NTM methods relative to the BACON detector, particularly with heavy-tailed data.

A final observation of note in the false alarm data is the decreasing number of false alarms for the FAST-MCD and SDE-NTM detectors as the level of contamination increases. We hypothesize that this phenomenon occurs for the following reason: as the contamination level increases, it is more likely that outliers are still contained in the set of observations used to estimate the shape matrix, since neither the FAST-MCD nor the SDE-NTM methods are guaranteed to generate a “clean” estimate. Though few in number, these outliers are sufficient to artificially increase the variance of the data. This increased variance will, in turn, result in a lower false alarm rate. A similar affect was demonstrated in Smetek and Bauer (2006).

The results given in Tables 8 through 11, lead to the conclusion that the BACON algorithm is the most effective outlier detection method of the four algorithms tested. Though the FAST-MCD, and SDE-NTM methods performed as well as BACON in terms

of true positive rate, BACON is more resistant to false alarms, particularly when heavy-tailed data is used. Another advantage of BACON relative to the other methods, is its computational speed. Though rigorous, controlled time trials were not performed, our experience with the algorithms in this experiment and with much larger datasets revealed BACON to be at least one to two orders of magnitude faster than the other methods. Based on these results, it would seem that the BACON algorithm is the most logical for use with hyperspectral data. In the following sections, we focus on adapting the BACON algorithm for this purpose, but for comparison sake, we adapt the FAST-MCD detector, as well.

Image Clustering

One of the fundamental assumptions of the BACON and FAST-MCD algorithms—and many other multivariate outlier detection methods—is the majority of the dataset comes from a single population with the remainder of the data belonging to one or more contaminating populations. Thus, these methods attempt to robustly estimate the mean vector and covariance matrix of the *good* population so that they can be used to identify the contaminating observations. If these methods are applied directly to a hyperspectral image, this single-population assumption is generally not valid and will lead to poor detection results. For example, a typical hyperspectral scene may consist of large areas of forest, open fields, road, urban areas, water, etc., none of which may be considered as outlying. Rather, they are simply different background constituents for the scene. If BACON or FAST-MCD are applied to such a scene, however, these methods will likely consider the most prevalent background material to be the good observations while discarding everything else as outliers.

To allow methods such as BACON and FAST-MCD to operate effectively on multiple background images, the image must first be clustered into homogeneous groups to which BACON or FAST-MCD are applied individually. If the number of clusters is set to coincide with the number of background materials in the scene, then any anomalous objects whose spectra occur relatively infrequently in the scene will be grouped with the most similar background material and should be amenable to detection by the BACON or FAST-MCD algorithms. Clustering a dataset prior to applying a multivariate outlier detection method is not a new idea. Woodruff and Reiners (2004) and Hardin and Rocke (2004) provide notable attempts to use cluster analysis for finding outliers in datasets with multiple good populations. However, there are two primary disadvantages with the methods they propose. First, they make use of robust cluster analysis, which adds considerable computational complexity to the clustering problem—relative to the ubiquitous *k*-means algorithm—in an attempt to minimize the effect of outliers on the clustering solution. A second disadvantage is the requirement to specify the number of groups as an input parameter to the clustering algorithm. This requirement makes the overall detection process less autonomous, which conflicts with the objectives of our research.

In the following sections, we develop a basic *k*-means-based clustering method for the BACON and FAST-MCD methods that can reasonably group a hyperspectral image into its major background materials prior to outlier detection. In so doing, we first explore the robustness of the *k*-means algorithm and argue that we can forego the computational complexity of robust clustering in favor of *k*-means without fear of inaccurate solutions caused by outlying observations. We then investigate methods for

automatically determining the number of background materials, *k*, in an image so that *k*-means can be employed in autonomous manner with minimal input from the user.

Robustness of the k-Means Algorithm

Our desire to employ the *k*-means algorithm as a preprocessor for the BACON and FAST-MCD detectors is its wide-spread use in the hyperspectral analysis community. The algorithm is relatively easy to use, it has been implemented in a number of image and statistical analysis software applications, and it can operate on large datasets in a reasonable amount of time. However, it would seem somewhat contradictory to argue the virtues of robust estimation elsewhere in this dissertation while blindly using the *k*-means algorithm without first ensuring it can accurately cluster a hyperspectral image when the image may be contain a number of outlying spectra. The issue of *k*-means robustness is thoroughly discussed by Garcia-Escudero and Gordaliza (1999) who show that the break-down point of the *k*-means algorithm is the worst it can be, $1/n$, where n is the number of observations in the dataset. In other words, it can take a single, well-positioned outlier to cause *k*-means to fail. The primary failure mode of the *k*-means algorithm, which we refer to as the *clumping effect*, is the assignment of the outlying observation to its own cluster while merging two other clusters into a single cluster.

This poor robustness property of *k*-means does not seem to favor this algorithm for clustering hyperspectral data that may contain anomalies. However, the additional computational complexity posed by robust clustering methods is not attractive either. Thus, we conduct the following experiment to determine the magnitude of the non-robustness problem in the context of hyperspectral image data:

1) For each of four background materials, generate 1000 observations for a multivariate Gaussian distribution using mean vectors and covariance matrices obtained from sample spectra taken from actual hyperspectral images.
2) Generate a specified number of outlier observations from a multivariate Gaussian distribution using the mean vector and covariance matrix obtained from sample spectra taken from an actual hyperspectral image. Combine these outlying observations with the 4000 background observations.
3) Apply the *k*-means algorithm to the contaminated dataset with *k* set to four, the actual number of background materials. Use a Cosine assignment rule in which each observation is assigned to the cluster whose mean vector forms the smallest angle with the observation vector.
4) Compute the accuracy of the clustering solution as the percent of background observations assigned to the correct group. We were not concerned with how the outliers were assigned.
5) Repeat Steps 1 through 4 50 times to obtain a mean estimate for the classification accuracy. Also compute the standard error for the mean estimate.
6) Repeat Steps 1 through 5 with an increased number of outliers. The number of outliers ranged from 0 to 1000 outliers in increments of 50 outliers.
7) Repeat Steps 1 through 6 after replacing the Cosine assignment rule in Step 3 with the Squared Euclidean rule by which each observation is assigned to the cluster to which it is closest in terms of squared Euclidean distance.

8) Repeat Steps 1 through 7 using a different set of four background materials in Step 1 derived from a different hyperspectral image. For this experiment, spectra were derived from the Fort A.P. Hill and D.C. Mall images used in previous experiments, as well as a 128-band HYMAP sensor image of the campus and surrounding community of Purdue University. The mean spectra and band standard deviations for the materials obtained from the Purdue image are shown in Figure 20. Table 12 lists the background and outlier materials used from each image.

9) Repeat Steps 1 through 8 using a multivariate-t distribution with twelve degrees of freedom to generate the background and outlier observations in Steps 1 and 2.

Before reviewing the results of this experiment, several clarifications need to be made. First, in order to avoid local clustering solutions, Step 3 of the experiment actually consisted of 30 applications of the *k*-means algorithm to the contaminated dataset, with each application using a different random starting point for the four cluster means. Second, each application of *k*-means was allowed to progress through 100 iterations to converge to the final solution. Third, the Cosine metric was used in Step 3 because inspection of actual hyperspectral image clustering solutions revealed that it was more effective in grouping observations with similar spectral shape, whereas the Squared Euclidean clusters for the same image tended to contain less-homogenous spectra. To see why this is so, consider a very simple two-cluster problem in three dimensions. Suppose that at an arbitrary iteration of the *k*-means algorithm the mean vector of cluster one is

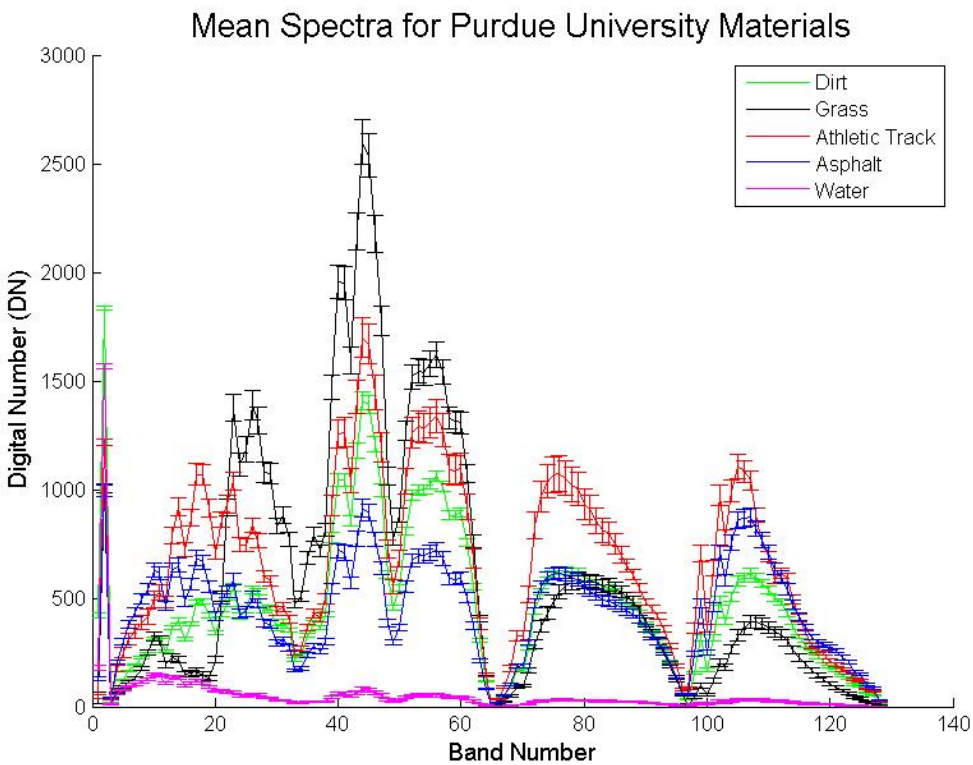


Figure 20. Mean Spectra for Purdue University Image Materials

$\boldsymbol{\mu}_1=(2,6,1)^T$ and the mean vector of cluster two is $\boldsymbol{\mu}_2=(2,1,1)^T$. Further, suppose we are attempting to assign a vector, $\mathbf{x}=(1,3,0.5)^T$, to the *closest* cluster. Upon inspection, it is clear the $\mathbf{x}$ is simply a mean shift of $\boldsymbol{\mu}_1$ which is a realistic scenario in hyperspectral imagery due to the effects of illumination on pixels containing the same materials. Under the Cosine rule, we obtain the desired outcome of assigning $\mathbf{x}$ to cluster one since the angle between $\mathbf{x}$ and $\boldsymbol{\mu}_1$ is zero. However, even though the spectral shape of $\mathbf{x}$ is much closer to $\boldsymbol{\mu}_1$ than $\boldsymbol{\mu}_2$, the Squared Euclidean rule dictates $\mathbf{x}$ be assigned to cluster two since the squared Euclidean distance between $\mathbf{x}$ and $\boldsymbol{\mu}_2$ is 5.25 as opposed 9.0 between $\mathbf{x}$ and $\boldsymbol{\mu}_1$.

A fourth clarification for the robustness experiment is that an excursion was also

Table 12. Materials used for *k*-Means Robustness Tests

Fort A.P. Hill	D.C. Mall	Purdue University
Grass	Grass	Dirt
Road	Asphalt	Grass
Dead Grass	Gravel	Track
Tree	Roof	Asphalt
Shadow (Outlier)	Water (Outlier)	Water (Outlier)

performed in which the contaminated dataset was normalized prior to applying the *k*-means algorithm in Step 3. By *normalize* we mean that each observation was divided by its respective vector norm. This preprocessing step has the effect of reducing the variability of spectral signatures corresponding to the same material and has been used in the hyperspectral literature to help minimize the effects of illumination (Healey and Slater, 1999). This normalization step is only used with the Squared Euclidean assignment rule, since normalizing the data prior to using the Cosine rule has no mathematical effect—the Cosine rule inherently normalizes the data.

A final clarification for the experiment is the that the hyperspectral data used to produce the mean vectors and covariance matrices in Steps 1 and 2 was reduced in dimensionality to p=15 prior to running the experiment. This reduction was performed by dividing the original image bands into 15 sequential blocks and using the block means as the new variables. This type of data reduction preserves the general shape of the original spectral signatures, while significantly reducing the computation time required for the *k*-means algorithm. In other tests not reported here, principal components analysis was also used as a data reduction technique; however, based on visual inspection, the PCA clusters appeared less-homogeneous than the band-aggregation clusters.

The results of the *k*-means robustness experiment are summarized in Tables 13 and 14 for the multivariate Gaussian and multivariate-*t* generated data, respectively. The first column of these tables specifies which *k*-means method was used for the experiment. The naming convention for the *k*-means method contains three parts. The first letter denotes the assignment rule used by *k*-means, where 'C' indicates the Cosine rule and 'S' indicates the Squared Euclidean rule. The second letter of the method designator specifies if normalized (N) or un-normalized (U) data was used. The third letter indicates if the outliers were generated from a single distribution (S), or if they were randomly selected from a set of multiple outlier materials (R)—this outlier generation method will be discussed in more detail momentarily. The remaining columns of Tables 13 and 14 give the mean accuracy of the *k*-means algorithm under different levels of contamination for the three underlying images used to generate the simulated data. The standard error for each mean estimate is also provided.

The first conclusion we draw from Table 13 is that the robustness of the *k*-means algorithm is dependent on the background materials that are being clustered. When the outliers come from a single population, the highest level of contamination achieved with 100% classification accuracy using the Fort A.P. Hill data is approximately 150 outliers out of 4150 total observations, or 3.6% (Methods C/U/S and S/N/S). For the D.C. Mall data, 100% accuracy can be obtained with contamination levels of 20% using the S/U/S method. In the case of the Purdue data, only a 1.2% contamination level can be tolerated before the classification accuracy drops below 100% (Method S/U/S). From Table 14 we

Table 13. *k*-Means Robustness Test Results (Gaussian Data)

Method	Outliers	A.P Hill Data		D.C. Mall Data		Purdue Data	
		Accuracy	S.E.	Accuracy	S.E.	Accuracy	S.E.
C/U/S	**0**	1.00	0.00	1.00	0.01	1.00	0.00
	50	1.00	0.01	1.00	0.01	0.75	0.06
	100	1.00	0.00	1.00	0.01	0.75	0.06
	150	1.00	0.00	1.00	0.01	0.75	0.06
	200	0.93	0.04	0.99	0.01	0.75	0.06
	300	0.75	0.06	0.99	0.01	0.75	0.06
	400	0.75	0.06	0.98	0.02	0.75	0.06
	500	0.75	0.06	0.85	0.05	0.75	0.06
	600	0.75	0.06	0.77	0.06	0.75	0.06
	1000	0.75	0.06	0.77	0.06	0.75	0.06
S/U/S	**0**	0.92	0.04	1.00	0.00	1.00	0.00
	50	0.93	0.04	1.00	0.00	1.00	0.00
	100	0.93	0.04	1.00	0.00	0.75	0.06
	150	0.92	0.04	1.00	0.00	0.75	0.06
	200	0.93	0.04	1.00	0.00	0.75	0.06
	300	0.93	0.04	1.00	0.00	0.75	0.06
	400	0.92	0.04	1.00	0.00	0.75	0.06
	500	0.73	0.06	1.00	0.00	0.75	0.06
	600	0.72	0.06	1.00	0.00	0.75	0.06
	1000	0.72	0.06	1.00	0.00	0.75	0.06
S/N/S	**0**	1.00	0.00	1.00	0.01	1.00	0.00
	50	1.00	0.00	1.00	0.01	0.75	0.06
	100	1.00	0.00	1.00	0.01	0.75	0.06
	150	1.00	0.00	1.00	0.01	0.75	0.06
	200	0.95	0.03	0.99	0.01	0.75	0.06
	300	0.75	0.06	0.99	0.01	0.75	0.06
	400	0.75	0.06	0.98	0.02	0.75	0.06
	500	0.75	0.06	0.86	0.05	0.75	0.06
	600	0.75	0.06	0.77	0.06	0.75	0.06
	1000	0.75	0.06	0.77	0.06	0.75	0.06
C/U/R	**0**	1.00	0.00	1.00	0.01	1.00	0.00
	50	1.00	0.00	1.00	0.01	1.00	0.00
	100	1.00	0.00	1.00	0.01	1.00	0.00
	150	1.00	0.00	1.00	0.01	1.00	0.00
	200	1.00	0.00	1.00	0.01	1.00	0.00
	300	1.00	0.00	1.00	0.01	1.00	0.00
	400	1.00	0.00	1.00	0.01	1.00	0.00
	500	0.86	0.05	0.99	0.01	1.00	0.00
	600	0.75	0.06	0.99	0.01	1.00	0.00
	1000	0.75	0.06	0.99	0.01	1.00	0.00
S/U/R	**0**	0.93	0.04	1.00	0.00	1.00	0.00
	50	0.93	0.04	1.00	0.00	1.00	0.00
	100	0.93	0.04	1.00	0.00	1.00	0.00
	150	0.93	0.04	1.00	0.00	1.00	0.00
	200	0.93	0.04	1.00	0.00	1.00	0.00
	300	0.93	0.04	1.00	0.00	1.00	0.00
	400	0.93	0.04	1.00	0.00	1.00	0.00
	500	0.93	0.04	1.00	0.00	0.99	0.01
	600	0.93	0.04	1.00	0.00	0.96	0.03
	1000	0.93	0.04	1.00	0.00	0.75	0.06
S/N/R	**0**	1.00	0.00	1.00	0.01	1.00	0.00
	50	1.00	0.00	1.00	0.01	1.00	0.00
	100	1.00	0.01	1.00	0.01	1.00	0.00
	150	1.00	0.00	1.00	0.01	1.00	0.00
	200	1.00	0.00	1.00	0.01	1.00	0.00
	300	1.00	0.00	1.00	0.01	1.00	0.00
	400	1.00	0.00	1.00	0.01	1.00	0.00
	500	0.84	0.05	0.99	0.01	1.00	0.00
	600	0.75	0.06	0.99	0.01	1.00	0.00
	1000	0.75	0.06	0.99	0.01	1.00	0.00

Table 14. *k*-Means Robustness Test Results (Multivariate-*t* Data)

Method	Outliers	A.P Hill Data		D.C. Mall Data		Purdue Data	
		Accuracy	S.E.	Accuracy	S.E.	Accuracy	S.E.
C/U/S	**0**	1.00	0.01	0.99	0.01	1.00	0.00
	50	1.00	0.01	0.99	0.01	0.75	0.06
	100	1.00	0.01	0.99	0.01	0.75	0.06
	150	1.00	0.01	0.99	0.01	0.75	0.06
	200	0.97	0.02	0.99	0.01	0.75	0.06
	300	0.75	0.06	0.99	0.02	0.75	0.06
	400	0.75	0.06	0.98	0.02	0.75	0.06
	500	0.75	0.06	0.80	0.06	0.75	0.06
	600	0.75	0.06	0.77	0.06	0.75	0.06
	1000	0.75	0.06	0.76	0.06	0.75	0.06
S/U/S	**0**	0.93	0.04	1.00	0.01	1.00	0.00
	50	0.93	0.04	1.00	0.01	1.00	0.00
	100	0.93	0.04	1.00	0.00	0.77	0.06
	150	0.93	0.04	1.00	0.01	0.75	0.06
	200	0.93	0.04	1.00	0.00	0.75	0.06
	300	0.93	0.04	1.00	0.00	0.75	0.06
	400	0.93	0.04	1.00	0.00	0.75	0.06
	500	0.74	0.06	1.00	0.01	0.75	0.06
	600	0.72	0.06	1.00	0.01	0.75	0.06
	1000	0.72	0.06	1.00	0.01	0.75	0.06
S/N/S	**0**	1.00	0.01	0.99	0.01	1.00	0.00
	50	1.00	0.01	0.99	0.01	0.75	0.06
	100	1.00	0.01	0.99	0.01	0.75	0.06
	150	1.00	0.01	0.99	0.01	0.75	0.06
	200	0.97	0.02	0.99	0.01	0.75	0.06
	300	0.75	0.06	0.99	0.02	0.75	0.06
	400	0.75	0.06	0.98	0.02	0.75	0.06
	500	0.75	0.06	0.80	0.06	0.75	0.06
	600	0.75	0.06	0.77	0.06	0.75	0.06
	1000	0.75	0.06	0.76	0.06	0.75	0.06
C/U/R	**0**	1.00	0.01	0.99	0.01	1.00	0.00
	50	1.00	0.01	0.99	0.01	1.00	0.00
	100	1.00	0.01	0.99	0.01	1.00	0.00
	150	1.00	0.01	0.99	0.01	1.00	0.00
	200	1.00	0.01	0.99	0.01	1.00	0.00
	300	1.00	0.01	0.99	0.01	1.00	0.00
	400	1.00	0.01	0.99	0.01	1.00	0.00
	500	0.85	0.05	0.99	0.01	1.00	0.00
	600	0.75	0.06	0.99	0.01	1.00	0.00
	1000	0.75	0.06	0.99	0.02	1.00	0.00
S/U/R	**0**	0.93	0.04	1.00	0.00	1.00	0.00
	50	0.93	0.04	1.00	0.01	1.00	0.00
	100	0.93	0.04	1.00	0.00	1.00	0.00
	150	0.93	0.04	1.00	0.00	1.00	0.00
	200	0.93	0.04	1.00	0.00	1.00	0.00
	300	0.93	0.04	1.00	0.00	1.00	0.00
	400	0.93	0.04	1.00	0.00	1.00	0.00
	500	0.93	0.04	1.00	0.00	0.99	0.01
	600	0.93	0.04	1.00	0.00	0.96	0.03
	1000	0.93	0.04	1.00	0.00	0.75	0.06
S/N/R	**0**	1.00	0.01	0.99	0.01	1.00	0.00
	50	1.00	0.01	0.99	0.01	1.00	0.00
	100	1.00	0.01	0.99	0.01	1.00	0.00
	150	1.00	0.01	0.99	0.01	1.00	0.00
	200	1.00	0.01	0.99	0.01	1.00	0.00
	300	1.00	0.01	0.99	0.01	1.00	0.00
	400	1.00	0.01	0.99	0.01	1.00	0.00
	500	0.85	0.05	0.99	0.01	1.00	0.00
	600	0.75	0.06	0.99	0.01	1.00	0.00
	1000	0.75	0.06	0.99	0.02	1.00	0.00

see that similar results are obtained when the data is generated using a heavier-tailed distribution. If we cast these contamination levels in the light of trying to detect 3m x 10m vehicle targets in a 150000-pixel image with one-meter spatial resolution, they correspond to 180, 1000, and 60 vehicles in a 0.15-km^2 area (approximately 37 acres) for the Fort A.P. Hill, D.C. Mall, and Purdue datasets, respectively, assuming the results in Tables 13 and 14 scale to the 150000-observation problem. Thus, depending on the number of single-population anomalies that are expected to exist in an image, the non-robust *k*-means algorithm may suffice as a pre-processor to the BACON or FAST-MCD algorithms.

In many cases, the assumption that all outliers come from a single population may not be realistic. For example, in a vehicle-detection problem, the vehicles may be painted in different schemes or be made from a number of different materials that have unique spectral signatures. To determine if such conditions affect the robustness of the *k*-means algorithm, we repeated the robustness experiment with a modification to Step 2. Specifically, for each image we collected an equal number of five different anomaly spectra. The materials and quantities collected from each image are listed in Table 15, and the mean spectra for these materials are given in Appendix A. In generating the specified number of outliers at Step 2, we randomly select spectra from the appropriate outlier set with replacement until the desired number of outliers is reached. This procedure effectively disperses the outliers in five general locations in the 15-dimensional space as opposed to only one location in the original experiment. The results obtained after this modification are indicated by the ·/·/R methods in Tables 13 and 14.

Table 15. Materials used for Dispersed Outliers

Image	Material	Number of Spectra Used
Fort A.P. Hill	Building Roof	10
	Target 1	10
	Target 2	10
	Shadow	10
	Dead Grass 2	10
D.C. Mall	Water	20
	Trees	20
	Marble Walk	20
	Museum Dome	20
	Shadow	20
Purdue University	Arena Dome	20
	Roof 1	20
	Roof 2	20
	Baseball Diamond	20
	Dirt 2	20

When the outliers are more widely dispersed in the high-dimensional space, we see that 100% classification accuracy can be achieved by the *k*-means algorithm at contamination levels of 9.1%, 20%, and 20% for the Fort A.P. Hill, D.C. Mall, and Purdue data, respectively. These results are achieved regardless if Gaussian or multivariate-*t* distributions are used to generate the data. Based on these results, it is evident that the *k*-means algorithm is reasonably robust to the presence of outliers in a practical hyperspectral analysis setting, and can be reasonably expected to cluster a hyperspectral image into its constituent background materials without significant modification to the algorithm. It should be cautioned, however, that our tests are somewhat limited in scope, particularly with respect to the number of observations in each background cluster. Our tests employed equal numbers of observations in each cluster, whereas unequal sized clusters may lead to different results; however, our

practical experience applying the *k*-means algorithm to actual images with unequal sized clusters has not revealed any alarming problems.

Thus far, we have concluded that the *k*-means algorithm is reasonably robust to the presence of outliers, but we have yet to address if normalizing the data serves any useful purpose, or whether it is better to use the Cosine rule or Squared Euclidean rule to ensure accurate clustering. To address these issues, we look more closely at the performance of the three basic *k*-means configurations tested: the Cosine rule applied to non-normalized data; the Squared Euclidean rule applied to non-normalized data; and the Squared Euclidean rule applied to normalized data. Regardless of whether multivariate Gaussian or multivariate-*t* data is used for the test, the Cosine rule only achieves 100% accuracy up to 3.6% contamination under the single-population outliers assumption for both the Fort A.P. Hill and D.C. Mall data. Though 3.6% contamination may seem low, as noted earlier, it is equivalent to approximately 180 vehicles in a 0.15-km^2 area which may be sufficient for most anomaly detection applications. More troubling, however, is the Cosine rule's apparent non-robustness when applied to the Purdue data. In this instance, less than 1.2% contamination resulted in two of the background materials being grouped in one cluster and the outliers being placed in their own cluster, producing only 75% classification accuracy. This catastrophic failure of *k*-means is, in fact, the clumping effect mentioned previously, which, if left unchecked, would lead to catastrophic failure of anomaly detection efforts with the BACON or FAST-MCD detectors. Countering this deficiency, however, is the Cosine rule's significantly improved performance when dispersed outliers are used. Under this outlier assumption, the Cosine rule is able to

tolerate at least 9.1% contamination across the three datasets, and 20% contamination for the Purdue data.

When the Squared Euclidean rule is applied to non-normalized data, it is seen that under the assumption of single-population outliers, 100% accuracy is never achieved for the Fort A.P. Hill data—even with 0.0% contamination. We also see, however, that the Squared Euclidean rule is more robust than the Cosine rule when applied to the D.C. Mall and Purdue datasets, though its robustness to the Purdue data is somewhat marginal. When dispersed outliers are used, we see that the Squared Euclidean rule is still unable to accurately classify the Fort A.P. Hill data, while its performance against the Purdue data improves considerably. Based on these results, the question arises as to why the Squared Euclidean rule is unable to correctly classify the Fort A.P. Hill data, regardless of the contamination level? Inspection of the classification results in these cases reveals that the misclassifications are between the Grass and Tree observations. Referring back to Figure 12, we see that these materials are spectrally similar, and even overlap in some bands, thereby leading to the possibility that some Grass spectra may be closer to the mean spectra of the tree material in terms of squared Euclidean distance, and vice versa. However, despite the fact that these materials are close enough together in Euclidean space to foil the Squared Euclidean rule, there is sufficient difference in the shape of the signatures for the Cosine rule to separate the two. This result supports our underlying rationale for considering the Cosine rule for use with hyperspectral data.

We now look at the Squared Euclidean rule applied to normalized data. From Tables 13 and 14 it is evident that this configuration achieves virtually the same robustness as the Cosine rule, regardless of the outlier assumption or the distribution of

the data. For the Fort A.P. Hill and Purdue data under the dispersed outlier assumption, this change is an improvement upon the Squared Euclidean rule applied to non-normalized data. For the D.C. Mall data, this change translates to a slight decrease in robustness. This decrease, however, means we only obtain 99% mean classification accuracy at any tested level of contamination versus 100% mean accuracy at any tested level of contamination. When considering the standard error of the mean accuracy estimates, however, this difference is not statistically significant. Under the single-population outlier assumption, normalizing the data improves the robustness of the Squared Euclidean rule for the Fort A.P. Hill data, but decreases robustness for the other two datasets. In general, then, we conclude that normalization prior to application of the Squared Euclidean rule should be considered when outliers are likely to be scattered throughout the Euclidean space, but should be used with caution if outliers are suspected to be relatively concentrated in one region. We will have more to say about this latter recommendation momentarily.

Based on the preceding discussion, we conclude that the Cosine rule provides the best alternative as a robust clustering method for hyperspectral data. Though the rule is not the most robust method in all cases tested, it performs extremely well under the assumption of dispersed outliers, a condition we feel is more realistic in practical hyperspectral clustering applications. We also find attractive the Cosine rule's ability to better-separate similar materials, as revealed by the tests with the Fort A.P. Hill dataset. This strength of the Cosine rule is further supported by our practical experience clustering hyperspectral images in which we have found the Cosine rule more effective in forming clusters of spectra with similar shapes, as opposed to the Squared Euclidean rule

that tends to contaminate clusters with spectra that are close to the cluster mean in terms of Euclidean distance, but clearly have different spectral shapes.

Though we favor the Cosine rule as the method to cluster hyperspectral data in the presence of outliers, it is obviously troubling that the rule can potentially fail depending on the nature of the outliers, as occurred with the Purdue dataset (though the Squared Euclidean rule does not offer much relief with this data.) As stated previously, this failure manifested itself as two background materials being grouped into a single cluster and the outliers being assigned to their own cluster—the clumping effect. Rather than blindly use the Cosine rule and hope that the clumping effect does not materialize, we now turn our attention to better understanding when this phenomenon is likely to occur. To begin, suppose we have a dataset containing k clusters and that the data is contaminated by n_0 outlying observations from a single population. Further, let C_1 and C_2 be the two clusters closest to the outlying observations with C_2 being the closest. We are interested in knowing something about the number, n_0, that will cause k-means to group the observations in C_1 and C_2 into one cluster and place the outlying observations in their own cluster.

To gain insight into values of n_0 that will lead to the clumping effect, we note that the objective of k-means, as implemented in our experiments, is to form clusters that minimize the total sum of Euclidean distances between all observations and their respective cluster mean vectors. That is, k-means attempts to form clusters that minimize

$$D = \sum_{i=1}^{k} D_i \tag{5.35}$$

where D_i is the sum of distances between the observations in cluster i and the mean vector of cluster i. For clumping to occur, it must be the case that $D>D_c$, where D_c is the

sum of distances when C_1 and C_2 are grouped together and the outliers are in their own cluster. In other words, the following must be true:

$$D = \sum_{i=1}^{k} D_i > D_0 + D_{12} + \sum_{i=3}^{k} D_i = D_c \tag{5.36}$$

where D_0 is the sum of distances for the outlier cluster, and D_{12} is the sum of distances for the single cluster containing the observations in C_1 and C_2. We can simplify (5.36) to

$$D_1 + D_2 > D_0 + D_{12} \tag{5.37}$$

We now expand (5.37) to reflect the number of observations in each cluster as well as the mean distance between a cluster observation and its respective cluster mean. In so doing, (5.37) becomes

$$\left(n_1 d_1 + n_2 d_{20} + n_0 d_{20}\right) > \left(n_1 d_{12} + n_2 d_{12} + n_0 d_0\right) \tag{5.38}$$

where

n_1 = the number of observations in C_1,
n_2 = the number of observations in C_2,
d_1 = the mean distance between observations and cluster mean for the cluster containing only C_1 observations,
d_{20} = the mean distance between observations and cluster mean for the cluster containing C_2 observations and outlier observations,
d_{12} = the mean distance between observations and cluster mean for the cluster containing C_1 and C_2 observations, and
d_0 = the mean distance between observations and cluster mean for the cluster containing only outlier observations.

Rearranging terms in (5.38), we can obtain a threshold for n_0 above which the clumping effect will occur. Specifically, clumping can be expected if the following is true:

$$n_0 > \frac{n_1\left(d_{12} - d_1\right) + n_2\left(d_{12} - d_{20}\right)}{d_{20} - d_0}. \tag{5.39}$$

In terms of determining the threshold for n_0 prior to an actual cluster analysis, (5.39) is of limited value because little may be known about the mean distances required by the formula. However, (5.39) offers considerable insight to the nature of the clumping effect. In particular, (5.39) leads to the following observations:

i) Highly concentrated outlier whose underlying population has relatively small variance will cause the clumping effect at smaller values of n_0 than more dispersed outliers. This is true since lower variance will give lower values of d_0.

ii) Large separation between the outliers and the background data will cause the clumping effect at relatively low contamination levels since d_{20} will increase while d_0 remains the same. The increase in d_{20} increases the denominator of (5.38) while decreasing the numerator.

iii) Clusters that are widely separated (large d_{12}) are more robust to outliers since an increase in d_{12} causes an increase in the numerator of (5.39) while leaving the denominator unchanged.

iv) A more concentrated cluster C_1 (the second closest cluster to the outliers) will decrease d_1, thereby increasing the critical value of n_0.

v) As the size of C_1 and C_2 increases, the number of outliers required to induce the clumping effect will also increase.

Based on these observations, we can more confidently use the Cosine rule with *k*-means if the following conditions hold true for a hyperspectral dataset: 1) the relative size

of the background clusters are significantly larger than the number of anomalies; 2) the anomalies are spectrally similar to at least one of the background materials; 3) the background materials are well-separated; and 4) the anomalies are not highly concentrated in Euclidean space. In many anomaly detection studies, particularly when military targets are involved, we maintain that conditions 1 and 2 are usually met. In cases were condition 3 may not be satisfied, the extremely large cluster sizes—typically on the order of several thousand spectra—dominates low separability between clusters. Similarly, large cluster sizes and the similarity of anomalies to one or more background materials generally reduces the need to satisfy condition 4. In short, we do not feel the clumping effect is likely to occur in most real-world anomaly detection studies, based on the nature of actual hyperspectral imagery.

To summarize the results of this section, we have demonstrated through simulated data experiments that the *k*-means algorithm is sufficiently robust to the presence of outliers, and can be used to cluster a hyperspectral image as a precursor to using the BACON or FAST-MCD algorithms. Additionally, our experiments indicate that using the Cosine rule with the *k*-means algorithm can produce 100% classification accuracy at higher levels of contamination than the Squared Euclidean rule across the datasets tested when the outliers are dispersed in the Euclidean space. Hence, we are led to using the *k*-means algorithm with the Cosine rule as the preferred method for clustering a hyperspectral image as a preprocessing step prior to using the BACON or FAST-MCD algorithm. In the following section, we turn our attention to automatically choosing a value for the number of clusters in an image.

Automatic Selection of k

As stated previously, one of the fundamental objectives of this research is to develop an anomaly detection method that requires minimal user input. It is our assertion that a detection methodology is of little operational value if the intended user—who may have limited technical training in applied statistics—must "read the tea leaves" at different stages of the detection methodology in order to get meaningful results. As a step towards reaching this objective of autonomy, we now seek a suitable method for determining the number of clusters in a hyperspectral dataset in order to smoothly integrate the *k*-means algorithm with the BACON and FAST-MCD detectors. As indicated by Everitt, Landau, and Leese (2001), a large number of informal and formal methods have been proposed over the years to address this difficult problem of determining *k*. Over 20 years ago, Milligan and Cooper (1985) identified no fewer than 30 formal methods, and the number has steadily grown since then. It is not our intention to extend this line of research, but rather to test the more promising methods suggested by Everitt et al. to determine which methods are useful for hyperspectral data. In the following paragraphs the methods tested are outlined, the experiments used to test the methods on both simulated and actual images are defined, and the significant conclusions derived from the experiments are presented.

The methods for choosing *k* that we evaluated are five methods suggested by Everitt et al. (2001). These methods were originally proposed by Calinski and Harabasz (1974), Marriott (1982), Kaufman and Rousseeuw (1990), Beale (1969), and Duda and Hart (1973). In addition to these five statistically-based methods, we also developed and evaluated a simple method that sets *k* equal to the number of major colors in a true-color

image of the hyperspectral dataset. This so-called Color Method attempts to mimic the manual process of visually inspecting an image to assess the number of background materials it contains. Each of these methods are outlined in the following paragraphs.

Calinski-Harabasz Method. This method proceeds by generating clustering solutions for a range of values for *k*. For each solution, the following metric is computed:

$$C(k) = \left(\frac{\text{trace}(\mathbf{B})}{k-1} \right) \Big/ \left(\frac{\text{trace}(\mathbf{W})}{n-k} \right) \tag{5.40}$$

where

$\mathbf{B}$ = the between-cluster scatter matrix,
$\mathbf{W}$ = the within-cluster scatter matrix, and
n = the total number of observations.

The between-cluster scatter matrix, **B**, and the within-cluster scatter matrix, **W**, are computed as

$$\mathbf{B} = \sum_{i=1}^{k} n_i \left(\overline{\mathbf{x}}_i - \overline{\mathbf{x}} \right) \left(\overline{\mathbf{x}}_i - \overline{\mathbf{x}} \right)^T \tag{5.41}$$

and

$$\mathbf{W} = \sum_{i=1}^{n} \left(\mathbf{x}_i - \overline{\mathbf{x}}_{ki} \right) \left(\mathbf{x}_i - \overline{\mathbf{x}}_{ki} \right)^T \tag{5.42}$$

where

n_i = the number of observations in cluster *i*,
$\overline{\mathbf{x}}_i$ = the mean vector of cluster *i*,
$\overline{\mathbf{x}}$ = the grand mean vector of the dataset,
$\mathbf{x}_i$ = the *i*th observation vector, and
$\overline{\mathbf{x}}_{ki}$ = the mean vector of the cluster to which the *i*th observation is assigned.

The final value for k under this method is the one that produces the largest value for (5.40) over the range of k values tested.

Marriott Method. In a process similar to the Calinski-Harabasz method, this method forms clustering solutions for different values of k over a specified range. For each solution, the metric in (5.40) is replaced by the following:

$$M(k) = k^2 \det(\mathbf{W}). \tag{5.43}$$

The solution with the smallest value of $M(k)$ determines the final value of k.

Kaufman-Rousseeuw (Silhouette) Method. Whereas the Calinski-Harabasz and Marriott methods focus on the scatter matrices of candidate clustering solutions, the Silhouette Method attempts to find the value k that ensures an observation is more similar to the other observations in it cluster than to observations in the next closest cluster. To accomplish this task, the Silhouette method also forms cluster solutions for a range of k values, and then computes the average value of s_i over all observations in the dataset. The Silhouette metric, s_i, for observation $\mathbf{x}_i$ is defined as

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \tag{5.44}$$

where

b_i = the average distance from observation i to the observations in the cluster closest to observation i's cluster,

a_i = the average distance from observation i to the other observations in its own cluster.

The average value of s_i will be in the range [-1,1], with positive values indicating observations are closer to their own cluster than to observations in other clusters. The

Silhouette method determines the final value of k by finding the corresponding cluster solution that produces the largest average s_i.

Beale Method. Whereas the previous three methods search for the number of clusters that minimize or maximize a summary statistic, the Beale method employs a more formal statistical test to determine the best number of clusters. Specifically, the method starts with k_1 clusters and then computes the following statistic to determine if $k_2>k_1$ (usually $k_2=k_1+1$) clusters offers a better solution:

$$F\left(k_1,k_2\right)=\frac{\left(S_{k_1}^2-S_{k_2}^2\right)/S_{k_2}^2}{\left[\left(n-k_1\right)/\left(n-k_2\right)\right]\left(k_2/k_1\right)^{2/p}-1} \tag{5.45}$$

where

$S_{k_i}^2$ = the sum of squared deviations from each observation to their respective cluster centroid when k_i clusters are used, and

p = the dimensionality of the data.

The statistic in (5.45) is compared to a critical value from an F-distribution with $p(k_2-k_1)$ and $p(n-k_2)$ degrees of freedom. If the computed value exceeds the critical value, k_2 is taken as a better value for k and the process is repeated with $k_1=k_2$ and k_2 set to some larger value than the new k_1. The method terminates when the computed value of (5.45) does not exceed the critical value, at which point k is set to the current value of k_1.

Duda-Hart Method. This method is similar to the Beale Method, but instead of testing if k_2 clusters is better than k_1 clusters, it starts with k_1 clusters and determines if any of the clusters should be split into two clusters. Specifically, the method computes the following statistic for the mth cluster in the starting set of k_1 clusters:

$$L(m) = \left\{ 1 - \frac{J_2^2}{J_1^2} - \frac{2}{\pi p} \right\} \left\{ \frac{n_m p}{2\left[1 - 8/\left(\pi^2 p\right)\right]} \right\} \tag{5.46}$$

where

J_1^2 = the within-cluster sum of squared distances between the observations in cluster m with the centroid of cluster m,

J_2^2 = the sum of within-cluster sum of squared distances when the cluster is divided into two clusters, and

n_m = the number of observations in cluster m.

If $L(m)$ exceeds a critical value from a standard normal distribution, the mth cluster is split into two clusters. For any new clusters produces during the first pass through the original k_1 clusters, the test is repeated. This process continues until no new clusters are formed, and the final value of k is the number of clusters when the method terminates.

The Color Method. The methods discussed to this point all use some statistic derived from the clustered data to determine if the number of clusters used adequately account for the structure in the data. A limitation with these methods is they are computationally expensive, since they generally require the k-means algorithm to be run multiple times. For very large hyperspectral datasets in high-dimensional space, the number of computations required to obtain an estimate for k can take on the order of tens of minutes for the Calinski-Harabasz, Marriott, Beale, and Duda-Hart methods, or hours for the Silhouette Method. These computation times are generally not practical in an operational setting when an anomaly detection analysis should preferably be completed in minutes or faster. As an alternative to these methods, we propose a basic method that sets k equal to the number of basic colors that account for 95% of the pixels in a true-

color representation of a hyperspectral image. As stated previously, this method attempts to automate the manual process of visually inspecting an image to determine the number of background materials. We feel this is a reasonable approach to determining *k* since most background materials, such as healthy grass, dead grass, trees, concrete, soil, water, asphalt, etc., have relatively unique colors that are discernable in a true-color image. Though this method is not applicable to generic datasets in which color information is either not available or meaningless, we contend that it is worthy of investigation for hyperspectral data. If nothing else, the fact that this method can produce an estimate of *k* in only a fraction of a second for even the largest of images makes it worthy of consideration.

The process that the Color Method employs is as follows:

1) For each image pixel, the digital numbers in the red (650nm to 750nm), green (550nm to 650nm), and blue (450nm to 550nm) bands of the original image are averaged to produce an RGB-triplet for the pixel.
2) The RGB values are converted to the Hue-Saturation-Intensity (HSI) color space using conversion equations found in Gonazalez, Woods, and Eddins (2004).
3) Each pixel is assigned to one of 54 color bins, where each bin is a region of the HSI color space. The regions are created by dividing the Hue component into 6 equal regions, the Saturation component into three equally spaced regions, and the Intensity component into three equally spaced regions. The 6 Hue regions correspond approximately to six major colors of red, yellow, green, cyan, blue, and magenta.

4) The number of bins that contain 95% of the pixels is determined and used for the estimate of *k*.

Though the Color Method is relatively straightforward, its simplicity comes with a price. The number of color bins we employ is somewhat subjective and will unquestionably impact the final value of *k*—using more (less) bins will generally give higher (lower) values of *k* for the same image. We chose the 54 bins as described because they define colors that are easily discernable to the human eye, and thus would reasonably model a human visually attempting to identify background materials. The use of 95% as the coverage to threshold in Step 4 is also subjective. Our use of this threshold was simply based on a preconceived notion of the percent of anomalies that may be present in an image. It would also seem reasonable to include information on the number of observations in a bin, though we did not test such a strategy here.

A final drawback with the Color Method is the well-known problem of rendering accurate true-color images, particularly when little is known about the measurement scale used to record the intensity of red, green, and blue light. We have found that in actual images that contain materials with reflectance values close to 0% and 100% for each color band, the Color Method performs relatively well. However, for images that do not adequately contain a full range of intensity values in each color band, the method tends to over-estimate the number of colors in the image because the absolute upper and lower bounds for each color band are set to those found in the image, rather than what the sensor is capable of detecting. This incorrect definition of the measurement scale for each band essentially performs a histogram stretch on the image, thereby generating more colors than would actually be present in an accurate true-color image. This problem can

be mitigated if the sensor's true dynamic range for each color band is used in the detection algorithm, but this information may not be readily available to the end-user.

k-Selection Tests with Simulated Data

In the preceding paragraphs, we defined six methods for determining the number of clusters in a dataset. We now describe the two experiments used to assess their relative merits for detecting clusters in hyperspectral imagery. The first experiment uses simulated multivariate Gaussian and multivariate-*t* data to measure each method's ability to detect the correct number of clusters in a dataset when we know definitively the true number of clusters. The second experiment tests the ability of the best performers from the first experiment to detect the number of clusters in actual images in which we are not able to determine the true number of clusters with absolute certainty. The first experiment is summarized as follows:

1) Starting with k=2, choose k background materials from a set of m mean vectors and covariance matrices representing m different background materials derived from actual hyperspectral spectra.
2) Generate 500 observations for each material selected in Step 1 using a multivariate Gaussian distribution and each material's respective mean vector and covariance matrix. Combine these observations to form the dataset.
3) Apply each of the six algorithms defined previously to the dataset, and record the number of clusters each algorithm detects.
4) Repeat Steps 1 through 3 30 times to account for the effects of random sampling.

Table 16. Materials used in Simulated Data *k*-Selection Tests

Fort A.P. Hill Materials		D.C. Mall Materials		Purdue Materials	
ID	**Material**	**ID**	**Material**	**ID**	**Material**
1	Grass	1	Asphalt	1	Grass
2	Road	2	Grass	2	Dead Grass
3	Dead Grass	3	Dead Grass	3	Asphalt
4	Trees	4	Gravel	4	Plowed Dirt
5	Shadow	5	Roof 1	5	Athletic Track
		6	Roof 2	6	Water
		7	Water		

5) Repeat Steps 1 through 4 for all remaining combinations of *k* background materials.
6) Repeat Steps 1 through 5 with k=3,4,…,m.
7) For each setting of *k*, record the percent of cases each algorithm correctly estimated *k*, the percent of cases each algorithm estimated k to be within plus or minus one of the true value, and the mode of the predicted value of *k*.
8) Repeat the entire experiment using a multivariate-*t* distribution with 11 degrees of freedom.

This experiment was run three different times for background datasets derived from the Fort A.P. Hill, D.C. Mall, and Purdue images. The materials used from each image are listed in Table 16. In order to alleviate the computational burden of this experiment, we also reduced the dimensionality of the data to p=15 using the same band aggregation scheme employed in the *k*-means robustness experiment. Finally, because the simulated datasets are not actual images, we explicitly specified the approximate dynamic range for the respective sensors in the Color Method to obtain a more realistic assessment of its performance.

Table 17. Results of *k*-Selection Test using Simulated Fort A.P. Hill Data (Multivariate Gaussian)

Method	Number of Clusters			
	2	3	4	5
Calinski-Harabasz				
% Correct	81.3	65.7	52.7	33.3
% W/I +/- 1	98.7	84.3	69.3	40.0
Mode	2.0	3.0	4.0	5.0
Marriott				
% Correct	10.0	0.0	0.0	0.0
% W/I +/- 1	10.0	0.0	0.0	0.0
Mode	14.0	14.0	14.0	13.0
Silhouette				
% Correct	90.3	64.7	39.3	100.0
% W/I +/- 1	100.0	100.0	78.7	100.0
Mode	2.0	3.0	3.0	5.0
Beale				
% Correct	100.0	0.0	0.0	0.0
% W/I +/- 1	100.0	100.0	0.0	0.0
Mode	2.0	2.0	2.0	2.0
Duda-Hart				
% Correct	100.0	0.0	0.0	0.0
% W/I +/- 1	100.0	100.0	0.0	0.0
Mode	2.0	2.0	2.0	2.0
Color Method				
% Correct	70.0	30.0	0.0	0.0
% W/I +/- 1	100.0	90.0	60.0	0.0
Mode	2.0	2.0	3.0	3.0

The results of the first *k*-selection experiment are listed in Tables 17 through 22. Tables 17 and 18 give the multivariate Gaussian and multivariate-*t* results, respectively, for the Fort A.P. Hill dataset, while Tables 19 and 20 report similar data for the D.C. Mall dataset, and Tables 21 and 22 list the results for the Purdue data. Relative to our objective of finding a suitable method for automatically determining *k*, these tables offer several suggestions. First, it would appear that the Marriott, Beale, and Duda-Hart methods are not useful alternatives. The Marriott method consistently chose values of *k* at or near the maximum value of 14 used in the search. Conversely, the Beale and Duda-Hart methods always selected the minimum value of 2 clusters used in the search process.

