

Analysis of Modeling, Training, and Dimension Reduction Approaches for Target Detection in Hyperspectral Imagery

A Dissertation
Presented to
The Academic Faculty

by

Michael D. Farrell Jr.

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Electrical and Computer Engineering
Georgia Institute of Technology
December 2005

Analysis of Modeling, Training, and Dimension Reduction Approaches for Target Detection in Hyperspectral Imagery

Approved by:

Professor Russell M. Mersereau, Advisor
School of ECE
Georgia Institute of Technology

Professor Aaron D. Lanterman
School of ECE
Georgia Institute of Technology

Professor John A. Copeland
School of ECE
Georgia Institute of Technology

Professor Derek T. Cunnold
School of EAS
Georgia Institute of Technology

Professor James H. McClellan
School of ECE
Georgia Institute of Technology

Date Approved: 1 November 2005

*To my parents: only with their absolute, unqualified love and support have
I achieved anything worthwhile.*

*And to my sister: she may never know how much she means to me and how
much I look up to her.*

ACKNOWLEDGEMENTS

I owe a debt of gratitude to my advisor for his guidance and trust. Professor Russ Mersereau is as true a gentleman and scholar as I will ever meet. He gave me the freedom to pursue my research, degree, and career plans while consistently sharing his sage advice.

The other members of my thesis committee also became friends and mentors during my time at Georgia Tech. Professors Jim McClellan and John Copeland are consummate professionals, and embody everything that makes Tech's School of ECE both spectacularly brilliant and well-grounded. Just being around the likes of them and Ron Schafer made me a better person; never before have I met four individuals (including Russ) who "walk softly and carry a big stick", intellectually speaking. Aaron Lanterman brought his high energy level to CSIP about the same time I arrived on the scene. He is a young professor who is leaving his mark both in and out of the classroom. Thanks, too, to Professor Derek Cunnold for serving on my committee. Professor Bruno Frazier deserves special recognition. He was the first person I knew on campus and gave me a warm introduction to the Institute as well as Atlanta in general. Funding support from the Army Research Office is also greatly appreciated.

There were a handful of others outside of Georgia Tech who had a significant impact on my research and career development. Drs. Gary Shaw, Randy Avent, and Dieter Willner of MIT Lincoln Laboratory are at the top of this list. They, along with others at MIT/LL such as Eric Crosson, gave me an opportunity to develop my expertise and grow professionally in one of the world's finest research organizations. Thanks also to friends in the Publications Group at MIT/LL, for their permission to use figures 2, 17, 18, and 20.

The staff at CSIP have been nothing short of courteous and helpful, and I would like to thank each of them for their hard work – especially Christy Ellis, Keith May, and Kay

Gilstrap.

I assert that grad school is much more like a job than most people realize; it's really not like undergraduate study at all. However, I was fortunate to meet and develop friendships with some truly talented and generous people at Tech. There are too many to name, yet I list a few here knowing that I will undoubtedly forget someone important: Ajay Kumar, Ghas-san Al-Regib, Greg Slabaugh, Jau-Horng Chen, Jinwoo Kang, Lisa Ehrman, Martin Tobias, Matt Lee, Nikko Papageorgiou, Paul Hong, Raviv Reich, Soumendu Bhattacharyya, Volkan Cevher, and Woojay Jeon.

There were a number of wonderful people that supported me from outside Atlanta, too. Heather selflessly gave me support and inspiration, and was a good sport about so many nights and weekends I spent doing research rather than us doing something more "fun." She has been the very definition of best friend and companion. I have to believe that Chris, Ben, Phil, Hasby, Jason, Beth, Melissa, and the rest of the gang will get a laugh every time they remember that I'm now "Dr. Farrell." The fact that I got a Ph.D is a testament to the unexpected nature of opportunity, and a good reason to believe that the best laid plans of mice and men often *do* go awry . . . and it can be a very good thing.

First, last, and anything but least in my life, has been faith. The Catholic Center at Georgia Tech is one of the Institute's hidden treasures, and God is truly at work in Father Mario. With God's continued blessings, and a little bit of good 'ole Irish luck, I'll be able to apply the knowledge and experience gained during this doctoral endeavor to do His will.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	ix
LIST OF FIGURES	x
SUMMARY	xiii
1 INTRODUCTION	1
1.1 Research Overview	1
2 BACKGROUND	4
2.1 Review of Hyperspectral Imaging	4
2.2 Framework for Adaptive Detection	6
2.3 Hyperspectral vs. Radar	8
3 STATISTICAL DATA MODELING AND PARAMETER ESTIMATION	10
3.1 Measuring Goodness-of-fit for Statistical Data Models	10
3.1.1 Kolmogorov-Smirnov Test	11
3.1.2 Chi-Square Test	12
3.1.3 Exceedance Metric	12
3.1.4 Other Variants and Modifications	13
3.1.5 Results for Normal	14
3.1.6 Results for Elliptically Contoured t	16
3.2 Finite Mixtures	17
3.3 Parameter Estimation for EC- t Mixture Models	18
3.3.1 Expectation-Maximization	19
3.3.2 Stochastic Expectation-Maximization	23
3.4 Finite Mixture Modeling Results for EM and SEM	26

4	A CONCISE DEVELOPMENT OF HSI DETECTION ALGORITHMS . .	36
4.1	Detection Algorithm Design	36
4.1.1	A Note on ROC Curves for HSI	37
4.2	Covariance-based Detectors	38
4.3	Subspace-based Detectors	43
4.4	Anomaly Detection	45
5	EVALUATION OF ADAPTIVE DETECTORS FOR GROUND TARGETS	47
5.1	Introduction to Land Mines	47
5.2	Mine Detection Using Hyperspectral Imaging	48
5.3	Experiments for HSI Mine Detection	49
5.4	Performance Comparison	51
5.5	Theoretical Performance	56
6	DETECTION AND IDENTIFICATION OF AIRBORNE GASES USING HSI	59
6.1	Collection of LWIR Hyperspectral Data	60
6.2	A Note on Signatures for Gas Plumes	61
6.3	Signature-based Gas Detection	62
6.4	HSI Gas Exploitation Without Signatures	64
6.4.1	Principal Components Analysis	64
6.4.2	Visualization Using 3-Band Composites and Scatter Plots	70
6.5	Comments on Airborne Gas Targets	71
7	EFFECTS OF COVARIANCE CONTAMINATION IN ADAPTIVE DETECTION	74
7.1	The Problem of Contaminated Covariance Estimates	74
7.2	Performance Comparison	74
7.3	Detector Robustness to Covariance Contamination	78
8	IMPACT OF DIMENSION REDUCTION ON DETECTOR OUTPUT . . .	82
8.1	Brief Survey of Techniques for HSI Data Reduction	83
8.1.1	PC-based Transforms for Dimension Reduction	84

8.1.2	A Note on Data Reduction vs. Dimension Reduction	84
8.2	Performance Comparison	85
8.3	Detector Robustness to Data Reduction	94
9	CONCLUSIONS	96
9.1	Contributions	98
9.2	Future Work	99
9.3	Concluding Remarks	101
	REFERENCES	102
	VITA	108

LIST OF TABLES

Table 1	Key differences between radar and hyperspectral for adaptive detection.	9
Table 2	Selected analysis of EM results for Fort Hood data.	32
Table 3	Selected analysis of SEM results for Fort Hood data.	32
Table 4	Comparison of EM and SEM results for Fort Hood data.	32
Table 5	Target info for AHI scene 1 with 145445 background pixels.	77
Table 6	Target info for AHI scene 2 with 128250 background pixels.	77
Table 7	Target info for AHI scene 3 with 161222 background pixels.	77
Table 8	Target info for AHI scene 4 with 116790 background pixels.	78
Table 9	Number of target and background pixels for dimension reduction tests.	86
Table 10	SNR and spectral angle between targets and mean background for dimension reduction tests.	86

LIST OF FIGURES

Figure 1	A breakdown of common spectral data exploitation tasks and their regions of support, from visible to LWIR.	2
Figure 2	Concept of hyperspectral imaging illustrated for a pushbroom sensor. . .	5
Figure 3	Exceedance metric goodness-of-fit test for the Normal distribution. . . .	15
Figure 4	Exceedance metric goodness-of-fit test for the EC- t distribution.	15
Figure 5	RGB image of Camp Pendleton, scene 7.	26
Figure 6	Exceedance metric for EC- t mixture model using EM technique.	27
Figure 7	Cluster image generated from EC- t mixture model using EM technique. .	27
Figure 8	Exceedance metric for EC- t mixture model using SEM technique.	28
Figure 9	Cluster image generated from EC- t mixture model using SEM technique.	28
Figure 10	RGB image of Fort Hood, scene 1.	29
Figure 11	RGB image of Fort Hood, scene 8.	29
Figure 12	K-S test for Fort Hood 1 using EM: $L = 15$, $K = 11$	30
Figure 13	K-S test for Fort Hood 8 using EM: $L = 15$, $K = 14$	31
Figure 14	SEM cluster images for Fort Hood 1 (a), $L = 15$, $K_{init} = 11$, $K_{end} = 4$, and Fort Hood 8 (b), $L = 15$, $K_{init} = 14$, $K_{end} = 13$	33
Figure 15	K-S test for Fort Hood 1 using SEM: $L = 15$, $K_{init} = 11$, $K_{end} = 4$, $\xi = 0.01$	34
Figure 16	K-S test for Fort Hood 8 using SEM: $L = 15$, $K_{init} = 14$, $K_{end} = 13$, $\xi = 0.01$	35
Figure 17	Illustration of detection and thresholding.	37
Figure 18	Receiver Operating Characteristic (ROC) elements.	38
Figure 19	Illustration of a scene with full and sub-pixel targets.	40
Figure 20	3-d geometric interpretation of the subspace GLRT detector.	44
Figure 21	Vis-SWIR, buried mines: (a) mine type 1, constant false alarm rate detection; (b) mine type 2, 100% detection.	52
Figure 22	Vis-SWIR, surface mines: (a) mine type 2, constant false alarm rate detection; (b) mine type 2, 100% detection.	53

Figure 23	LWIR, buried mines: (a) mine type 2, constant false alarm rate detection; (b) mine type 2, 100% detection.	54
Figure 24	LWIR, surface mines: (a) mine type 2, constant false alarm rate detection; (b) mine type 2, 100% detection.	55
Figure 25	Theoretical ROC for matched filter detector on sub-pixel targets.	56
Figure 26	Theoretical ROC for GLRT detector on targets of varying dimensionality and with varying training support.	57
Figure 27	VNIR linescanner RGB images for three AHI scenes: (a) scene A (b) scene B (c) scene C.	61
Figure 28	Subspace GLRT vs. Anomaly Detector, benzene, scene C: (a) Subspace GLRT, (b) Anomaly Detector, (c) color scale.	63
Figure 29	Subspace GLRT vs. Anomaly Detector, ethylene glycol, scene A: (a) Subspace GLRT (b) Anomaly Detector.	64
Figure 30	PCA images, scene A: (a) RGB of all 3 components (b) component 1 (c) component 2 (d) component 3.	66
Figure 31	MNF images, scene A: (a) RGB of all 3 components (b) component 1 (c) component 2 (d) component 3.	67
Figure 32	PCA images, scene B: (a) RGB of all 3 components (b) component 1 (c) component 2 (d) component 3.	68
Figure 33	MNF images, scene B: (a) RGB of all 3 components (b) component 1 (c) component 2 (d) component 3.	69
Figure 34	Scene B, SO ₂ analysis (a) RGB image using 9.2-10.7-8.6 μm bands (b) principal component 4.	71
Figure 35	Scene B, SO ₂ RGB image using 9.2-10.7-8.6 μm bands: (a) 3-d RGB colorspace, (b)-(d) 3-d scatter plots from various view angles.	72
Figure 36	Three dimensional vector space illustration of small spectral angle between target 3 and background in AHI scene 3.	76
Figure 37	Detection statistic output using contaminated covariance estimate (left, blue) and uncontaminated estimate (right, red) for AHI data.	80
Figure 38	Graphical presentation of residual covariance error (clean - contaminated) for Sensor X. Run 6300: (a) and (b). Run 5700: (c) and (d).	81
Figure 39	Target detectability for dimension reduced data: (a) AHI, scene 1; (b) AHI, scene 4; (c) Sensor X, run 1068; (d) Sensor X, run 6300.	88

Figure 40	Matched filter detector output for dimension reduced data: (a) AHI, scene 1; (b) AHI, scene 4; (c) Sensor X, run 1068; (d) Sensor X, run 6300.	90
Figure 41	GLRT detector output for dimension reduced data: (a) AHI, scene 1; (b) AHI, scene 4; (c) Sensor X, run 1068; (d) Sensor X, run 6300.	91
Figure 42	ACE detector output for dimension reduced data: (a) AHI, scene 1; (b) AHI, scene 4; (c) Sensor X, run 1068; (d) Sensor X, run 6300.	92
Figure 43	Subspace GLRT detector output for dimension reduced data: (a) AHI, scene 1; (b) AHI, scene 4; (c) Sensor X, run 1068; (d) Sensor X, run 6300.	93

SUMMARY

Whenever a new sensor or system comes online, engineers and analysts responsible for processing the measured data turn first to methods that are tried and true on existing systems. This is a natural, if not wholly logical approach, and is exactly what has happened in the advent of hyperspectral imagery (HSI) exploitation. However, a closer look at the assumptions made by the approaches published in the literature has not been undertaken.

This thesis analyzes three key aspects of HSI exploitation: statistical data modeling, covariance estimation from training data, and dimension reduction. These items are part of standard processing schemes, and it is worthwhile to understand and quantify the impact that various assumptions for these items have on target detectability and detection statistics.

First, the accuracy and applicability of the standard Gaussian (i.e., Normal) model is evaluated, and it is shown that the elliptically contoured t -distribution (EC- t) sometimes offers a better statistical model for HSI data. A finite mixture approach for EC- t is developed in which all parameters are estimated simultaneously without a priori information. Then the effects of making a poor covariance estimate are shown by including target samples in the training data. Multiple test cases with ground targets are explored. They show that the magnitude of the deleterious effect of covariance contamination on detection statistics depends on algorithm type and target signal characteristics. Next, the two most widely used dimension reduction approaches are tested. It is demonstrated that, in many cases, significant dimension reduction can be achieved with only a minor loss in detection performance.

In addition, a concise development of key HSI detection algorithms is presented, and the state-of-the-art in adaptive detectors is benchmarked for land mine targets. Methods for detection and identification of airborne gases using hyperspectral imagery are discussed, and this application is highlighted as an excellent opportunity for future work.

CHAPTER 1

INTRODUCTION

Since the first flight of NASA’s Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) instrument 20 years ago, interest in algorithms for exploitation of hyperspectral imaging (HSI) data has grown by leaps and bounds. A variety of sensors have been built to satisfy the research interests of both commercial and government sponsors, and practitioners from all over the world have shifted their attention to this new form of remote sensing. The ability of this new technology to discriminate spectral signals (i.e., targets) of interest is intriguing to a wide audience: academia, commercial industry, and the military alike. As a passive technique hyperspectral imaging offers the advantage of not being detectable by the objects or adversaries being sensed, and requires electronics that consume relatively little power. To illustrate the variety of applications for HSI, figure 1 breaks down common spectral exploitation tasks and their region of support, from visible to long-wave infrared (LWIR). It is clear that there is tremendous utility across a wide spectral region, and as such there is a real need for robust algorithms in exploitation tasks such as target detection.

1.1 Research Overview

This thesis presents a series of analyses on various aspects on the signal processing chain for hyperspectral image data. Target detection is by far the most common HSI exploitation task. This research pokes and prods standard assumptions of the adaptive detection problem in order to gain both a qualitative and quantitative feel for their accuracy and importance in maintaining the overall fidelity of the final information product – the detection statistic. Along the way, a few peripheral but altogether relevant and interesting items are discussed, such as detecting land mines and hazardous airborne gases.

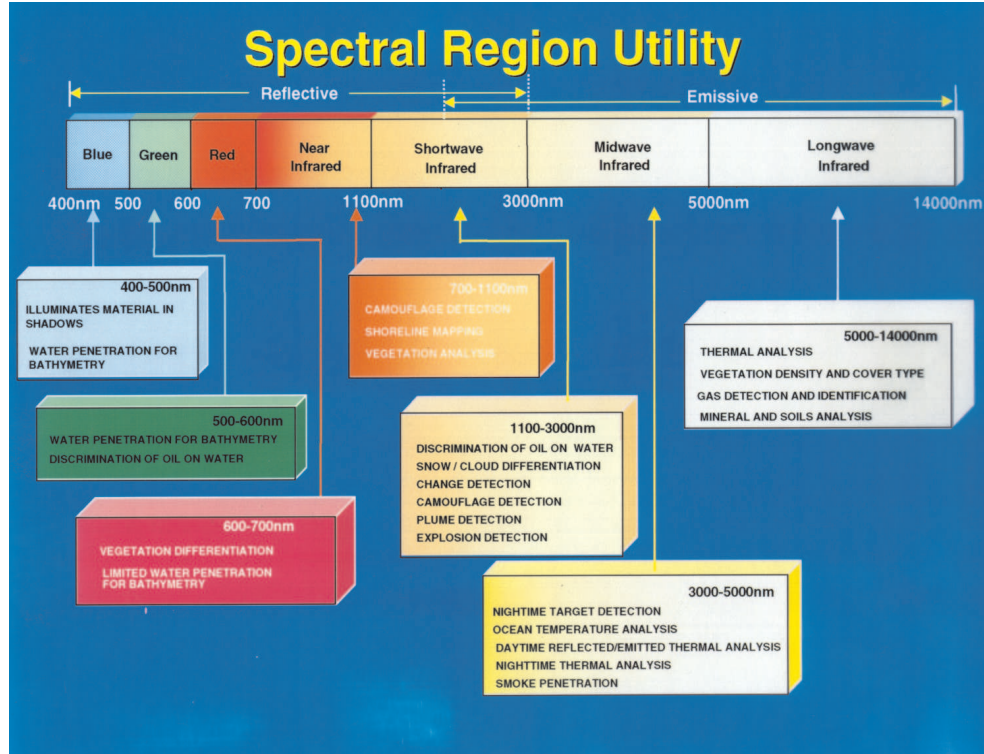


Figure 1: A breakdown of common spectral data exploitation tasks and their regions of support, from visible to LWIR.

Specifically, this thesis is organized as follows:

Chapter 2 presents a brief background of hyperspectral imaging, frames the problem of adaptive detection, and outlines key similarities and differences between HSI and radar.

Chapter 3 discusses statistical modeling of spectral data and ways to measure to goodness-of-fit of statistical models. A scheme for simultaneously estimating all the parameters of an EC- t finite mixture model is covered in detail. Results show that the elliptically contoured t -distribution offers a valid modeling alternative to those based on Normal distributions.

Chapter 4 provides a direct, concise treatment of algorithms for target detection in hyperspectral imaging. A signal processing perspective, something rarely found in the

remote sensing literature, is taken in the description of signal models and performance evaluation for state-of-the-art algorithms.

Chapter 5 evaluates the set of detectors presented in Chapter 4 for multiple, real-world data sets. Land mines, one of the most difficult ground targets to detect passively, are the backdrop for this evaluation.

Chapter 6 further explores some of the algorithms presented in Chapter 4, but this time applied to the task of detecting airborne gases and chemical plumes. Other, non-algorithmic methods for identifying a gas specimen are also presented.

Chapter 7 provides an analysis of the deleterious effect of covariance estimate contamination. Qualitative (e.g., visual) and quantitative (e.g., empirical) results clearly show that poor training data including target-like samples can have a significant impact on detection results.

Chapter 8 offers analysis of the impact that the most widely-used dimension reduction methods have on detector output. Reducing dimensionality is a common pre-processing step, yet very little is known on the impact that dimension reduction transformations have on target detectability and detection performance.

Chapter 9 concludes the thesis with a summary of contributions and future work.

CHAPTER 2

BACKGROUND

2.1 Review of Hyperspectral Imaging

Generally speaking, electro-optical (E-O) remote sensing involves the acquisition of information about a scene or object without making physical contact with it. Hyperspectral imagers are a class of E-O imaging spectroscopy sensors in which the waveband of interest is divided into hundreds of contiguous narrow bands (i.e., image channels) for the purpose of signature analysis. Figure 2 shows the concept of hyperspectral imaging. Hyperspectral imagers offer high spectral resolution that preserves important aspects of the spectrum (i.e., the shape of narrow absorption bands) and makes it possible to differentiate distinct materials on the ground. The basic principle is that materials reflect, absorb, and emit electromagnetic radiation in ways characteristic of their molecular composition and shape [1], [2], [3], [4]. The spatially and spectrally sampled information is typically visualized as a ‘data cube’, whose face is a function of the spatial coordinates and whose depth is a function of spectral band (i.e., wavelength). In the wavelength dimension, each image pixel is a vector that provides a spectrum characterizing the materials within the pixel. Conversely, the data in each band corresponds to a narrowband image of the surface covered by the field of view of the sensor. Progress in multi-channel (i.e., spectral band) imaging has been evolutionary, with the width and number of channels steadily improving as the quality of focal plane technology has increased [5].

Many commercial and a significant number of government-funded hyperspectral imagers operate in the reflective regime of the electromagnetic spectrum. Ranging from

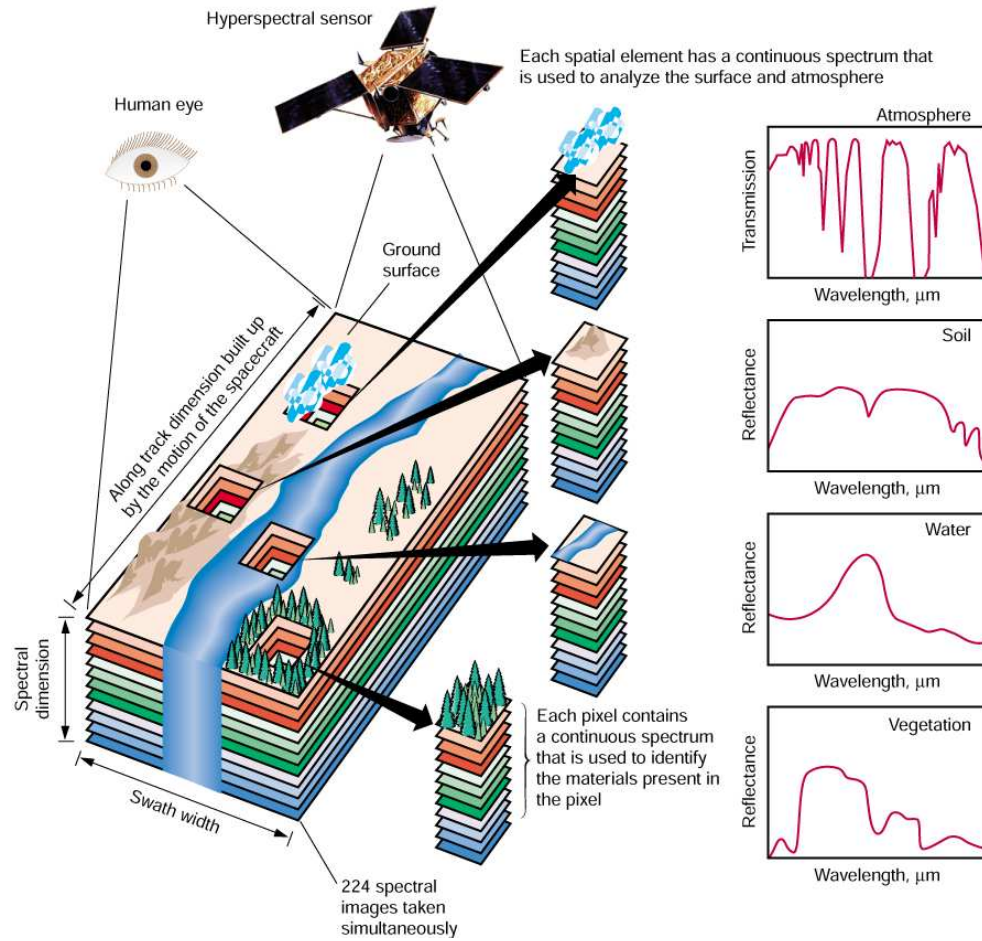


Figure 2: Concept of hyperspectral imaging illustrated for a pushbroom sensor.

approximately $0.4\text{--}2.5\ \mu\text{m}$ ($400\text{--}2500\ \text{nm}$), this portion of the spectrum covers the visible through short-wave infrared (SWIR) wavelengths. Hyperspectral imaging systems designed for the emissive regime, also called the thermal or long-wave IR region, typically operate in the range from $7\text{--}14\ \mu\text{m}$. These sensors are less prevalent in the commercial and academic research communities than in government since they require more sensitive optics and complicated electronics that are more costly.

While the concept of hyperspectral imaging is straightforward, there are a number of practical considerations that must be dealt with in fielding a hyperspectral sensor and processing its data. Although not the focus of this research, it is worth mentioning that environmental factors play an enormous role. Atmospheric effects such as absorption and

scattering are chief among these. Also, viewing angle, secondary illumination, and shadowing come into play. Spatial and spectral resolution trade-offs are constantly considered by system engineers [6]. From a signal processing perspective, the spectral variability exhibited by a given material (largely resulting from surface roughness) is probably the greatest challenge for algorithm developers.

2.2 Framework for Adaptive Detection

Despite its lifespan of only 30 years, the area of adaptive array signal processing has a rich history. This is evident from the bodies of literature and numerous conferences devoted to the topic, as well as from technologies spawned or furthered by successes in the field. Specifically, the contemporary formulations of adaptive array detection trace their roots back to Reed, Mallett, and Brennan [7] and later to Kelly. These individuals (along with a few others) were instrumental not only in formalizing this area of signal processing, but also in making it more widely known and popular with those outside a narrow community of researchers tackling problems for the defense establishment. Of particular interest to this thesis is the Generalized Likelihood Ratio Test approach taken by Kelly and a series of detection statistics that resulted. The problem statement offered by Kelly is summarized below.

Adaptive array detection considers the problem of determining signal presence in a single $L \times 1$ array observation \mathbf{x} . The radar community refers to \mathbf{x} as a ‘snapshot’ or the primary data vector. In the case of multi-channel (e.g., hyperspectral) imaging, the sample is a pixel vector. This primary data vector has an unknown covariance denoted by $\mathbf{\Gamma}$, which, at times, may also be scaled by a known constant κ . Essentially, it is desired to declare the observation as one of two things:

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

$$H_1 : \mathbf{x} = \mathbf{S}\mathbf{a} + \mathbf{v};$$

either the primary data vector consists of interference only (i.e., background¹ plus noise), represented by null hypothesis H_0 , or it consists of signal plus noise, represented by hypothesis H_1 . When the target signal is not present in the case of H_0 , there is still other signal energy measured. However, it is not of interest and is included along with the additive noise, together denoted by \mathbf{v} , such that it encapsulates both noise and background. When the target signal is present in the primary data vector in the case of H_1 , the signal model consists of an $L \times P$ matrix \mathbf{S} , multiplied by an unknown vector of target signal parameters \mathbf{a} , plus an additive noise term \mathbf{v} . The matrix \mathbf{S} can be thought of as a system transfer function. This is analogous to the radar world, where \mathbf{S} contains the steering vectors for multiple pulses of radar echo returns. In the case of hyperspectral detection, \mathbf{S} contains the a priori information available about the target. As the number of columns of \mathbf{S} decreases (i.e., dimension of target subspace), the information on the target increases. This is to say that there is less variability in the target signal model. In fact, for a deterministic target $P = 1$, and since there is only a single column, the target is represented by \mathbf{s} .

This detection problem has two unknowns, $\mathbf{\Gamma}$ and \mathbf{a} . To accommodate the ignorance of these ‘nuisance’ parameters, it is assumed that a secondary data set (i.e., a training set) is available. Training data is assumed to be independent and identically distributed (i.i.d.) to the test data. In denoting the training data with N samples as $\mathbf{X} = [\mathbf{x}_1 | \mathbf{x}_2 | \dots | \mathbf{x}_N]$, it is assumed that each pixel vector \mathbf{x}_i contains background only and shares the same covariance as the primary data vector. As such, the determination on signal presence will not be made on observing the primary data vector \mathbf{x} alone, but rather based on the totality of the data summarized by \mathbf{X} and the primary data vector. Under both hypotheses, it is assumed that $N \geq L$.

¹The term ‘background’ is frequently used in the HSI literature and is synonymous with the term clutter in the radar literature. Both systems collect measurements that contain unwanted energy. Data collected by hyperspectral sensors is ultimately still an image, and as such those samples (e.g., pixels) containing anything other than the target signal of interest are called background. While the term interference is really more apt since it means anything that is not wanted, the term background is more popular in the HSI literature, and we may use the terms interchangeably.

An important quantity that dominates the detection schemes discussed in this thesis is the covariance matrix, also referred to as the sample covariance. Throughout this document, a known quantity such as the covariance is simply denoted by Γ , whereas an estimated quantity, such as the maximum likelihood (ML) estimate of the covariance matrix, is denoted by $\hat{\Gamma}$. Of course, for the Normal case, this is

$$\hat{\Gamma} = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \hat{\boldsymbol{\mu}})(\mathbf{x}_n - \hat{\boldsymbol{\mu}})^T, \quad (1)$$

and $\hat{\boldsymbol{\mu}}$ is the ML estimate of the mean of the data set. Keeping with standard notation, boldface upper-case letters and symbols are matrices and boldface lower-case letters and symbols are vectors.

2.3 *Hyperspectral vs. Radar*

Spawned by research and development successes in the mid 1970s and late 1980s, a significant body of work now exists in the area of radar array processing, including adaptive detection. This research was motivated by the need for more accurate and robust radar target detection, driven by the production of increasingly more capable radar systems. Despite the maturation of other remote sensing technologies in recent years, however, the majority of publications on detection algorithm development within the statistical signal processing community are still focused on radar.

One theme of this thesis is to revisit the classic adaptive detection problem developed for radar array processing and apply it to hyperspectral imaging. While key parallels exist that make this a promising proposition, a number of important items must be addressed to successfully employ adaptive detection concepts to HSI. Some items have already been discussed in the literature; others have not.

