

# NAVAL POSTGRADUATE SCHOOL

## Monterey, California



## THESIS

### A SIGNAL PROCESSING PERSPECTIVE OF HYPERSPECTRAL IMAGERY ANALYSIS TECHNIQUES

by

Marcus Stavros Stefanou

June, 1997

Thesis Advisors:

Richard C. Olsen  
Roberto Cristi

Approved for public release; distribution is unlimited.

19980102 117

QUALITY IMPROVEMENT  
DEVELOPMENT

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE June 1997		3. REPORT TYPE AND DATES COVERED Master's Thesis
4. TITLE AND SUBTITLE TITLE OF THESIS A SIGNAL PROCESSING PERSPECTIVE OF HYPERSPECTRAL IMAGERY ANALYSIS TECHNIQUES			5. FUNDING NUMBERS	
6. AUTHOR(S) Marcus Stavros Stefanou				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey CA 93943-5000			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (maximum 200 words) A new class of remote sensing data with great potential for the accurate identification of surface materials is termed hyperspectral imagery. Airborne or satellite imaging spectrometers record reflected solar or emissive thermal electromagnetic energy in hundreds of contiguous narrow spectral bands. The substantial dimensionality and unique character of hyperspectral imagery require techniques which differ substantially from traditional imagery analysis. One such approach is offered by a signal processing paradigm, which seeks to detect signals in the presence of noise and multiple interfering signals. This study reviews existing hyperspectral imagery analysis techniques from a signal processing perspective and arranges them in a contextual hierarchy. It focuses on a large subset of analysis techniques based on linear transform and subspace projection theory, a well established part of signal processing. Four broad families of linear transformation-based analysis techniques are specified by the amounts of available <i>a priori</i> scene information. Strengths and weaknesses of each technique are developed. In general, the spectral angle mapper (SAM) and the orthogonal subspace projection (OSP) techniques gave the best results and highest signal-to-clutter ratios (SCRs). In the case of minority targets, where a small number of target pixels occurred over the entire scene, the low probability of detection (LPD) technique performed well.				
14. SUBJECT TERMS Hyperspectral, Digital Imagery Analysis, Signal Processing, Remote Sensing			15. NUMBER OF PAGES 239	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	



Approved for public release; distribution is unlimited

# **A SIGNAL PROCESSING PERSPECTIVE OF HYPERSPECTRAL IMAGERY ANALYSIS TECHNIQUES**

Marcus Stavros Stefanou  
Captain, United States Marine Corps  
B.S., United States Naval Academy, 1990

Submitted in partial fulfillment of the  
requirements for the degree of

**MASTER OF SCIENCE IN ELECTRICAL ENGINEERING**

from the

**NAVAL POSTGRADUATE SCHOOL**

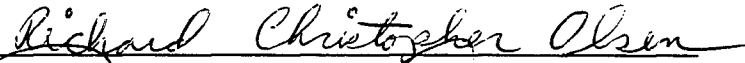
**June, 1997**

Author:

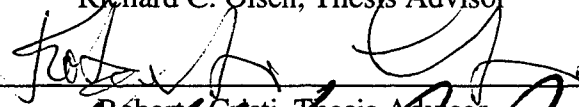


Marcus Stavros Stefanou

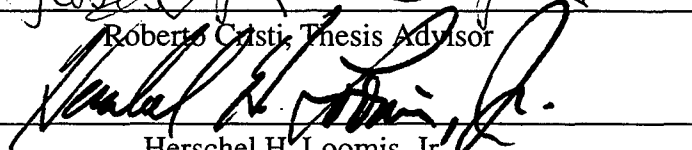
Approved by:



Richard C. Olsen, Thesis Advisor



Roberto Cristi, Thesis Advisor



Herschel H. Loomis, Jr.,

Chairman, Department of Electrical and  
Computer Engineering





## ABSTRACT

A new class of remote sensing data with great potential for the accurate identification of surface materials is termed hyperspectral imagery. Airborne or satellite imaging spectrometers record reflected solar or emissive thermal electromagnetic energy in hundreds of contiguous narrow spectral bands. The substantial dimensionality and unique character of hyperspectral imagery require techniques which differ substantially from traditional imagery analysis. One such approach is offered by a signal processing paradigm, which seeks to detect signals in the presence of noise and multiple interfering signals.

This study reviews existing hyperspectral imagery analysis techniques from a signal processing perspective and arranges them in a contextual hierarchy. It focuses on a large subset of analysis techniques based on linear transform and subspace projection theory, a well established part of signal processing. Four broad families of linear transformation-based analysis techniques are specified by the amounts of available *a priori* scene information. Strengths and weaknesses of each technique are developed. In general, the spectral angle mapper (SAM) and the orthogonal subspace projection (OSP) techniques gave the best results and highest signal-to-clutter ratios (SCRs). In the case of minority targets, where a small number of target pixels occurred over the entire scene, the low probability of detection (LPD) technique performed well.



## TABLE OF CONTENTS

I. INTRODUCTION.....	1
II. BACKGROUND .....	5
A. PROBLEM STATEMENT .....	5
B. STRATEGIES FOR SPECTRAL IMAGERY ANALYSIS .....	5
C. AN OVERVIEW OF SPECIFIC METHODS WITHIN THE STRATEGIES..	9
D. HISTORY.....	13
E. CREATION OF A TAXONOMY FOR THE LINEAR TRANSFORMATION AND PROJECTION STRATEGY .....	15
III. DEFINITIONS.....	17
A. SPECTRAL IMAGERY.....	17
B. STATISTICAL INTERPRETATION.....	22
C. RELATED SIGNAL PROCESSING AND LINEAR ALGEBRA CONCEPTS .....	34
1. Linear Transformations of Random Vectors.....	34
2. Eigenvectors and Eigenvalues.....	35
3. Unitary Transformations .....	36
4. A Geometric Interpretation of the Unitary Transform .....	36
5. Simultaneous Diagonalization of Two Covariance Matrices.....	38
IV. THE PRINCIPAL COMPONENTS ANALYSIS FAMILY OF TECHNIQUES.....	41
A. DESCRIPTION.....	41
B. BACKGROUND DEVELOPMENT.....	42
1. Multivariate Data Analysis View.....	42
2. Signal Processing View.....	44
3. Pattern Recognition View .....	46
C. OPERATION.....	48
1. Basic Principal Components Analysis (PCA).....	48
2. Maximum Noise Fraction (MNF) or Noise Adjusted Principal Components (NAPC) Transform.....	66
3. Standardized Principal Components Analysis (SPCA).....	72
V. THE MATCHED FILTER FAMILY OF TECHNIQUES.....	81
A. DESCRIPTION.....	81
B. BACKGROUND DEVELOPMENT.....	81
1. Linearly Additive Spatially Invariant Image Sequences.....	81
2. The Theory of Least Squares: the <i>A Priori</i> Model .....	83
3. The Theory of Least Squares: the <i>A Posteriori</i> Model.....	86
4. The Matched Filter .....	87

C. OPERATION.....	89
1. The Simultaneous Diagonalization (SD) Filter .....	89
2. Orthogonal Subspace Projection (OSP) .....	96
3. Least Squares Orthogonal Subspace Projection (LSOSP) .....	107
4. The Filter Vector Algorithm (FVA).....	109
VI. THE UNKNOWN BACKGROUND FAMILY OF TECHNIQUES.....	113
A. DESCRIPTION .....	113
B. BACKGROUND DEVELOPMENT.....	113
1. Determining the Intrinsic Dimensionality Based on Information Theoretic Criteria .....	113
2. Beamforming.....	116
C. OPERATION.....	119
1. Low Probability of Detection (LPD).....	119
2. Constrained Energy Minimization (CEM) .....	134
3. Adaptive Multidimensional Matched Filter .....	142
VII. THE LIMITED IMAGE ENDMEMBERS FAMILY OF TECHNIQUES.....	145
A. DESCRIPTION .....	145
B. BACKGROUND DEVELOPMENT.....	145
1. Multiple Signal Classification (MUSIC) Technique.....	146
2. Convex Sets.....	148
3. The Correlation Detector.....	149
4. Singular Value Decomposition (SVD).....	150
C. OPERATION.....	152
1. MUSIC-Based Endmember Identification .....	152
2. Partial Unmixing .....	154
3. Spectral Angle Mapper (SAM) .....	159
4. Unmixing via SVD.....	163
VIII. SUMMARY.....	167
IX. CONCLUSIONS.....	173
APPENDIX A. COLOR FIGURES .....	177
APPENDIX B. SPECIAL COLOR FIGURES .....	205
LIST OF REFERENCES .....	211
INITIAL DISTRIBUTION LIST.....	217

## LIST OF FIGURES

Figure 1.1: Components of a Remote Sensing System .....	1
Figure 2.1: Taxonomy of Spectral Imagery Analysis Strategies.....	5
Figure 2.2: The Mixed Pixel Concept .....	6
Figure 2.3: Spectral Variability Within Material Species .....	7
Figure 2.4: Hierarchy of Techniques in the Linear Transformation and Projection Strategy .....	10
Figure 2.5: Hierachy of Techniques Within the Classification Strategy.....	12
Figure 2.6: Major Imagery Analysis Paradigms.....	14
Figure 3.1: A Typical Multispectral Image Produced by Landsat TM.....	18
Figure 3.2: A Typical Hyperspectral Image Cube.....	19
Figure 3.3: The Concept of a Pixel Vector.....	20
Figure 3.4: Typical Pixel Vectors From Multispectral and Hyperspectral Images.....	20
Figure 3.5: Radiance and Reflectance Spectra of Aberdeen HYDICE Scene.....	21
Figure 3.6: Mean Spectrum with One Standard Deviation of Landsat Image and Mean Spectrum with Representation of Variance of HYDICE Image.....	23
Figure 3.7: Mean Removal Illustrated With Scatter Plots.....	25
Figure 3.8: Histograms of Four of the Boulder Landsat TM Bands.....	28
Figure 3.9: Scatter Plots of Boulder Landsat TM Data Showing Highly Correlated and Uncorrelated Band Combinations.....	29
Figure 3.10: Second Order Statistics of the HYDICE Aberdeen Scene.....	31
Figure 3.11: HYDICE Scene of Davis Monthan Air Force Base.....	33
Figure 3.12: Davis Monthan Radiance Covariance and Correlation Matrices.....	34
Figure 3.13: Linear Transformation of a Two-Dimensional Vector .....	35
Figure 3.14: The Unitary Transformation as a Rotation of Axes.....	37
Figure 3.15: Simultaneous Diagonalization of Signal and Noise Covariance Matrices ...	38
Figure 4.1: PC Transformation Depicted as a Linear Transformation.....	41
Figure 4.2: The Karhunen-Loeve Expansion in Terms of Discrete Signals.....	44
Figure 4.3: Paradigm of the Pattern Recognition Process.....	46
Figure 4.4: First 25 PC Images of Davis Monthan Radiance Scene.....	49
Figure 4.5: First Ten PC Images of Aberdeen Radiance and Reflectance Scenes.....	52
Figure 4.6: Eigenvalue Behavior of the Davis Monthan and Aberdeen Radiance Scene Covariance Matrices.....	53
Figure 4.7: Eigenvalue Behavior of the Aberdeen Reflectance Scene Covariance Matrix .....	54
Figure 4.8: Variance and Entropy Behavior of Davis Monthan Radiance Covariance Matrix .....	55
Figure 4.9: Variance and Entropy Behavior of Aberdeen Radiance Covariance Matrix.....	56
Figure 4.10: Variance and Entropy Behavior of Aberdeen Reflectance Covariance Matrix.....	57
Figure 4.11: Eigenvectors and Traces of the Covariance Matrices of Davis Monthan and Aberdeen Radiance and Reflectance.....	58
Figure 4.12: Eigenvectors of the Davis Monthan Radiance Covariance Matrix.....	59
Figure 4.13: Eigenvectors of Aberdeen Radiance Covariance Matrix.....	60

Figure 4.14: Eigenvectors of Aberdeen Reflectance Covariance Matrix.....	61
Figure 4.15: First Eight Eigenvectors of Davis Monthan Scene Superimposed on a Random Slice Across the Hypercube.....	63
Figure 4.16: First Eight Eigenvectors of Aberdeen Radiance Scene Superimposed on a Random Slice Across the Hypercube.....	64
Figure 4.17: First Eight Eigenvectors of Aberdeen Reflectance Scene Superimposed on a Random Slice Across the Hypercube.....	65
Figure 4.18: The MNF Transform.....	67
Figure 4.19: Eigenvalues of the Estimated Noise and Noise-whitened Covariance Matrices of the Davis Monthan Scene. ....	68
Figure 4.20: First 25 MNF Component Images of the Davis Monthan Scene.....	69
Figure 4.21: The NAPC Transform.....	70
Figure 4.22: Eigenvalue Behavior of the Davis Monthan Correlation Matrix.....	73
Figure 4.23: First 25 Standardized PC images of the Davis Monthan Scene Using the Correlation Matrix.....	74
Figure 4.24: Variance and Entropy Behavior of the Davis Monthan Correlation Matrix.	75
Figure 4.25: Eigenvectors and Trace of the Correlation Matrix of Davis Monthan Scene	75
Figure 4.26: Eigenvectors of the Davis Monthan Correlation Matrix.....	76
Figure 4.27: First Eight Eigenvectors of Davis Monthan Normalized Scene Superimposed on a Random Slice Across the Hypercube. ....	77
Figure 4.28: Comparison of Pixel Vectors from Component Images of the PCA Family of Techniques.....	78
Figure 5.1: Linear Mixing of Target and Background Spectra .....	83
Figure 5.2: Least Squares Illustrated Geometrically.....	85
Figure 5.3: Simple Linear Filter Showing Signal and Noise Components.....	87
Figure 5.4: Sub-scene of Davis Monthan Image with Boxes Showing Chosen Pixels. ..	93
Figure 5.5: Spectra of Selected Davis Monthan Sub-scene Pixel Vectors.....	94
Figure 5.6: The SD Filter Vector .....	94
Figure 5.7: The Orthogonal Complement Projector .....	95
Figure 5.8: Scatter Plot of Davis Monthan Sub-scene Indicating Various Pixel Vector Locations and the Subspace Orthogonal to the Undesired Pixel Vector.....	98
Figure 5.9: Davis Monthan Sub-scene Scatter Plot in Projected Space and Histogram of Final OSP Image.....	100
Figure 5.10: Davis Monthan Sub-scene Scatter Plot in Projected Space and Histogram of the Final OSP Image Using Mixed a Pixel Vector as the Matched Filter..	102
Figure 5.11: Histogram of the Davis Monthan Sub-scene OSP Output Image.....	103
Figure 5.12: Davis Monthan Sub-scene OSP Output Image.....	104
Figure 5.13: Histogram of Output OSP Image for Davis Monthan Scene.....	105
Figure 5.14: Davis Monthan OSP Output Image.. .....	106
Figure 5.15: Comparison of LSOSP and OSP Simulation Results for Three SNRs.....	109
Figure 5.16: FVA Results for a Simulated Spectrum.....	111
Figure 6.1: Uniform Linear Array. ....	116
Figure 6.2: Whitening of the Davis Monthan Sub-scene Noise Covariance Matrix.....	122
Figure 6.3: Whitening of the Observed Davis Monthan Sub-scene Data Covariance Matrix .....	122

Figure 6.4: MDL Criterion Applied to Davis Monthan Subscene .....	123
Figure 6.5: Histogram of Davis Monthan Sub-scene LPD Output.....	124
Figure 6.6: Davis Monthan Sub-scene LPD Output.....	125
Figure 6.7: Histogram of LPD Scalar Image Using the First Eigenvector.....	126
Figure 6.8: Davis Monthan LPD Output Image Using the First Eigenvector .....	127
Figure 6.9: Histogram of LPD Scalar Image Using the First Five Eigenvectors.....	128
Figure 6.10: LPD Projector Matrices Created with the First Eigenvector and the First Five Eigenvectors .....	129
Figure 6.11: LPD Operators and Output Images for the First Ten Eigenvectors.....	130
Figure 6.12: Aberdeen Scene LPD Output Image Using One Eigenvector.....	131
Figure 6.13: Aberdeen Scene LPD Output Image Using Ten Eigenvectors .....	132
Figure 6.14: Effect of the CEM Weight Vector Components on the Observed Pixel Vector Components.....	134
Figure 6.15: Histogram of Davis Monthan Sub-scene CEM Output .....	137
Figure 6.16: Davis Monthan Sub-scene CEM Output .....	137
Figure 6.17: Histogram of CEM Scalar Image Using P-3 Pixel Vector as the Target....	138
Figure 6.18: Davis Monthan CEM Output Image Using P-3 Pixel Vector as the Target	139
Figure 6.19: Histogram of CEM Scalar Image Using B-52 Pixel Vector as the Target..	140
Figure 6.20: Comparison of the CEM operator for Three Different Target Pixel Vectors.....	141
Figure 7.1: Geometric Representation of the Observed Signal Model with Three Antenna Elements.....	146
Figure 7.2: Correlation Detector.....	149
Figure 7.3: Mixing Simplices from Zero to $l$ Dimensions .....	156
Figure 7.4: Purest Pixels in the Davis Monthan Sub-scene .....	156
Figure 7.5: Pure Pixel Spectra from Davis Monthan Sub-scene.....	157
Figure 7.6: Results of Partial Unmixing.....	158
Figure 7.7: SAM Output for Davis Monthan Scene.....	160
Figure 7.8: Histogram of the Davis Monthan SAM Image.....	161