Table 18. Results of *k*-Selection Test using Simulated Fort A.P. Hill Data (Multivariate-*t*)

Method	Number of Clusters			
	2	3	4	5
Calinski-Harabasz				
% Correct	71.7	28.9	41.1	15.2
% W/I +/- 1	98.1	52.9	42.1	30.3
Mode	2.0	3.0	4.0	12.0
Marriott				
% Correct	10.3	0.0	0.0	0.0
% W/I +/- 1	10.3	0.0	0.0	0.0
Mode	12.0	10.0	13.0	13.0
Silhouette				
% Correct	91.7	64.3	46.2	100.0
% W/I +/- 1	100.0	100.0	76.3	100.0
Mode	2.0	3.0	4.0	5.0
Beale				
% Correct	100.0	0.0	0.0	0.0
% W/I +/- 1	100.0	100.0	0.0	0.0
Mode	2.0	2.0	2.0	2.0
Duda-Hart				
% Correct	100.0	0.0	0.0	0.0
% W/I +/- 1	100.0	100.0	0.0	0.0
Mode	2.0	2.0	2.0	2.0
Color Method				
% Correct	70.0	30.0	0.0	0.0
% W/I +/- 1	100.0	90.0	60.0	0.0
Mode	2.0	2.0	3.0	3.0

We believe the Marriott method's poor performance can be attributed to a rapidly decreasing value of det(**W**) as k increases, relative to the increasing value of k^2. Since the statistic in (5.43) is designed to use the k^2 term to counter the det(**W**) term (which will always decrease with larger k), a rapidly decreasing det(**W**) can be expected to produce the results we obtained. In regards to the Beale and Duda-Hart method results, it is possible that using a higher significance level to test the respective test statistics—we set α=0.05 for both methods—may produce better results; however, we did not explore this option further.

Table 19. Results of *k*-Selection Tests using Simulated D.C. Mall Data (Multivariate Gaussian)

Method	Number of Clusters					
	2	3	4	5	6	7
Calinski-Harabasz						
% Correct	90.3	59.6	32.6	14.1	6.2	0.0
% W/I +/- 1	98.3	96.4	86.4	66.3	36.7	6.7
Mode	2.0	3.0	5.0	6.0	8.0	9.0
Marriott						
% Correct	0.0	0.0	0.0	0.0	0.0	0.0
% W/I +/- 1	0.0	0.0	0.0	0.0	0.0	0.0
Mode	14.0	14.0	14.0	14.0	14.0	13.0
Silhouette						
% Correct	100.0	77.1	72.1	79.7	91.0	96.7
% W/I +/- 1	100.0	100.0	97.0	95.1	100.0	100.0
Mode	2.0	3.0	4.0	5.0	6.0	7.0
Beale						
% Correct	100.0	0.0	0.0	0.0	0.0	0.0
% W/I +/- 1	100.0	100.0	0.0	0.0	0.0	0.0
Mode	2.0	2.0	2.0	2.0	2.0	2.0
Duda-Hart						
% Correct	100.0	0.0	0.0	0.0	0.0	0.0
% W/I +/- 1	100.0	100.0	0.0	0.0	0.0	0.0
Mode	2.0	2.0	2.0	2.0	2.0	2.0
Color Method						
% Correct	28.6	26.8	58.7	29.0	0.0	0.0
% W/I +/- 1	57.1	80.9	100.0	93.8	56.2	0.0
Mode	4.0	4.0	4.0	4.0	5.0	4.0

Whereas the Marriott, Beale, and Duda-Hart methods appeared to falter with the datasets used in the experiment, the Silhouette and Calinski-Harabasz methods faired somewhat better. For the Fort A.P. Hill and D.C. Mall datasets, the Silhouette method was the stronger performer in its ability to choose the correct value of *k* and to provide a value within one cluster of the correct value. For the Purdue dataset, the Calinski-Harabasz method clearly gave the best performance. The Silhouette method also did well with this dataset, but in the multivariate Gaussian case faltered with 5 of the 20 combinations of three materials and 5 of the 15 combinations of four materials. If these combinations were omitted from the statistics, the Silhouette method would have

Table 20. Results of *k*-Selection Tests using Simulated D.C. Mall Data (Multivariate-*t*)

Method	Number of Clusters					
	2	3	4	5	6	7
Calinski-Harabasz						
% Correct	39.6	33.7	27.8	15.7	21.9	0.0
% W/I +/- 1	72.8	70.7	55.7	38.3	43.8	10.0
Mode	2.0	3.0	5.0	6.0	8.0	10.0
Marriott						
% Correct	0.0	0.0	0.0	0.0	0.0	0.0
% W/I +/- 1	0.0	0.0	0.0	0.0	0.0	0.0
Mode	14.0	13.0	12.0	11.0	13.0	14.0
Silhouette						
% Correct	100.0	77.1	12.7	47.5	47.1	71.4
% W/I +/- 1	100.0	100.0	93.5	91.5	100.0	100.0
Mode	2.0	3.0	3.0	5.0	6.0	7.0
Beale						
% Correct	100.0	0.0	0.0	0.0	0.0	0.0
% W/I +/- 1	100.0	100.0	0.0	0.0	0.0	0.0
Mode	2.0	2.0	2.0	2.0	2.0	2.0
Duda-Hart						
% Correct	100.0	0.0	0.0	0.0	0.0	0.0
% W/I +/- 1	100.0	100.0	0.0	0.0	0.0	0.0
Mode	2.0	2.0	2.0	2.0	2.0	2.0
Color Method						
% Correct	41.3	23.1	40.9	48.5	0.0	0.0
% W/I +/- 1	89.2	73.2	100.0	91.2	50.4	0.0
Mode	3.0	5.0	4.0	5.0	5.0	5.0

estimated the correct number of clusters 100% of the time when three and four clusters were used. It is also apparent that, upon consideration of the distribution used to generate the test data, the Silhouette method has a better capacity to deal with heavy-tailed data. Though the Silhouette method's performance is clearly affected when multivariate-*t* data is used, its degradation is not as severe as the decline experienced with the Calinski-Harabasz method.

Based on the preceding discussion, it would appear that the Silhouette method is the preferable method for automatically determining *k*. However, this conclusion should be viewed with caution for several reasons. First, the *k*-selection experiment used simulated data with equal cluster sizes. Actual hyperspectral data and background

Table 21. Results of *k*-Selection Test using Simulated Purdue Data (Multivariate Gaussian)

Method	Number of Clusters				
	2	3	4	5	6
Calinski-Harabasz					
% Correct	100.0	99.8	99.8	97.8	100.0
% W/I +/- 1	100.0	100.0	100.0	100.0	100.0
Mode	2.0	3.0	4.0	5.0	6.0
Marriott					
% Correct	31.8	3.3	0.0	0.0	0.0
% W/I +/- 1	31.8	3.7	0.2	0.0	0.0
Mode	14.0	14.0	14.0	14.0	13.0
Silhouette					
% Correct	100.0	75.0	60.0	32.8	0.0
% W/I +/- 1	100.0	100.0	86.7	88.9	100.0
Mode	2.0	3.0	4.0	4.0	5.0
Beale					
% Correct	100.0	0.0	0.0	0.0	0.0
% W/I +/- 1	100.0	100.0	0.0	0.0	0.0
Mode	2.0	2.0	2.0	2.0	2.0
Duda-Hart					
% Correct	100.0	0.0	0.0	0.0	0.0
% W/I +/- 1	100.0	100.0	0.0	0.0	0.0
Mode	2.0	2.0	2.0	2.0	2.0
Color Method					
% Correct	26.7	25.0	35.1	66.7	100.0
% W/I +/- 1	100.0	100.0	100.0	100.0	100.0
Mode	3.0	4.0	5.0	5.0	6.0

materials that appear in unequal proportions may give different results. Second, the number of distances that need to be computed by the Silhouette method is

$$\sum_{i=1}^{k}\left[\frac{n_i\left(n_i-1\right)}{2}+n_i n_{i0}\right] \tag{5.47}$$

where

n_i = the number of observations in cluster *i*, and
n_{i0} = the number of observations in the cluster closest to cluster *i*.

Thus, the Silhouette method can become quite computationally expensive for large datasets unless some action is taken to alleviate the problem. A final precaution with

Table 22. Results of *k*-Selection Test using Simulated Purdue Data (Multivariate-*t*)

Method	Number of Clusters				
	2	3	4	5	6
Calinski-Harabasz					
% Correct	100.0	54.2	51.0	59.2	100.0
% W/I +/- 1	100.0	100.0	100.0	100.0	100.0
Mode	2.0	3.0	4.0	5.0	6.0
Marriott					
% Correct	11.9	6.7	0.0	0.0	0.0
% W/I +/- 1	11.9	10.4	11.3	0.0	0.0
Mode	12.0	9.0	11.0	11.0	11.0
Silhouette					
% Correct	100.0	75.0	25.6	47.7	0.0
% W/I +/- 1	100.0	100.0	75.2	89.8	100.0
Mode	2.0	3.0	3.0	5.0	5.0
Beale					
% Correct	100.0	0.0	0.0	0.0	0.0
% W/I +/- 1	100.0	100.0	0.0	0.0	0.0
Mode	2.0	2.0	2.0	2.0	2.0
Duda-Hart					
% Correct	100.0	0.0	0.0	0.0	0.0
% W/I +/- 1	100.0	100.0	0.0	0.0	0.0
Mode	2.0	2.0	2.0	2.0	2.0
Color Method					
% Correct	26.7	6.2	32.6	66.7	100.0
% W/I +/- 1	100.0	100.0	100.0	100.0	100.0
Mode	3.0	4.0	5.0	5.0	6.0

using the Silhouette method to choose *k* is that our tests indicate it is very good at estimating *k* within one cluster of the correct value, but its ability to find the exact answer can be somewhat lacking. This limitation is also evident with the Calinski-Harabasz method, which leads to the question: How accurate does the *k*-selection algorithm need to be? Certainly, to obtain the most accurate anomaly detection results we would like to know the true value of *k* with certainty; however, we demonstrate later in this dissertation that using different values of *k* within a small range generally does not significantly change anomaly detection results. Thus, if a selection algorithm can estimate *k* within

one or two clusters of the true value, it is generally sufficient to detect a range of militarily significant targets.

Before moving on to the second *k*-selection test, several comments are in order concerning the Color Method. As indicated by Tables 17 through 22, this method only performed well when applied to the Purdue dataset. Though this marginal performance seems to argue against using this method, one must consider the scaling problem discussed previously. Because we used artificial datasets in our tests, it was difficult to define the proper measurement scales to accurately represent the background materials in the RGB color space. Though we attempted to manually specify the dynamic range for each color band, we are not confident that we were successful in this effort. In light of this problem, we will continue to investigate this method further in our next experiment that uses actual imagery.

k-Selection Tests using Actual Images

In the preceding *k*-selection test, our primary goal was to identify useful *k*-selection methods under controlled conditions. In our second *k*-selection test, we now investigate the performance of the Silhouette, Calinski-Harabasz, and Color methods when applied to actual images. Additionally, we used this experiment to gain insight into data pre-processing steps that may either reduce the computational burden of a method, or improve its accuracy. With these objectives in mind, the second *k*-selection experiment is outlined as follows:

1) Set the *k*-selection method to be tested.
2) Set the image to be used for the test.

3) Preprocess the image data using a technique relevant to the *k*-selection method established in Step 1.
4) Apply the *k*-selection method to the pre-processed image and record the number of clusters detected.
5) Repeat Steps 3 and 4 for all preprocessing techniques defined for the *k*-selection method.
6) Repeat Steps 2 through 5 for all images to be tested.
7) Repeat Steps 1 through 6 for all *k*-selection methods to be tested.

Six hyperspectral images were used in Step 2 of the experiment. True-color versions of these images are shown in Appendix B. These images are chip-outs from larger images and were chosen to give a range of challenges for the selection methods. Table 23 lists the parent image for each image chip, as well as the respective sensor, number of pixels, and major background materials contained in the image. Admittedly, determining the background materials for each image chip is somewhat subjective and was conducted by simple visual inspection of each image.

The preprocessing steps used in Step 3 of the experiment were tailored to each selection method. For the Calinski-Harabasz method, the general preprocessing factors included dimensionality reduction, data normalization, and random sampling. The dimensionality reduction factor was used to determine the most useful way to reduce the number of variables in the dataset in an effort to increase processing speed. The data normalization factor tested if normalizing each pixel in the image helped to increase separation between background materials, thereby improving the selection method's ability to correctly estimate the number of clusters. Finally, the random sampling factor

Table 23. Description of Images Used in *k*-Selection Test

Chip ID	Parent Image	Sensor	Pixels	Major Background Materials
1	Forest Radiance I	HYDICE	23085	Trees, Shadow, Dead Grass, Healthy Grass, Road 1, Road 2, Bushes
2	Desert Radiance II	HYDICE	20447	Brush, Road, Light Soil, Medium Soil, Dark Soil
3	Forest Radiance I	HYDICE	3721	Dead Grass, Road, Healthy Grass
4	D.C. Mall	AVIRIS	11948	Water, Dead Grass, Healthy Grass, Trees, Gravel, Asphalt, Roof, Concrete
5	Purdue	HYMAP	12000	Dead Grass, Healthy Grass, Road, Roof 1, Roof 2, Roof 3, Stadium Seats
6	Purdue	HYMAP	11760	Trees, Plowed Dirt, Dead Grass, Road, Water

Table 24. Factor Definition for *k*-Selection Test

Calinski-Harabasz Factors		
Dimensionality Reduction	**Normalization**	**Random Sampling**
Full Dimensionality	Don't Normalize Spectra	All Data
5 Principal Components	Normalize Spectra	2000 Spectra
10 Principal Components		5000 Spectral
Band Aggregation to 15 Bands		
Band Aggregation to 30 Bands		
Silhouette Factors		
Full Dimensionality	**Normalization**	**Random Sampling**
5 Principal Components	Don't Normalize Spectra	All Data
10 Principal Components	Normalize Spectra	2000 Spectra
Band Aggregation to 15 Bands		4000 Spectra
Band Aggregation to 30 Bands		
Color Method Factors		
Normalization	**Square-Root Transform**	
Don't Normalize Spectra	Don't use Transform	
Normalize Spectra	Use Transform	

tested if the performance of the Calinski-Harabasz method is affected when applied to a randomly drawn subset of the image pixels. If performance is not affected, this preprocessing step offers another strategy for reducing the computational burden of the

selection method in addition to dimensionality reduction. The levels used for each of the three factors are listed in Table 24. These factors were combined in a full factorial design and applied to each image chip. The same factors and experimental design were used for the Silhouette method, with the exception that different factor levels were used for the random sampling factor. Because the Silhouette method is so computationally expensive, the method was not applied to the full dataset, and the sample size of 5000 spectra was replaced by a sample size of 4000 spectra—the practical limit that the method could handle.

For the Color Method, only two factors were used: data normalization and the square-root transformation. The normalization factor again was used to determine if normalization better-separated the background materials so that their number could be better determined. The square-root transformation factor tests if taking the square root of all the spectra sufficiently minimizes the hazing effect caused by atmospheric particles, thereby improving the contrast between different colors in the image. This transformation is a well-known technique commonly used in gray-scale and color image processing. The levels for these two factors are simply using the preprocessing method or not. As with the other two selection methods, these factors are combined in a factorial design and applied to each image.

The results of the second *k*-selection test for the Calinski-Harabasz, Silhouette, and Color methods are listed in Tables 25, 26, and 27, respectively. Each row of these tables lists the factor combination tested and the corresponding number of clusters the respective method detected in each image chip. As a reference, the manually estimated number of background materials are listed in parentheses after each image chip name.

Table 25. Actual Image *k*-Selection Test Results (Calinski-Harabasz Method)

Design Point	Image Chip					
	1 (7)	2 (5)	3 (3)	4 (8)	5 (7)	6 (5)
Full/---/All	2	3	3	3	2	2
P5/---/All	3	2	3	3	6	3
P10/---/All	3	2	3	3	6	3
A15/---/All	2	6	3	3	2	2
A30/---/All	2	6	3	3	2	2
Full/N/All	3	3	3	10	4	3
P5/N/All	3	2	3	10	4	2
P10/N/All	3	2	3	3	4	2
A15/N/All	8	3	3	13	4	3
A30/N/All	6	3	3	13	4	3
Full/---/2000	2	6	3	3	2	2
P5/---/2000	3	2	3	3	5	3
P10/---/2000	3	2	3	3	3	3
A15/---/2000	2	6	3	3	2	2
A30/---/2000	2	3	3	3	2	2
Full/N/2000	6	3	3	13	4	3
P5/N/2000	3	2	3	3	4	2
P10/N/2000	11	2	3	10	4	2
A15/N/2000	6	3	8	13	4	3
A30/N/2000	6	3	3	13	4	3
Full/---/5000	2	6	3	3	2	2
P5/---/5000	3	2	3	3	6	3
P10/---/5000	3	2	3	3	6	3
A15/---/5000	2	6	3	3	2	2
A30/---/5000	2	3	3	3	2	2
Full/N/5000	3	3	3	10	4	3
P5/N/5000	3	2	3	11	4	2
P10/N/5000	3	2	3	3	4	2
A15/N/5000	8	3	3	13	4	3
A30/N/5000	6	3	3	13	4	3

For the Calinski-Harabasz and Silhouette methods, the design point naming convention is in three parts—for example, "P5/N/All." The first part designates the variable reduction method used and can assume the values Full, P5, P10, A15, and A30 which stand for full dimensionality, five principal components, ten principal components, band aggregation into 15 variables, and band aggregation into 30 variables, respectively. The

Table 26. Actual Image *k*-Selection Test Results (Silhouette Method)

Design Point	Image Chip					
	1 (7)	2 (5)	3 (3)	4 (8)	5 (7)	6 (5)
Full/---/2000	2	6	3	3	2	2
P5/---/2000	3	2	3	3	6	3
P10/---/2000	3	2	3	3	6	3
A15/---/2000	2	7	3	3	2	2
A30/---/2000	2	6	3	3	2	2
Full/N/2000	5	3	3	14	4	3
P5/N/2000	3	3	3	10	4	2
P10/N/2000	3	2	3	11	4	2
A15/N/2000	8	3	3	12	4	3
A30/N/2000	9	3	3	13	4	3
Full/---/4000	2	7	3	3	2	2
P5/---/4000	3	2	3	3	5	3
P10/---/4000	3	2	3	3	6	3
A15/---/4000	2	7	3	3	2	2
A30/---/4000	2	7	3	3	2	2
Full/N/4000	3	3	3	10	4	3
P5/N/4000	3	2	3	3	4	2
P10/N/4000	3	2	3	10	4	2
A15/N/4000	6	3	3	13	4	3
A30/N/4000	3	3	3	13	4	3

second part of the design point name indicates if normalization was performed on the data, and is set to "---" if the data was not normalized or N if it was. The third field of the design point name indicates if the selection method was applied to a random subset of the image data. The use of All in this field specifies that the entire dataset was used. A numeric value in this field indicates the size of the random subset. Note that in the case of Image Chip 3, which only contained 3721 pixels, a random subset size of 4000 or 5000 indicates the entire image was used.

For the Color method, the design point naming convention uses only two parts to indicate whether or not normalization and the square root transformation were used. Use of a "---" in the first field indicates that the data was not normalized, and Norm indicates

Table 27. Actual Image *k*-Selection Test Results (Color Method)

Design Point	Image Chip					
	1 (7)	2 (5)	3 (3)	4 (8)	5 (7)	6 (5)
---/---	6	3	5	4	4	5
Norm/---	4	2	3	3	3	3
---/Sqrt	4	2	3	4	5	4
Norm/Sqrt	3	1	2	3	3	3

that it was normalized. Similarly, a "---" in the second field indicates the square root transformation was not employed, and Sqrt specifies that it was used.

Addressing the Calinski-Harabasz method first, we notice in Table 25 that this method's ability to estimate the same number of clusters as estimated manually is dependant on the preprocessing action applied to the image. Further, the preprocessing action that works well for one image does not necessarily improve performance for another image. For example, we see that band aggregation and normalization improve the method's performance on Image Chip 1, but have minimal effect with Image Chips 2, 3, 5, and 6, while producing too high a value of *k* for Image Chip 4. Likewise, we see that using principal component analysis to reduce the number of variables works well with Image Chip 5, but has minimal benefit when applied to the other images. Moreover, we see that none of the preprocessing actions give satisfactory results for Image Chips 4 and 6. Where these results complicate the issue of how best to use the Calinski-Harabasz method, it is encouraging that applying the method to a randomly drawn subset of the image does not appear to significantly affect performance relative to using the entire image. Thus, it is reasonable to conclude that random sampling of the dataset offers a means to reduce the computational burden of the method.

As seen in Table 26, the performance of the Silhouette method in the second *k*-selection experiment is similar to that of the Calinski-Harabasz method. Specifically, we see that different preprocessing actions are useful with different images, and that none of the preprocessing combinations appear to benefit Image Chips 4 and 6. The benefit of using random sampling with the Silhouette method cannot be determined with certainty because the methods computational complexity precluded its application to the full images except for Image Chip 3. However, we can say that the using a sample size of 2000 spectra as opposed to 4000 spectra did not significantly change performance of the algorithm, which is an encouraging result for this computationally intensive method.

For both the Calinski-Harabasz and Silhouette methods, the underlying causes that dictate which preprocessing methods are better suited to specific images, is not known at this time. It is reasonable to believe that relative sizes of background clusters, separability of spectra, and spectra variability are the primary factors that determine the benefits of one preprocessing technique over another, but further research is required to justify this hypothesis.

Moving on to the Color Method, it seen in Table 27 that this method's performance is also affected by image-dependent preprocessing. For example, normalizing the data or applying the square root transformation improves the method's performance on Image Chips 3 and 5, but tends to degrade performance on the other images. A more notable finding is that when no preprocessing is applied to the images, the Color method tends to give estimates of *k* closer to the manually estimated number of materials than the Calinski-Harabasz or Silhouette methods. This result is significant for two reasons. First, we would prefer to use a *k*-selection method that does not require the

user to deliberate over the best choice of image preprocessing in order to conduct an anomaly detection analysis. The Color method would appear to satisfy this goal better than the other two methods, though certainly not in a flawless manner. Second, the Color method arrives at its estimate in a fraction of a second as opposed to tens of minutes or longer for the other two methods. In short, the Color method arrives at better estimates of *k* when no preprocessing is applied, and does so in a fraction of the time. Though it is possible to obtain comparable answers with the Calinski-Harabasz and Silhouette methods after applying an appropriate combination of preprocessing techniques, it is unknown at this time how to select the appropriate techniques for an arbitrary image.

Conclusions

In the preceding paragraphs we evaluated several methods for automatically determining the number of clusters in a hyperspectral dataset with the goal of identifying the most useful method for automatically clustering an image prior to applying the BACON or FAST-MCD outlier detectors. This evaluation was performed using two experiments that tested the candidate selection methods with simulated and actual hyperspectral data, respectively. In the first *k*-selection experiment which used simulated data, the Calinski-Harabasz and Silhouette methods performed the best of the algorithms tested. However, in the second *k*-selection test in which actual images were used, the Color method gave more favorable results. The question arises, then, as to which method should be used in our overall anomaly detection methodology?

Though none of the methods tested can be viewed as clearly superior to the others, we suggest the Color Method as the method of choice for several reasons to which we have already alluded. First, it is extremely fast, an important attribute for practical

anomaly detection analyses. Second, the Color Method does not perform any worse than the Calinski-Harabasz or Silhouette methods when used on actual images. Though the latter two methods handled the simulated datasets better than the Color Method, this better performance is due in-part to scaling problems imposed by the simulated data that conspire against the Color Method. Finally, the Color Method performs better against the real images without any image preprocessing. The Calinski-Harabasz and Silhouette methods, however, generally require some form of preprocessing to provide useful estimates of *k*. Because the effect of different preprocessing actions depends on the image, we feel that the relative simplicity of the Color Method is more conducive to an autonomous anomaly detection methodology. Before departing this topic to discuss the integration of image clustering with the BACON and FAST-MCD algorithms, we note that this problem of adequately estimating the number of clusters for use with *k*-means is an unresolved problem. Though we choose to use the Color Method for the reasons stated, a more expansive investigation into this topic may lead to better solutions.

The AutoDet Anomaly Detector

Thus far in this chapter, we have identified the BACON and FAST-MCD multivariate outlier detection methods as suitable methods for detecting anomalies in large, high-dimensional datasets representative of hyperspectral data. We have also determined that the *k*-means clustering algorithm is reasonably robust to datasets contaminated by outliers and can be used as a means to group the spectra of a hyperspectral image into homogenous groups so that the BACON and FAST-MCD methods can be applied to these groups. Finally, we have argued that the Color Method is preferable to other *k*-selection methods for automating the specification of *k* in the *k*-

means algorithm. In this section we combine these components into an overall methodology—which we refer to as the AutoDet detector—that is well-suited to autonomously detect anomalies in a range of hyperspectral images. By *autonomous* we mean that the only required input from the user is the hyperspectral image itself. In the following paragraphs, we first outline the AutoDet method and then use Taguchi robust parameter design methods to configure the detector to produce accurate anomaly detection results across a range of hyperspectral images. We conclude by demonstrating the superior performance of AutoDet relative to benchmark anomaly detectors found in the anomaly detection literature.

Algorithm Overview

The AutoDet anomaly detector can be run using either the BACON algorithm or the FAST-MCD algorithm and consists of the following basic steps:

1) Ensure all atmospheric absorption bands and any bands with significant noise or artifacts have been removed from the hyperspectral image cube. These bands are quite capable of causing false alarms and contribute no useful information for the AutoDet method.
2) Apply the Color Method to the red, green, and blue color bands of the image to obtain an estimate of the number of background materials in the image.
3) Reduce the dimensionality of the data by aggregating the original image bands into 30 bands using the band aggregation method described earlier in this chapter.
4) Use the estimate obtained in Step 2 in the *k*-means algorithm to cluster the image. Use the Cosine assignment rule within *k*-means to better ensure

spectra from the same material but under different illumination conditions are included in the same cluster.

5) Apply the BACON or FAST-MCD algorithm to each cluster to obtain robust MSDs for the observations in the clusters. Produce a gray-scale image that displays the relative magnitude of the MSDs for each pixel.
6) Threshold the robust MSDs to detect anomalous spectra and produce a binary image that indicates the location of anomalies.

Though the AutoDet procedure is relatively straightforward, there are several algorithm specific implementation details in Steps 5 and 6 that require further explanation. For the BACON algorithm, we modified the original algorithm's criteria for adding new observations to the basic subset. Specifically, at each iteration of the algorithm, we included an observation in the basic subset if the observation's MSD was less than the 0.9999-quantile of the Chi-Square distribution with 30 degrees of freedom as opposed to the $(1-\alpha/n)$-quantile as originally proposed, where n is the cluster size. This modification was necessary because the extremely large cluster sizes—sometimes as large as 30000 spectra—common in hyperspectral images results in cut-off values so large that the target spectra we hope to detect are included in the basic subset, thereby defeating the objective of the BACON algorithm.

In addition to modifying the basic subset threshold, we also use the component-wise median vector for each cluster in forming the initial basic subset in our implementation of the BACON algorithm. As stated by Billor et al., this option for forming the initial basic subset is non-affine equivariant, but provides a higher break-down point than the alternative option of using classical MSDs in forming the initial

basic subset. Based on cursory trials with the two methods, we confirmed that using the more robust component-wise median in forming the initial basic subset generally detects outliers that the non-robust method misses, and hence we choose to use this method in the AutoDet implementation of BACON.

In configuring the FAST-MCD algorithm for use in Steps 5 and 6 of AutoDet, we set the so-called half-sample size to h=0.75n, where n is the cluster size. We use this value of h as opposed to the higher-breakdown version of h=[(n+p+1)/2] because we feel that the contamination level in most anomaly detection studies is well below 25%, and that the algorithm better estimates the true shape of the cluster covariance matrix if larger values of h are used. The nesting procedure used in FAST-MCD to deal with very large datasets is invoked when cluster sizes exceed 1000 spectra. When the nesting procedure is used, we configured the algorithm to initialize the nesting procedure with 100 random samples. These algorithms settings are based on recommendations given in Rousseeuw and van Driessen (1999).

A notable modification we made to the FAST-MCD algorithm for use in AutoDet, deals with the scaling of the MSDs that it produces. In the original description of the algorithm, the primary output of the method is the mean vector and covariance matrix of the half-sample of observations giving the smallest covariance determinant. In searching for the half-sample with smallest covariance determinant, the algorithm will naturally exclude observations far from the centroid of the data. If the actual number of outliers in the data is considerably less than (1-h), the excluded observations will consist of both good observations and outliers. Hence, the covariance matrix of the half sample will tend to underestimate the total variance of the good data, which, in turn, will lead to

MSDs that are too large. Under the assumption of Gaussian data, Rousseeuw and van Driessen suggest countering this problem by multiplying the half-sample covariance matrix by the scaling factor given in (5.10) followed by a one-step re-weighted estimate of the mean and covariance matrix using the procedure given by Rousseeuw and von Zomeren. Only after the mean vector and covariance matrix of the half-sample are adjusted in this manner are the MSDs for the observations computed and compared to a suitable threshold for outlier detection. Our experience with this method, however, indicates that it still tends to underestimate the true variance of the good data, thereby leading to large MSDs. These inaccurate MSDs, in turn, lead to a large number of false alarms, as demonstrated in our simulated data tests of the FAST-MCD algorithm presented earlier. This problem was also identified in Smetek and Bauer (2007).

To correct for this problem, it was proposed in Smetek and Bauer (2006) that the original half-sample produced by FAST-MCD should be input to the BACON algorithm as a basic subset and "grown" using BACON's iteration scheme. Though this process significantly reduces false alarms relative to the original FAST-MCD algorithm, it is essentially just the BACON algorithm using FAST-MCD to generate the initial basic subset; therefore, outliers that tend to elude the original BACON algorithm may also elude the hybrid FAST-MCD/BACON algorithm. In an effort to maintain the FAST-MCD algorithm as a distinct method for detecting outliers, we explored additional methods for scaling the FAST-MCD MSDs. Specifically, we implemented scaling methods described in Maronna and Yohai (1995) and Hardin and Rocke (2005), but met with the same limitations as the original FAST-MCD scaling method. Additional methods proposed in the literature include those described in Rocke and Woodruff (1996)

and Atkinson (1994); however, these methods require the use of Monte Carlo simulations to determine dataset-dependent threshold values, a process we do not view as computationally practical given the size of hyperspectral datasets.

Based on research conducted by Meidunas (2006), it is likely that the MSD scaling methods proposed in the literature fail to work satisfactorily with hyperspectral data because the data is not multivariate Gaussian; therefore, the MSDs are not Chi-Square distributed, an assumption upon which these scaling factors are built. For the BACON algorithm, which does a better job estimating the total variance in the data, using an inaccurate Gaussian assumption can still produce descent detection results, as indicated in our previous tests. For the FAST-MCD algorithm, however, the underestimated variance evidently tolerates smaller departures from normality. As an alternative to using an inappropriate Gaussian assumption, Meidunas offers methods for modeling the distributions of MSDs obtained from hyperspectral data. However, these methods assume that the MSDs have been computed from an accurate covariance matrix; when using the FAST-MCD half-sample covariance matrix to compute the MSDs, this is not the case. Though it may be possible to estimate the distribution of the half-sample-based MSDs, standard methods for distribution estimation are not practical because MSDs corresponding to outlying observations will lead to distribution estimates with unusually heavy tails. If these distribution estimates are used to threshold MSDs for anomaly detection, it is likely that outliers will be masked by the heavy tails.

Though further research is certainly required to adequately resolve this MSD scaling dilemma, we are still in need of a procedure for use in Step 6 of the AutoDet procedure when using the FAST-MCD algorithm in Step 5. As a simple solution, we use

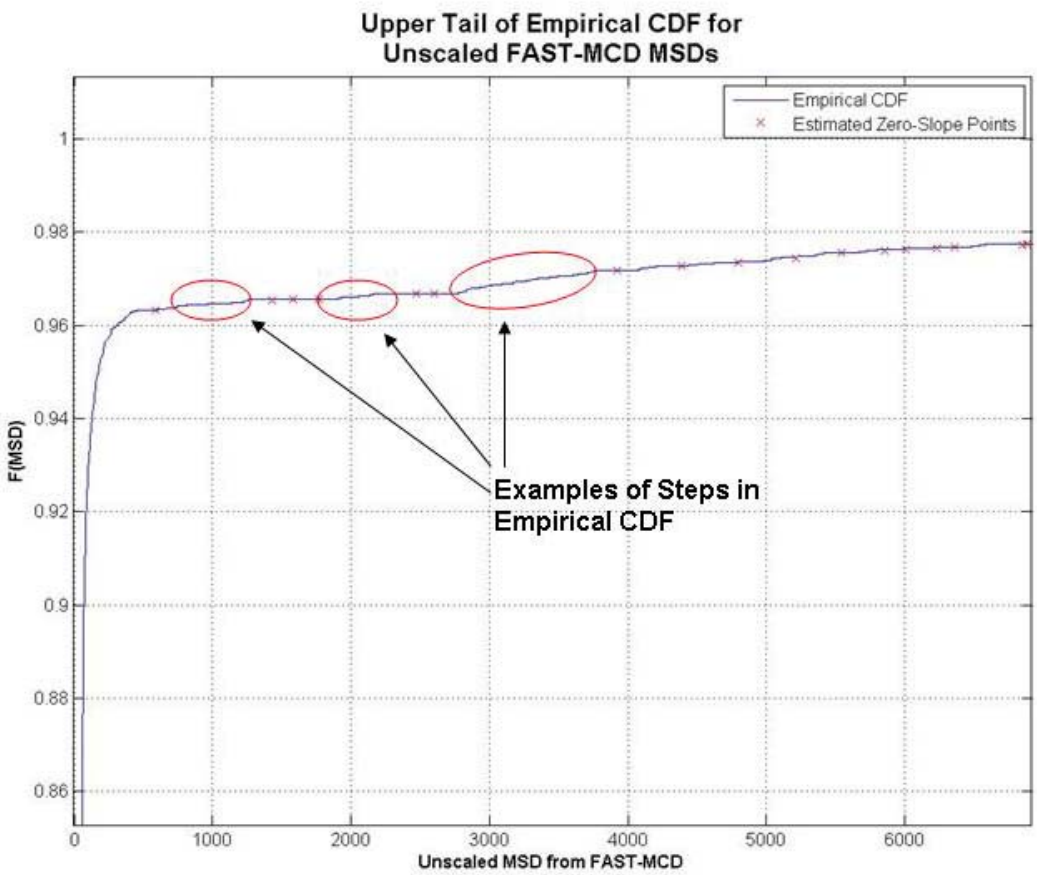


Figure 21. Example of Steps in CDF Caused by Anomalies

the following: 1) construct the empirical cumulative distribution function (CDF) of the MSDs generated using the half-sample mean vector and covariance matrix estimate; 2) identify the MSD value above the 0.75-quantile of the empirical CDF at which the slope of the empirical CDF is close to zero; and 3) designate any observations whose MSDs exceed this zero-slope-point as anomalies. The reasoning behind this method is that anomalies which are well-separated from the background spectra in hyperspectral data will tend to cause "steps" in the empirical CDF as illustrated in Figure 21. These steps are characterized by the slope of the empirical CDF converging to zero before the step, followed by a sudden increase. As will be seen later in this chapter, this method for

thresholding FAST-MCD MSDs performs well for some images, but can fail if the anomalous spectra are not well-separated from the background material. As previously stated, further research is required to improve upon this method.

Robust Parameter Design

In constructing the AutoDet methodology, a number of decisions were required concerning how to set various parameters in the methodology in a manner that allows the detector to achieve good detection performance across a range of hyperspectral images. Some of these parameters include the number of features to use with BACON and FAST-MCD, how these features should be constructed, and the cut-off threshold used in BACON's iteration scheme. Additionally, an analysis of the false-negatives produced by the anomaly detection tests conducted in Smetek and Bauer (2007) reveal that normalizing and/or standardizing a cluster's data prior to applying BACON or FAST-MCD could help reveal anomalies that otherwise evade detection. In short, many alternatives present themselves as possible ways to implement the AutoDet methodology.

To be consistent with our objective of developing an autonomous anomaly detector, we felt it necessary to remove the burden of configuring AutoDet from the user by providing a configuration that achieves consistently good performance across a range of images that the user may encounter. Thus, we seek a combination of settings for a group of controllable algorithm parameters that produces detection results that are robust to a noise variable, namely the different hyperspectral images presented to the algorithm. This problem is none other than a robust parameter design problem, a solution to which can be obtained using classical Taguchi robust parameter design methods. The following

Table 28. Factors and Levels for Taguchi Experiments

Detector	Factor	Levels	Level Description
AutoDet-BACON	Normalization (A)	-1	Do not normalize cluster data
		+1	Normalize cluster data
	Standardization (B)	-1	Do not standardize cluster data
		+1	Standardize cluster data
	Threshold (C)	-1	0.9999
		+1	0.999999
	Features (D)	1	15-Band Aggregation
		2	30-Band Aggregation
		3	5 Principal Components
		4	10 Principal Components
	Noise (E)	1	Scene 1
		2	Scene 2
		3	Scene 3
		4	Scene 4
		5	Scene 8
		6	Scene 9
AutoDet-FASTMCD	Normalization (A)	-1	Do not normalize cluster data
		+1	Normalize cluster data
	Standardization (B)	-1	Do not standardize cluster data
		+1	Standardize cluster data
	Features (C)	1	15-Band Aggregation
		2	30-Band Aggregation
		3	5 Principal Components
		4	10 Principal Components
	Noise (D)	1	Scene 1
		2	Scene 2
		3	Scene 3
		4	Scene 4
		5	Scene 8
		6	Scene 9

paragraphs describe how the Taguchi method, as defined in Myers and Montgomery (1995), was used to configure the AutoDet methodology for both the BACON and FAST-MCD methods.

For the AutoDet-BACON algorithm, the Taguchi method was used to configure the four factors listed in Table 28. The Normalization factor (Factor A) is used to

determine if normalizing the spectra in a cluster prior to applying BACON improves anomaly detection. This factor can assume one of two levels: either the data is normalized or it is not normalized. The Standardization factor (Factor B) assesses the benefit of standardizing the cluster data prior to using BACON. This factor also assumes one of two levels: the data is standardized or it is not standardized. To standardize the data, we used the robust standardization method described by Chiang, Pell, and Seashotlz in which the value, x_{ij}, of observation i for variable j is replaced by

$$d_{ij} = \frac{x_{ij} - \tilde{x}_j}{\underset{i}{\mathrm{med}}\left(x_{ij} - \tilde{x}_j\right)} \tag{5.48}$$

where

$$\tilde{x}_j = \text{the median value of variable } j.$$

The Threshold factor (Factor C) is used to study the effect of changing the cut-off value used in each iteration of BACON to form the basic subset. As mentioned previously, the original BACON algorithm uses the (1-α/n)-quantile of the Chi-Square distribution with p degrees of freedom as the cut-off. In our experiment we allow this factor to take-on the values of either the 0.9999-quantile or the 0.999999-quantile of the Chi-Square distribution. The Features factor (Factor D) defines the manner in which the dimensionality of the hyperspectral data is reduced. Four levels were considered for this factor: 1) band aggregation into 15 bands; 2) band aggregation into 30 bands; 3) PCA reduction using five principal components; and 4) PCA reduction using 10 principal components.

The noise factor (Factor E) used in the Taguchi experiment defines the image to which the AutoDet-BACON algorithm is applied. This factor has six levels

corresponding to six hyperspectral images. These images are Scenes 1, 2, 3, 4, 8, and 9 found in Appendix C. Scenes 1, 2, and 4 are HYDICE sensor images taken from the Forest Radiance I data collection effort, and represent different degrees of clutter and target complexity with background materials comprised of different types of trees, grass, soil, and asphalt. Scene 3 is taken from a COMPASS image of Fort A.P. Hill, VA, and represents a scene with similar background materials as those in Forest Radiance I scenes, but acquired using a different sensor. Scenes 8 and 9 are also HYDICE images, but were taken from the Desert Radiance II dataset. These two images contain similar targets as the Forest Radiance I images, but the background materials are considerably different, consisting primarily of barren soil with some sparse vegetation. Collectively, these images provide a range of challenges for the AutoDet algorithm, and provide useful settings for the noise variable in the Taguchi experiment. Conceptually, additional images can be added to this experiment at a later time to broaden the scope of this experiment.

The experimental design used for the robust parameter design of the AutoDet-BACON algorithm consists of a full factorial design in Factors A, B, C, and D nested with the six levels of the noise factor. Table 40 in Appendix D lists the factor combinations used for each design point. The response variable measured at each of the 192 design points is the area under the Operating Characteristic (OC) curve computed over the false-positive fraction interval [0.0, 0.01]. This interval was used because false-positive fractions exceeding 0.01 are generally too high to be of value for hyperspectral anomaly detection applications. In a manner consistent with Taguchi robust parameter design, a signal-to-noise ration (*SNR*) was computed for each unique combination of

Factors A, B, C, and D using the responses for the respective combination across the six levels of Factor E. The *SNR* used in this experiment is

$$SNR = -10\log\sum_{i=1}^{6}\left[\frac{1}{y_i^2}\right]\Big/6 \tag{5.49}$$

where y_i is the measured response at the *i*th level of the noise variable. As indicated by Myers and Montgomery (1995), this expression for the *SNR* is applicable when the objective of the robust parameter experiment is to simultaneously find parameter settings that maximize the response variable while minimizing its variance across the levels of the noise variable. Since we are attempting to find a configuration of AutoDet-BACON that maximizes the area under the OC curve in a consistent manner across multiple images, the *SNR* given in (5.49) appears reasonable. It should be noted, however, that many expressions for the *SNR* have been proposed in the literature, with each possessing different strengths and weaknesses. Further research may reveal an *SNR* expression that is more useful than what we use here.

The results of the robust parameter design experiment conducted with the AutoDet-BACON algorithm are shown in Figure 22 and in Figures 74 through 79 in Appendix E. Figure 22 plots the SNR value obtained for each of the 32 combinations of Factors A, B, C, and D. From this plot, it is evident that design points 9 and 11 give the highest SNR values. Design point 9 calls for using band aggregation into 30 features as a data reduction technique, as well as using the 0.9999-quantile threshold with the BACON iteration. Design point 11 is the same as 9 with the additional preprocessing step of standardizing the data. To confirm the validity of these results, we refer to the main effects and interaction plots displayed in Figures 74 through 75. The main effects plot in Figure 74 confirms that band aggregation into 30 variables (level 2 of the Features factor)

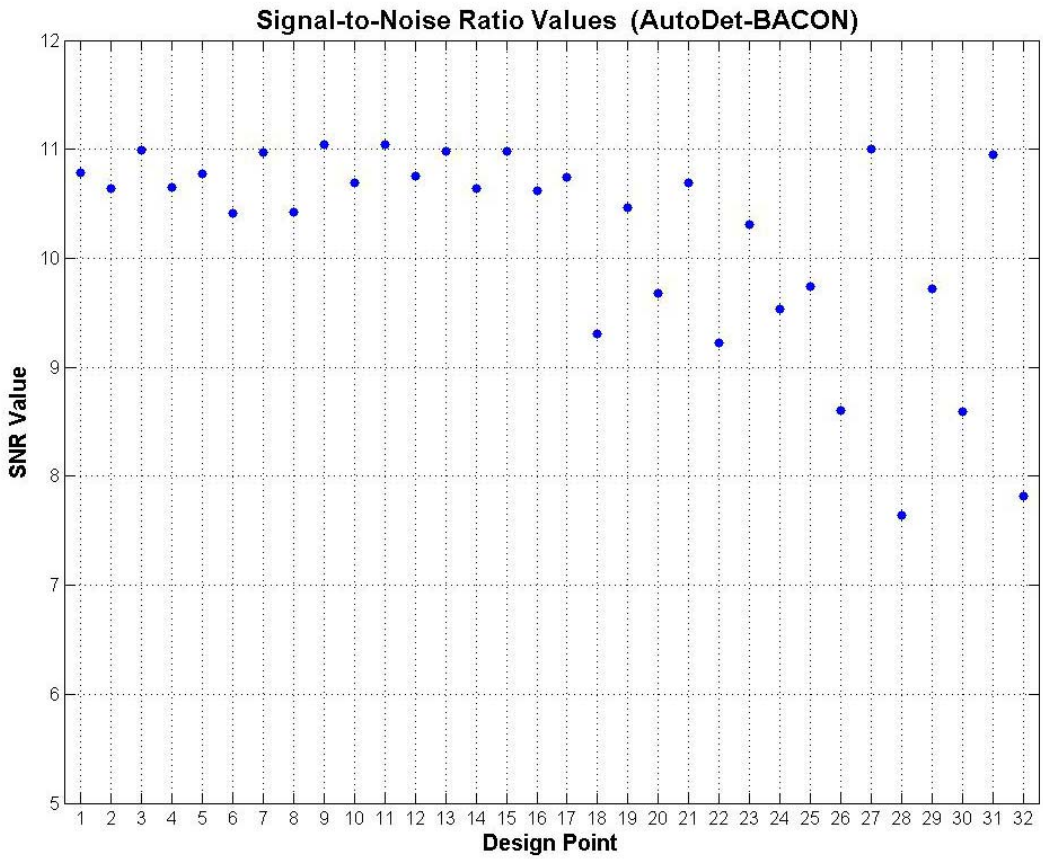


Figure 22. SNR Values for AutoDet-BACON Taguchi Experiment

produces the highest mean value of the response variable for the four feature reduction methods tested. We also see that using the 0.9999-quantile threshold is the better option of the two levels tested. A somewhat confusing result is the decrease in the mean response when the data is standardized. This fact would seem to conflict with the SNR plot; however, inspection of the interaction plot in Figure 77 shows that standardizing the data slightly reduces the variability of the response. This reduced variability is sufficient to make design point 11 attractive from an SNR perspective even though the mean response across the levels of noise decreases when standardization is used. Another note in this regard, however, is that the interaction plots in Figure 75 indicate that

standardization provides a slight improvement in the mean response when band aggregation into 30 variables and the 0.9999-quantile threshold are used—which are the settings for design point 11.

Based on the preceding discussion, there is no compelling reason to choose design point 9 over design point 11, or vice versa, as the preferred configuration of AutoDet-BACON. However, in the interest of simplifying further experiments with the algorithm, we use design point 9 as the configuration of AutoDet-BACON for the remainder of this dissertation.

We now turn our attention to the robust parameter design of AutoDet-FAST-MCD. The factors and levels used for this experiment are identical to those used for AutoDet-BACON with the exception that the threshold factor is omitted. Thus, the experimental design consists of a full factorial design in the normalization, standardization, and features factors nested with the six levels of the noise variable. The 96 design points constituting this experimental design are listed in Table 41 of Appendix D, while the SNR, main effects, and interaction plots produced from the experiment are contained in Appendix E.