Many of these differences are a direct result of the remote sensing phenomenology. Radar is an active system, illuminating the target with coherent pulses. HSI is a passive technique, relying on incoherent solar illumination from the sun to provide energy in the

scene. Radar transmitters are instruments that send out electromagnetic pulses in a well-defined, controlled fashion. For this reason, radar signals can be, and often are, constructed to be zero mean. Data measured by hyperspectral sensors are electro-optical signals that are the result of many complicated interactions solar radiated energy has with the atmosphere and ground. These signals are decidedly not zero mean. Another major difference is that radar data are complex, while HSI data are real valued. Further, even though it is not a requirement of the adaptive detection structures developed for radar, the literature is dominated by work where both the dimensionality of the problem and the number of samples are small. On the contrary, HSI data sets often contain many hundred thousand samples and are of a dimension in the hundreds.

When considered together, these factors amount to substantial and significant differences in processing hyperspectral imaging data for adaptive detection. Table 1 offers a concise summary of these differences.

Table 1: Key differences between radar and hyperspectral for adaptive detection.

HSI	RADAR
natural illumination	man made illumination
incoherent energy	coherent energy
passive technique	active technique
electro-optical	electro-magnetic
many pixel vectors	few snapshots
high dimensionality (100's)	low dimensionality (10's)
real-valued	complex-valued
never zero-mean	almost always zero-mean

CHAPTER 3

STATISTICAL DATA MODELING AND PARAMETER ESTIMATION

For this thesis, research has been conducted in a number of areas related to multivariate statistical analysis. First, existing statistical models for hyperspectral imaging data of different types were evaluated for goodness-of-fit using both traditional and contemporary metrics. As a result, it was confirmed that models based on a Normal distribution can be inaccurate. We then showed that densities from the family of elliptically contoured (EC) distributions can lead to more accurate models with smaller residual error, specifically models based on the elliptically contoured t (EC- t) distribution. Two automated techniques for generating models based on a mixture of EC- t distributions were developed, both of which are novel in that they require no manual manipulation of parameters during the process. Also they do not require any a priori information.

3.1 Measuring Goodness-of-fit for Statistical Data Models

A statistical test in which the validity of one hypothesis is tested without specification of an alternative hypothesis is called a goodness-of-fit test. The general procedure consists of defining a test statistic, which is some function of the data measuring the distance between the hypothesis and the data (in fact, the ‘badness-of-fit’), and then calculating the probability of obtaining data that have a still larger value of this test statistic than the value observed, assuming the hypothesis is true. For the case of modeling hyperspectral imaging data, goodness-of-fit tests are used to see whether a group of pixels $\mathbf{X} = [\mathbf{x}_1 | \mathbf{x}_2 | \dots | \mathbf{x}_N]$ matches a theoretical distribution such as the multivariate Normal. If so, algorithms can be designed with significant assumptions and can take advantage of desirable statistical properties. Goodness-of-fit tests can be employed on a variety of distributions for a given data set, quickly providing the residual error from the model to the data. The distribution that

yields the smallest residual error is the best fit.

The Mahalanobis distance is a familiar quadratic term and is defined for the Normal distribution as

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

According to [8] the distribution of the estimated Mahalanobis distance $\hat{\Delta}$ (using ML estimates for the mean and covariance) is well-modeled using a Chi-Square distribution when the underlying data are multivariate Normal.

As such, goodness-of-fit tests can be used to evaluate how well the univariate statistic $\hat{\Delta}$ follows a theoretical curve. To test data that supposedly follow a Normal distribution, we compare $\hat{\Delta}$ to a Chi-Square. This is a powerful approach, since goodness-of-fit tests for multivariate distributions are naturally more complicated and since the Mahalanobis distance is a familiar quantity that can be easily computed using equation (2).

3.1.1 Kolmogorov-Smirnov Test

Given a data set \mathbf{y} and computing its empirical cumulative distribution function $F_d(\mathbf{y})$, the Kolmogorov-Smirnov test [9] is

$$D = \max_i |F_d(y_i) - F(y_i)|, \quad (3)$$

where $F(y_i)$ is the cumulative distribution function (cdf) under test at point i . The Kolmogorov-Smirnov test compares the empirical cdf of the given data set with that of a known cdf by computing the maximum difference between the theoretical cdf and the empirical cdf (ecdf) for all points in \mathbf{y} . The result of the test D is the maximum difference between the two values at all points in the data set.

A closer examination of the Kolmogorov-Smirnov test reveals a possible weakness for the goal of paying careful attention to the tails of the distribution when measuring goodness-of-fit. At the tails of the distribution, the values are small, so even though the difference between the theoretical cdf and the ecdf of the data at a given point in the tail may be *relatively* very large – and therefore significant to target detection because precious false

alarm probability (P_{FA}) depends largely on the tails – it is quite possible that the magnitude of a difference between two values at a given point in the main body of the distribution may be larger, despite being relatively insignificant.

3.1.2 Chi-Square Test

An alternative to the Kolmogorov-Smirnov test is the Chi-Square test. It has the attractive feature of being applicable to any univariate distribution for which the cumulative distribution function can be calculated. For a given distribution, the Chi-Square test [10] compares the actual number of observations in an interval to the given number of observations in the same interval. Here, K equiprobable intervals are used to cover the univariate probability density function (pdf) for the given distribution, and each of these intervals has a probability of $1/K$. For the number of intervals K , the number of expected data points in each interval C , and the number of actual data points in the i^{th} interval G_i , the Chi-Square test is

$$D = \sum_{i=1}^K \frac{(G_i - C)^2}{C}. \quad (4)$$

Each of the intervals represents a region of equiprobability and the test gives equal weight to each of the K intervals. Recall, however, that the tails of a Normal distribution are regions of low probability. As such, the tail of the distribution is covered by only a few intervals, while the main body of the distribution is covered by several intervals. This means that the Chi-Square test tends to reliably match the main body of the given data set with the distribution under test, but offers a poor evaluation of the tails. Again, this is not very desirable for a goodness-of-fit test whose ultimate application (e.g., target detection) cares about the tails of the background distribution.

3.1.3 Exceedance Metric

Recognizing the limitations of the previous tests in measuring goodness-of-fit for situations where the tails of the distribution are important, Marden [11] was the first in the remote sensing literature to identify a test that properly evaluates the fit of the tails of the empirical

distribution. Called the Exceedance metric, it compares the exceedance curve of the given data with that of a theoretical distribution. The function for probability of exceedance $E(\mathbf{y})$ for a cdf $F(\mathbf{y})$ is

$$E(\mathbf{y}) = 1 - F(\mathbf{y}). \quad (5)$$

The inverse of the exceedance function is then

$$E^{-1}(P) = \{\mathbf{y} : 1 - F(\mathbf{y}) = P\} \quad (6)$$

which is the value of \mathbf{y} where the exceedance curve evaluates to P . If we take the inverse of the exceedance curve of the proposed distribution at the i^{th} point we have $E^{-1}(P_i)$. The inverse of the exceedance curve of the actual data is $E_n^{-1}(P_i)$ and is evaluated at the point y_i , where $1 - F(y_i) = P_i$. The points P_i are K equally log-spaced steps on the probability axis of the exceedance curves. Together these terms are used to construct the Exceedance metric

$$D = \sum_{i=1}^K |[E_n^{-1}(P_i) - E^{-1}(P_i)]|. \quad (7)$$

When trying to show the shape of the tails of a distribution, exceedance curves prove to be quite useful. Compared with other tests, the Exceedance metric as a goodness-of-fit test does a better job of modeling the tails of a distribution.

3.1.4 Other Variants and Modifications

It is possible to modify the Chi-Square test so that it uses only a desired fraction of the upper part of the data set, for example 10%. This corresponds to the upper tail of the distribution and assumes that the distribution is one-sided. Similar to the description in Section 3.1.2, the modified Chi-Square test divides the data set's tail into K equiprobable intervals, where each interval has a probability of $0.1/K$ (in the case of 10%). Each interval then has a corresponding number of data points C , which is constant, and the number of actual data points in each interval G_i is computed. Goodness-of-fit is computed using equation (4) as before.

A modification of the Kolmogorov-Smirnov (K-S) test, which tries to give more weight to the tails of the distribution, is the Anderson-Darling test. The K-S test is distribution free in the sense that the critical values do not depend on the specific distribution being tested. The Anderson-Darling test makes use of the specific distribution in calculating critical values. This has the advantage of allowing a more sensitive test and the disadvantage that critical values must be calculated for each distribution.

Certainly, the three goodness-of-fit tests presented here do not constitute an exhaustive set. Over the years, minor modifications have been made to fundamental tests and variations may or may not work better in a specific application of goodness-of-fit testing or model-fitting. However, the general approaches taken by the Kolmogorov-Smirnov, Chi-Square, and Exceedance tests are representative of fundamental goodness-of-fit techniques. A comprehensive treatment of goodness-of-fit techniques can be found in [12].

3.1.5 Results for Normal

Data collected by the Airborne Hyperspectral Imager (AHI) [13], a long-wave IR spectrum sensor built and operated by the University of Hawaii, was used to assess the accuracy of the multivariate Normal distribution in modeling hyperspectral data. As discussed above, the Mahalanobis distance Δ of a Normal random vector \mathbf{x} follows a Chi-Square distribution. However, as the Exceedance metric in Figure 3 clearly shows, the data do not come close to following the χ^2 curve. Tests on other data sets from the visible/short-wave IR spectrum show similar results; the Normal often does a poor job of fitting hyperspectral imaging data.

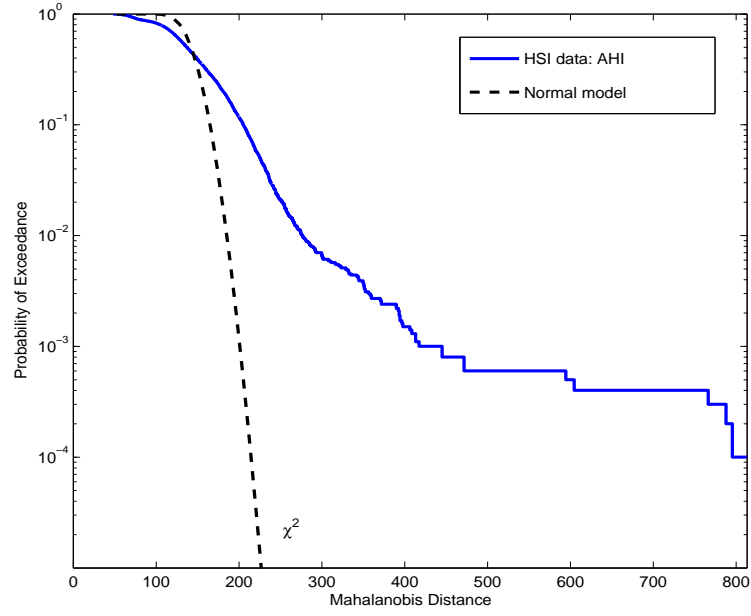


Figure 3: Exceedance metric goodness-of-fit test for the Normal distribution.

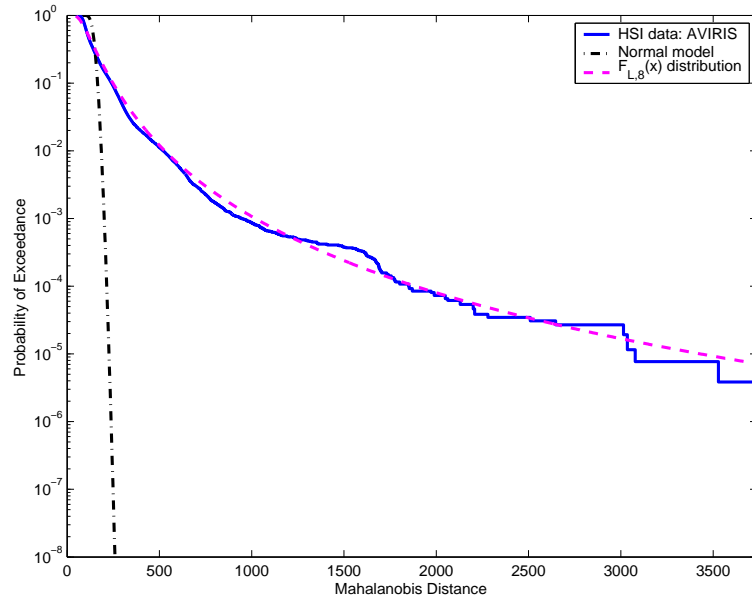


Figure 4: Exceedance metric goodness-of-fit test for the EC- t distribution.

3.1.6 Results for Elliptically Contoured t

After testing a number of different distributions, the elliptically contoured t distribution offered the best fit on all the data sets tested. The family of EC distributions includes the Normal as a special case, as well as the Weibull, K, Cauchy and others [14]. All of these distributions share the familiar symmetry of the Normal distribution and are characterized by their contours of equiprobability [15]. One of the things that makes the EC- t a good match for hyperspectral data is a third parameter, the degrees of freedom [16]. This parameter can be used to tune the tails of the distribution so that it nicely matches the heavier tails exhibited by HSI data.

Certainly, the t and others in the family of EC distributions are not the only alternative for dealing with long, heavy tails not handled by the Normal. The family of (symmetric) alpha-stable distributions also has shown promise in modeling heavy-tailed radar clutter [17]. However, many of the distributions in the alpha-stable family have infinite variance and/or do not have closed-form expressions for their moments. Further, statistical inference of the type we are concerned with in adaptive detection is extremely complicated for the alpha-stable family [18]. As such, the EC- t is a more practical alternative for this research.

The reason it is important to accurately model the tails of the background distribution in target detection is related to false alarm performance and constant false alarm rate (CFAR) operation. Threshold selection, an important task in the overall detection process, is predicated on the background distribution. In CFAR operation, the tail of the background distribution is used to integrate out a constant value and the threshold is set at the point where that value is achieved.

Figure 4 again shows an Exceedance metric goodness-of-fit test for the Mahalanobis distance, but this time the data is from the AVIRIS sensor [19] in the reflective regime, and the Chi-Square curve representing the Normal model is joined by an F distribution curve representing an EC- t model. Clearly, the F curve, which characterizes the quadratic term (i.e., Mahalanobis distance) of the EC- t distribution, is a much better fit for hyperspectral

imaging data.

3.2 Finite Mixtures

Hyperspectral images are inherently spectrally inhomogeneous. Even though many hyperspectral imaging sensors have a narrow field of view, the altitudes from which the data are collected dictate that the imaged scenes contain many different physical materials. For example, at an altitude of 705 km, each pixel captured by the Hyperion satellite hyperspectral imager [20] covers 30 meters of ground on a side.

As indicated in Section 2.1, HSI sensors are employed because of their ability to differentiate one material from another by spectral properties. In ground cover classification, the goal is to create a thematic map, or simply a color-coded image that represents what material is present at each pixel location. The variability of each material type must be first characterized, using either probabilistic or geometric constructs. When applying a decision rule in the classification procedure, the results will be valid and useful if the class definitions are accurate. Similar logic applies for target detection applications; the background data must be modeled accurately so the response of these pixels is well-separated from target pixels in the output detection statistic.

Since inhomogeneity of the data reduces the robustness of spectral models, accurately representing the variability present in the data is important.

Parameterized models are popular, with the simplest being the univariate Normal distribution. As seen in Section 3.1.5, very rarely does a single Normal distribution accurately characterize the variability of data collected by operational hyperspectral imaging sensors. Instead, mixtures of distributions have shown promise to more precisely model the data than a single distribution [21] [22]. Intuitively, this mixture approach to data modeling has appeal since it can be thought of as combining individual probability density functions (pdfs) to get a multi-modal pdf; one mode is assumed for each class of spectrally homogeneous material present in the image. Without a priori spectral information about

the materials present, and spatial information on the fraction of the image each material covers, directly specifying parameters for each component of a mixture model is difficult. In many if not most remote sensing scenarios, such a priori information is not available, and parameters for a mixture model must be estimated from the data.

As such, a reasonable probabilistic model for spectral data is provided by the mixture density

$$f(\mathbf{x}) = \sum_{i=1}^K \pi_i f(\mathbf{x}; \boldsymbol{\theta}_i). \quad (8)$$

This is known as a finite-mixture model and has been espoused by many authors for a number of different statistical contexts, in particular [23] [24]. The key assumptions driving the use of finite-mixture models in hyperspectral image analysis are that unique materials exhibit a representative spectrum and that the inter-class spectral variability of different materials can be used to separate an entire scene into multiple homogeneous classes. These groups of pixels can then, in turn, be accurately characterized by a single uni-modal multivariate pdf. Each mode of the mixed pdf corresponds to a unique material and, in combination, account for the spectral variability of the entire inhomogeneous scene within a single function.

3.3 Parameter Estimation for EC- t Mixture Models

The two major items to be addressed in constructing such a mixture model are the form for each pdf component in the mixture and the method for estimating all the parameters of each component's distribution. The first issue has been chosen based on preliminary research findings presented in Section 3.1.6; that is, EC- t densities will be used. The second issue is discussed in this section.

When trying to compute maximum likelihood estimates, a number of methods can be employed. Newton-Raphson, quasi-Newton, and modified-Newton are all Newton-type methods that can be used to find maximum likelihood estimates. However, these methods

can quickly become quite complicated for many estimation problems and a stable alternative is required. An iterative approach that is widely applicable to the computation of ML estimates is the Expectation-Maximization algorithm. This algorithm has demonstrated its utility in a variety of so-called incomplete data problems. On each iteration of the EM algorithm there are two steps: called the expectation step (E step) and the maximization step (M step). Because of these two steps, Dempster et al. so named the algorithm ‘EM’ in their paper [25].

3.3.1 Expectation-Maximization

The EM algorithm augments the observed data \mathbf{Y}_{obs} (i.e., incomplete data) to the larger \mathbf{Y}_{aug} (i.e., complete data). Starting with an initial value $\Psi^{(0)} \in \Psi$, it then finds Ψ^* , a stationary point of $L(\Psi|\mathbf{Y}_{obs})$, by iterating the following two steps for $(j = 0, 1, 2, \dots)$ iterations:

E step – impute the augmented data (log-)likelihood $L(\Psi|\mathbf{Y}_{aug})$ by

$$Q(\Psi|\Psi^{(j)}) = E[L(\Psi|\mathbf{Y}_{aug})|\mathbf{Y}_{obs}, \Psi^{(j)}], \quad (9)$$

M step – determine $\Psi^{(j+1)}$ by maximizing the imputed (log-)likelihood $Q(\Psi|\Psi^{(j)})$

$$Q(\Psi^{(j+1)}|\Psi^{(j)}) \geq Q(\Psi|\Psi^{(j)}) \quad \forall \Psi \in \Psi. \quad (10)$$

Again, the idea is to select \mathbf{Y}_{aug} such that $\Psi^{(j+1)}$ is easy to compute, thereby providing a simple, stable algorithm.

Here, EM is used to estimate four parameters, the weight of each component in the mixture and three parameters for each pdf. The elliptically contoured t distribution also has a parameter that controls the shape of its tail, in addition to the well-known mean and covariance parameters of the Normal. The degree of freedom parameter specifies the rate of decay for the tails of the t distribution and allows for heavier tails to accommodate hyperspectral data.

Recall that \mathbf{x} is an L -dimensional column vector representing a pixel in the hyperspectral image and Θ is the set of parameters for the multivariate uni-modal pdf f_L , also of dimension L . Both Θ and π range from $i = 1 \dots K$, where K is the number of components in the mixture model, and the π_i are mixture weights (i.e., priors) for each component. A unity sum is enforced for the mixture weights, $\sum_{i=1}^K \pi_i = 1$. The entire set of parameters (collection of all Θ 's and π 's) for the mixture model is denoted by Ψ . The multivariate t density is

$$f_L(\mathbf{x}; \boldsymbol{\mu}, \mathbf{C}, \nu) = \frac{\Gamma[\frac{L+\nu}{2}]}{\Gamma[\frac{\nu}{2}](\nu\pi)^{\frac{L}{2}}} |\mathbf{C}|^{-\frac{1}{2}} [1 + \frac{1}{L}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu})]^{-\frac{L+\nu}{2}} \quad (11)$$

$$f_L(\mathbf{x}; \Theta_i) \sim t_L(\mathbf{x}; \boldsymbol{\mu}_i, \boldsymbol{\Gamma}_i, \nu_i)$$

where L is the length of \mathbf{x} , ν is the dof, $\frac{\nu}{\nu-2}\mathbf{C} = \boldsymbol{\Gamma}$ is the covariance matrix, and $\boldsymbol{\mu}$ is the mean vector. Not to be confused with the covariance matrix, unbolded Γ is the Gamma function in equation (11).

Once the number of components K is set, the next step is determining how to initialize the mixture model. For ML estimation using EM, initialization is extremely important to finding the global maximum of the likelihood function in addition to rate of convergence. This is because EM ensures finding a maximum of the likelihood function using equations (9) and (10), but since the function often has multiple maxima (i.e., many peaks), there is no guarantee that the root found is the global maximum (i.e., tallest peak).

We use K randomly chosen pixels to seed the segmentation process, one for each mixture component. Random seeding is a simple mechanism and indeed not optimal, but choosing random pixels as a starting point actually has a number of practical advantages over Euclidean distance metrics or spectral angle measures for initialization. While using the centroids found by pre-clustering the data may or may not lead to starting points that are nearer to the global maximum in the likelihood space, such operations require additional computation. They are also randomly initialized themselves, and as such ultimately provide no certainty that the initialization will be close to the global maximum.

Given the data \mathbf{X} for N number of pixels, the objective is to estimate the mixture model parameters $\Psi = [\pi_1, \pi_2, \dots, \pi_K, \mu_1, \mu_2, \dots, \mu_K, \Gamma_1, \Gamma_2, \dots, \Gamma_K, \nu_1, \nu_2, \dots, \nu_K]$. Seeking ML estimates for Ψ , the pixel likelihood function is used:

$$L(\mathbf{x}_n; \Psi) = \sum_{k=1}^K \pi_k f_k(\mathbf{x}_n; \mu_k, \Gamma_k, \nu_k). \quad (12)$$

The data likelihood function is then

$$L(\mathbf{X}; \Psi) = \prod_{n=1}^N L(\mathbf{x}_n; \Psi). \quad (13)$$

Since the likelihood space is complicated and unfriendly to the application of brute force to directly solve equation (13), the use of the EM technique is practical to find $\hat{\Psi}$.

Once the statistics for each component of the model $\hat{\Psi}_i^{(0)}$ have been initialized, the expectation step is executed. Here, the posterior probability P is computed for the i^{th} component at the j^{th} iteration by

$$P_i(\mathbf{x}_n; \hat{\Psi}_i^{(j)}) = \frac{\hat{\pi}_i f_i(\mathbf{x}_n; \hat{\mu}_i, \hat{\Gamma}_i, \hat{\nu}_i)}{L(\mathbf{x}_n; \hat{\Psi}_i^{(j)})}. \quad (14)$$

This posterior computes the probability that the n^{th} pixel is a member of the i^{th} data cluster¹. At each iteration, maximization is the second half of EM. Here, the parameter estimates are updated according to the posterior just computed in equation (14). The mean, covariance, and mixture weights are respectively updated by

$$\hat{\mu}_i^{(j+1)} = \frac{\sum_{n=1}^N \mathbf{x}_n P_i(\mathbf{x}_n; \hat{\Psi}_i^{(j)})}{\sum_{n=1}^N P_i(\mathbf{x}_n; \hat{\Psi}_i^{(j)})} \quad (15)$$

$$\hat{\Gamma}_i^{(j+1)} = \frac{\sum_{n=1}^N (\mathbf{x}_n - \hat{\mu}_i^{(j+1)})(\mathbf{x}_n - \hat{\mu}_i^{(j+1)})^T P_i(\mathbf{x}_n; \hat{\Psi}_i^{(j)})}{\sum_{n=1}^N P_i(\mathbf{x}_n; \hat{\Psi}_i^{(j)})} \quad (16)$$

$$\hat{\pi}_i^{(j+1)} = \frac{\sum_{n=1}^N P_i(\mathbf{x}_n; \hat{\Psi}_i^{(j)})}{N}. \quad (17)$$

¹To be precise, each component of the model is a pdf. The term cluster refers to a set of data samples that is the result of using the model to assign each sample to one and only one cluster. Further, these pixel sets are naturally grouped together (i.e., “clustered”) in spectral space.

To find the ML estimate of the dof ν , as with the other parameters, a derivative (e.g., a “root”) of the likelihood function is located:

$$\frac{dL(\mathbf{X}; \hat{\Psi})}{d\nu} = \mathbf{0}. \quad (18)$$

Unlike the other parameters, however, expanding the left side of equation (18) results in a nonlinear expression. It turns out that $\nu_i^{(j+1)}$ is a solution of

$$-\psi\left(\frac{1}{2}\nu_i^{(j)}\right) + \log\left(\frac{1}{2}\nu_i^{(j)}\right) + 1 + \left[\frac{1}{N} \sum_{n=1}^N (\log u_n^{(j)} - u_n^{(j)})\right] + \psi\left(\frac{\nu_i^{(j)} + L}{2}\right) - \log\left(\frac{\nu_i^{(j)} + L}{2}\right) = 0 \quad (19)$$

where $u_n^{(j)}$ is

$$u_n^{(j)} = \frac{\nu_i^{(j)} + L}{\nu_i^{(j)} + [(\mathbf{x}_n - \hat{\boldsymbol{\mu}}_i^{(j)})\boldsymbol{\Gamma}^{-1}(\mathbf{x}_n - \hat{\boldsymbol{\mu}}_i^{(j)})^T]} \quad (20)$$

and ψ is the Digamma function in equation (19),

$$\psi(x) = \frac{d\Gamma(x)/dx}{\Gamma(x)}. \quad (21)$$

3.3.1.1 Solving for the Degrees of Freedom Numerically

While there are no references to the use of full, automatic t mixtures with unknown dof in the remote sensing literature, it has been noted in the statistics literature [26] that the convergence of EM can be slow for unknown ν . This is due to the need for a one-dimensional search in determining ν at each iteration. Unfortunately, there is no way around this, and numerical optimization of equation (19) is required to find a solution. Newton-Raphson and similar gradient or steepest-descent methods are often used to solve nonlinear equations numerically. However, Newton-Raphson can be very sensitive to its starting point. Given that, for the first few iterations especially (because of random seeding) the estimate of $\nu^{(j-1)}$ for each cluster is extremely inaccurate, it is highly likely that it will make a poor starting point for the search and an optimum may not be found.

Instead, the bisection method is employed, a fundamental approach that ensures that a solution will be found. An equation $f(w) = 0$, where w is a real continuous function, has

at least one root in the interval $[w_l, w_u]$ if $f(w_l)f(w_u) < 0$. Beginning with an interval that is large enough to ensure that it contains at least one solution, the binary search begins by halving the interval:

$$w_m = \frac{w_l + w_u}{2}. \quad (22)$$

The function is then evaluated at the boundaries to determine in which half interval the solution lies.

$$f(w_l)f(w_m) < 0 \rightarrow [w_l, w_m] \quad (23)$$

$$f(w_l)f(w_m) > 0 \rightarrow [w_m, w_u] \quad (24)$$

$$f(w_l)f(w_m) = 0 \rightarrow w_m \quad (25)$$

The boundaries are reset to the appropriate interval and the function is evaluated again, with the recursion ending when the product is exactly zero or when the interval is sufficiently small (i.e., $w_u - w_l < \epsilon$).

Overall, the EM algorithm stops when a maximum in the likelihood space is found Ψ^* . Once the complete data likelihood no longer increases after each iteration, the procedure terminates and the current parameter estimates for each cluster are recorded. Each cluster now represents a spectrally homogeneous class and is one of the K components of the final mixture model. The assignment of each pixel to one of the model components is actually determined by applying the *maximum a posteriori* (MAP) rule:

$$\max_i \{P_i(\mathbf{x}_n; \hat{\Psi}_i)\} \Rightarrow z_{i,n} = 1, \quad (26)$$

where \mathbf{z} is an indicator variable with $z_{i,n}$ set to 1 when pixel n belongs to cluster i , and 0 otherwise.

Our initial trials using this approach were published in [27].

3.3.2 Stochastic Expectation-Maximization

In some applications of the EM algorithm, the E step is complicated and does not yield a closed-form solution to the computation of conditional expectation of the complete data

(log-)likelihood. One way to get around this problem is to resort to numerical integration. However, in some situations, especially when the complete data density does not belong to the exponential family, numerical integration over the missing data density does not always preserve the function [28]. Thus, executing the E step by a Monte Carlo process may be a viable and attractive alternative. Such a method was introduced in [29]. An EM algorithm, where the E step is executed by Monte Carlo, is known as a Monte Carlo EM (MCEM) algorithm. It applies whether the ML or MAP estimate is being sought.

Even before the MCEM algorithm, others considered a modified version of the EM algorithm in the context of computing the ML estimate of parameters for finite-mixture models. It was called the Stochastic EM (SEM) algorithm [30] and it is the same as the MCEM algorithm with $M = 1$.

However, with the SEM algorithm, the current posterior probabilities are calculated using a Stochastic E step, wherein a single draw is made from the current conditional distribution of \mathbf{z} given the observed data \mathbf{x} . Because of the assumption of independence of the complete data observations, this is done by conducting a draw for each j ($j = 1, 2, \dots$). That is, a draw $\mathbf{z}_n^{(j)}$ is made from the multivariate distribution with the number of categories having probabilities specified by equation (14). This effectively assigns each observation outright to one of the components of the mixture. The M step then consists of finding the ML estimate of the parameter vector as if the observations were deterministically classified according to \mathbf{z} . This contrasts with the EM algorithm, where these computations are weighted with respect to *all* components of the mixture according to the current posterior probabilities. Note that with the SEM algorithm, there is only one Monte Carlo sample taken, so $M = 1$ always. This algorithm prevents the sequence from staying near an unstable stationary point of the likelihood function. It also then avoids the cases of slow convergence observed in some uses of the EM algorithm, such as in the mixture problem considered here.

As before, initializing $\Psi^{(0)}$ is necessary to seed the moments and mixture weights for

each component. For SEM this is uniformly random. But in this case, K is not fixed before initialization; instead, it is an initial guess that will be refined along with the estimation of weights and density parameters.

Now, $\Psi = [K, \pi_1, \pi_2, \dots, \pi_K, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_K, \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \dots, \boldsymbol{\Gamma}_K, \nu_1, \nu_2, \dots, \nu_K]$.

The first step in SEM stands for stochastic (or perhaps more appropriately segmentation), where equation (26) is used to assign cluster membership.

Next is the maximization step, which occurs in three parts. The cardinality of each cluster is computed by

$$\sum_{i=1}^K C_i^{(j)} = N \quad (27)$$

where $C_i^{(j)}$ is a count of the pixels in cluster i . This clustering is done in an attempt to encourage larger increases in equation (13), with the hope of moving out of a path toward a local maximum and onto a path to the global maximum.