## LIST OF TABLES

Table 3.1: Covariance and Correlation Matrices of Landsat TM Image.....	27
-------------------------------------------------------------------------	----



## LIST OF ABBREVIATIONS, SYMBOLS, AND NOTATION

### Sensors

AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
FSS	Field Spectrometer System
GER	Geophysical and Environmental Research
HYDICE	Hyperspectral Digital Imaging Collection Experiment
PHILLS	Portable Hyperspectral Imager for Low Light Spectroscopy
TIMS	Thermal Infrared Multispectral Scanner
TM	Thematic Mapper

### Remote Sensing

DN	Digital Number
GIFOV	Ground Instantaneous Field of View
$\lambda$	Wavelength
pixel	Picture Element

### Spectral Imagery Analysis

CEM	Constrained Energy Minimization
FVA	Filter Vector Algorithm
IDL	Interactive Data Language
LPD	Low Probability Detection
LSOSP	Least Squares Orthogonal Subspace Projection
MNF	Maximum Noise Fraction
NAPC	Noise Adjusted Principal Components Analysis
OSP	Orthogonal Subspace Projection
PC	Principal Component
PCA	Principal Components Analysis
SAM	Spectral Angle Mapper
SCR	Signal-to-Clutter Ratio
SD	Simultaneous Diagonalization
SPCA	Standardized Principal Components Analysis

### Probability and Statistics

$\text{COV}[\mathbf{x}]$	Covariance matrix calculation of random vector $\mathbf{x}$
$f(\mathbf{x}   \theta)$	Conditional probability density function
$E\{\mathbf{x}\}$	Expectation of the random vector $\mathbf{x}$
$H$	Entropy
pdf	Probability density function
pr	<i>a priori</i>
ps	<i>a posteriori</i>
$N(\mu, \sigma^2)$	Normal Distribution with mean $\mu$ and variance $\sigma^2$
$\text{VAR}[\mathbf{x}]$	Variance of random vector $\mathbf{x}$

### Linear Algebra

$\mathbf{x}$	Vector
$\mathbf{X}$	Matrix
$\langle \rangle$	Subspace
$\langle \mathbf{x}, \mathbf{y} \rangle$	Inner product
$\lambda$	Eigenvalue or Lagrange multiplier
$\partial$	Partial derivative
$\mathbf{I}$	Identity matrix
$\  \cdot \ $	Euclidean norm
$\oplus$	Subspace addition
$\Pi$	Product
$\Sigma$	Summation
$\text{tr}[\mathbf{X}]$	Trace of matrix $\mathbf{X}$
$\nabla$	Gradient
$\mathcal{L}$	Lagrangian
$\text{T}$	Transpose
$\Rightarrow$	Implies
SVD	Singular Value Decomposition

### Signal Processing

*	Complex Conjugate
DKLT	Discrete Kahunen-Loeve Transform
KLT	Kahunen-Loeve Transform
J	Cost Function
MUSIC	Multiple Signal Classification
MSE	Mean Square Error
MVDR	Minimum Variance Distorionless Beamformer
SNR	Signal-to-Noise Ratio
$\hat{\mathbf{x}}$	Estimate of the vector $\mathbf{x}$
$\mathbf{x}(n)$	Discrete Sequence

### Notation of this Study

$\alpha$	Abundance vector
$\mathbf{d}$	Desired pixel vector
$\mathbf{E}$	Matrix of eigenvectors arranged as columns
$\mathbf{e}$	Eigenvector
$\Lambda$	Diagonal matrix of eigenvalues
$\lambda$	Eigenvalue
$l$	Number of spectral bands
$N$	Number of spatial samples in an image
$\mathbf{n}$	Additive noise vector ( $l \times 1$ )
$\mathbf{s}$	Signal Vector ( $l \times 1$ ), or endmember
$\Phi$	Correlation matrix (signal processing version)
$\Sigma_{\mathbf{x}}$	Observed data covariance matrix
$\Sigma_{\mathbf{n}}$	Noise covariance matrix
$\mathbf{R}$	Standardized covariance matrix (statistics version of correlation matrix)
$p$	Intrinsic dimensionality
$\mathbf{M}$	Matrix containing desired and undesired endmembers
$\mathbf{U}$	Matrix containing undesired endmembers or left hand eigenvectors in
$\mathbf{u}$	Undesired pixel vector or left hand singular vector
$\mathbf{w}$	Filter weight vector
$\mathbf{x}$	Observed $l \times 1$ Pixel Vector
$\mathbf{X}$	$l \times N$ Observed data matrix
$y$	Scalar output resulting from a transformation
$\mathbf{y}$	Vector output resulting from a transformation



## ACKNOWLEDGEMENTS

The author would like to thank Mr. Joe McKinney of the Davis Monthan Aerospace Maintenance and Regeneration Center for his assistance in aircraft type identification. Special thanks to Mr. Bernie Brower of the Commercial and Government Systems Division of the Eastman Kodak Company for his support. The author would also like to thank Professor Charlie Therrien of the Naval Postgraduate School Electrical and Computer Engineering Department for his time and insight into signal processing concepts. Thanks to Capt. Steve Finney, USAF, for assistance in drawing parallels between signal processing and spectral imagery analysis. Thanks to the HYMSO Program Office for financial support and for the HYDICE data. Finally, the author would like to thank his wife, Carrie, for her devotion and patience.



## I. INTRODUCTION

A remote sensing system may be viewed in its broadest context as three parts. The first is the scene, that is, the earth's surface and the atmosphere through which energy passes. The second is the sensor system, which is designed so that the scene will be adequately represented for the extraction of desired information. The third component is the processing system, which is optimized with respect to specific information extraction applications. This overarching view of the remote sensing philosophy was introduced by Swain and Davis (1978). Figure 1.1 illustrates the remote sensing system concept, with additional details in the sensor and processing systems. It is important to note that the

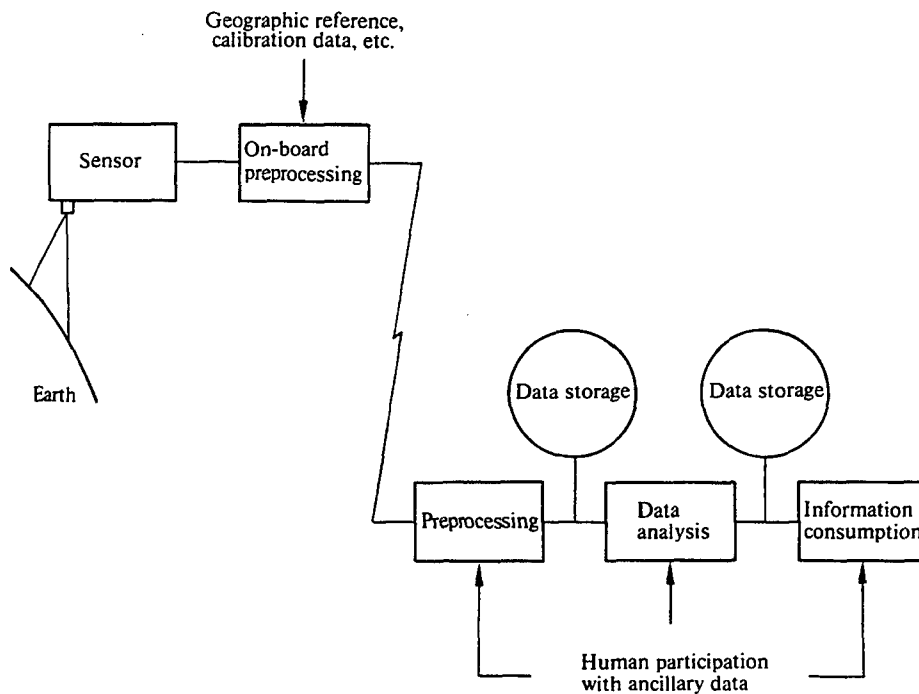


Figure 1.1: Components of a Remote Sensing System.  
From Swain and Davis, 1978, p. 337.

ordering of these elements reflects our increasing level of control over them. The focus of this study is on the element of remote sensing systems over which we have the most control, the data processing. This study will review all currently known techniques for analyzing data from one of the newest family of remote sensing systems, hyperspectral imagery. The particular emphasis of the study is a detailed examination of the techniques

with signal processing origins that have been applied to the specific task of target detection.

The advent of imaging spectroscopy with the Airborne Imaging Spectrometer in 1982 established a new tool for immediate application to several topics in the earth sciences but also created a fundamentally new class of data requiring new approaches to information extraction (Vane and Goetz, 1988, p. 1). This new class of data measures the spectral character of materials on the ground and is referred to as spectral imagery throughout this study. Hyperspectral data, a particular type of spectral imagery, is produced when solar electromagnetic energy reflected from the earth's surface is dispersed into many contiguous narrow spectral bands by an airborne spectrometer (Vane and Goetz, 1988, p. 3). Each picture element (pixel) of a hyperspectral image can be thought of as a high resolution trace of radiation versus wavelength, or a spectrum (Rinker, 1990, p. 6). The characteristic wavelength dependent changes in the emissivity and reflectivity of a given material can be related to the chemical composition and types of atomic and molecular bonds present in that material (Gorman, Subotic, and Thelen, 1995, p. 2805). The chemical composition of different materials is thus manifested in the spectral properties of these materials, and can serve as a means of differentiating materials observed in a hyperspectral image with great detail.

The task of analyzing hyperspectral imagery is complicated by several factors, however. The first is the sheer amount of data inherent in a hyperspectral image. A typical 224-band Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) image, considered to be the state-of-the-art in hyperspectral imaging systems, occupies about 134 Mbytes (Roger and Cavenor, 1996, p. 713). Algorithms for processing such vast quantities of data must be computationally efficient to be of any service, and must seek to eliminate redundant data prior to processing. The second factor is that the radiances recorded at the spectrometer output are subjected to additive noise from the atmosphere, the sensor instrumentation, the data quantization procedure, and transmission back to earth. The cumulative effect of these noise terms is a spectrum that has been corrupted by noise, and detecting a target is no longer a simple proposition. It is here where a signal processing point of view is helpful, as the problem has now become the classical signal in noise problem. The third factor is that owing to the finite spatial resolution of the imaging spectrometer and the actual ground scene, the observed spectrum for a pixel may not be that of a single material. Rather, it could be a mixture of several different materials which exist within the spatial dimensions of the sensor's ground instantaneous

field of view (GIFOV). The GIFOV of the AVIRIS sensor at sea level is nominally 20 m x 20 m (Farrand and Harsanyi, 1995, p. 1566), and the implication is that several materials could contribute to the observed spectrum for that pixel depending on the complexity of the ground scene. A fourth factor that complicates analysis efforts is that spectra of the same type of material often appear very different. This variability within the spectra of a species dictates a statistical approach vice a deterministic one.

There are many types of data processing techniques which address the unique issues raised by hyperspectral imagery. Many of them grew out of earlier techniques which had been successfully applied to multispectral imagery, the precursor of hyperspectral imagery. Others have a foundation in the discipline of pattern recognition. A newer approach, which is naturally suited to the task of detecting signals in the presence of noise and multiple interfering signals, is based on signal processing. It efficiently handles the data by viewing it from the vantage of vectors and matrices, and performs processing by various linear transformations.

A major goal of this study is to logically order the many techniques available for the analysis of hyperspectral data in such a manner that potential users understand the optimum situation for specific techniques to be applied. This goal is best stated in the words of E. T. Jaynes, a pioneer in the field of maximum entropy spectral estimation research. In describing the importance of considering the problem to be solved prior to applying specific techniques he writes:

There are many different spectral analysis problems, corresponding to different kinds of prior information, different kinds of data, different kinds of perturbing noise, and different objectives. It is, therefore, quite meaningless to pass judgment on the merits of any proposed method unless one specifies clearly: "In what class of problems is this method intended to be used?" Today, programming and running a computer is much easier than actually thinking about a problem, so one may program an algorithm appropriate to one kind of problem, and then feed in the data of an entirely different problem. If the result is unsatisfactory, there is an understandable tendency to blame the algorithm and the method that produced it rather than the faulty application (Jaynes, 1982, p. 939).

Jaynes' argument is appropriate to the issue at hand, and a desired major end result of this study is a clear picture in the reader's mind of the capabilities of various hyperspectral analysis techniques. The signal processing approach will assist in the objective

evaluation of the theoretical concepts behind each technique and the circumstances in which the technique is best applied.

This study is organized in a manner that will facilitate the goal of an orderly approach to many different hyperspectral analysis techniques. Chapter II presents an overview of all currently known methods that can be applied to the analysis of hyperspectral imagery based on various user community paradigms. It also develops the historical context for these paradigms. Finally, the chapter narrows the scope of the study to those techniques with a signal processing flavor, and establishes a method of categorizing these techniques based on the amount of information available to the user at the start of the problem. Chapter III defines some basic statistical and linear algebra concepts using spectral imagery as illustrative examples. This chapter is important in that it introduces the mathematical foundation that underlies this study. The next four chapters address four major families of techniques that have been grouped according to *a priori* knowledge. Each of these chapters includes the broad concepts that motivated each technique as well as specific examples of the operation and applicability of each technique to real data sets. Chapter IV discusses the principal components family of techniques, Chapter V considers the matched filter family of techniques, Chapter VI studies the unknown background family of techniques, and Chapter VII examines the limited image endmember family of techniques. Chapter VIII is a summary of the results of the previous chapters. Chapter IX concludes the paper. It seeks to solidify the connections between specific families of techniques and emphasize the situations in which the techniques are most appropriate.

## II. BACKGROUND

### A. PROBLEM STATEMENT

This study concentrates on the application of hyperspectral imagery analysis techniques to the particular task of target detection. Some of the techniques considered have this as their original goal. Others have never been applied in this context. Before beginning an overview of all techniques, it is appropriate to define the problem at hand. The nature of hyperspectral data is such that the detection of a target in the image is best achieved by using the large amount of information inherent in the observed spectra. Thus, the problem is to localize the spectrum which is characteristic of the target material. Although this seems simple enough, there are a myriad of methods that can be applied to solve this problem. The reason for the multiplicity of methods is due to factors such as the amount of *a priori* information assumed, the view of the type of data, and the data model assumed. These factors dictate which approach is optimally suited to the particular task of target detection. The various approaches or strategies to the problem are highlighted in the next section.

### B. STRATEGIES FOR SPECTRAL IMAGERY ANALYSIS

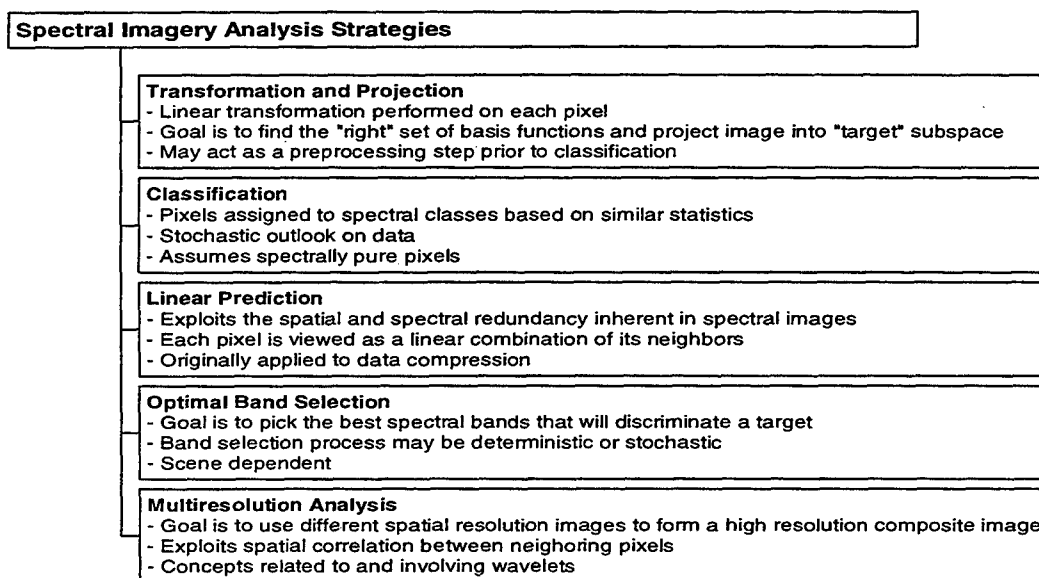


Figure 2.1: Taxonomy of Spectral Imagery Analysis Strategies.