The SNR plot shown in Figure 23 points to similar conclusions as before in that band aggregation into 30 variables with or without standardization produces the best results. These configurations are represented by design points 5 and 7. As with the AutoDet-BACON algorithm, the main effects and interaction plots indicate that, on average, standardization tends to reduce detection accuracy across the tested images, but when combined with band aggregation into 30 variables, it produces a slight improvement. Based on these results, there is again no strong argument for using design

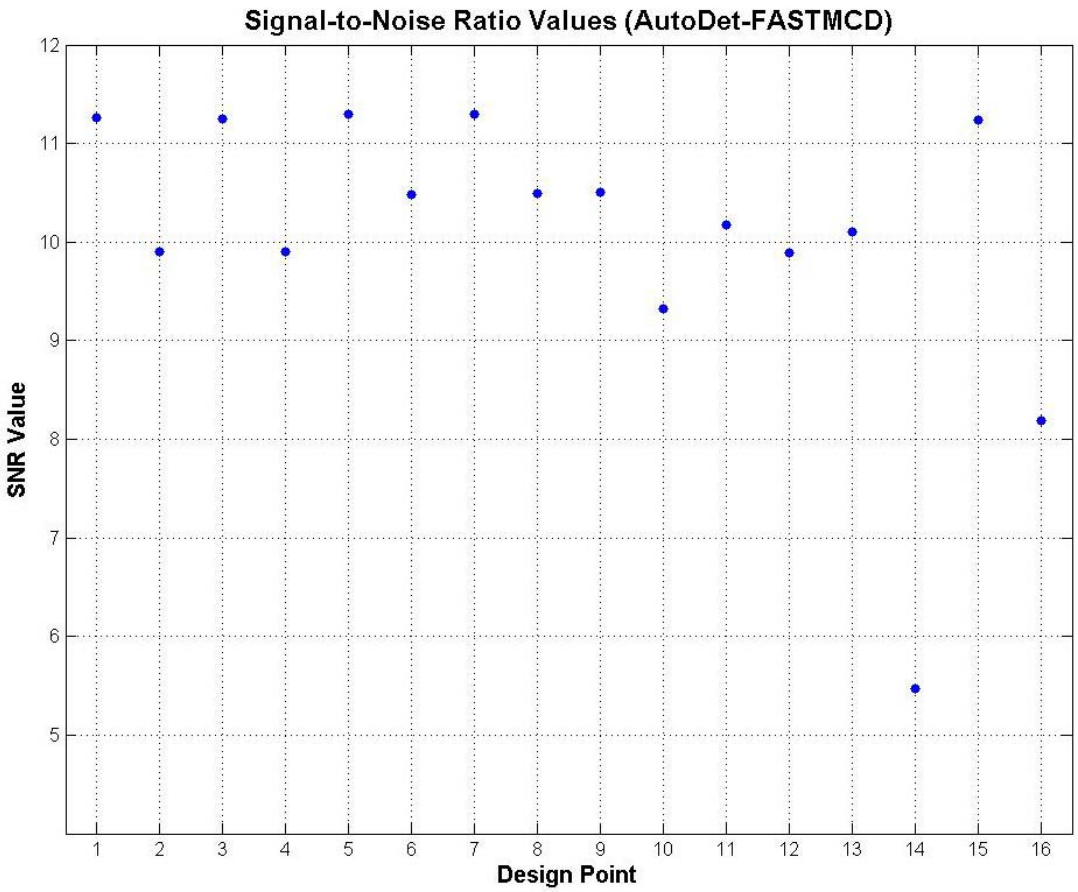


Figure 23. SNR Values for AutoDet-FASTMCD Taguchi Experiment

point 5 as opposed to design point 7 for the AutoDet-FASTMCD configuration. To be consistent with AutoDet-BACON, however, we use design point 5.

Comparison Tests

In the preceding sections we introduced the AutoDet anomaly detection methodology and discussed how Taguchi robust parameter design was used to configure the detector to provide good classification accuracy across a range of images. In this section we compare the performance of the AutoDet-BACON and AutoDet-FAST-MCD algorithms to each other and to the Sub-Space RX algorithm (SSRX) currently used for

real-world anomaly detection applications, and to a cluster-based anomaly detector similar to that proposed by Carlotto (2005). Through this comparison, we show that the AutoDet methodology outperforms these two benchmark detectors when applied to a range of different images from those used in the Taguchi experiment. We begin this discussion with a description of the images used for the comparison, followed by a brief description of the two benchmark detectors. We then outline the test procedure used to compare the detectors, and conclude with a summary of the comparison test results.

Seven images were used to perform the comparison tests, and can be found in Appendix C. The images are referred to as Scenes 5, 6, 7, 12, 13, 17, and 19 based on a larger collection of images to which they belong. Scenes 5 and 6 are derived from the Desert Radiance II dataset, and both contain target panels composed of different materials. The primary difference between these two scenes are the addition of several larger targets and vegetation clutter in Scene 5. Scene 7 is also from the Desert Radiance II datasets, but contains several vehicle and other manmade targets. This scene also has a considerable amount of vegetation clutter in both the upper and lower portions of the image. Scenes 12, 13, and 17 are derived from the Forest Radiance I dataset and contain the same targets as Scenes 2 and 4 used in the robust parameter design experiments. However, Scenes 12, 13, and 17 were acquired at a considerably higher altitude then Scenes 2 and 4, and therefore contain more sub-pixel targets. Additionally, the targets in Scene 17 are placed closer to the tree-line than in the other images in an effort to make them more challenging to detect. The final image, Scene 19, was provided by the Air Force Research Lab and contains several partially concealed targets with background materials consisting of trees, bare soil, asphalt, concrete, and other natural clutter.

Collectively, these scenes represent a range of target types, target sizes, background materials, scene clutter, acquisition altitudes, and collection sensors which provide considerable challenges for all the algorithms tested.

As stated previously, the benchmark detectors to which AutoDet is compared are the Sub-Space RX detector described in Schaum (2004), and what we refer to as the Cluster Based Anomaly Detector (CBAD) which is similar to a detector by the same name originally described in Carlotto (2005). The SSRX algorithm is essentially the RX detector described in Chapter 3 that is applied to the hyperspectral dataset after principal component data reduction. Schaum suggests that the SSRX detector works best when applied to the lower variance principal component variables; however, our initial tests with this detector gave better performance using the first 10 principal components to reduce the dataset; therefore it is that implementation of SSRX we used for the comparison test. To give the SSRX detector a fair chance at detecting a range of targets, we used both 21x21 and 41x41-pixel local processing windows. We refer to these two versions of SSRX as SSRX-21 and SSRX-41.

Our version of the CBAD detector is essentially the same as AutoDet. The only difference between the two methods is that CBAD computes classical MSDs for each observation in a cluster as opposed to the robust MSDs computed with BACON or FAST-MCD. Thus, our CBAD detector can be viewed as a non-robust version of AutoDet, and therefore provides a good benchmark by which to measure the benefits of using robust outlier detection methods to find anomalies. It should be noted that our CBAD method differs from the original method given by Carlotto in two ways. First, we use the Color Method to automatically estimate the number of clusters in the image,

whereas Carlotto's implementation requires a manual selection of *k*. Second, we use the *k*-means algorithm to perform the clustering as opposed to a quantization-based method used in the original detector. Carlotto's preference for the latter method is based primarily on computational speed, and no information is given on the method's robustness to outlying observations.

The actual comparison test procedure used to compare AutoDet-BACON, AutoDet-FAST-MCD, SSRX-21, SSRX-41, and CBAD is relatively simple. Each method was applied to each of the seven images. For a given image, we then used the resulting MSDs from each detector along with the image truth mask to construct OC curves for the detectors. The range of false-positive fractions used in the OC curves is [0.0, 0.01] because false-positive fractions beyond this range generally have no operational value. In fact, our research goal is to maximize the true-positive fraction at a false-positive fraction of 0.001. It should also be noted that only the portion of an image that was analyzed by the SSRX-41 detector was used to generate the OC curves for all the detectors—a disadvantage of local processing methods is they cannot analyze the pixels at the periphery of an image.

In addition to the OC curves, we also produced grey-scale images of the MSDs produced by each detector, as well as *target images* showing the location of anomalies determined by applying a threshold to the MSD images. For the AutoDet-BACON, SSRX-21, SSRX-41, and CBAD detectors, the MSDs were compared to the 0.99999-quantile of the Chi-square distribution with the appropriate degrees of freedom to produce the target images. For the AutoDet-FAST-MCD detector, the slope method was used to determine an MSD threshold, as described earlier in this chapter. Though these

images are illustrative of the effectiveness of each detector in finding targets, we do not regard them as a primary indicator of performance because of the subjective nature of the thresholds used. That is to say, raising or lowering the MSD threshold that produces the target images will give different impressions of detector performance, and it was not practical for us to generate a number of these images for each detector to effectively use them as a comparison metric.

It should also be noted that in comparing the AutoDet-BACON, AutoDet-FAST-MCD, and CBAD detectors, the exact same clustering solution was used for each method. Therefore, the possibility of *k*-means generating different solutions for each method was removed as a possible confounding factor for the experiment, making it easier to judge the benefits of each method.

The seven OC curves produced by the comparison experiment are shown in Figures 24 through 30, and the MSD and target images for each detector are contained in Appendix F. Upon inspection of the OC curves, several conclusions are evident. First, either AutoDet-BACON—referred to as AutoDet in the OC curves—or AutoDet-FAST-MCD outperform the benchmark detectors in all images tested. Second, AutoDet-BACON exhibits better OC curve performance than CBAD for all tested images, clearly demonstrating the benefit of using robust mean vector and covariance matrix estimates to compute MSDs for anomaly detection. Third, though the AutoDet-FAST-MCD method was the best detector for Scenes 7 and 13, and the second best detector for Scenes 6 and 12, it also demonstrated the capacity to be less accurate then the non-robust benchmark detectors when applied to Scenes 5, 17, and 19. It is our hypothesis that this fluctuation in performance is due to the number of initial subsets used in the FAST-MCD nesting

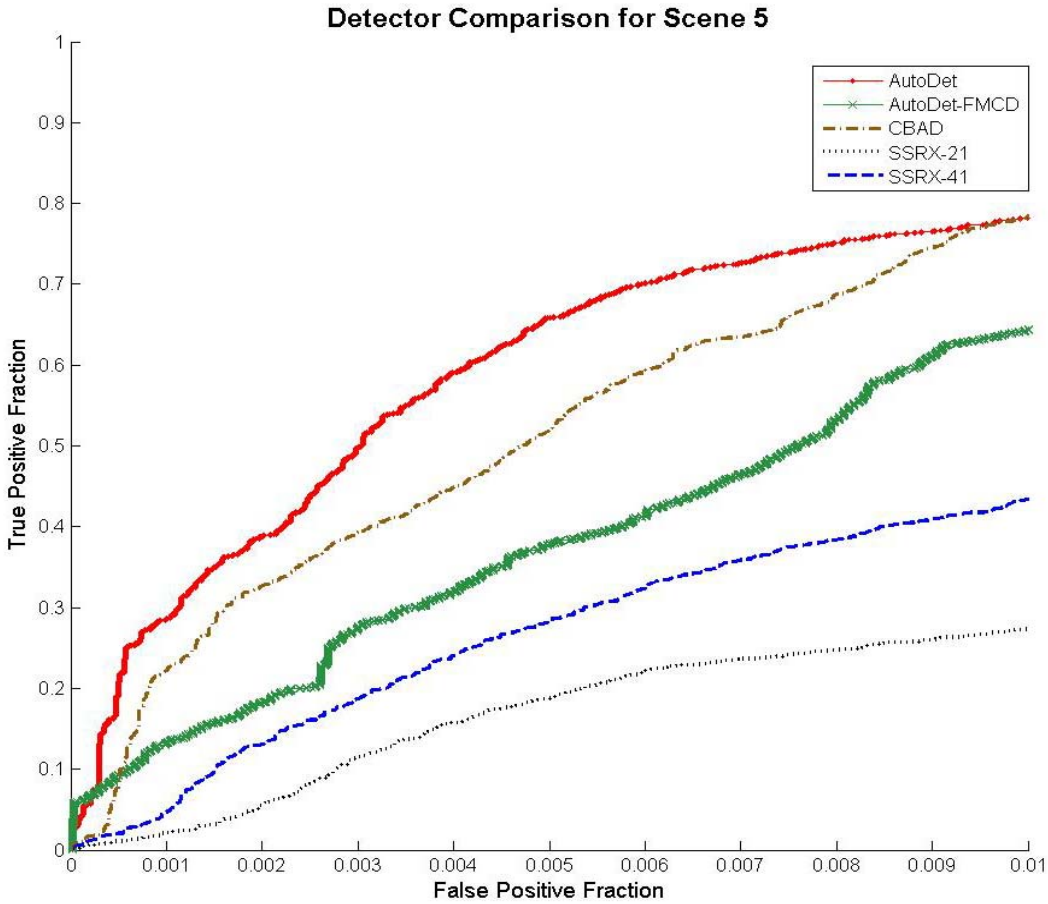


Figure 24. Operating Characteristic Curves for Detector Comparisons (Scene 5)

procedure. In our tests, we set this value to 100 subsets. Further experimentation with this parameter may lead to a setting that gives more consistent detection performance.

A final conclusion drawn from the OC curves is that the AutoDet detectors struggle to obtain good OC curve performance—which we subjectively determine to be a true-positive fraction of at least 0.60 at a false-positive fraction of 0.001—for Scenes 5, 7, and 17. Our analysis of the false alarms generated for these scenes indicate that they are caused predominantly by *natural anomalies* that are anomalous relative to the major background materials in the respective scenes, but are not the manmade targets we are

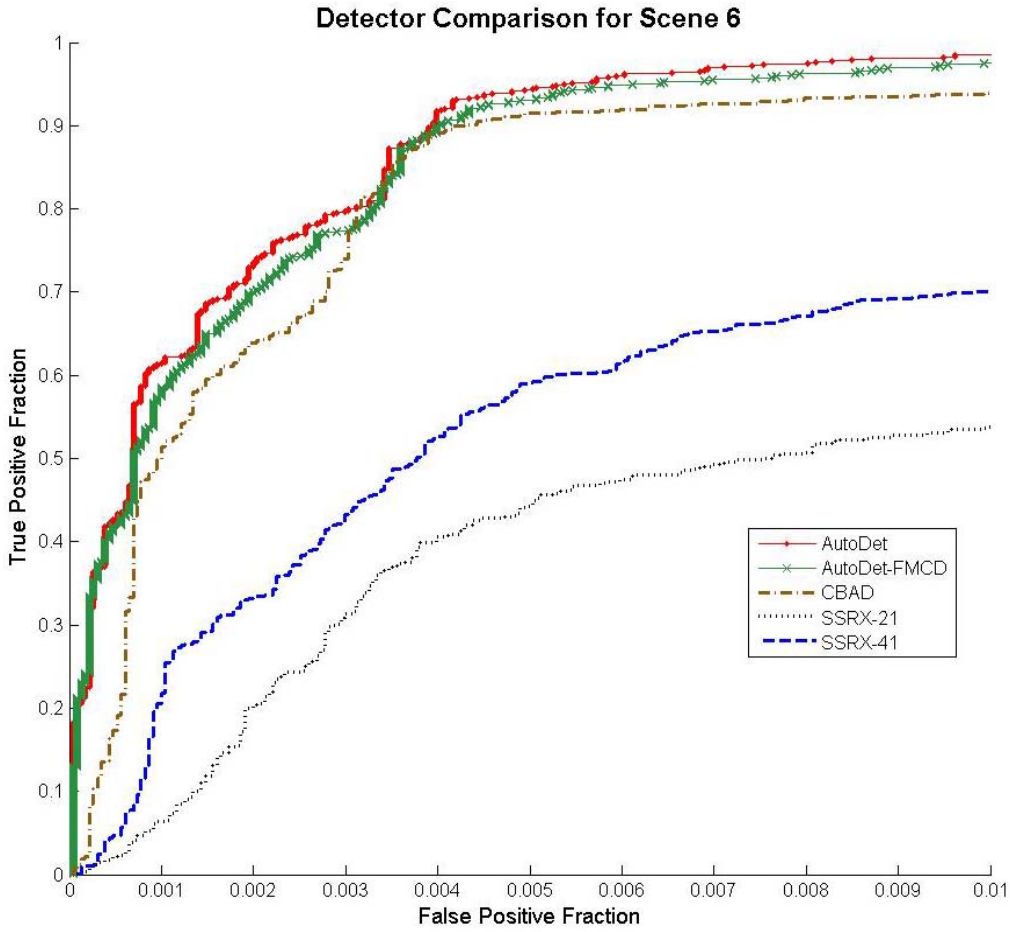


Figure 25. Operating Characteristic Curves for Detector Comparisons (Scene 6)

interested in detecting. A clear example of these natural anomaly false alarms can be seen in the AutoDet-BACON MSD image for Scene 5 shown in Figure 86. In the upper left corner of this image, it is clearly evident that the algorithm assigned high MSD values to the sparse vegetation contained in the scene. The reason the vegetation is considered anomalous is the Color Method estimated the presence of only one background cluster for this scene—namely, soil. Thus, the MSDs for the vegetation spectra were computed using the mean vector and covariance matrix for the soil, resulting in expectedly high values. A confirmation of this hypothesis is given by the results from Scene 6. This scene is essentially the same as Scene 5, with a notable difference being the absence of the vegetation clutter. As seen in the OC curves for Scene 6, removal of this clutter leads to significantly better performance for all the detectors evaluated.

In addition to the OC curves, the MSD and target images produced in the comparison test also offer insights into how the different detectors perform. In both the MSD and target images it is clear that a major strength of the AutoDet and CBAD methods is their ability to detect large anomalies, which is expected based on the discussion of global anomaly detectors in Chapter III. It is evident that using a larger processing window improves SSRX detection in this regard, but comes with the obvious cost of being able to analyze less of the image and increasing the chance that the processing window will contain non-homogenous materials.

We can see the effect that non-homogeneous, or contaminated, window data has on the SSRX detector by inspecting the Scene 5 MSD image for SSRX-21 shown in Figure 86. In this image, we notice a very strong anomaly located in the left, upper-middle region of the image that is surrounded by a band of dark pixels.

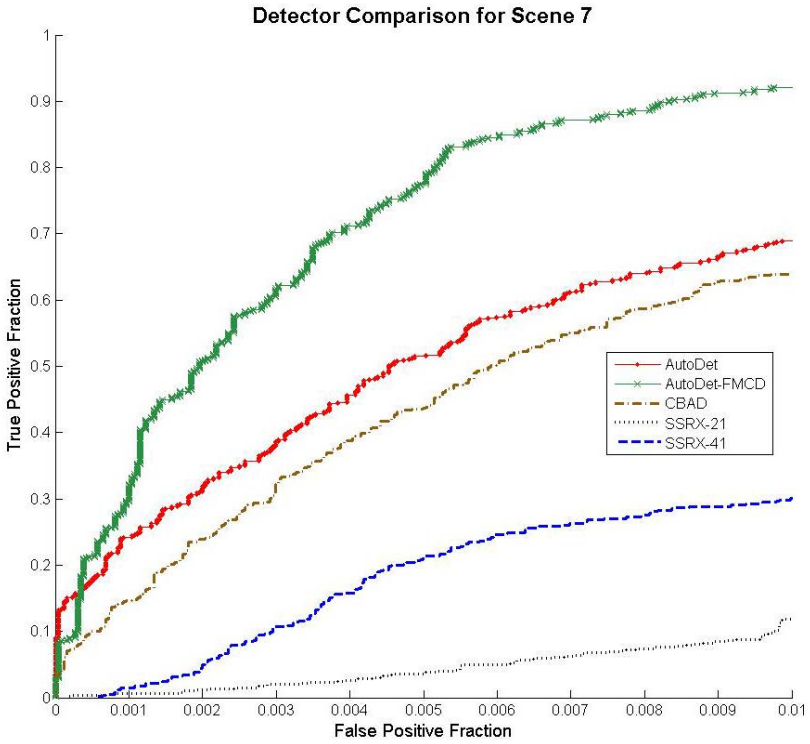


Figure 26. Operating Characteristic Curves for Detector Comparisons (Scene 7)

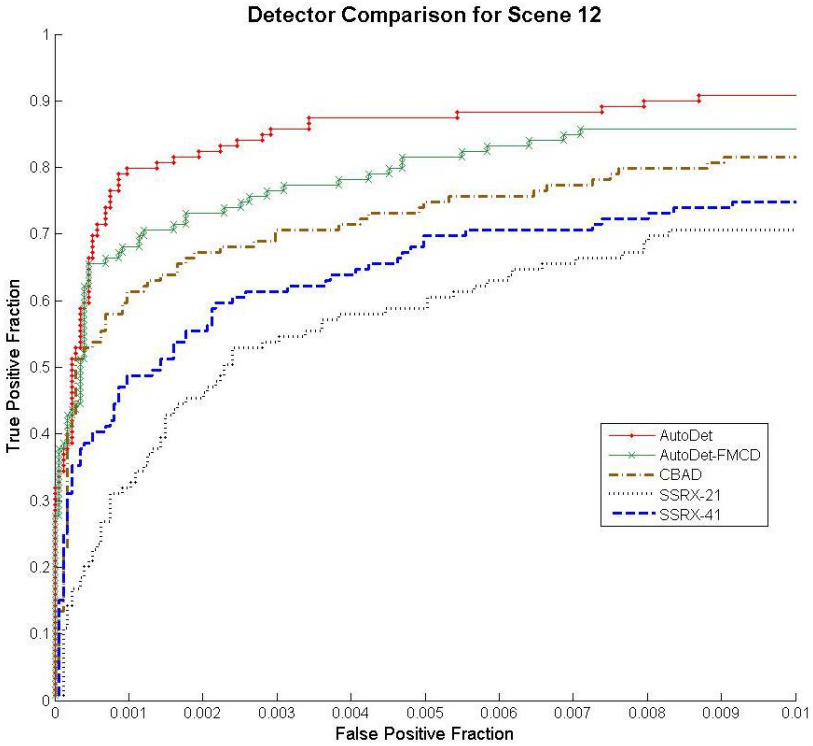


Figure 27. Operating Characteristic Curves for Detector Comparisons (Scene 12)

This *shading* of pixels in the immediate vicinity of anomalies can also be perceived to a lesser degree with other anomalies in the image—both manmade and natural. This shading effect is caused by distorting effects the anomaly has on the window statistics of its neighboring pixels. In particular, the anomalies tend to inflate the measured variance of the window pixels, thereby decreasing the MSDs of the *shade* pixels. Though this distortion effect is most obvious in Figure 86, it occurs in varying degrees whenever a processing window overlaps multiple materials, as also noted by Schaum (2006).

A final observation we make concerning the MSD images is that all the detectors are susceptible to false alarms caused by sensor artifacts. In the images used for the comparison tests, these artifacts take the form of vertical stripes running through the images, and are most evident in Figures 86, 90, and 92. In general, these artifacts are band-dependent, which means they can be eliminated by simply removing the effected band from the dataset. In more severe instances in which many bands are affected, image smoothing may be an appropriate remedial action, though using these methods could sufficiently alter the spectral signatures in the image such that the distinctions between target and background spectra are blurred.

Based on the preceding discussion, we conclude that the AutoDet methodology, and the AutoDet-BACON detector, in particular, is a superior alternative to the benchmark detectors for hyperspectral anomaly detection. Moreover, this superiority is achieved through the combined benefits of robust MSD estimation as well as global anomaly detection. Though we consider AutoDet to be an improvement over existing

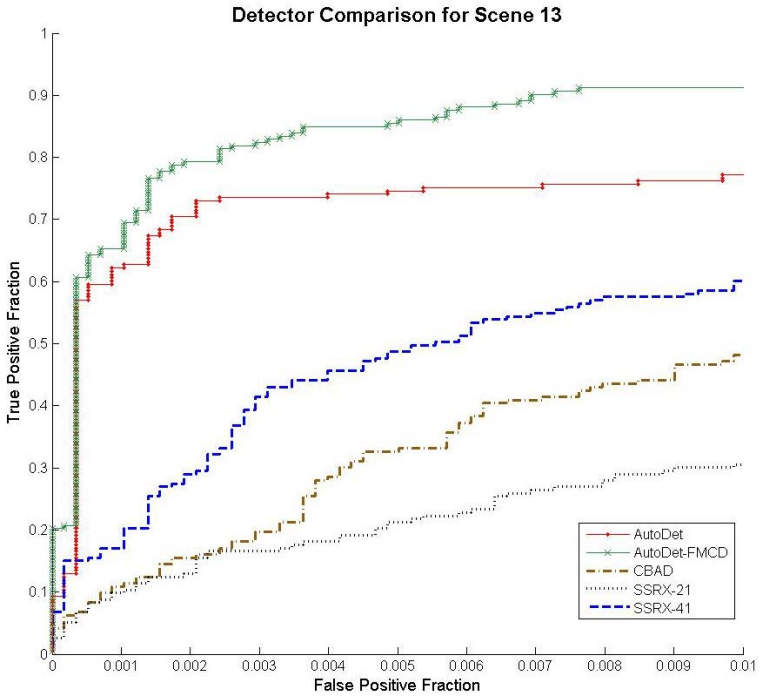


Figure 28. Operating Characteristic Curves for Detector Comparisons (Scene 13)

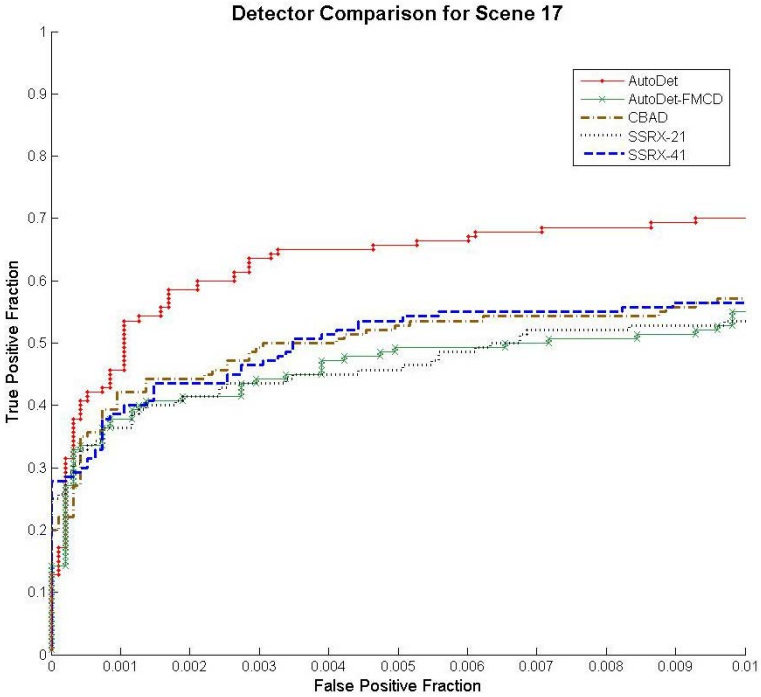


Figure 29. Operating Characteristic Curves for Detector Comparisons (Scene 17)

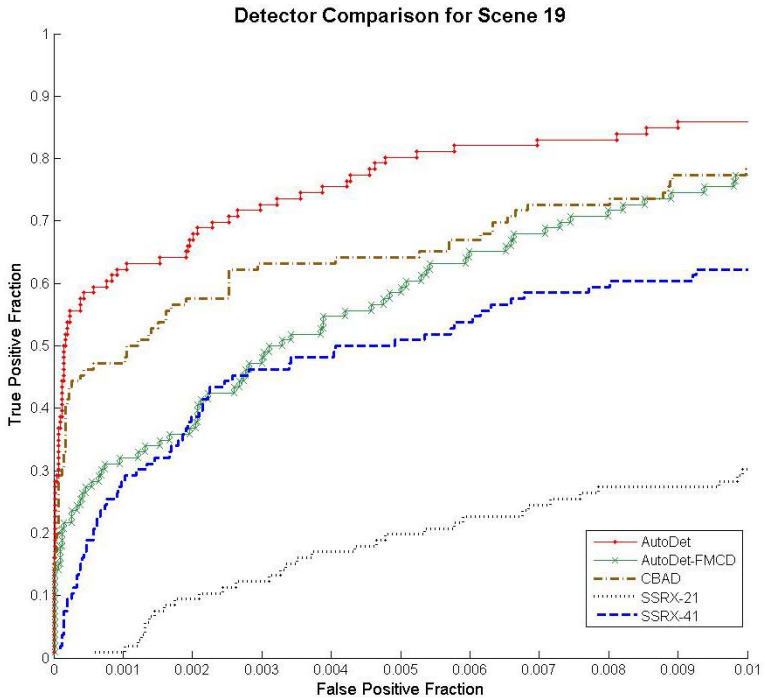


Figure 30. Operating Characteristic Curves for Detector Comparisons (Scene 19)

anomaly detectors, the method is not without limitations. We address these limitations further in the following section.

Limitations of AutoDet

We have demonstrated through experimental tests with actual hyperspectral images that AutoDet can be an effective method for hyperspectral anomaly detection. However, our experience with the method has also revealed several areas in which AutoDet struggles. Specifically, we recommend further research to improve AutoDet's clustering methodology, its ability to deal with targets that are outlying in only a small subset of dimensions, and in the method used to threshold MSDs for anomaly detection. We address each of these areas in the following paragraphs.

Clustering Limitations. As discussed earlier in this chapter, clustering an image into groups that represent the major background materials in the scene is a key

component of the AutoDet methodology. In its current implementation, AutoDet uses the Color Method and *k*-means to accomplish this task. As we have seen, this process works reasonably well for a range of images, but it is not perfect. Specifically, if a particular background material constitutes a relatively small percentage of the image, it can either be grouped with another background material, or with a group of anomaly pixels. In the former case, the less predominant material is likely to be considered an anomaly relative to the more prevalent material, thus leading to the natural anomalies encountered in the comparison tests. In the latter case, it is possible the background spectra and anomaly spectra are proportional in frequency, in which case the background spectra may be declared as anomalies while the anomaly spectra are declared as background. Along similar lines, it is also theoretically possible, via the clumping effect discussed earlier in this chapter, that strong anomalies are placed in their own cluster. Should this phenomenon occur—which we have yet to experience in practice—the AutoDet method would completely fail to detect the anomalies.

As potential solutions to these clustering problems, the following options may prove useful. In terms of minimizing natural anomalies, the use of spatial information describing the shape of the anomaly may help to better distinguish manmade objects such as vehicles from natural objects such as bushes or patches of dirt. To counter the problem of anomalies being clustered in their own group, it may be useful to again use spatial information to post-process unusually small clusters to determine if they exhibit target-like characteristics. For example, if the spectra in a small cluster form a number of spatially-connected groups similar in size to targets of interest, this might be enough evidence to designate the spectra as targets.

Marginally Outlying Targets. A basic strategy of AutoDet is to use cluster analysis to group targets with the background material to which they are most similar, and then use BACON or FAST-MCD to separate the targets from background spectra. In many instances, target spectra are sufficiently different from the background spectra in their assigned cluster such that this strategy works well. However, some target material spectra we encountered differed from background spectra in only a small subset of dimensions. These marginally outlying target spectra pose two problems. First, the more similar a target spectra is to the background spectra, the more likely it is to be included in the mean vector and covariance matrix computed by BACON or FAST-MCD, thereby reducing the effectiveness of these algorithms.

The second problem encountered with marginal outliers that may only be outlying in a small subset of dimensions is that they exploit a fundamental flaw with Mahalanobis Squared Distance detectors. In particular, setting the threshold of a MSD detector so that it can detect outliers in a small subset of dimensions implies that it will also designate as outliers those spectra that are shifted a relatively small distance from the background mean vector in all dimensions. This problem is significant since the spectral variability of background materials in hyperspectral imagery tends to manifest itself as shifts in the background mean spectra caused by changes in material illumination. As a simple example of this problem, suppose we have a cluster of background and target spectra in dimension p=30, and that these p dimensions are uncorrelated. Further, suppose the background spectra are distributed as $N(\mathbf{0},\sigma_{ii}\mathbf{I})$, and that the target spectra only deviate significantly from the background mean spectra in the last three dimensions. Thus, we want our MSD detector to be able to find a target vector, $\mathbf{t}$, that has zeros in the first 27

elements and values that are three standard deviations from the background mean spectra in the last three elements—assuming a value of three standard deviations from the mean is considered an outlying measurement. In such a scenario, the MSD of the target vector would be

$$\begin{aligned} MSD(\mathbf{t}) &= (\mathbf{t}-\mathbf{0})^T \Sigma^{-1} (\mathbf{t}-\mathbf{0}) \\ &= \sum_{i=1}^{27} \left(\frac{0}{\sigma_{ii}} \right)^2 + \sum_{i=28}^{30} \left(\frac{3\sigma_{ii} - 0}{\sigma_{ii}} \right) \\ &= 9 \end{aligned} \tag{5.50}$$

and we would declare any observation with an MSD of 9 or higher as an outlier. However, under this rule we would also declare as an outlier any observation vector, **b**, that is only 0.54 standard deviations from the background mean vector in all 30 dimensions, since *MSD*(**b**)=9. Therefore, in our effort to detect targets with signatures similar to **t**, we are also likely to generate false alarms from pixels with signatures similar to **b**.

To help mitigate the problem of marginal outliers contaminating the robust mean vector and covariance matrix computed by BACON or FAST-MCD, the criteria for allowing observations to be used in these computations can be tightened. In the case of BACON, this action entails using a lower quantile of the Chi-square distribution in forming the basic subset, which was shown to improve detector performance in the preceding robust parameter design experiments. For FAST-MCD, it may be possible to use the smaller half-sample size of h=[(n+p-1)/2] as opposed to h=0.75n to decrease the chance of marginal outliers being included in the final half-sample, though we did not conduct any research to confirm this hypothesis. In regards to detecting outliers that are outlying in a small number of dimensions without generating mean-shift false alarms, it

may be helpful to perform a univariate outlier detection on each dimension in addition to evaluating MSDs; however, for such a method to be practical, it would also be useful to objectively determine the extent to which an observation must be outlying in a subset of bands for it also to be considered a multivariate outlier. If this latter methodology is not developed, then bad sensor readings in a single dimension may unnecessarily trigger false alarms.

MSD Thresholds. A fundamental problem faced by hyperspectral anomaly detection algorithms is determining the proper threshold for a detectors output metric in order to distinguish anomalies from background spectra. The AutoDet methodology currently has no advantage over other detectors in this regard. In the case of AutoDet-BACON, a Chi-Square threshold is used to separate anomalies from background, and it generally provides reasonable results despite evidence in the technical literature that hyperspectral data is typically not Gaussian. In other words, the data is close enough to being Gaussian that modeling the BACON-derived MSDs with a Chi-square distribution can still provide useful results, though there is a clear risk in assuming good results will always be obtained. For the AutoDet-FAST-MCD MSDs, the Gaussian assumption is more clearly violated because the method discards the tails of the data, as mentioned previously. Thus, we are left contemplating the distribution of MSDs produced from a truncated, elliptically contoured multivariate distribution. To avoid estimating the form of such a distribution, we instead choose to use our slope method to look for signs of anomalies in the empirical distribution of FAST-MCD MSDs. Though this method works in some instances, it can also produce relatively poor results, as seen in the AutoDet-FAST-MCD target images produced in the detector comparison tests. As an

example, we see in Figure 26, that AutoDet-FAST-MCD had the best OC curve performance for Scene 7, but the respective target image in Figure 89 would seem to indicate it performed no better than AutoDet-BACON or CBAD.

As alternatives to blind faith in using BACON's Chi-Square threshold and to using the slope method for FAST-MCD, we suggest developing more accurate methods for estimating the MSD distributions for these two detectors when applied to hyperspectral images. A starting point in this endeavor is the research conducted by Meidunas (2006) which offers several possibilities for modeling hyperspectral data MSDs. The challenge in using Meidunas' distribution-fitting methods, however, is implementing them in an autonomous fashion so as to maintain the autonomy of the AutoDet methodology.

k-Selection Methods. We indicated earlier in this chapter that the Color Method was preferred over the other *k*-selection methods we tested because it is fast and produces better results with no data pre-processing when applied to actual hyperspectral imagery. However, our tests also indicate that this method has difficulty producing estimates that are the same as manually estimated *k* values. In the six images used to test the Color Method, its estimates were within one cluster of the manual estimate for two images, within two clusters for two images, within three clusters for one image, and within four clusters for the remaining image. Though this performance is certainly not ideal, is it good enough? To shed some light on this matter, we applied AutoDet-Bacon to the same images used in the detector comparison test while varying the number of clusters used in the *k*-means algorithm from one to 12. For each image, we constructed an OC curve

Table 29. Manual and Color Method *k*-Estimates for Comparison Test Scenes

Scene	Color Method *k* Estimate	Manual *k* Estimate	Best Values of *k* for FPF of 0.001 (Listed in descending order by TPF)
5	1	6-7	5, 6, 7, 4, 8, 3
6	1	3-5	5, 4, 3, 1, 7
7	3	4-5	1, 4, 3
12	6	8-9	7, 8, 9, 3, 6, 5
13	9	5-6	8, 1, 2, 6, 3, 9, 4, 10, 11
17	5	6-7	1, 4, 9, 10, 12, 11
19	4	6-8	11, 9, 10, 8

for the detector at each value of *k*. These curves are shown in Figures 31 through 37. Additionally, we manually estimate the number of clusters in each image. These manual estimates are listed in Table 29, along with the Color Method estimates for *k*. Table 29 also provides the value of *k* that resulted in the highest true-positive fraction at a false-positive fraction of 0.001, as well as values of *k* that produce a true-positive fraction within 0.05 of the best *k* value (these values are listed in descending order by true-positive fraction.)

From Table 29, it is again evident that the Color Method does not provide estimates of *k* that are in agreement with manual estimates. However, it is also apparent that when the Color Method is applied to four out of the seven images, the *k* estimate it provides works nearly as well as the best *k* value for the respective image at an FPF of 0.001. Further, if we had used our manual estimates for *k* rather than those provided by the Color Method, we would have only obtained significantly better results in two of the images tested (Scenes 1 and 19). Thus, based on the images used in this test, the Color Method has some value in automating the *k*-selection process and removing this burden

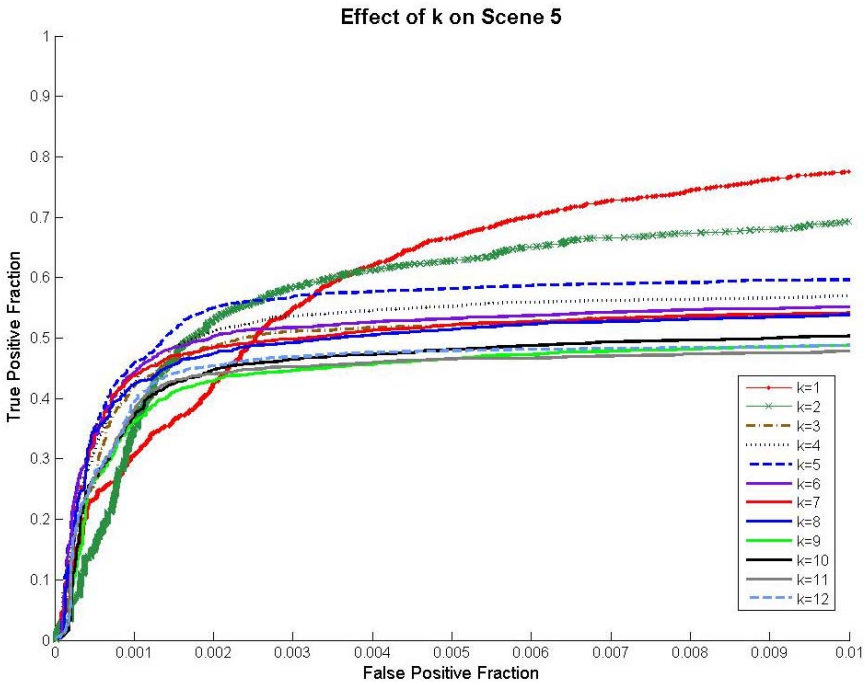


Figure 31. Effect of *k*-Estimate on Scene 5 Detection

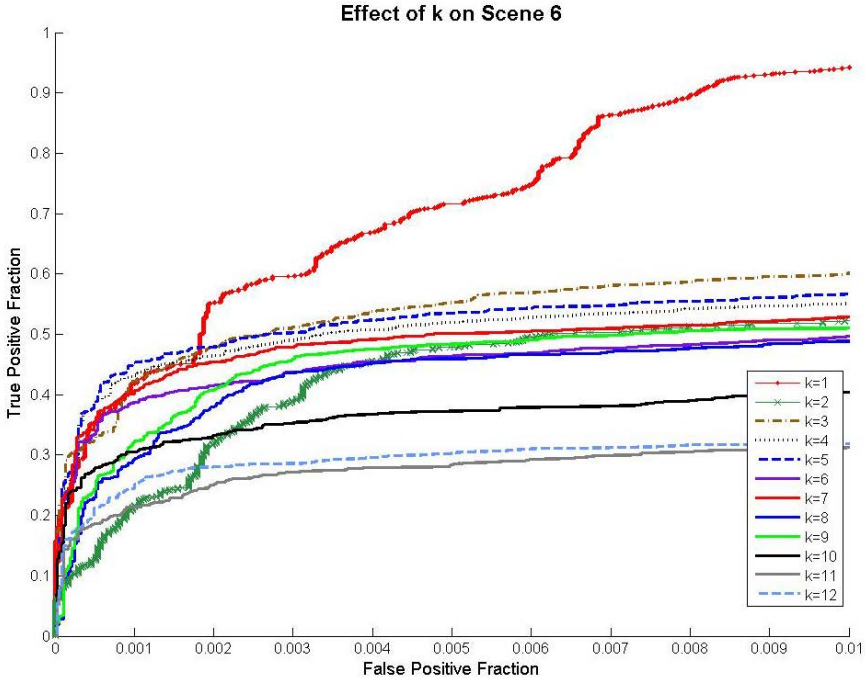


Figure 32. Effect of *k*-Estimate on Scene 6 Detection

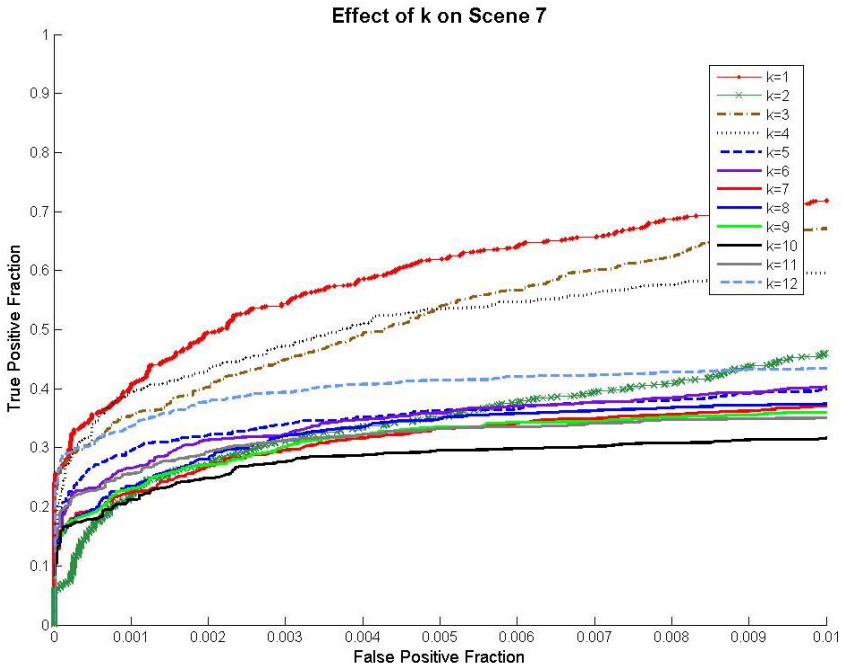


Figure 33. Effect of *k*-Estimate on Scene 7 Detection

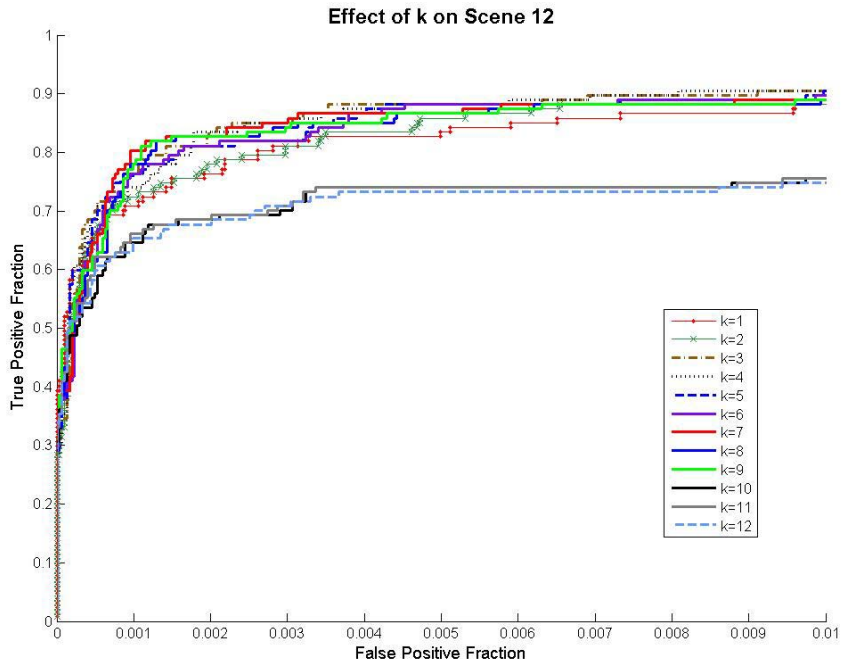


Figure 34. Effect of *k*-Estimate on Scene 12 Detection

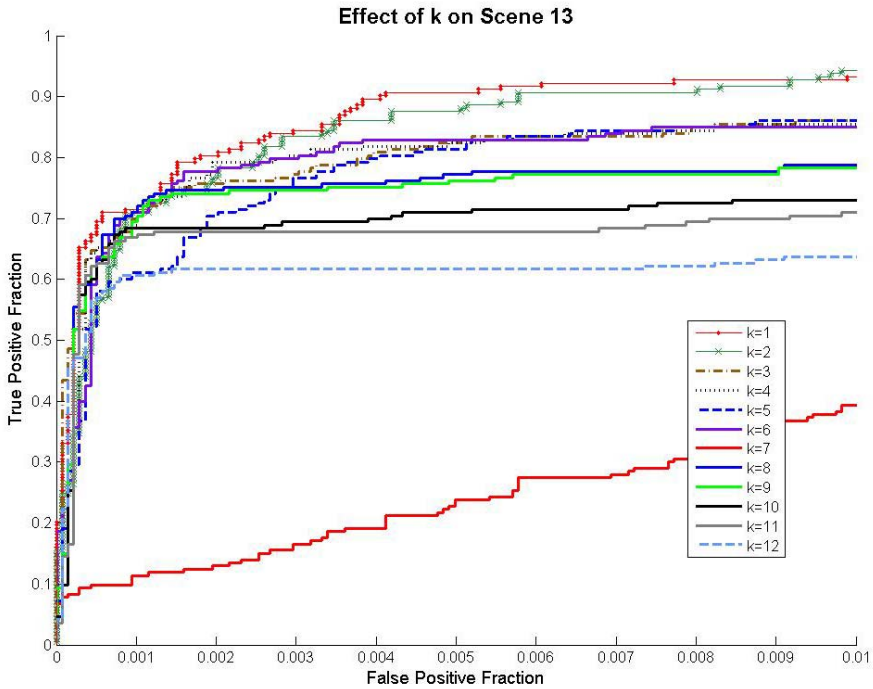


Figure 35. Effect of *k*-Selection on Scene 13 Detection

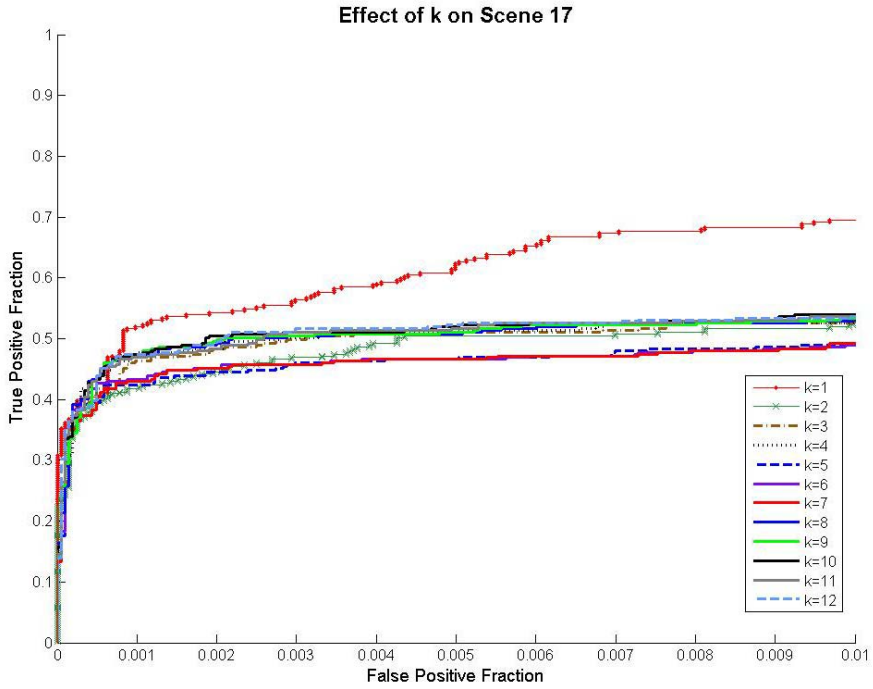


Figure 36. Effect of *k*-Selection on Scene 17 Detection

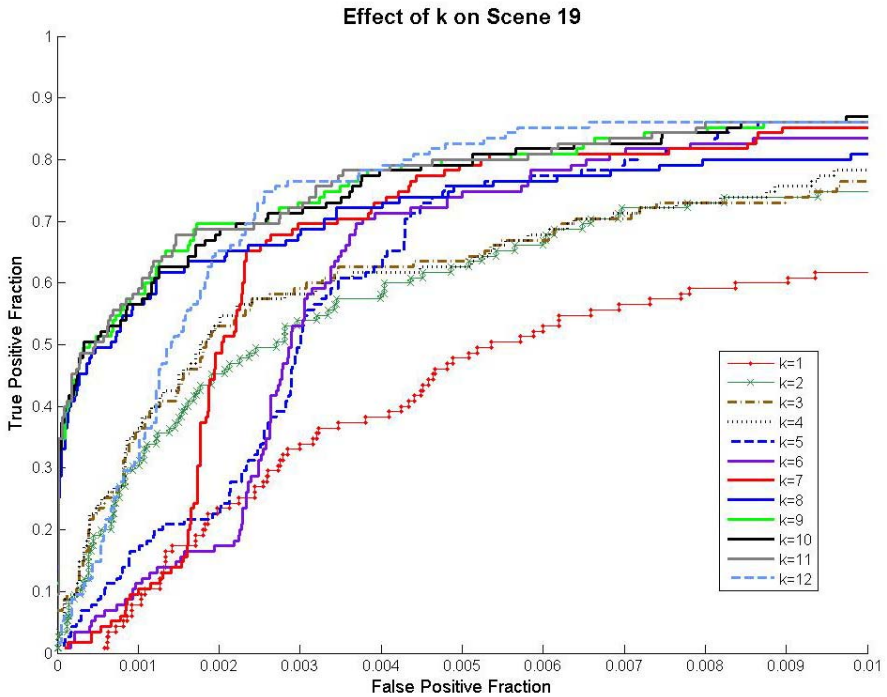


Figure 37. Effect of *k*-Selection on Scene 19 Detection

from the user. Of course, the fact that the Color Method also generated *k* values that are not best for two of the seven images clearly indicates the need for further research to find a better *k*-selection method.