The mean, covariance, and mixture weights are respectively computed using the cardinality of the current cluster rather than all pixels (where $\mathbf{x}_{n,i}$ is the n^{th} pixel of the i^{th} cluster)

$$\hat{\boldsymbol{\mu}}_i^{(j+1)} = \frac{1}{C_i^{(j)}} \sum_{n=1}^{C_i^{(j)}} \mathbf{x}_{n,i} \quad (28)$$

$$\hat{\boldsymbol{\Gamma}}_i^{(j+1)} = \frac{1}{C_i^{(j)}} \sum_{n=1}^{C_i^{(j)}} (\mathbf{x}_{n,i} - \hat{\boldsymbol{\mu}}_i^{(j+1)})(\mathbf{x}_{n,i} - \hat{\boldsymbol{\mu}}_i^{(j+1)})^T \quad (29)$$

$$\hat{\pi}_i^{(j+1)} = \frac{C_i^{(j)}}{N}. \quad (30)$$

The third part of the M step is when each component is checked to see if its contribution to the overall mixture is significant enough. Before initialization, a constraint is set that all mixture components must satisfy. At each iteration, the mixture weights $\hat{\pi}_i^{(j+1)}$ must be greater than a minimum ξ . If

$$\hat{\pi}_i^{(j+1)} < \xi, \quad (31)$$

the number of mixture components is decreased by one, $K = K - 1$. If only a tiny fraction of the scene is being represented by a given component, not only is it not statistically

significant, but it is not worth the computational burden of keeping the component in the model. Finally, the E step in SEM is computed the same as in equation (14).

3.4 *Finite Mixture Modeling Results for EM and SEM*

Our preliminary experiments using the SEM technique in comparison with EM were published in [31]. Figures 6 and 8 offer goodness-of-fit results from these experiments using EM and SEM, respectively. The value of ν exhibited by the data can be gleaned from reading the plot in figures 6 and 8, and this value should be compared with the value of ν_{est} shown in the legend. For example, in figure 6, the dark blue line in the plot is the actual value of ν for the data in cluster 7. This line is almost directly on top of the theoretical dashed black line for $\nu = 10$. Looking at the ν_{est} value in the legend of figure 6 shows an estimated dof value of 10.3. This means that our estimate using the EM technique is close to the actual value. For this data set, both parameter estimation techniques were reasonably accurate in estimating the appropriate value for ν and ultimately modeling the tails of the data.

The data was collected by the AVIRIS sensor at Camp Pendleton, a large U.S. Marine Corps base in California. Figure 5 shows an RGB image of the scene. Both models were initialized with $K = 9$ components. However, the result using the SEM technique had only $K = 6$ components. The associated cluster images are shown in Figures 7 and 9, where the assignment of pixels can be seen to be different for the SEM case of only six clusters.

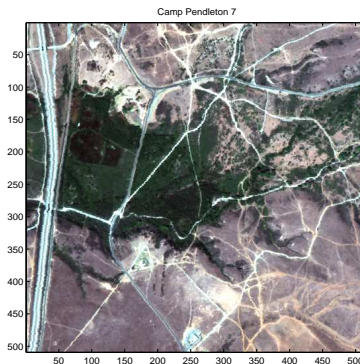


Figure 5: RGB image of Camp Pendleton, scene 7.

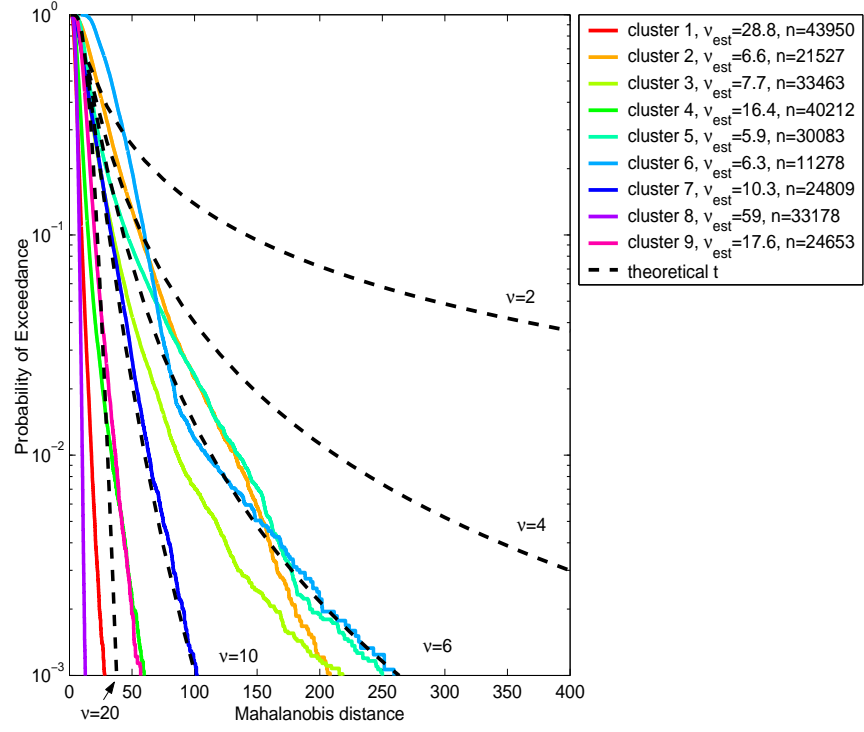


Figure 6: Exceedance metric for EC- t mixture model using EM technique.

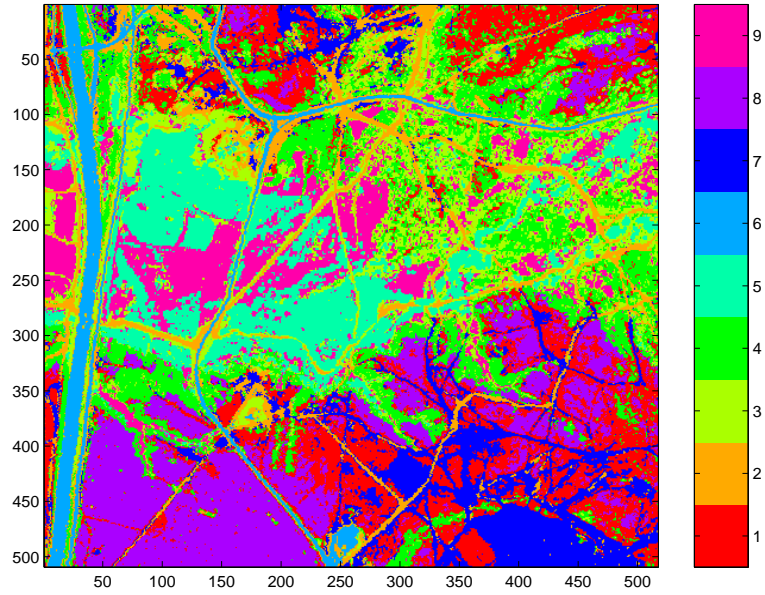


Figure 7: Cluster image generated from EC- t mixture model using EM technique.

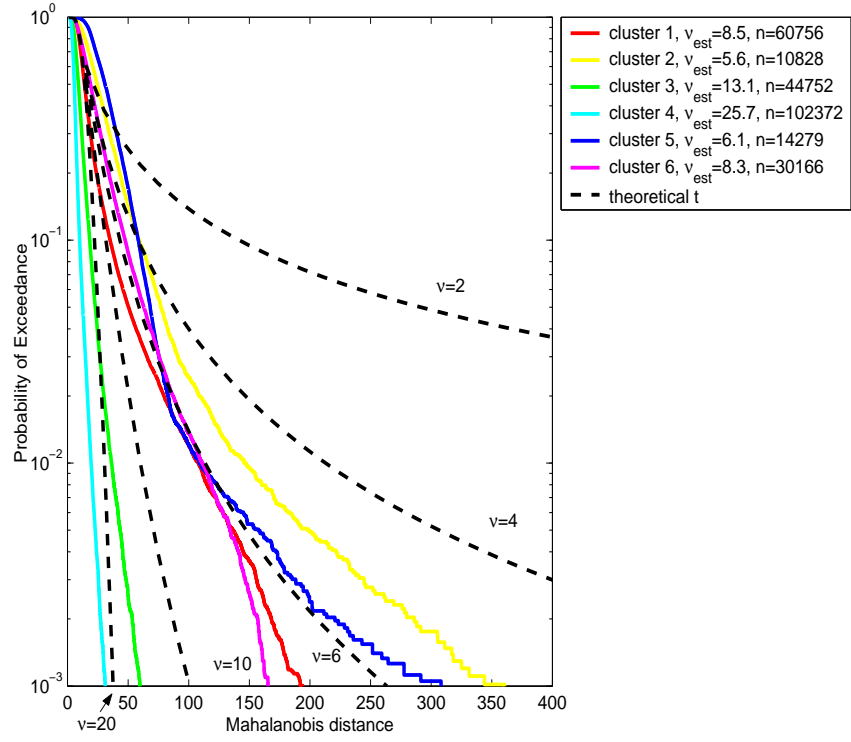


Figure 8: Exceedance metric for EC- t mixture model using SEM technique.

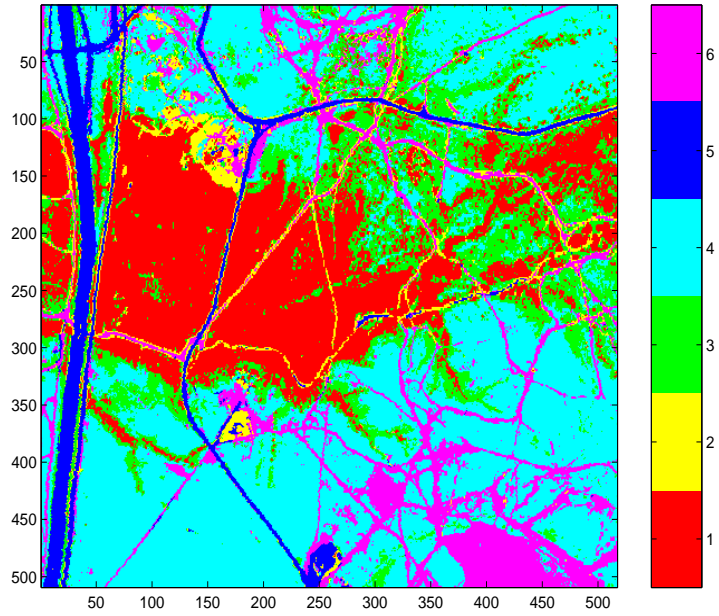


Figure 9: Cluster image generated from EC- t mixture model using SEM technique.

We published a more in-depth treatment of the entire problem in [32], including statistical analysis of the estimation process and use of a different goodness-of-fit test. In this work, AVIRIS data from Fort Hood, Texas was tested. RGB images of Fort Hood scenes 1 and 8 can be seen in figures 10 and 11, respectively.

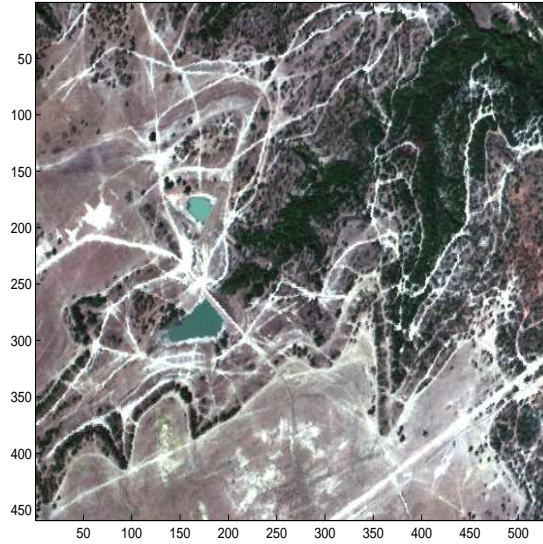


Figure 10: RGB image of Fort Hood, scene 1.

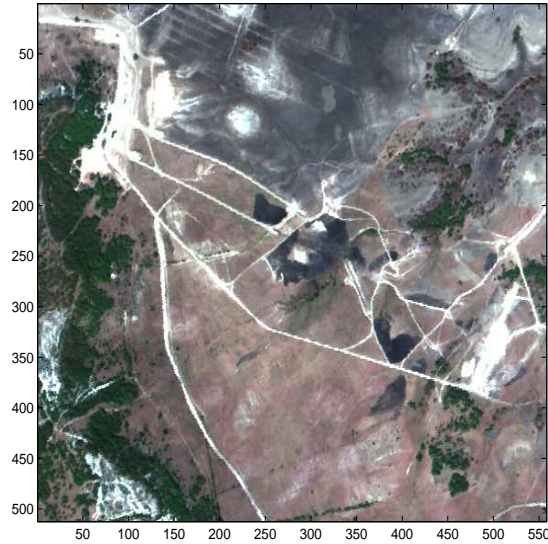


Figure 11: RGB image of Fort Hood, scene 8.

For Fort Hood scene 1, the results can be seen graphically in figures 12, 14(a), and 15. For Fort Hood scene 8, the results can be seen graphically in figures 13, 14(b), and 16.

These results show that using the K-S goodness-of-fit test, the EC- t did a reasonably good job at modeling the data. It is not perfect, however. Figures 12(a), 15(d), and 16(b) all show data with notable deviation from the model, despite using the EC- t distribution.

As such, our conclusion is that EC- t models offer a viable and tractable alternative for hyperspectral imagery. Such a modeling approach is well-advised for HSI data exhibiting heavy tails. The EC- t approach is not a panacea for all ills, however. The added complexity of deviating from the simpler Normal models should be considered, especially if the residual model error using the EC- t is only marginally smaller.

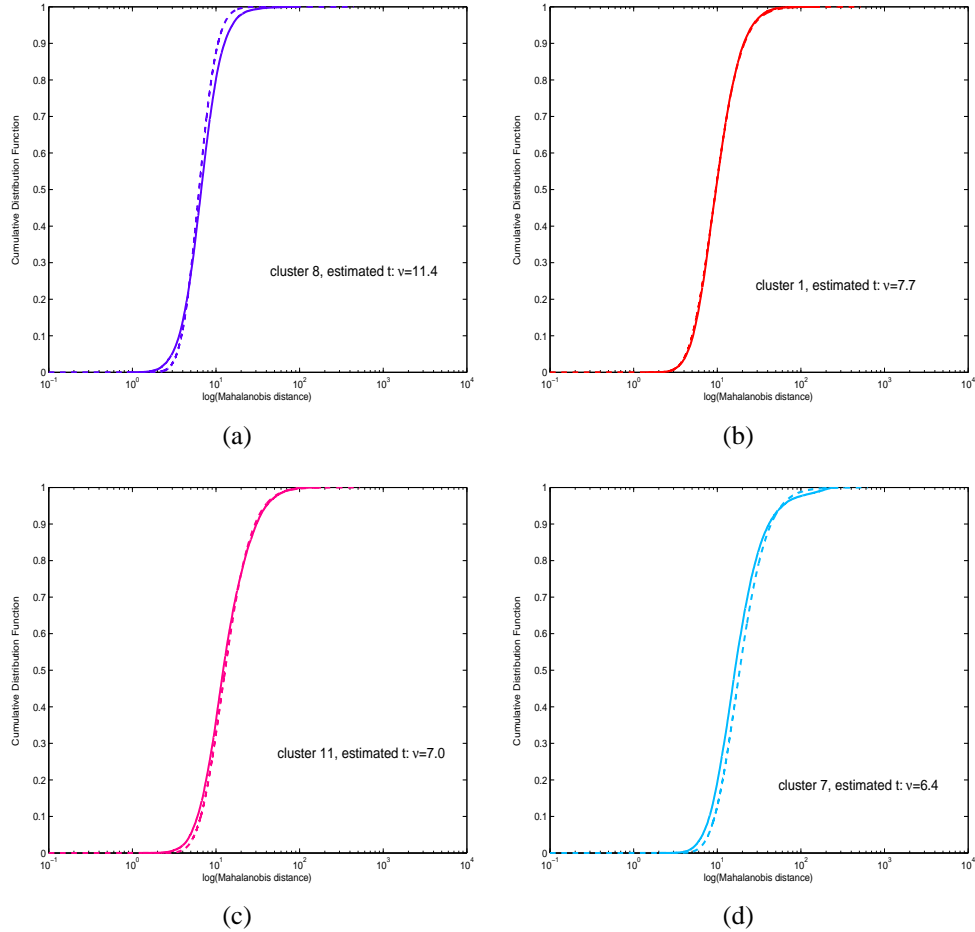


Figure 12: K-S test for Fort Hood 1 using EM: $L = 15$, $K = 11$.

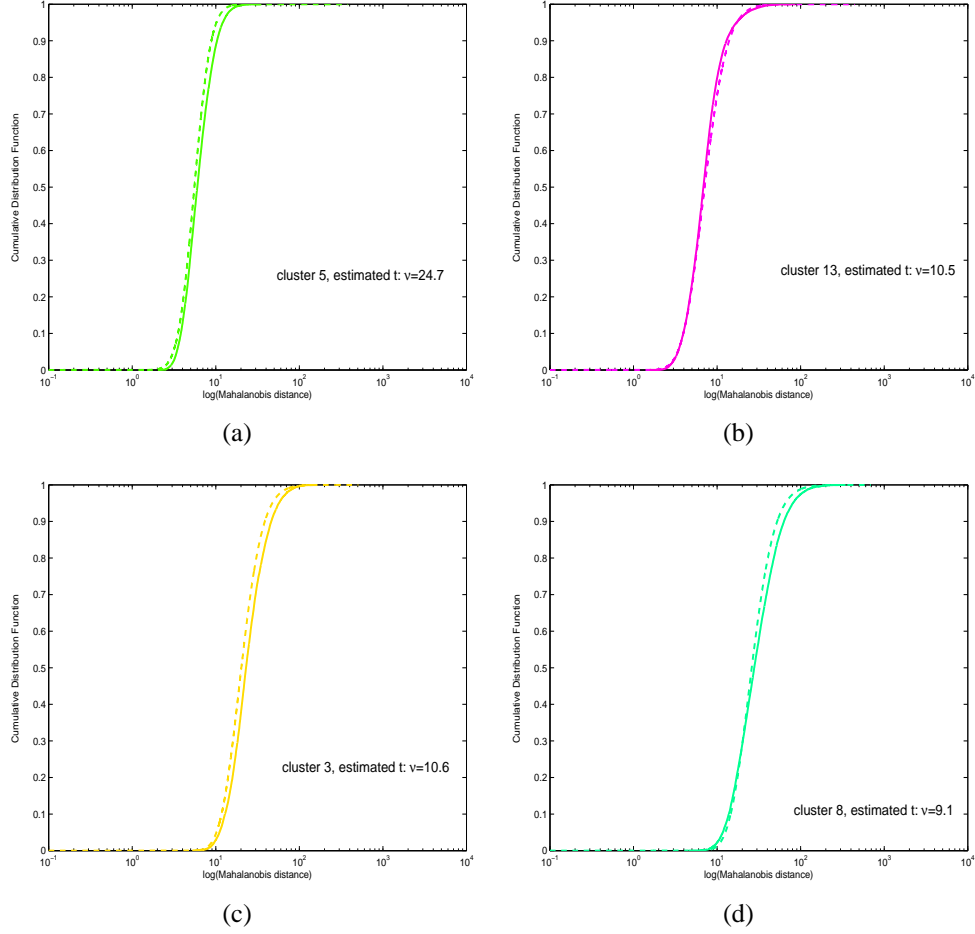


Figure 13: K-S test for Fort Hood 8 using EM: $L = 15$, $K = 14$.

On the topic of parameter estimation, there is more of a clear cut result. Table 4 synthesizes tables 2 and 3, and compares EM with SEM. Table 4 shows that EM requires almost an order of magnitude more iterations to achieve a fractionally larger improvement in $\log[L_c(\Psi)]$, equation (13). Further, EM must start and end with the same number of model components, forcing the user to be very confident in his model-order selection. SEM reduces the degree of certainty required in the initial number of model components. The only drawback with SEM is that it does not guarantee an absolute increase in data log-likelihood at each iteration. Still, it appears that SEM is clearly a fine parameter estimation choice for many HSI data scenarios.

Table 2: Selected analysis of EM results for Fort Hood data.

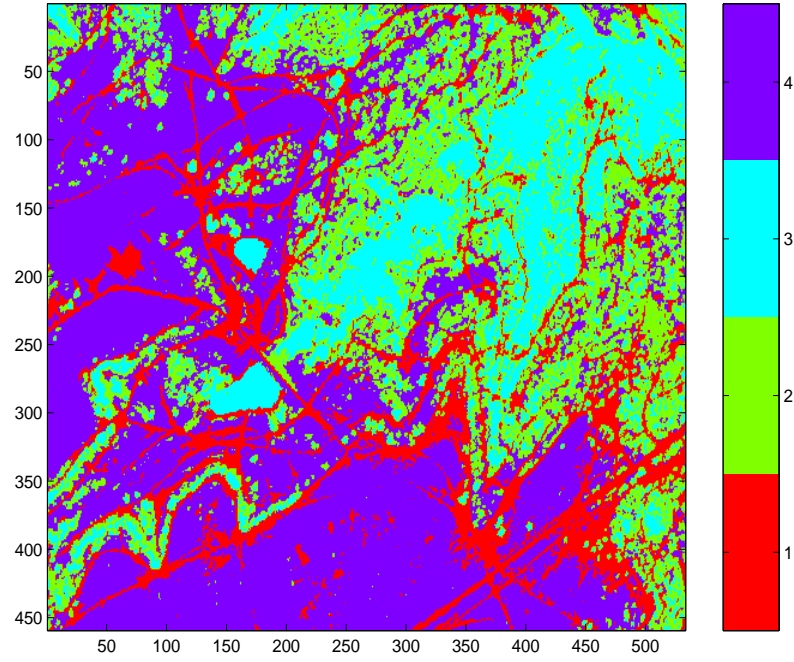
Scene	Fort Hood 1	Fort Hood 8
Log-likelihood: mean	20688878	23413725
Log-likelihood: std dev	3158	9259
Log-likelihood: maximum	20693077	23426014
Iterations: mean	864	590
Iterations: std dev	282	61
Iterations: at maximum $\log[L_c(\Psi)]$	913	535

Table 3: Selected analysis of SEM results for Fort Hood data.

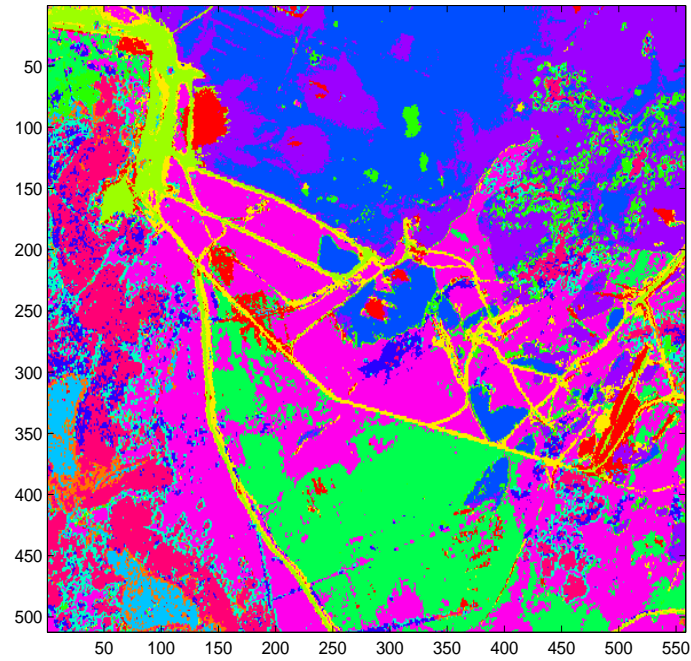
Scene	Fort Hood 1	Fort Hood 8
Log-likelihood: mean	20870657	23563042
Log-likelihood: std dev	159182	41594
Log-likelihood: maximum	21073599	23607418
Iterations: mean	132	143
Iterations: std dev	46.1	39
Iterations: at maximum $\log[L_c(\Psi)]$	88	103
Components: mean	4.8	9.3
Components: std dev	1.3	0.8
Components: at maximum $\log[L_c(\Psi)]$	7	10

Table 4: Comparison of EM and SEM results for Fort Hood data.

Scene	FH 1	FH 8
Max log-likelihood, EM > SEM in value	0.88%	0.64%
Max log-likelihood, EM > SEM in num. iterations	1037%	519%
Mean log-likelihood, EM > SEM in value	1.81%	0.77%
Mean log-likelihood, EM > SEM in num. iterations	655%	414%



(a)



(b)

Figure 14: SEM cluster images for Fort Hood 1 (a), $L = 15$, $K_{init} = 11$, $K_{end} = 4$, and Fort Hood 8 (b), $L = 15$, $K_{init} = 14$, $K_{end} = 13$.

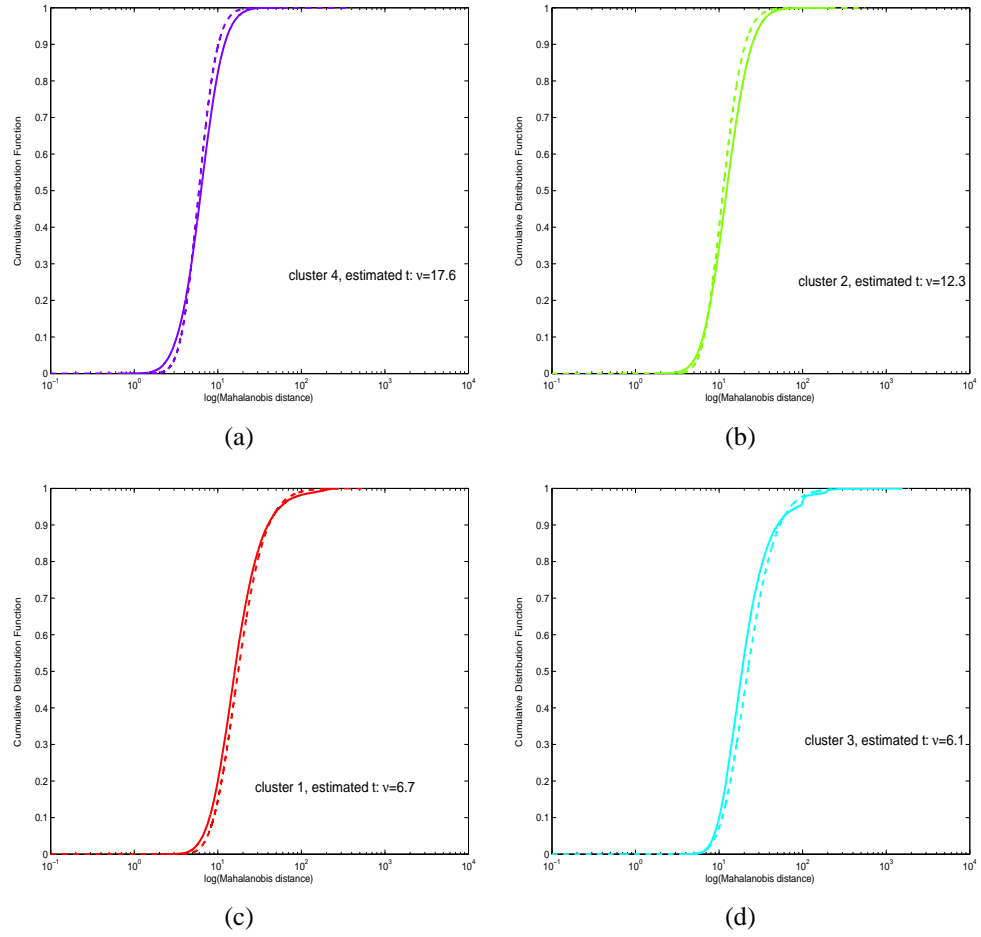


Figure 15: K-S test for Fort Hood 1 using SEM: $L = 15$, $K_{init} = 11$, $K_{end} = 4$, $\xi = 0.01$.

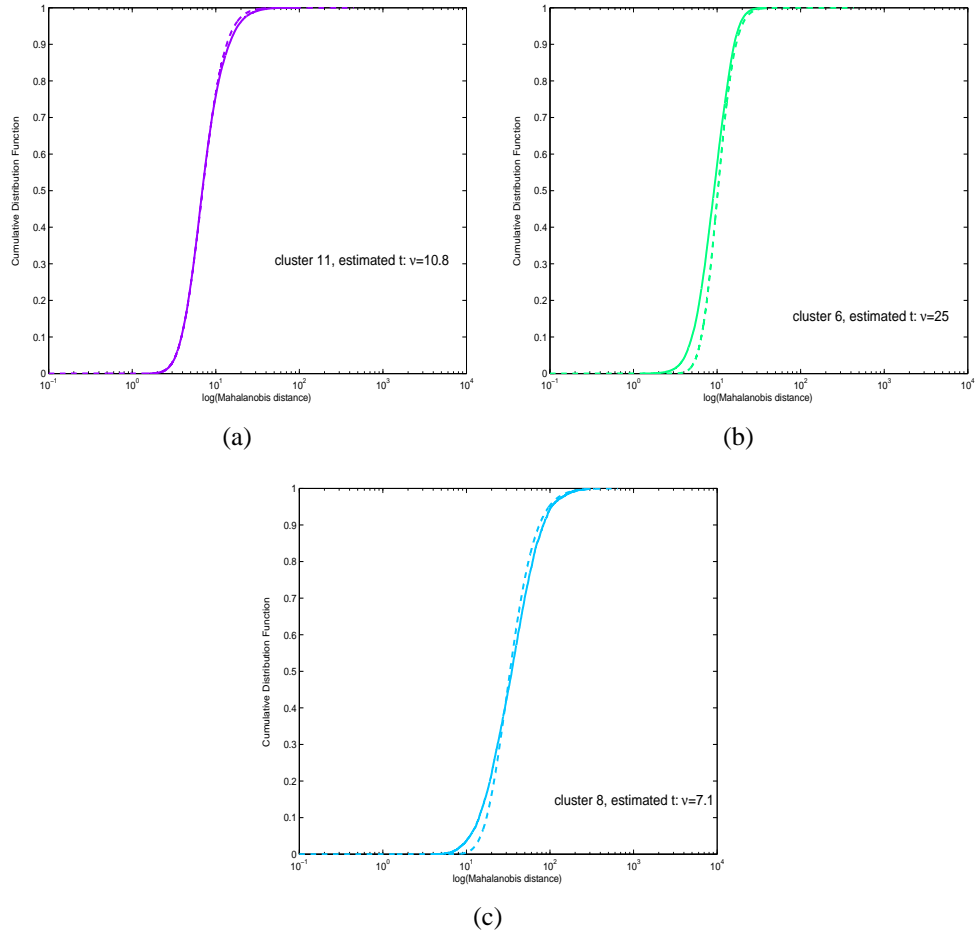


Figure 16: K-S test for Fort Hood 8 using SEM: $L = 15$, $K_{init} = 14$, $K_{end} = 13$, $\xi = 0.01$.