In a broad survey of the pertinent literature, five major strategies are perceived into which the many methods for hyperspectral imagery target detection can be placed. Figure 2.1 illustrates these major strategies and the noteworthy aspects of each. In this context, the discussion of each strategy is simplified. The creation of this particular taxonomy is driven by four major determinants. The first is the model of the data. As mentioned above, the observed spectrum of each pixel recorded at the hyperspectral sensor can be viewed as a combination of multiple spectra within the spatial boundaries of the GIFOV. The individual spectra contributing to the observed spectrum are assumed to represent spectrally pure materials called endmembers and are assumed to mix in a linear fashion. When this model of the data is assumed, as is often the case with hyperspectral imagery, the data model is called the linear mixture model or mixed pixel model. An illustration of this concept is seen in Figure 2.2, where the imaged pixel is composed of three unique material types. The finite spatial resolution of the sensor is the

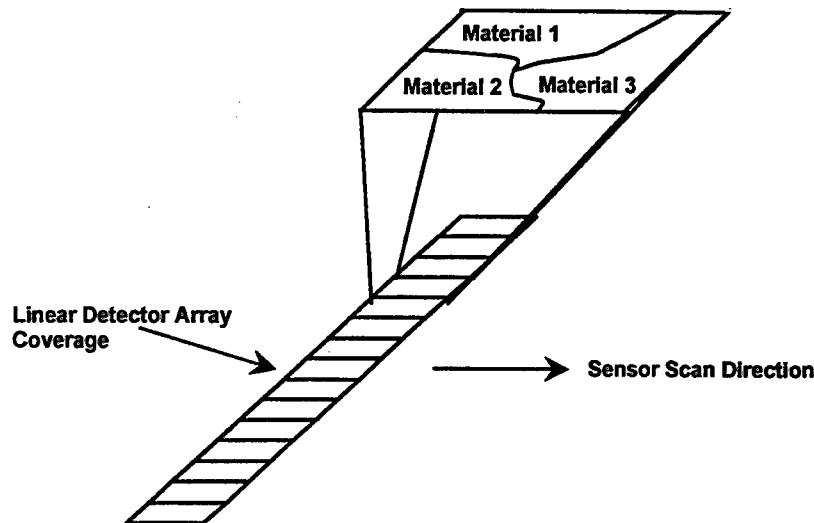


Figure 2.2: The Mixed Pixel Concept. After Harsanyi, 1993, p. 17.

cause of this situation. Each small square in Figure 2.2 represents the observed pixel on the ground. The expanded view of one of the pixels shows that it is composed of three different endmembers. The alternative to this model is the assumption that each observed pixel spectrum represents a unique material, which will be classified according to its own properties with respect to all other observed spectra in the scene. This model actually predates the mixed pixel model and is a simpler view of the matter. The second determinant of the type of strategy is the nature of the data. In considering hyperspectral

data, one can either view each observed pixel spectrum as a totally deterministic vector, a deterministic vector with additive random noise, or a random vector. It can be argued that the individual vector which represents the observed spectrum at each pixel is deterministic since it represents a unique material with its own peculiar absorption features. Likewise, the case can be made that the observed spectral vectors simply represent realizations of a random process because of the large amount of variability that one encounters when looking at various specimens of the same material. Figure 2.3 illustrates the high degree of variation that can exist in the spectra of the same material.

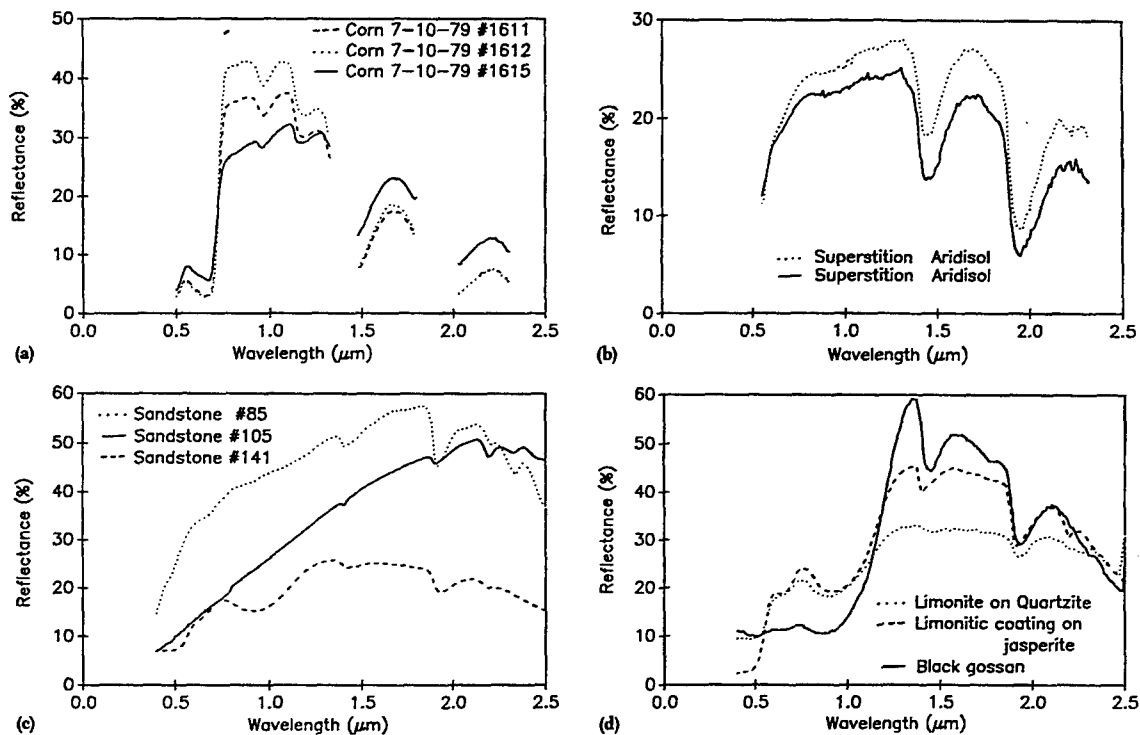


Figure 2.3: Spectral Variability Within Material Species. From Price, 1994, p. 183.

The spectra in Figure 2.3 were produced by field measurements using radiometers, and constitute a rough idea of laboratory spectra. Note that the high within-species variability makes deterministic identification of spectra a challenging matter, even with a spectral reference library. The third determinant of type of strategy is related to the scene itself. In a naturally occurring homogenous background, adjacent pixel spectra should have very similar statistical qualities. They could be expected to exhibit a high degree of correlation between neighbors. The occurrence of an object not belonging to the uniform

background would display a lower correlation with its neighbors. If no relation between neighbors is assumed, then each pixel spectrum is treated independently in the processing. Thus, whether or not the strategy assumes that such a relation between neighboring pixel spectra is a discriminant among strategies. The fourth determinant deals with the amount of *a priori* information provided at the outset of the problem. The *a priori* knowledge ranges from complete knowledge of the target and the background to no knowledge at all. The *a priori* knowledge categorization will be detailed in a subsequent section, as it is the most important distinguishing characteristic in the linear transform and projection strategy. It is important to note that these discriminants of the strategies do not produce mutually exclusive sets of strategies. In many cases, there is overlap, and the assignment of a particular technique to this strategy or another can be argued either way. This study seeks to be consistent in assigning techniques to strategies, and attempts to strictly follow the above discriminants as guidelines. Also, more than one strategy can be applied to the analysis of an image, and is often done in practice. The strategies may be viewed as building blocks which allow the user flexibility in implementation.

The first strategy, the techniques of which will be examined in detail by this study, is that of linear transformations and projections from the signal processing perspective. Data is visualized as belonging to either a signal subspace or a noise subspace, where a subspace is a linear algebra term which describes vectors with similar characteristics. It will be seen that the general approach is to project the observed image into a subspace where possible targets are easily discriminated. The key to the proper projection is having the right basis functions to construct a projection operator. The mixed pixel problem is assumed in most techniques of this strategy, and the statistics of the data, particularly the covariance matrix, play a major role in determining the proper basis functions for a projection operator. Each pixel is treated independently, as no assumption is made concerning the spatial arrangement of the pixel vectors. It should be noted that this strategy is often applied as a preprocessing step which aids in the later classification of image pixels using another strategy. This observation underscores the fact that these strategies do not necessarily have to be applied independently.

The second strategy in analyzing spectral imagery is a classification approach. It assigns observed pixel spectra to classes based on similar statistical characteristics. This strategy necessarily assumes a stochastic outlook of the data. The mixed pixel problem is not assumed, and the target spectrum is discriminated based on its membership in a class



separate from the background pixels. Techniques are further differentiated based on the need for training pixels from known classes and on the assumptions made about the statistical distribution of the pixels in each class.

The third strategy is based on the ideas of linear prediction. It assumes that each pixel vector is a linear combination of its neighbors, and seeks to exploit this relationship. The idea is to create a residual image in which there is less redundancy spectrally and spatially. Although applied to compression, this strategy has potential for target detection. Data is viewed statistically and modeled as such.

The fourth strategy is optimal band selection. The intent is to select the best original bands of the hyperspectral image that can be used to discriminate the target. The band selection process can be guided by a deterministic or statistical view of the target and background spectra. No explicit assumption about the linear mixing model is made.

The fifth strategy involves the use of multiresolution techniques. Concepts such as wavelets are used to pick out varying levels of detail from the image. The spatial correlation between neighbors is exploited, though a statistical outlook is not necessarily required.

### **C. AN OVERVIEW OF SPECIFIC METHODS WITHIN THE STRATEGIES**

This subsection provides the reader an overview of where specific techniques belong in the taxonomy of the above strategies. The techniques are described in only the briefest detail, with the pertinent references included for further research. The categorization of techniques within strategies is guided by the discussion of the previous subsection.

The transform and projection strategy has as the underlying assumption the mixed pixel problem in most cases. The segregation of techniques within this family is determined by the *a priori* knowledge that one has of the image endmembers. Figure 2.4 shows the breakdown of various techniques within this strategy. An endmember is defined as the spectrum associated with a pure material which is a constituent of the scene. The next four paragraphs address the techniques found within each family of the linear transformation and projection strategy.

If nothing is known about the data prior to processing, then the principal components analysis (PCA) family of techniques is best suited to the task. The basic

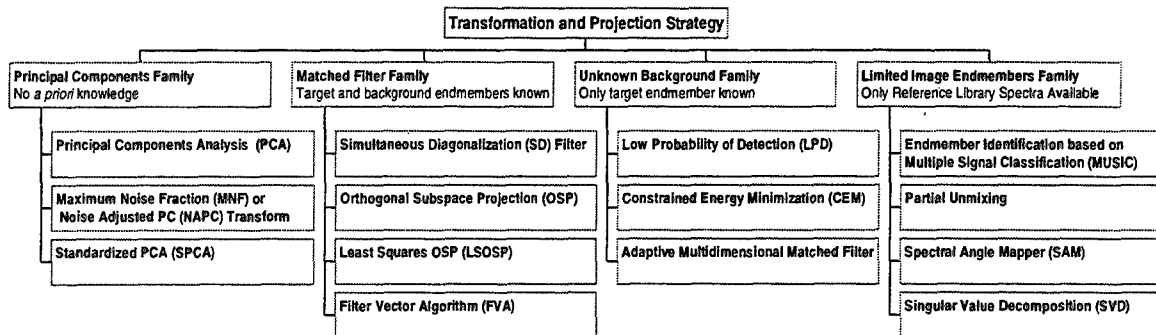


Figure 2.4: Hierarchy of Techniques in the Linear Transformation and Projection Strategy.

PCA is well described by Richards (1986). The noise adjusted principal components (NAPC) transform is a variant of PCA that orders images based on signal to noise ratio by whitening the additive noise and is described by Green, Berman, Switzer, and Craig (1988) and redefined by Lee, Woodyatt, and Berman (1990) as the maximum noise fraction (MNF) transform. Standardized principal components analysis is a technique that uses the standardized covariance matrix instead of the covariance matrix and is developed in Singh and Harrison (1985). Though the PCA does not explicitly assume the mixed pixel model, an interesting application by Smith, Adams, and Johnson (1985) allows the determination of the relative abundance of endmembers in each pixel spectrum using PCA. This technique represents a move towards the linear mixture model, and is important in that it sets the stage for the family of techniques which operates on no *a priori* knowledge except the presence of a reference library of endmember spectra.

The second major family of the linear transformation and projection strategy is generalized as the matched filter family because of the similarity to the signal processing concept of a matched filter. If all endmembers are known including the target, then the simultaneous diagonalization (SD) filter (Miller, Farrison, Shin, 1992), a special case of the matched filter, is applied. The SD filter was developed for a wide range of applications, of which spectral imagery analysis is a subset. A special case of the SD filter when the noise term is assumed to be of zero variance is the orthogonal subspace projection (OSP) first introduced by Harsanyi (1993). The OSP technique takes an *a priori* least squares approach to the data, and represents an extension of ideas from the array processing community. A further improvement of the OSP proposed by Tu, Chen, and Chang (1997) is the least squares OSP (LSOSP), which assumes an *a posteriori*

model of the data to improve the signal to noise ratio of the OSP technique. Filter vectors (Palmadesso, Antoniadou, Baumbach, Bowles, and Rickard, 1995) are a similar technique based upon the use of a matched filter to detect the target spectrum.

The third family of techniques is characterized by no *a priori* knowledge of the background endmembers. Only the target endmember is known. If the target signal is assumed to occur with a very low probability in the scene, then the low probability detection (LPD) technique (Harsanyi, 1993) can be applied. This technique is based on the concept of eigenfiltering. If the low probability of target occurrence is relaxed, then another technique also developed by Harsanyi (1993), called the constrained energy minimization (CEM) technique, may be employed. This technique is developed from the concept of beamforming in array processing. The matched filter can also be derived from a hypothesis test approach, which is more commonly associated with statistically based classification approaches. Stocker, Reed, and Yu (1990) derive such a matched filter which exploits spatial and spectral differences between a target and the background. Winter (1995) gives an interesting application of this spectral matched filter to the problem of hyperspectral mine detection.

The last family of techniques is one in which no *a priori* knowledge of any endmembers exists, but a spectral library or limited ground truth is available. If no knowledge of endmembers exists, then they may be estimated using a technique similar to the multiple signal classification (MUSIC) technique used in signal direction of arrival estimation problems. This technique, proposed by Harsanyi, Farrand, Hejl, and Chang (1994), employs elements of the OSP technique and principal components analysis, and requires a reference library of spectra. The spectral angle mapper (SAM) described by Yuhas, Goetz, and Boardman (1992) is a technique which treats spectra in a deterministic manner and attempts to measure the closeness of an observed pixel vector to one from a reference library. This technique does not assume mixed pixels. The partial unmixing technique developed by Boardman, Kruse, Green (1995) uses the ideas of convex sets to isolate the pixels most representative of pure endmembers in the scene, and then constructs a subspace which is orthogonal to the background endmembers in order to isolate the target. This technique builds upon the PCA endmember identification technique of Smith, Adams, and Johnson (1985). Another technique described by Danaher and O'Mongain (1992) and Herries, Selige, Danaher (1996) employs the linear algebra concept of the singular value decomposition (SVD) to efficiently estimate the abundance of a target material in the image. The technique requires ground truth in one

instance to create an operator, or key vector, which will isolate the target spectrum. Following the one time derivation of the key vector, no *a priori* knowledge of the image endmembers is required for subsequent applications.