Before closing this discussion on AutoDet limitations, we note several perplexing results in Table 29 and Figures 31 through 37.. First, we note that for four of the seven images, using only one cluster generated very good results, if not the best. Second, we note that for many of the scenes, a fairly large number of *k* values gave nearly the same results. Finally, the k values that worked best for a particular image are not always sequential, as one might expect. For example, the values that worked well for Scene 12 were 3, 5, 6, 7, 8, and 9, but not 4. Though we have not conducted any formal tests to confirm our hypothesis, we believe that these peculiar results are caused by the relative orientation of the outliers to the background data, as well as the nature of the *k*-means algorithm itself. For example, if the outliers are sufficiently *outside* the data cloud

formed by the background observations, the covariance ellipsoid for the pooled background data may still exclude the outliers, thus giving good results with only one cluster. In such a scenario, forcing *k*-means to then divide the background data into a less-than-optimal number of clusters will generate more false alarms since many dissimilar background materials will be assigned to the same clusters. When the correct value for *k* is used, similar background observations are more likely to be clustered together, reducing the number of false alarms. A simple illustration of this suspected phenomenon is shown in Figure 38.

The manner in which *k*-means operates further contributes to counterintuitive results that can be caused by outlier orientation. In our experimentation, for example, we noticed that increasing the value of *k* for the *k*-means algorithm often had the effect dividing high-variance, homogeneous clusters containing a large number of observations rather than dividing low-variance, heterogeneous clusters containing significantly fewer observations. The reason for this phenomenon is that *k*-means simply tries to minimize the total sum of distances to cluster centroids without any regard for cluster structure. Relative to this objective, splitting a large, high-variance cluster is more advantageous than splitting a heterogeneous cluster whose respective sum of distances is relatively low. Thus, it may take a relatively large value of *k* to split some heterogeneous clusters since *k*-means will continue to divide the data from the homogenous cluster until no additional benefit is achieved. If the heterogeneous cluster contains outliers that are masked due to the multiple background materials contained in the cluster, the value of *k* that gives the best OC curve performance may be considerably higher than the number of background

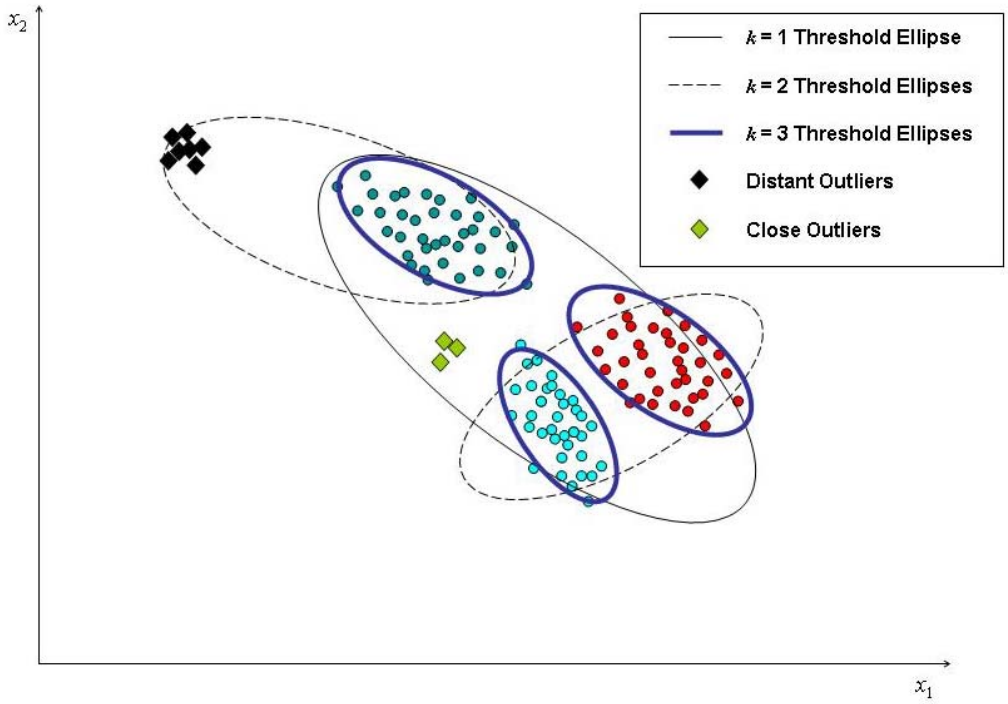


Figure 38. Possible Effect of Different *k*-Values on Outlier Detection

materials actually contained in the image. Additionally, the values of *k* that do nothing other than sub-divide the homogenous data will all produce similar OC curves. We believe that this phenomenon occurred in Scenes 13, 17, and 19, though further tests are required to confirm this hypothesis.

Summary of Conclusions and Areas for Further Research

The purpose of this chapter has been to develop a new hyperspectral anomaly detection algorithm using multivariate anomaly detection methods. In pursuing this objective, we: 1) reviewed the multivariate outlier detection literature; 2) used simulated hyperspectral data experiments to show the problems outliers pose to existing anomaly detectors that use non-robust statistics; 3) evaluated the capability of different multivariate outlier methods to detect outliers in simulated hyperspectral data; 4)

explored the robustness of the *k*-means clustering algorithm to determine its worth as a preprocessing tool for multivariate outlier methods; 5) evaluated different methods for automatically selecting *k* for the *k*-means algorithm; 6) employed Taguchi robust parameter design to effectively join the *k*-means algorithm with the BACON and FAST-MCD detectors to form the AutoDet methodology; and 7) showed that the AutoDet detector is superior to current benchmark detectors for detecting anomalies in actual hyperspectral imagery. In the following paragraphs, we summarize the significant conclusions obtained from this effort.

1) Based on our simulated Multivariate Gaussian tests, outliers can be masked from classical MSD detectors with as little as 2.4% contamination, depending on the background and outlier materials.
2) For background-outlier material combinations that appear to be resistant to masking, covariance matrix distortions are still present. In our principal component axis rotation tests, we showed that the background material's covariance structure can significantly change with as little as 0.5% contamination.
3) In our simulated data tests it was shown that the number of false alarms generated from a classical MSD detector may actually decrease in the presence of outliers due to the artificial inflation of the estimated variance of the data. As contamination levels become large, however, the number of false alarms can be expected to increase.
4) When the background data comes from heavy-tailed multivariate-*t* distributions, our tests indicate that multivariate Gaussian outlier detection

methods are more effective at finding outliers than non-robust MSD detectors, even if the distribution of the MSDs for the multivariate-*t* data is known with certainty.

5) Controlled experiments using simulated Gaussian and multivariate-*t* data indicate that the BACON and FAST-MCD outlier detectors are effective at detecting outliers in datasets with high-dimensionality and large numbers of observations. In the tests we conducted, these detectors found 100% of the outliers. For the BACON detector, these outliers were found with virtually no false alarms and with the least computational effort of the methods tested. The FAST-MCD detector produced significantly more false alarms than the BACON algorithm due to its tendency to underestimate the true variance of the background data.

6) The *k*-means clustering algorithm using the Cosine assignment rule is adequate for accurately clustering hyperspectral data into homogenous groups when the data is contaminated by outlying observations. In our simulated data tests using both Gaussian and multivariate *t*-distributed data, this method accurately clustered data with contamination levels up to 9.1% when outliers are dispersed in the high-dimensional space. In the presence of highly-concentrated, distant outliers, the tolerated contamination level can drop to 3.6% or less, depending on the background data.

7) The primary failure mode of the *k*-means algorithm using the Cosine assignment rule is the clumping effect in which outliers are assigned to their own cluster and background materials are grouped together. The clumping

effect is more likely to occur when: a) outliers are highly concentrated and distant from the background materials; b) background material are spectrally similar to on another; and c) the background clusters closest to the outliers contain relatively few observations.

8) When applied to simulated hyperspectral datasets, the Silhouette and Calinski-Harabasz methods are most effective at estimating the number of clusters in a dataset. This conclusion was verified using datasets with different background materials and a range of known values of *k*.

9) When applied to actual hyperspectral images, our proposed Color Method, which estimates *k* based on the number of colors detected in the visible region of the hyperspectral image, performed comparably to the Silhouette and Calinski-Harabasz methods without the need for any data preprocessing. The Color Method also produced its estimates in fractions of a second, as opposed to minutes for the other two methods. Based on these results, we view the Color Method as more practical in achieving an autonomous anomaly detection method.

10) The AutoDet methodology, which combines the Color Method, *k*-means clustering, and the BACON or FAST-MCD algorithms into a single, autonomous anomaly detection method, out-performs the SSRX and CBAD detectors when applied to a range of hyperspectral images, as indicated by OC curve analysis.

11) Comparisons between the AutoDet-BACON and CBAD detectors clearly indicate the benefit using robustly estimated mean vectors and covariance

matrices in an MSD detector to more accurately detect hyperspectral anomalies.

12) The primary limitations with the AutoDet methodology are: a) its ability to detect marginally outlying targets; b) inconsistent estimation of the number of clusters in an image, leading to degraded detection accuracy; c) the potential for degraded performance due to the effect natural anomalies have on the clustering methodology; and d) a limited theoretical basis for accurately specifying a critical value that can be used to threshold the MSDs from the BACON and FAST-MCD detectors.

To remedy the limitations of the AutoDet methodology, we recommend further research in the following areas:

1) Develop alternative methods to the *k*-means algorithm for robustly clustering hyperspectral images. Implementation of the method described by Hardin and Rocke (2004) may be a first step in this direction; however, this method may prove to be too computationally intensive for hyperspectral images.
2) Develop improved methods for estimating the number of clusters in an image. Enhancing the Color Method so that it is more resistant to the scaling problems mentioned earlier in this chapter may improve its accuracy while maintaining the considerable speed advantage it has over statistically-based methods.
3) Study the distribution of MSDs generated by the BACON and FAST-MCD methods in order to better threshold these MSDs for anomaly detection. As an alternative, fast Monte Carlo simulation-based methods should also be

explored for estimating image-based thresholds *on the fly*. Starting points for this research are Meidunus (2006) and Hardin and Rocke (2005).

4) Investigate methods for detecting marginal outliers that are only outlying in a small subset of dimensions. As stated earlier, these types of outliers are difficult to detect using MSD detectors because setting the MSD threshold low-enough to reveal them will increase the false alarms generated by spectra shifted a small amount from the background mean vector.
5) Expand the Taguchi experimental design used to configure the AutoDet detector to include more images and additional algorithm parameters. For the FAST-MCD detector, such parameters may include the half-sample size and the number of starting subsets.
6) Explore the use of spatial information to screen anomalies that are likely to be natural clutter as opposed to manmade targets of interest. Use of phenomenology information for manmade materials may also be useful in this screening process.

Of course, even if these research efforts are fruitful, there will always be a limit on the target detection accuracy an anomaly detection method can achieve, simply because hyperspectral images are likely to contain many non-target objects that also occur with low frequency in the image. In a military context, the use of decoys can further complicate this problem. Thus, as proposed in the target detection framework defined in Chapter 1, anomaly detection should be augmented with signature matching techniques to further improve the chance of detecting targets of interest. In the following

chapter, we pursue this idea further by developing an autonomous signature matching algorithm to complement the AutoDet methodology.

VI. Signature Matching using In-Scene Calibration

In the previous chapter, we developed the AutoDet methodology as an autonomous anomaly detection method that performs well on a range of hyperspectral data with minimum user input. However, it is also evident that natural anomalies and targets that are outlying in only a small subset of dimensions are problematic for AutoDet. This limitation is not unique to AutoDet, but rather is a fundamental weakness of anomaly detection, in general. Thus, the detection of anomalies in hyperspectral data is often one piece of the target detection puzzle, and must often be combined with other information to locate targets of interest. Specifically, as our target detection framework in Chapter 1 indicates, we believe that the fusion of anomaly detection and signature matching analysis leads to a superior target detection methodology than either of the two methods used independently. To this end, we now turn our attention to developing a signature matching methodology to complement the AutoDet anomaly detection method. As with AutoDet, it is our objective to construct a signature matching method that is as autonomous as possible in the sense that minimum information and interaction is required from the user, thus making the methodology accessible to a wide range of operational users. In particular, we are interested in developing a signature matching algorithm that removes from the user the burden of atmospheric calibration and all the scene-specific knowledge of viewing geometry and atmospheric conditions that it requires.

As indicated in the literature review of Chapter 4, the quest for signature matching algorithms that minimize the complications imposed by atmospheric calibration is not new. However, the common thread winding through the invariant methods of Healey and Slater (1999), Pan, Healey and Slater (2000), Suen, Healey and Slater (2001), Thai and

Healey (2002), Liu and Healey (2004), Bajorski, Ientilucci and Schott (2004), and Bajorski and Ientilucci (2004), is the requirement to use atmospheric modeling software, such as MODTRAN4, to estimate the path radiance, sky irradiance, solar irradiance, and atmospheric transmission coefficients for different imaging scenarios in order to develop target subspaces for a material of interest. As indicated by Ientilucci and Bajorski (2006), estimation of these parameters using MODTRAN4 can be a time-consuming process that requires a substantial amount of computing resources, motivating their research into the use of statistical regression models to make these invariant signature matching methods more computationally practical.

As an alternative to these MODTRAN-based methods, we expand upon a concept proposed by Eismann (2006) and suggest the use of in-scene atmospheric calibration information to convert a target's reflectance signatures into a set of possible image signatures the target may exhibit in the hyperspectral image. These image signatures can be used to form target subspaces in the same manner as suggested by Healey and Slater and Thai and Healey, or they can be used in other signature matching methods such as the Linearly Constrained Minimum Variance (LCMV), Constrained Energy Minimization (CEM), or Target-Constrained Interference-Minimized Filter (TCIMF) methods described in Chang (2003). The advantages of using in-scene information to perform the reflectance-to-radiance conversion are: 1) the need to run MODTRAN is eliminated; 2) the user only needs to provide target reflectance signatures to use the algorithm; 3) the variability in target image signatures can potentially be reduced since they are no longer based on a range of viewing geometries and atmospheric conditions that cannot possibly all exist for a specific scene; and 4) target radiance variability due to

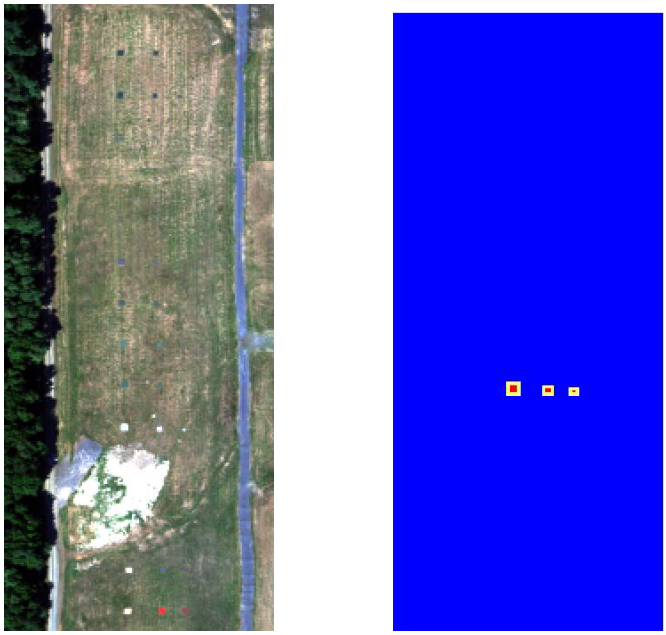
signature noise may be implicitly incorporated in the reflectance-to-radiance conversion, whereas it is omitted in existing invariant signature matching methods.

In the following paragraphs, we further describe our proposed signature matching methodology through an actual target detection example. We then summarize the major steps of the final algorithm—which we refer to as AutoMatch—and demonstrate its performance relative to the spectral angle mapper (SAM) applied to atmospherically calibrated imagery and to alternative configurations of AutoMatch. We conclude with a discussion of the limitations of our signature matching scheme and suggestions for future research.

Proposed Signature Matching Process

Perhaps the best way to describe our proposed AutoMatch method is to explain how it is applied to a basic signature matching problem. The problem we use in this effort is the detection of the F2 target material contained in the hyperspectral image depicted in Figure 39 which is derived from the Forest Radiance I HYDICE dataset. The actual location of this target is also indicated in Figure 39, and the ground-truth reflectance signatures for the target material are shown in Figure 40. The task at hand, then, is to show how the AutoMatch method is used to detect the F2 material using its reflectance curves as the only input parameter.

To begin, we note the fundamental obstacle to signature matching: our knowledge of the target material is expressed by reflectance signatures, while our image gives us the radiance signatures—or possibly just a relative energy intensity curve—detected by the



(a) (b)

Figure 39. Image Scene and Target Mask for Signature Matching Example
The image scene in (a) is Scene 4 in Appendix C. The target mask in (b) shows background in blue, target material in red, and buffer pixels in yellow.

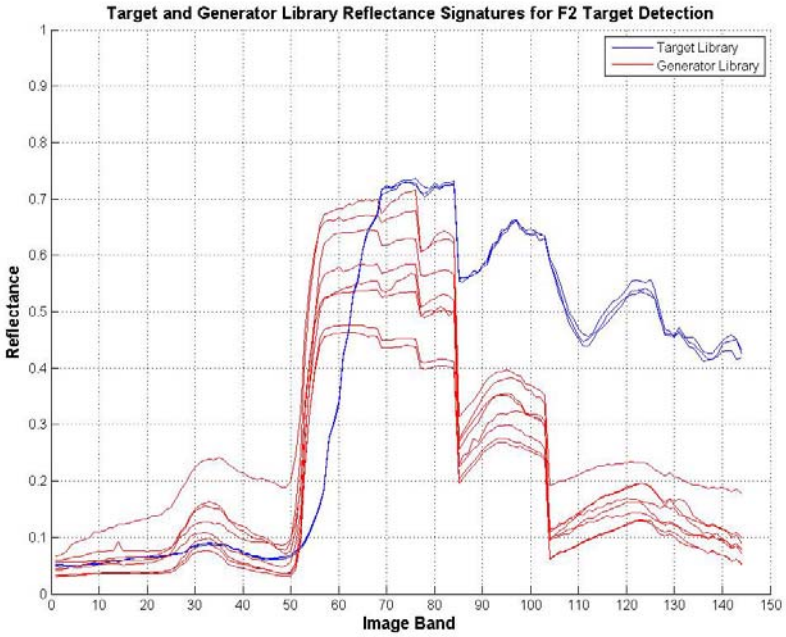


Figure 40. Reflectance Signatures for the Target and Generator Libraries

sensor at each pixel location in the image. To effectively compare the target reflectance signatures to the radiance signatures in the image, we must convert the reflectance signatures for the target material into a image signatures. For sensors operating in the reflective region of the electro-magnetic spectrum, this conversion is performed via the following equation derived by Schott (1997) that gives the effective radiance reaching the sensor for wavelength band *p* as a function of the wavelengths, λ, in the band:

$$L_p(\lambda) = \int_{\lambda} \beta_p(\lambda) \left[\left(E_s^{'}(\lambda)\tau_1(\lambda)\cos\theta + FE_d(\lambda) \right) \tau_2(\lambda) \frac{r(\lambda)}{\pi} + L_u(\lambda) \right] d\lambda \qquad (6.1)$$

where

$\beta_p(\lambda)$ = the normalized spectral response of the pth wavelength band of the sensor,
$E_s^{'}(\lambda)$ = the exoatmospheric irradiance from the sun,
$\tau_1(\lambda)$ = the Sun-Target atmospheric transmission,
$\tau_2(\lambda)$ = the Target-Sensor atmospheric transmission,
θ = the Sun zenith angle measured from the target surface normal,
$E_d(\lambda)$ = the spectral irradiance from the sky above the target,
F = the fraction of the sky above the target that is visible,
$r(\lambda)$ = the reflectance factor of the target as specified by the target's reflectance spectra, and
$L_u(\lambda)$ = the spectral path radiance.

For the MODTRAN-based signature matching methods, the MODTRAN4 model is used to generate estimates of the atmospheric terms in (6.1) at discrete wavelengths in band *p* for different atmospheres and viewing geometry. With these atmospheric terms, a numerical approximation to the integral in (6.1) is computed for band *p* for each atmosphere-viewing geometry permutation. Rather than proceed down this road, however, we note that if the response function for band *p* is assumed constant, and if the

wavelength-dependent terms are assumed to be relatively constant over the band, then (6.1) can be simplified to

$$L_p \cong \left(E_{sp}^{'}\tau_{1p}\cos\theta + FE_{dp}\right)\tau_{2p}\frac{r_p}{\pi} + L_{up} \tag{6.2}$$

where

$E_{sp}^{'}$ = the average exoatmospheric irradiance over band p,
τ_{1p} = the average Sun-Target transmission over band p,
τ_{2p} = the average Target-Sensor transmission over band p,
E_{dp} = the average sky irradiance over band p,
r_p = the average target reflectance over band p, and
L_{up} = the average path radiance over band p.

Assuming a sensor gain, g_p, and dark current, d_p, for the sensor in band p, the sensor reading, x_p, for band p of a target material with an average reflectance, r_p, in band p can be approximated by

$$\begin{aligned} x_p &= g_p L_p + d_p \\ &= g_p\left[\left(E_{sp}^{'}\tau_{1p}\cos\theta + FE_{dp}\right)\tau_{2p}\frac{r_p}{\pi} + L_{up}\right] + d_p \\ &= \left[\frac{g_p\left(E_{sp}^{'}\tau_{1p}\cos\theta + FE_{dp}\right)\tau_{2p}}{\pi}\right] r_p + g_p L_{up} + d_p \\ &= t_1 r_p + t_0 \end{aligned} \tag{6.3}$$

Thus, we can generate possible image spectra for a target of interest if we can determine values of t_0 and t_1 in (6.3) instead of the more computationally intensive approach taken by the original invariant subspace methods. In other words, if we know the average reflectance in band p for a target, and we have estimates of the coefficients t_0 and t_1 for the image, we can estimate the sensor reading produced by the target in band p of the

image. If we perform this estimate for all bands in the image, we can obtain an estimate for the target material image signatures in terms of sensor digital numbers. Note that by using (6.3) we are converting from target reflectance signature to a *digital number signature*, rather than a target radiance signature. If the sensor output is calibrated to generate radiance readings, there is no difference between the two signatures. However, if the sensor is not calibrated, or if the user simply does not know the sensor output units, (6.3) will automatically produce target image signatures in whatever units are applicable for the image.

With (6.3) in-hand, the next step in the signature matching process is to estimate values for the coefficients t_0 and t_1. Stated in vector form, we seek vectors $\mathbf{t}_0$ and $\mathbf{t}_1$ that produce a target image signature $\mathbf{x}_t$ from a target reflectance signature $\mathbf{r}_t$ through the linear equation

$$\mathbf{x}_t = \mathbf{t}_1 \otimes \mathbf{r}_t + \mathbf{t}_0 \tag{6.4}$$

To estimate the vector $\mathbf{t}_0$, we use the simple approach of setting each component of the vector to the minimum sensor reading in the corresponding band. For our target detection example, the $\mathbf{t}_0$ vector for the image is shown in Figure 41. This approach assumes that the image contains materials that collectively have zero reflectance across all image bands. If this assumption is valid, then the minimum value in each band should be the product of the band's sensor gain and path radiance summed with the sensor dark current. Obviously, not all images contain zero-reflectance materials for all bands, in which case this simple estimation method will tend to use values corresponding to shadow pixels. In such instances, the sky radiance will also contribute to the minimum values in each band, thereby overestimating the respective component of $\mathbf{t}_0$.

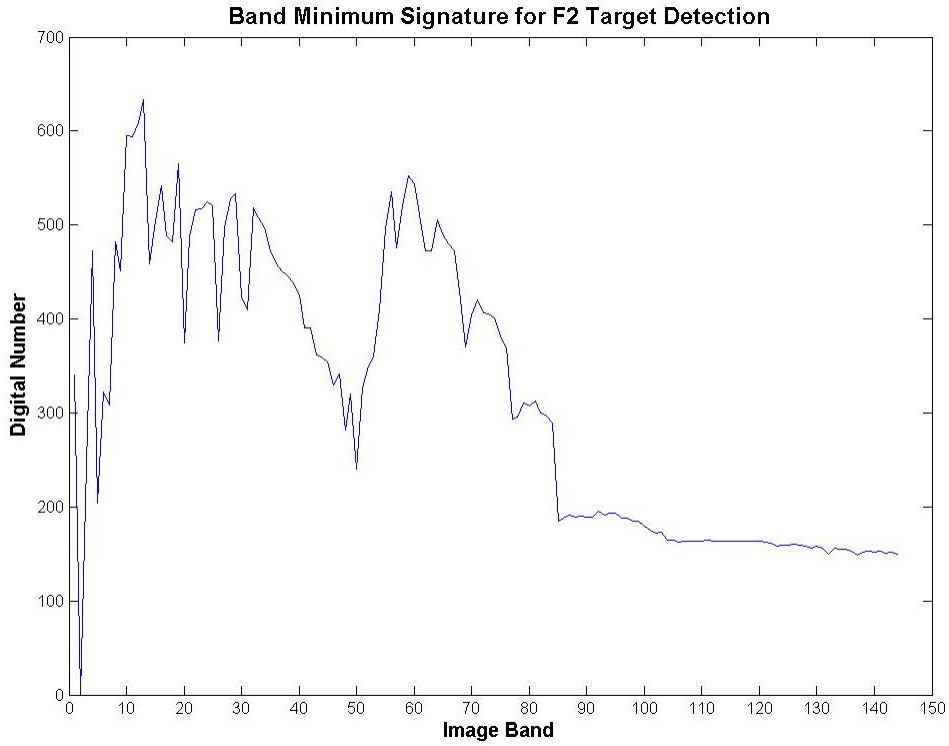


Figure 41. Band Minimum Signature, $\mathbf{t_0}$, for Signature Matching Example

Schott (1997) outlines more sophisticated methods for estimating $\mathbf{t}_0$ that avoid this problem; however, these methods require the user to identify known materials in the image or even transitions regions between shaded and illuminated pixels containing the same material. Due to the added user interaction imposed by these methods, we opt for the simpler band-minimum method for this research.

To determine the vector $\mathbf{t}_1$, we employ a technique commonly used in hyperspectral analysis for approximate in-scene atmospheric calibration. In particular, we attempt to find a background material in the image whose reflectance signature, $\mathbf{r}_b$, is known, and then use the image spectra, $\mathbf{x}_b$, for this material to find the components of $\mathbf{t}_1$ using the following:

$$t_{1p} = \frac{x_{bp} - t_{0p}}{r_{bp}} \tag{6.5}$$

where

x_{bp} = the component of the background material image spectra for band p,
t_{0p} = the component of $\mathbf{t}_0$ for band p, and
r_{bp} = the component of the background material reflectance spectra for band p.

Of course, to be consistent with our objective of an autonomous signature matching algorithm, we need to locate a suitable background material without user intervention. To accomplish this task, we use either the normalized-difference vegetation index (NDVI) originally proposed by Rouse et al (1973), or the Bare-Soil Index (BI) proposed by Chen et al (2004) to identify image pixels containing vegetation or bare soil, respectively—this process is similar to the method used in the ARCHER system described by Stevenson et al. (2005). These indices, which are computed for every pixel vector in an image, are defined as

$$NDVI = \frac{NIR - R}{NIR + R} \tag{6.6}$$

where

NIR = the average sensor reading at the pixel location across the NIR bands, and
R = the average sensor reading at the pixel location across the red light bands,

and

$$BI = \frac{(SWIR + R) - (NIR + B)}{(SWIR + R) + (NIR + B)} \tag{6.7}$$

where

$SWIR$ = the average sensor reading for the pixel location across the short-wave IR (SWIR) bands, and

B = the average sensor reading for the pixel location across the blue light bands.

In computing the NDVI and BI for the pixel vectors, we define the NIR, SWIR, red, and blue band ranges to be 700-1100nm, 1100-2500nm, 600-700nm, and 450-500nm, respectively. Figures 42 and 43 show gray-scale images based on the NDVI and BI, respectively, for our target detection example. In both images, brighter pixels indicate higher values of the indices.

As seen in the NDVI and BI images, these indices appear reasonably effective at assigning high values to materials for which they are intended to locate. In the NDVI image, trees and healthy grass obtain the highest values, while in the BI image, the large patch of dirt in the lower portion of the image generates the highest BI values. So which pixel vector should be selected for use in (6.5)? To answer this question, we make the observation that using only a single pixel vector in (6.5) is somewhat limiting and does not adequately describe the variability of a material's image spectra due to different illumination conditions and sensor noise. Thus, rather than select the single pixel vector corresponding to the highest index reading, we instead choose a number of pixel vectors with high index readings to better account for signature variability. To ensure the selected vectors correspond to the same material, we first run the AutoDet algorithm on the image to cluster the image spectra into similar groups as well as to identify outlying pixels. This latter step is important since both the NDVI and BI images indicate that potential targets may also receive a high index value. Using the index, cluster, and

Figure 42. Gray-scale Image of Pixel NDVI Values

Figure 43. Gray-scale Image of Pixel BI Values

anomaly information, we then select our background spectra in the following way: 1) identify the pixel vector with the highest reading that is not an anomaly; 2) identify the cluster to which the pixel vector is assigned; and 3) select additional pixel vectors from the same cluster identified in Step 2 that have the highest index values and are not anomalies. For the AutoMatch method, we select a total of 200 pixel vectors in the manner described.

For our target detection example, we use the NDVI as the selection index due to the large amount of vegetation in the image. Figure 44 shows the 200 pixel vectors selected based on NDVI, cluster assignment, and anomaly status, and Figure 45 shows the location of these 200 spectra in the image. As seen in these figures, the selection method appears to have chosen pixel vectors corresponding to a single material—namely, trees—and that the vectors appear to represent the same material under different conditions. It is hoped that the spectral variability evident in Figure 44 is due to sensor noise and different illumination conditions; however, the validity of this assumption is difficult to confirm or guarantee.

Using the selected pixel vectors—which we refer to as generator signatures—we are nearly in position to use (6.5) to estimate a set of possible $\mathbf{t}_1$ vectors corresponding to the generator signatures. However, we first must determine the reflectance signature for the background material that produced the generator signatures. Without any ground-truth information for an image, there is generally no way of determining the reflectance signature of the background material with any certainty. Instead, we use a library of reflectance signatures corresponding to similar materials and that is likely to contain the reflectance signature of the generator material. For example, if we are searching for

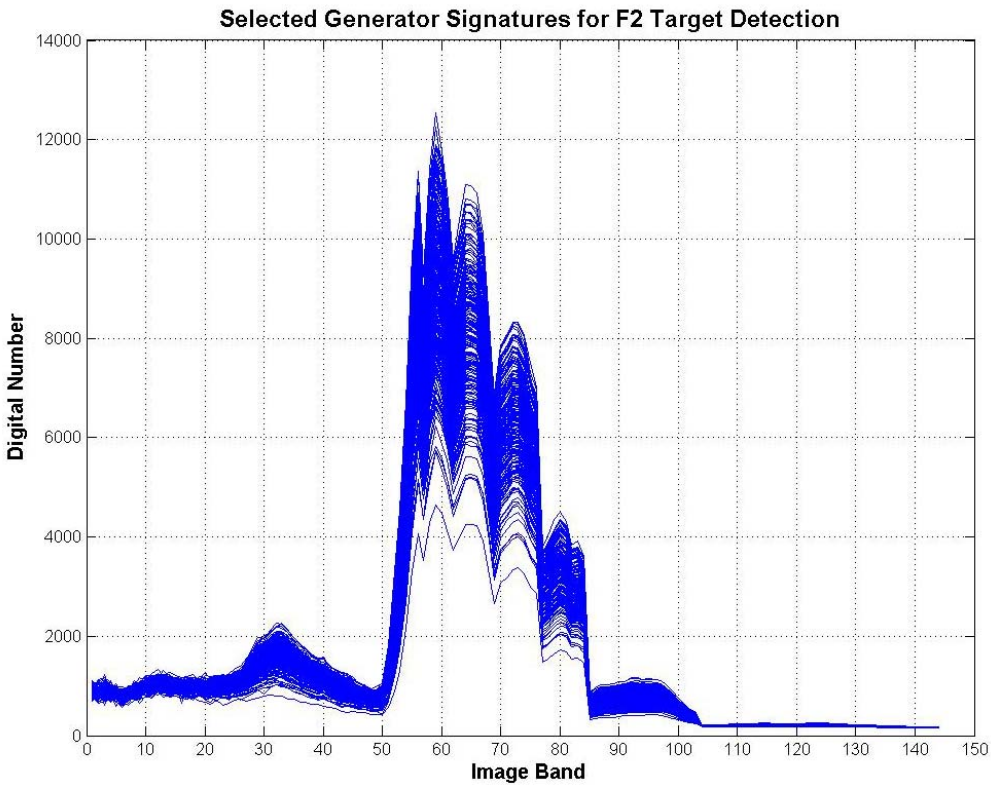


Figure 44. Generator Signatures Obtained using NDVI Values

targets in a desert scene and have selected generator signatures using the BI, then a soil library containing different desert soil reflectance spectra may be appropriate. Ideally, this library should match the image scene as much as possible, but, as we demonstrate later in this chapter, a generic vegetation or soil library that is not tailored to the hyperspectral image can also give good detection results.

For our target detection example, we use a reflectance library containing signatures for different broadleaf trees. These signatures, shown in Figure 40, were taken from the USGS and Johns Hopkins University (JHU) spectral libraries included with the ENVI software package, and correspond to aspen, maple, walnut, blue oak, leather oak, live oak, and a generic deciduous tree signature. It is important to note that the sweet

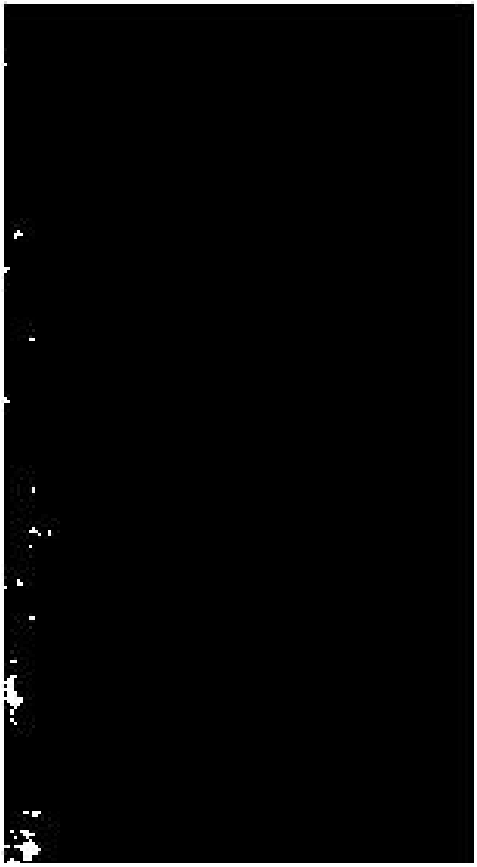


Figure 45. Image Showing Pixel Location for Generator Signatures

gum and locust tree signatures that are actually contained in the example image are not included in this library.

At this point, we have a set of reflectance signatures, a set of generator signatures, and the $\mathbf{t}_0$-vector. Hence, it is now possible to compute possible $\mathbf{t}_1$ vectors using (6.5). To do so, we use every combination of generator-reflectance signature pairs as well as $\mathbf{t}_0$ in (6.5) to produce a total of $n_1=n_g n_r$ $\mathbf{t}_1$ vectors, where n_g is the number of generator signatures, and n_r is the number of reflectance signatures. In our example, n_g=200 and n_r=8, establishing n_1 to be 1600. By computing this set of vectors, we hope to implicitly account for the variability in the target image spectra due to illumination, sensor noise, and our uncertainty of the true identity of the generator material.

Having determined the $\mathbf{t}_0$ vector and a set of $\mathbf{t}_1$ vectors, we return to (6.4) and generate a set of image signatures that may represent possible realizations of the target material. These target image signatures are generated by substituting all combinations of the target reflectance signatures and $\mathbf{t}_1$ vectors into (6.4) along with the estimate of $\mathbf{t}_0$. In our example, we have three reflectance signatures for the F2 target material, which, when combined with the 1600 $\mathbf{t}_1$ vectors, produces 4800 possible target image signatures that may indicate what the F2 target looks like spectrally in the image. Figure 46 plots these 4800 image signatures in blue, and also shows the actual ground-truth image signatures for the F2 target in green. As seen in this plot, the generated signatures serve as a reasonable approximation to the actual target signatures in the image, though the predicted sensor readings slightly underestimate the actual readings in the visible region of the EM spectrum. Possible reason for this underestimation are discussed later in this chapter.

At this point in the AutoMatch methodology, we have essentially arrived at the same point in the original invariant signature matching scheme of Healey and Slater (1999) at which we can use the generated image signatures to define a target subspace. The primary difference between the two methods is the manner in which the target image signatures are derived. Therefore, if we are dealing with pure-pixel targets we can proceed with the original Healey-Slater method to complete the target detection, or use the method proposed by Thai and Healey for the case of sub-pixel targets. Instead, we choose to use the TCIMF method of Ren and Chang (2000) which is a finite impulse response (FIR) filter that incorporates target and background signature information to detect targets in hyperspectral imagery. We use the TCIMF filter because it is relatively

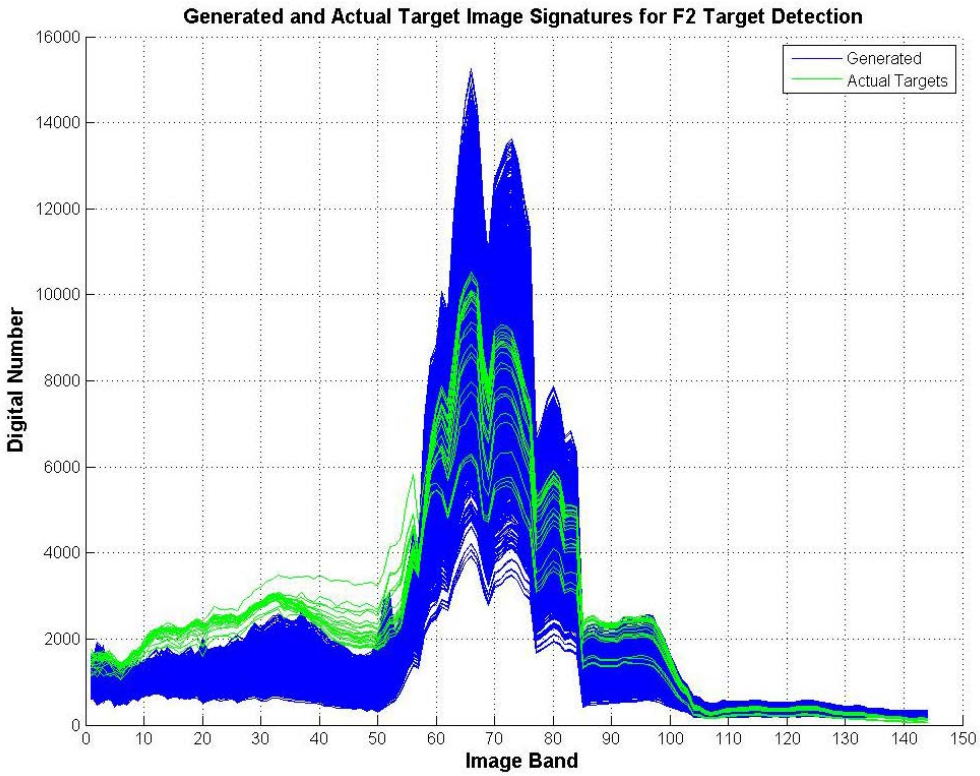


Figure 46. Generated and Actual Target Image Signatures for F2 Target

easy to implement, and more importantly, it produced better detection results when compared to the Healey-Slater and Thai-Healey methods, as we will see later in this chapter.

The objective of the TCIMF filter is to find a filter vector, $\mathbf{w}$, that when applied to an image vector, $\mathbf{x}_i$, produces a filter output of unity if $\mathbf{x}_i$ is a target, a filter output of zero if $\mathbf{x}_i$ is an undesired background signature, and minimizes the average output energy over all the signatures in the image. Specifically, let y_i represent the filter output when applied to $\mathbf{x}_i$. That is to say

$$y_i = \mathbf{w}^T \mathbf{x}_i = \mathbf{x}_i^T \mathbf{w}. \tag{6.8}$$

The average output energy is then defined to be

$$
\begin{aligned}
E_{avg} &= \frac{1}{N}\sum_{i=1}^{N} y_i^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{x}_i^T\mathbf{w}\right)^T\mathbf{x}_i^T\mathbf{w} \\
&= \mathbf{w}^T\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}^T\right)\mathbf{w} \\
&= \mathbf{w}^T\mathbf{R}\mathbf{w},
\end{aligned}
\tag{6.9}
$$

where **R** is referred to as the sample autocorrelation matrix of the matrix

$$
\mathbf{X} = \left\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\right\}. \tag{6.10}
$$

If we form a matrix, **D**, whose *d* columns contain the image signatures of the *d* targets we wish to detect, and a matrix, **U**, whose *u* columns contain the image signatures of *u* undesirable background signatures, construction of the TCIMF filter involves finding a filter vector, **w**, that solves the optimization problem

$$
\min_{\mathbf{w}}\left\{\mathbf{w}^T\mathbf{R}\mathbf{w}\right\} \text{ subject to } \left[\mathbf{D}\ \mathbf{U}\right]^T\mathbf{w} = \begin{bmatrix}\mathbf{1}_{d\times 1}\\ \mathbf{0}_{u\times 1}\end{bmatrix}. \tag{6.11}
$$

The solution to (6.11) is found by Ren and Chang to be

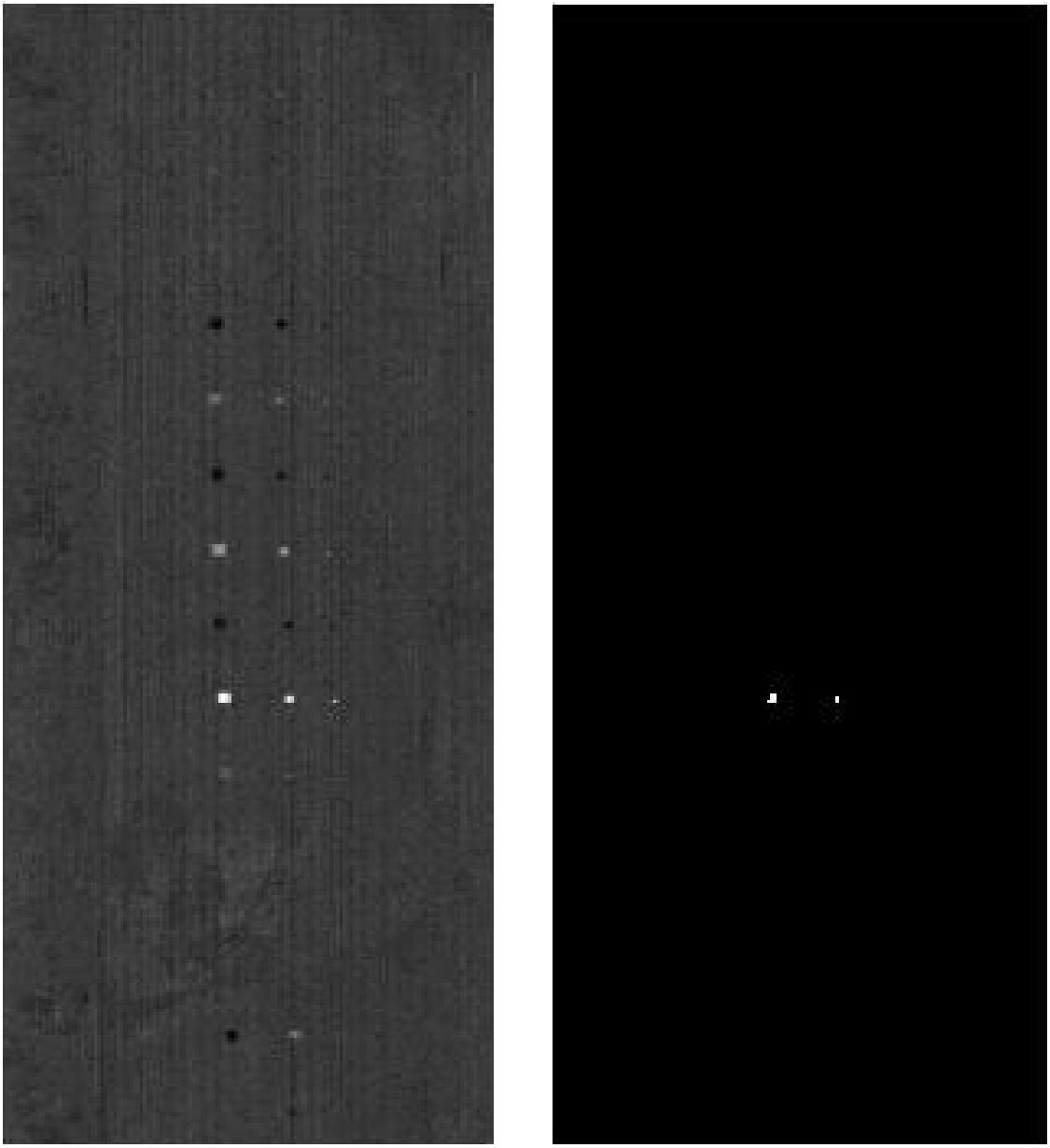
$$
\mathbf{w}_{TCIMF} = \mathbf{R}^{-1}\left[\mathbf{D}\ \mathbf{U}\right]\left(\left[\mathbf{D}\ \mathbf{U}\right]^T\mathbf{R}^{-1}\left[\mathbf{D}\ \mathbf{U}\right]\right)^{-1}\begin{bmatrix}\mathbf{1}_{d\times 1}\\ \mathbf{0}_{u\times 1}\end{bmatrix}. \tag{6.12}
$$

To apply the TCIMF filter, the **D** and **U** matrices are populated with desired target image signatures and undesired background image signatures, respectively, w_{TCIMF} is computed, and the filter is applied to each pixel vector in the image. The resulting filter outputs are then compared to a threshold value to identify target pixels. Unfortunately, the distribution of the filter outputs is not known, making this last step somewhat subjective. Ren and Chang use plots of the filter output to identify targets, or produce gray-scale images generated by the filter outputs in which bright pixels indicate pixels with high TCIMF values.