CHAPTER 4

A CONCISE DEVELOPMENT OF HSI DETECTION ALGORITHMS

Despite years of research into exploitation of hyperspectral imagery, the remote sensing literature lacks references on the comprehensive development of target detection algorithms. This chapter is a synthesis of items in detection theory textbooks, combined with conference and journal papers on individual detection schemes, and my own additions of theory and explanation to unify the discussion and keep it concise. The detection statistics presented in this chapter provide a foundation from which to conduct experiments and analyses on both synthetic and measured data.

Throughout this thesis, the focus is on detection algorithms that exploit only spectral information. It will become clear from the concepts presented in this chapter, and the experiments in later chapters, that it is necessary for the target spectrum to be distinguishable from the background spectrum in order for a detector to be effective. The degree of spectral contrast between target and background is a determining factor in the utility of hyperspectral imagery for target detection – along with fundamental factors such as SNR.

4.1 Detection Algorithm Design

As mentioned briefly in Section 2.2, the task of a detection algorithm is to decide if a signal of interest exists in a pixel under test, based solely on the observed spectrum vector \mathbf{x} . The optimum decision strategy is to maximize the probability of detection (P_D) while keeping the probability of false alarm (P_{FA}) under a fixed value. This is known as the Neyman-Pearson criterion and is embodied in the likelihood ratio test

$$\Lambda(\mathbf{x}) = \frac{f(\mathbf{x}|H_1 = \text{target signal present})}{f(\mathbf{x}|H_0 = \text{target signal absent})} \underset{H_1}{\overset{H_0}{\geq}} \eta, \quad (32)$$

where the probability of observing \mathbf{x} under the null hypothesis is $f(\mathbf{x}|H_0)$, and the probability of observing \mathbf{x} under the alternative hypothesis is $f(\mathbf{x}|H_1)$. The desired P_{FA} is achieved by setting the threshold η to appropriately include only a set amount of false alarms. Figure 17 illustrates this concept. Determining (32) requires knowledge of the conditional probabilities (pdfs), and these are estimated from the data. When this approach is taken, signal models are used that lead to the construction of practical (although suboptimal) detectors.

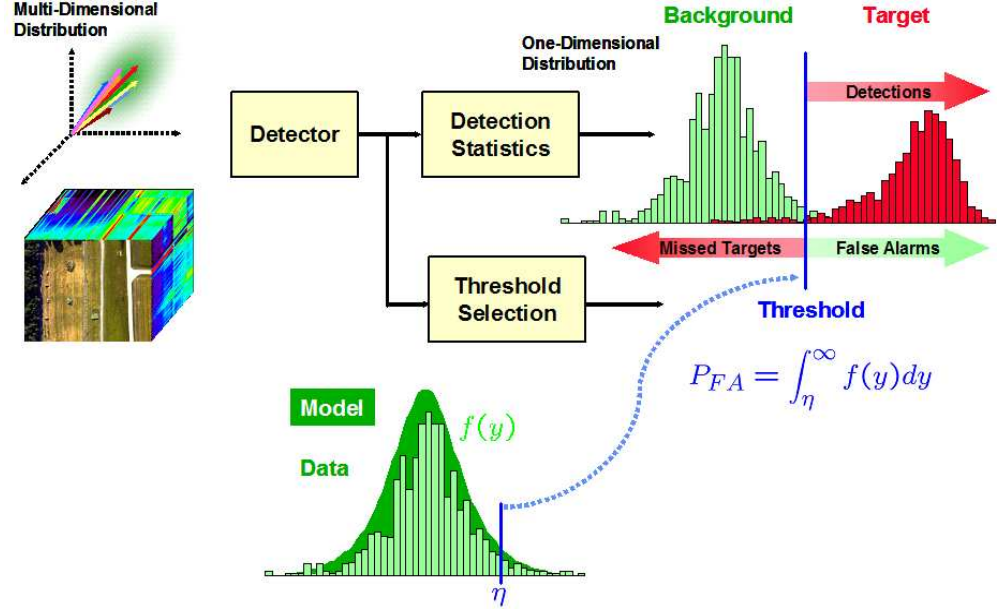


Figure 17: Illustration of detection and thresholding.

4.1.1 A Note on ROC Curves for HSI

Receiver operating characteristic (ROC) curves are typically employed in the evaluation of detector performance. ROC curves plot the P_D versus P_{FA} as a function of the threshold η . If assumptions are made about the pdfs under the two hypotheses, theoretical ROC curves can be generated. However, we are forced to use measured data to *estimate* the density functions. This makes performance evaluation of detection algorithms challenging due to the limitations imposed by a small amount of target data – typically less than 10^2

target pixels exist in a data set of 10^5 background pixels. An estimate of P_D values for a ROC curve using only 100 target pixels is not robust. As a result the establishment of accurate ROC curves on real data sets is quite difficult, and figure 18 highlights differences between theory and practice. Indeed, it is well known that as a rule of thumb the minimum number of N samples used to estimate a probability P should be at least $10/P$, or better yet $100/P$ [33]. Monte Carlo techniques are suitable for theoretical comparison, and such results are presented in Section 5.5.

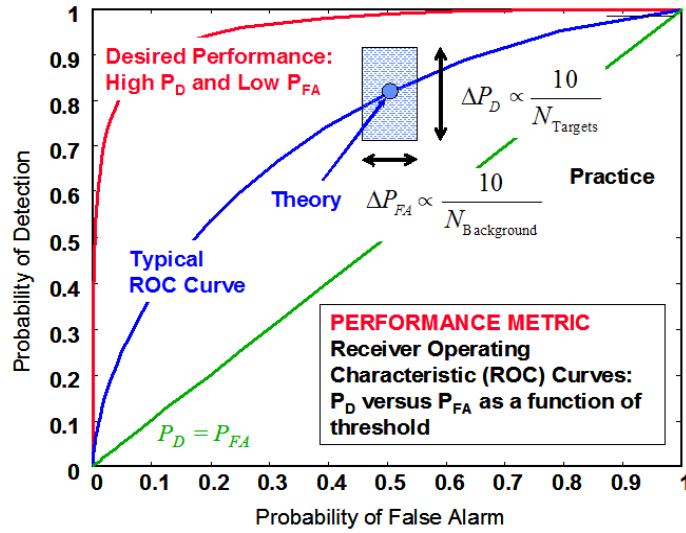


Figure 18: Receiver Operating Characteristic (ROC) elements.

4.2 Covariance-based Detectors

Unfortunately, as alluded to above, practical situations prevent the conditional densities from being known due to a lack of perfect a priori knowledge of the background and target signal parameters, θ . As such, a standard approach is to replace the unknown parameters with their ML estimates, $\hat{\theta}$. While there is no optimality associated with the Generalized Likelihood Ratio, it has proven widely effective in many applications

$$\Lambda_{GLR}(\mathbf{x}) = \frac{f(\mathbf{x}; \hat{\theta}_1 | H_1)}{f(\mathbf{x}; \hat{\theta}_0 | H_0)}. \quad (33)$$

Parametric signal models are needed to design target detectors based on the Generalized Likelihood Ratio (GLR) approach. For the case when we can safely assume that, aside from noise, the target fills the entire pixel (i.e., there are no other interfering signals in our measurement), algorithm performance is primarily a function of background and target variability. As noted above, the detection problem is formulated as a binary hypothesis test of target presence.

Here it is assumed both classes can be well-characterized by multivariate Normal distributions¹. Since the background and target are different physical materials they have different means and covariances. This leads to a non-linear decision boundary in spectral measurement space and a quadratic detector.

In the special case when the two classes can be described by a common covariance matrix (i.e., $\Gamma_0 = \Gamma_1 \equiv \Gamma$), the detection statistic becomes linear. In the signal processing (i.e., radar and communications) literature, this is known as the matched filter, whereas in the pattern classification literature this is called Fisher’s linear discriminant. For hyperspectral imaging, the case of equal covariance for target and background rarely happens.

When the target of interest is at most the same size as the spatial extent of a pixel, and possibly smaller, any remaining background that fills the pixel becomes an interfering signal. Figure 19 is a simple illustration of this notion. The nature of sub-pixel targets leads to a replacement signal model, and physically speaking we expect the target fraction of the pixel (α_t) and the background fraction of the pixel (α_b) to sum to one. Further, we expect physical conditions dictate that there is no negative fraction of either class, i.e., $\alpha_t, \alpha_b > 0$. However, enforcing the sum-to-unity and non-negativity constraints make algorithm development challenging, and most of the literature stays away from the fully-constrained approach.

¹Detectors using EC- t background models have exactly the same functional form, and only differ from their Normal counterparts in distribution of the output statistic. See [34] and [35] for details. As such, this treatment is relevant for all EC distributions, including the t and Normal.

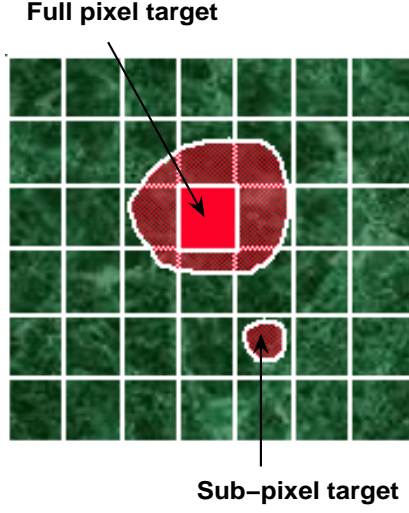


Figure 19: Illustration of a scene with full and sub-pixel targets.

In practice a set of simpler assumptions are used. First and most important is the existence of a linear mixing model,

$$\mathbf{x} = \alpha_t \mathbf{s} + \alpha_b \mathbf{b}, \quad (34)$$

where \mathbf{b} is a background spectrum and \mathbf{s} is a deterministic target spectrum – meaning it shows no variability in the *shape* of its spectral signature, only amplitude variability. If the target is not deterministic, we assume it lies in a linear subspace whose dimension is at most d (where $d \ll L$, the number of spectral bands). In the extreme case, the target signal is represented by the sum

$$\sum_{k=1}^d \alpha_{t(k)} \mathbf{s}_{(k)}. \quad (35)$$

Next we assume the background component of each pixel is randomly distributed according to a multidimensional Normal distribution of dimension d , i.e., $\mathbf{b} \sim N_d(\mathbf{0}, \mathbf{\Gamma})$. One reason this component of the signal model is assumed random is in order to account for *all* forms of interference, including sensor noise, etc. Finally, we do not enforce the unity and additivity constraints. This now yields an additive, rather than replacement, signal model

$$\begin{aligned} H_0 : \mathbf{x} &= \mathbf{b}, \\ H_1 : \mathbf{x} &= \mathbf{S}\mathbf{a} + \mathbf{b}, \end{aligned} \quad (36)$$

where the pixel is distributed under the two hypotheses as

$$H_0 : \mathbf{x} \sim N(\mathbf{0}, \mathbf{\Gamma})$$

$$H_1 : \mathbf{x} \sim N(\mathbf{S}\mathbf{a}, \mathbf{\Gamma}). \quad (37)$$

As touched upon earlier, one key difference between HSI and radar data is that it is real and non-negative. This means that HSI data will never be zero-mean, and prior to processing we must remove the estimated background mean from the entire data cube as well as the target signature.

In the mid 1980's, Kelly used the above linear model for radar data and applied the GLR approach in order to develop what is commonly referred to as the Generalized Likelihood Ratio Test (GLRT) [36], [37]

$$D_K(\mathbf{x}) = \frac{\mathbf{x}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{S}(\mathbf{S}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{S})^{-1} \mathbf{S}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{x}}{N + \mathbf{x}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{x}} \geq \eta_K. \quad (38)$$

We denote this detector with a subscript K to identify it as Kelly's algorithm, because in the literature there are now many detectors that employ the GLR approach. We will simply refer to it as the GLRT from here on.

In the early 1990's another algorithm surfaced called the Adaptive Matched Filter (AMF) [38]. It takes the form

$$D_{AMF}(\mathbf{x}) = \frac{\mathbf{x}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{S}(\mathbf{S}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{S})^{-1} \mathbf{S}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{x}}{N + \mathbf{s}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{s}} \geq \eta_{AMF}. \quad (39)$$

In the mid to late 1990's it was noted that a key aspect implicit in the structure of equations (38) and (39) is their assumption of equality for the background covariance matrix under the H_0 and H_1 hypotheses. Physically speaking, this would mean that despite different amounts of background being present in each pixel under H_0 and H_1 , the background covariance is still the same. Clearly, for sub-pixel targets this is not intuitive. Instead, the following hypotheses incorporate a minor adjustment to account for the difference in background fraction of each pixel:

$$H_0 : \mathbf{x} = \rho \mathbf{b},$$

$$H_1 : \mathbf{x} = \mathbf{S}\mathbf{a} + \rho\mathbf{b}. \quad (40)$$

Now the pixel distributions are as

$$H_0 : \mathbf{x} \sim N(\mathbf{0}, \rho^2\mathbf{\Gamma})$$

$$H_1 : \mathbf{x} \sim N(\mathbf{S}\mathbf{a}, \rho^2\mathbf{\Gamma}). \quad (41)$$

By making this adjustment we acknowledge that the training data used to generate the estimate $\hat{\mathbf{\Gamma}}$ has the same structure for its covariance matrix as the pixel under test \mathbf{x} , but the magnitudes of the variance values differ. The fraction of the pixel that is filled by the target has a direct bearing on ρ .

Employing the modified model in (40), the Adaptive Cosine Estimator (ACE) algorithm [39] has the form

$$D_{ACE}(\mathbf{x}) = \frac{\mathbf{x}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{S} (\mathbf{S}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{S})^{-1} \mathbf{S}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{x}}{\mathbf{x}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{x}} \geq \eta_{ACE}. \quad (42)$$

The number of pixels N no longer appears in the denominator as it did for Kelly's statistic in (38).

For each of these detectors, the theoretical signal-to-noise (SNR) ratio is

$$\text{SNR}_0 = (\mathbf{S}\mathbf{a})^T \hat{\mathbf{\Gamma}}^{-1} (\mathbf{S}\mathbf{a}). \quad (43)$$

We can rewrite these three detectors for the case where we have a deterministic target signature \mathbf{s} . Respectively for GLRT, AMF, and ACE, [40] they are

$$D_K(\mathbf{x}) = \frac{(\mathbf{s}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{x})^2}{(\mathbf{s}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{s})(1 + \frac{1}{N} \mathbf{x}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{x})} \geq \eta_K, \quad (44)$$

$$D_{AMF}(\mathbf{x}) = \frac{(\mathbf{s}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{x})^2}{(\mathbf{s}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{s})} \geq \eta_{AMF}, \quad (45)$$

$$D_{ACE}(\mathbf{x}) = \frac{(\mathbf{s}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{x})^2}{(\mathbf{s}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{s})(\mathbf{x}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{x})} \geq \eta_{ACE}, \quad (46)$$

Radar engineers will notice no magnitude signs in these expressions because HSI data are real, not complex valued.

When using a predefined spectral library or target database, often the algorithms using \mathbf{s} are utilized. When experimental data or new target info is being utilized, typically the versions with \mathbf{S} are used.

4.3 Subspace-based Detectors

As a contrast to these three covariance-based detectors, we also examine subspace-based algorithms.

The Orthogonal Subspace Projector (OSP) [41] was designed for HSI applications and is motivated by some basic concepts of multidimensional Euclidean geometry. The background variability is determined from the data and is modelled as a set of vectors that make up the matrix \mathbf{B} . Assuming the three coefficient vectors, \mathbf{a} , $\mathbf{a}_{b,0}$, and $\mathbf{a}_{b,1}$ are unknown constants, we differentiate between the background under the two hypotheses: $\mathbf{a}_{b,0}$ and $\mathbf{a}_{b,1}$. This leads to a decision structure that is

$$\begin{aligned} H_0 : \mathbf{x} &= \mathbf{B}\mathbf{a}_{b,0} + \mathbf{w}, \\ H_1 : \mathbf{x} &= \mathbf{S}\mathbf{a} + \mathbf{B}\mathbf{a}_{b,1} + \mathbf{w}, \end{aligned} \quad (47)$$

where \mathbf{w} is a random term for additive noise of unknown variance σ_w^2 , i.e., $\mathbf{w} \sim N(\mathbf{0}, \sigma_w^2 \mathbf{I})$.

The OSP algorithm can be thought of in two parts. First, the test pixel \mathbf{x} is projected onto the subspace orthogonal to the background by the operator $\mathbf{P}_B^\perp \mathbf{x}$. If we were to take the 2-norm of this quantity we would have the Euclidean distance from the test pixel to the background free subspace. This quantity is then multiplied by the target signature \mathbf{s} , making the OSP detector

$$D_{OSP}(\mathbf{x}) = \mathbf{s}^T \mathbf{P}_B^\perp \mathbf{x} \geq \eta_{OSP}. \quad (48)$$

The projection matrix onto the column space of the background \mathbf{B} is computed as

$$\mathbf{P}_B = \mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \quad (49)$$

and its orthogonal complement is

$$\mathbf{P}_B^\perp = \mathbf{I} - \mathbf{P}_B. \quad (50)$$

Those familiar with least squares theory will recognize $(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T$ as the pseudoinverse of \mathbf{B} . Also, equation (50) is sometimes referred to as a signal blocking matrix in the array processing literature.

It is also possible to obtain an expression for a subspace version of the GLRT [42]. This detector takes the same approach as in Section 4.2, but uses a geometric method rather than a probabilistic method to model spectral variability.

$$D(\mathbf{x})_{GLRTSB} = \frac{\mathbf{x}^T \mathbf{P}_B^\perp \mathbf{x}}{\mathbf{x}^T \mathbf{P}_{SB}^\perp \mathbf{x}} \geq \eta_{SB}. \quad (51)$$

The notation $[\mathbf{SB}]$ refers to a combined subspace of target plus background, which spans the space between our signal of interest and that which we believe to interfere with it. The geometric concepts of the subspace algorithms are depicted in figure 20 for three dimensions, specifically the subspace GLRT.

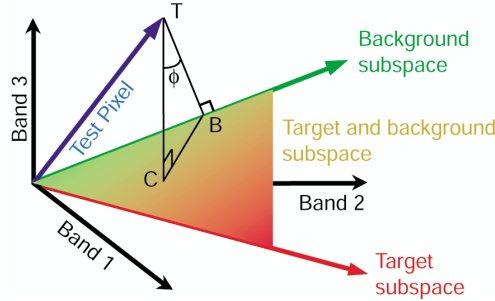


Figure 20: 3-d geometric interpretation of the subspace GLRT detector.

Some algorithm developers prefer to notice that the 2-norm, $\|\mathbf{P}_B^\perp \mathbf{x}\|$, is the Euclidean distance from the pixel under test to the background subspace – line segment TB in figure 20. Similarly, line segment TC in figure 20 is the distance from the test pixel to the combined subspace $\|\mathbf{P}_{SB}^\perp \mathbf{x}\|$. These terms appear in the numerator and denominator of equation (51), respectively. As such, the subspace GLRT can be rewritten in a 2-norm format as

$$D(\mathbf{x})_{GLRTSB} = \frac{\|\mathbf{P}_B^\perp \mathbf{x}\|^2}{\|\mathbf{P}_{SB}^\perp \mathbf{x}\|^2} \geq \eta_{SB}. \quad (52)$$

In some survey papers, the subspace GLRT is referred to as a “cosine” detector, however,

the *reciprocal* of equation (51) is actually a \cos^2 function. That is,

$$1/D(\mathbf{x})_{GLRTSB} = \cos^2 \phi. \quad (53)$$

There is a popular monotonic function of equation (51) that is known in the statistical literature as the F-statistic,

$$D(\mathbf{x})_{GLRTSB} = \frac{\mathbf{x}^T \mathbf{P}_B^\perp \mathbf{x} - \mathbf{x}^T \mathbf{P}_{SB}^\perp \mathbf{x}}{\mathbf{x}^T \mathbf{P}_{SB}^\perp \mathbf{x}}. \quad (54)$$

It is called the F-statistic because the detector output is distributed as a noncentral F-distribution, where the noncentrality parameter is given by SNR_0 . For the subspace GLRT this quantity is computed as

$$\text{SNR}_0 = \frac{(\mathbf{S}\mathbf{a})^T \mathbf{P}_B^\perp (\mathbf{S}\mathbf{a})}{\sigma_w^2} = \frac{\|\mathbf{P}_B^\perp (\mathbf{S}\mathbf{a})\|^2}{\sigma_w^2}. \quad (55)$$

What makes the subspace GLRT desirable is that it maximizes SNR for any distribution of \mathbf{w} and can operate in constant false alarm rate (CFAR) mode for Normal noise.

4.4 Anomaly Detection

Finally, it is not uncommon in operational scenarios for the target to be ill-defined. Sometimes there is simply not enough information about the target, or in the case of wide-area reconnaissance, there may not be a predefined target at all (just the desire to look for “objects of interest”). When this lack of a priori target info occurs, we go back to the linear matched filter approach. But instead of using the mean of the target class (μ_1) in our equation, we use the pixel under test, \mathbf{x} . That is,

$$D_{MF}(\mathbf{x}) = (\mu_1 - \mu_0)^T \Gamma^{-1} (\mathbf{x} - \mu_0) \geq \eta_{MF} \quad (56)$$

becomes

$$D_{AD}(\mathbf{x}) = (\mathbf{x} - \mu_0)^T \Gamma^{-1} (\mathbf{x} - \mu_0) \geq \eta_{AD}. \quad (57)$$

Those familiar with the statistical literature will recognize (57) as the Mahalanobis distance of the pixel under test to the background mean. Equation (57) is called the anomaly detector (AD) in HSI literature and is given a full treatment for hyperspectral imaging in [43].

It is worth noting that while anomaly detection might seem uninteresting from a research standpoint, it is a practical approach that can be used in an operational mission. This can be useful for operating in areas where spectral libraries are sparse or unreliable.

CHAPTER 5

EVALUATION OF ADAPTIVE DETECTORS FOR GROUND TARGETS

Now that a series of detection algorithms have been presented, attention is turned to evaluating which of these detectors are the best performers. In this thesis, the target of interest used for this benchmarking process is the land mine.

5.1 Introduction to Land Mines

Land mines are among the smallest and most difficult ground targets facing developers of imagery exploitation algorithms. Almost as ubiquitous as they are powerful, land mines remain an issue for today's military ground forces despite their introduction more than 60 years ago in the World Wars of the first half of the twentieth century. What's more, leftover mines and other unexploded ordinance (UXO) have also become a serious civilian problem. When these items are forgotten about or inadvertently moved (due mostly to natural phenomenon such as weather), they become a serious risk to unwitting civilians living in the area. The task of addressing this issue, known as 'humanitarian demining', has received increased attention in recent years [44], [45], [46].

A variety of electro-optical and radar sensors have been tested and evaluated for the detection of land mines, especially buried mines. In this chapter hyperspectral imaging sensors are considered for land mine detection. By looking at two different portions of the infrared spectrum – reflective (i.e., visible-SWIR) and emissive (i.e., LWIR) – for both buried and surface mines, we seek to evaluate the performance of previously developed detection algorithms. This study is novel in the sense that multiple test sites, multiple sensors, and multiple targets are used in comparing detector performance – an “apples to apples” comparison of algorithms is made for each target case.

5.2 Mine Detection Using Hyperspectral Imaging

In the summer of 1995, DARPA sponsored a series of experiments known as the Hyperspectral Mine Detection program. This activity is considered the impetus for the last ten years of research in mine detection using HSI sensors [47], which continues today. Two elementary, yet important, findings of these initial experiments are as follows:

1. Recent disturbances of the ground surface usually can be observed as a localized texture change in the surface, which can be detected by a broadband IR sensor. However, such single band approaches suffer severely from false alarms caused by vegetation and/or rocks.
2. The act of burying a mine will bring to the surface some subsurface material that can be seen as a spectral “scar.”

The first item is the motivation for using hyperspectral, rather than single band or multispectral sensors in the mine detection application. The second item confirms the idea that the presence of a localized difference between a land mine and its surroundings, caused by the mine itself or the emplacement of the mine, can be used as the key detection feature – a land mine target signature.

A surface-laid land mine produces a signature that is a direct result of the mine’s size, shape, composite material, and thermal properties. The background objects (i.e., clutter) such as rocks, grass, and dirt surrounding the mine have inherently different properties. When viewed in the thermal IR region these properties manifest themselves as an apparent temperature contrast. Specifically, it is difficult to make sweeping characterizations about the thermal contrast between a land mine and its background because mines come in a variety compositions, sizes, and shapes – metal vs. plastic, diameter of inches vs. feet, thin discs vs. cube shapes. It is also important to note that observed target signatures change with atmospheric (diurnal) conditions; time of day and location are important factors [48].

On the other hand, the signature of a buried land mine is not due as much to the features of the mine itself, but rather the impact the mine (and the process of its emplacement) has on the background. Target signatures for buried mines are the result of an apparent contrast between the temperature of the surface soil above the mine and the temperature of the surface soil surrounding the mine. This contrast comes from a complicated interplay of events, but it can be generalized by two effects. First is a surface effect of disturbing the soil directly above the mine during its burial, which changes the soil's density and lowers its conductivity. Second is the volume effect of the existence of a thermal mass (i.e., buried mine) in the soil. The volume of soil directly above the buried mine does not heat up and cool down at the same rate as the surrounding soil thanks to the presence of the mine's thermal mass. Naturally, this effect is greater at shallower depths and lesser as time passes from initial emplacement [49] (i.e., the distribution of soil in the vicinity of the mine becomes more consistent and soil properties even out).

The bottom line is that the phenomenology of target signatures for mines are quite different – visible/near IR/SWIR vs. LWIR sensors, and surface vs. buried mines. However, from an algorithm and signal processing standpoint it is enough to know that signatures can be developed throughout the infrared regime for various mine scenarios.

5.3 Experiments for HSI Mine Detection

Tests were made on a variety of measured data with actual targets so that a realistic comparison could be made. It is worth noting, however, that great strides have recently been made in synthetic data generation including high fidelity land mine scenes [50].

The first set of data used in our tests was collected by a sensor in the reflective regime, which we call Sensor X. Sensor X measures 256 fine spectral channels ranging from $0.4 - 2.35\mu m$, each nominally $8nm$ wide. This is a popular regime for many HSI sensors used in environmental remote sensing since the waveband covers the visible, near infrared (NIR), and short-wave infrared (SWIR). The data were collected in November 2002 at a test range

consisting of forest, grassy meadows, and dirt roads. A field of short grass and a dirt area are the two scenes on which we focus our tests.

The second data set was collected in April 2003 by the University of Hawaii's Airborne Hyperspectral Imager (AHI) [51], a long-wave infrared (LWIR) sensor that measures 256 bands in the range of $7.0 - 11.5\mu m$. An optional, but popular, pre-processing option for AHI data includes discarding fringe bands at the beginning and end of the spectral region (bands 1-10 and 211-256). The remaining 210 bands are binned by three to yield 70 final spectral channels, each about $50nm$ wide. This sort of pre-processing is done to improve the SNR and reduce the computational burden. As noted earlier, the phenomenology in the LWIR region is very different from the vis-SWIR region, which makes this AHI data very useful for trend comparison with results from Sensor X. In addition, the LWIR scene is a desert terrain that is very different from the other scenes while sharing the same mine types.

Two mine types were studied. Mine type 1 is a plastic-cased land mine that is square shaped and on the order of one foot on a side. Mine type 2 is a metal mine that is circular with roughly the same diameter. When buried, the mines were placed at a depth of a few inches. The ground sample distance (GSD, referring to the spatial extent covered by a pixel) of the two sensors are different, with Sensor X having a slightly better GSD in these data sets. It is fair to say that these targets occupy either slightly more or slightly less than a single pixel, depending on platform altitude. The AMF, GLRT, and ACE algorithms are all capable of detecting sub-pixel targets.

In processing the data, two straightforward thresholding techniques were used. After computing the detection statistic for all five algorithms, the first threshold applied was one that achieved a constant false alarm rate (CFAR). While not an optimal strategy such as Neyman-Pearson, CFAR operation has proven useful in radar and a number of other detection tasks. The second thresholding scheme guarantees 100% detection. This is possible because ground truth is available for each scene; that is, the location of each of the targets

in the image is known. By examining the detector output for all target pixels and finding the lowest value, we can ensure $P_D = 1$ by setting the threshold to this lowest target value. Counting the number of false alarms that result from this threshold choice allows us to calculate a false alarm rate.

5.4 Performance Comparison

Figures 21 - 24 show results from our tests. In each figure, part (a) plots the number of detected targets using a CFAR thresholding scheme. For all test cases, this value was set to 10^{-4} . The dashed line near the top of the chart indicates the number of possible detections (i.e., total mines of that type) in each scene. For example, in figure 21(a) there were 13 mines in the dirt area and 11 mines in the short grass area. Part (b) of each figure uses the 100% detection thresholding scheme. These charts can be read as the algorithm being able to achieve a false alarm rate of “10 to the ...” using the y-axis value. For example, in figure 21(b) the ACE algorithm is able to achieve a false alarm rate of 10^{-4} in the dirt area while still detecting all targets.

There are a number of different perspectives from which we can draw interesting conclusions from these results.

With regards to the first choice of threshold selection, CFAR at 10^{-4} , the anomaly detector failed to detect a single mine. The OSP algorithm also did poorly, detecting zero mines in five of eight CFAR test cases – and the other three cases had only a few detections. The AMF and GLRT algorithms performed similarly to each other in CFAR mode. At a level of 10^{-4} these two detectors found about half the targets in the scene. The ACE algorithm performed well in all CFAR tests, finding all mines in the scene in three of eight cases.