The classification strategy is well documented in the literature of remote sensing data processing. It is a technique primarily used in two-dimensional photointerpretation and is an effective means of analyzing multispectral imagery. Figure 2.5 shows the various techniques within this strategy. The first discriminator of techniques in this family

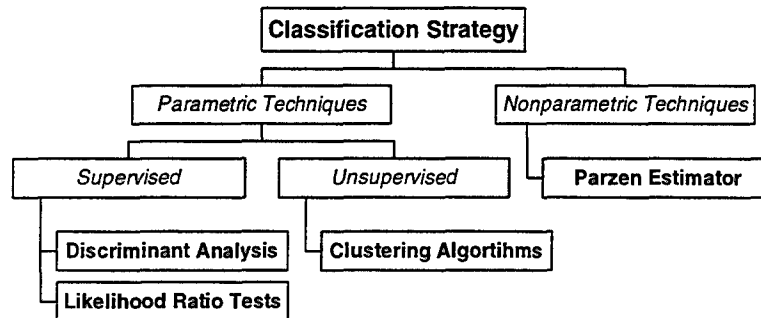


Figure 2.5: Hierarchy of Techniques Within the Classification Strategy.

is whether or not a form for the statistical model of the data is assumed. If no assumptions are made then the techniques are termed nonparametric. The Parzen estimator is a nonparametric method applied by Nedeljkovic and Pendock (1996) as a means of finding spectral anomalies in hyperspectral imagery. If the statistics of the data are assumed to be Gaussian, then classification techniques are termed parametric. A further division of parametric techniques is achieved by considering the availability of training pixels at the start of the problem. A training pixel is a pixel which is known to belong to a specific class. If training pixels are used, then the techniques are categorized as supervised classification techniques. Richards (1986) describes the most common supervised classification techniques, those based on maximum likelihood classification, in the context of multispectral imagery. These techniques are based on the statistical concept of Bayesian estimation which is described in detail with respect to two-dimensional image processing by Therrien (1989). Another technique called discriminant analysis, described by Fukunaga (1971) for application to pattern recognition, and Hoffbeck and Landgrebe (1996) for hyperspectral imagery, seeks to maximize the separation of classes. If no training pixels are used to assist in the classification process,

then the techniques are termed unsupervised classifiers. These techniques are described by Mather (1987) as clustering techniques which seek to classify pixels by the compactness of their groupings in multispectral space. Further discussions of classification techniques can be found in Swain and Davis (1978).

The linear prediction strategy seeks to capitalize upon the spatial and spectral redundancy inherent in hyperspectral imagery. The purpose of these techniques is to remove this redundancy for data compression. The techniques can take advantage of spatial redundancy as described by Therrien, Quatieri, and Dudgeon (1986) for application to two-dimensional image object detection. Other techniques in this strategy seek to reduce spectral redundancy as Rao and Bhargava (1996) describe. Techniques can also attempt to reduce the redundancy in both the spatial and spectral dimensions, as discussed by Wang, Zhang, and Tang (1995), and objectively evaluated by Roger and Cavenor (1996).

The optimal band selection strategy can be implemented by a technique introduced by Solberg and Egeland (1993) that uses Markov chain theory to select an optimal set of bands which is subsequently used for classification purposes. A different technique that has been proposed if all endmembers are known is to select the optimal bands to enhance the target signature using wavelets (Gorman, Subotic, Thelen, 1995).

The multiresolution strategy of hyperspectral imagery analysis is closely related to the concept of the wavelet transform. Burt (1992) has pioneered several applications of multiresolution techniques to the problems of image fusion and alignment. The application of these techniques to hyperspectral imagery analysis is described by Wilson, Rogers, and Meyers (1995). The techniques associated with multiresolution analysis lend themselves to the use of neural networks as tools in classifying hyperspectral images (Moon and Merenyi, 1995, p. 726).

#### **D. HISTORY**

In order to fully appreciate the significance of the hierarchy of hyperspectral imagery analysis strategies, a review of the historical perspective and paradigms in the analysis of hyperspectral images is necessary. Figure 2.6 illustrates the major image analysis paradigms over the past seventy years. This is by no means an all inclusive history, but rather a quick synopsis of the major ideas that led to the area specifically addressed in this study. The analysis of imagery began in the early part of this century

with photointerpretation. This analysis of aerial photographs to extract information of interest was a strictly human operation. The strength of the human element in interpretation was the ability to recognize large scale patterns (Richards, 1986, p. 75) and make inferences based on these patterns. The weakness of the human element was the inability to accurately quantify the results in a consistent manner. The computing power that began to become available in the 1960's and the ability to represent data in a digital fashion provided the impetus for automation of the photointerpretation task into digital imagery analysis. Here, the computer was programmed to work within narrow parameters, such as counting the number of occurrences of certain brightness values, a

**Photointerpretation (1930s -)**

- 2D Images
- good qualitative analysis (human)
- poor quantitative analysis



**Digital Imagery (1960s -)**

- 2D Images
- Pattern Recognition/ Computer Vision
- Emphasis on Classification Techniques



**Multispectral Imagery (1970s -)**

- 3D Images
- Principal Components Analysis
- Land Usage Classification



**Hyperspectral imagery (1980s -)**

- Need to reduce data dimensionality
- Software Packages with Spectral Libraries
- Need efficient processing techniques

Figure 2.6: Major Imagery Analysis Paradigms.

job that it performed better than any human analysts. The fields of pattern recognition and computer vision became important, and a statistical description of the data was needed to form the basis of classification schemes which could accurately determine the number of pixels in the scene belonging to a certain class. Linear prediction and principal components analysis (PCA) were tools that could assist in the automated detection of a target in the two-dimensional digital images. The advent of multispectral imagery with Landsat data in the 1970's added the spectral dimension to the problem of imagery

analysis. PCA played a significant role in reducing the dimensionality of the data and assisted in the classification of large land areas. The relationship between PCA techniques and classification techniques was a sequential operation, in that PCA was first applied to an image to remove the redundant information or create a better class separation and then a classifier was applied. This preprocessing application of PCA continues today. Improved classification techniques helped separate classes more consistently and accurately, but the majority of the techniques continued to be those found in pattern recognition disciplines. The 1980's and hyperspectral imagery ushered in a new challenge to the existing methods of analyzing data. Compression became an important concern. The search for new techniques to deal with the large amount of information and commensurate amount of redundancy prompted new views of the analysis paradigm. Ideas from the signal processing community provided a means of handling the large amount of data and confronting the mixed pixel problem. Software packages dedicated to the analysis of hyperspectral imagery incorporated spectral libraries, and found particular interest in the geological remote sensing community. The generation of innovative approaches and techniques is continuing as computing power increases.

#### **E. CREATION OF A TAXONOMY FOR THE LINEAR TRANSFORMATION AND PROJECTION STRATEGY**

As discussed above, the guiding rule for establishing the families within the linear transformation and projection strategy is the amount of *a priori* knowledge available to the user at the start of the problem. This is the analyst's perspective, which seeks to use the appropriate tool for the job. The four major divisions of *a priori* knowledge are: no knowledge of the scene endmembers, knowledge of all scene endmembers, knowledge of the target endmember only, limited knowledge of endmembers through a reference library or ground truth. The detailed discussion of these families of techniques constitutes the Chapters IV, V, VI, and VII. Each family is viewed as part of a hierarchy and has been categorized according to the taxonomy based on *a priori* knowledge. In the interest of a consistent approach, each family is treated using a common framework, which is reflected in the sections of each of the four chapters, and quickly described here. First, the techniques within a family are given a general description, which seeks to highlight some common major concepts. Second, the background concepts of the family are presented.

Statistical, linear algebra, and signal processing concepts are developed in detail to provide a good idea of the impetus for the specific techniques that follow. Third, the operation of the specific techniques are discussed and illustrated with examples. This framework is intended to concentrate all of the information required to fully understand a technique in one location. In an effort to enhance the basic concepts involved with the some of the statistical descriptions of the data, Chapter III discusses and illustrates some important definitions used frequently throughout this study.



### III. DEFINITIONS

An understanding of the fundamental ideas behind the various spectral imagery analysis techniques is important since it leads to the intelligent application of these techniques. The fundamental ideas involve concepts from statistics, linear algebra, and signal processing theory. Discussion of these ideas in the context of spectral imagery sets the stage for the detailed discussion of specific techniques that follow. This section presents multispectral and hyperspectral images as a means of further highlighting certain properties of the spectral concept. The images are also characterized from a statistics view, which assists in better understanding the image content and the statistical principles used in spectral imagery analysis techniques. Some concepts from linear algebra and signal processing are defined to provide a framework through which to understand certain spectral imagery analysis techniques. These perspectives offer a means of defining key concepts that appear throughout this study. An effort has been made to make these definitions simple yet comprehensive through the use of illustrative examples.

#### A. SPECTRAL IMAGERY

Spectral imagery is the acquisition of images at multiple wavelengths by spectrometers onboard aircraft or spacecraft. Two primary classes of such measurements are the traditional multispectral images, as with those produced by the thematic mapper (TM) radiometer on the Landsat satellites, and hyperspectral imagery, produced by imaging spectrometers in the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and Hyperspectral Digital Imaging Collection Experiment (HYDICE) systems. Typical images from Landsat and HYDICE data will be used here to introduce many of the concepts needed for this study. These data sets will also be used to illustrate analysis techniques in future sections. The Landsat TM scene is a seven-band 1000 x 1000 pixel image of Boulder, Colorado, made in August, 1985. The scene includes urban and mountainous areas. The presentation of the data as seven distinct image planes representing the various wavelength ranges is highlighted by Figure 3.1. The color version of this figure may be found in Appendix A. Notice how objects which appear bright in one band may appear dark in another band. The Flatiron Mountains of the Front Range, found in the left third of the image, illustrate this effect. Through this sort of contrasting effect, Landsat imagery offers a very basic means of discerning the spectral

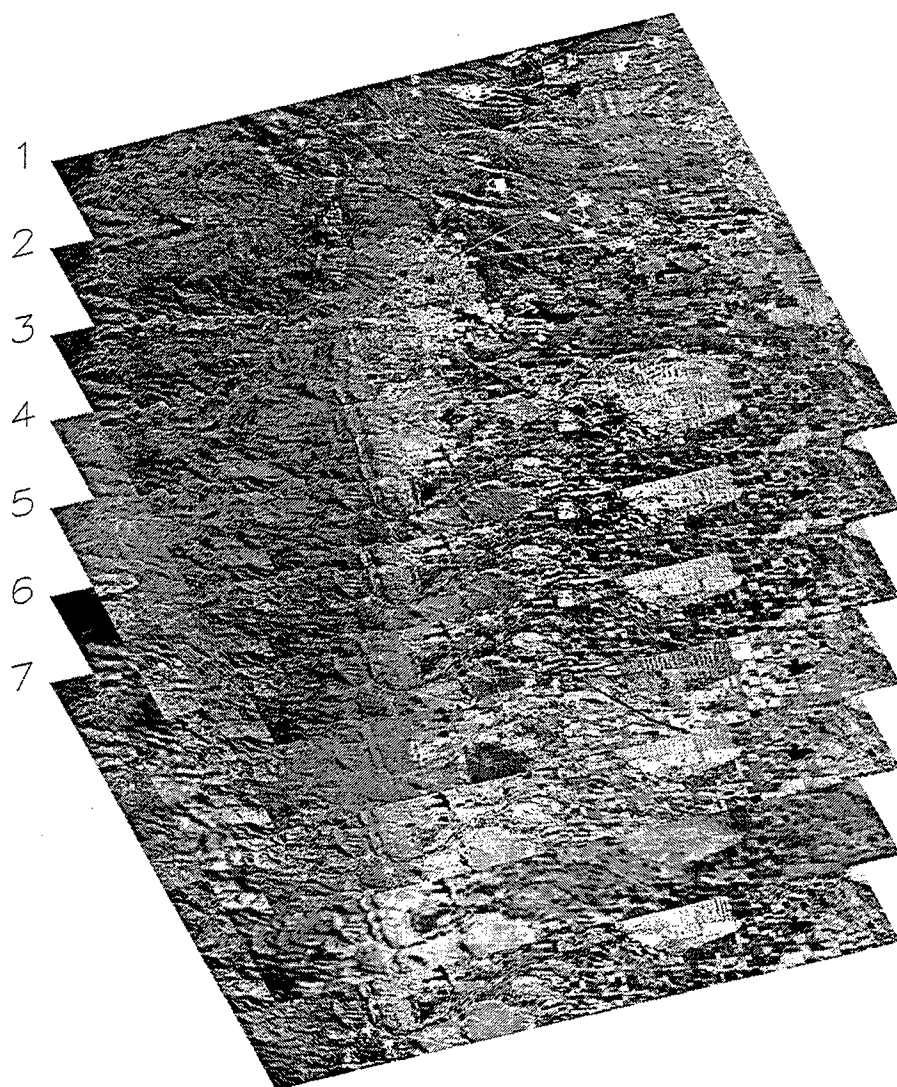


Figure 3.1: A Typical Multispectral Image Produced by Landsat TM.

character of a particular class of material.

A representative HYDICE scene was chosen from the FOREST RADIANCE I collect of Aberdeen Proving Grounds, MD, made in 1995 from a Convair CV-580 aircraft flying at 20,000 feet. The scene shows multiple vehicles parked in a field and treeline with roads running predominantly vertically through the scene. Figure 3.2 shows the hyperspectral image consisting of 320 samples, 320 lines, and 210 bands. A color version of this figure may be found in Appendix A. This image is a red,green,blue composite formed using bands 176 (2198.1 nm), 91 (1172.3 nm) and 31 (518.4 nm). One way of

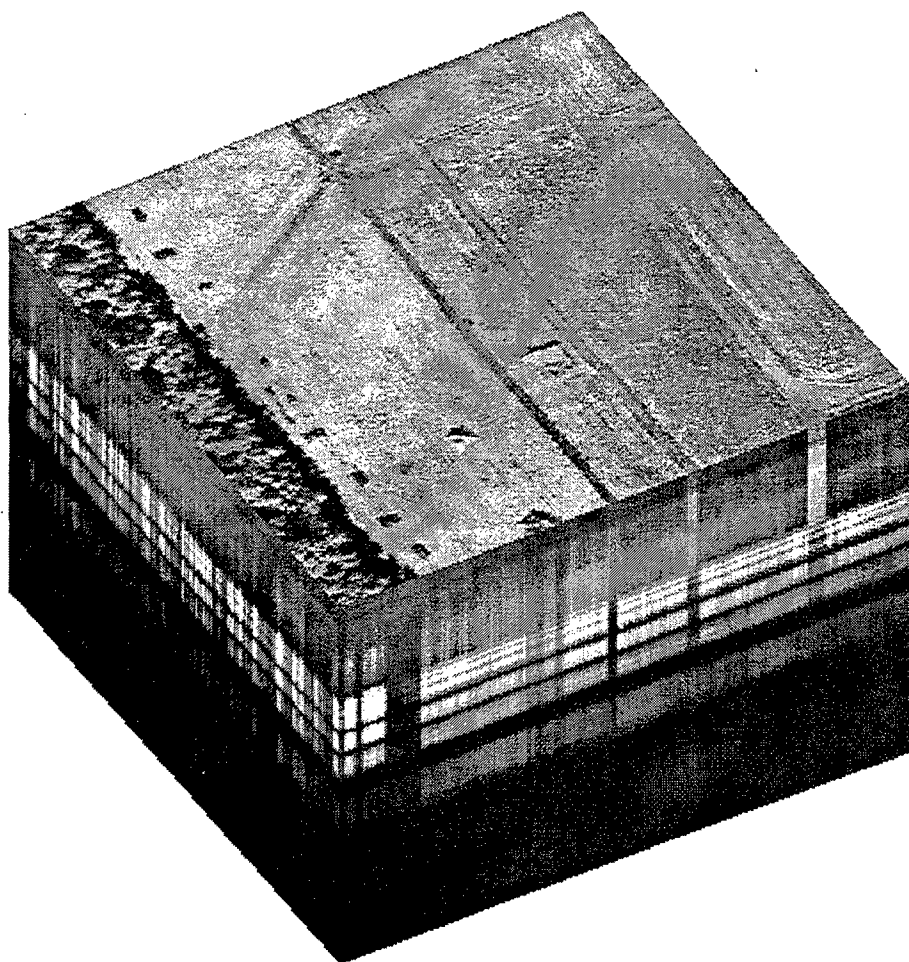


Figure 3.2: A Typical Hyperspectral Image Cube.