In the context of AutoMatch, we initialize the **D** matrix to contain a single vector corresponding to the mean vector of the generated target image signatures. For the **U** matrix, we use the mean vectors of the clusters produced by the AutoDet algorithm under the assumption that the clusters contain background materials whose filter output should be suppressed. By forming the **D** matrix with only the mean vector of the generated target image signatures, it would seem that much of the information contained in the set of generated target signatures is being lost. Whether or not this hypothesis is true warrants further research, but for now, we are content to use only the mean vector.

Upon applying the TCIMF filter to the image, AutoMatch produces a gray-scale image of the relative magnitudes of the filter outputs, as well as a binary target image showing the pixels that exceed the (1-α)-quantile of the filter outputs. In practice, α is set to the maximum tolerable false alarm rate measured as a fraction of the total number of image pixels. Viewed in another light, this threshold strategy returns the maximum number of pixels one is willing to manually inspect without finding an actual target. Though this threshold method lacks theoretical rigor, it is used in operational target detection systems and is a reasonable approach to designating targets when nothing is known of the output metric's distribution.

For our target detection example, the TCIMF gray-scale image and the binary target image with α=0.0001 are shown in Figure 47. Additionally, Figure 48 shows the OC curve for the TCIMF output values. As seen in these figures, the AutoMatch algorithm is quite effective in detecting the F2 target material. Moreover, the TCIMF filter does well in suppressing the background materials in the image making it



TCIMF Image **Binary Target Image**

Figure 47. TCIMF and Target Image for Target Detection Example

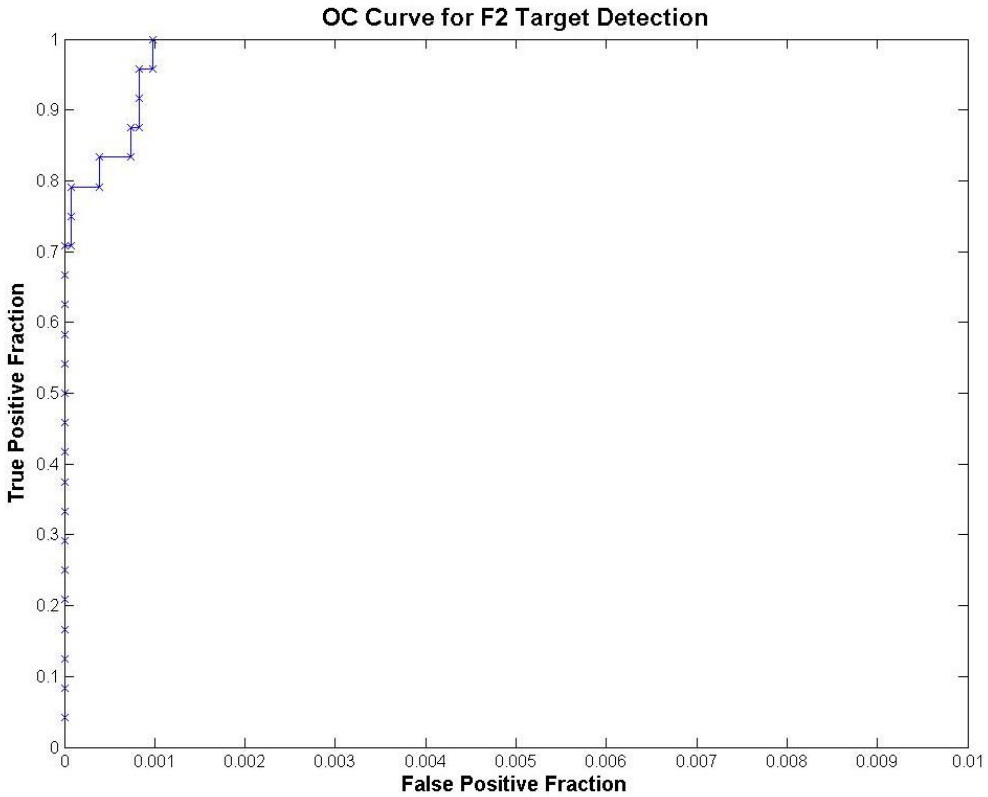


Figure 48. OC Curve for Target Detection Example

easier to visually detect candidate targets. We also see that in this particular example the threshold strategy used in AutoMatch was useful in highlighting the correct targets with no false-alarms. The primary reason the threshold method was so successful in this example is the high degree of separation between the target and background signatures in terms of the TCIMF filter output, as indicated by the OC curve.

Summary of the AutoMatch Target Detector

In the previous section, we explained the basic components of the AutoMatch detector and demonstrated its application to a an actual target detection problem. We now summarize the steps comprising the AutoMatch detector as follows:

1) Specify the reflectance signatures for the target that is to be detected.

2) Specify either a soil or vegetation library of reflectance signatures for use in (6.5) to generate $\mathbf{t}_1$ vectors. The choice of soil or vegetation signatures should be based on the expected background materials in the scene—i.e., use soil signatures for a desert scene and use vegetation spectra for a woodland scene.
3) Compute the $\mathbf{t}_0$ vector as the band minimum values.
4) Depending on the library specified in Step 2, compute either the NDVI or BI for every pixel vector in the image using (6.6) or (6.7), respectively.
5) Run the AutoDet algorithm to cluster the image pixels and identify anomalous pixels that may also be targets.
6) Identify the non-anomalous pixel vector with the highest index value from Step 4, as well as the cluster to which it belongs.
7) Select the n_g pixel vectors with the highest index values from the cluster identified in Step 6. This set of vectors is referred to as the generator signatures.
8) Use the generator signatures, the $\mathbf{t}_0$ vector, the n_r reflectance signatures from the library in Step 2, and (6.5) to generate $n_1=n_g n_r$ $\mathbf{t}_1$ vectors.
9) Use the $\mathbf{t}_1$ signatures and the n_t reflectance signatures from the target library specified in Step 1 to generate $N_t=n_t n_1$ target image signatures.
10) Use the mean vector of the target image signatures as the single column of the matrix **D**.
11) Use the mean vectors of the clusters found in Step 5 to form the columns of the matrix **U**.

12) Use **D** and **U** in (6.12) to form the TCIMF filter vector, $\mathbf{w}_{TCIMF}$, and apply the filter to all the pixel vectors in the image.

13) Produce a gray-scale image of the TCIMF filter outputs.

14) Designate any TCIMF filter outputs that exceed the (1-α)-quantile of the outputs as targets.

As indicated by this outline of the AutoMatch algorithm, the only parameters that need to be specified by the user are the target reflectance signatures, whether soil or vegetation should be used for $\mathbf{t}_1$ generation, either a soil or vegetation library, and a value for α. Thus we have eliminated the need for any knowledge of viewing geometry and atmospheric conditions, and we avoid the complexity of running the MODTRAN4 model or managing any database containing its output. Moreover, the user need not be concerned of the units of the hyperspectral data, making the AutoMatch detector easy to apply to an arbitrary hyperspectral image. Though these attributes are desirable in our quest for an autonomous signature matching algorithm, the accuracy of the detector still must be assessed relative to other alternative methods. We address this issue in the following sections.

Detector Comparisons

In the preceding sections of this chapter, we introduced the AutoMatch target detection algorithm as a method for locating targets in a hyperspectral image without the need for detailed atmospheric correction. We now investigate if the detection methodology employed by AutoMatch can be expected to perform as well as standard signature matching methods applied to atmospherically calibrated imagery. Additionally, we determine if using the TCIMF filter with the generated target image signatures is the

most useful implementation of AutoMatch relative to alternative configurations, and if the reflectance library used in Step 2 of AutoMatch significantly impacts detection accuracy. To achieve these objectives, the following test was performed:

1) Set the detection method to one of six methods.
2) Set the hyperspectral image to one of seven images.
3) Set the target material to a known target in the image whose location is verified by ground truth information.
4) Set the reflectance library used in Step 2 of AutoMatch.
5) Apply the detection method selected in Step 1 to the image.
6) Construct an OC curve for the detection results.
7) Repeat Steps 4 through 6 for additional libraries that are relevant to the image.
8) Repeat Steps 3 through 7 for additional targets in the image.
9) Repeat Steps 2 through 8 for each of the seven images.
10) Repeat Steps 1 through 9 for each of the six detection methods.

Additional implementation details for this test are described in the following paragraphs.

Detection Methods Tested

To assess the merits of the AutoMatch detector, we compared its performance to five other detectors: the Spectral Angle Mapper (SAM) applied to imagery calibrated using the FLAASH atmospheric calibration algorithm; the Healey-Slater method using the scene-derived target signatures; the Thai-Healey method using the scene-derived target signatures; the Thai-Healey method using scene-derived target signatures and the cluster means from AutoDet to form the background subspace; and an iterative version of

AutoMatch designed to improve detection through better estimates of the **R** matrix. The following paragraphs describe each of these methods in more detail.

FLAASH-SAM Detector

This detection method applies the commonly used Spectral Angle Mapper signature matching algorithm to the test images calibrated with the FLAASH atmospheric calibration algorithm originally developed by the Air Force Research Laboratory (AFRL). The SAM detector computes the following statistic for each pixel vector, **x**, in the image:

$$SAM\left(\mathbf{x}\right) = \arccos\left(\frac{\mathbf{r}_t\mathbf{x}}{\|\mathbf{r}_t\|\|\mathbf{x}\|}\right) \tag{6.13}$$

where $\mathbf{r}_t$ is the reflectance signature of the target. For our experiments, we used the mean target reflectance signature for the target, $\mathbf{r}_t$, whenever multiple reflectance signatures were available for the target. The pixels generating the smallest values of the *SAM*-statistic are designated as targets.

When using reflectance signatures in the SAM detector, the image must be calibrated to units of reflectance. For our tests, we applied the SAM detector to images that were calibrated using the FLAASH algorithm as part of the Forest Radiance I and Desert Radiance II data collection effort. This calibration was performed using atmospheric conditions and viewing geometry existing at the time of image acquisition, and hence represents an ideal target detection scenario. We therefore use the FLAASH-SAM detector as the benchmark to which we measure the relative merits of the in-scene calibration methods.

Healey-Slater Variant of AutoMatch (HS Method)

For this detection method, we follow the detection methodology proposed in Healey and Slater (1999) as closely as possible with the only difference being the generation of the target image signatures used to form the target subspace. Specifically, we use the in-scene generation process discussed earlier as opposed to the MODTRAN4-based method used in the original methodology.

Thai-Healey Variant of AutoMatch (TH Method)

This variant of the AutoMatch detector attempts to improve upon the HS Method by incorporating background information in the detection process to better detect sub-pixel targets. In implementing this detection method, we follow the detection methodology originally proposed by Thai and Healey (2002) as closely as possible. As in the case of the HS Method, the primary difference between the original methodology and our implementation is the use of the in-scene generated target image signatures.

We also note that we follow Thai and Healey's suggestions for setting the t_1 and t_2 parameters, and for their proposed *non-negativity test*. The t_1 and t_2 parameters control the number of basis vectors used in defining the background subspace for the target model given in (4.9)—that is, the number of columns in the **B** matrix. Thai and Healey suggest using values of t_1=0.0002 and t_2=0.000045, though no detailed explanation is provided for these recommendations. The non-negativity test is used in the original implementation to discard pixel vectors that achieve a high likelihood ratio score but have a negative target component, $\mathbf{T\theta}$, in (4.9). Specifically, Thai and Healey suggest discarding any pixel vector that has an estimated $\mathbf{T\theta}$ vector whose "sum of absolute

values of the negative elements in the normalized unit length spectrum" is less than 0.01. Again, no detailed explanation is given for this non-negativity threshold.

Thai-Healey Variant of AutoMatch using Cluster Means (TH-CL Method)

In implementing the TH Method, we encountered several details of the original methodology that were somewhat awkward to implement. As mentioned previously, use of the suggested t_1 and t_2 parameter settings led to the uncertainty of whether or not these values are applicable to a range of images or just those used in Thai and Healey's original experiments. Additionally, the original method calls for dividing the image into rectangular regions and estimating the **B** matrix for each region. This suggestion increases the complexity of the algorithm, but more importantly, it opens a debate on the region size to use for best detection results.

To avoid these implementation issues, we modified the TH Method to use a single **B** matrix estimated from the k cluster means generated by the AutoDet method. Specifically, we form a matrix, **M**, using the AutoDet cluster means as the columns. We then perform a singular value decomposition on **M** and set **B** equal to the first k columns of the resulting SVD **U** matrix, since these columns are a basis for the column-space of **M** (assuming the column-space of **M** has a rank of k). We then use this single **B** matrix for all pixel vectors in the image when computing the respective likelihood ratio. Besides this revised computation of **B**, the TH-CL Method is the same as the TH Method.

Iterative TCIMF (TCIMF-I Method)

As described by Chang (2003), the **R** matrix used in the TCIMF filter has the effect of minimizing the filter output of background signatures. However, if the signatures of targets that we are trying to detect are included in the computation of **R**, the

filter output for target signatures will also be reduced. To counter this problem, Chang suggests screening the image for pixel signatures that are similar to the target signature using a metric such as *SAM* given in (6.13) with the mean target image signature from the **D** matrix replacing $\mathbf{r}_t$. Any pixel vectors with a sufficiently small *SAM* value are then excluded from the **R** matrix computation. A limitation with this method, however, is determining how small the *SAM* statistic must be to exclude a pixel vector from the computation. To avoid this problem, we instead modify the original TCIMF method in the following manner: 1) compute the **R** matrix using all pixel vectors in the image; 2) apply the TCIMF filter to the entire image; 3) remove any pixel vectors producing significantly high TCIMF values and recompute **R**; and 4) reapply the TCIMF filter using the updated version of **R**. In Step 3, pixel vectors are removed if their TCIMF value is more than five standard deviations from the mean value, where the number of standard deviations, z_i, for the ith pixel vector is computed robustly as

$$z_i = \frac{\left(T_i - \tilde{T}\right)}{\operatorname*{med}_i\left(T_i - \tilde{T}\right)} \tag{6.14}$$

where

T_i = the TCIMF output for pixel vector i, and
$\tilde{T}$ = the median TCIMF output for the image.

Our choice of five standard deviations as the threshold value is based on typical non-iterative TCIMF values of actual targets obtained using the AutoMatch detector. Our tests in this area were not comprehensive, and further research may produce a better threshold.

Test Images and Targets

The images used in the comparison test are Scenes 2, 4, 5, 6, 7, 8, and 9 found in Appendix C. Scenes 2 and 4 are subsets of a single, larger scene from the Forest Radiance I dataset, while the remaining scenes are subsets from a single, larger scene from the Desert Radiance II dataset. Our reasons for using these subsets as opposed to the two larger scenes are two-fold: first, we are interested in determining the effect of different degrees of scene clutter on detection performance; and second, the original images are too large to analyze in a timely manner. In the case of the FLAASH-SAM detector, the scenes are exactly the same, but the units have been converted to reflectance using the FLAASH algorithm. As stated previously, this atmospheric calibration was performed as part of the Forest Radiance I and Desert Radiance II collection efforts.

The targets that we attempted to detect in each image are listed in Table 30 using the same nomenclature as in the Forest Radiance I and Desert Radiance II datasets. These targets represent a variety of different man-made materials ranging from sub-pixel to multiple-pixel sizes. The reflectance signatures for these targets were ground-truthed at the time of the image acquisition using multiple field spectrometer measurements. Truth masks depicting the location of the targets and used in our OC curve computations, were also obtained from the Forest Radiance I and Desert Radiance II datasets. For a given target, these truth masks indicate the location of pure pixels, sub-pixels, shade pixels, glare pixels, and pixels for the which the respective material could not be verified (guard pixels). In computing our OC curves, we only measured a detector's ability to detect pure pixel targets and sub-pixel targets; the shade and glare pixels were merged into the guard pixel category, and were not used in the OC curve computations.

Table 30. List of Targets Contained in Detector Comparison Scenes

Scene	Target List
2	c5, c6, dv3, dv4, v1, v3, vf1, vf5, vf6, vf7
4	f2, f3, f4, f6, f7, f8, f11, f12, f13, f14, t1, t2
5	cb1, cb2, cb3, cb4, cb5, cb6, cr1, cr2, cr3, e2, e4, e5, f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, m4, m5, m6, m7, m9, m10, m12, pp1, pp2, pp3, pr2, pw1, pw2, pw3, t1, t2
6	cb1, cb2, cb3, cb4, cb5, cb6, cr1, cr2, cr3, f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, m4, m5, m6, m7, m9, m10, m12, pp1, pp2, pp3, pr2, pw1, pw2, pw3, t1, t2
7	e2, v5, v6, v10, v11, v12
8	v5, v6
9	cb1, cb2, cb3, cb4, cb5, cb6, cr1, cr2, cr3, f1, f2, f3, f4, f5, f7, m6, m7, pp1, pp3, pw1, pw2, pw3, t1, t2

Table 31. Summary of Generator Reflectance Signature Libraries

Library	Scenes	Signatures
FR Trees	2, 4	Sweet Gum, Locust
Generic Trees	2, 4	Aspen 1, Aspen 2, Maple, Walnut, Blue Oak, Leather Oak, Live Oak, Generic Deciduous
FR Soil	2, 4	Soil 1, Soil 2
DR Soil	5, 6, 7, 8, 9	Soil 1, Soil 2, Soil 3, Soil 4, Soil 5, Soil 6
Generic Soil	2, 4, 5, 6, 7, 8, 9	Silty Loam 1, Silty Loam 2, Sandy Loam 1, Sandy Loam 2, Sandy Loam 3, Sandy Loam 4, Sandy Loam 5, Sandy Loam 6, Sand 1, Sand 2, Grayish Brown Loam
Generic Brush	7, 8	Coyote Bush, Rabbit Brush, Sage Brush, Salt Brush

Reflectance Libraries

A total of six reflectance libraries where used in Step 4 of the test procedure. The signatures contained in these libraries are shown in Appendix G. Table 31 summarizes the contents of each library and lists the scenes with which they were used in our tests. As indicated in the table, we only used a library with a particular scene if the materials in the library matched those expected to be present in the scene. For example, we only used libraries containing broad-leaf tree signatures with the Forest Radiance I scenes since the

Desert Radiance II are unlikely to contain broad-leaf trees. For each image, we also used both a generic library and a library containing only ground truth signatures obtained during the Forest Radiance and Desert Radiance collections. By using these generic and specialized libraries we are able to assess the impact of library accuracy on AutoMatch performance.

Test Results

Our comparison tests generated 1848 detection scenarios over the seven images, six detection methods, six generator libraries, and range of targets tested. Rather than plot 1848 different OC curves, we summarize the OC curve results for each image scene in Tables 32 through 38. In each of these tables, we list the number of targets in the scene that were detected at a true-positive fraction of greater than 0.4, 0.6, or 0.8 while only producing a false-positive fraction of less than 0.0005. For each image, we list this information for each detection method and each generator library applied to the image. As an example of how to interpret this data, we refer to the (AutoMatch, TPF=0.040, FR Soils)-entry in Table 32. The value of 8 in this entry indicates that for eight of the 11 target materials that we attempted to detect in Scene 2, the AutoMatch method detected at least 40% of the respective target pixels while generating an FPF less than 0.0005. By reporting our results in this manner, we make the assumption that an FPF of 0.0005 is a reasonably small number of false alarms. Based on the range of TPF values that we use, we also assume that 40-80% target coverage at a FPF of 0.0005 is sufficient to visually distinguish a target from background materials.

Using the information in Tables 32 through 38, we address the major questions this experiment is designed to answer: 1) Does the AutoMatch algorithm achieve similar

Table 32. Signature Matching Comparison Results for Scene 2

Method	TPF	Library			
		FR Trees	Gen. Trees	FR Soil	Gen. Soil
FLAASH-SAM	**0.40**	7	7	7	7
	0.60	4	4	4	4
	0.80	2	2	2	2
HS	**0.40**	3	5	3	4
	0.60	2	4	2	2
	0.80	2	1	1	2
TH	**0.40**	2	0	3	0
	0.60	0	0	2	0
	0.80	0	0	0	0
TH-CL	**0.40**	6	1	4	4
	0.60	2	1	3	3
	0.80	2	1	2	2
AutoMatch (TCIMF)	**0.40**	4	4	8	8
	0.60	2	1	4	3
	0.80	1	1	2	1
TCIMF-I	**0.40**	5	4	7	7
	0.60	2	2	3	3
	0.80	2	2	3	3

Table 33. Signature Matching Comparison Results for Scene 4

Method	TPF	Library			
		FR Trees	Gen. Trees	FR Soil	Gen. Soil
FLAASH-SAM	**0.40**	7	7	7	7
	0.60	6	6	6	6
	0.80	4	4	4	4
HS	**0.40**	5	6	7	7
	0.60	4	4	5	6
	0.80	3	3	5	5
TH	**0.40**	8	6	11	9
	0.60	7	5	10	7
	0.80	6	5	9	7
TH-CL	**0.40**	9	9	9	8
	0.60	9	9	9	8
	0.80	7	6	7	5
AutoMatch (TCIMF)	**0.40**	10	9	10	10
	0.60	9	7	7	7
	0.80	3	4	6	6
TCIMF-I	**0.40**	10	9	9	8
	0.60	8	6	7	7
	0.80	4	3	7	5

performance as detectors using more sophisticated atmospheric correction methods; 2) Is the TCIMF method the most useful detector to use in AutoMatch relative to the other

Table 34. Signature Matching Comparison Results for Scene 5

Method	TPF	Library	
		DR Soil	Generic Soil
FLAASH-SAM	**0.40**	22	22
	0.60	15	15
	0.80	7	7
HS	**0.40**	22	18
	0.60	19	12
	0.80	12	6
TH	**0.40**	20	13
	0.60	14	12
	0.80	13	9
TH-CL	**0.40**	17	8
	0.60	13	8
	0.80	11	7
AutoMatch (TCIMF)	**0.40**	28	23
	0.60	23	15
	0.80	16	11
TCIMF-I	**0.40**	24	19
	0.60	19	14
	0.80	15	10

Table 35. Signature Matching Comparison Results for Scene 6

Method	TPF	Library	
		DR Soil	Generic Soil
FLAASH-SAM	**0.40**	22	22
	0.60	18	18
	0.80	10	10
HS	**0.40**	21	21
	0.60	19	13
	0.80	13	5
TH	**0.40**	16	13
	0.60	13	9
	0.80	10	8
TH-CL	**0.40**	15	10
	0.60	11	9
	0.80	9	8
AutoMatch (TCIMF)	**0.40**	23	20
	0.60	20	15
	0.80	16	9
TCIMF-I	**0.40**	20	13
	0.60	17	11
	0.80	10	9

methods tested; and 3) Does the accuracy of the reflectance library used by AutoMatch in the target signature generation process have a significant impact on detection accuracy? In regards to the first question, it is evident that the AutoMatch detector performed as well or better than the FLAASH-SAM detector at the FPF level of 0.0005 for Scenes 4, 5, 7, 8, and 9, regardless of the generation library. For Scenes 6, AutoMatch peformed better than FLAASH-SAM when the specialized library was used, and only slightly worse when the generic library was used. For Scene 2, the performance of AutoMatch relative to FLAASH-SAM is dependent on the type of reflectance library used rather than the accuracy of the signatures. That is to say, when the soil libraries are used, AutoMatch performs comparably to FLAASH-SAM, while using the vegetation libraries in AutoMatch significantly degrades its performance relative to FLAASH-SAM. Based on these results, we conclude that the AutoMatch detector is a useful alternative to methods using more sophisticated atmospheric calibration algorithms. Because these latter methods can only be used when atmospheric conditions and viewing geometry for an image are known or estimated, the value of AutoMatch—which requires none of this information—is further increased.

To answer the second question, we note that for the five desert images, AutoMatch using TCIMF performs the best of all the alternative AutoMatch configurations tested, though the HS Method gives comparable performance depending on the generation reflectance library used. For the Forest Radiance I scenes, AutoMatch with TCIMF is the better performer in some cases, but not all. Specifically, AutoMatch with TCIMF performs the best for Scene 2 when the soil libraries are used, but lags

Table 36. Signature Matching Comparison Results for Scene 7

Method	TPF	Library		
		Gen. Brush	DR Soil	Generic Soil
FLAASH-SAM	**0.40**	2	2	2
	0.60	0	0	0
	0.80	0	0	0
HS	**0.40**	1	4	1
	0.60	0	2	0
	0.80	0	1	0
TH	**0.40**	0	4	1
	0.60	0	2	0
	0.80	0	0	0
TH-CL	**0.40**	0	1	0
	0.60	0	0	0
	0.80	0	0	0
AutoMatch (TCIMF)	**0.40**	0	4	4
	0.60	0	3	2
	0.80	0	1	0
TCIMF-I	**0.40**	0	2	2
	0.60	0	0	1
	0.80	0	0	0

Table 37. Signature Matching Comparison Results for Scene 8

Method	TPF	Library		
		Gen. Brush	DR Soil	Generic Soil
FLAASH-SAM	**0.40**	0	0	0
	0.60	0	0	0
	0.80	0	0	0
HS	**0.40**	0	1	1
	0.60	0	1	0
	0.80	0	0	0
TH	**0.40**	0	0	0
	0.60	0	0	0
	0.80	0	0	0
TH-CL	**0.40**	0	0	0
	0.60	0	0	0
	0.80	0	0	0
AutoMatch (TCIMF)	**0.40**	0	1	1
	0.60	0	1	1
	0.80	0	1	1
TCIMF-I	**0.40**	0	1	1
	0.60	0	1	1
	0.80	0	1	1

Table 38. Signature Matching Comparison Results for Scene 9

Method	TPF	Library	
		DR Soil	Generic Soil
FLAASH-SAM	**0.40**	14	14
	0.60	11	11
	0.80	6	6
HS	**0.40**	15	10
	0.60	12	6
	0.80	10	3
TH	**0.40**	12	7
	0.60	9	4
	0.80	7	3
TH-CL	**0.40**	15	8
	0.60	11	8
	0.80	7	6
AutoMatch (TCIMF)	**0.40**	15	14
	0.60	14	12
	0.80	11	5
TCIMF-I	**0.40**	14	13
	0.60	12	11
	0.80	10	8

behind the HS and TH-CL methods when using the vegetation libraries. In Scene 4, the performance of AutoMatch with TCIMF is comparable to the TH, TH-CL, and TCIMF-I methods with none of them being clearly superior to the rest. Based on these results, it is evident that AutoMatch using TCIMF is either the best or among the best detectors across all the images tested, and therefore is the most logical choice for use as the final AutoMatch detector.

Turning our attention to the third question, we see in each of the output tables that the use of a generic reflectance library—as opposed to signatures of materials that are known to exist in the image scene—degrades the performance of all the AutoMatch variants (FLAASH-SAM performance is not affected because it does not generate target signatures). The degree of degradation due to the generic libraries is detector-dependent. AutoMatch using TCIMF is the least sensitive to the reflectance library, while the TH and TH-CL methods can be significantly impacted by the type of reflectance library. The

reason TCIMF is not as sensitive to the accuracy of the reflectance library is its use of the mean of the generated target signatures as opposed to the subspace defined by these signatures. When a generic library is used, the variance of the reflectance signatures it contains is generally larger (this is definitely the case for the libraries we used). This higher variance, in-turn, leads to higher variance in the generated target signatures. Where this increased variance in the generated target signatures may have little or no effect on the target signature mean vector, it will certainly inflate the size of the target subspace, thereby increasing the chance for false alarms.

Beyond the three questions the comparison test was intended to answer, several other observations can be made from the test results. To begin with, it is evident that none of the detectors tested—including the FLAASH-SAM method—are effective at detecting all the target materials used in the test. With the false-positive fraction held at 0.0005, we see that the best detectors can only achieve 40% target coverage for 8 out of 11, 11 out of 12, 28 out of 38, 23 out of 35, 4 out of 6, 1 out of 2, and 15 out of 23 target materials in Scenes 2, 4, 5, 6, 7, 8, and 9, respectively. Upon further analysis of our test output, we found that, for a given image, a subset of the target materials are difficult for all methods to detect. In Table 39, we list the target materials that evaded 40% detection at a false positive fraction of 0.0005 by all detectors, regardless of the generator reflectance library used. Table 39 also lists the targets that evaded 40% detection at a false positive fraction of 0.0005 for four or five out of the six methods tested. In general, these problematic target materials fall in one of three categories: 1) their spectral signatures are very similar to background signatures in the image; 2) their signatures are very similar to other target materials in the scene, thereby causing many false

Table 39. List of Targets that are Difficult to Detect

Scene	Targets with less than 40% coverage at an FPF of 0.0005 for all 6 methods	Targets with less than 40% coverage at an FPF of 0.0005 for 4 or 5 methods
2	v5, vf6	dv4, vf5, vf7
4	f2	f8, f13
5	cr1, e5, m5, pr2	cr2, e2, e4, f8, m4, m6, m7, m9, m10, m12, pp1, pp2, pw2, t1, t2
6	m5, m6, pr2, t2	cb1, cr1, cr2, f8, m4, m9, m10, m12, pp1, pw2, pw3, t1
7	None	e2, v5, v6, v11, v12
8	None	v5, v6,
9	m6	cb1, cb4, cb5, cr1, cr2, m7, pw3, t2

alarms by other targets of similar materials; or 3) their reflectance signatures have high variability, thereby increasing the size of the target subspace used in the HS, TH, and TH-CL methods. In regards to the third category, some of the problematic targets are vehicles for which reflectance signatures were measured at different points on the vehicle. In some instances the measurement points were over different materials or under different levels of illumination, thus contributing to the high variance of the reflectance signatures. Focusing on a single part of these types of targets where the reflectance signature is relatively constant, may improve the detection of these targets. Further research is required to confirm this hypothesis and to improve detection of the other two categories of problem targets.

A second observation we make from the test results is the inconsistency of the TH and TH-CL methods. In particular, we notice that for Scenes 4 and 9 these methods are comparable to the AutoMatch method, but their performance lags for the remaining scenes. The primary reason of this inconsistency is the non-negativity test recommended by Thai and Healey in the original implementation. In our experiment, there were several instances in which target pixels failed this non-negativity test along with a large number

of background pixels. When these failures occurred, the pixels were discarded, per the recommendation of Thai and Healey, with the unfortunate consequence that some target pixels could no longer be detected, thereby corrupting the OC curve computations. It is possible that using a different threshold than 0.01 in this non-negativity test will improve the performance of these detectors, but there is no practical guidance on what this threshold should be for an arbitrary image. Thus, we are left with the conclusion that this non-negativity test is a limitation of the TH and TH-CL methods that must be resolved before these methods can be used in practice.

An additional consequence of this non-negativity test problem is that the TH and TH-CL methods cannot be accurately compared due to the confusing effects produced by the non-negativity tests. However, based on the results we obtained, the TH-CL method performed comparably or better than the TH method in Scenes 2, 4, 6, and 9, and only slightly worse for Scene 5—both methods performed poorly with Scene 8. Thus, it would appear as though using the background cluster means to define the background subspace may lead to a less subjective strategy than the original method proposed by Thai and Healey.

As a final observation from the comparison test, we note that there is little or no benefit gained by the TCIMF-I method as we have implemented it. For a very small number of targets in Scenes 2 and 4, the iterative scheme improved the separability between the target and background materials, but in general, performance was either unaffected or even decreased. Further research is required to determine if a better threshold for removing suspected targets on the first pass of the algorithm can improve upon these results.

AutoMatch Limitations

In the previous section we demonstrated that the AutoMatch detector can effectively detect a range of targets in different hyperspectral images without the need of detailed atmospheric correction. However, the detector is not without limitations, some of which we have alluded to already. In particular, the method may produce inaccurate target signatures, it is limited in its ability to use target signature variance information, and its performance can be degraded when applied to images with non-homogeneous atmospheric conditions. Each of these limitations is discussed further in the following paragraphs.

Inaccurate Target Signatures

In generating target image signatures, AutoMatch uses several pieces of information whose potential inaccuracies can lead to inaccurate target signatures. To understand this limitation better, suppose that we are attempting to generate a target image signature using the generator signature, $\mathbf{t}_1$, target reflectance signature, $\mathbf{r}_t$, and band-minimum vector, $\mathbf{t}_0$. Then for band λ, the generated target signature value is

$$G_\lambda = t_{1\lambda} r_{t\lambda} + t_{0\lambda} \tag{6.15}$$

where

$$\begin{aligned} t_{1\lambda} &= \text{the value in band } \lambda \text{ of } \mathbf{t}_1, \\ r_{t\lambda} &= \text{the value in band } \lambda \text{ of } \mathbf{r}_t, \text{ and} \\ t_{0\lambda} &= \text{the value in band } \lambda \text{ of } \mathbf{t}_0. \end{aligned}$$

Further, suppose that $\mathbf{t}_1$ was computed using generation signature, $\mathbf{g}$, and generator reflectance signature, $\mathbf{r}_g$. Then $t_{1\lambda}$ has the value

$$t_{1\lambda} = \frac{g_\lambda - t_{0\lambda}}{r_{g\lambda}} \tag{6.16}$$

where

$$g_\lambda = \text{the value in band } \lambda \text{ of } \mathbf{g}, \text{ and}$$
$$r_{g\lambda} = \text{the value in band } \lambda \text{ of } \mathbf{r}_g.$$

Substituting (6.16) into (6.15) gives

$$\begin{aligned} G_\lambda &= \left(\frac{g_\lambda - t_{0\lambda}}{r_{g\lambda}} \right) r_{t\lambda} + t_{0\lambda} \\ &= \frac{g_\lambda r_{t\lambda}}{r_{g\lambda}} + t_{0\lambda} \left(1 - r_{t\lambda}\right). \end{aligned} \tag{6.17}$$

Upon consideration of (6.17), it is evident that our estimate of G_λ can be adversely affected in the following ways:

1) If the band minimum, $t_{0\lambda}$, overestimates (underestimates) the actual product of the path radiance and sensor gain, summed with the dark-current in band λ, then G_λ will be too high (low), assuming the other parameters in (6.17) are accurate. However, in bands where the target reflectance is relatively high, this problem is less of a concern.
2) If $r_{g\lambda}$ is higher (lower) than the true reflectance of the material producing **g**, then G_λ will be too low (high), assuming the other parameters in (6.17) are accurate.
3) If $r_{t\lambda}$ is higher (lower) than the reflectance signature of the actual target signatures in the scene due to effects such as weathering, G_λ will be too high (low), assuming the other parameters in (6.17) are accurate.

In addition to these potential problems in generating the target signature vectors, we also notice for a given $\mathbf{t}_1$ and $\mathbf{t}_0$ the variance of (6.15) is

$$
\begin{aligned}
V\left(G_{\lambda}\right) &= V\left(t_{1\lambda} r_{t\lambda} + t_{0\lambda}\right) \\
&= V\left(t_{1\lambda} r_{t\lambda}\right) + V\left(t_{0\lambda}\right) \\
&= V\left(t_{1\lambda} r_{t\lambda}\right) + 0 \\
&= t_{1\lambda}^{2} V\left(r_{t\lambda}\right).
\end{aligned}
\tag{6.18}
$$

Thus, if the target reflectance signatures have a high variance, so will the variance of the generated signatures. Additionally, the variance of the generated signatures in band λ will grow as the square of the value of $\mathbf{t}_1$ in band λ. Finally, if the target reflectance signatures are relatively constant, but the $\mathbf{t}_1$ vectors we generate have high variance due to the generator reflectance signatures, the variance of target signatures will again increase. If the variance of the generated signatures grows via either of these mechanisms, the subspace defined by these signatures will also grow, thereby increasing the chance for false alarms if the HS, TH, or TH-CL methods are used in AutoMatch. Because the TCIMF method does not use target subspaces, it is not affected by these potential variance problems, though the other generation problems listed above, which can impact the shape of the generated signatures, will affect all the methods, including TCIMF.

Loss of Variance Information Using TCIMF

Though TCIMF is somewhat resistant to increases in generated target signature variance due to potential inaccuracies in the generation process, its complete ignorance of the target signature variance can also lead to detection errors, particularly if the actual target signatures have high variance. Because we only use the mean vector of the generated targets to form the **D** matrix of TCIMF, extreme observations from the actual target signature distribution may produce unintended results. As an example, suppose the mean of the generated target signatures is $\mathbf{d}_t$, and that some of the actual target pixels in

the image have signatures equal to $\mathbf{d}_t$ while other target pixels have signatures approximately equal to some multiple of $\mathbf{d}_t$, say $\kappa\mathbf{d}_t$, due to differences in target orientation. For the targets with signatures equal to $\mathbf{d}_t$, TCIMF should produce values of $\mathbf{w}^T\mathbf{d}_t=1$. For the other target signatures, the output from TCIMF will be $\mathbf{w}^T\kappa\mathbf{d}_t = \kappa\mathbf{w}^T\mathbf{d}_t=\kappa$. If $\kappa>1$, the TCIMF output for the target will obviously be greater than one, while if $\kappa<1$, the TCIMF output will be less than one. Though the former case may not pose a problem if we are simply looking for pixels with the highest TCIMF values, the latter case may result in missed targets depending on how we threshold the output values.

In order to avoid this problem, we suggest further research to determine how better to account for the variability of target signatures using the TCIMF construct. There is no restriction on the number of signatures that can be used in the **D** matrix; however, the best way to populate this matrix from the generated target signatures requires further investigation. Likewise, a method for constructing the **U** matrix beyond using simply the background cluster mean vectors may also help to better suppress background materials, thereby increasing the separation between target and background TCIMF values.

Non-Homogeneous Atmospheric Conditions

The final limitation of the AutoMatch detector that we address is its current assumption of homogeneous atmospheric conditions throughout the image scene. Depending on the size of the geographic region that the image covers, it is likely that the atmospheric conditions vary throughout the region due to different concentrations of airborne particles and water vapor. If the generator signatures selected by AutoMatch are distributed throughout the image, these differences in atmospheric conditions may not

pose a problem. However, if the generator signatures are concentrated in one region of the image, the $\mathbf{t}_1$ vectors computed from the signatures may not capture the atmospheric variations. Should this event occur, the set of target signatures generated from the $\mathbf{t}_1$ vectors may not adequately describe all the possible forms of the target image signatures that may be in the image. A simple solution to this problem may be to divide the image into smaller regions for which the assumption of a homogeneous atmosphere is valid, and apply AutoMatch to each region. Further research is required to assess the usefulness of this proposal.

Summary of Conclusions and Areas for Further Research

The preceding sections presented the AutoMatch detector as a new hyperspectral signature matching algorithm that can be applied to an arbitrary image without the need for atmospheric calibration. We also compared the AutoMatch detector to alternative configurations of the methodology and to a benchmark atmospheric-calibration-based method. Finally, we identified limitations of the algorithm and indicated areas in which it can be improved. We summarize the significant conclusions from this research in the following paragraphs:

1) When applied to a range of targets in different hyperspectral scenes, AutoMatch performed as well or better than the FLAASH-SAM detector which uses more sophisticated atmospheric calibration methods and requires more detailed knowledge of the hyperspectral image. This result demonstrates the validity of the in-scene calibration methods used by AutoMatch.

2) When using generic generation reflectance libraries that where not well-matched to the materials in the image scene, AutoMatch still performed as well or better than the FLAASH-SAM method. Thus, AutoMatch is a useful detection tool when the only information available to the user are the reflectance signatures for the target of interest. This characteristic of AutoMatch is consistent with our objective to develop an autonomous signature matching algorithm to complement the AutoDet anomaly detector.
3) The TCIMF method was consistently the best, or among the best, detectors tested for use in the AutoMatch methodology. Other methods we tested based on the algorithms of Healey and Slater and Thai and Healey performed well in some instances, but proved to be inconsistent across the range of targets and images tested. The TH and TH-CL methods we tested proved to be limited by the non-negativity test employed in these methods.
4) Inaccuracies in the generator reflectance libraries, the target reflectance libraries, and the estimation of the $\mathbf{t}_0$ vector, can lead to highly variable or inaccurate generated target image signatures. Depending on the degree of similarity between the target and background materials, these inaccurate generated signatures can lead to a large number of false alarms.

In order to improve upon the proposed AutoMatch detector, we suggest further research in the following areas:

1) Investigate methods for incorporating the variability of the generated target image signatures and the background signatures into the TCIMF detector.

2) Adapt AutoMatch to account for the possibility of non-homogeneous atmospheric conditions in a hyperspectral scene. This extension of AutoMatch is particularly important if the method is to be applied to scenes covering large geographic areas.
3) Investigate the use of constrained least-squares methods for use with the TH and TH-CL variants of AutoMatch to eliminate the non-negativity test. A starting point in this effort may be the Non-negativity Constrained Least-Squares (NCLS) algorithm discussed in Chang (2003).
4) Investigate the use of other methods or indices besides the NDVI and BI to select generator signatures. Ideally, such methods would narrow the list of possible identities of the generator material, thereby improving the accuracy of the generator reflectance library and AutoMatch detection performance.
5) Develop methods for fusing the output of the AutoDet and AutoMatch detectors to improve overall target detection accuracy, particularly against targets that are marginally detectable by either algorithm.

This last recommendation is, as stated in Chapter 1, the final objective of our proposed target detection framework. It is a particularly important research area since, as we have seen in this chapter and in the preceding chapter, some targets are exceedingly difficult to detect exclusively with anomaly detection or with signature matching methods. The primary difficulty with many of these challenging targets is that they lie in a region of uncertainty in which they are not clearly anomalous or signature matches, but they cannot be clearly designated as background materials either. It is hoped that the fusion of signature matching information with anomaly information will provide a

synergy that pushes the target declaration decision conclusively in one direction or another. Further research in this area should strive to validate this hypothesis.

VII. Summary of Contributions

In Chapter 1, we stated that the objectives of this research are to: develop a new anomaly detection methodology using multivariate outlier detection concepts; develop a signature matching target detection method that eliminates the need for atmospheric calibration; and ensure that both of these methodologies minimize the technical expertise and level of intervention required by the user. In meeting these objectives we have contributed to the technical body of knowledge in several areas, as summarized in the following paragraphs.

Anomaly Detection Contributions

Our primary contribution to the field of hyperspectral anomaly detection is the AutoDet methodology developed in Chapter 5. This methodology combines multivariate outlier detection methods with an automated *k*-means clustering scheme to improve anomaly detection accuracy relative to existing benchmark detectors. Secondary contributions stemming from the development of AutoDet include:

1) We demonstrated through simulated multivariate Gaussian data tests that anomalies can be masked from classical MSD detectors with as little as 2.4% contamination. Additionally, our tests showed that the shape of a material's covariance matrix estimate, as represented by the orientation of the first principal component axis, can be significantly distorted with as little as 0.5% contamination.
2) It was confirmed through experiments with heavy-tailed, multivariate *t*-distributed data, that multivariate Gaussian outlier detection methods are more

effective at finding outliers in this heavy-tailed data than non-robust MSD detectors, even if the distribution of the MSDs for the multivariate *t*-distributed data is known with certainty. In other words, when searching for anomalies in heavy-tailed data, it may be better to incorrectly make a Gaussian assumption for the data distribution as opposed to computing the MSDs using a contaminated covariance matrix estimate.

3) Controlled experiments using simulated Gaussian and multivariate-*t* data showed that the BACON and FAST-MCD outlier detectors are effective at detecting outliers in datasets with high-dimensionality and large numbers of observations. This contribution is significant since none of the algorithms proposed in the multivariate outlier detection literature have been shown to be scaleable to datasets comparable in size and dimension to hyperspectral data.
4) Our use of Taguchi robust parameter design to determine a robust configuration for AutoDet that performs well across a range of images and targets is a novel approach to anomaly detector design. Based on our literature review of anomaly detection methods, guidance is often lacking on the best way to set the input parameters of proposed algorithms, and if settings that are useful for one image are also useful for other images. Through our tests, we show that robust parameter design methods can be a useful tool in this endeavor.
5) Our comparison tests between the AutoDet methodology and benchmark anomaly detection methods show the superior performance of multivariate outlier detection methods in finding anomalies relative to detectors that

employ non-robust statistical methods. Specifically, our tests showed that AutoDet outperformed the SSRX and CBAD detectors when applied to a range of images containing a variety of targets.

Image Clustering Contributions

As stated in Chapter 5, multivariate outlier detection methods generally assume that the *good* data in a sample comes from a single distribution and that any outliers in the sample come from one or more different distributions. Since a typical hyperspectral image consisting of multiple background materials does not satisfy this assumption, the image data must be clustered into homogeneous groups of signatures and the outlier detection methods applied to each group. In justifying the ability of the commonly used *k*-means algorithm for this purpose, we produced the following contributions:

1) We used empirical tests to demonstrate the *k*-means clustering algorithm with Cosine assignment rule is adequate for accurately clustering hyperspectral data into homogenous groups when the data is contaminated by outlying observations. In our simulated data tests using both Gaussian and multivariate *t*-distributed data, this method accurately clustered data with contamination levels up to 9.1% when outliers are dispersed in the high-dimensional space. In the presence of highly-concentrated, distant outliers, the tolerated contamination level can drop to 3.6% or less, depending on the background data.
2) Our tests confirmed that a primary failure mode of the *k*-means algorithm using the Cosine assignment rule is the *clumping effect* in which outliers are assigned to their own cluster and background materials are grouped together.

We showed that the clumping effect is more likely to occur when: a) outliers are highly concentrated and distant from the background materials; b) background material are spectrally similar to one another; and c) the background clusters closest to the outliers contain relatively few observations.

3) Five statistically-based *k*-selection methods were compared using simulated multivariate Gaussian and multivariate *t*-distributed data. These experiments revealed the Silhouette and Calinski-Harabasz methods to be most effective at estimating the number of clusters in a dataset, relative to the other methods tested. This conclusion was verified using datasets with different background materials and a range of known values of *k*.
4) When applied to actual hyperspectral images, our tests demonstrate that our proposed Color Method, which estimates *k* based on the number of colors detected in the visible region of the hyperspectral image, performed comparably to the Silhouette and Calinski-Harabasz methods without the need for any data preprocessing. Though the appropriate choice of image pre-processing allowed the Silhouette and Calinski-Harabasz methods to give reasonable results, we found that the choice of pre-processing method was image-dependent. A further advantage of the Color Method is its ability to produce estimates of *k* in fractions of a second, as opposed to minutes or hours for the other methods.

Signature Matching Contributions

The AutoMatch target detection algorithm is our primary contribution to the field of hyperspectral signature matching. This algorithm is unique in its use of the NDVI and

BI metrics, as well as cluster and anomaly information, to select background materials for in-scene calibration. AutoMatch also employs a novel approach for generating a set of possible image signatures for the target of interest that captures both the variability of target reflectance signatures and the uncertainty of the true identity of background materials used for the in-scene calibration. The most significant contribution of AutoMatch is its ability to detect a range of target materials while requiring the user to only specify target reflectance signatures and a generic library of either vegetation or soil reflectance signatures. Thus, AutoMatch bypasses the complexity of both detailed atmospheric calibration and the MODTRAN4-based methods introduced by Healey and Slater. Additional contributions stemming from the development of AutoMatch are as follows:

1) Our comparison tests between AutoMatch and the FLAASH-SAM algorithm demonstrate the ability of a nearly-autonomous, in-scene calibration signature matching algorithm to perform as well or better than an algorithm using detailed atmospheric correction. These tests used 64 types of target materials and seven hyperspectral images to verify performance results, making our tests more comprehensive than any tests presented in the technical literature.
2) We demonstrate through experimental testing with actual hyperspectral imagery that the Target-Constrained Interference-Minimized Filter (TCIMF) proposed by Ren and Change (2000) achieves better detection results with our generated target signatures than target subspace methods based on the methods of Healey and Slater and Thai and Healey. We also revealed that the non-negativity test contained in the Thai-Healey method can limit the utility

of the algorithm due to its subjective nature and potentially detrimental impact on detection results.

3) Analysis of our tests results revealed that certain target materials are extremely difficult to detect with any of the methods we tested. This insight is valuable in guiding future research into detection methods that are more effective in dealing with these targets.

Areas for Further Research

In both Chapters 5 and 6, we identify research areas that may lead to improvements in the AutoDet and AutoMatch methodologies. From these suggested research areas, we feel the following are most worthy of consideration:

1) Develop methods to accurately threshold the MSDs produced by the BACON and FAST-MCD detectors. The original method given by Billor, Hadi, and Velleman for BACON, and the scaling methods proposed for FAST-MCD assume the multivariate data is Gaussian, and hence use a quantile from the Chi-Square distribution as the MSD threshold. For hyperspectral data that deviates from the Gaussian assumption, a Chi-Square threshold can lead to an increase in false alarms and decrease the confidence the detection results.
2) Identify more accurate methods for automatically clustering hyperspectral image data. As we illustrated in Chapter 5, our combination of the Color Method with *k*-means provides a satisfactory means for automatic clustering, but more accurate solutions can further increase the detection accuracy of the AutoDet methodology.