In the trials where threshold selection for 100% detection was used, the anomaly detector again had the worst performance. A false alarm rate of about 10^{-1} was seen when no signature information (anomaly detection) was used. The OSP compared a bit more

favorably with the other three algorithms in the 100% detection setting, with false alarm rates between 10^{-1} and 10^{-2} . The AMF, GLRT, and ACE algorithms had false alarm rates that went from 10^{-2} in the worst cases to better than 10^{-4} in the best cases.

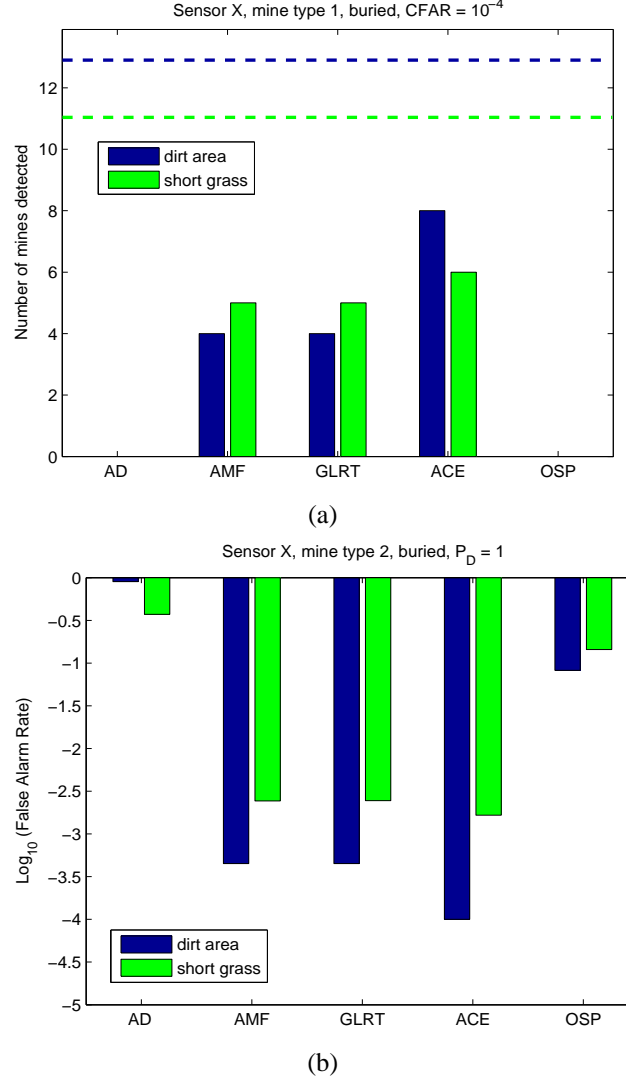
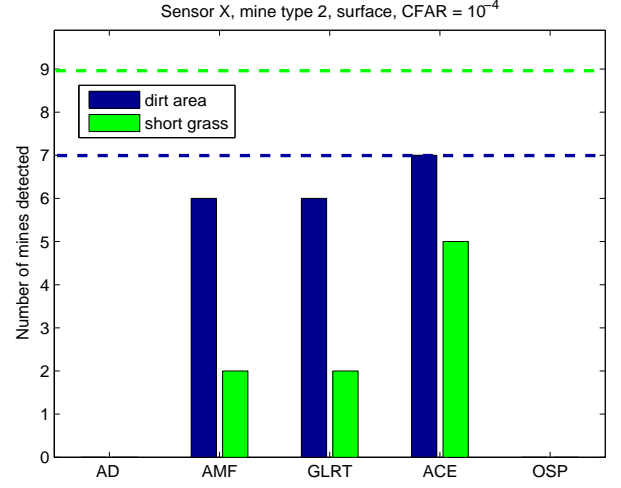
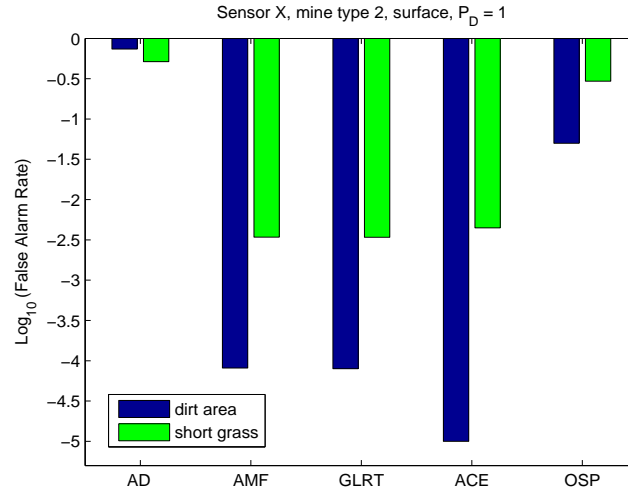


Figure 21: Vis-SWIR, buried mines: (a) mine type 1, constant false alarm rate detection; (b) mine type 2, 100% detection.

Upon further review of the covariance-based detectors, the reason for their performance similarity becomes clear. When N is large (i.e., many background pixels), the second term in the denominator of Kelly's algorithm (44) becomes negligible. That is, as $N \rightarrow \infty$, the term $\{(\frac{1}{N})\mathbf{x}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{x}\} \rightarrow 0$. In the case of many background pixels, which often happens in HSI detection, the GLRT devolves into the AMF. Conversely, when there are



(a)

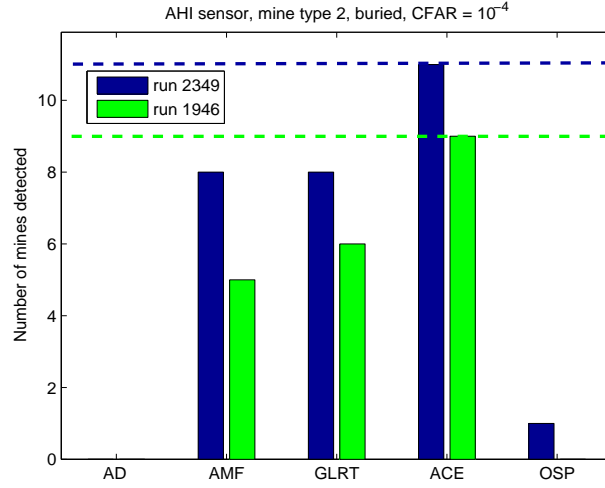


(b)

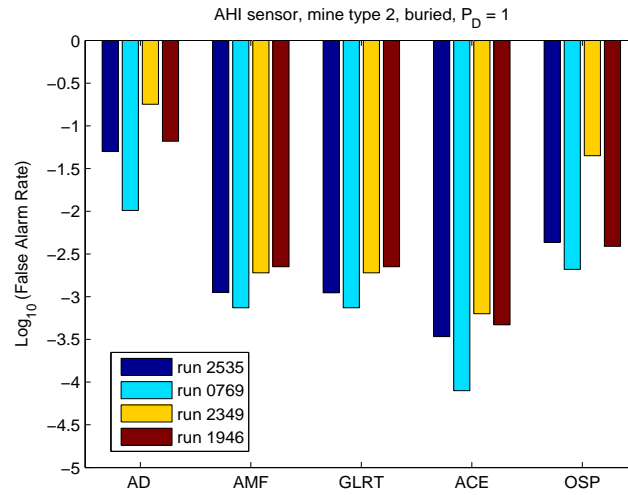
Figure 22: Vis-SWIR, surface mines: (a) mine type 2, constant false alarm rate detection; (b) mine type 2, 100% detection.

few background pixels, the normalization of the second denominator term by N becomes meaningless and the GLRT behaves like the ACE. What's more, the superior performance of ACE in most of our test cases is directly related to it's elegant property of scale invariance [39]. Simply put, this means that the training data and test data may be scaled differently without altering the detecting statistics. However, for the detection statistics of the AMF and GLRT to remain unchanged, the training and test data must be scaled identically.

When looking at the SWIR versus LWIR regimes, the results were mostly as expected. In the LWIR, the AHI sensor was able to detect more buried mines than surface mines. This



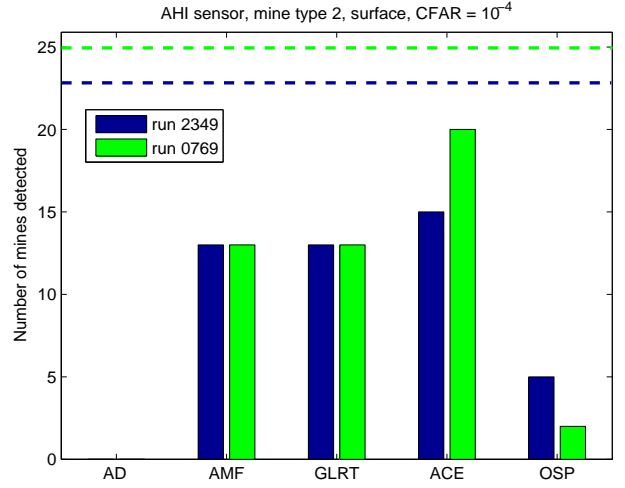
(a)



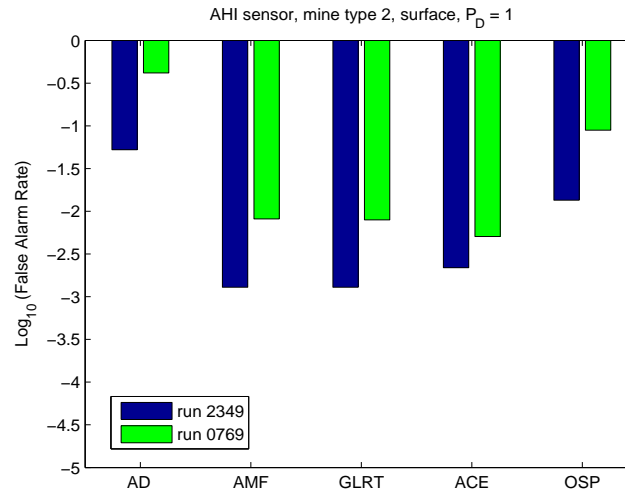
(b)

Figure 23: LWIR, buried mines: (a) mine type 2, constant false alarm rate detection; (b) mine type 2, 100% detection.

is due to the thermal nature of the spectral signature, which is not based on the mine itself, but rather on the effect it's emplacement and presence in the ground has on the surface. Also, the AMF, GLRT, and ACE algorithms in the LWIR region detected at least half the mines (both surface and buried) in all cases. On the other hand, Sensor X, operating in the visible-SWIR bands, performed better on surface mines. This is due to the fact that spectral characteristics of the mine's composite material are readily observed since the target is at least flush (if not slightly protruding from the ground) surface. It was somewhat surprising that Sensor X also did reasonably well in detecting buried mines.



(a)



(b)

Figure 24: LWIR, surface mines: (a) mine type 2, constant false alarm rate detection; (b) mine type 2, 100% detection.

The AHI sensor collected data at an arid test site, where each run was over essentially the same desert clutter scene. The primary difference between runs was the altitude at which the instrument was flown. In figure 23, the minor difference in performance in part (a) between runs is likely a function of platform altitude and thus ground sample distance. Run 2349 was made at 700 feet while run 1946 was made at 1400 feet. This change means a larger GSD (worse spatial resolution) for run 1946. Generally speaking this translates into fewer pixels on target – including the case where the mine is now a sub-pixel target. In figure 24 the altitude for both runs was 700 feet and the performance in part (a) was almost

the same.

The Sensor X data used in these tests was collected at a woodland test range, and two different sites were examined. The short grass area proved to be a tougher setting than the dirt area for both buried and surface mines. For both CFAR and 100% detection threshold schemes, most cases showed more mines detected in the dirt area.

5.5 Theoretical Performance

As a contrast to the results using real data presented earlier in this chapter, theoretical ROC curves were generated for the matched filter and Kelly's GLRT. Monte Carlo simulations were used assuming Normal statistics similar to those seen in many AVIRIS data sets.

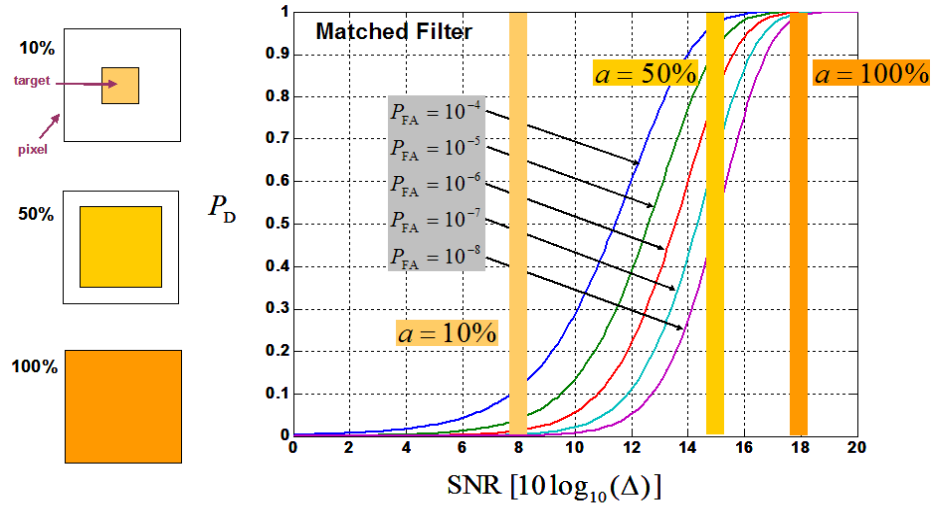


Figure 25: Theoretical ROC for matched filter detector on sub-pixel targets.

Figure 25 shows expected performance for subpixel targets using the matched filter. There are three levels of fractional fill for the target material, 100%, 50%, and 10%. As expected, there is a 3 dB difference between a pixel exactly filled with target and half-filled with target. The one-tenth pixel target yields a 10 dB lower SNR than a full pixel target. In order to achieve a $P_D = 0.5$ for a $P_{FA} = 10^{-6}$, the matched filter requires

13.5 dB SNR. This translates to a fill fraction of approximately 40%. It is fair to say that for sole source exploitation of hyperspectral imagery, subpixel targets are very difficult to detect. Realistically, data consumers should expect poor results for targets that fill less than half a pixel. This underscores the need for sensors with high resolution *both* spatially and spectrally.

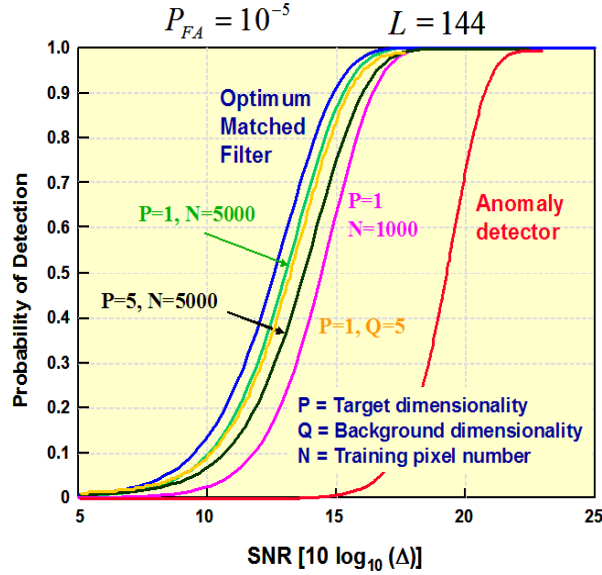


Figure 26: Theoretical ROC for GLRT detector on targets of varying dimensionality and with varying training support.

Figure 26 shows the expected performance for the GLRT under a variety of training conditions, assuming $L = 144$ measured spectral bands. The dimensionality of the target (going from $P = 1$ in the case of \mathbf{s} , to $P = 5$ in the case of \mathbf{S}) is varied, representing the increased degree of variability as P increases. The number of training samples also varies, highlighting the impact of the N term in the denominator of equation (38). For a target \mathbf{s} with only amplitude variability ($P = 1$), a 2 dB increase in SNR is required to achieve $P_D = 0.5$ at $P_{FA} = 10^{-5}$ when the number of training samples is reduced from 5000 to 1000. This is intuitive since a smaller number of samples used to calculate an estimate means it will be less accurate than an estimate of the same quantity that used a larger

number of samples. For the same number of 5000 training samples, a 1 dB increase in SNR is required to achieve $P_D = 0.5$ at $P_{FA} = 10^{-5}$ when the target dimensionality increases from $P = 1$ to $P = 5$. Note that the blue curve for the linear matched filter at the left of the figure represents a sort of upper bound for GLRT performance, and the red curve at the right represents a sort of lower bound (anomaly detection – no signature). An interesting point to note is that while training support has a significant impact on detection performance, HSI collection efforts rarely lack background samples. Instead, the more frequent condition is a lack of certainty in the target signature. The use of a span of vectors (\mathbf{S}) to characterize the target is common. The GLRT showed only a minor reduction in detection performance for a target with a 5-dimensional signal template versus a 1-dimensional signal.

CHAPTER 6

DETECTION AND IDENTIFICATION OF AIRBORNE GASES USING HSI

In addition to ground-focused hyperspectral exploitation tasks such as environmental assessment of ground cover or detection of military targets as seen in Chapter 5, there are other interesting applications. Atmospheric monitoring is a relatively new task for HSI sensors. While most atmospheric monitoring applications require long wave infrared (LWIR) hyperspectral imagery, many traditional HSI applications use data in the reflective regime (visible-SWIR).

HSI data analysis for gas plumes presents a set of problems different from those the remote sensing algorithm development community face with terrestrial data. Terrestrial objects and airborne gases differ in the way the underlying environment impacts the target observation. In the case of data with terrestrial targets, it is typically assumed that the intervening atmosphere (air between the sensor and ground) varies little throughout the scene, and that temperatures of observed objects vary slowly over the image. This is not the case for LWIR sensing of airborne industrial gas plumes. The radiance of the ground below the plume changes significantly over the scene, which impacts the observed composition of the plume in different pixels. Put another way, ground targets are *looked at*, whereas gas plumes are *looked through*. Much also depends on gas concentration and its optical thickness. This means spectral contrast (i.e., target detectability) depends not only on gas concentration, but also on composition of background material on the ground underneath the gas in a given pixel.

Identification or detection of hazardous gases using standoff sensors with a wide coverage area is a desirable capability. Often, the gases of interest are colorless, possibly odorless, have no set spatial structure once released, and may be released from a point that

is hard for humans to reach (i.e., high elevation, remote location, prohibited area, dangerous setting, etc.). Under these circumstances, gases make challenging targets to detect. Hyperspectral imaging sensors provide the ability for more precise identification of gas plumes than previously possible, and also give users the ability to determine spatial extent of release over a wide area. Conceptually, the consumers of HSI data in this context are many; including regulatory enforcement and homeland security officials.

The work in this chapter seeks to fill in gaps between other references on this topic and perhaps unify some key exploitation concepts. Messinger [52] uses synthetic data and applies only standard PCA and standard matched filtering to detect gas plumes. O'Donnell et. al [53] again use only synthetic data and apply a signature-based detection algorithm built on the principle of maximum distance between spectra in order to create linearly independent basis vectors for the necessary subspace (i.e., projection pursuit). In a report by Young [54], real data collected with the SEABASS sensor were used. However, in that work different gases are sought. Also, only standard PCA and standard matched filtering are applied.

6.1 Collection of LWIR Hyperspectral Data

Once again, the Airborne Hyperspectral Imager (AHI) was utilized for data collection. AHI was flown over central Texas in April 2004 as part of an EPA scientific investigation. The flight lines used in this work were collected over petrochemical and energy facilities, and we took three scenes from the data (labeled scenes A-C). Note that these images have not been roll-corrected. This was done deliberately to demonstrate that compensating for platform motion is not a prerequisite for non-literal HSI exploitation such as this case. RGB images from an on-board VNIR color linescan camera can be seen for the three scenes in figure 27. On collection dates of April 19-20, the sensor was flown in the afternoon local time under high scattered clouds, moderate winds, high humidity ($90 + \%$), and warm temperatures (avg. $82^{\circ}F$). From a platform altitude of ~ 2000 feet, the flight geometry as a function of AHI's instantaneous field of view (IFOV) yielded a pixel size that is

asymmetric. In the along-track direction, ground sample distance (GSD) was ~ 1 feet, and cross-track GSD was ~ 4 feet.

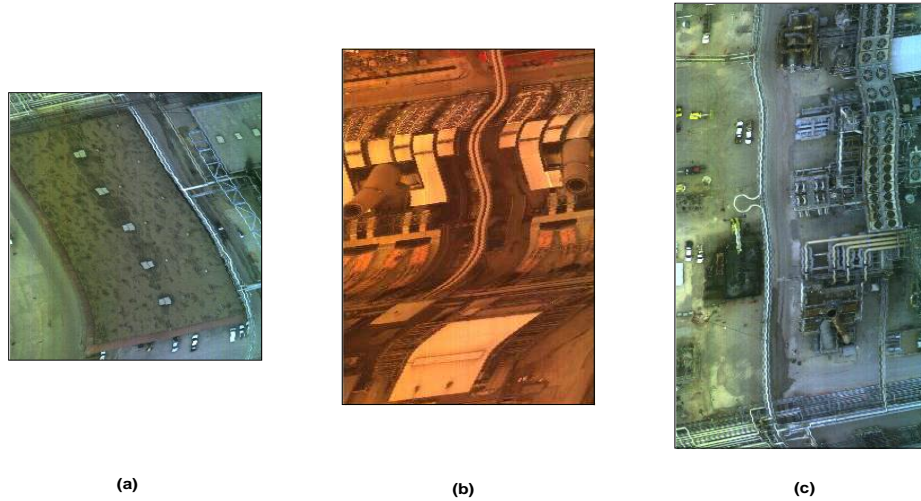


Figure 27: VNIR linescanner RGB images for three AHI scenes: (a) scene A (b) scene B (c) scene C.

6.2 *A Note on Signatures for Gas Plumes*

Inherent in this discussion is the notion that we can define a target “signature”, a spectral response that uniquely represents our material of interest. Unlike ground targets whose signature depends solely on reflectance in the SWIR, gas plumes can be identified by either emission or absorption features in the LWIR. At the emission source, a noticeable feature comes from the hot gas hitting the outside air. Further away from the emission source, a key signature feature can come from the plume cooling to meet ambient air temperature.

Laboratory spectra are often used as references for signature-based detection algorithms, and data collected in the field must be compensated for the modulation of the intervening atmosphere between the sensor and the scene. While this is a very important aspect

of hyperspectral data processing, we assume in this thesis that all data has been properly calibrated at the sensor and that a robust compensation approach has been applied. For the AHI sensor a discussion of calibration can be found in [55]. Examples of atmospheric compensation routines for the LWIR include MODTRAN [56] and ISAC [57].

6.3 Signature-based Gas Detection

As a representative of the set of detectors that use a target signal template, the subspace GLRT was applied for two different effluent compounds. The subspace GLRT detection approach differs from the approach most often taken in the LWIR gas detection literature – matched filtering. The matched filter is both simple and straightforward, and is even optimal under the right circumstances [33]. However, its application must be made with care for hyperspectral detection. Recall that the linear matched filter makes two key assumptions. First is that the variability present in the target and background classes are exactly equal, meaning they share a common covariance matrix. Second, the target signal must be deterministic. Practical conditions prevent both of these things from being true in almost all HSI operating scenarios, which is one reason why subspace-based algorithms such as equation (51) are often a better choice. Schaum revisits matched filtering for HSI in [58] and discusses some of these points in detail for hyperspectral target detection.

The subspace GLRT detector was applied to scene C using a target signature for benzene. From the RGB image in figure 27(c), it is clear this is a section of a petrochemical facility. Figure 28(a) shows the results from the subspace GLRT and part (b) of that same figure shows the anomaly detector result. As indicated in figure 28(c), red indicates a high value of the detection statistic (target present), and white indicates a low value (target absent). Clearly, the anomaly detector (AD) in figure 28(b) found many pixels of interest, however, these are very likely to be false alarms since they are an exhaust stack and cooling fans. The pipe junctions detected in figure 28(a) are much more likely to be fugitive emission sources, since this facility does pipe benzene between areas of processing and

storage.

The same two algorithms were run for scene A using an ethylene glycol target signature. Again, the subspace GLRT outperformed the AD. Figure 29(b) shows the anomaly detector had a much higher average value throughout the scene, as indicated by the greenish and blueish pixels that dominate the image. Figure 29(a), on the other hand, shows the subspace GLRT finding all five of the vents in the holding area as well as the flame tower in the upper-right corner. Not only do these detection results agree with gases identified in an initial screening of this data [59], but both of these gases are known to be present at the site under study [60] – either as recognized leaks or noted fugitive emissions. While this is not definitive ground truth, these results are certainly plausible and in agreement with visual analysis of the scene.

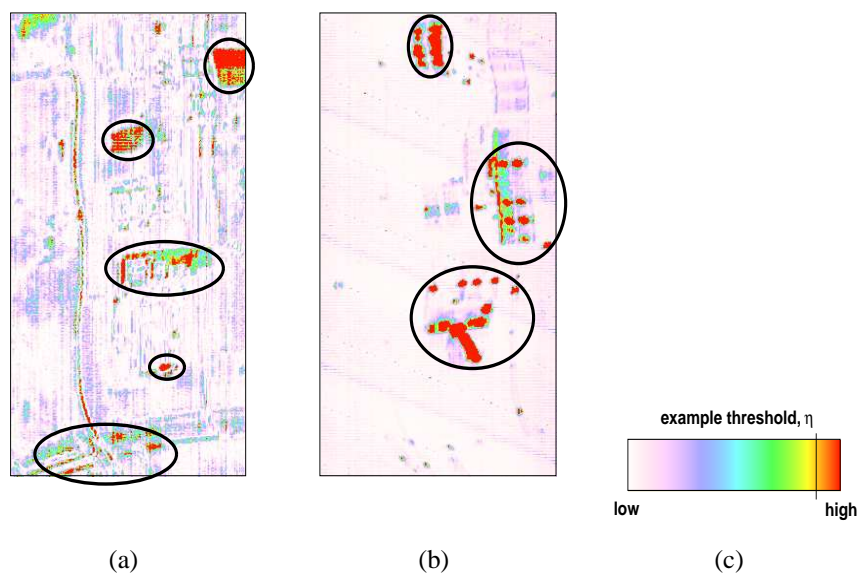


Figure 28: Subspace GLRT vs. Anomaly Detector, benzene, scene C: (a) Subspace GLRT, (b) Anomaly Detector, (c) color scale.

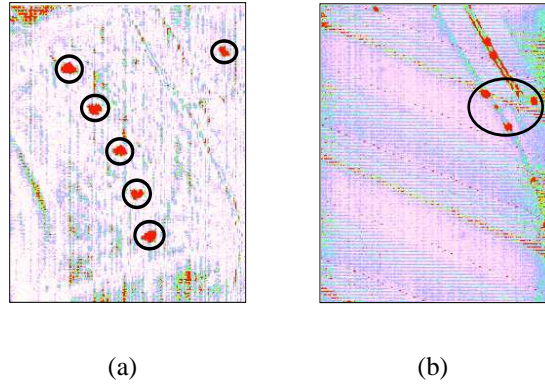


Figure 29: Subspace GLRT vs. Anomaly Detector, ethylene glycol, scene A: (a) Subspace GLRT (b) Anomaly Detector.

6.4 *HSI Gas Exploitation Without Signatures*

Target detection implies there is a defined material of interest that is being sought. However, there are other methods by which a user can search for materials of interest in a hyperspectral image cube. On the whole these techniques are more supervised and less automated in nature than signature-based or even anomaly detection algorithms. As such, more iteration and hands-on visual analysis are typically required for accurate interpretation.

6.4.1 Principal Components Analysis

The Karhunen-Loeve transform (KLT) is a well-known technique from signal processing used to obtain a new decorrelated signal that retains as much energy from the original signal as possible while only using a few components. The KLT is defined for continuous signals and its discrete-time counterpart is the principal components transform, known in the remote sensing literature as principal components analysis (PCA) [61]. The transform identifies orthogonal axes by way of an eigendecomposition of the data covariance matrix. The magnitude of the resulting ordered eigenvalues indicates the variability (energy) residing in the data along the component parallel to the corresponding eigenvector [62]. This becomes useful in dimension reduction for remotely sensed data because selecting only a few of the first basis components from the transformed space means the user can operate on

a greatly reduced data set that retains most all of the energy in the original data. Of course, this is in a statistical sense since PCA takes a wholly statistical interpretation of the data. The data are pre-multiplied by the selected PC vectors and the transform rotates the data into a new spectral space. This is done to optimize a squared-error criterion of $(\mathbf{x} - \boldsymbol{\mu}_0)$. Standard PCA assumes no noise in the signal model and uses all data in the scene.

A different flavor of PC transform was published in [63]. In the maximum noise fraction (MNF) version of PCA, additive observation noise is assumed and as such requires an estimate of the noise covariance. The criterion used for optimization in MNF is signal-to-noise ratio (SNR), and the basis vectors selected are the left-hand eigenvectors of the noise covariance \times data covariance product. That is, the bases are taken from

$$\hat{\mathbf{\Gamma}}_w \hat{\mathbf{\Gamma}}_0 \quad (58)$$

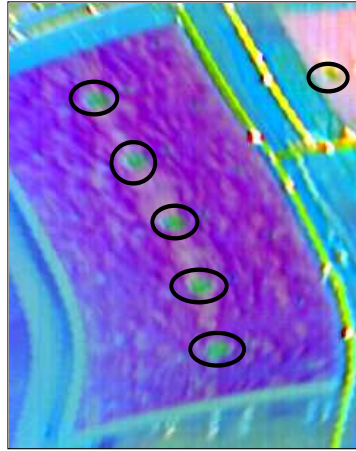
where the selected eigenvectors optimize the ratio

$$\frac{\mathbf{x}^T \hat{\mathbf{\Gamma}}_w \mathbf{x}}{\mathbf{x}^T \hat{\mathbf{\Gamma}}_0 \mathbf{x}}, \quad (59)$$

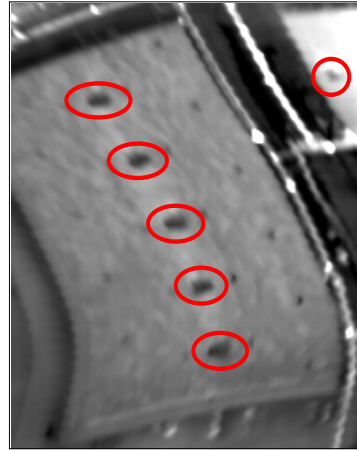
where $\hat{\mathbf{\Gamma}}_w$ is the estimated noise covariance matrix and $\hat{\mathbf{\Gamma}}_0$ is the estimated data covariance. It is worth noting that the resulting component vectors do not form an orthogonal basis as is the case for PCA. However, in the case of MNF, the components are ordered in terms of decreasing SNR. Even though the MNF formulation was realized by a different approach, it yields a mathematically equivalent result as noise-adjusted principal components (NAPC) [64].