visualizing this type of data that has two spatial and one spectral dimension is as a cube. The vertical axis of the 'hypercube' represents the spectral response of individual spatial locations. The two dark bands which stretch horizontally across the spectral response faces of the hypercube correspond to atmospheric absorption bands. The ability to identify materials based on spectral detail is clearly more effective with hyperspectral imagery as opposed to multispectral imagery. As an example, note on the hypercube how the spectra associated with the road pixels appear clearly different than the spectra associated with the field or the trees. Figure 3.3 emphasizes the high spectral resolution of hyperspectral data by extracting information in the spectral dimension, or downward in

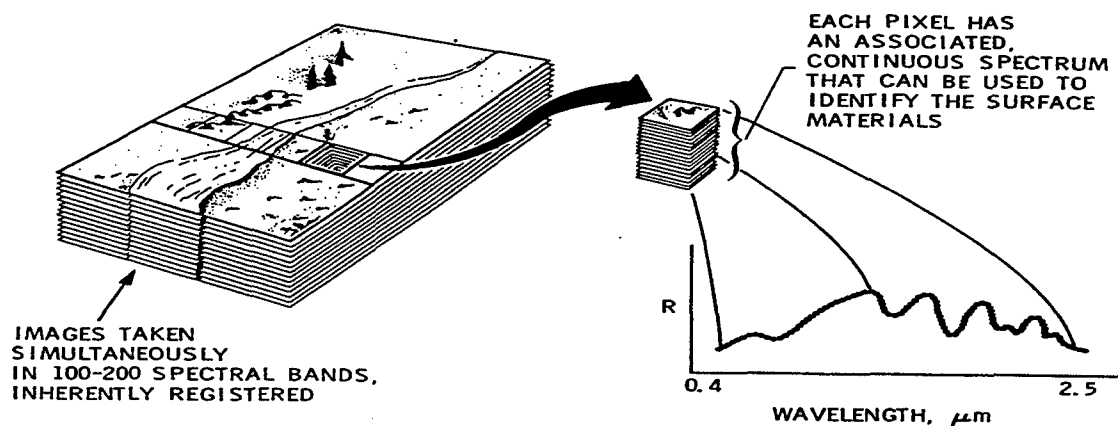


Figure 3.3: The Concept of a Pixel Vector. From Vane and Goetz, 1988, p. 2.

the axes of the cube. It shows the construction of an observed spectrum associated with a particular spatial location, called a pixel vector. The pixel vector is central to the discussion which follows, since the pixel vector may be viewed as a unique signal associated with a material of interest. Figure 3.4 further illustrates the pixel vector

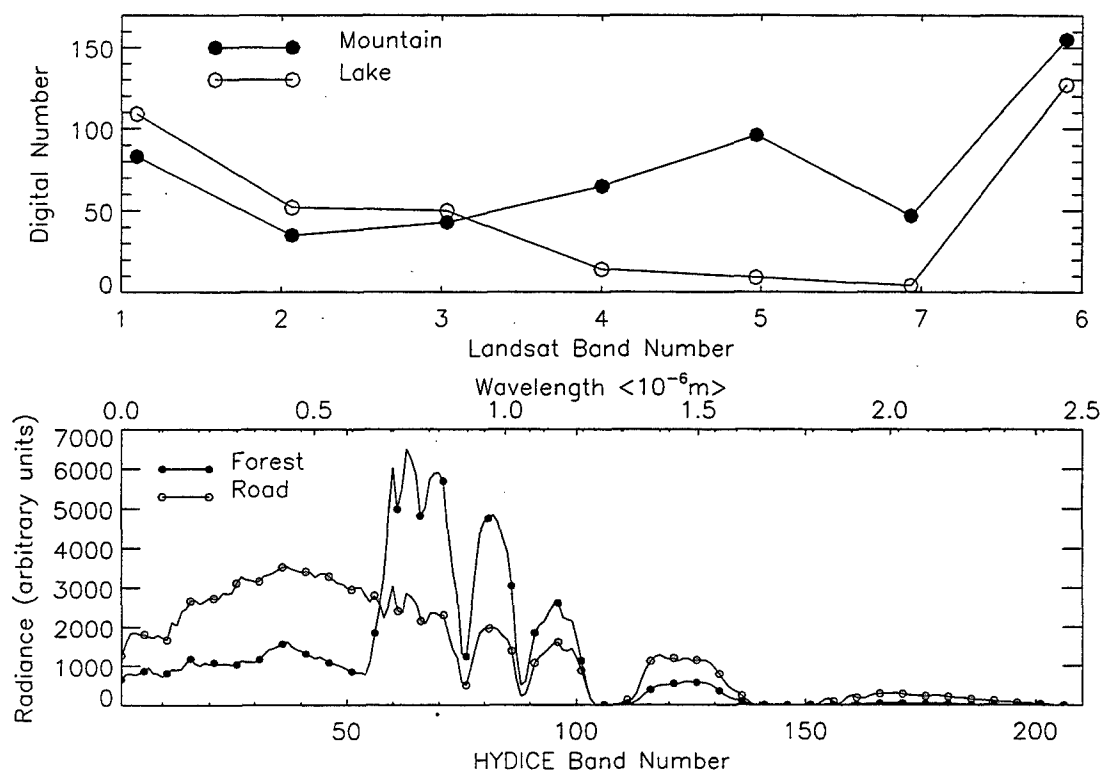


Figure 3.4: Typical Pixel Vectors From Multispectral and Hyperspectral Images.

concept using randomly chosen observed spectra from the Landsat and HYDICE images. The fine spectral detail that can be discerned in the hyperspectral image spectrum is a stark contrast to the coarse detail that comes from seven data points, as in the Landsat observed spectrum. Band seven precedes band six in the Landsat data to accurately reflect the corresponding wavelengths. The implication is that the characteristic shape of the pixel vectors obtained using hyperspectral imagery allows a more definitive identification of material based on unique spectral characteristics. Note also that the range of brightness values for the Landsat data is from zero to 255, corresponding to eight bit quantization of the data by the sensor. The HYDICE sensor has 12-bit quantization of the data.

In a hyperspectral sensor such as HYDICE, the spectral bands are configured to cover a range of 400 to 2500 nm. The observations of this reflected energy at the sensor are measured in terms of radiance, which has units of watts per square meter. A significant portion of the spectrum imaged in the HYDICE system is dominated by solar energy reflected from the earth's surface. This solar energy accounts for the characteristic "hump" in roughly the 50<sup>th</sup> to the 70<sup>th</sup> bands. At times, it is desirable to mitigate the effect of the dominant solar curve so that other spectral details may be discerned. One means of doing so entails converting radiance measurements to reflectance measurements by dividing the radiance observations by the scene average spectrum. Other methods include an offset based on in-scene brightness calibration points. The net effect is to normalize the radiance measurements in such a manner that the solar bias is removed and the resulting reflectance spectrum appears flatter. Figure 3.5 shows typical

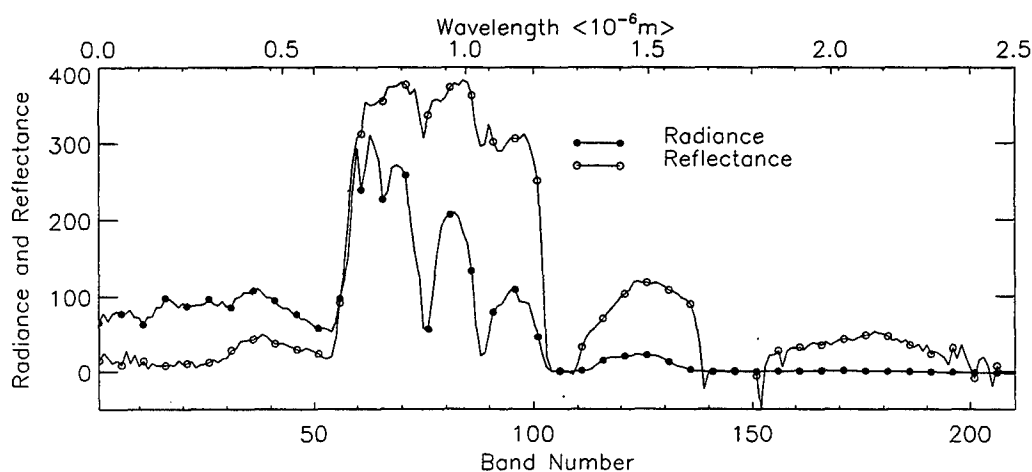


Figure 3.5: Radiance and Reflectance Spectra of Aberdeen HYDICE scene.

reflectance and radiance measurements for the same pixel of the HYDICE Aberdeen scene. The radiance data has been divided by a factor of ten in order to give it a dynamic range closer to that of the reflectance data. In spite of the scaling, note how the large peaks in the radiance data have been smoothed in the reflectance data. In both Figures 3.4 and 3.5, the wavelength range of the particular sensor has been annotated on the upper horizontal axis. This accentuates the fact that the HYDICE sensor employs narrower spectral bandwidths than does the Landsat TM.

## B. STATISTICAL INTERPRETATION

In order to assist in the quantitative discussion of characterizing the data statistically, we need to formally define the concept of the observed pixel vector. Assume that the observed pixel vector  $\mathbf{x}$  is a real valued random vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_l \end{bmatrix} \quad (3.1)$$

where the components  $\{x_1, \dots, x_l\}$  correspond to measured brightness values in each of  $l$  spectral bands. Since a stochastic view of the data assumes that these vectors are random entities, one means of characterizing them is to describe their behavior using statistical concepts. Exact statistical descriptions of their behavior are unavailable in real applications, so we must rely on methods that estimate the statistics of the observed random vectors. There are three major statistical definitions of interest in this respect. The first is the concept of expectation. The expectation of a random vector is called the mean or the average value that the random vector assumes, and is denoted as  $E\{\mathbf{x}\}$ . The mean is also called the first moment since it involves only the random vector itself and not products of the components of the vector  $\mathbf{x}$  (Therrien, 1992, p. 33). In using the observed data, is desirable that the statistical expectation of the estimated mean equal the actual mean. This is called an unbiased estimate of the mean. The framework for this estimation is to view the spectral image or scene as a collection of  $N$  random pixel vectors. This implies that the scene is comprised of  $N$  pixel vectors, each consisting of an  $l$ -band spectrum. The unbiased estimate of the mean spectrum for the scene is given by:

$$\mathbf{m} = \frac{1}{N} \sum_{j=1}^N \mathbf{x}_j = \begin{bmatrix} m_1 \\ \vdots \\ m_l \end{bmatrix} \quad (3.2)$$

where  $\mathbf{x}_j$  represents the spectrum of the  $j^{\text{th}}$  pixel of the scene. The mean spectrum vector,  $\mathbf{m}$ , of Equation 3.2 can also be interpreted as a  $l$ -dimensional vector with each component representing the average brightness value over the entire image for one particular band. Figure 3.6 illustrates the scene mean spectra for the Boulder Landsat TM and Aberdeen HYDICE images. It also shows the standard deviations for the Landsat image as error bars and the spectra of fifty randomly chosen pixel vectors for the HYDICE image as dots. These additional statistical characteristics of the data will be defined shortly. The important characteristic of Figure 3.6 with respect to the definition of mean spectra is the degree of similarity that exists between the mean spectra and the randomly chosen spectra of Figure 3.4. By the nature of its definition, the mean spectrum will appear to match the shape of the pixel vectors which occur most frequently in the scene.

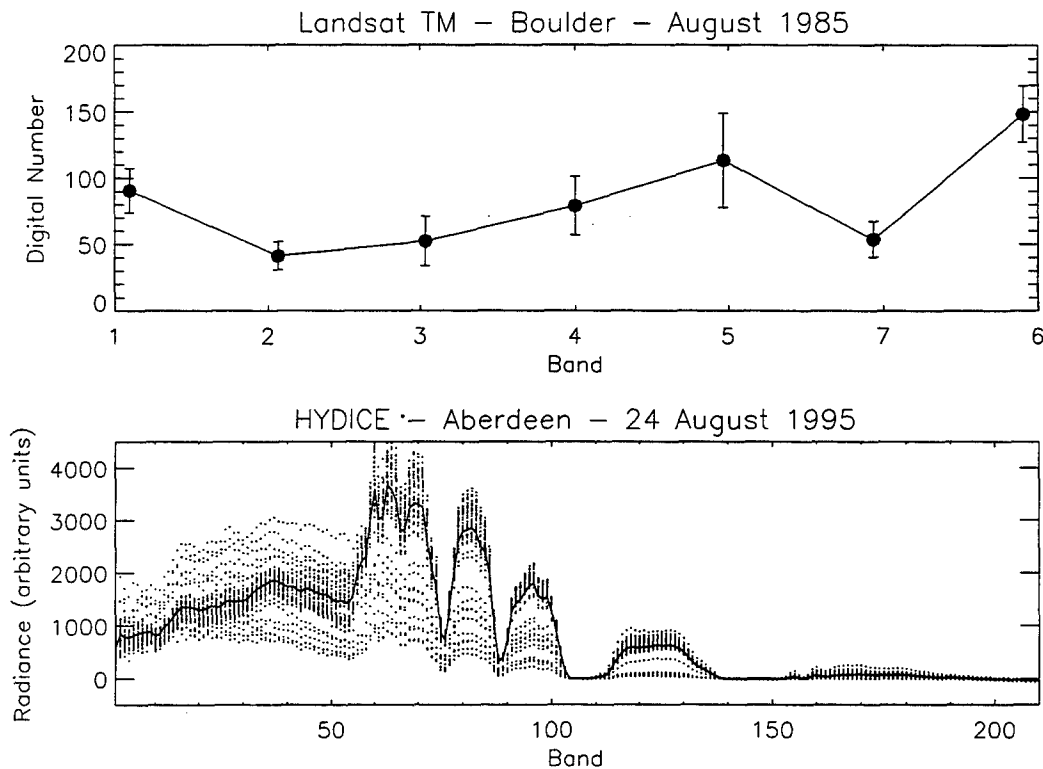


Figure 3.6: Mean Spectrum with One Standard Deviation of Landsat Image and Mean Spectrum with Representation of Variance of HYDICE Image.

The second definition of importance in characterizing random vectors is that of the covariance matrix. The covariance matrix is defined in vector and expanded component form as:

$$\Sigma_x = E\{(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T\} =$$

$$\begin{bmatrix} E\{(x_1 - m_1)^2\} & E\{(x_1 - m_1)(x_2 - m_2)\} & \cdots & E\{(x_1 - m_1)(x_l - m_l)\} \\ E\{(x_2 - m_2)(x_1 - m_1)\} & E\{(x_2 - m_2)^2\} & \cdots & E\{(x_2 - m_2)(x_l - m_l)\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{(x_l - m_l)(x_1 - m_1)\} & E\{(x_l - m_l)(x_2 - m_2)\} & \cdots & E\{(x_l - m_l)^2\} \end{bmatrix} \quad (3.3)$$

where  $\mathbf{m}$  is the mean vector of the entire image defined in Equation 3.2. The covariance matrix is symmetric and the elements of the main diagonal represent the variances associated with each of the component variables of the random vector  $\mathbf{x}$ . In the case of spectral imagery, the variance is a measure of how the brightness value of a particular band varies over all spatial image pixels. Figure 3.6 gives a rough idea of variance as the amount of distance between the mean spectrum and the 50 randomly chosen spectra plotted with it. It is also considered to be a measure of the power or contrast associated with each band. The off-diagonal elements are called the covariances, and measure how different variables vary with respect to each other. In the spectral sense, this is a measure of how much a band varies compared to another band over the image. When the covariance of two random variables is zero, then the random variables are said to be uncorrelated, which implies that those random variables were generated by separate random processes (Leon-Garcia, 1994, p. 337). The covariance matrix is the set of second central moments of the distribution, which are also referred to as moments about the mean since the mean component is subtracted from each random variable. The unbiased estimate of the covariance matrix is generated by:

$$\Sigma_x = \frac{1}{N-1} \sum_{j=1}^N (\mathbf{x}_j - \mathbf{m})(\mathbf{x}_j - \mathbf{m})^T \quad (3.4)$$

where  $\mathbf{x}_j$  is again the pixel vector associated with the  $j^{\text{th}}$  spatial location (Richards, 1986, p. 128). This is an outer product operation, which is performed  $N$  times, and is in a sense the average outer product of the vector  $\mathbf{x}_j - \mathbf{m}$ . In the calculation of the unbiased estimates of statistical quantities, the computational expense of the covariance matrix for a large number of samples,  $N$ , must be balanced with the desired degree of accuracy for the



estimate. More samples imply better estimates, and in order to ensure sufficient accuracy, the number of samples must be sufficiently large (Fukunaga, 1971, p. 242) .