3) Investigate methods for incorporating target and background signature variability in the TCIMF method used in AutoMatch. In the current implementation, we only use the mean vectors of the generated target signatures and background clusters in TCIMF. Though we achieved good detection results with this approach, the TCIMF detector is theoretically capable of accommodating a better representation of the target and background materials. The challenge in exploiting this capability, however, is determining the optimal representation of target and background variability that leads to the detection of more targets without reducing the separability between the two classes to the extent that false alarms increase.

In addition to these three research areas, we also advocate further development of the target detection framework described in Chapter 1. By using AutoDet and AutoMatch as the anomaly detection and signature matching components of this framework, the fusion methodology is the remaining piece of the architecture requiring development. Should a successful fusion method be devised, we believe that the completed target detection framework will provide an autonomous target detection method that is practical for a diverse set of users and that achieves higher detection accuracy than can be attained by either anomaly detection or signature matching alone.

Appendix A: Signatures of Dispersed Outliers Used in *k*-Means Robustness Tests

This appendix contains the mean vectors of the materials used as dispersed outliers in the *k*-means robustness tests presented in Chapter 5. Figures 49, 50, and 51 give the mean vectors for the signatures taken from the Fort A.P. Hill, D.C. Mall, and Purdue University images, respectively. The error bars in each figure denote one standard deviation above and below the mean in each spectral band for the respective material signatures.

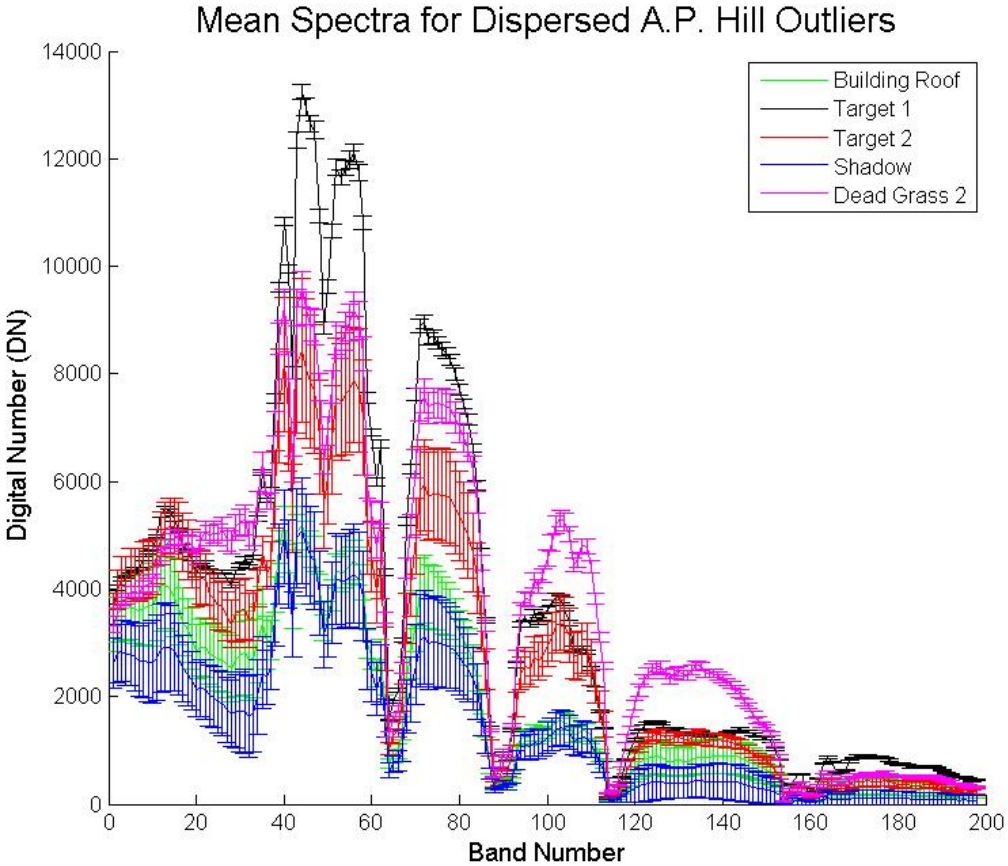


Figure 49. Signature Mean Vectors for Dispersed Fort A.P. Hill Outliers

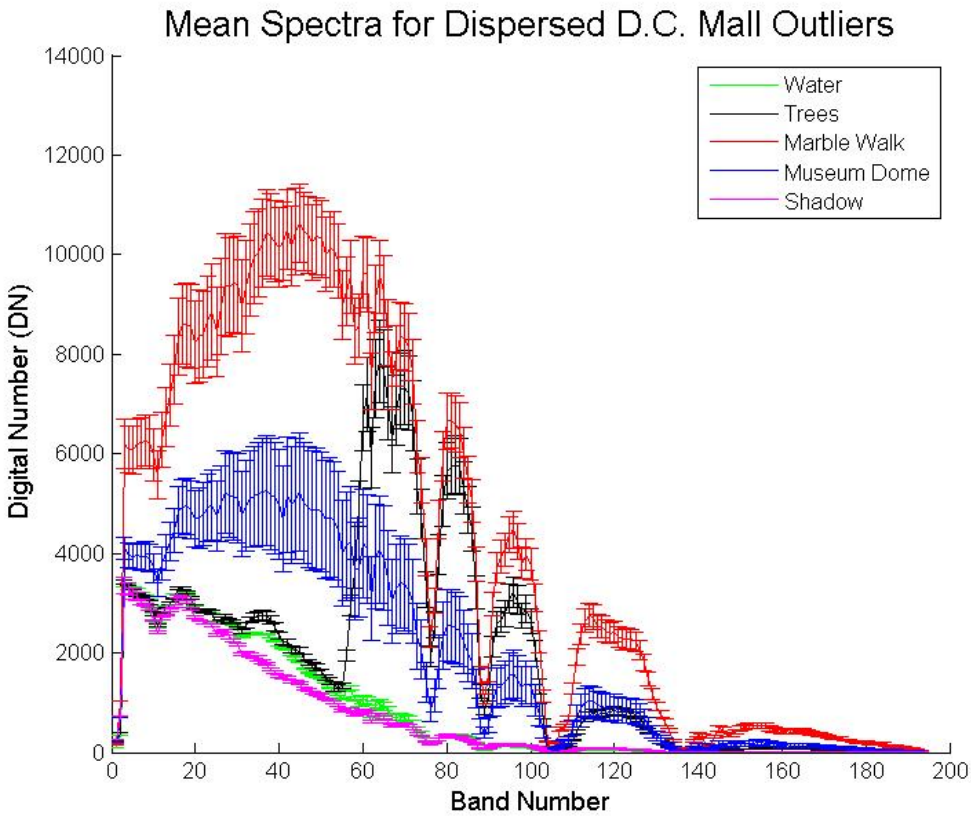


Figure 50. Signature Mean Vectors for Dispersed D.C. Mall Outliers

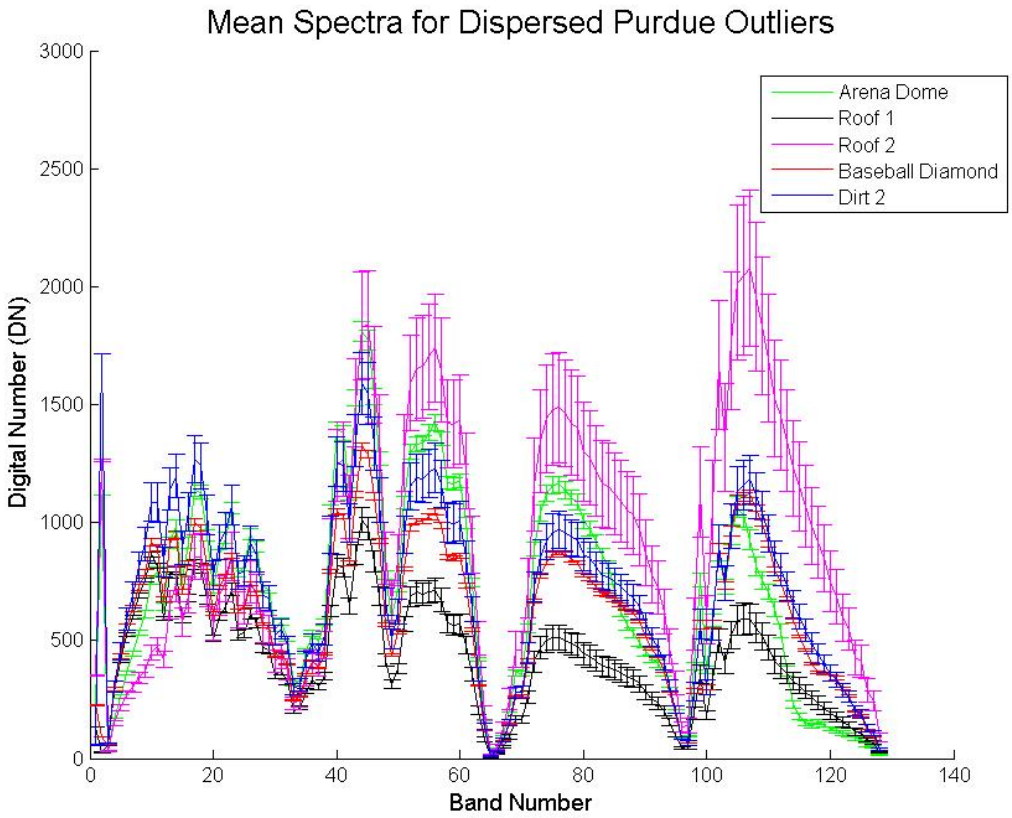


Figure 51. Signature Mean Vectors for Dispersed Purdue Outliers

Appendix B: Image Chips Used for *k*-Selection Tests

This appendix contains true color representations of the six hyperspectral images used to compare the Calinski-Harabasz, Silhouette, and Color methods in the *k*-selection tests of Chapter 5.

Figure 52. Image Chip 1 (Taken from Forest Radiance I Dataset)

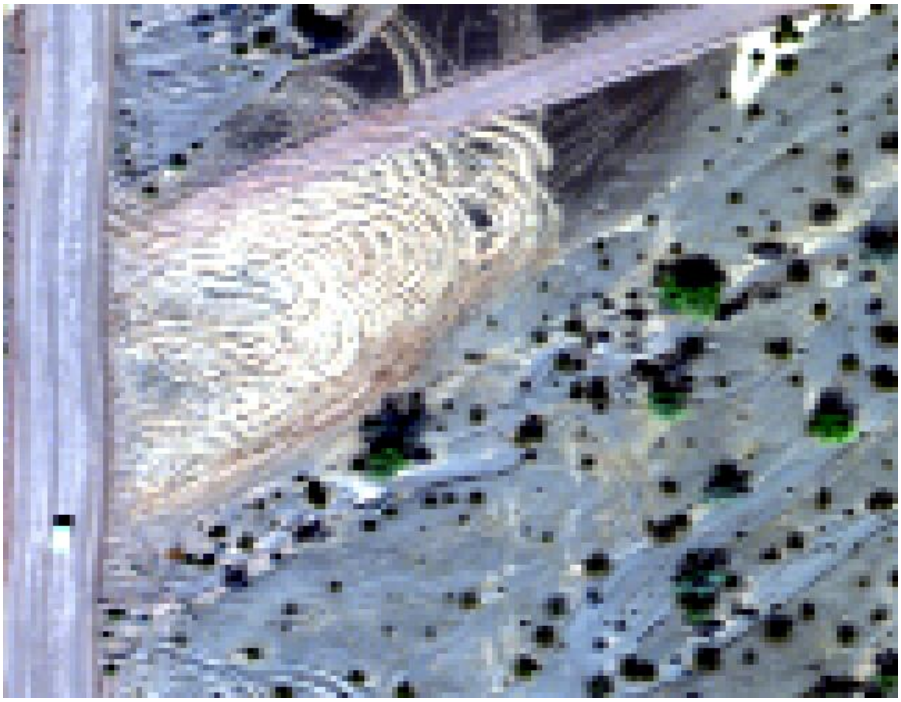


Figure 53. Image Chip 2 (Taken from Desert Radiance II Dataset)

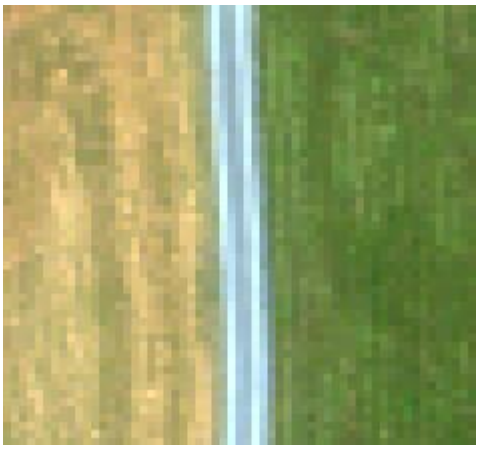


Figure 54. Image Chip 3 (Taken from Forest Radiance I Dataset)

Figure 55. Image Chip 4 (Taken from D.C. Mall AVIRIS Image)

Figure 56. Image Chip 5 (Taken from Purdue HYMAP Image)

Figure 57. Image Chip 6 (Taken from Purdue HYMAP Image)

Appendix C: Image Scenes

This appendix contains color renditions of the hyperspectral image scenes used throughout this dissertation. The numbering scheme used with some of the images (Scene 1, Scene 2, etc.) refers to a large set of images, some of which were not used in this dissertation; therefore, it may appear as though some images are missing. The originally naming convention was retained due to the extensive use of these names in other documents and computer code.

Figure 58. Fort A.P. Hill Image

Figure 59. D.C. Mall Image

Figure 60. Purdue University Image

Figure 61. Scene 1 (Taken from Forest Radiance I Dataset)

Figure 62. Scene 2 (Taken from Forest Radiance I Dataset)

Figure 63. Scene 3 (Taken from Fort A.P. Hill Image)

Figure 64. Scene 4 (Taken from Forest Radiance I Dataset)

Figure 65. Scene 5 (Taken from Desert Radiance II Dataset)

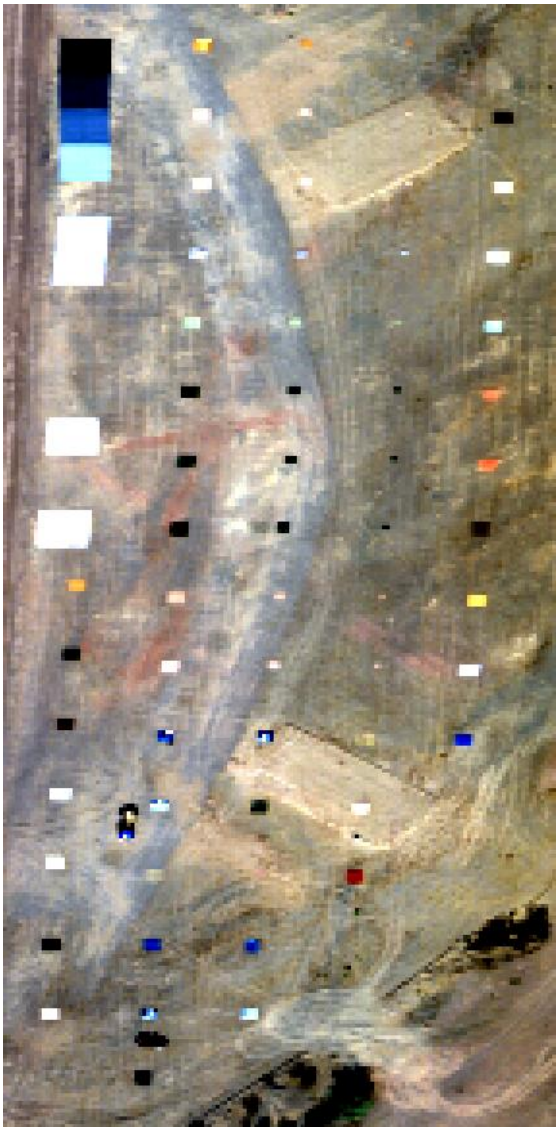


Figure 66. Scene 6 (Taken from Desert Radiance II Dataset)

Figure 67. Scene 7 (Taken from Desert Radiance II Dataset)

Figure 68. Scene 8 (Taken from Desert Radiance II Dataset)

Figure 69. Scene 9 (Taken from Desert Radiance II Dataset)

Figure 70. Scene 12 (Taken from Forest Radiance I Dataset)

Figure 71. Scene 13 (Taken from Forest Radiance I Dataset)

Figure 72. Scene 17 (Taken from Forest Radiance I Dataset)

Figure 73. Scene 19 (Taken from the MAD 98 Site 19 Data Fusion Dataset)

Appendix D: Taguchi Experimental Designs

This appendix contains the Taguchi experimental designes used in the robust parameter designs of the AutoDet-BACON and AutoDet-FASTMCD methods in Chapter 5. Definitions for the factors and levels can be found in Chapter 5.

Table 40. Experimental Design for AutoDet-BACON Robust Parameter Design

Design Point	Factor				
	A	B	C	D	E
1	-1	-1	-1	1	1
2	1	-1	-1	1	1
3	-1	1	-1	1	1
4	1	1	-1	1	1
5	-1	-1	1	1	1
6	1	-1	1	1	1
7	-1	1	1	1	1
8	1	1	1	1	1
9	-1	-1	-1	2	1
10	1	-1	-1	2	1
11	-1	1	-1	2	1
12	1	1	-1	2	1
13	-1	-1	1	2	1
14	1	-1	1	2	1
15	-1	1	1	2	1
16	1	1	1	2	1
17	-1	-1	-1	3	1
18	1	-1	-1	3	1
19	-1	1	-1	3	1
20	1	1	-1	3	1
21	-1	-1	1	3	1
22	1	-1	1	3	1
23	-1	1	1	3	1
24	1	1	1	3	1
25	-1	-1	-1	4	1
26	1	-1	-1	4	1
27	-1	1	-1	4	1
28	1	1	-1	4	1
29	-1	-1	1	4	1
30	1	-1	1	4	1
31	-1	1	1	4	1
32	1	1	1	4	1
33	-1	-1	-1	1	2
34	1	-1	-1	1	2
35	-1	1	-1	1	2
36	1	1	-1	1	2
37	-1	-1	1	1	2
38	1	-1	1	1	2
39	-1	1	1	1	2
40	1	1	1	1	2

Table 40 (con't)

Design Point	Factor				
	A	B	C	D	E
41	-1	-1	-1	2	2
42	1	-1	-1	2	2
43	-1	1	-1	2	2
44	1	1	-1	2	2
45	-1	-1	1	2	2
46	1	-1	1	2	2
47	-1	1	1	2	2
48	1	1	1	2	2
49	-1	-1	-1	3	2
50	1	-1	-1	3	2
51	-1	1	-1	3	2
52	1	1	-1	3	2
53	-1	-1	1	3	2
54	1	-1	1	3	2
55	-1	1	1	3	2
56	1	1	1	3	2
57	-1	-1	-1	4	2
58	1	-1	-1	4	2
59	-1	1	-1	4	2
60	1	1	-1	4	2
61	-1	-1	1	4	2
62	1	-1	1	4	2
63	-1	1	1	4	2
64	1	1	1	4	2
65	-1	-1	-1	1	3
66	1	-1	-1	1	3
67	-1	1	-1	1	3
68	1	1	-1	1	3
69	-1	-1	1	1	3
70	1	-1	1	1	3
71	-1	1	1	1	3
72	1	1	1	1	3
73	-1	-1	-1	2	3
74	1	-1	-1	2	3
75	-1	1	-1	2	3
76	1	1	-1	2	3
77	-1	-1	1	2	3
78	1	-1	1	2	3
79	-1	1	1	2	3
80	1	1	1	2	3

Table 40 (con't)

Design Point	Factor				
	A	B	C	D	E
81	-1	-1	-1	3	3
82	1	-1	-1	3	3
83	-1	1	-1	3	3
84	1	1	-1	3	3
85	-1	-1	1	3	3
86	1	-1	1	3	3
87	-1	1	1	3	3
88	1	1	1	3	3
89	-1	-1	-1	4	3
90	1	-1	-1	4	3
91	-1	1	-1	4	3
92	1	1	-1	4	3
93	-1	-1	1	4	3
94	1	-1	1	4	3
95	-1	1	1	4	3
96	1	1	1	4	3
97	-1	-1	-1	1	4
98	1	-1	-1	1	4
99	-1	1	-1	1	4
100	1	1	-1	1	4
101	-1	-1	1	1	4
102	1	-1	1	1	4
103	-1	1	1	1	4
104	1	1	1	1	4
105	-1	-1	-1	2	4
106	1	-1	-1	2	4
107	-1	1	-1	2	4
108	1	1	-1	2	4
109	-1	-1	1	2	4
110	1	-1	1	2	4
111	-1	1	1	2	4
112	1	1	1	2	4
113	-1	-1	-1	3	4
114	1	-1	-1	3	4
115	-1	1	-1	3	4
116	1	1	-1	3	4
117	-1	-1	1	3	4
118	1	-1	1	3	4
119	-1	1	1	3	4
120	1	1	1	3	4

Table 40 (con't)

Design Point	Factor				
	A	B	C	D	E
121	-1	-1	-1	4	4
122	1	-1	-1	4	4
123	-1	1	-1	4	4
124	1	1	-1	4	4
125	-1	-1	1	4	4
126	1	-1	1	4	4
127	-1	1	1	4	4
128	1	1	1	4	4
129	-1	-1	-1	1	5
130	1	-1	-1	1	5
131	-1	1	-1	1	5
132	1	1	-1	1	5
133	-1	-1	1	1	5
134	1	-1	1	1	5
135	-1	1	1	1	5
136	1	1	1	1	5
137	-1	-1	-1	2	5
138	1	-1	-1	2	5
139	-1	1	-1	2	5
140	1	1	-1	2	5
141	-1	-1	1	2	5
142	1	-1	1	2	5
143	-1	1	1	2	5
144	1	1	1	2	5
145	-1	-1	-1	3	5
146	1	-1	-1	3	5
147	-1	1	-1	3	5
148	1	1	-1	3	5
149	-1	-1	1	3	5
150	1	-1	1	3	5
151	-1	1	1	3	5
152	1	1	1	3	5
153	-1	-1	-1	4	5
154	1	-1	-1	4	5
155	-1	1	-1	4	5
156	1	1	-1	4	5
157	-1	-1	1	4	5
158	1	-1	1	4	5
159	-1	1	1	4	5
160	1	1	1	4	5

Table 40 (con't)

Design Point	Factor				
	A	B	C	D	E
161	-1	-1	-1	1	6
162	1	-1	-1	1	6
163	-1	1	-1	1	6
164	1	1	-1	1	6
165	-1	-1	1	1	6
166	1	-1	1	1	6
167	-1	1	1	1	6
168	1	1	1	1	6
169	-1	-1	-1	2	6
170	1	-1	-1	2	6
171	-1	1	-1	2	6
172	1	1	-1	2	6
173	-1	-1	1	2	6
174	1	-1	1	2	6
175	-1	1	1	2	6
176	1	1	1	2	6
177	-1	-1	-1	3	6
178	1	-1	-1	3	6
179	-1	1	-1	3	6
180	1	1	-1	3	6
181	-1	-1	1	3	6
182	1	-1	1	3	6
183	-1	1	1	3	6
184	1	1	1	3	6
185	-1	-1	-1	4	6
186	1	-1	-1	4	6
187	-1	1	-1	4	6
188	1	1	-1	4	6
189	-1	-1	1	4	6
190	1	-1	1	4	6
191	-1	1	1	4	6
192	1	1	1	4	6

Table 41. Experimental Design for AutoDet-FASTMCD Robust Parameter Design

Design Point	Factor			
	A	B	C	D
1	-1	-1	1	1
2	1	-1	1	1
3	-1	1	1	1
4	1	1	1	1
5	-1	-1	2	1
6	1	-1	2	1
7	-1	1	2	1
8	1	1	2	1
9	-1	-1	3	1
10	1	-1	3	1
11	-1	1	3	1
12	1	1	3	1
13	-1	-1	4	1
14	1	-1	4	1
15	-1	1	4	1
16	1	1	4	1
17	-1	-1	1	2
18	1	-1	1	2
19	-1	1	1	2
20	1	1	1	2
21	-1	-1	2	2
22	1	-1	2	2
23	-1	1	2	2
24	1	1	2	2
25	-1	-1	3	2
26	1	-1	3	2
27	-1	1	3	2
28	1	1	3	2
29	-1	-1	4	2
30	1	-1	4	2
31	-1	1	4	2
32	1	1	4	2
33	-1	-1	1	3
34	1	-1	1	3
35	-1	1	1	3
36	1	1	1	3
37	-1	-1	2	3
38	1	-1	2	3
39	-1	1	2	3
40	1	1	2	3

Table 41 (con't)

Design Point	Factor			
	A	B	C	D
41	-1	-1	3	3
42	1	-1	3	3
43	-1	1	3	3
44	1	1	3	3
45	-1	-1	4	3
46	1	-1	4	3
47	-1	1	4	3
48	1	1	4	3
49	-1	-1	1	4
50	1	-1	1	4
51	-1	1	1	4
52	1	1	1	4
53	-1	-1	2	4
54	1	-1	2	4
55	-1	1	2	4
56	1	1	2	4
57	-1	-1	3	4
58	1	-1	3	4
59	-1	1	3	4
60	1	1	3	4
61	-1	-1	4	4
62	1	-1	4	4
63	-1	1	4	4
64	1	1	4	4
65	-1	-1	1	5
66	1	-1	1	5
67	-1	1	1	5
68	1	1	1	5
69	-1	-1	2	5
70	1	-1	2	5
71	-1	1	2	5
72	1	1	2	5
73	-1	-1	3	5
74	1	-1	3	5
75	-1	1	3	5
76	1	1	3	5
77	-1	-1	4	5
78	1	-1	4	5
79	-1	1	4	5
80	1	1	4	5

Table 41 (con't)

Design Point	Factor			
	A	B	C	D
81	-1	-1	1	6
82	1	-1	1	6
83	-1	1	1	6
84	1	1	1	6
85	-1	-1	2	6
86	1	-1	2	6
87	-1	1	2	6
88	1	1	2	6
89	-1	-1	3	6
90	1	-1	3	6
91	-1	1	3	6
92	1	1	3	6
93	-1	-1	4	6
94	1	-1	4	6
95	-1	1	4	6
96	1	1	4	6

Appendix E: Taguchi Main Effects and Interaction Plots

This appendix contains the main effects and interaction plots for the robust parameter design experiments presented in Chapter 5. Detailed definitions of the factors and levels used in these plots can be found in Chapter 5.

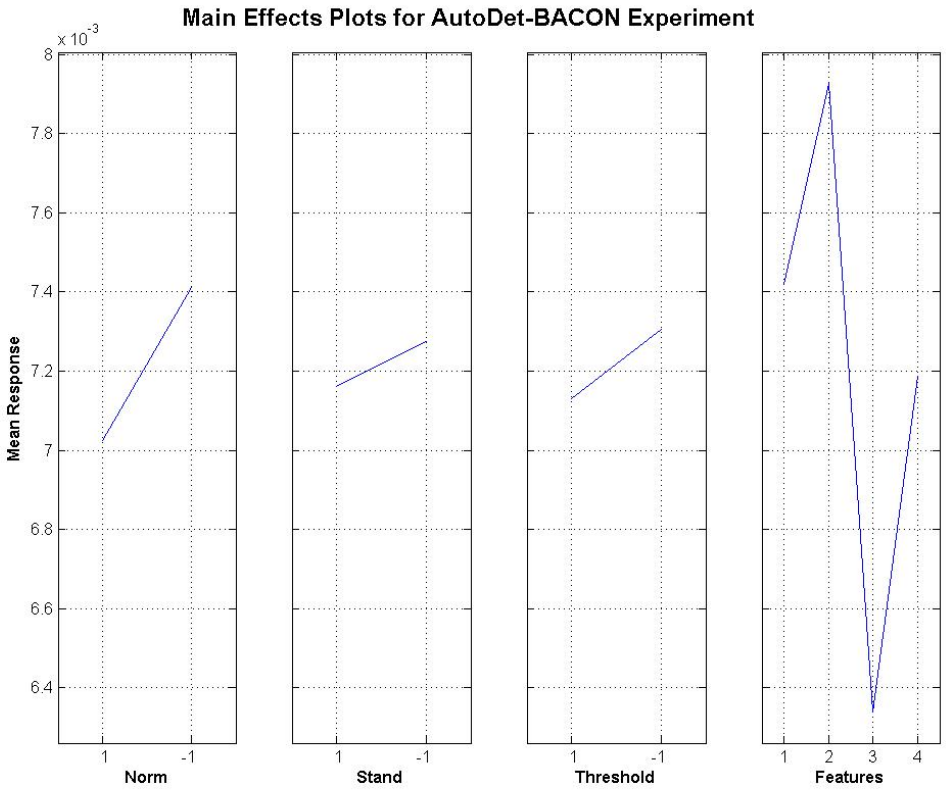


Figure 74. Main Effect Plot for AutoDet-BACON Experiment

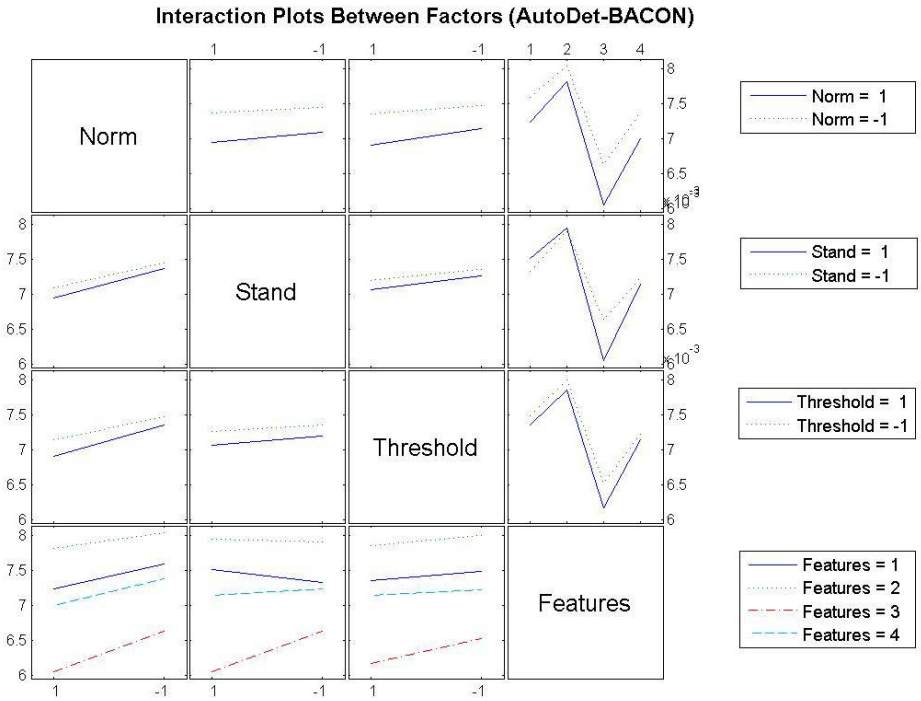


Figure 75. Interaction Plots for Main Factors (AutoDet-BACON)

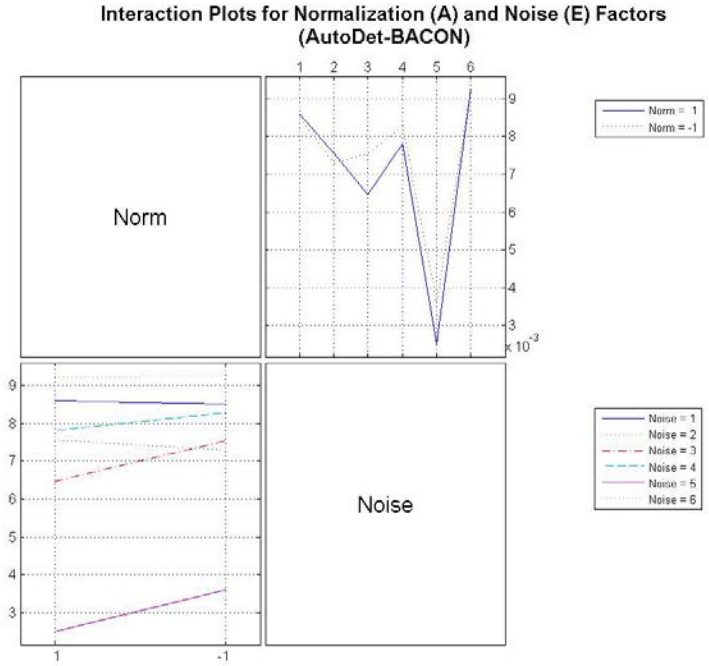


Figure 76. Normalization-Noise Interaction Plots (AutoDet-BACON)

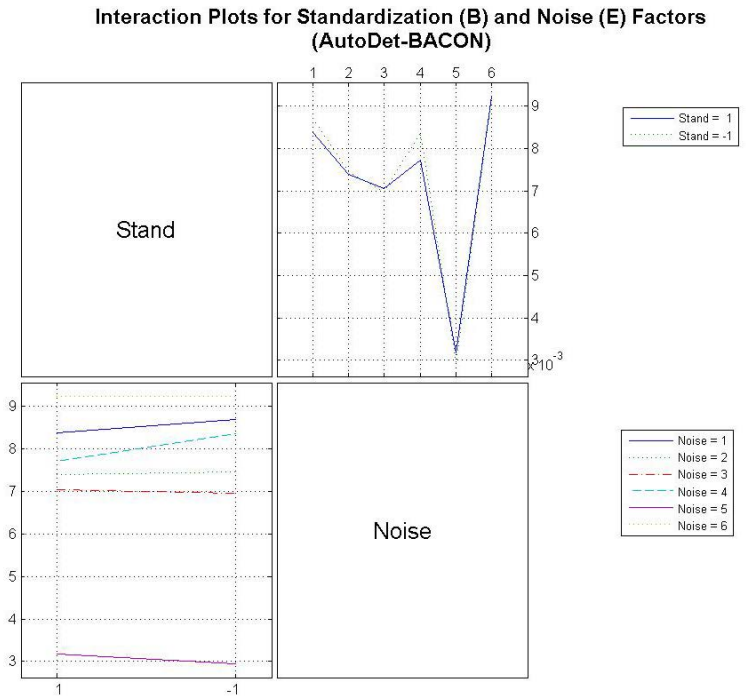


Figure 77. Standardization-Noise Interaction Plots (AutoDet-BACON)

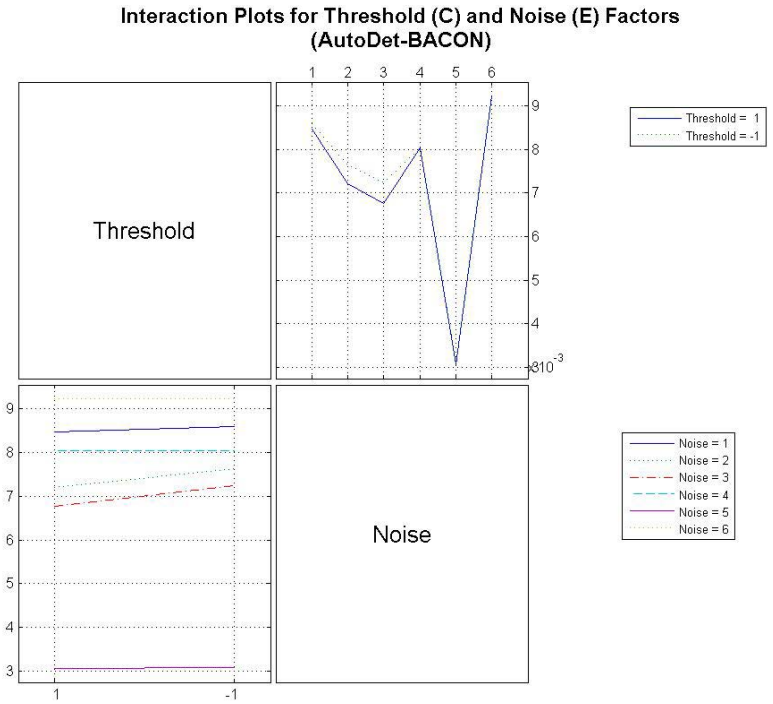


Figure 78. Threshold-Noise Interaction Plots (AutoDet-BACON)

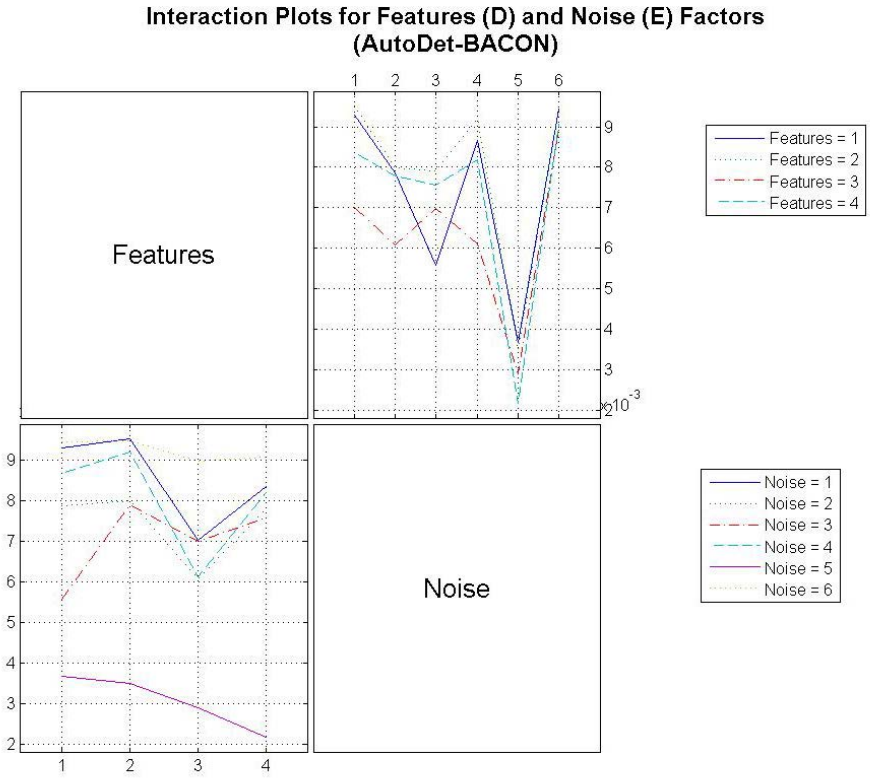


Figure 79. Features-Noise Interaction Plots (AutoDet-BACON)

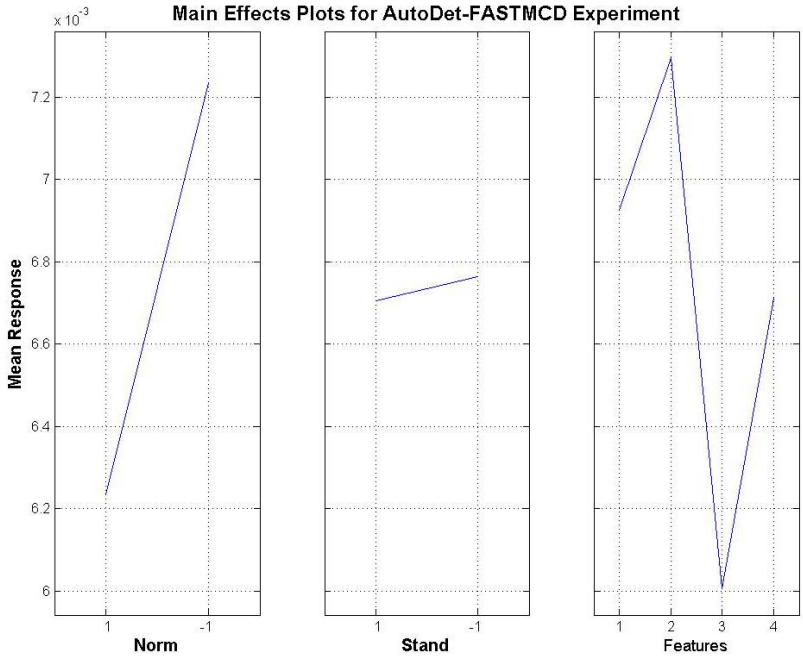


Figure 80. Main Effects Plot for AutoDet-FASTMCD Experiment)

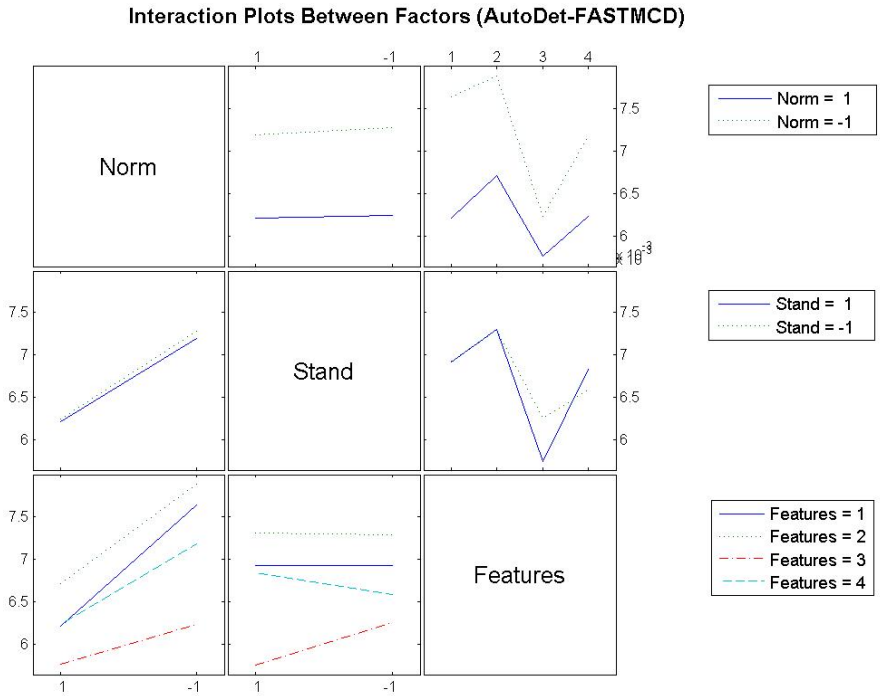


Figure 81. Interaction Plots for Main Effects (AutoDet-FASTMCD)

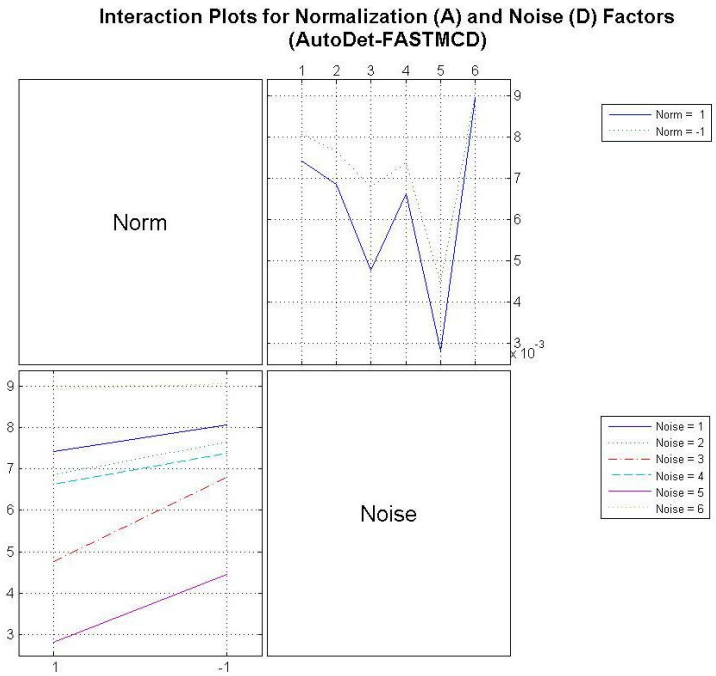


Figure 82. Normalization-Noise Interaction Plots (AutoDet-FASTMCD)

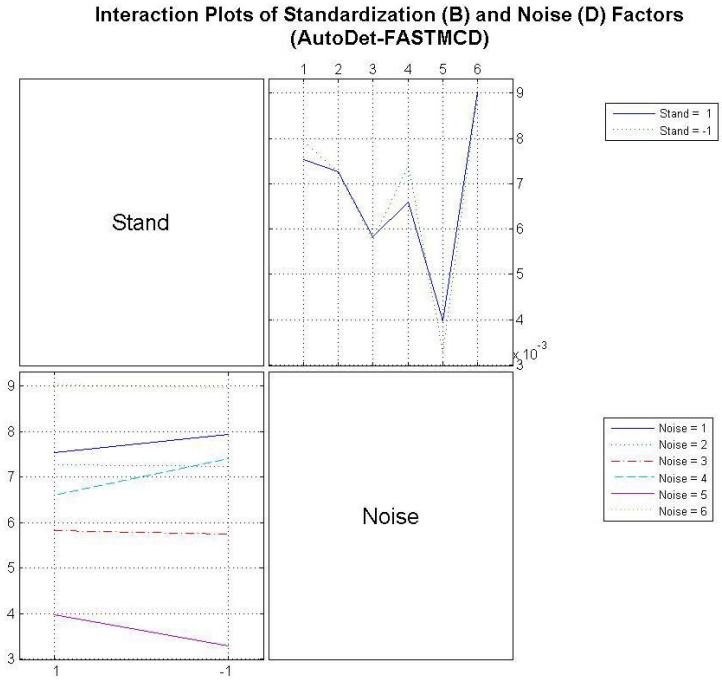


Figure 83. Standardization-Noise Interaction Plots (AutoDet-FASTMCD)

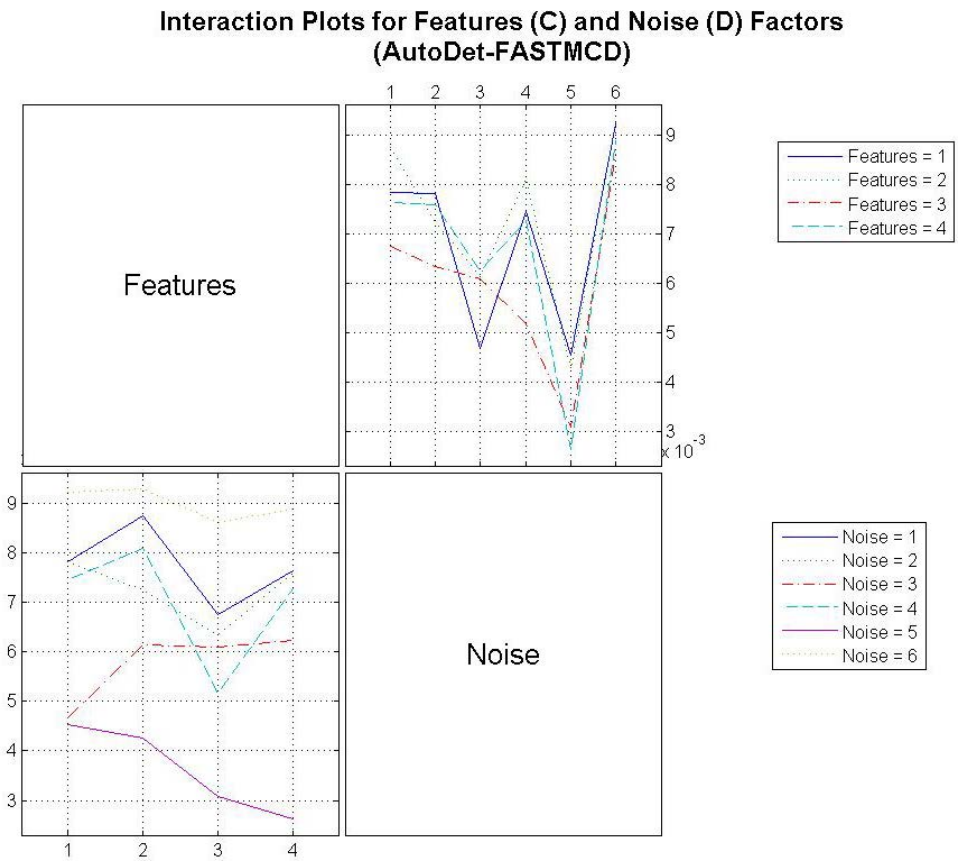


Figure 84. Features-Noise Interaction Plots (AutoDet-FASTMCD)

Appendix F: Anomaly Detector Comparison Test Output Images

This appendix contains the output images for the anomaly detection comparison tests performed in Chapter 5. For each of the images used in the test, the Mahalanobis Squared Distance (MSD) images, and binary target images produced by each detector are given. The target mask for each image is also provided. When viewing these images, several factors must be considered. First, the target images are based on a threshold value that may not be optimal for the respective detection methods. For the AutoDet-BACON, CBAD, and SSRX methods, the threshold is set to the 0.9999-quantile of the Chi-Square distribution with p=30 degrees of freedom. This distribution was used based on a Guassian assumption for the data that may have varying degrees of validity for the different detectors. For the AutoDet-FASTMCD method, the threshold is established using the zero-slope method discussed in Chapter 5 since the MSDs from this detector are produced from a trimmed elliptically-contoured distribution for which the distribution of corresponding MSDs is not known. The consequence of thresholding the MSDs in these ways is that the resulting thresholds will likely lie at different quantiles of the actual MSD distributions for the different detectors. In other words, the target images generally do not represent the same region of the respective OC curves for each detector.