PCA was applied to scene A in an attempt to identify the small ethylene glycol concentrations. Figure 30(a) shows an RGB composite of principal components 1, 2, and 3, respectively. Parts (b) - (d) of the same figure show the same 3 components individually in grayscale. Figure 30(a) seems to display possible ethylene glycol locations in a dark green, and parts (b) and (d) similarly show good contrast for these same locations. When MNF was applied to scene A in figure 31, the resulting RGB composite of the first 3 components in part (a) shows some of the same contrast for the vents, but not for the flame tower. Only

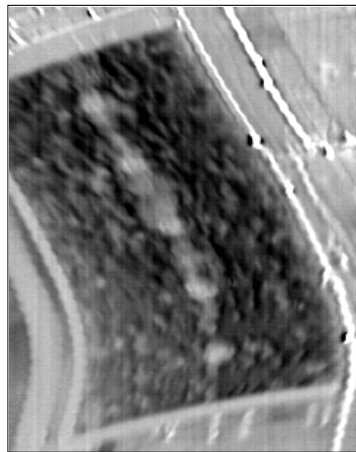
the first MNF component shows any contrast on an individual basis. Figures 32 and 33, respectively, show results of PCA and MNF applied to scene B. Plume-like structures are weakly visible in the RGB images of figures 32 - 33(a).



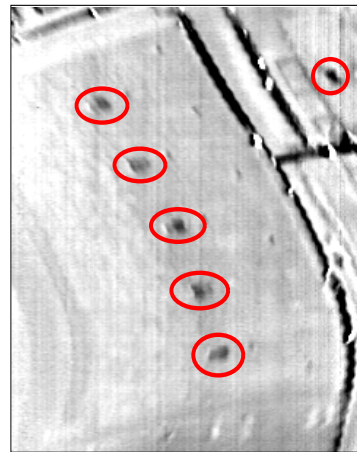
(a)



(b)



(c)



(d)

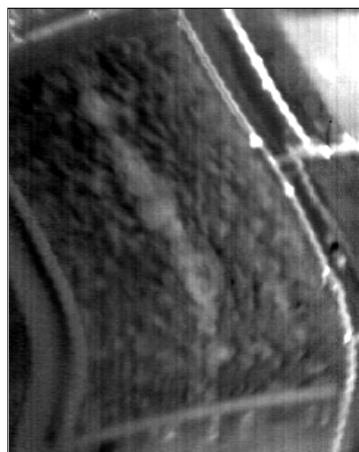
Figure 30: PCA images, scene A: (a) RGB of all 3 components (b) component 1 (c) component 2 (d) component 3.



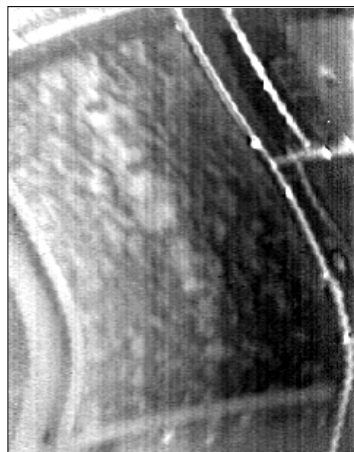
(a)



(b)

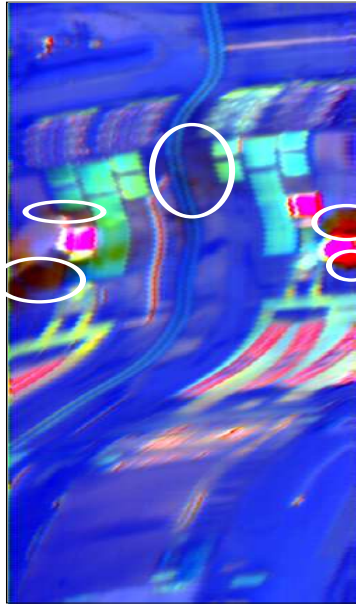


(c)



(d)

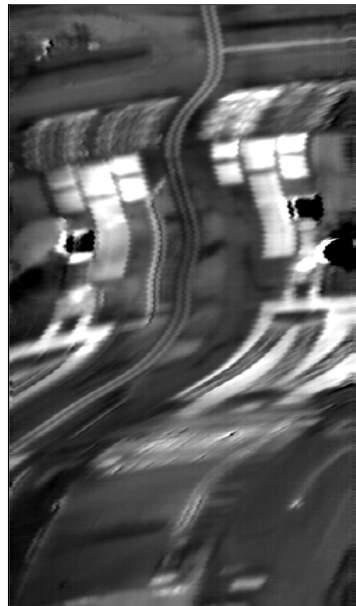
Figure 31: MNF images, scene A: (a) RGB of all 3 components (b) component 1 (c) component 2 (d) component 3.



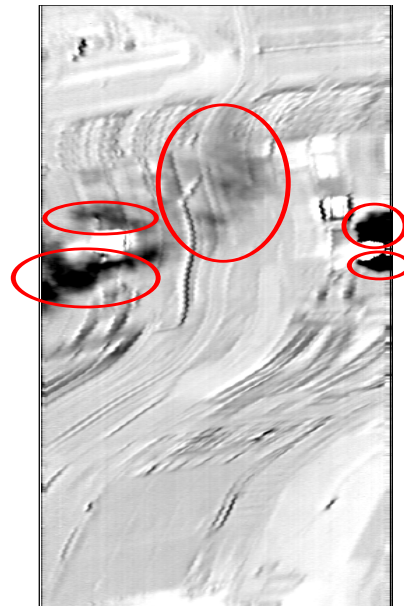
(a)



(b)



(c)



(d)

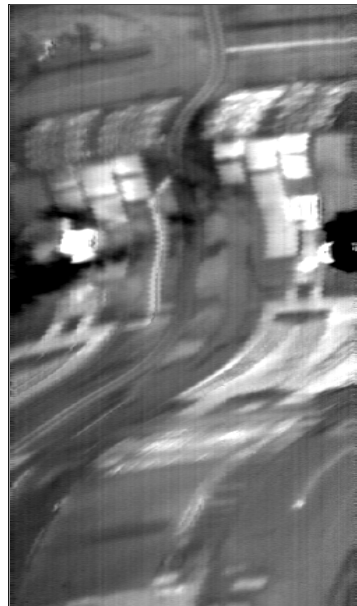
Figure 32: PCA images, scene B: (a) RGB of all 3 components (b) component 1 (c) component 2 (d) component 3.



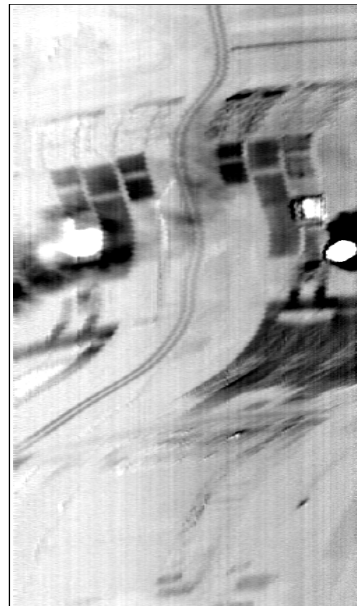
(a)



(b)



(c)



(d)

Figure 33: MNF images, scene B: (a) RGB of all 3 components (b) component 1 (c) component 2 (d) component 3.

6.4.2 Visualization Using 3-Band Composites and Scatter Plots

Data collected outside the visible spectrum is difficult to present in color format since our eyes are accustomed to the natural reflectance properties of everyday materials (i.e., we expect dirt to be brown and grass to be green). Since there is no “right” or “wrong” way to choose color combinations for multi-band IR imagery, it is intuitive to seek color combinations that highlight specific spectral features of our material of interest and make the target visually distinguishable. This approach has been taken for volcanic sulfur dioxide (SO_2) plumes in [65], choosing a combination that makes SO_2 gas appear yellow. Judiciously selecting the red, green, and blue color planes to show bands that have distinct emission/absorption lines can make trace gases stand out in a bright color.

Once a favorite 3-band RGB combination has been found, another way to determine the number of pixels containing a material of interest is to generate a 3-d scatter plot. By counting the number of samples that exist in a certain region of the RGB colorspace (a box in three dimensions), a user can quantitatively assess the spatial extent of a gas plume. Also, a qualitative evaluation can be made as to how pure these target pixels are: the brighter a pixel is in the RGB composite, the further it will be towards an extreme in the 3-d RGB colorcube.

Figure 34(b) shows the fourth principal component for scene B in which the two suspected SO_2 plumes are clearly visible near the top of the stacks. Figure 34(a) shows an RGB composite using bands $R=89$, $B=175$, and $G=55$ for scene B. This was done in an attempt to highlight the $8.6\mu\text{m}$ SO_2 absorption feature in the color pink. While the areas of pink might be difficult to see in the composite image of figure 34(a), the pink pixels are clearly visible in a 3-d scatter plot of the image. Figure 35(a) begins by illustrating the RGB colorspace, which is a cube in three dimensions, showing where pixels of a certain color are located. Parts (b) - (d) of figure 35 offer different angles of the 3-d scatter plot and allow the user to easily see the number of bright pink pixels in the image. Note the tight cluster of pixels along the edge of the box in parts (b) and (d) of figure 35; this indicates an

optically thicker, more consistent concentration of SO_2 gas. Further, it is easy to notice the pixels that are darker pink and lie closer to the main cluster of dark pixels which represent the background. Such pixels are likely to be part of a gas plume that is optically thin, which means more of the terrestrial background was captured in the spectral measurement.

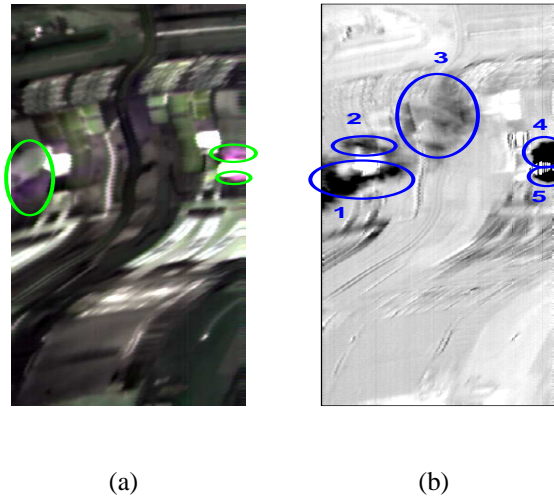


Figure 34: Scene B, SO_2 analysis (a) RGB image using 9.2-10.7-8.6 μm bands (b) principal component 4.

6.5 Comments on Airborne Gas Targets

If there is adequate and accurate information to specify a target signature, then a signature-based detection algorithm such as the subspace GLRT is a good choice for gas detection. Anomaly detection may be all that is possible under certain circumstances; however, it should be used in conjunction with other target identification methods since the AD is prone to high false alarm rates. In the LWIR, anomaly detectors find “hot” objects that may or may not have spectral similarity to the material of interest.

Traditionally, PCA and MNF are used for dimension reduction as will be discussed in Chapter 8, but they are also useful for visually identifying the dominant image components. Again, these components are not guaranteed to be a gas plume or any other object of interest, yet spectral transformations are a helpful tool in gaining a perspective on the key

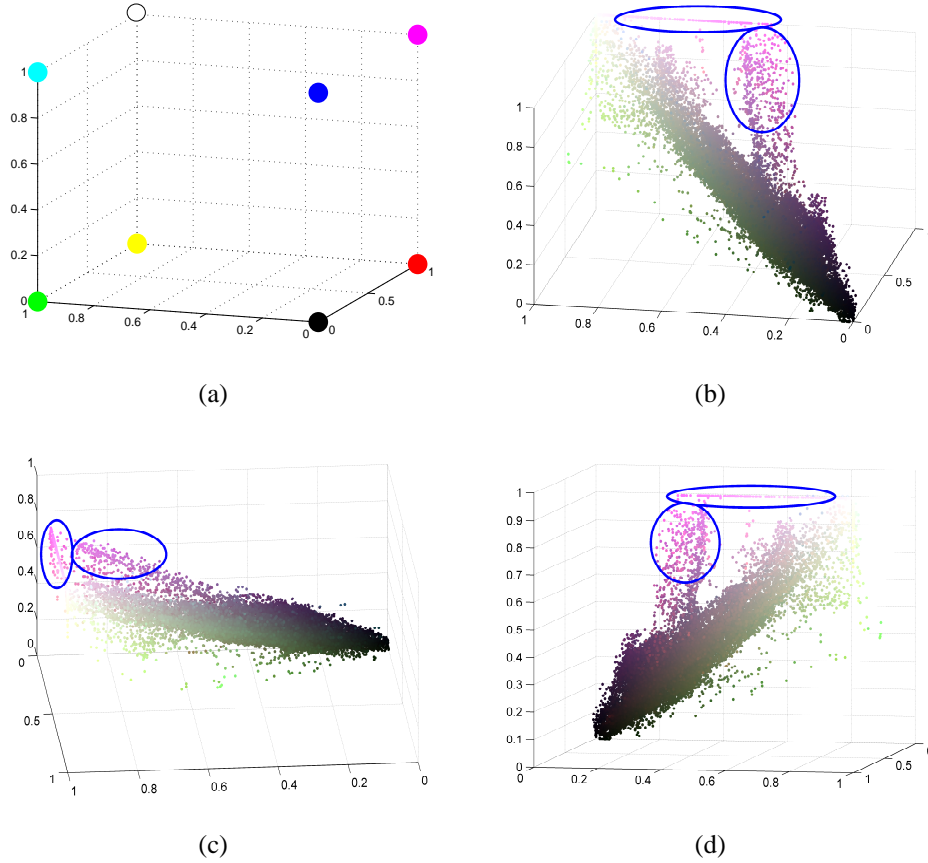


Figure 35: Scene B, SO_2 RGB image using $9.2\text{-}10.7\text{-}8.6\mu\text{m}$ bands: (a) 3-d RGB color space, (b)-(d) 3-d scatter plots from various view angles.

constituents in a scene.

RGB composite images using well-chosen 3-band combinations offer a way to make a specific gas stand out visually. This method of interpretation is good for gases with very distinct absorption/emission lines. Plotting that RGB data as a 3-d scatter plot can also be a valuable visualization tool. It allows a user to make a rough quantitative estimate of the number of pixels in a gas plume and also helps to qualitatively assess the changing optical thickness of the gas in the scene.

This chapter's investigation into airborne gases as targets offers a contrast to ground-based objects for hyperspectral image detection. Obviously, the results here are preliminary and this application is rich for follow-on work in the area of algorithm development. As yet, there is no single, simple technique for consistent and accurate gas plume detection.

Applying signature-based detection algorithms is a valid approach, but only when the gas signature is well-defined for the scene of interest.

CHAPTER 7

EFFECTS OF COVARIANCE CONTAMINATION IN ADAPTIVE DETECTION

7.1 The Problem of Contaminated Covariance Estimates

One practical constraint that comes into play for timely hyperspectral target detection is the quality of training data. Adaptive detectors require a secondary i.i.d. data set, training data, which is used to estimate the covariance of interference against which the target signal is being sought. However, in many cases of actual operation of HSI sensors, proper training data is not available. Sometimes archived data is used or the test data is carved into two pieces, where one part is used for training and the other for test. Either way, it is highly likely that target or target-like samples are unwittingly included amongst the background data that is used to estimate a spectral covariance. In the radar literature this is referred to as self-nulling. This contamination of the covariance matrix is obviously in violation of a key assumption (i.e., i.i.d.) under which adaptive detectors are developed. Moreover, engineering intuition leads us to assume that it is also detrimental to the detection statistic output. In this chapter we investigate the impact that covariance contamination has on a set of adaptive detectors for hyperspectral imaging and the general detectability of targets.

7.2 Performance Comparison

Actual hyperspectral imaging data were used in our experiments, not simulations. Scenes in the images are mostly a desert terrain. The data was collected by Sensor X in the reflective regime and Airborne Hyperspectral Imager (AHI) in the emissive regime. The scenes used for these experiments all have accompanying ground truth. This means we have a priori signatures for the targets of interest, and knowledge of the number and location of target pixels in the hyperspectral images. For each data set, the background covariance was

first computed by removing all targets and target-like samples, thus leaving only interference. Then the covariance was again estimated using the entire data set that included a specified number of target-bearing samples. Detection statistics were computed using the contaminated covariance and the uncontaminated covariance estimate for each target type in all four scenes.

Three different ground targets whose spectral signatures are similar to the background, as measured by spectral angle, were chosen. The same targets (denoted as targets 1, 2, and 3) are in all four scenes, except for target 3 that is not present in scene 4. Evaluating three detectors in four scenes gives us 12 test cases. The SNR in dB is given for each of the targets in tables 5-8. Note that each of the targets has a relatively low SNR value, with nine of the 11 targets being less than 2 dB.

In addition to a low SNR value, another factor that makes a given target difficult to detect using a hyperspectral sensor is how spectrally similar it is to the background against which it is being sought. One way of determining spectral similarity between two spectra of length L is by measuring the angle between the spectra in an L -dimensional vector space. The spectral angle mapper (SAM), while simple to compute, has demonstrated widespread utility and continues to be part of HSI algorithm research [66]. Spectral angle is computed for two spectra, vectors \mathbf{s} and \mathbf{b} , as

$$\theta(\mathbf{s}, \mathbf{b}) = \arccos \left(\frac{\langle \mathbf{s}, \mathbf{b} \rangle}{\|\mathbf{s}\| \|\mathbf{b}\|} \right), \quad (60)$$

where $\langle \cdot, \cdot \rangle$ is the dot product and $\|\cdot\|$ is the 2-norm.

Spectrally similar targets are of keen interest since they present HSI detection algorithms with only a small amount of spectral contrast, which makes them difficult to detect. Tables 5-8 also show the spectral angle between the mean background spectra and the particular target signatures for each scene. Ideally for detection, a target would have a spectral response that is very different from the background and thus be orthogonal. In three dimensions, it is easy to visualize that a 90° angle between the target and background vectors would be ideal. Targets that are nearly orthogonal are called ‘well-separated.’ The angles

for all targets used in this study are very small, with all but one of eleven being separated from the background by less than 1° . Illustrating this point is figure 36, which depicts the spectral angle between target 3 and the background in scene 3. Considering these small angles, and the modest SNR of these targets, it is reasonable to consider these targets difficult.

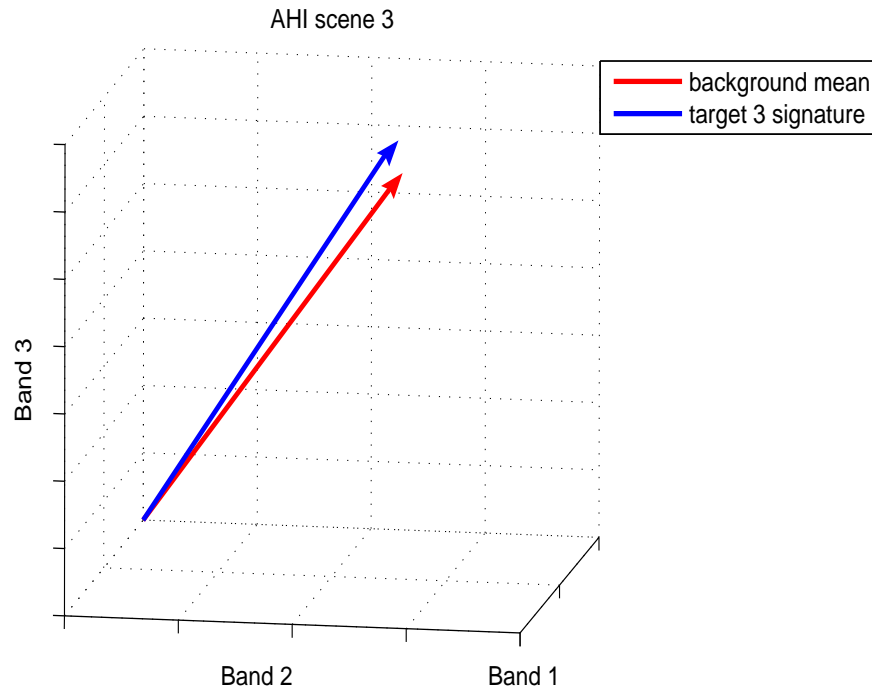


Figure 36: Three dimensional vector space illustration of small spectral angle between target 3 and background in AHI scene 3.

Another reason for using these targets is experimental control. The high degree of spectral similarity for the targets chosen assures us that the experiments are not overly influenced by one or two target samples that are extraordinarily different from the background. Instead, these target pixels are only slightly different from the mean background pixel. This allows us to illustrate that covariance contamination, even with only a few samples that are not very different from typical interference, has a significant impact on adaptive detection output.

Tables 5-8 show the number of target and background pixels for each scene. In all

scenes, target pixels account for no more than a fraction of 1% of total samples. Figure 37 shows the results of our experiments. For each scene, three detection statistics (AMF, GLRT, ACE) are evaluated separately for the target set. At each target, the value of the detection statistic using a contaminated covariance is plotted on the left and stands next to the value of the detection statistic using a proper, uncontaminated covariance on the right. Above each pair is the value of how much greater the uncontaminated detector output is with respect to the contaminated output. These differences range from only a few percent to over 50% in some cases. It can be seen that, as expected, different targets in different backgrounds display varying degrees of a reduced response. This change also varies with algorithm type.

Table 5: Target info for AHI scene 1 with 145445 background pixels.

Item	Num. Pixels	Spectral Angle	SNR
target 1	26	$\theta = 0.85^\circ$	3.63
target 2	8	$\theta = 0.47^\circ$	2.48
target 3	30	$\theta = 0.23^\circ$	1.81

Table 6: Target info for AHI scene 2 with 128250 background pixels.

Item	Num. Pixels	Spectral Angle	SNR
target 1	44	$\theta = 0.55^\circ$	1.77
target 2	42	$\theta = 0.65^\circ$	1.88
target 3	23	$\theta = 0.70^\circ$	1.61

Table 7: Target info for AHI scene 3 with 161222 background pixels.

Item	Num. Pixels	Spectral Angle	SNR
target 1	45	$\theta = 0.77^\circ$	1.70
target 2	56	$\theta = 0.55^\circ$	1.57
target 3	29	$\theta = 1.13^\circ$	1.58

Table 8: Target info for AHI scene 4 with 116790 background pixels.

Item	Num. Pixels	Spectral Angle	SNR
target 1	19	$\theta = 0.91^\circ$	1.71
target 2	24	$\theta = 0.76^\circ$	1.73

Figure 38 graphically shows the structure and composition of residual contamination in covariance matrices for two Sensor X data sets. In part (a) for one scene and part (c) for another scene, the residual error from target contamination of the background covariance estimate is shown as a color-scaled image. The first spectral bands ($0.35\mu m$) are in the lower left corner, and the last spectral bands ($2.4\mu m$) are in the upper right corner of the image. Parts (b) and (d) of figure 38 show the 3-d structure of the residual error due to covariance contamination. Overall, this figure shows that contamination most severely affects the visible and NIR bands in the reflective regime.

The results shown in figures 37 and 38 are averages. This means the percent change in figure 37, and the residual error graphically shown in figure 38, is the average reduction over all target pixels. The maximum degradation found is as follows for each scene. AHI, scene 4: 3.14 dB. AHI, scene 1: 2.53 dB. Sensor X, run 6300: 1.77 dB. Sensor X, run 5700: 1.85 dB.

7.3 Detector Robustness to Covariance Contamination

Even using ground targets with low SNR and high spectral similarity with the background, it is clear that popular adaptive detectors are negatively impacted when only a few target samples are included in the estimate of the interference covariance. We first reported this finding in [67]. This result seems to be consistent with other studies done from a theoretical perspective on similar topics [68], [69]. The Adaptive Matched Filter, Generalized Likelihood Ratio Test, and Adaptive Coherence Estimator all use a maximum likelihood (ML) estimate of the background covariance matrix to characterize interference. Figure 37 shows that among these three detection statistics, the ACE algorithm is the most robust to

covariance contamination. The performance of ACE was the least affected for all targets in all test cases, typically on the order of a 10-15% reduction in detector output when the covariance is contaminated. Separately, the response of the AMF and GLRT algorithms was very similar, for both contaminated and uncontaminated estimates. This is due to the fact that for a large number of samples, which is often the case in HSI (e.g., 100000-250000+ pixels), the contribution of the term $\mathbf{x}^T \hat{\mathbf{\Gamma}}^{-1} \mathbf{x}$ in the denominator of equation (38) is negligible and the statistic becomes more like equation (39) as $N \rightarrow \infty$. Clearly, whichever algorithm is used, data consumers should expect a reduction in detector output for target samples when the interference covariance is contaminated.

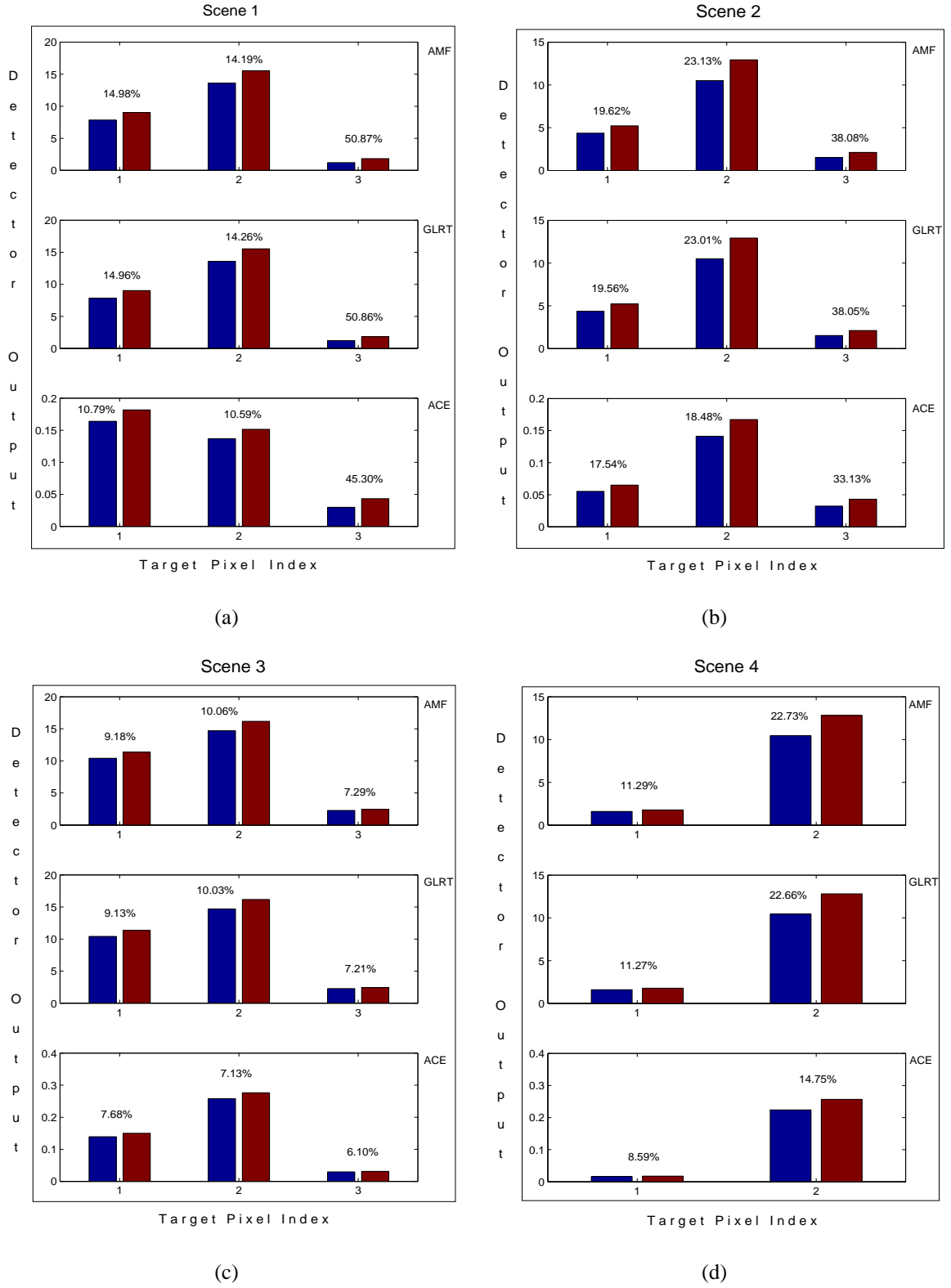


Figure 37: Detection statistic output using contaminated covariance estimate (left, blue) and uncontaminated estimate (right, red) for AHI data.