The third statistical definition involves an issue that requires clarification regarding the term “correlation” matrix. In signal processing terminology, the correlation matrix stated as  $E\{\mathbf{xx}^T\}$  is formed exactly as the covariance matrix, except that the mean vector is not subtracted from the random vector  $\mathbf{x}$  (Therrien, 1992, p. 33). Figure 3.7 demonstrates the concept of mean removal using the scatter plots of two bands of Landsat data. The scatter plots are a representation of many two-dimensional random vectors which have a two-dimensional mean vector. The subtraction of this mean vector from every random vector results in a centering of the data about the origin. This introduces negative numbers into the previously positive data values. While the correlation matrix is more frequently used in signal processing where zero mean signals are the norm, remote sensing uses the covariance matrix since negative brightness values do not have a clear physical significance.

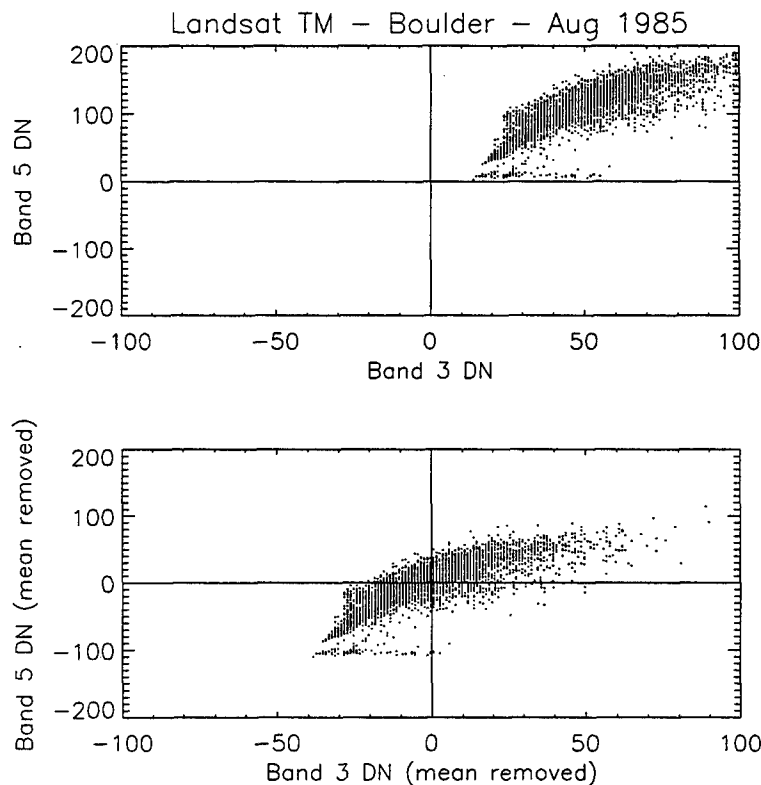


Figure 3.7: Mean Removal Illustrated With Scatter Plots.

In statistical and remote sensing applications, the correlation matrix is defined in terms of the covariance matrix. The  $ij^{\text{th}}$  element of the statistical version of the correlation matrix is:

$$\rho_{ij} = \frac{\sigma_{ij}^2}{\sqrt{\sigma_{ii}^2 \sigma_{jj}^2}} \quad (3.5)$$

where  $\sigma_{ij}^2$  is the covariance between bands  $i$  and  $j$  in  $\Sigma_x$ ,  $\sigma_{ii}^2$  represents the variance of the  $i^{\text{th}}$  band of data, and the square root of variance is defined as the standard deviation (Richards, 1986, p. 128). The statistical and signal processing versions of correlation do not produce the same matrix. The statistical definition produces a matrix which has a unit main diagonal and can be represented as:

$$\mathbf{R}_x = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1N} \\ \rho_{21} & 1 & \cdots & \rho_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N1} & \rho_{N2} & \cdots & 1 \end{bmatrix} \quad (3.6)$$

(Searle, 1982, p. 348). It is apparent that dividing the covariance matrix elements by their standard deviations has the effect of reducing all the variables to an equal importance since all have unit variance. The signal processing definition does not produce a unit diagonal matrix, though it is symmetric. The off-diagonal elements of  $\mathbf{R}_x$ , represented by  $\rho_{ij}$ , are called correlation coefficients. They range between -1 and +1 in value, and provide a measure of how well two random variables vary jointly by quantifying the degree of fit to a linear model (Research Systems, Inc., 1995, p. 20-6). A value near +1 or -1 represents a high degree of fit between the random variables to a positive or negative linear model, whereas a values near zero implies that the random variables exhibit a poor fit to the model. The conclusion that may be drawn is that a high degree of fit implies well-correlated random variables, whereas a correlation coefficient of zero is indicative of statistically orthogonal random variables. We will assume that we are dealing with the statistical definition of the correlation matrix, though a more descriptive term for the "correlation" matrix might be the "normalized" or "standardized" covariance matrix.

The definitions of statistical properties become clearer when they are linked to a physically observable phenomenon. The next few illustrations attempt to show the large amount of information revealed by the statistics of the data. Table 3.1 shows the covariance and correlation matrices for the Landsat data. In examining the Landsat

covariance matrix, we see that the highest variance results from band five, the lowest covariance is between bands four and six, and the highest covariance is between bands five and seven. The correlation coefficient is highest between bands one and two and is lowest between bands four and six. We can draw some conclusions from these statistics. First, band five has more variance, or contrast over the scene, than any other band. Before we assume that this means that band five can detect some sort of unique information better than other bands, we must ask if this variance was caused by signal coming from the ground or if it was noise introduced by our sensor or the atmosphere in that particular band. If we know the signal-to-noise ratio of our sensor in band five then we can answer the question. Signal-to-noise ratio (SNR) is the ratio of signal power to noise power, and

COVARIANCE MATRIX OF LANDSAT IMAGE:

BAND	BAND 1	BAND 2	BAND 3	BAND 4	BAND 5	BAND 6	BAND 7
1	276.6						
2	173.1	114.5					
3	293.3	194.7	350.3				
4	86.3	71.8	96.2	487.7			
5	407.3	282.2	534.9	340.2	1265.7		
6	124.7	77.4	146.9	36.6	318.5	181.3	
7	277.1	186.6	353.7	113.9	709.9	185.8	447.4

CORRELATION MATRIX OF LANDSAT IMAGE:

BAND	BAND 1	BAND 2	BAND 3	BAND 4	BAND 5	BAND 6	BAND 7
1	1						
2	.973	1					
3	.942	.972	1				
4	.235	.304	.233	1			
5	.688	.741	.803	.433	1		
6	.557	.537	.583	.123	.665	1	
7	.788	.824	.893	.244	.943	.653	1

Table 3.1: Covariance and Correlation Matrices of Landsat TM Image.

can be obtained using the variances as the power. Second, band four exhibits the lowest correlation coefficient when compared to all other bands. Again, before we assume that band four detects unique information, we must ask about the signal-to-noise characteristics of band four. For example, if band four were purely noise, then it would exhibit an even lower correlation with other bands, perhaps even zero. This is because it is independent of the other bands, not because it carries any information. A further explanation of these effects is seen in examining the histograms of the individual bands. Figure 3.8 shows the histograms of four of the Landsat bands. The histogram of band four indicates that an anomaly of some sort exists which places a sizable number of pixels at a lower brightness value than the rest. The “different” nature of band four brightness

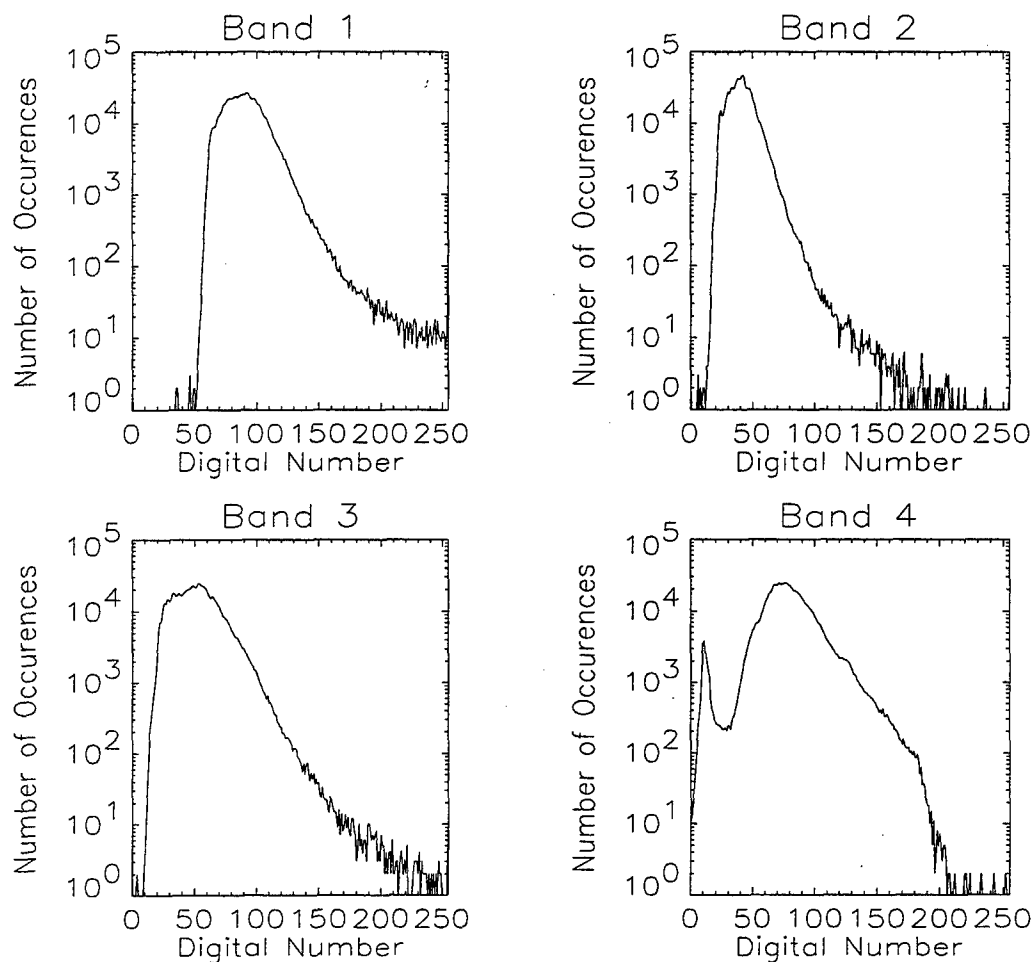


Figure 3.8: Histograms of Four of the Boulder Landsat TM Bands.

values accounts for the low correlation coefficient.

The scatter plot is another means of characterizing the statistics of the data by visually presenting the two-dimensional histogram using two selected bands. The scatter plot is a means of visualizing two of the seven dimensions of Landsat data and is shown two band combinations in Figure 3.9. It is a representation of all of the two-dimensional random pixel vectors formed by the two bands of interest. By plotting the data of one band against that of another, information regarding the statistical similarity of bands may be inferred. The scatter plots for the Landsat image show a definite linear feature when

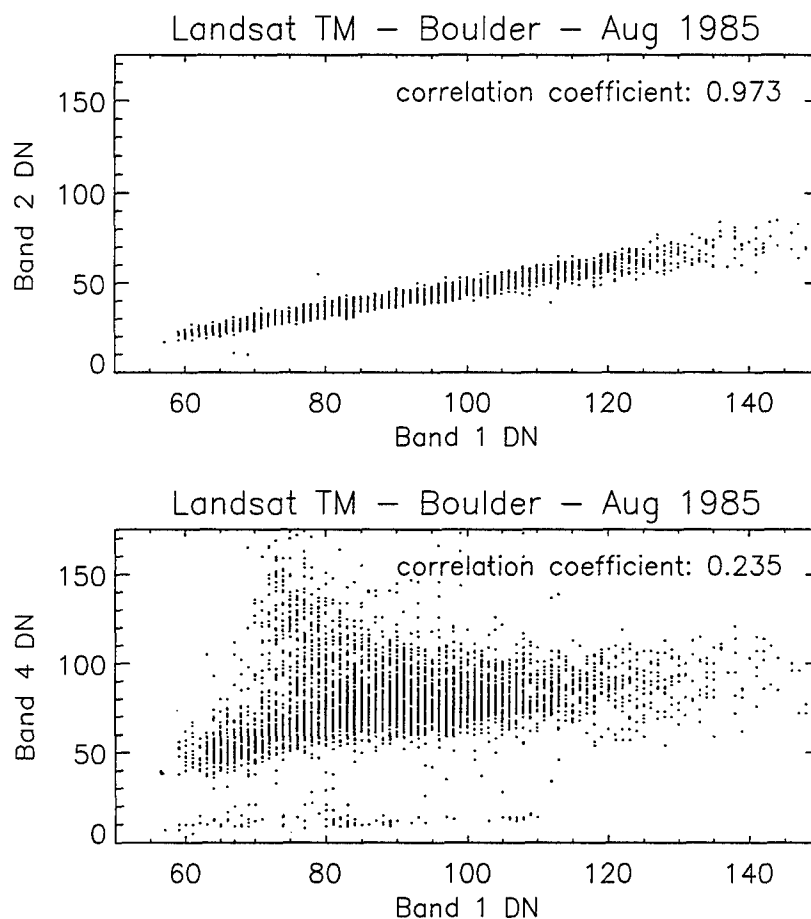


Figure 3.9: Scatter Plots of Boulder Landsat TM Data Showing Highly Correlated and Uncorrelated Band Combinations.

a high correlation coefficient exists, as between bands one and two. Thus, bands one and two are statistically similar, to the extent that there appears to be a near linear relationship between their random variables. The correlation coefficient of 0.973 substantiates this

observation. This is a sharp contrast to the more distributed shape for the scatter plot of band four data versus band one. This graphically depicts the independent and uncorrelated nature of the data in band four, as evidenced by the low correlation coefficient of 0.235. The scatter plot has also historically provided those involved in image classification with a method of grouping pixels with statistically similar characteristics into a statistical class. This can be seen in the Figure 3.9 scatter plot of band four and band one. The bottom of the plot reveals a smaller cluster of points away from the main body of points. This is an indicator that the pixels corresponding to these points belong to a different spectral class. In this case, these points are known to correspond to ground water that appears in the scene.