The second factor to consider when viewing the images in this appendix is that the gray-scale MSD images are also based on a threshold that establishes the bin widths for the 256 shades of gray. Because the MSDs for some outliers are extremely high, setting the bin widths based on the lowest and highest MSD values will generally produce a black image with only one or two white dots. To make better use of the dynamic range of the gray-scale, we use the 0.9999-quantile of the Chi-Square distribution with p=30

degrees of freedom as the maximum value for the AutoDet-BACON, CBAD, and SSRX methods. This threshold seems reasonable since it will cause all pixels that appear as outliers in the target image to appear white, while better illustrating the relative MSD values for the remaining pixels. For the AutoDet-FASTMCD method we use the 0.97-quantile of all the MSDs since, as before, the Chi-Square distribution assumption for the FAST-MCD MSDs is much less valid. The end result of this strategy for creating the gray-scale MSD images is that comparisons between them should be avoided. Rather, these images should be used to better understand the strengths and weaknesses of the respective detectors.

As a final note concerning the images in the appendix, the color scheme used in the target masks is as follows: black represents non-target pixels in the image; white denotes target pixels used in the OC curve computations in Chapter 5; and red signifies border pixels for which the identity of the pixel could not be verified. This latter category of pixels is not included in OC curve computations.

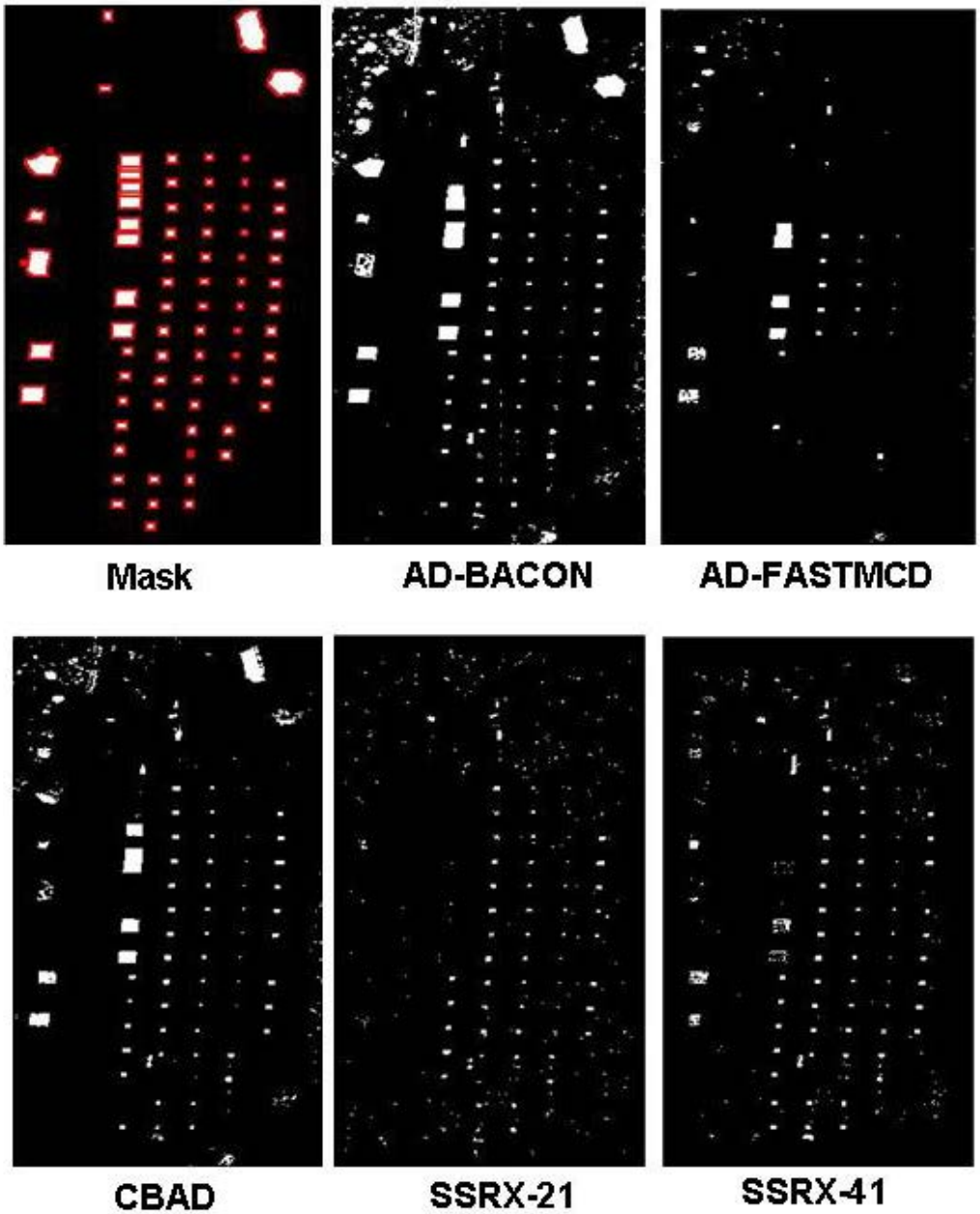


Figure 85. Target Images for Anomaly Detector Comparisons (Scene 5)

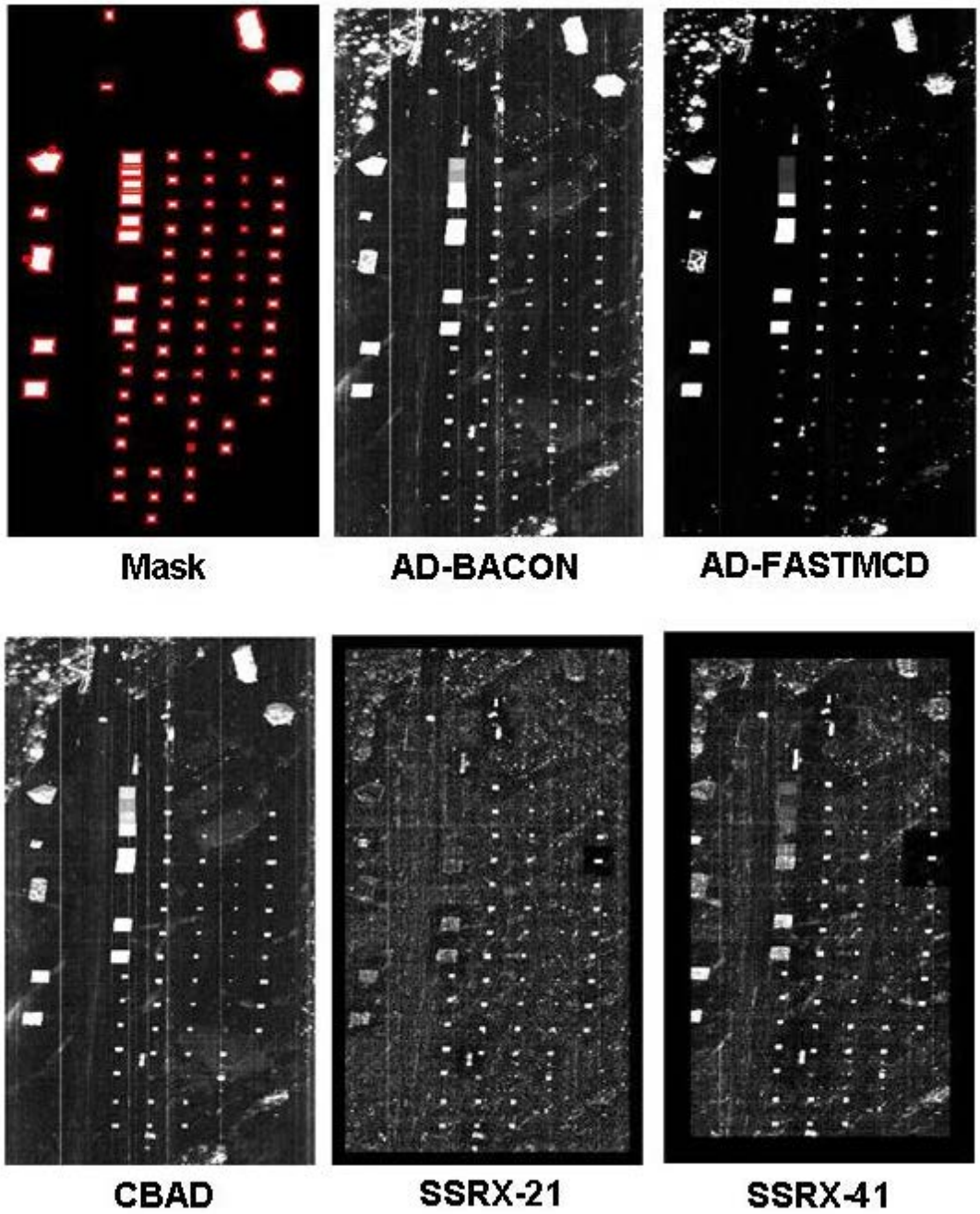


Figure 86. MSD Images for Anomaly Detector Comparisons (Scene 5)

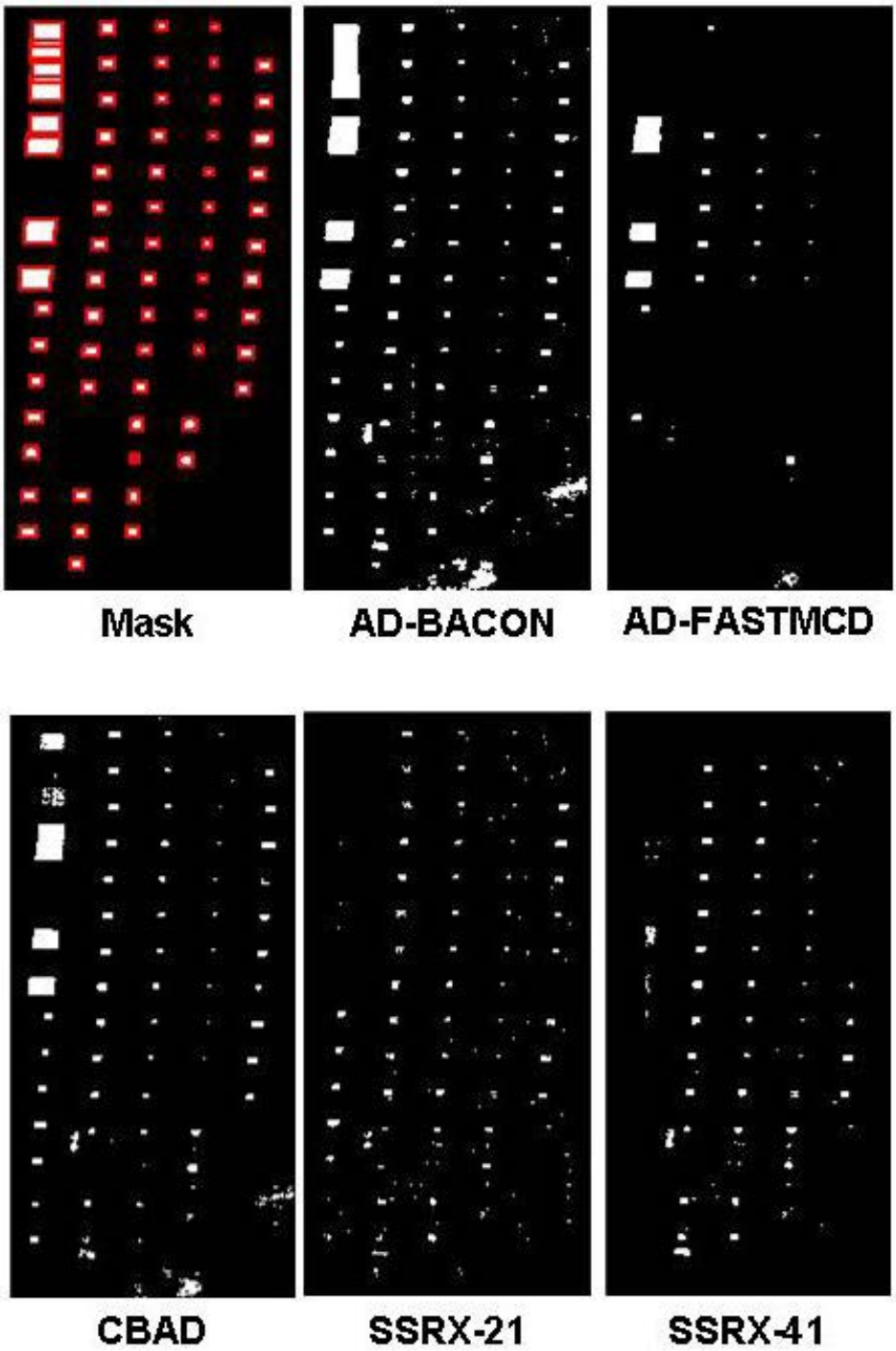


Figure 87. Target Images for Anomaly Detector Comparisons (Scene 6)

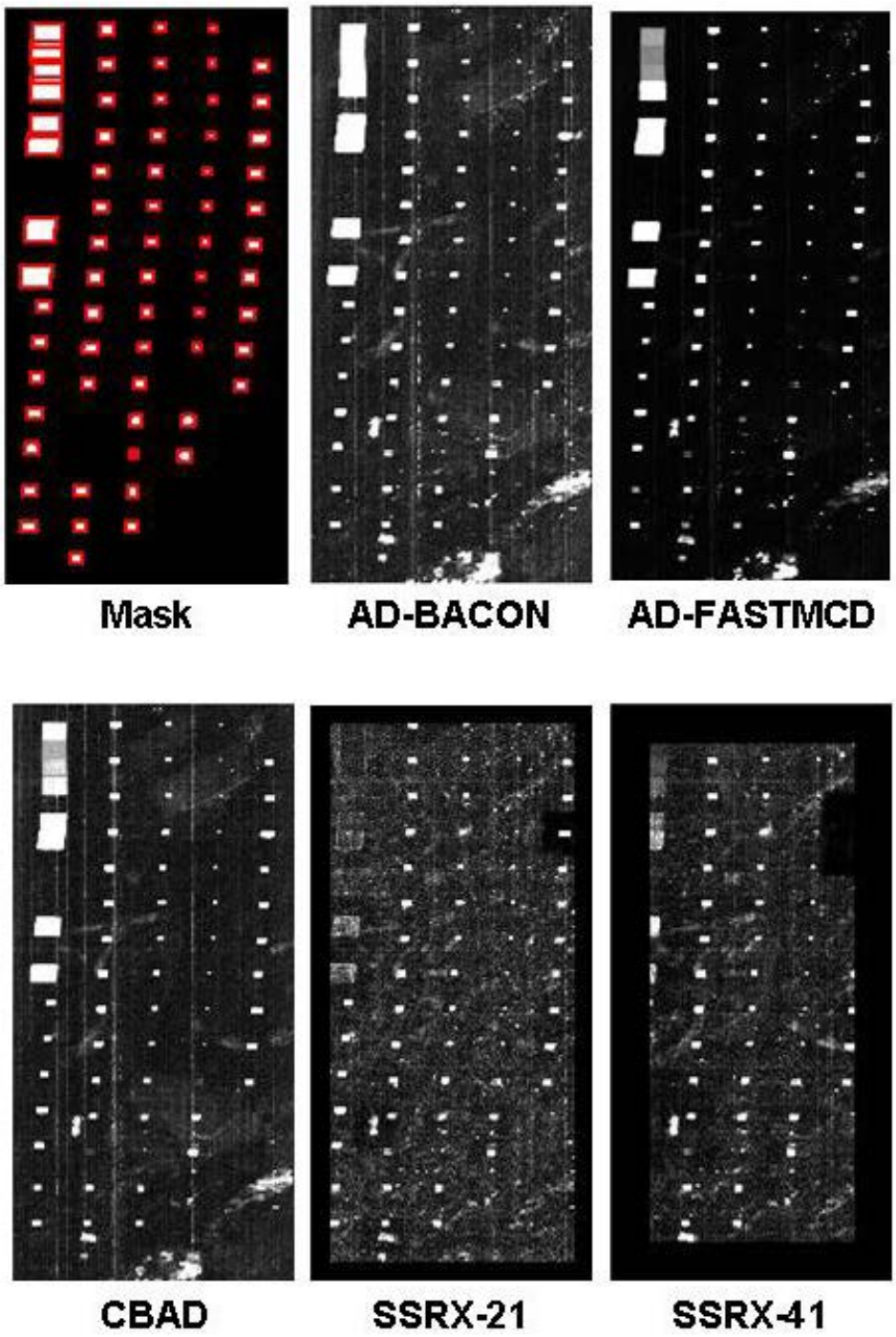


Figure 88. MSD Images for Anomaly Detector Comparisons (Scene 6)

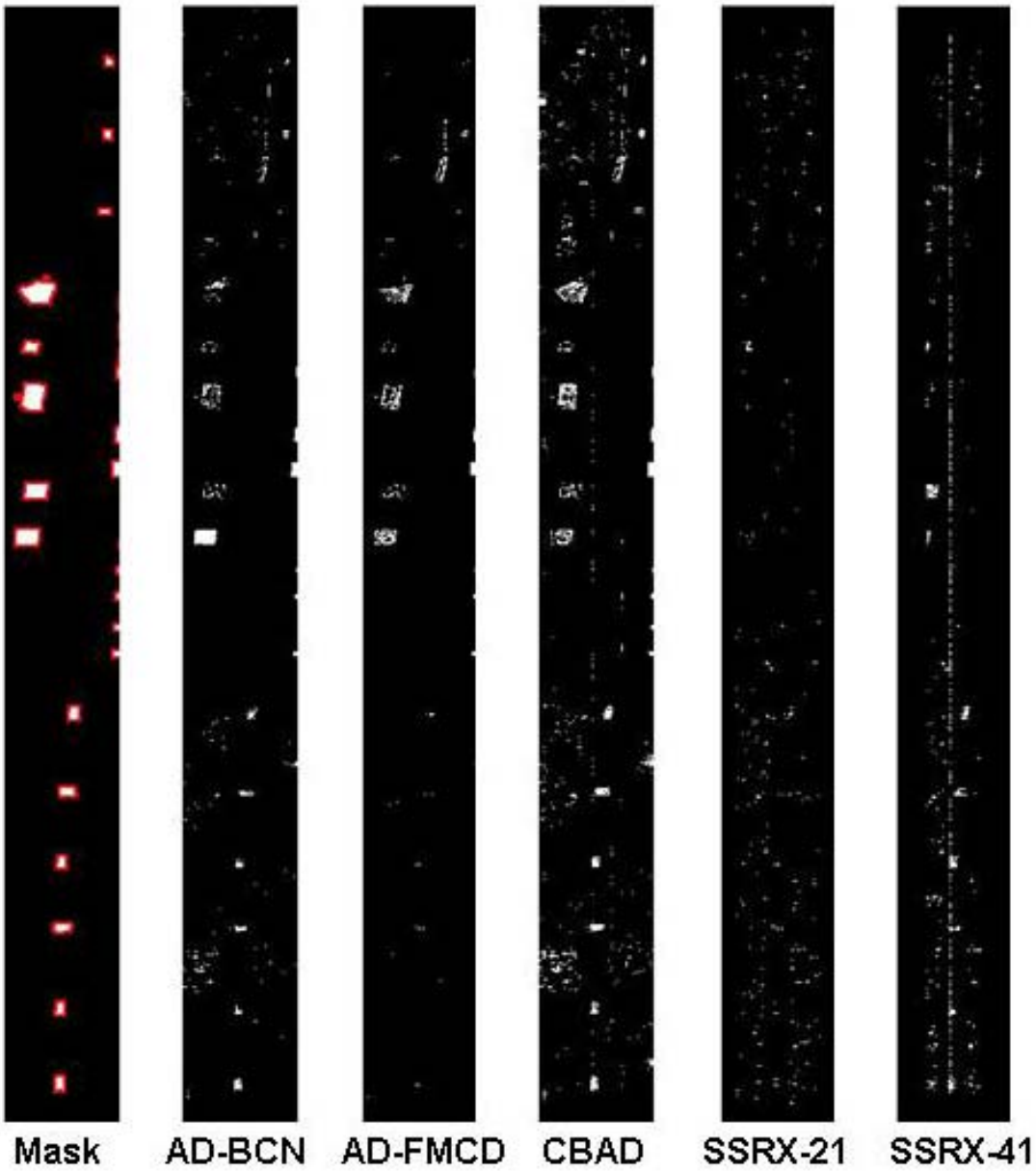


Figure 89. Target Images for Anomaly Detector Comparisons (Scene 7)

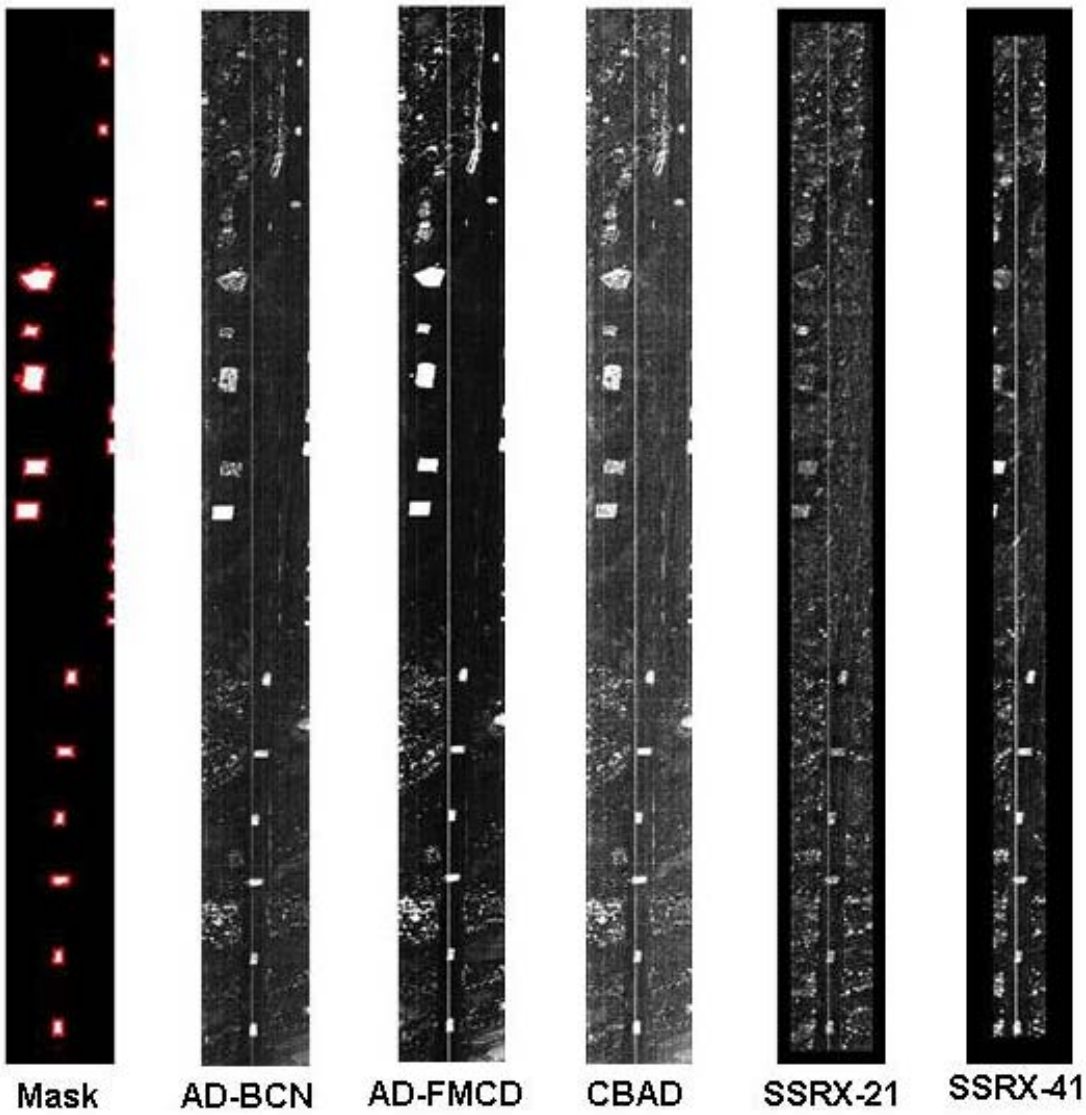


Figure 90. MSD Images for Anomaly Detector Comparisons (Scene 7)

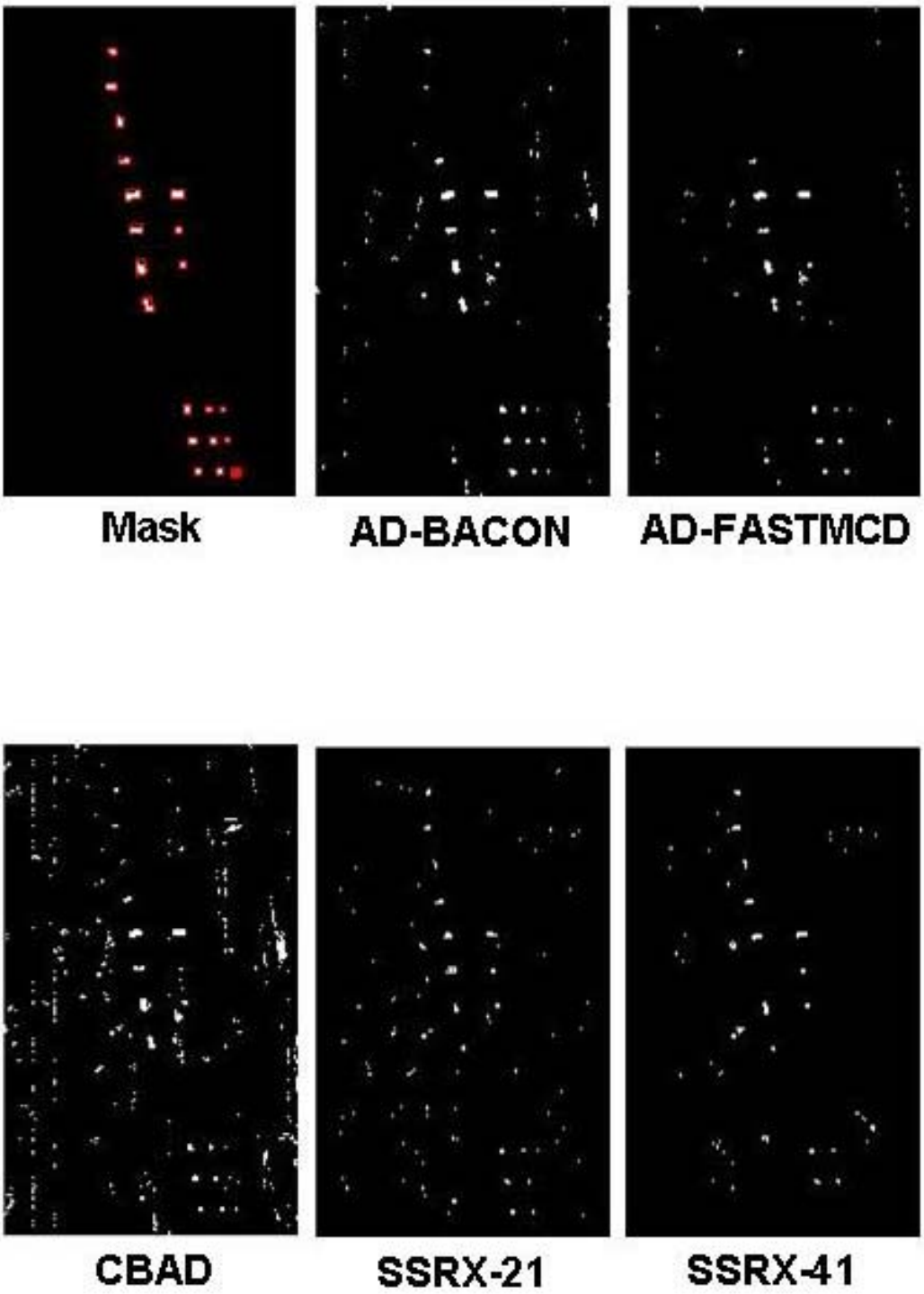


Figure 91. Target Images for Anomaly Detector Comparisons (Scene 12)

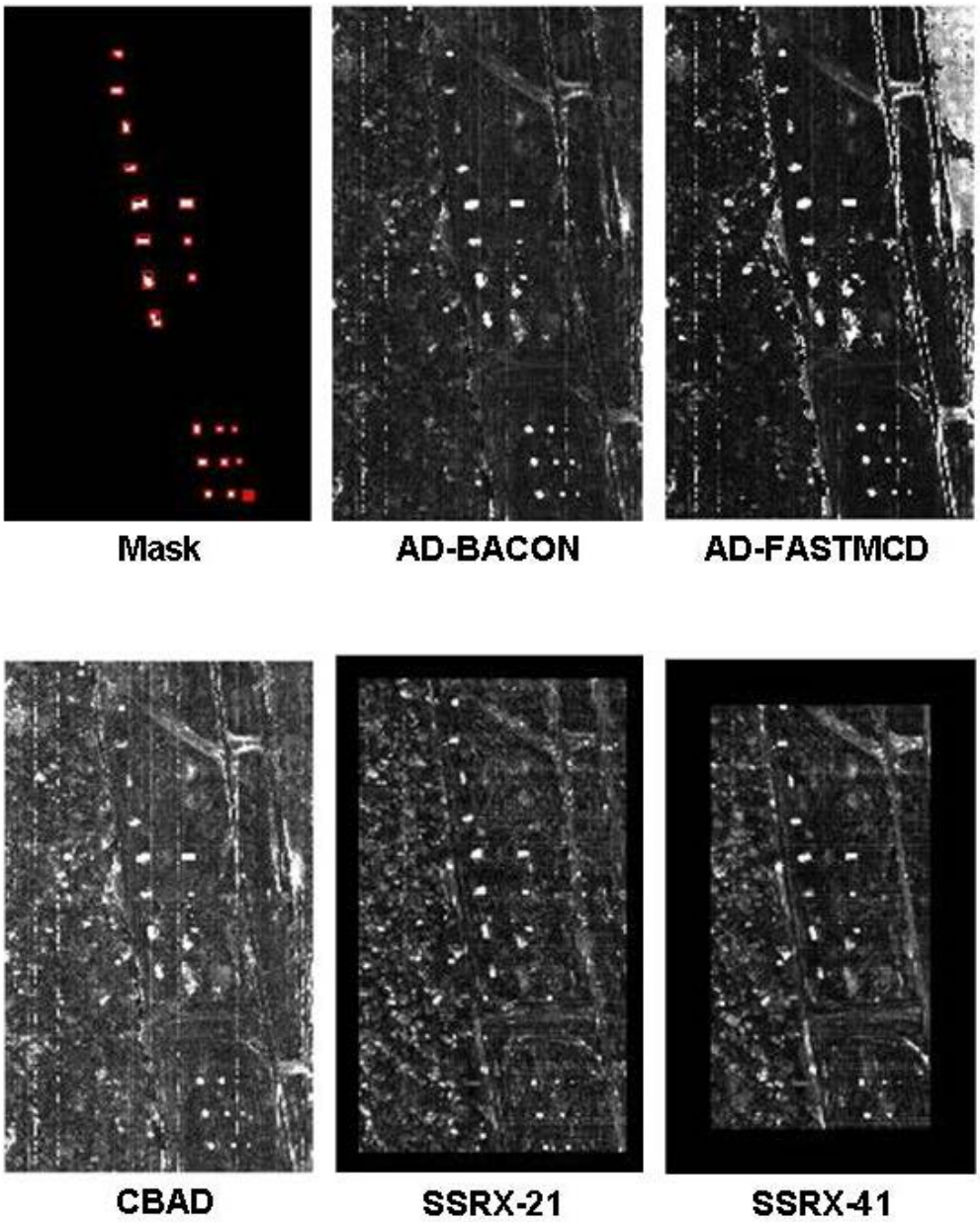


Figure 92. MSD Images for Anomaly Detector Comparisons (Scene 12)

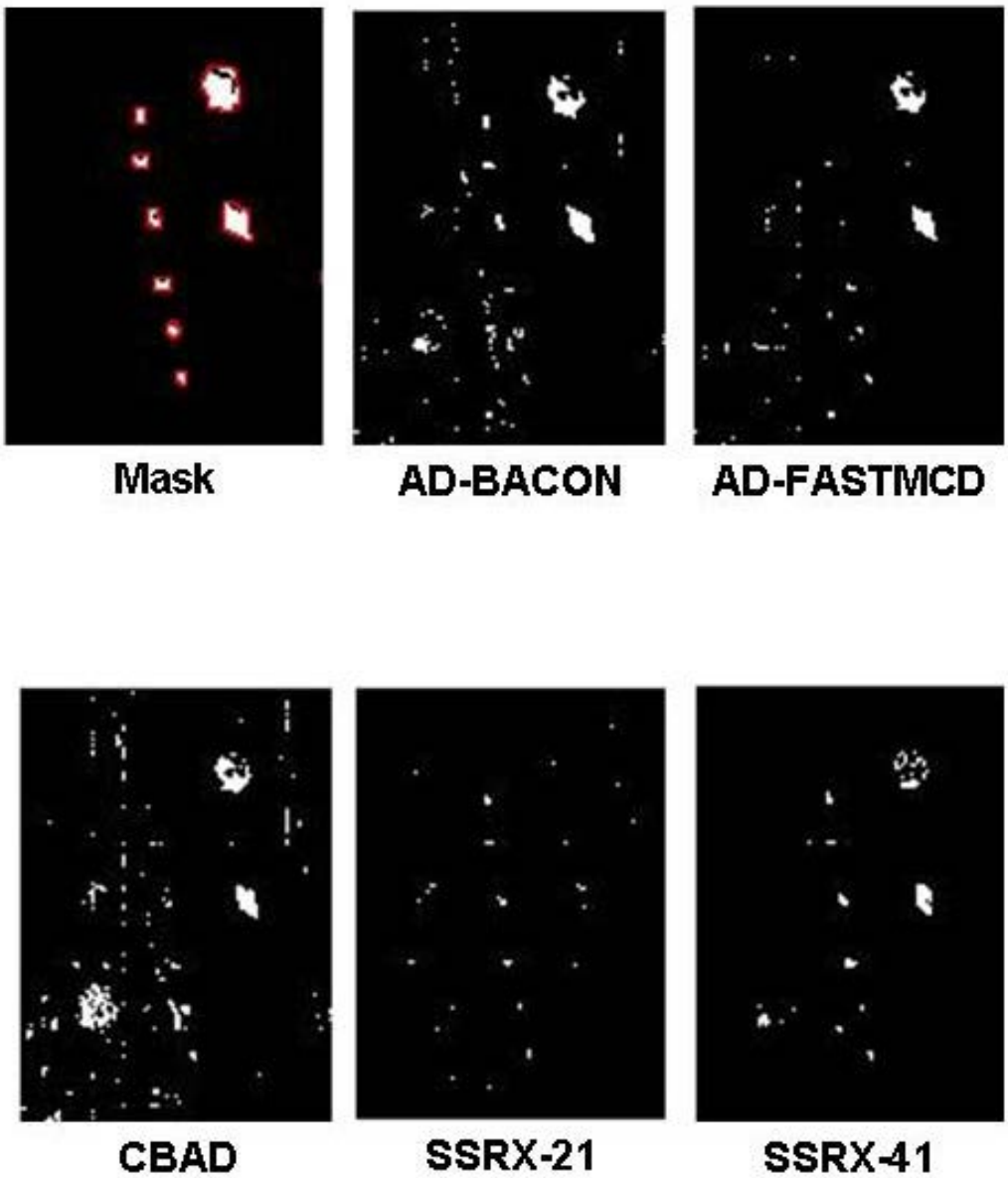


Figure 93. Target Images for Anomaly Detector Comparisons (Scene 13)

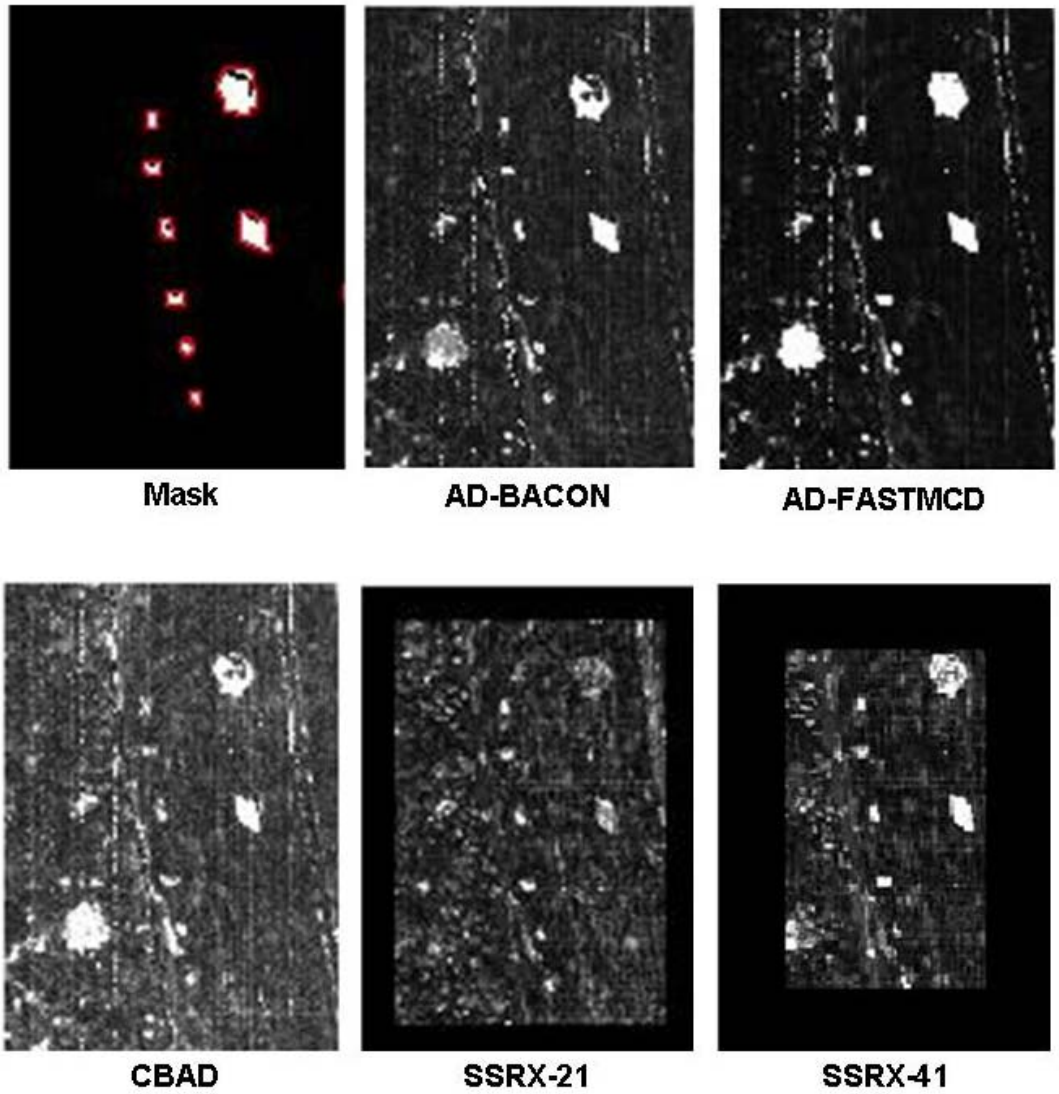


Figure 94. MSD Images for Anomaly Detector Comparisons (Scene 13)

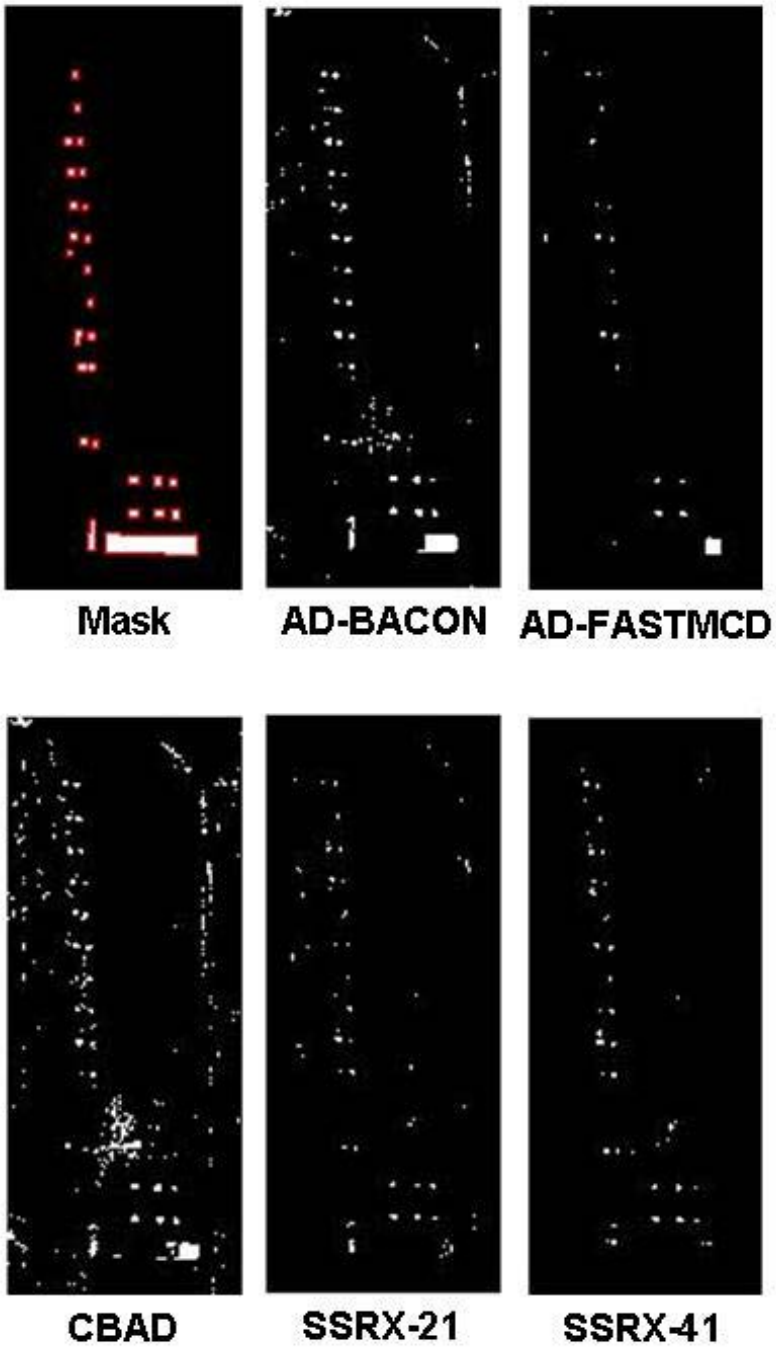


Figure 95. Target Images for Anomaly Detector Comparisons (Scene 17)

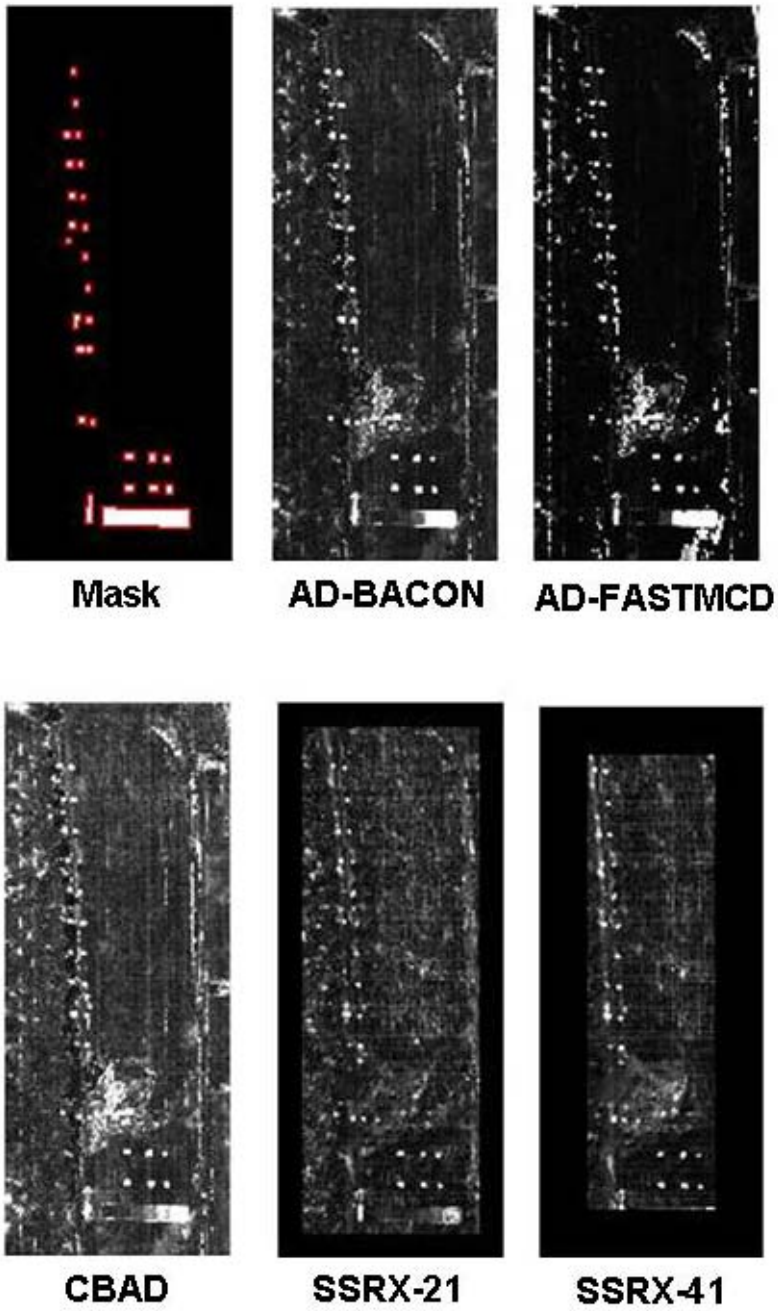


Figure 96. MSD Images for Anomaly Detector Comparisons (Scene 17)

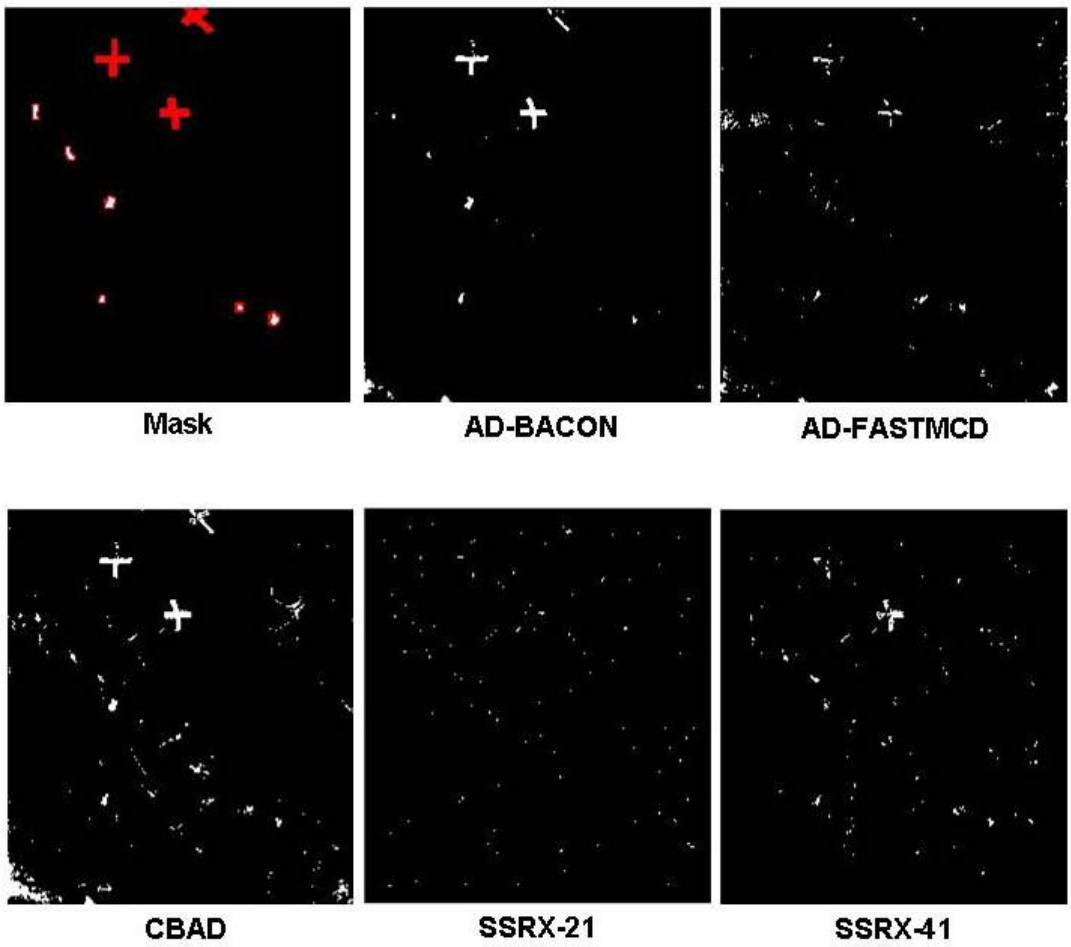


Figure 97. Target Images for Anomaly Detector Comparisons (Scene 19)

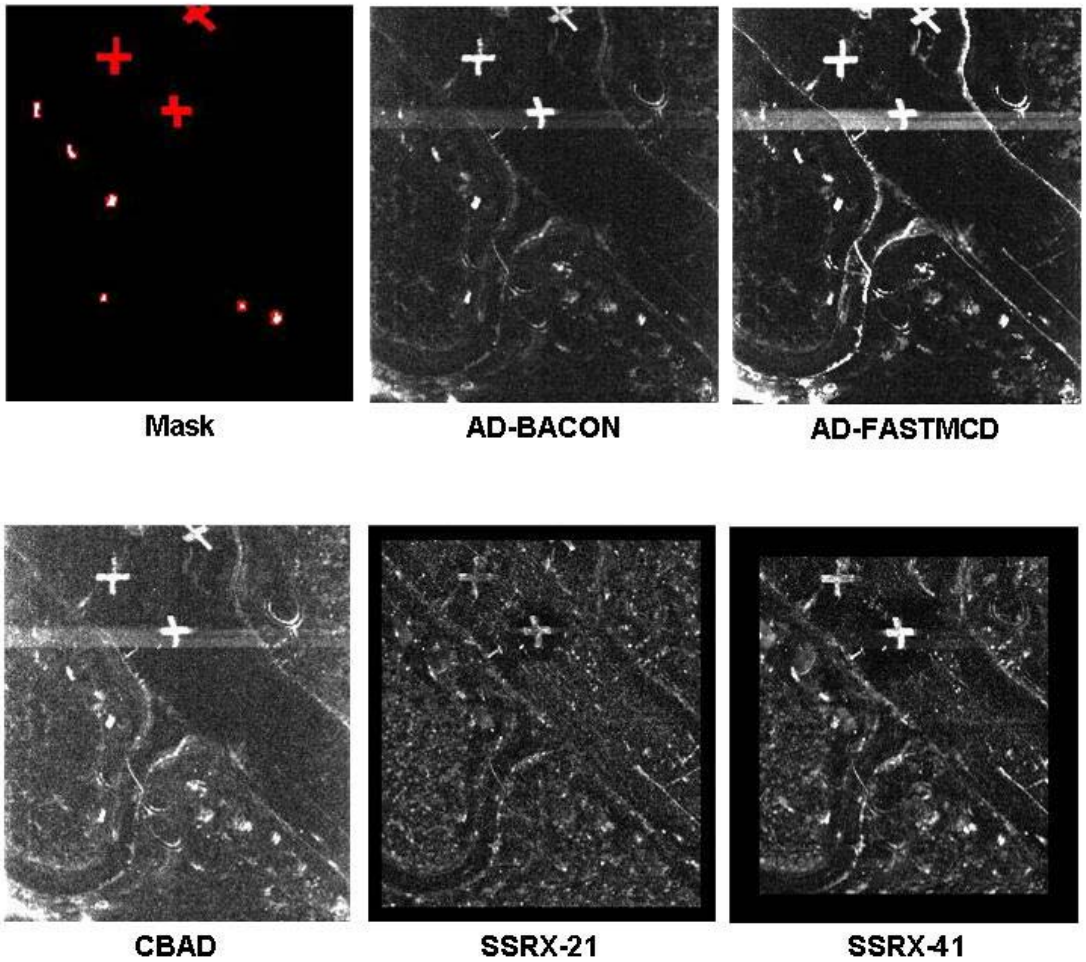


Figure 98. MSD Images for Anomaly Detector Comparisons (Scene 19)

Appendix G: Generator Reflectance Signature Libraries

This appendix contains plots of the reflectance signatures contained in the generator signature libraries used in the AutoMatch detector described in Chapter 6. These signatures are used in Equation (6.5) to generate target image signatures for the target material being detected.