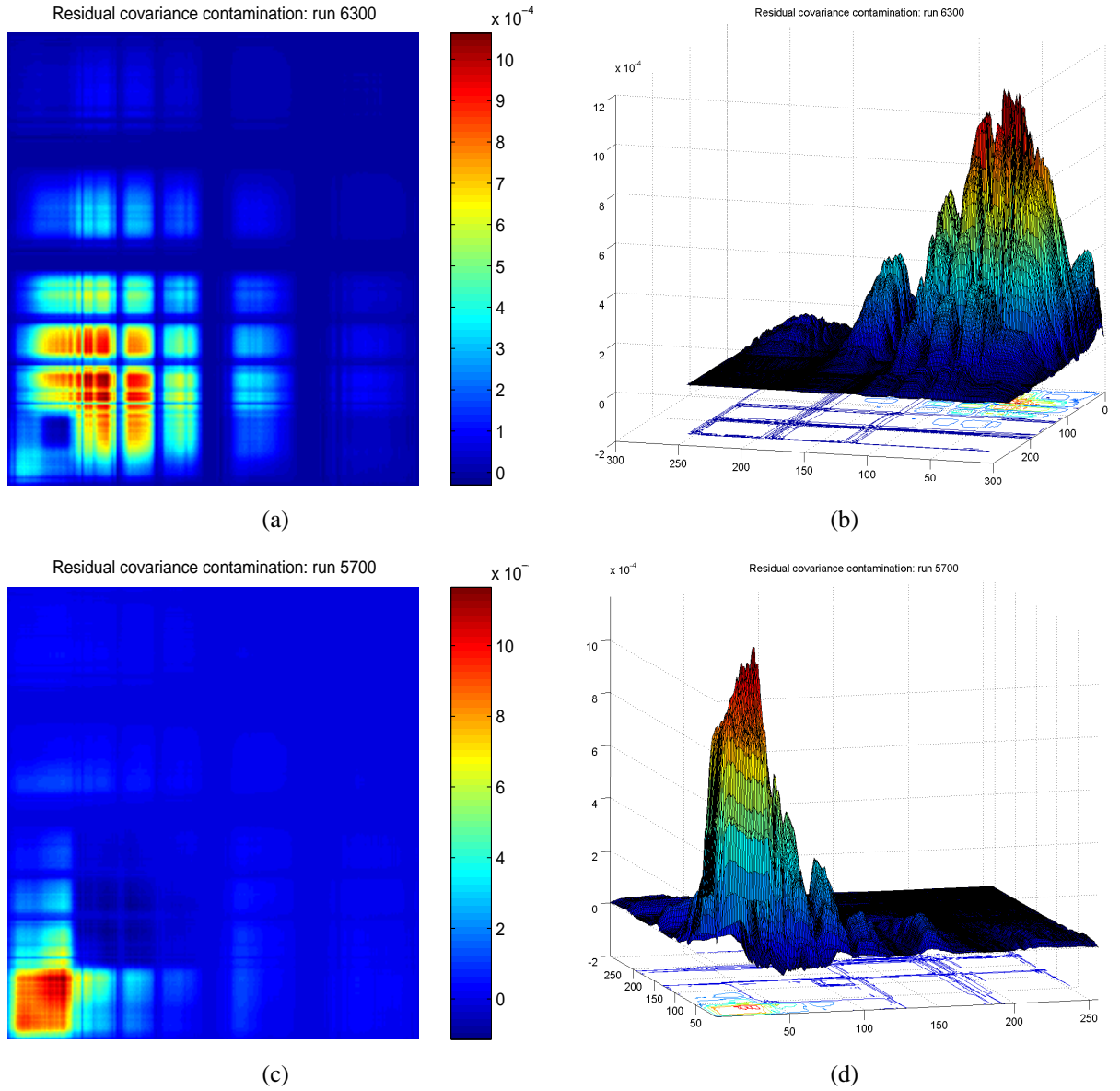


Figure 38: Graphical presentation of residual covariance error (clean - contaminated) for Sensor X. Run 6300: (a) and (b). Run 5700: (c) and (d).

CHAPTER 8

IMPACT OF DIMENSION REDUCTION ON DETECTOR OUTPUT

Whatever the application, those responsible for fielding hyperspectral sensors and exploiting HSI data face a set of similar challenges. Chief among these issues are the storage, transfer, and computer processing of large files generated by hyperspectral sensors. Operational instruments on airborne platforms, such as unmanned air vehicles (UAVs), airplanes, and helicopters, are being designed to have small form factors – leaving little room for large on-board storage capability. Communications links on tactical platforms have limited bandwidth for real-time data transmission. The same can be said for scientific platforms such as NASA’s EO-1 satellite, which carries the Hyperion hyperspectral imager. Ground stations and other processing centers frequently rely on commercial off-the-shelf (COTS) computer workstations for digital data computations, and such hardware struggles to yield exploitation results in real-time or even near real-time for full HSI data cubes.

To deal with these constraints, dimension reduction is often performed at some point in the image chain of TCPED: tasking \rightarrow collection \rightarrow processing \rightarrow exploitation \rightarrow dissemination. Generally speaking, thanks to numerous narrow image channels, a significant amount of spectral redundancy exists in HSI data. Some level of signal compression or dimension reduction is appropriate. Careful attention must be given, however, to the impact that such pre-processing operations can have on exploitation algorithms. In this chapter, we seek to characterize and quantify the impact that the two most widely-used dimension reduction techniques have on adaptive detection statistics. We again do this for the case of difficult targets – signals of interest that are spectrally very similar to the background against which these targets are being sought.

8.1 Brief Survey of Techniques for HSI Data Reduction

There are a number of dimension reduction techniques for hyperspectral imaging data that have appeared in the literature. One of those worth noting that will not be covered in this thesis is [70], where wavelet spectral analysis is employed for automated dimensionality reduction of pixel vectors. Projection pursuit – related to the geometry of convex sets – is discussed in [71], where the authors claim to find image features that can be used to unmix pixel vectors into a smaller data set.

Utilizing similar concepts of convexity, the N-FINDR algorithm [72] finds the set of pixels with the largest possible volume by “inflating” a simplex within the data. The input for the algorithm is the full image cube, with no previous pre-processing. A random set of vectors is initially selected. In order to refine the initial estimate of endmembers, every pixel in the image must be evaluated in terms of pixel purity likelihood or nearly pure statehood. To achieve this, the volume is calculated for every pixel in the place of each endmember. A trial volume is calculated for every pixel in each endmember position by replacing that endmember and finding the volume. If the replacement results in a volume increase, the pixel replaces the endmember. This procedure is repeated until there are no more replacements of endmembers. Once the pure pixels are found, their spectra can be used to unmix the original image. This produces a set of images that show the abundance of each endmember. While the endmember determination step of N-FINDR has been optimized and can be executed rapidly, the computational performance of the algorithm depends on the accuracy of the initial random selection of endmembers.

Another successful approach has been the pixel purity index (PPI) [73], which is again based on the geometry of convex sets [74]. PPI considers spectral pixels as vectors in an L -dimensional space (where L is the number of spectral bands). The algorithm proceeds by generating a large number of random vectors, also called “skewers” [75], through the dataset. Every data point is projected onto each skewer, along which the position of each

point is recorded for every skewer. The data points that correspond to extrema in the direction of a skewer are identified and placed on a list. As more skewers are generated the list grows, and the number of times a given pixel is placed on this list is also tallied. The pixels with the highest tallies are considered the purest ones, since a pixel count provides a “pixel purity index.”

8.1.1 PC-based Transforms for Dimension Reduction

The principal component (PC) transform (also called PCA as in Chapter 6) is arguably the most popular dimension reduction technique for hyperspectral image processing. As indicated in Section 6.4.1, the PC transform is the discrete counterpart of the continuous Karhunen-Loeve transform (KLT). In HSI exploitation, PCA offers a straightforward approach for computation and is optimal in a statistical sense of preserving a maximal amount of the variability (i.e., energy) present in the original data. PCA does not take into account any information about noise or the target signal of interest in the case of detection applications. On the other hand, the Minimum Noise Fraction (MNF) version of PCA does use a statistical estimate of noise in order to produce new data vectors whose elements are ordered in terms of SNR. This, too, was covered in Section 6.4.1, and the utility of PC transforms in various applications is covered in [76].

8.1.2 A Note on Data Reduction vs. Dimension Reduction

To be clear, the notion of data reduction refers to any transformation that results in fewer (supposedly representative) pixels or in an HSI cube with a smaller number of elements per image pixel vector. PCA and MNF are truly focused on the elimination of spectral redundancy (i.e., reducing dimensionality) in the image signal, whereas PPI and N-FINDR attempt to generate reduced data sets by way of endmember determination. As such, the objective functions are obviously different. Nonetheless, all the techniques mentioned above are widely used during pre-processing and prior to the application of detection algorithms.

Our initial investigations into the impact of PCA on adaptive detection statistics were

published in [77]. In the next section, the detectability of targets is analyzed for both PCA and MNF dimension reduction.

8.2 *Performance Comparison*

Data from AHI and Sensor X were once again used to empirically analyze the effects of PCA and MNF dimension reduction. All 256 Sensor X bands were used and 70 final AHI bands were used. The fact that this is actual data and ground truth is available for these scenes makes these measurements well-suited to algorithmic experiments.

Target detectability is a function of SNR and spectral contrast. As discussed in Chapter 7, targets with similar spectral characteristics as the background exhibit a small spectral angle when the target and background mean are viewed as vectors in Euclidean band space. As discussed in Section 7.2, spectral angle is computed by equation (60). This simple quantity has demonstrated its utility and proven effective in not only in HSI detection, but also classification [78], [79]. In order to narrow our investigation to focus on the impact that dimension reduction has on SNR, we use targets with modest spectral angle but high SNR. For these particular tests, tables 9 and 10 describe the data involved.

To be precise, the listed SNR throughout this thesis is a first-order approximation computed using the Mahalanobis distance,

$$\text{SNR} = 10 \log_{10}(\hat{\Delta}), \quad (61)$$

where for each pixel vector \mathbf{x}

$$\hat{\Delta} = (\mathbf{x} - \hat{\boldsymbol{\mu}})^T \hat{\boldsymbol{\Gamma}}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}). \quad (62)$$

In equation (62), the estimated covariance of the background is denoted by $\hat{\boldsymbol{\Gamma}}$ and the estimated background mean is $\hat{\boldsymbol{\mu}}$. Recall that the maximum likelihood (ML) estimates are, respectively,

$$\hat{\boldsymbol{\Gamma}} = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \hat{\boldsymbol{\mu}})(\mathbf{x}_n - \hat{\boldsymbol{\mu}})^T, \quad (63)$$

$$\hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n. \quad (64)$$

Table 9: Number of target and background pixels for dimension reduction tests.

Sensor	Description	Target	Background
Sensor X	run 6300: grassland	target A = 29	121690
Sensor X	run 1068: dirt area	target A = 30	125580
AHI	run 0769: desert	target D = 30	145445
AHI	run 2349: desert	target D = 31	142887

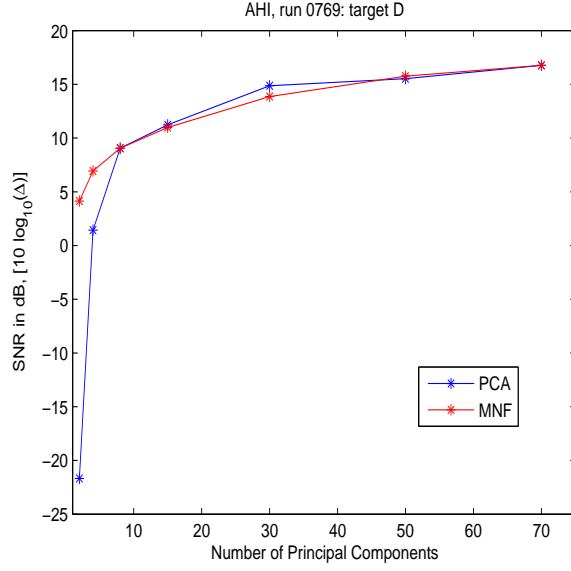
Table 10: SNR and spectral angle between targets and mean background for dimension reduction tests.

Sensor	Description	Spectral Angle	SNR
Sensor X	run 6300: grassland	target A: $\theta = 13.29^\circ$	16.77
Sensor X	run 1068: dirt area	target A: $\theta = 14.58^\circ$	16.87
AHI	run 0769: desert	target D: $\theta = 1.92^\circ$	23.69
AHI	run 2349: desert	target D: $\theta = 2.44^\circ$	22.83

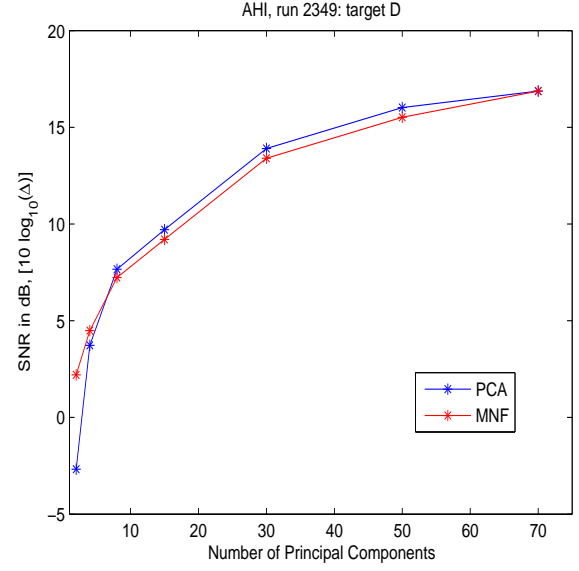
Figure 39 shows the effects of dimension reduction on target detectability for the four test cases. In each part of the figure, the red curve plots the SNR as a function of number of principal components used for the MNF transform. The blue curve in each figure plots the SNR as a function of number of principal components used for the standard PCA transform.

A few things are clear from these results. First, the MNF transform preserves more target SNR when only a few principal components are used. That is to say that when 10 or fewer dimensions are retained, the MNF transform produced a smaller reduction in target SNR in the new (reduced) data set. For moderate levels of dimensionality reduction, between 10 and 50 principal components of the original data, there was no clear winner. In fact, standard PCA often performed better in preserving target SNR than MNF at these reduction levels. While the results for moderate reduction levels were inconclusive, this suggests an interesting point. When a moderate number of components are used, typically accounting for 99 + % of the original energy, it seems that specifics of the target signature

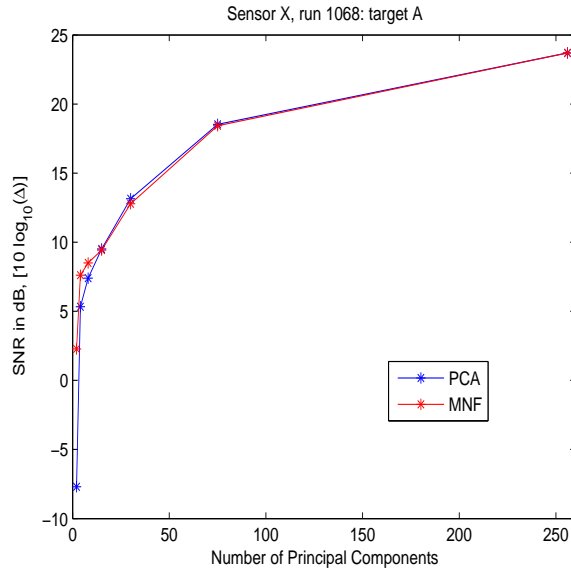
really dictate which objective function (i.e., transform) is more effective. As expected, when more components are used, the preserved SNR approaches the original SNR of the target in the full data cube. For more than 50 principal components, there was little, if any, difference between PCA and MNF for any of our test cases.



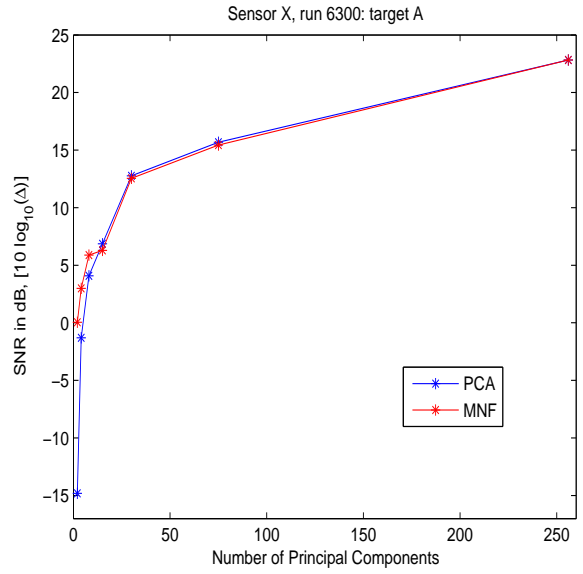
(a)



(b)



(c)

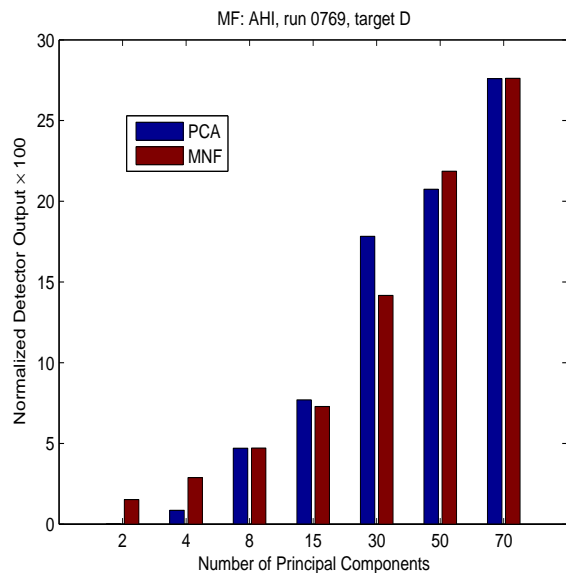


(d)

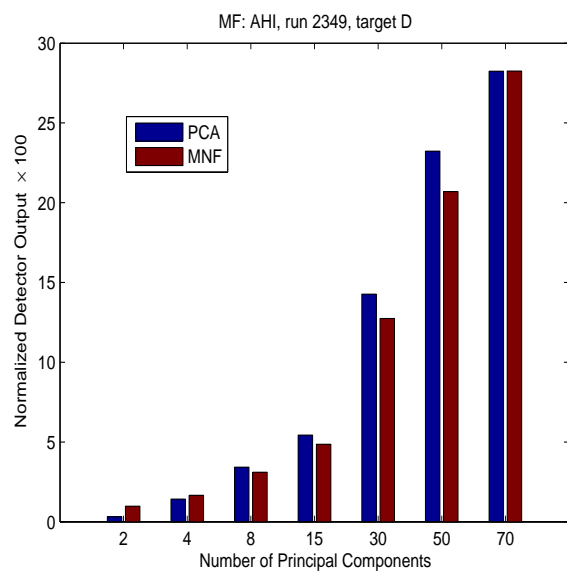
Figure 39: Target detectability for dimension reduced data: (a) AHI, scene 1; (b) AHI, scene 4; (c) Sensor X, run 1068; (d) Sensor X, run 6300.

Looking at detector output for reduced data, results mirrored what was seen in terms of target SNR performance. As fewer principal components were used to create a dimension reduced data set, detector output went down. Also, data reduced using the MNF transform has slightly higher detection statistics than data reduced using standard PCA for extreme levels of dimension reduction. Figures 40 - 43 make comparisons on an algorithm by algorithm basis, showing PCA and MNF results side by side for each scene. Again MNF is in red and PCA is colored blue. The figures plot detector output – normalized and scaled to be between 1-100 – as a function of principal components used.

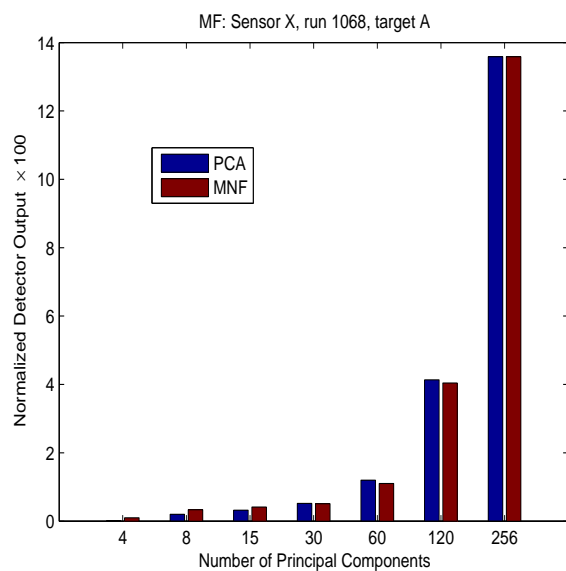
Similar to the results for land mines in Chapter 5 and airborne gases in Chapter 6, the ACE algorithm was a superior performer followed by the subspace GLRT. Despite a slightly higher SNR, the targets in the Sensor X scenes proved difficult due to their limited spectral contrast (i.e., small spectral angle).



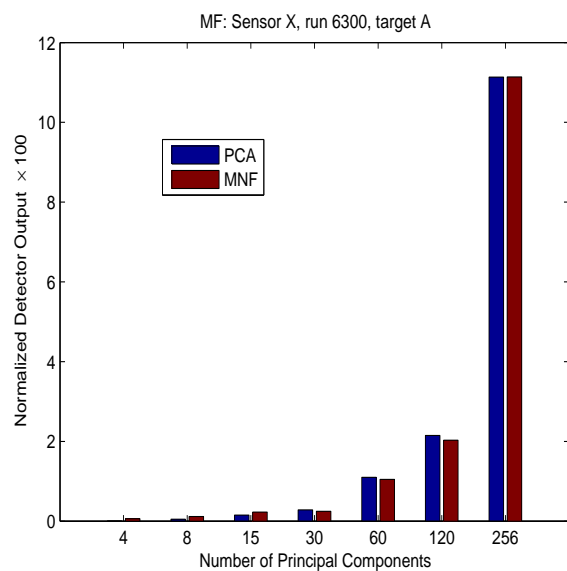
(a)



(b)

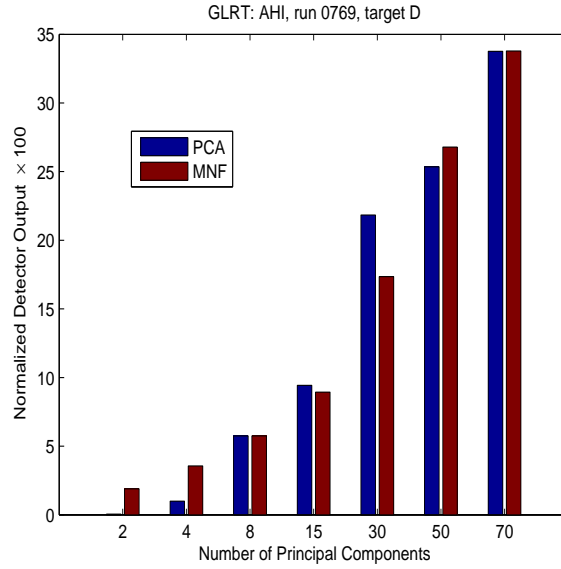


(c)

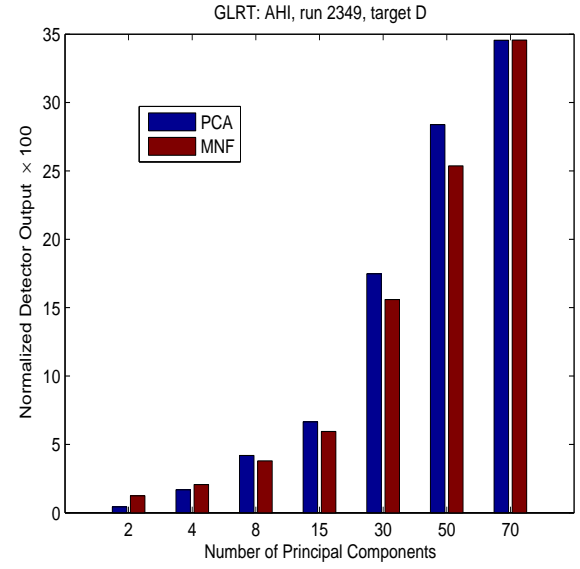


(d)

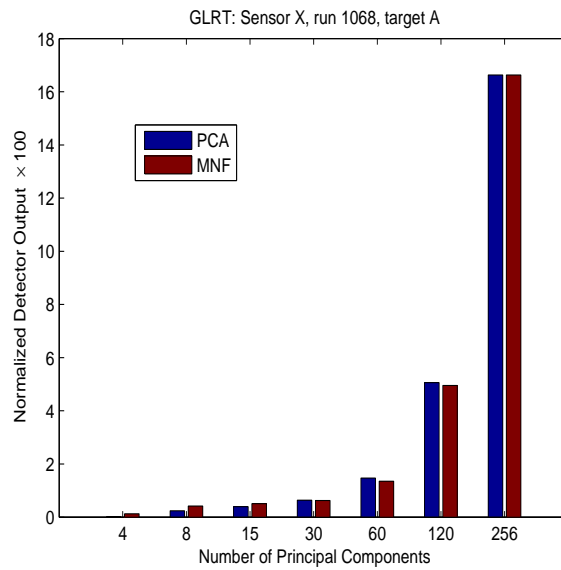
Figure 40: Matched filter detector output for dimension reduced data: (a) AHI, scene 1; (b) AHI, scene 4; (c) Sensor X, run 1068; (d) Sensor X, run 6300.



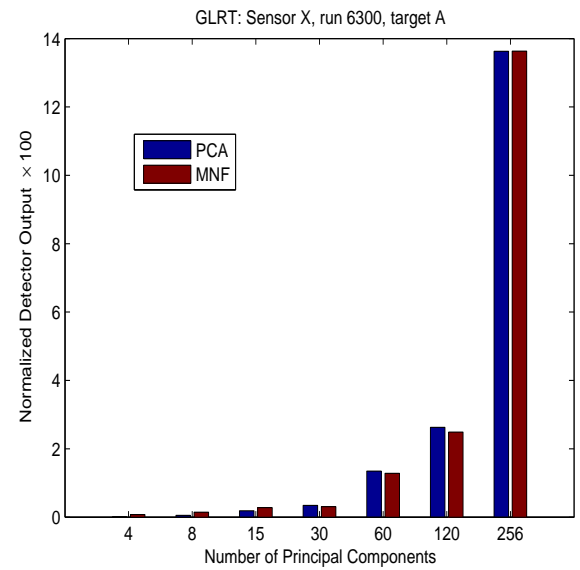
(a)



(b)

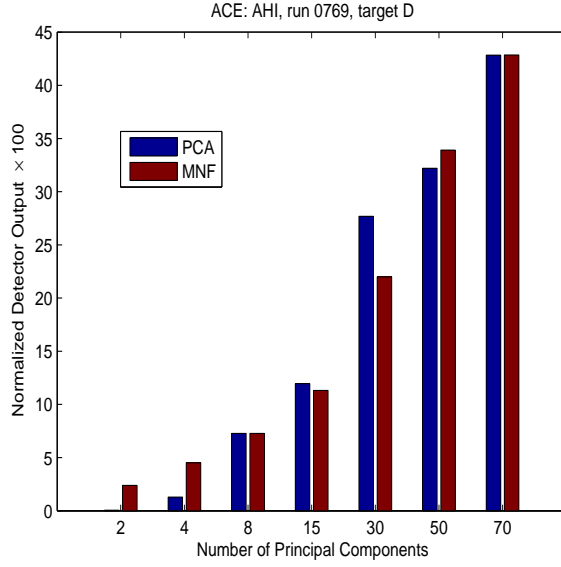


(c)

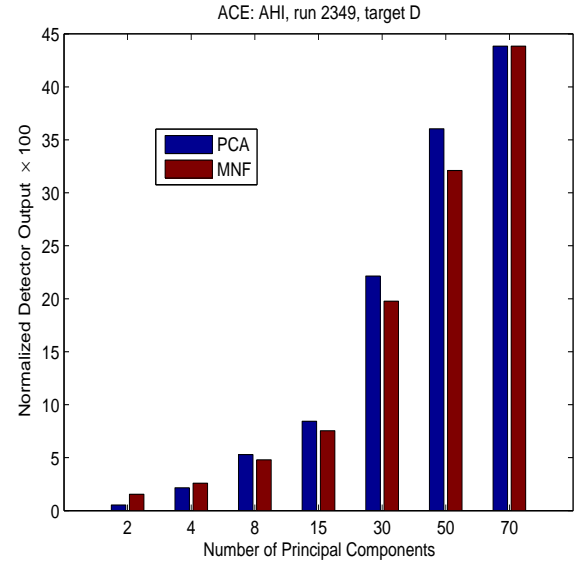


(d)

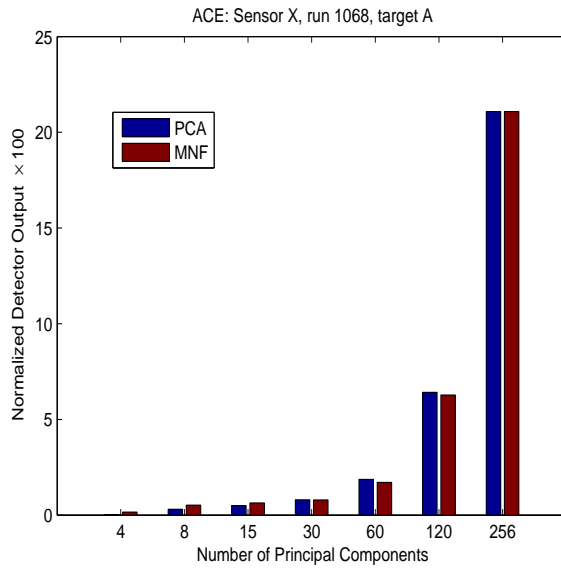
Figure 41: GLRT detector output for dimension reduced data: (a) AHI, scene 1; (b) AHI, scene 4; (c) Sensor X, run 1068; (d) Sensor X, run 6300.