In order to show the second order statistics of a hyperspectral image, another visualization technique is introduced. With 210 bands, manually examining the covariance matrix would be tedious, and comparing two bands at a time with scatter plots would be similarly ineffective. For hyperspectral data statistics, the elements in the covariance matrices are assigned color values corresponding to their value. The result is a matrix which helps in explaining trends. Figure 3.10 illustrates the covariance and correlation matrices for both radiance and reflectance data in the HYDICE Aberdeen scene. A color version of this figure may be found in Appendix A. There are several notable features which are worth discussion in the four matrices. In the radiance covariance matrix, we see the effect of the sun on bands 50 to 70 manifested in the higher (redder) variance and covariance values. This is because the covariance matrix is constructed in a manner that uses the absolute radiance values, which are very large in these bands for radiance data. The correlation matrix of the radiance does not show this uneven weighting of variances. Instead, the correlation coefficients closest to the main diagonal exhibit a fairly similar value over all image bands, indicating that the correlation matrix has normalized the variances and covariances with respect to their standard deviations. The high values in the vicinity of the main diagonal are indicative of an important characteristic of hyperspectral imagery, namely the high correlation between adjacent bands. The covariance matrix of the reflectance data exhibits more distributed variances over the main diagonal than the radiance covariance matrix. This is due to the fact that the reflectance data has removed the bias of the sun from the data. The correlation matrix of the reflectance is in some sense the most unbiased estimate of the statistics since the effects of the sun and unequal variances have been eliminated. All of the matrices show the effects of the absorption bands as areas of very low covariances

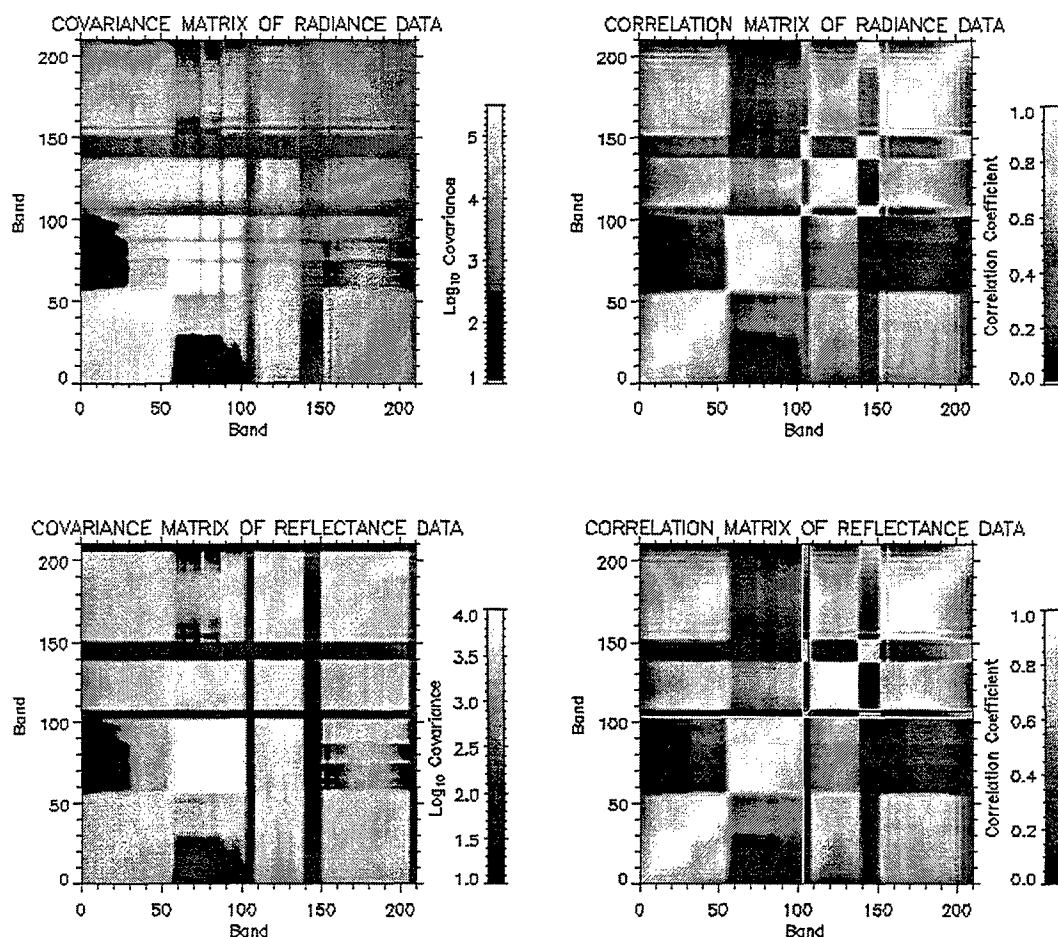


Figure 3.10: Second Order Statistics of the HYDICE Aberdeen Scene.

and correlation coefficients. This is intuitively pleasing, since the absorption bands should be very uncorrelated with all other bands. These dark vertical and horizontal features on the matrices represent the presence of atmospheric absorption features and are a good illustration of the effect of additive noise. The bands corresponding to these absorption features have had the “signal” drowned out by “noise” introduced by the atmosphere. Note also that the main diagonal or trace of these matrices represents the variance associated with each band.

The blocky, segmented nature of the second order statistics matrices reveals important details about the scene. The low covariances in the absorption bands are easily

explained because the brightness values in those bands are so statistically different than all other bands. More subtly, these matrices show the degree of difference or similarity between the brightness values in other parts of the observed spectra. In order to illustrate this concept, another HYDICE data set is introduced. This is a scene made during the DESERT RADIANCE collect in 1994. Figure 3.11 shows a red-green-blue false color composite image of Davis Monthan Air Force Base formed using bands 119 (1567.4 nm), 81 (1023.4 nm), and 57 (713.8 nm). Figure 3.11 is annotated with the different type aircraft that are found in the scene. The color version of this image may be found in Appendix A. The scene is a good contrast to the Aberdeen image because the predominant background material is sand instead of grass. Recalling the plots of various pixel vectors seen in Figure 3.4, note how the spectrum of the trees sharply spiked up at band 55 whereas the spectrum of the road remained smooth. This corresponds to a wavelength of about  $0.7 \mu\text{m}$ , and is referred to as the "infrared ledge". In Figure 3.10 note how a "block" of high covariances rapidly transitions to a "block" of low covariances at band 55. This feature is an indicator of the fact that there are significant differences in the spectral shapes of the observed pixel vectors which start at band 55. This can be interpreted to mean that the scene consists of both vegetation and non-vegetation pixel vectors. If the pixel vectors did not possess significantly different shapes, then this feature would not have manifested itself. Figure 3.12 shows such an instance, and the color version of it may be found in color in Appendix A. The Davis Monthan scene has predominantly sandy background, and as a result, the area between bands one and 100 appears to have high covariances and correlation coefficients without a sharp transition at band 55. The blocky appearance in the first hundred bands, evident when vegetation was present, is now not apparent.

While these observations are very cursory, they demonstrate how the statistics of the scene reveal a great deal of useful information. A more refined study of scene statistics, such as that pursued by the Rochester Institute of Technology, finds that the scene statistics can be used to differentiate urban and rural areas (Brower, Haddock, Reitz, and Schott, 1996, p. 56). This idea can be carried further to the problem of differentiating small man-made objects in a natural background. The challenge is that in



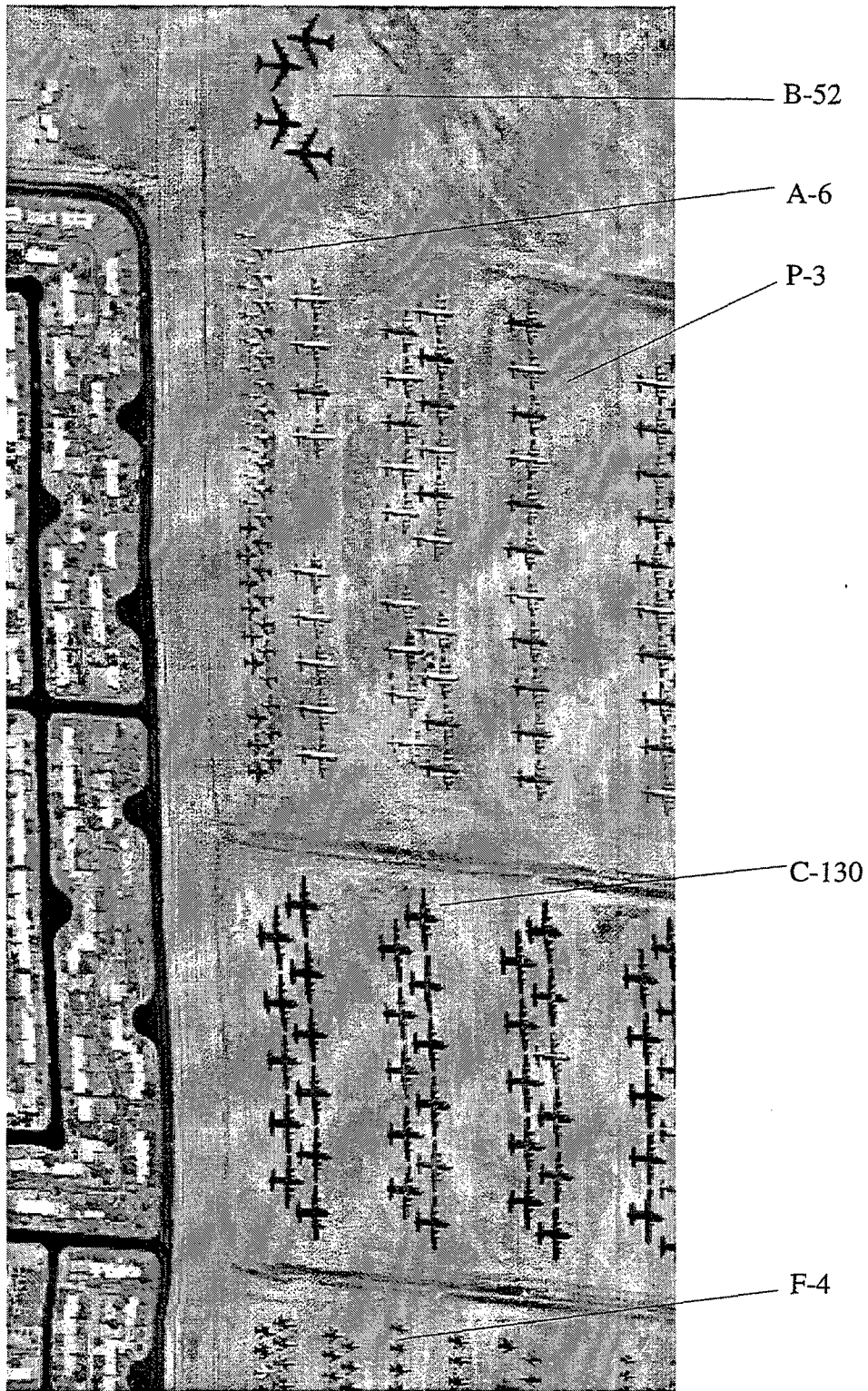


Figure 3.11: HYDICE Scene of Davis Monthan Air Force Base.

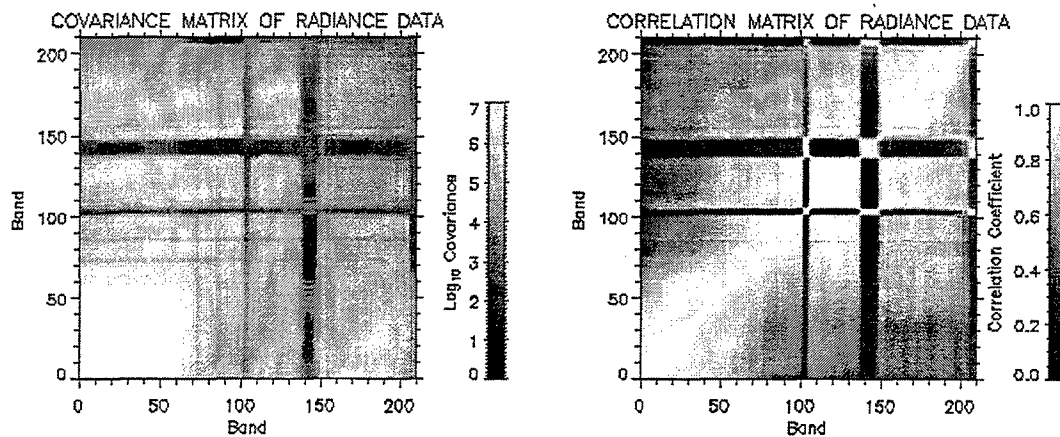


Figure 3.12: Davis- Monthan Radiance Covariance and Correlation Matrices.

order to be statistically significant, the target material must occur in many of the observed pixels. Considered independently, the scene statistics are interesting in that they provide further perspective and understanding into the nature of the scene. More importantly, they bring us closer to the target detection problem by setting the stage for an understanding of the techniques which use statistics to describe the background.

### C. RELATED SIGNAL PROCESSING AND LINEAR ALGEBRA CONCEPTS

#### 1. Linear Transformations of Random Vectors

The fundamental basis of the hyperspectral image analysis techniques addressed by this study is that of linear transformations. Our statistical definitions of the data using the covariance matrix and its standardized form, the correlation matrix, are important. Understanding the effect of a linear transformation on these matrices is also important. A linear transformation of a vector  $\mathbf{x}$  into a vector  $\mathbf{y}$  is accomplished by the matrix  $\mathbf{A}$  in the relation  $\mathbf{y} = \mathbf{A}\mathbf{x}$ . Figure 3.13 illustrates this concept using two-dimensional vectors. The transformation matrix  $\mathbf{A}$  rotates and scales the vector  $\mathbf{x}$  into the new vector  $\mathbf{y}$ . Since we are working with symmetric matrices in the second order moments of random vectors, we

may assume that  $\mathbf{A}$  is symmetric. The expectation operator is linear, which implies that the mean of the random vector  $\mathbf{x}$  is transformed as:

$$E\{\mathbf{y}\} = E\{\mathbf{A}\mathbf{x}\} = \mathbf{A}E\{\mathbf{x}\} \quad (3.7)$$

which can be restated as  $\mathbf{m}_y = \mathbf{A}\mathbf{m}_x$ , where the subscript on the mean vector denotes

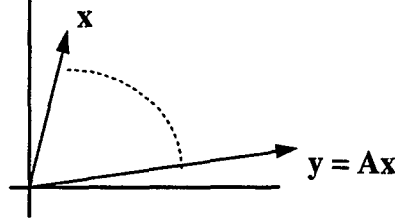


Figure 3.13: Linear Transformation of a Two-dimensional Vector.

which random vector the mean vector represents. Similarly, using the definition of the second order moment, the covariance matrix is transformed by the matrix  $\mathbf{A}$  so that

$$\Sigma_y = \mathbf{A}\Sigma_x\mathbf{A}^T \quad (3.8)$$

(Therrien, 1992, p. 45).

A particularly interesting and useful transformation is one which transforms a random vector,  $\mathbf{x}$ , into another random vector,  $\mathbf{y}$ , whose  $k^{\text{th}}$  and  $l^{\text{th}}$  components have the property of statistical orthogonality such that:

$$E\{y_k y_l\} = 0 \quad k \neq l \quad (3.9)$$

(Therrien, 1992, p. 50). The statistically orthogonal or uncorrelated random variables which result from such a transformation cause the transformed data covariance matrix to be diagonal. The means of achieving such a transformation which diagonalizes the covariance matrix is provided by the idea of eigenvectors and eigenvalues.

## 2. Eigenvectors and Eigenvalues

The eigenvalues of a  $l \times l$  matrix  $\mathbf{A}$  are the scalar roots of its characteristic equation, and are denoted as  $\{\lambda_1, \dots, \lambda_l\}$ . The nonzero vectors,  $\{\mathbf{e}_1, \dots, \mathbf{e}_l\}$  which satisfy the equation:

$$\mathbf{A}\mathbf{e}_k = \lambda_k \mathbf{e}_k \quad (3.10)$$

are called the eigenvectors of  $\mathbf{A}$ . Stated another way, an eigenvector defines a one-dimensional subspace that is invariant with respect to premultiplication by  $\mathbf{A}$  (Golub and

Van Loan, 1983, p. 190). In applying the above definitions of the eigenvalue and eigenvector to the  $l$ -band  $\times$   $l$ -band covariance matrix, we obtain:

$$\Sigma_x \mathbf{e}_k = \lambda_k \mathbf{e}_k \quad (3.11)$$

The covariance matrix in this relation may be viewed as a linear transformation which maps the eigenvector  $\mathbf{e}_k$  into a scaled version of itself (Therrien, 1992, p. 50). Because of the symmetry of the real covariance matrix, the  $l$  eigenvalues are guaranteed to be real (Searle, 1982, p. 274). It is also possible to find  $l$  orthonormal eigenvectors  $\{\mathbf{e}_1, \dots, \mathbf{e}_l\}$ , that correspond to the  $l$  eigenvalues (Therrien, 1992, p. 50).

### 3. Unitary Transformations

Suppose that the eigenvectors of the  $l \times l$  covariance matrix  $\Sigma_x$  are packed into a matrix  $\mathbf{E}$  as column vectors. Then, because of the orthonormality of the eigenvectors, the matrix  $\mathbf{E}$  transforms the covariance matrix in the following manner:

$$\mathbf{E}^T \Sigma_x \mathbf{E} = \begin{bmatrix} \leftarrow & \mathbf{e}_1^T & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{e}_l^T & \rightarrow \end{bmatrix} \Sigma_x \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{e}_1 & \cdots & \mathbf{e}_l \\ \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_l \end{bmatrix} = \Lambda \quad (3.12)$$

following the rules of linear transformations (Therrien, 1992, p. 45). The transformation matrix  $\mathbf{E}^T$  defines a linear transformation of a random vector  $\mathbf{x}$  into a random vector  $\mathbf{y}$ , by the relation

$$\mathbf{y} = \mathbf{E}^T \mathbf{x} \quad (3.13)$$

in which the covariance matrix of  $\mathbf{y}$  is a diagonal matrix represented by  $\Lambda$ . This diagonalization of the covariance matrix  $\Sigma_x$  is another manner of stating that the components of random vector  $\mathbf{y}$  are now uncorrelated since all off-diagonal elements of  $\Lambda$  are zero. The orthonormal columns of  $\mathbf{E}$  imply that the transformation matrix  $\mathbf{E}^T$  represents a unitary transformation defined by:

$$\mathbf{E}^T \mathbf{E} = \mathbf{E} \mathbf{E}^T = \mathbf{I} \quad (3.14)$$

(Therrien, 1992, p. 51).