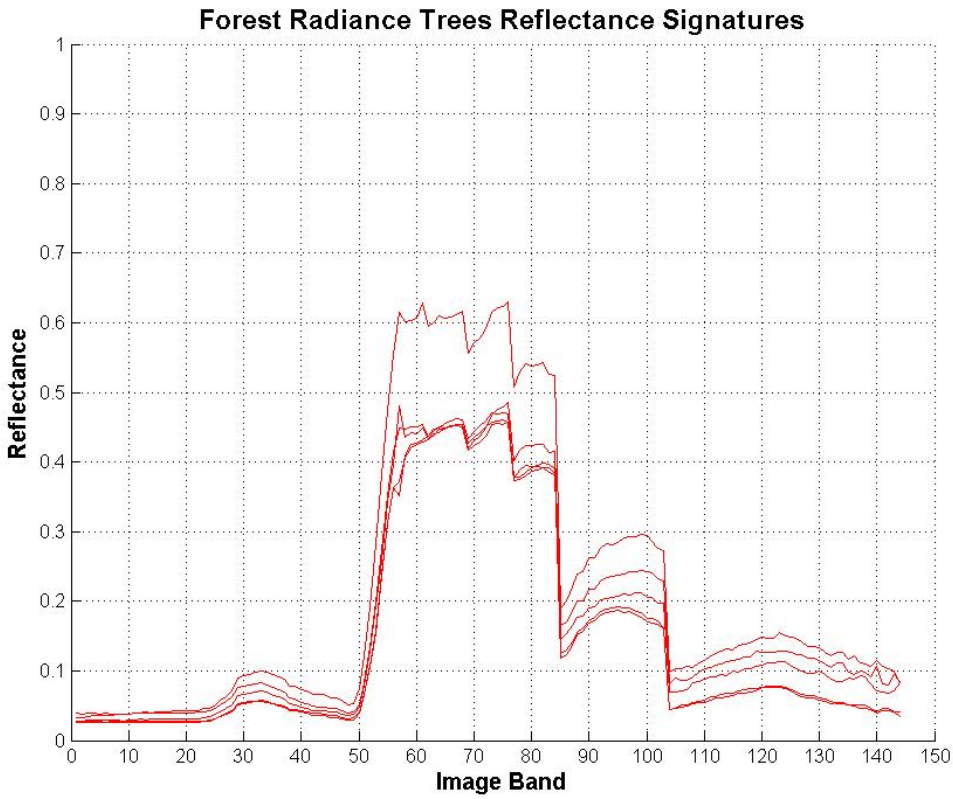


Figure 99. Forest Radiance Tree Library Reflectance Signatures

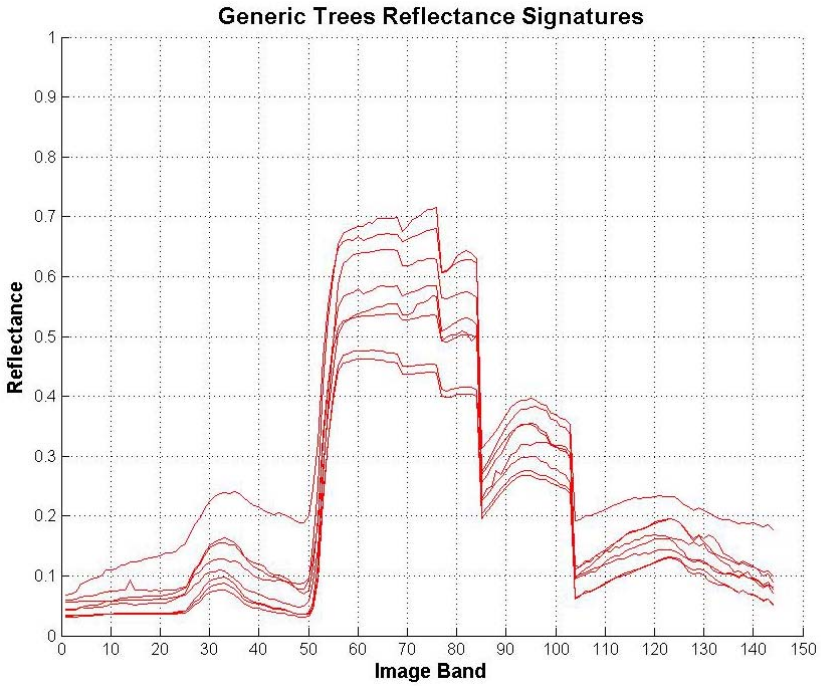


Figure 100. Generic Tree Library Reflectance Signatures

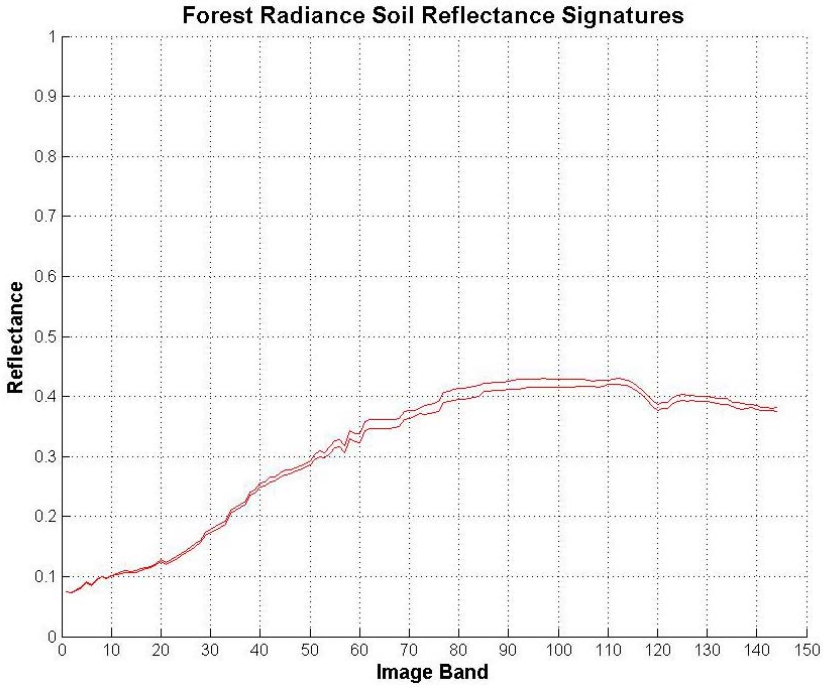


Figure 101. Forest Radiance Soil Reflectance Library Reflectance Signatures

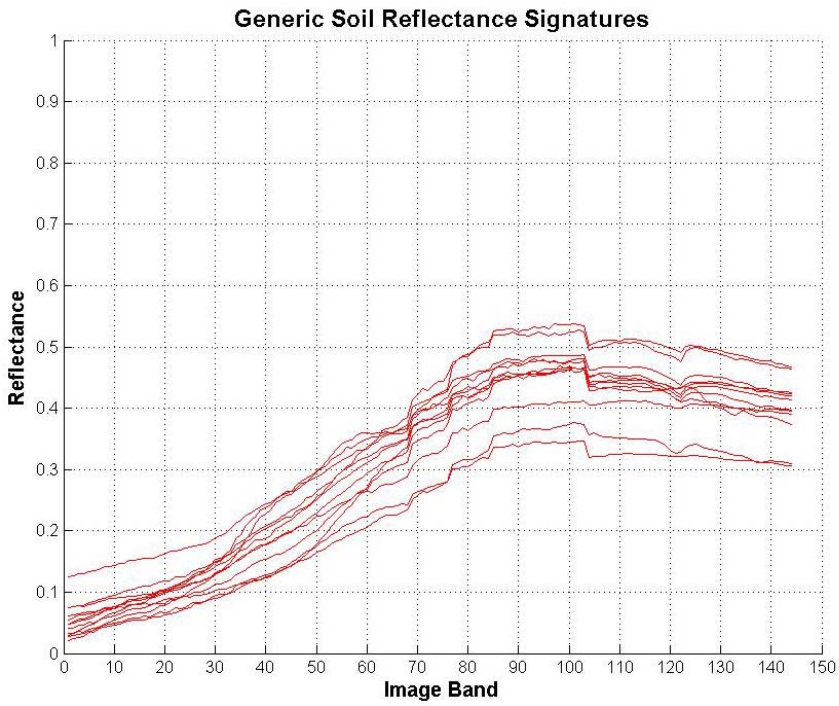


Figure 102. Generic Soil Library Reflectance Signatures

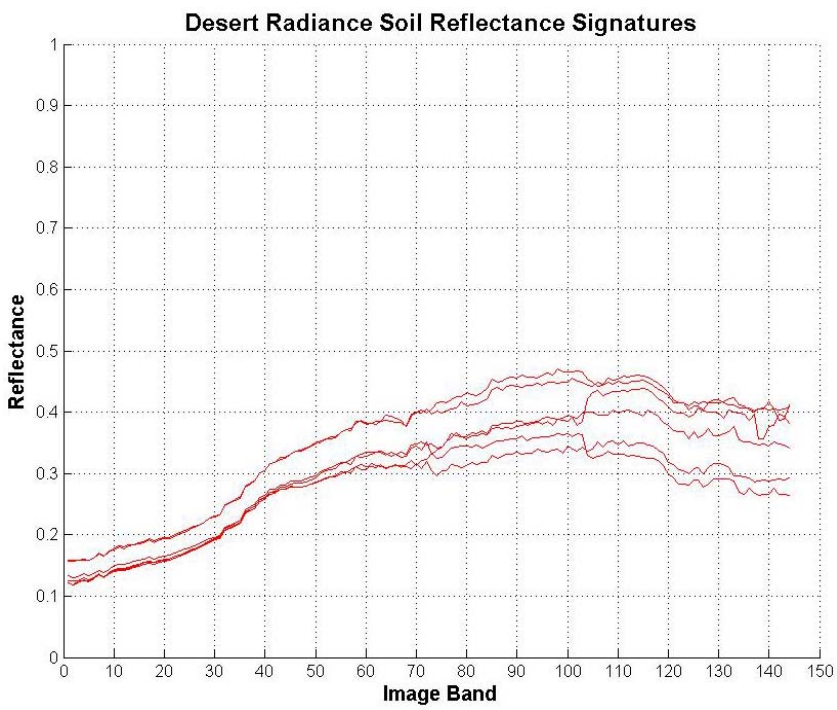


Figure 103. Desert Radiance Soil Library Reflectance Signatures

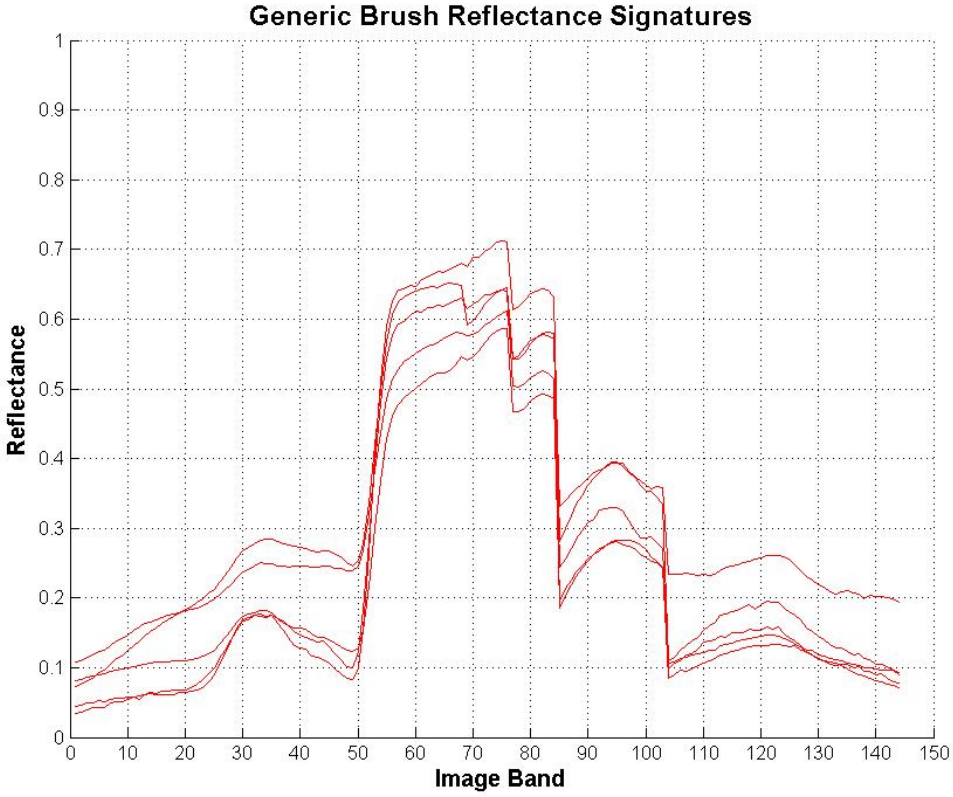


Figure 104. Generic Brush Library Reflectance Signatures

Bibliography

Achard, V., A. Landrevie and J.C. Fort. "Anomalies Detection in Hyperspectral Imagery Using Projection Pursuit Algorithm," *SPIE Conference on Image and Signal Processing for Remote Sensing X*, 5573: 193-202 (2004).

Ammeraal, L. *Programming Principles in Computer Graphics*. New York: Wiley, 1992.

Atkinson, Anthony C. "Stalactite Plots and Robust Estimation for the Detection of Multivariate Outliers," in *New Directions in Statistical Data Analysis and Robustness*, S. Morgenthaler, E. Ronchetti and W.A. Stahel, Eds, Basel: Birkhauser, 1993, pp. 1-8.

---. "Fast Very Robust Methods for the Detection of Multiple Outliers," *Journal of the American Statistical Association*, 89:1329-1339 (December 1994).

Bajorski, Peter and Emmett J. Ientilucci. "Geometric Basis-Vector Selection Methods and Subpixel Target Detection as Applied to Hyperspectral Imagery," *IEEE International Geoscience and Remote Sensing Symposium, 2004 (IGARSS '04)*, 5: 3211-3214 (2004).

Bajorski, Peter, Emmett J. Ientilucci and John R. Schott. "Comparison of Basis-Vector Selection Methods for Target and Background Subspaces as Applied to Subpixel Target Detection," *SPIE Conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X*, 5425: 97-108 (2004).

Barnett, Vic. "The Ordering of Multivariate Data," *Journal of the Royal Statistical Society, Series A*, 138:318-344 (1976).

Barnett, Vic and Toby Lewis. *Outliers in Statistical Data, 3rd Ed.* Chichester, UK: John Wiley & Sons, Inc., 1994.

Beale, E.M.L. "Euclidean Cluster Analysis," *Bulletin of the International Statistical Institute: Proceedings of the 37th Session*, 2: 92-94 (1969).

Bebbington, A.C. "A Method of Bivariate Trimming for Robust Estimation of the Correlation Coefficient," *Applied Statistics*, 27:221-226 (1978).

Becker, Claudia and Ursula Gather. "The Masking Breakdown Point of Multivariate Outlier Identification Rules," *Journal of the American Statistical Association*, 94:947-955 (September 1999).

Beckman, R.J. and R.D. Cook. "Outlier...s," *Technometrics*, 25:119-163 (May 1983).

Bernoulli, Daniel and C.G. Allen. "The Most Probable Choice Between Severel Discrepant Observations and the Formation Therefrom of the Most Likely Induction," *Biometrika*, 48:3-13 (1961).

Billor, Nedret, Ali S. Hadi and Paul F. Velleman. "BACON: Blocked Adaptive Computationally Efficient Outlier Nominators," *Computational Statistics & Data Analysis*, 34:279-298 (2000).

Boardman, J.W., F.A. Kruse and R.O. Green. "Mapping Target Signatures via Partial Unmixing of AVIRIS Data," *Fifth JPL Airborne Earth Science Workshop*, JPL Publication 95-1: 23-26 (1995).

Butler, R.W., P.L. Davies and M. Jhun. "Asymptotics for the Minimum Covariance Determinant Estimator," *The Annals of Statistics*, 21:1385-1400 (1993).

Calinski, R.B. and J. Harabasz. "A Dendrite Method for Cluster Analysis," *Communications in Statistics*, 3:1-27 (1974).

Campbell, N.A. "Robust Procedures in Multivariate Analysis I: Robust Covariance Estimation," *Applied Statistics*, 29:231-237 (1980).

Carlotto, Mark J. "A Cluster-based Approach for Detecting Man-made Objects and Changes in Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, 43:374-387 (February 2005).

Caroni, C. and P. Prescott. "Sequential Application of Wilks's Multivariate Outlier Test," *Applied Statistics*, 41:355-364 (1992).

Catterall, Stephen. "Anomaly Detection Based on the Statistics of Hyperspectral Imagery," *SPIE Conference on Imagery Spectroscopy X*, 5546: 171-178 (2004).

Chang, Chein-I and Shao-Shan Chiang. "Anamoly Detection and Classification for Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, 40:1314-1325 (June 2002).

Chang, Chein-I. *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. New York: Kluwer Academic/Plenum Publishers, 2003.

Chen, Wanhui, Liangyun Liu, Chao Zhang, Jihua Wang, Jindi Wang and Yuchun Pan. "Monitoring the Seasonal Bare Soil Areas in Beijing Using Multi-Temporal TM Images," *Proceedings of the 2004 International Geoscience and Remote Sensing Symposium*, 5: 3379-3382

Chiang, Leo H., Randy J. Pell and Mary Beth Seasholtz. "Exploring Process Data with the Use of Robust Outlier Detection Algorithms," *Journal of Process Control*, 13:437-449 (2003).

Chiang, Shao-Shan, Chein-I Chang and I.W. Ginsberg. "Unsupervised Target Detection in Hyperspectral Images Using Projection Pursuit," *IEEE Transactions on Geoscience and Remote Sensing*, 39:1380-1391 (July 2001).

Clare, Phil, Mark Bernhardt, William Oxford, Sean Murphy, Peter Godfree and Vicky Wilkinson. "A New Approach to Anomaly Detection in Hyperspectral Images," *SPIE Conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*, 5093: 17-28 (2003).

Croux, Christophe and A. Ruiz-Gazen. "A Fast Algorithm for Robust Principal Components Based on Projection Pursuit," *COMPSTAT 96*, 211-216 (1996).

Croux, Christophe and Gentiane Haesbroeck. "Influence Function and Efficiency of the Minimum Covariance Determinant Scatter Matrix Estimator," *Journal of Multivariate Analysis*, 71:161-190 (1999).

Cui, H.J. and Y.B. Ting. "Projected Median of Absolute Deviation and Its Applications," *Journal of Systems Science and Mathematical Science*, 14:63-72 (1994).

David, H.A. *Order Statistics*. New York: Wiley, 1981.

Davies, L. "The Asymptotics of Rousseeuw's Minimum Volume Ellipsoid Estimator," *The Annals of Statistics*, 20:1828-1843 (1992).

Donoho, D. L. "Breakdown Properties of Multivariate Location Estimators," PhD Qualifying Paper, Department of Statistics, Harvard University, Cambridge, MA, 1982.

Donoho, D. L. and P.J. Huber. "The Notion of Breakdown Point," in *A Festschrift for Erich L. Lehmann*, P.J. Bickel, K.A. Doksum and J.L. Hodges, Eds, Belmont, CA: 1983, pp. 157-184.

Duda, Richard O. and Peter E. Hart. *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

Eaton, M.L. "Isotropic Distributions," in *Encyclopedia of Statistical Sciences*, S. Kotz, N.L. Johnson and C.B. Read, Eds, New York: Wiley, 1983, pp. 265-267.

Egan, William J. and Stephen L. Morgan. "Outlier Detection in Multivariate Analytical Chemical Data," *Analytical Chemistry*, 70:2372-2379 (June 1998).

Eismann, Michael T. *Strategies for Hyperspectral Target Detection in Complex Background Environments*. Draft Manuscript, Air Force Research Laboratory, Wright-Patterson AFB, OH,

Everitt, Brian S., Sabine Landau and Morven Leese. *Cluster Analysis, 4th Ed.* London: Arnold, 2001.

Fang, Kai-Tai and Y. Wang. *Number Theoretic Methods in Statistics*. London: Chapman and Hall, 1994.

Farrell, Michael D. and Russell M. Mersereau. "On the Impact of Covariance Contamination for Adaptive Detection in Hyperspectral Imaging," *IEEE Signal Processing Letters*, 12:649-652 (September).

Friedman, J.H. and J.W. Tukey. "A Projection Pursuit Algorithm for Exploratory Data Analysis," *IEEE Transactions on Computers*, C-23:881-889 (1974).

Gao, Shaogen, Guoying Li and Dongqian Wang. "A New Approach for Detecting Multivariate Outliers," *Communications in Statistics--Theory and Methods*, 34:1857-1865 (2005).

Gasko, M. and D. L. Donoho. "Influential Observation in Data Analysis," *American Statistical Association Proceedings of the Business and Economic Statistics Section*, 1: 104-109 (1982).

Gaucel, J.-M., M. Guillaume and S. Bourennane. "Whitening Spacial Correlation Filtering for Hyperspectral Anomaly Detection," *2005 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, 1: 333-336 (2005).

Gnanadesikan, R. and J.R. Kettenring. "Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data," *Biometrics*, 28:81-124 (March 1972).

Gonzalez, Rafael C., Richard E. Woods and Steven L. Eddins. *Digital Image Processing Using Matlab*. Upper Saddle River, NJ: Pearson Prentice Hall, 2004.

Goovaerts, Pierre. "Factorial Kriging Analysis: A Useful Tool for Exploring the Structure of Multivariate Spatial Soil Information," *Journal of Soil Science*, 43:597-619 (1992).

Goovaerts, Pierre, Geoffrey Jacquez, Amanda Warner, Bob Crabtree and Andrew Marcus. "Detection of Local Anomalies in High Resolution Hyperspectral Imagery Using Geostatistical Filtering and Local Spatial Statistics," *2003 IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, 1: 385-394 (2004).

Grossman, John M., Jeffrey Bowles, Daniel Haas, John A. Antoniades, Mitchell R. Grunes, Peter Palmadesso, David Gillis, Kwok Y. Tsang, Mark Baumbeck, Mark Daniel, John Fisher and Ioana Triandaf. "Hyperspectral Analysis and Target Detection System for the Adaptive Spectral Reconnaissance Program," *SPIE*

Conference on Algorithms for Multispectral and Hyperspectral Imagery, IV, 3372: 2-13 (April 1998).

Grubel, R. and David M. Rocke. "On the Cumulants of Affine-Equivariant Estimators in Elliptical Families," *Journal of Multivariate Analysis*, 35:203-222 (1990).

Hadi, Ali S. "Identifying Multiple Outliers in Multivariate Data," *Journal of the Royal Statistical Society, Series B*, 54:761-771 (1992).

---. "A Modification of a Method for the Detection of Outliers in Multivariate Samples," *Journal of the Royal Statistical Society, Series B*, 56:393-396 (1994).

Hampel, F.R. "Contributions to the Theory of Robust Estimation," PhD Thesis, University of California, Berkeley, Berkeley, CA, 1968.

---. "A Generalized Qualitative Definition of Robustness," *Annals of Mathematical Statistics*, 42:1887-1896 (1971).

Hampel, F.R., E. Ronchetti, Peter J. Rousseeuw and W.A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley, 1986.

Hardin, Johanna and David M. Rocke. "Outlier Detection in the Multiple Cluster Setting Using the Minimum Covariance Determinant Estimator," *Computational Statistics & Data Analysis*, 44:625-638 (2004).

---. "The Distribution of Robust Distances," *Journal of Computational and Graphical Statistics*, 14:928-946 (2005).

Hawkins, Douglas M. "The Feasible Solution Algorithm for the Minimum Covariance Determinant Estimator in Multivariate Data," *Computational Statistics & Data Analysis*, 17:197-210 (1994).

Hawkins, Douglas M. and David J. Olive. "Improved Feasible Solution Algorithms for High Breakdown Estimation," *Computational Statistics & Data Analysis*, 30:1-11 (1999).

Hazel, Geoffrey. "Multivariate Gaussian MRF for Multispectral Scene Segmentation and Anomaly Detection," *IEEE Transactions on Geoscience and Remote Sensing*, 38:1199-1211 (May 2000).

Healey, Glenn and David Slater. "Models and Methods for Automated Material Identification in Hyperspectral Imagery Acquired Under Unknown Illumination and Atmospheric Conditions," *IEEE Transactions on Geoscience and Remote Sensing*, 37:2706-2717 (1999).

Helbling, J.M. "Ellipsoides Minimaux de Couverture en Statistique Multivariee," PhD Thesis, Ecole Polytechnique Federale de Lausanne, Switzerland, 1983.

Helge, H., Y. Liang and O.M. Kvalheim. "Trimmed Object Projections: A Nonparametric Robust Latent-Structure Decomposition Method," *Chemometrics and Intelligent Laboratory Systems*, 27:33-40 (1995).

Hodges, J.L. "Efficiency in Normal Samples and Tolerance of Extreme Values for Some Estimates of Location," *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, (1967).

Hoffbeck, Joseph P. and David A. Landgrebe. "Covariance Matrix Estimation and Classification with Limited Training Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:763-767 (1996).

Hsueh, Mingkai and Chein-I Chang. "Adaptive Causal Anomaly Detection for Hyperspectral Imagery," *2004 IEEE International Geoscience and Remote Sensing Symposium (IGARSS '04)*, 5: 3222-3224 (2004).

Huber, P.J. *Robust Statistics*. New York: Wiley, 1989.

Hubert, M., Peter J. Rousseeuw and S. Verboven. "A Fast Method for Robust Principal Components with Applications to Chemometrics," *Chemometrics and Intelligent Laboratory Systems*, 60:101-111 (2002).

Ientilucci, Emmett J. and Peter Bajorski. "Statistical Models for Physically Derived Target Sub-spaces," *SPIE Proceedings on Imaging Spectrometry XI*, 6302: (September 2006).

Jackson, J.E. and G.S. Mudholkar. "Control Procedures for Residuals Associated with Principal Component Analysis," *Technometrics*, 21:341-349 (1979).

Jackson, Qiong and David A. Landgrebe. "An Adaptive Method for Combined Covariance Estimation and Classification," *IEEE Transactions on Geoscience and Remote Sensing*, 40:182-1087 (2002).

Jimenez, Luis O. and David A. Landgrebe. "Supervised Classification in High-Dimensional Space: Geometrical, Statistical, and Asymptotical Properties of Multivariate Data," *IEEE Transaction on Systems, Man, and Cybernetics, Part C*, 28:39-54 (1998).

---. "Hyperspectral Data Analysis and Supervised Feature Reduction via Projection Pursuit," *IEEE Transactions on Geoscience and Remote Sensing*, 37:2653-2667 (1999).

Juan, Jesus and Francisco J. Prieto. "Using Angles to Identify Concentrated Multivariate Outliers," *Journal of the American Statistical Association*, 43:311-322 (August 2001).

Kaufman, L. and Peter J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley-Interscience, 1990.

Kerekes, John P. and Dimitris Manolakis. "Improved Modeling of Background Distributions in an End-to-End Spectral Imaging System Model," *Proceedings of the 2004 IEEE International Geoscience and Remote Sensing Symposium*, 2: 972-975 (2004).

Kim, Myung Geun. "Multivariate Outliers and Decompositions of Mahalanobis Distance," *Communications in Statistics--Theory and Methods*, 29:1511-1526 (2000).

Kosinski, Andrzej S. "A Procedure for the Detection of Multivariate Outliers," *Computational Statistics & Data Analysis*, 29:145-161 (1999).

Kwon, Heesung, S.Z. Der and Nasser M. Nasrabadi. "Adaptive Anomaly Detection Using Subspace Separation for Hyperspectral Imagery," *Optical Engineering*, 42:3342-3351 (November 2003).

Kwon, Heesung and Nasser M. Nasrabadi. "Kernel RX-Algorithm: A Nonlinear Anomaly Detector for Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, 43:388-397 (February 2005).

Landgrebe, David A. "Hyperspectral Image Data Analysis," *IEEE Signal Processing Magazine*, 19:17-28

---. *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2003.

Lee, Kyungsuk. "A Subpixel Scale Target Detection Algorithm for Hyperspectral Imagery," PhD Dissertation, Rochester Institute of Technology, Rochester, NY, 2003.

Li, Guoying and Z. Chen. "Unknown," *Journal of the American Statistical Association*, 80:759-766 (1985).

Liu, Weimin and Chein-I Chang. "A Nested Spatial Window-Based Approach to Target Detection for Hyperspectral Imagery," *2004 IEEE International Geoscience and Remote Sensing Symposium (IGARSS '04)*, 1: 266-268 (2004).

Liu, Yong and Glenn Healey. "Using Nonparametric Distribution Estimates for Subpixel Detection of 3D Objects," *SPIE Conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X*, 5425: 91-96 (2004).

Lopuhaa, Hendrik P. "On the Relation Between S-Estimators and M-Estimators of Multivariate Location and Covariance," *The Annals of Statistics*, 17:1662-1683 (December 1989).

Lopuhaa, Hendrik P. and Peter J. Rousseeuw. "Breakdown Points of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices," *The Annals of Statistics*, 19:229-248 (March 1991).

Lopuhaa, Hendrik P. "Highly Efficient Estimators of Multivariate Location with High Breakdown Point," *The Annals of Statistics*, 20:398-413 (March 1992).

MacGregor, J.F. and T. Kourti. "Statistical Process Control of Multivariate Process," *Control Engineering Practice*, 3:403-414 (1995).

Manolakis, D., C. Siracusa and G. Shaw. "Hyperspectral Subpixel Target Detection Using the Linear Mixing Model," *IEEE Transactions on Geoscience and Remote Sensing*, 39:1392-1409 (July 2001).

Manolakis, D. and D. Marden. "Non Gaussian Models for Hyperspectral Algorithm Design and Assessment," *IEEE International Geoscience and Remote Sensing Symposium, 2002 (IGARSS '02)*, 1: 1664-1666 (June 2002).

Manolakis, D., M. Rossacci, J. Cipar, R. Lockwood, T. Cooley and J. Jacobson. "Statistical Characterization of Natural Hyperspectral Backgrounds Using t-Elliptically Contoured Distributions," *SPIE Conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, 5806: 56-65 (April 2005).

Maronna, Ricardo A. "Robust M-Estimators of Multivariate Location and Scatter," *The Annals of Statistics*, 4:51-67 (January 1976).

Maronna, Ricardo A. and Victor J. Yohai. "The Behavior of the Stahel-Donoho Robust Multivariate Estimator," *Journal of the American Statistical Association*, 90:330-341 (March 1995).

Marriott, F.H.C. "Optimization Methods of Cluster Analysis," *Biometrika*, 69:417-421 (1982).

Meidunas, Eduardo. "Robust Estimation of Mahalanobis Distances in Hyperspectral Images," PhD Dissertation, Department of Electrical and Computer Engineering, Air Force Institute of Technology, Wright-Patterson AFB, December 2006.

Moon, T.K. "The Expectation-Maximization Algorithm," *IEEE Signal Processing Magazine*, 31:47-60 (May 1993).

Myers, Raymond H. and Douglas C. Montgomery. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. New York: John Wiley &Sons, Inc., 1995.

Neville, R.A., K. Staenz, T. Szeredi, J. Lefebvre and P. Hauff. "Automatic Endmember Extraction from Hyperspectral Data for Mineral Exploration," *Fourth International Airborne Remote Sensing Conference/21st Canadian Symposium on Remote Sensing*, 1: 891-896 (June 1999).

Oigard, T.A. and A. Hanssen. "The Multivariate Normal Inverse Gaussian Heavy-Tailed Distribution: Simulation and Estimation," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2: 1489-1492 (2002).

Pan, Jian-Xin, Wing-Kam Fung and Kai-Tai Fang. "Multiple Outlier Detection in Multivariate Data Using Projection Pursuit Techniques," *Journal of Statistical Planning and Inference*, 83:153-167 (2000).

Pan, Zhihong, Glenn Healey and David Slater. "Modeling the Spectral Variability of Ground Irradiance Functions," *SPIE Conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, 4049: 82-93 (2000).

Rao, C.R. "The Use and Interpretation of Principle Component Analysis in Applied Research," *Sankhya*, 26:329-358 (1964).

Reed, Irving S. and Xiaoli Yu. "Adaptive Multiple-Band CFAR Detection of an Optical Pattern with Unknown Spectral Distribution," *IEEE Proceedings on Acoustics, Speech, and Signal Processing*, 38:1760-1770 (October 1990).

Ren, Hsuan and Chein-I Chang. "Target-Constrained Interference-Minimized Approach to Subpixel Target Detection for Hyperspectral Images," *Optical Engineering*, 39:3138-3145 (December 2000).

Richards, John A. and Xiuping Jia. *Remote Sensing Digital Image Analysis: An Introduction, 3rd Ed.* Berlin: Springer-Verlag, 1999.

Riley, Ronald, Rob K. Newsom and Aaron K Andrews. "Anomaly Detection in Noisy Hyperspectral Imagery," *SPIE Conference on Imaging Spectrometry X*, 5546: 159-170 (2004).

Rocke, David M. and David L. Woodruff. "Computation of Robust Estimates of Multivariate Location and Shape," *Statistica Neerlandica*, 47:27-42 (1993).

Rocke, David M. "Robustness Properties of S-Estimators of Multivariate Location and Shape in High Dimension," *The Annals of Statistics*, 24:1327-1345 (1996).

Rocke, David M. and David L. Woodruff. "Identification of Outliers in Multivariate Data," *Journal of the American Statistical Association*, 91:1047-1061 (September 1996).

Rosario, Dalton S. "Highly Effective Logistic Regression Model for Signal (Anomaly) Detection," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, 1: 817-820 (2004).

Rouse, J.W., R.H. Haas, J.A. Schell and D.W. Deering. "Monitoring Vegetation Systems in the Great Plains with Third ERTS," *ERTS Symposium*, NASA No. SP-351: 309-317

Rousseeuw, Peter J. "Multivariate Estimation with High Breakdown Point," *Fourth Pannonian Symposium on Mathematical Statistics and Probability*, (September 4, 1983).

Rousseeuw, Peter J. and Annick M. Leroy. *Robust Regression and Outlier Detection*. New York: John Wiley & Sons, Inc., 1987.

Rousseeuw, Peter J. and Bert C. van Zomeren. "Unmasking Multivariate Outliers and Leverage Points," *Journal of the American Statistical Association*, 85:633-639 (September 1990).

Rousseeuw, Peter J. and C. Croux. "Alternatives to the Median Absolute Deviation," *Journal of the American Statistical Association*, 88:1273-1283 (1993).

Rousseeuw, Peter J. and Katrien van Driessen. "A Fast Algorithm for the Minimum Covariance Determinant Estimator," *Technometrics*, 41:212-223 (August 1999).

Schaum, Alan P. and Alan D. Stocker. "The Stochastic Mixing Model," *1997 International Symposium on Spectral Sensing Research*, (14-17 December 1997).

Schaum, Alan P. "Joint Subspace Detection of Hyperspectral Targets," *Proceedings of the 2004 IEEE Aerospace Conference*, 3: 1818-1824 (6-13 March 2004).

---. "A Remedy for Nonstationarity in Background Transition Regions for Real Time Hyperspectral Detection," *Proceedings of the 2006 IEEE Aerospace Conference*, (4-11 March 2006).

Schott, John R. *Remote Sensing: The Image Chain Approach*. New York: Oxford University Press, 1997.

Schott, John R., Kyungsuk Lee, Rolando Raqueno and Gary Hoffman. "Use of Physics Based Models in Hyperspectral Imagery," *Proceedings of the 31st Applied Imagery Pattern Recognition Workshop (AIPR '02)*, 1: 36-42 (October 2002).

Schweizer, Susan M. and Jose M.F. Moura. "Hyperspectral Imagery: Clutter Adaptation in Anomaly Detection," *IEEE Transactions on Information Theory*, 46:1855-1871 (August 2000).

---. "Efficient Detection in Hyperspectral Imagery," *IEEE Transactions on Image Processing*, 10:584-597 (April 2001).

Shi, Miaohong and Glenn Healey. "Using Multiband Correlation Models for the Invariant Recognition of 3-D Hyperspectral Textures," *IEEE Transactions on Geoscience and Remote Sensing*, 43:1201-1209 (May 2005).

Slater, David and Glenn Healey. "Material Classification for 3D Objects in Aerial Hyperspectral Images," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999*, 2: 268-273 (June 1999).

---. "Physics-based Model Acquisition and Identification in Airborne Spectral Images," *Proceedings of the Eighth IEEE International Conference on Computer Vision (ICCV), 2001*, 2: 257-262 (July 2001).

Smetek, Timothy E. and Kenneth W. Bauer. "A Comparison of Multivariate Outlier Detection Methods for Finding Hyperspectral Anomalies," *Final Report of the 75th MORSS*, (June 2006).

---. "Finding Hyperspectral Anomalies Using Multivariate Outlier Detection," *Proceedings of the 2007 IEEE Aerospace Conference*, (March 2007).

Stahel, W.A. "Robuste Schatzungen: Infinitesimale Optimalitat und Schatzungen von Kovarianzmatrizen," PhD Thesis, ETH Zurich, Zurich, Switzerland, 1981.

Stein, David W.J., Scott G. Beaven, Lawrence E. Hoff, Edwin M. Winter, Alan P. Schaum and Alan D. Stocker. "Anomaly Detection for Hyperspectral Imagery," *IEEE Signal Processing Magazine*, 19:58-69 (January 2002).

Stevenson, Brian, Rory O'Connor, William Kendall, Alan D. Stocker, William Schaff, Rick Holasek, Detlev Even, Drew Alexa, John Salvador, Michael T. Eismann, Robert Mack, Pat Kee, Steve Harris, Barry Karch and John Kershenstein. "The Civil Air Patrol ARCHER Hyperspectral Sensor System," *SPIE Proceedings on Airborne ISR Systems and Applications II*, 5787: 17-28 (May 2005).

Suen, Pei-hsiu and Glenn Healey. "Invariant Mixture Recognition in Hyperspectral Images," *Eighth IEEE International Conference on Computer Vision*, 1: 262-267 (July 2001).

Suen, Pei-hsiu, Glenn Healey and David Slater. "The Impact of Viewing Geometry on Matrial Discriminability in Hyperspectral Images," *IEEE Transactions on Geoscience and Remote Sensing*, 39:1352-1359 (July 2001).

Tadjudin, Saldju and David A. Landgrebe. "Covariance Estimation with Limited Training Samples," *IEEE Transactions on Geoscience and Remote Sensing*, 37:2113-2118 (1999).

Thai, Bea and Glenn Healey. "Using a Linear Subspace Approach for Invariant Subpixel Material Identification in Airborne Hyperspectral Imaging," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999*, 1: 567-572 (June 1999).

---. "Invariant Subpixel Material Detection in Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, 40:599-608 (March 2002).

Titterington, D.M. "Estimation of Correlation Coefficients by Ellipsoidal Trimming," *Applied Statistics*, 27:227-234 (1978).

Tyler, David E. "Some Results on the Existence, Uniqueness, and Computation of the M-Estimates of Multivariate Location and Scatter," *SIAM Journal on Scientific and Statistical Computing*, 9:354-362 (March 1988).

Viljoen, H. and J.H. Venter. "Identifying Multivariate Discordant Observations: A Computer-Intensive Approach," *Computational Statistics & Data Analysis*, 40:159-172 (2002).

Walczak, B. and D.L. Massart. "Robust Principle Component Regression as a Detection Tool for Outliers," *Chemometrics and Intelligent Laboratory Systems*, 27:41-54 (1995).

West, Jason E., David W. Messinger, Emmett J. Ientilucci, John P. Kerekes and John R. Schott. "Matched Filter Stochastic Background Characterization for Hyperspectral Target Detection," *SPIE Conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, 5806: 1-12 (28 March 2005).

Wilk, M.B. and R. Gnanadesikan. "Graphical Methods for Internal Comparisons in Multiresponse Experiments," *Annals of Mathematical Statistics*, 35:613-631

Wilks, S.S. "Multivariate Statistical Outliers," *Sankhya*, 25:407-426 (1963).

Winter, Edwin M. "Detection of Surface Mines Using Hyperspectral Sensors," *2004 IEEE International Geoscience and Remote Sensing Symposium (IGARSS '04)*, 3: 1597-1600 (2004).

Woodruff, David L. and David M. Rocke. "Heuristic Search Algorithms for the Minimum Volume Ellipsoid," *Journal of Computational and Graphical Statistics*, 2:69-95 (1993).

---. "Computable Robust Estimation of Multivariate Location and Shape in High Dimension Using Compound Estimators," *Journal of the American Statistical Association*, 89:888-896 (September 1994).

Woodruff, David L. and Torsten Reiners. "Experiments With, and On, Algorithms for Maximum Likelihood Clustering," *Computational Statistics & Data Analysis*, 47:237-253 (2004).

Zani, Sergio, Marco Riani and Aldo Corbellini. "Robust Bivariate Boxplots and Multiple Outlier Detection," *Computational Statistics & Data Analysis*, 28:257-270 (1998).

Zhang, Ye and Yanfeng Gu. "Kernel-Based Invariant Subspace Method for Hyperspectral Target Detection," *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004 (ICASSP '04)*, 5: 801-804 (May 2004).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
06-14-2007	**Doctoral Dissertation**	Jun 2004 – Jun 2007

4. TITLE AND SUBTITLE

Hyperspectral Imagery Target Detection Using Improved Anomaly Detection and Signature Matching Methods

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Smetek, Timothy E., Major, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Street, Building 642
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/DS/ENS/07-07

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

N/A

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This research extends the field of hyperspectral target detection by developing autonomous anomaly detection and signature matching methodologies that reduce false alarms relative to existing benchmark detectors. The proposed anomaly detection methodology adapts multivariate outlier detection algorithms for use with hyperspectral datasets containing thousands of high-dimensional spectral signatures. In so doing, the limitations of existing, non-robust anomaly detectors are identified, an autonomous clustering methodology is developed to divide an image into homogeneous background materials, and competing multivariate outlier detection methods are evaluated. To arrive at a final detection algorithm, robust parameter design methods are employed to determine parameter settings that achieve good detection performance over a range of hyperspectral images and targets. The final anomaly detection algorithm is tested against existing local and global anomaly detectors, and is shown to achieve superior detection accuracy when applied to a diverse set of hyperspectral images.

The proposed signature matching methodology employs image-based atmospheric correction techniques in an automated process to transform a target reflectance signature library into a set of image signatures. This set of signatures is combined with an existing linear filter to form a target detector that is shown to perform as well or better relative to detectors that rely on complicated, information-intensive atmospheric correction schemes. The performance of the proposed methodology is assessed using a range of target materials in both woodland and desert hyperspectral scenes.

15. SUBJECT TERMS

Hyperspectral Imagery, Spectrum Analysis, Target Detection, Hyperspectral Anomaly Detection, Hyperspectral Signature Matching, Multivariate Outlier Detection, Robust Parameter Design, Atmospheric Correction, Multispectral

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Kenneth W. Bauer, Jr., Professor (ENS)
U	U	U	UU	388	19b. TELEPHONE NUMBER *(Include area code)* (937) 255-6565, ext 4328; e-mail: Kenneth.Bauer@afit.edu

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18