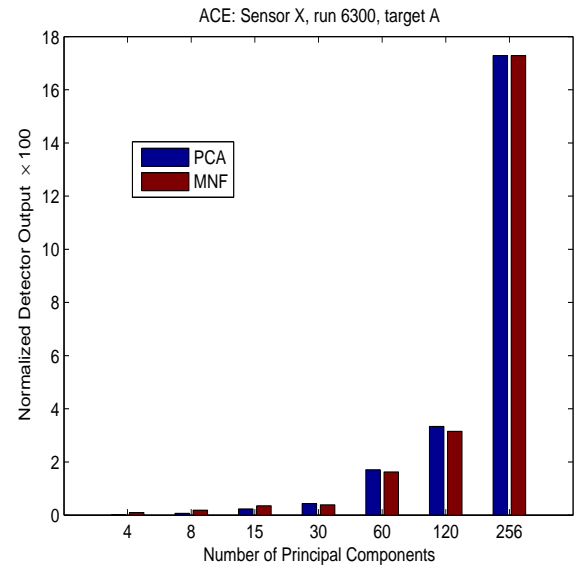
(a)



(b)

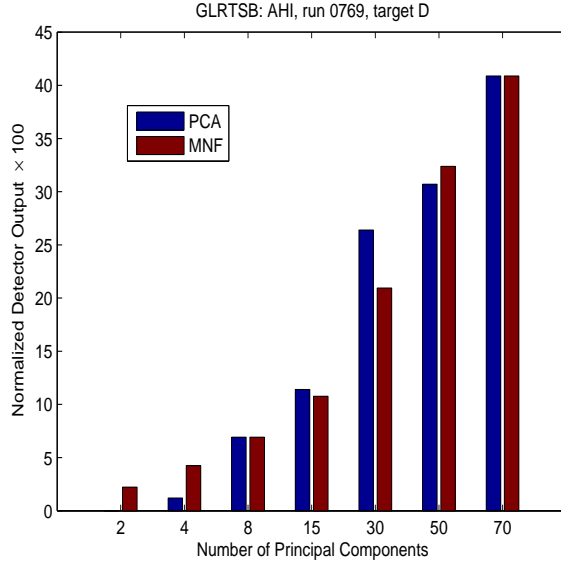


(c)

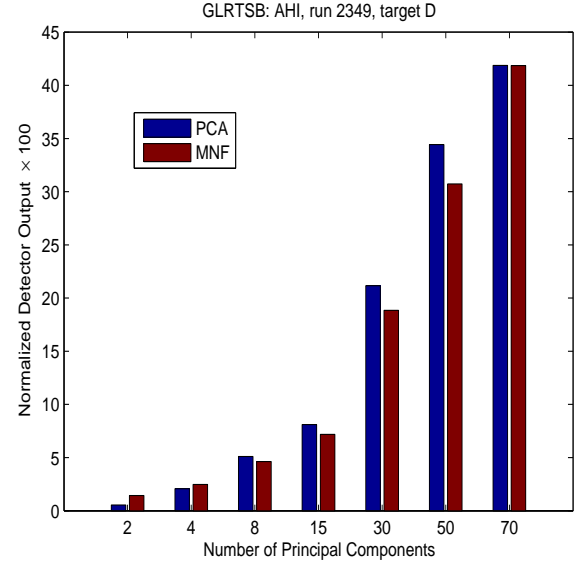


(d)

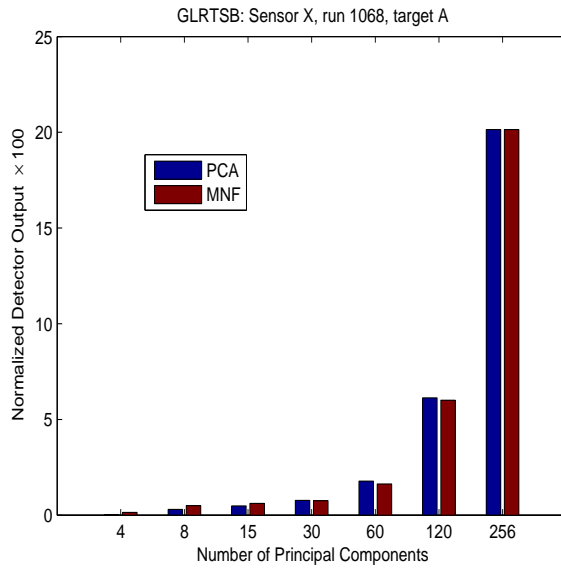
Figure 42: ACE detector output for dimension reduced data: (a) AHI, scene 1; (b) AHI, scene 4; (c) Sensor X, run 1068; (d) Sensor X, run 6300.



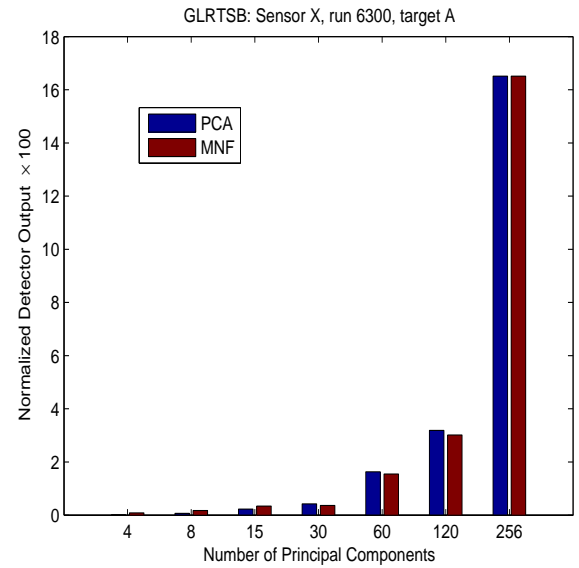
(a)



(b)



(c)



(d)

Figure 43: Subspace GLRT detector output for dimension reduced data: (a) AHI, scene 1; (b) AHI, scene 4; (c) Sensor X, run 1068; (d) Sensor X, run 6300.

8.3 Detector Robustness to Data Reduction

In this chapter we have evaluated the impact that two principal components transforms have on target detection performance in hyperspectral imaging, and empirically quantified the change in detection statistic output and target detectability that results for a wide variety of captured energy levels.

Detection performance for reduced data was dependent on algorithm, and this result seems to be consistent with other studies done from a different perspective on similar topics [80], [81]. But a few new and interesting trends can be gleaned from our experiments.

Intuitively, it may be expected that detector output would increase and decrease proportional to the number of principal components included in the reduced data. However, this wasn't always the case in our experiments. For example, figure 40(a) shows that the output for eight components was more than half the output for 15 components. Yet the output for 30 components was almost exactly double that of 15 components. In figure 41(c), the detection statistic computed for 60 principal components was only 35% of that for 120 components. Likewise, the value for 120 components was only 30% of the value computed using 256 components. This type of inconsistency was seen in both data regimes, SWIR and LWIR, as well as in different detectors, including the matched filter and GLRT. The ACE algorithm in figure 42 is reasonably indicative. Part (a) shows that for 70 principal components the detection value is roughly 42, and for eight components the value is about six. That means for almost nine times the number of spectral bands, the detector output is about seven times greater. This sometimes nonlinear proportionality is the same for the PCA and MNF transforms.

As such, it is difficult to proclaim any sort of “optimal” level of dimension reduction. A suitable level of detector output must be determined on an individual basis per specific target and scene characteristics.

The number of target pixels relative to the total number of pixels in the scene is another important factor. With only a small number of pixels present that follow a particular

spectral signature (i.e., target), dimension reduction transforms such as PCA are likely to discard their contribution to the reduced image if a small portion of the overall image variability is present in those samples. This fact has direct implication for HSI exploitation tasks such as wide-area detection and reconnaissance, where the targets are often difficult and/or sparse.

CHAPTER 9

CONCLUSIONS

In this research, a series of pre-processing steps for the exploitation of hyperspectral imaging (HSI) data were investigated. Traditional assumptions for statistical data modeling, covariance estimation from training data, and dimensionality reduction were evaluated to see what happens when assumptions for these items are varied or violated altogether.

First, background was provided on the history and use of HSI in remote sensing. Adaptive detection was highlighted as a primary application for HSI sensing, and parallels were drawn between target detectors for radar and HSI systems.

Then the task of statistically modeling HSI data was covered in detail. The common choice of modeling data using Normal distributions was shown to be insufficient in some cases. As such, a finite mixture modeling approach was developed for elliptically contoured t (EC- t) distributions. Assuming no a priori knowledge, two parameter estimation techniques were presented to simultaneously estimate all unknown parameters from the data. Results demonstrated that EC- t mixtures can be an accurate and attractive alternative to Normal models in many data sets, but it is not a silver bullet solution to the statistical data modeling problem. The Stochastic Expectation-Maximization (SEM) parameter estimation technique proved quite valuable, and seems well-suited to this and similar situations.

A concise treatment of adaptive detectors for HSI was given, covering algorithms that use both structured and unstructured approaches for characterizing the background. In experiments on both SWIR and LWIR data, the Adaptive Coherence Estimator (ACE) algorithm was consistently the best performer. Its property of scale invariance allowed the ACE to detect both full and sub-pixel land mine targets [82] more robustly than other algorithms. Such scale invariance may be why the ACE detector was less prone to the effects of

shadowing and secondary illumination differences in test and training data. All in all, it is apparent that land mines are difficult targets to detect passively, and HSI can play a unique role in mine detection.

Next, attention was given to non-terrestrial targets, specifically airborne gas plumes. Signature-based algorithms performed better than generic anomaly detection [83]. However, it was noted that gases are difficult specimens to generate precise target signatures for, and anomaly detection may be used frequently in practice. Other methods for gas identification were discussed, including three-dimensional scatter plots and special RGB composites of carefully chosen bands that highlight distinct emission/absorption features.

The issue of covariance contamination was then addressed, and the problems for target detection associated with poor training data were made clear. Including target samples in background covariance estimates is known as covariance contamination, and its negative impact on detector output was quantified experimentally using multiple data sets. Also the structure and magnitude of residual error in the covariance matrix itself was illustrated. Reduction in performance due to covariance contamination was shown to depend largely on algorithm type, with the ACE again showing the most resiliency. Bands in the visible and near infrared (NIR) were the most prone to contamination error. Despite the feeling of many in the remote sensing community that inclusion of target pixels in training data is not a problem for HSI detection scenarios, the results presented in chapter 7 show that covariance contamination can be a very real problem and leads to noticeable performance degradation.

Finally, the topic of dimension reduction was investigated. This is a widely used step in pre-processing hyperspectral imagery, and is done to reduce computational complexity for exploitation algorithms and reduce file size for data storage. Yet most HSI practitioners believe that dimension reduction greatly reduces detector output, and only tolerate this to speed up execution. After a brief overview of other data reduction and endmember determination methods, two types of principal components (PC) transforms were discussed.

Standard PCA and minimum noise fraction (MNF) PCA were used on real data sets to quantify the impact that dimension reduction has on target detectability. The MNF transforms the data into new spectral axes ordered by decreasing SNR, and this had a profound impact on target detectability for cases of extreme dimensionality reduction. MNF preserved target SNR better than PCA for extreme dimension reduction. Standard PCA does not take noise into account, and rotates the data to a new set of orthogonal axes that place a maximal amount of variability (i.e., energy) from the original image signal into the first few rotated bands. For moderate dimension reduction, PCA and MNF had the same impact on target detectability and ultimately detector output. Surprisingly to some, the results of chapter 8 demonstrate that significant dimension reduction can be achieved with a minimal impact on target detectability.

9.1 Contributions

Contributions of the thesis include the following:

- Extension of a finite-mixture model using elliptically contoured t distributions to accommodate heavy-tailed hyperspectral imaging data.
- Development of methods for simultaneously estimating the parameters of an EC- t mixture model (including dof) using the Expectation-Maximization and Stochastic Expectation-Maximization algorithms.
- An empirical analysis of the performance characteristics of five state-of-the-art adaptive detection algorithms for both reflective and emissive HSI data. These detectors were originally designed for radar systems and have only recently been applied to HSI exploitation.
- A fair, “apples to apples” comparison of hyperspectral detection algorithms for land mines. These results are the only performance benchmark of their kind in the literature for mine detection using HSI.

- Utility assessment of principal component-based transforms for use in special RGB composite images to help visually identify gas plumes in hyperspectral imagery.
- Development of a three-dimensional scatter plot approach for qualitatively evaluating the concentration and quantitatively evaluating the abundance of gas plumes in HSI data. This is based on a flexible, but judicious selection of spectral bands for RGB color planes to highlight a distinct emission/absorption feature of the gas specimen.
- An experimental, quantitative analysis of the deleterious impact that covariance contamination has on adaptive detection performance. Even for relatively weak targets with limited spectral contrast, a significant reduction in detector output was demonstrated for multiple data sets.
- An evaluation of the effect dimension reduction techniques have on target detectability. The Minimum Noise Fraction (MNF) transform maintains target detectability better than does standard PCA in the cases of extreme dimension reduction. Considerable dimension reduction is possible without a noticeable loss in detection performance.
- Fully automated software implementations of the discussed modeling, detection, and dimension reduction algorithms.

9.2 *Future Work*

Algorithms that assume a priori info As the name implies, adaptive detectors all require a secondary data set to train the algorithm on what not to look for in the scene. While this has been a practical and successful approach to date, theoretical work is possible for HSI in which additional a priori information is assumed. In such cases, Bayesian inference can be brought to bear on this problem. Despite the fact it is unlikely that complete a priori information would be available for many HSI exploitation scenarios, it is an interesting theoretical exercise for a novel sensor application.

Non-terrestrial targets In this thesis an initial investigation was performed on detection of airborne gases, and results showed some limited success. To date, the majority of HSI detection work has focused on ground targets. While this is an important application, there is tremendous potential in exploiting hyperspectral imagery to detect and classify scenes with airborne materials of interest. The emissive regime is where most of this work will be done, and detection work in the LWIR for airborne targets will require a deeper understanding of the atmospheric physics involved. However, with bio-terrorism and homeland security a growing source of both concern and funding, it would be foolish to ignore airborne gas targets for HSI.

Inclusion of Morphology A fundamental assumption made in all the detection work presented in this thesis is that only spectral information was used in determining target signal presence. But as the spatial resolution of HSI sensors improves, there is a better opportunity to include spatial structure and shape information of targets in a decision process. Classical image processing techniques including image segmentation and edge detection might play a role in introducing morphology into the hyperspectral target detection procedure.

Signature phenomenology and band selection Selecting the “best” bands to support the exploitation of a particular signature is not currently done. If band selection is done in such a way that takes into account the unique properties of the target signature and background – rather than simply a statistical generality of the entire data cube – there is opportunity for a major benefit. This thesis talked briefly about the vastly different nature of target signatures for buried vs. surface land mines. These differences are largely dependent on the spectral regime in which the data is collected, SWIR vs. LWIR. Signature phenomenology can play a larger role in band selection.

Data fusion All the work in this thesis, as well as the vast majority of HSI exploitation

work in the literature, is based upon sole source data. That is, detection or classification results are based on the processing output from a single hyperspectral imaging sensor. This is valid in order to assess the utility of HSI in general, however, it is unlikely that for many applications the end user will be able to rely solely upon the information product of an HSI sensor with high confidence. As such, it makes sense to look at novel ways of fusing the data collected from multiple remote sensors (e.g., synthetic aperture radar (SAR), broadband IR imagers, high resolution panchromatic cameras, etc.) in order to produce a more confident and accurate detection result. Each of these sensors have algorithms that are developed in a “stovepipe” fashion, and do not take into account the features made available by other sensors for detection and estimation.

9.3 Concluding Remarks

In this work, we have attempted to answer two difficult questions: “Are typical assumptions for HSI exploitation useful and valid?” and “What happens if you change or violate these assumptions?” Although these questions are somewhat subjective and depend on the specific exploitation scenario, it has been shown that target detection depends very much on assumptions made in problem formulation and choices made during pre-processing. There may never be a sufficient level of agreement among the remote sensing community as to which steps are best taken for detecting targets, but through the models and analysis presented here, a series of guideposts have been developed that can hopefully be used toward making sound decisions in the performance prediction and exploitation of hyperspectral imaging data.

REFERENCES

- [1] E.C. Barrett and L.F. Curtis, *Introduction to environmental remote sensing*, 4th ed., Stanley Thornes, Cheltenham, 1999.
- [2] W.G. Rees, *Physical Principles of Remote Sensing*, 2nd ed., Cambridge University Press, Cambridge, 2001.
- [3] C. Elachi, *Introduction to the Physics and Techniques of Remote Sensing*, 2nd ed., Wiley & Sons, New York, 1987.
- [4] J.B. Campbell, *Introduction to Remote Sensing*, 3rd ed., Guilford Press, London, 2002.
- [5] G.A. Shaw and H.K. Burke, "Spectral Imaging for Remote Sensing," *Lincoln Laboratory Journal*, vol. 14, pp. 3–28, 2003.
- [6] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," in *IEEE Signal Processing Magazine*, pp. 12–16, vol. 19, no. 1, 2002.
- [7] I.S. Reed, J.D. Mallett, and L.E. Brennan, "Rapid Convergence Rate in Adaptive Arrays," *IEEE Trans. Aerospace and Electronic Systems*, vol. 10, pp. 853–863, 1974.
- [8] W. Krzanowski, *Principles of Multivariate Analysis*, Oxford Statistical Science Series. Oxford University Press, Oxford, UK, 2000.
- [9] I.M. Chakravarti, R.G. Laha, and J. Roy, *Handbook of Methods of Applied Statistics: Volume I*, John Wiley and Sons, 1967.
- [10] G. Snedecor and W. Cochran, *Statistical Methods*, 8th Ed., Iowa State University Press, 1989.
- [11] D. Marden and D. Manolakis, "Non-Gaussian models for hyperspectral algorithm design and assessment," in *IEEE Int. Conf. Geosci. and Remote Sensing*, Toronto, 2002.
- [12] R.B. D'Agostino and M.A. Stephens, *Goodness-of-Fit Techniques*, Marcel Dekker Inc., New York, 1986.
- [13] P. Lucey, T. Williams, and M. Winter, "Recent results from AHI: a LWIR hyperspectral imager," in *Proc. SPIE*, vol. 5159, pp. 361–369, 2003.
- [14] K.T. Fang and Y.T. Zhang, *Generalized Multivariate Analysis*, Springer-Verlag, Berlin, 1990.
- [15] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 3rd ed., Wiley, New York, 2003.

- [16] K.T. Fang, S. Kotz, and K. Ng, *Symmetric Multivariate and Related Distributions*, Chapman & Hall, London, 1990.
- [17] C. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*, Wiley, New York, 1995.
- [18] J. How and H. Leung, “No evidence of stable distributions in radar clutter,” in *IEEE Signal Processing Workshop on Higher Order Statistics*, Banff, Canada, 1997.
- [19] G. Vane, “First Results from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS),” JPL Publication, vol. 87-38, no. 15, 1987.
- [20] J. Pearlman, C. Segal, L. Liao, S. Carman, M. Folkman, B. Browne, L. Ong, and S. Ungar, “Development and operations of the EO-1 Hyperion imaging spectrometer,” in *Proc. SPIE*, vol. 4135, pp. 243–253, 2000.
- [21] M. Dundar and D. Landgrebe, “A model-based mixture-supervised classification approach in hyperspectral data analysis,” *IEEE Trans. Geoscience and Remote Sensing*, vol. 40, pp. 2692–2699, 2002.
- [22] D. Manolakis and D. Marden, “Statistical modeling of hyperspectral imaging data and their applications,” *MIT Lincoln Laboratory Project Report*, HTAP-16, 2004.
- [23] G. McLachlan and D. Peel, *Finite Mixture Models*, Wiley, New York, 2000.
- [24] D. Peel and G.J. McLachlan, “Robust mixture modeling using the t distribution,” *Statistics and Computing*, vol. 10, pp. 335–344, 2000.
- [25] A. Dempster, N. Laird, and D. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society, B*, vol. 39, no. 1, pp. 1–38, 1977.
- [26] K. Lange, R. Little, and J. Taylor, “Robust statistical modeling using the t distribution,” *Journal of the American Statistical Association*, vol. 84, pp. 881–896, 1989.
- [27] M.D. Farrell Jr. and R.M. Mersereau, “Estimation of elliptically contoured mixture models for hyperspectral imaging data,” in *IEEE Int. Conf. Geosci. and Remote Sensing*, Anchorage, AK, 2004.
- [28] X. Meng and D. Rubin, “Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm,” *Journal of the American Statistical Association*, vol. 86, pp. 899–909, 1991.
- [29] G. Wei and M. Tanner, “A Monte Carlo implementation of the EM algorithm and the poor man’s data augmentation,” *Journal of the American Statistical Association*, vol. 85, pp. 699–704, 1990.
- [30] G. Celeux and J. Diebolt, “The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem,” *Computational Statistics Quarterly*, vol. 2, pp. 73–82, 1985.

- [31] M.D. Farrell Jr. and R.M. Mersereau, "Robust automatic clustering of hyperspectral imagery using non-Gaussian mixtures," in *Proc. SPIE*, vol. 5573, pp. 161–172, 2004.
- [32] M.D. Farrell Jr. and R.M. Mersereau, "Hyperspectral parameter estimation of elliptically contoured t mixture models using Expectation-Maximisation," in press *Int. Journal of Remote Sensing*, November, 2005.
- [33] S.M. Kay, *Fundamentals of Statistical Signal Processing*, Prentice Hall, Englewood Cliffs, 1998.
- [34] C.D. Richmond, "Adaptive array processing in non-Gaussian environments," in *IEEE Asilomar Conf. on Signals and Systems*, Pacific Grove, CA, 1996.
- [35] C.D. Richmond, "A note on non-Gaussian adaptive array detection and signal parameter estimation," in *IEEE Trans. Signal Processing Letters*, vol. 3, pp. 251–252, 1996.
- [36] E.J. Kelly, "An adaptive detection algorithm," in *IEEE Trans. Aerospace and Electronic Systems*, vol. 22, pp. 115–127, 1986.
- [37] E.J. Kelly, "Adaptive detection in non-stationary interference," in *MIT Lincoln Laboratory Technical Report 761*, DTIC# ADA-185622, 1986.
- [38] F.C. Robey, D.R. Fuhrmann, E.J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," in *IEEE Trans. Aerospace and Electronic Systems*, vol. 28, pp. 208–216, 1992.
- [39] S. Kraut and L.L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," in *IEEE Trans. Signal Processing*, vol. 47, pp. 2538–2541, 1999.
- [40] L.L. Scharf and L.T. McWhorter, "Adaptive matched subspace detectors and adaptive coherence," in *IEEE Asilomar Conf. on Signals and Systems*, Pacific Grove, CA, 1996.
- [41] J.C. Harsanyi and C.I. Chang, "Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach," in *IEEE Trans. Geoscience and Remote Sensing*, vol. 32, pp. 779–785, 1994.
- [42] L.L. Scharf and B. Friedlander, "Matched subspace detectors," in *IEEE Trans. Signal Processing*, vol. 42, pp. 2146–2157, 1994.
- [43] D. Stein, S. Beaven, L. Hoff, E. Winter, A. Schaum and A. Stocker, "Anomaly detection from hyperspectral imagery," in *IEEE Signal Processing Magazine*, pp. 58–69, vol. 19, no. 1, 2002.
- [44] J.A. Craib, "Survey of mine clearance technology," *Occasional Paper Series*, United Nations University, New York, 1994.
- [45] H. Hambric and W. Schneck, "The antipersonnel mine threat," in *Technology and the Mine Problem Symposium*, pp. 3/11–3/45, Naval Postgraduate School, Monterey, CA, 1996.

- [46] K. Tsipis, "Report on the landmine brainstorming workshop of August 25-30," in *MIT Program in Science & Technology for International Security*, report no. 27, Cambridge, MA, 1996.
- [47] E. Winter, M. Schlangen, A. Bowman, M. Carter, C. Bennett, D. Fields, W. Aimonetti, P. Lucey, J. Johnson, K. Horton, T. Williams, A. Stocker, A. Oshagan, A. DePersia and C. Sayre, "Experiments to support the development of techniques for hyperspectral mine detection," in *Proc. SPIE*, vol. 2759, pp. 139–148, 1996.
- [48] G. Maksymenko, B. Ware, and D. Poole, "A characterization of diurnal and environmental effects of mines and the factors influencing the performance of mine detection ATR algorithms," in *Proc. SPIE*, vol. 2496, pp. 140–151, 1995.
- [49] J. Simard, "Improved landmine detection capability (ILDC): systematic approach to the detection of buried mines using passive IR imaging," in *Proc. SPIE*, vol. 2765, pp. 489–500, 1996.
- [50] E. Peterson, S. Brown, T. Hattenberger, and J. Schott, "Surface and buried landmine scene generation and validation using the digital imaging and remote sensing image generation model," in *Proc. SPIE*, vol. 5546, pp. 312–323, 2004.
- [51] P. Lucey, T. Williams, M. Mignard, J. Julian, D. Kobubun, G. Allen, D. Hampton, W. Schaff, M. Schlangen, E. Winter, W. Kendall, A. Stocker, K. Horton, and A. Bowman "AHI: an airborne long-wave infrared hyperspectral imager," in *Proc. SPIE*, vol. 3431, pp. 36–43, 1998.
- [52] D.W. Messinger, "Gaseous plume detection in hyperspectral image: a comparison of methods," in *Proc. SPIE*, vol. 5425, pp. 592–603, 2004.
- [53] E. O'Donnell, D. Messinger, C. Salvaggio, and J. Schott, "Identification and detection of gaseous effluents from hyperspectral imagery using invariant algorithms," in *Proc. SPIE*, vol. 5425, pp. 573–582, 2004.
- [54] S. Young, "Detection and quantification of gases in industrial stack plumes using thermal infrared hyperspectral imaging," in *Technical Report ATR-2002-8407-1*, Aerospace Corporation, 2002.
- [55] P. Lucey, T. Williams, M. Winter, and E. Winter, "Performance of the AHI airborne thermal infrared hyperspectral imager," in *Proc. SPIE*, vol. 4132, pp. 186–194, 2000.
- [56] A. Berk, G. Anderson, P. Acharya, J. Chetwynd, L. Bernstein, E. Shettle, M. Matthew, and S. Adler-Golden, "MODTRAN4 User's Manual," in Air Force Research Laboratory documentation, Hanscom AFB, 1999.
- [57] S. Young, B. Johnson, and J. Hackwell, "An in-scene method for atmospheric compensation of thermal infrared hyperspectral data," in *Journal of Geophysical Research*, vol. 107, no. D24, pp. 4774–4786, 2002.

- [58] A. Schaum, "Spectral subspace matched filtering," in *Proc. SPIE*, vol. 4381, pp. 1–17, 2001.
- [59] D. Williams, B. Feldman, T. Williams, A. Pilant, P. Lucey, and L. Worthy, "Detection and identification of toxic air pollutants using airborne LWIR hyperspectral imaging," in *Proc. SPIE*, vol. 5655, pp. 134–141, 2004.
- [60] United States Environmental Protection Agency, *National Emission Inventory Data and Documentation*, <http://www.epa.gov/ttn/chief/net/2002inventory.html>, 2002.
- [61] J.A. Richards, *Remote Sensing Digital Image Analysis*, Springer-Verlag, Berlin, 1993.
- [62] G. Strang, *Linear Algebra and Its Applications*, 3rd ed., Harcourt College Publishers, Fort Worth, 1988.
- [63] A. Green, M. Berman, P. Switzer, and M. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," in *IEEE Trans. Geosci. Rem. Sens.*, vol. 26, pp. 65–74, 1998.
- [64] J. Lee, S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote sensing data by a noise adjusted principal components transform," in *IEEE Trans. Geosci. Rem. Sens.*, vol. 28, pp. 295–304, 1990.
- [65] A. Mares, R. Olsen, and P. Lucey, "LWIR spectral measurements of volcanic sulfur dioxide plumes," in *Proc. SPIE*, vol. 5425, pp. 266–272, 2004.
- [66] N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," in *IEEE Trans. Geoscience and Remote Sensing*, vol. 42, no. 7, pp. 1552–1565, 2004.
- [67] M.D. Farrell Jr. and R.M. Mersereau, "On the impact of covariance contamination for adaptive detection in hyperspectral imaging," in *IEEE Signal Processing Letters*, vol. 12, no. 9, pp. 649–652, 2005.
- [68] K. Gerlach, "The effects of signal contamination on two adaptive detectors," in *IEEE Trans. Aerospace and Electronic Systems*, vol. 31, pp. 297–309, 1995.
- [69] K.F. McDonald and R.S. Blum, "Performance characterization of space-time adaptive processing algorithms for distributed target detection in non-ideal environments," in *IEEE Int. Radar Conference*, Long Beach, CA, 2002.
- [70] S. Kaewpijit, J. LeMoigne, and T. El-Ghazawi, "Automatic Reduction of Hyperspectral Imagery Using Wavelet Spectral Analysis," in *IEEE Trans. Geoscience and Remote Sensing*, vol. 41, no. 4, pp. 863–871, 2003.
- [71] L. Jimenez and D. Landgrebe, "Hyperspectral Data Analysis and Feature Reduction Via Projection Pursuit," in *IEEE Trans. Geoscience and Remote Sensing*, vol. 37, no. 6, pp. 2653–2667, 1999.

- [72] M.E. Winter, “N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data,” in *Proc. SPIE*, vol. 3753, pp. 266–275, 1999.
- [73] J. Boardman, F. Kruse, and R. Green, “Mapping target signatures via partial unmixing of AVIRIS data,” in *5th JPL Airborne Earth Science Workshop*, Pasadena, CA, 1995.
- [74] A. Ifarraguerri and C.I. Chang, “Multispectral and hyperspectral image analysis with convex cones,” in *IEEE Trans. Geoscience and Remote Sensing*, vol. 37, pp. 756–770, 1999.
- [75] J. Theiler, D. Lavenier, N. Harvey, S. Perkins, and J. Szymanski, “Using blocks of skewers for faster computation of pixel purity index,” in *Proc. SPIE*, vol. 4132, pp. 61–71, 2000.
- [76] J.E. Jackson, *A User’s Guide to Principal Components*, Wiley and Sons, New York, 1991.
- [77] M.D. Farrell Jr. and R.M. Mersereau, “On the impact of dimension reduction using PCA for hyperspectral detection of difficult targets,” in *IEEE Geoscience and Remote Sensing Letters*, vol. 2, no. 2, pp. 192–195, 2005.
- [78] S. Rand and D. Keenan, “Spatially smooth partitioning of hyperspectral imagery using spectral/spatial measures of disparity,” in *IEEE Trans. Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1479–1490, 2003.
- [79] G. Mercier, and M. Lennon, “Support vector machines for hyperspectral image classification with spectral-based kernels,” in *IEEE Int’l Geoscience and Remote Sensing Symposium (IGARSS)*, Toulouse, France, 2003.
- [80] S. Shen and B. Beard, “Effects of hyperspectral compression on nonliteral exploitation,” in *Proc. SPIE*, vol. 3438, pp. 191–195, 1998.
- [81] S. Beaven, D. Stein, and S. Stewart, “Effect of compression on detection in hyperspectral data,” in *IEEE Asilomar Conf. on Signals and Systems*, Pacific Grove, CA, 2000.
- [82] M.D. Farrell Jr. and R.M. Mersereau, “Comparison of Adaptive Array Processing Schemes for Land Mine Detection Using Hyperspectral Imagery,” in *Proc. SPIE*, vol. 5913, pp. 1–11, 2005.
- [83] M.D. Farrell Jr. and R.M. Mersereau, “Detection of Industrial Gaseous Chemical Plumes Using Hyperspectral Imagery in the Emissive Regime,” in *Proc. SPIE*, vol. 5890, pp. 90–101, 2005.

VITA

Michael D. Farrell Jr. was born and raised in the great state of Wisconsin. After graduating high school as a National Honor Society member, he attended The University of Iowa. There he earned a B.S.E. in Industrial & Systems Engineering as well as a Regents' Certificate in Technological Entrepreneurship. After stints in industry with Intel Corporation, EY Consulting Services, and Inventa Technologies, Michael came to the Georgia Institute of Technology in Atlanta to pursue graduate study. He earned the M.S. in Electrical & Computer Engineering and went on for the Ph.D. degree in the same school at Georgia Tech, where he graduated in December 2005.

His technical interests include statistical & adaptive signal processing for imagery, communications, and radar systems. His personal interests include scuba diving, golf, collecting baseball hats, eating Thai food, and watching sports.