### 4. A Geometric Interpretation of the Unitary Transform

If we assume that our data has a Gaussian distribution, then we can describe its probability density function (pdf) with a family of ellipsoids as:

$$(\mathbf{x} - \mathbf{m}_x)^T \Sigma_x^{-1} (\mathbf{x} - \mathbf{m}_x) = \text{constant} \quad (3.15)$$

Because the matrix  $\mathbf{E}$  is orthonormal, the implication is that the eigenvectors of  $\Sigma_{\mathbf{x}}$  are the same as those of its inverse, and the eigenvalues of  $\Sigma_{\mathbf{x}}^{-1}$  are simply the reciprocals of those of  $\Sigma_{\mathbf{x}}$  (Jolliffe, 1986, p. 14). Thus, the inverse transformation may be written as

$$\mathbf{x} = \mathbf{E}^T \mathbf{y} \quad (3.16)$$

and the equation defining the contours of constant density may be rewritten as:

$$(\mathbf{x} - \mathbf{m}_{\mathbf{x}})^T \mathbf{E} \mathbf{\Lambda}^{-1} \mathbf{E}^T (\mathbf{x} - \mathbf{m}_{\mathbf{x}}) = (\mathbf{y} - \mathbf{m}_{\mathbf{y}})^T \mathbf{\Lambda}^{-1} (\mathbf{y} - \mathbf{m}_{\mathbf{y}}) = \sum_{k=1}^l \frac{|y_k - m_k|^2}{\lambda_k} = \text{constant} = C \quad (3.17)$$

which is the equation for an ellipse with the principal axes of the ellipse being aligned with the eigenvectors and the magnitudes proportional to  $\lambda_k^{1/2}$  (Jolliffe, 1986, p. 19). This geometrically illustrates the role that eigenvalues and eigenvectors play in the unitary transform. Figure 3.14 shows that the unitary transformation is equivalent to a rotation of the coordinate axes. The tilt of the ellipse with respect to the original coordinate system is indicative of the fact that correlation exists between the original vector components (Therrien, 1992, p. 59). In the new coordinate system defined by the unitary transform,

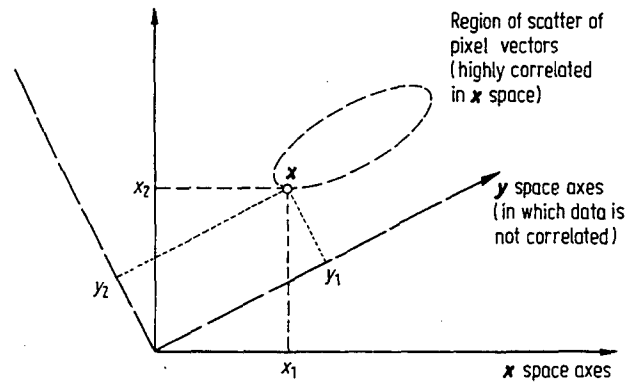


Figure 3.14: The Unitary Transformation as a Rotation of Axes.  
From Richards, 1986, p. 131

the axes of the ellipse are parallel to the new axes, showing that the vector components are indeed uncorrelated in this coordinate system. Although the assumption was made that the data was Gaussian, this concept of two-dimensional ellipsoids is a useful one in understanding the workings of the transformation. In this context, the scatter plots of the Landsat data are useful in portraying a rough idea of the distribution of the probability density function of the random vectors.

## 5. Simultaneous Diagonalization of Two Covariance Matrices

Often times, we cannot make the assumption of additive white Gaussian noise of equal SNR in all bands. At such times, in order to pose the problem in terms of a signal in white noise, we employ a diagonalization technique that transforms the noise covariance matrix into the identity matrix, or in effect whitens it. This transformation is referred to as the whitening transformation (Therrien, 1992, p. 60). It is assumed that the noise can be characterized by a covariance matrix  $\Sigma_N$  and the signal by the covariance matrix  $\Sigma_S$ , depicted as Gaussian ellipses in Figure 3.15(a). The transformation begins with the diagonalization of the noise covariance matrix  $\Sigma_N$  by a unitary transformation created from the noise covariance eigenvalues packed in the matrix  $\Lambda_N$  and eigenvectors packed in the matrix  $\mathbf{E}_N$  as follows:

$$\Sigma_N = \mathbf{E}_N \Lambda_N \mathbf{E}_N^T \quad (3.18)$$

The whitening transformation is formed as:

$$\mathbf{y} = (\Lambda_N^{-1/2} \mathbf{E}_N^T) \mathbf{x} \quad (3.19)$$

where the matrix  $\Lambda_N^{-1/2}$  is the inverse of the diagonal matrix  $\Lambda_N^{1/2}$ , whose diagonal elements are the square roots of the eigenvalues (Therrien, 1992, p. 60). The whitened covariance matrix is formed by applying the whitening transform to the original  $\Sigma_N$ :

$$\Sigma_{WN} = (\Lambda_N^{-1/2} \mathbf{E}_N^T) \Sigma_N (\Lambda_N^{-1/2} \mathbf{E}_N^T)^T = \mathbf{I} \quad (3.20)$$

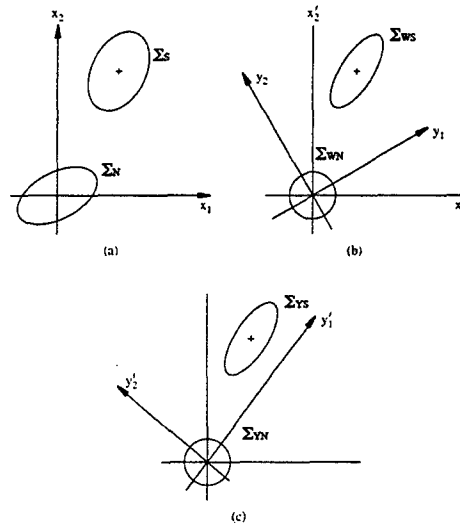


Figure 3.15: Simultaneous Diagonalization of Signal and Noise Covariance Matrices.  
After Therrien, 1992, p. 61.

The whitening transformation is also applied to the signal covariance matrix to yield the new signal covariance:

$$\Sigma_{ws} = (\Lambda_N^{-1/2} \mathbf{E}_N^T) \Sigma_s (\Lambda_N^{-1/2} \mathbf{E}_N^T)^T = \Lambda_N^{-1/2} \mathbf{E}_N^T \Sigma_s \mathbf{E}_N \Lambda_N^{-1/2} \quad (3.21)$$

The important effect of this transformation is to rotate the coordinate system and scale the noise covariance matrix to the identity matrix as shown in Figure 3.15(b). The final step in simultaneous diagonalization entails diagonalizing the noise-whitened signal covariance matrix. It is depicted in Figure 3.15(c), and accomplished by finding the eigenvectors and eigenvalues of  $\Sigma_{ws}$  and packing them into the matrices  $\mathbf{E}_{ws}$  and  $\Lambda_{ws}$ . The unitary transformation which represents the last step of simultaneous diagonalization is applied as:

$$\mathbf{y}' = \mathbf{E}_{ws}^T \mathbf{y} \quad (3.22)$$

The transformed signal covariance matrix is then:

$$\Sigma_{ys} = \mathbf{E}_{ws}^T \Sigma_{ws} \mathbf{E}_{ws} = \Lambda_{ws} \quad (3.23)$$

and the transformed whitened noise covariance matrix is still the identity matrix because of the nature of the unitary transform:

$$\Sigma_{yn} = \mathbf{E}_{ws}^T \mathbf{I} \mathbf{E}_{ws} = \mathbf{I} \quad (3.24)$$

Note that in Figure 3.15, the labeling of the coordinate axes corresponds to the two scalar components of the vectors  $\mathbf{x}$ ,  $\mathbf{y}$ , and  $\mathbf{y}'$ . The simultaneous diagonalization of the noise and signal covariance matrices is sometimes written as a one step transformation:

$$\mathbf{z} = (\mathbf{E}_N \Lambda_N^{-1/2} \mathbf{E}_{ws}) \mathbf{x} \quad (3.25)$$

The simultaneous diagonalization technique lies at the heart of several hyperspectral imagery analysis techniques.





## IV. THE PRINCIPAL COMPONENTS ANALYSIS FAMILY OF TECHNIQUES

### A. DESCRIPTION

Principal components analysis (PCA) as applied in multispectral and hyperspectral remote sensing is an analytical technique based on the linear transformation of the observed spectral axes to a new coordinate system in which spectral variability is maximized. The impetus for such a transformation is the high correlation that exists between adjacent bands in spectral imagery. The spectral overlap of the sensors and the wide frequency range of the energy reflected from the ground account for this high correlation (Rao and Bhargava, 1996, p. 385). This implies that a great deal of spectral redundancy exists in the data. The principal components transformation decorrelates the information in the original bands and allows the significant information content of the scene to be represented by a smaller number of new bands called principal components. The transformation effected by the PCA is a unitary transformation and is graphically depicted in Figure 4.1 as operating on observed

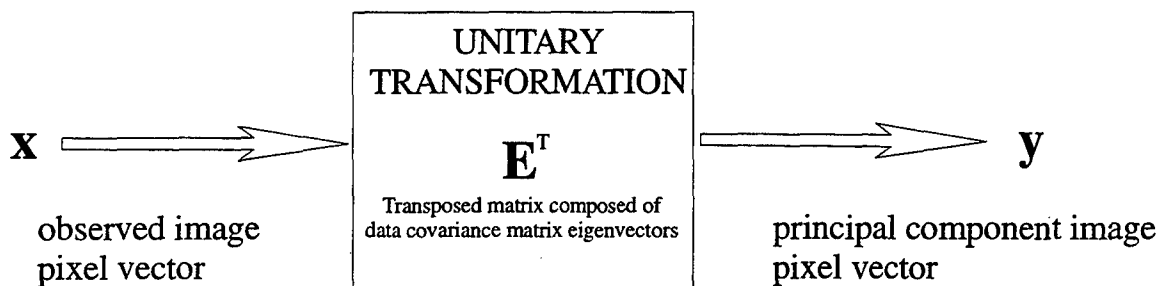


Figure 4.1: PC Transformation Depicted as a Linear Transformation.

pixel vectors to produce new pixel vectors with uncorrelated components. This basic linear transformation lies at the heart of the PCA family of techniques. Two immediate applications of the principal components transformation are data compression and information extraction. In the problem of target detection, the latter is of considerable interest. The PCA family of techniques is based exclusively on the statistics of the observed variables, requiring no *a priori* deterministic or statistical information about

the variables in the image. The techniques that are included in this family of techniques are the basic PCA, the noise adjusted principal components (NAPC) or maximum noise fraction (MNF) transform, and the standardized PCA (SPCA). The PCA family of techniques serves as a building block in the formulation of more elaborate analysis techniques, and is fully explored in this chapter to emphasize its importance.

## **B. BACKGROUND DEVELOPMENT**

Principal components analysis is an extremely versatile tool in the analysis of multidimensional data. In tracing the historical roots of this technique, it is clear that it is based upon ideas drawn from the fields of statistics and linear algebra. The mathematical underpinnings of PCA deal with the diagonalization of the covariance matrix of the data by unitary transform. This diagonalization is accomplished through an eigendecomposition of the covariance matrix to form a unitary transform and serves as a bridge between matrix algebra and stochastic processes (Haykin, 1996, p. 187). The wide applicability of PCA is due to the fact that it assumes a stochastic outlook of the data, which is fundamental to the analysis of data in many scientific disciplines. We will investigate the views of three disciplines which employ PCA to better understand some of the mechanics of this seemingly simple transformation. The three views are those of multivariate data analysis, signal processing, and pattern recognition. A thorough understanding of the ideas that motivate the PCA will assist in understanding why it is such a commonly used technique in remotely sensed imagery analysis, and when it is most appropriately applied.

### **1. Multivariate Data Analysis View**

PCA was described by Pearson in 1901 and introduced as the Hotelling transform in 1933 by Hotelling for application in educational psychology (Singh and Harrison, 1985, p. 884). Hotelling's goal was to find a fundamental set of independent variables of smaller dimensionality than the observations that could be used to determine the underlying nature of the observed variables (Hotelling, 1933, p. 417). In many scientific experiments, the large number of variables makes the problem of determining the relative importance of specific variables intractable. Hotelling's method makes the problem manageable by discarding the linear combinations of variables with small variances, and studying only those linear combinations with large variances. Since the

important information in the data is usually contained in the deviation of the variables from a mean value, it is logical to seek a transform which provides a convenient means of identifying the combinations of variables most responsible for the variances (Anderson, 1984, p. 451). The linear combination of the original variables which behave sufficiently similarly are combined into new variables called principal components. In this context, principal components analysis studies the covariance relationships within a data set by investigating the number of independent variables, and identifies the natural associations of the variables.

Mathematically represented, each principal component is a scalar formed by a linear combination of the elements of the observed random vector  $\mathbf{x}$ , where each vector component corresponds to a random variable. The principal components are constructed in such a manner as to be uncorrelated with all other principal components and ordered so that variance is maximized (Jolliffe, 1986, p. 2). The  $k^{\text{th}}$  principal component is obtained by multiplying the transposed  $k^{\text{th}}$  eigenvector of the covariance matrix of  $\mathbf{x}$  by the data vector  $\mathbf{x}$ , as depicted in the equation

$$y_k = \mathbf{e}_k^T \mathbf{x} \quad (4.1)$$

The  $k^{\text{th}}$  principal component is also called a score, and the components of the eigenvector are called loadings because they determine the contribution of each original variable to the principal component. Generalizing the scalar result of Equation 4.1 to a vector result:

$$\mathbf{y} = \mathbf{E}^T \mathbf{x} \quad (4.2)$$

we obtain a vector of  $l$  principal components when we take the product of all of the transposed eigenvectors of  $\Sigma_{\mathbf{x}}$  and the data vector,  $\mathbf{x}$ .

While the property of the unitary transform to produce new uncorrelated variables has been previously discussed, the property of the unitary transform to maximize the variance, which is central to the PCA, merits further discussion. The best illustration of this property is the algebraic derivation of the PCA. The goal is to maximize the variance of the first principal component, denoted as  $\text{VAR}[y_1]$  or  $\text{VAR}[\mathbf{e}_1^T \mathbf{x}]$ . By the definition of variance as a second order moment, this is equivalent to maximizing  $\mathbf{e}_1^T \Sigma_{\mathbf{x}} \mathbf{e}_1$ , where the eigenvectors are orthonormal, so that  $\mathbf{e}_1^T \mathbf{e}_1 = 1$ . The method of Lagrange multipliers is employed so that the expression to be maximized is differentiated with respect to the eigenvector and set equal to zero as:

$$\frac{\partial}{\partial \mathbf{e}_1} [\mathbf{e}_1^T \Sigma_{\mathbf{x}} \mathbf{e}_1 - \lambda (\mathbf{e}_1^T \mathbf{e}_1 - 1)] = 0 \Rightarrow (\Sigma_{\mathbf{x}} - \lambda \mathbf{I}) \mathbf{e}_1 = 0 \quad (4.3)$$

In Equation 4.3,  $\lambda$  is a Lagrangian multiplier in the left hand expression and corresponds to the largest eigenvalue of  $\Sigma_x$  in the right hand expression, and  $e_1$  is the eigenvector corresponding to the largest eigenvalue (Jolliffe, 1986, p. 4). Thus, the eigenvalues of  $\Sigma_x$  represent the variances of the principal components, and are ordered from largest to smallest magnitude. If the original variables have significant linear intercorrelations, as spectral imagery does, then the first few principal components account for a large part of the total variance. (Singh and Harrison, 1985, p. 883).

## 2. Signal Processing View

In the analysis of random signals, the key is to have a set of basis functions that also make the components of the signal statistically orthogonal or uncorrelated (Therrien, 1992, p. 173). The Karhunen-Loeve Transform (KLT) was introduced in 1947 for the analysis of continuous random processes, and is developed here in its discrete form, the DKLT. It is the same unitary transform previously presented, but is posed to solve the problem from a different perspective. The motivation for the DKLT is actually an expansion, best seen by Figure 4.2, which shows a discrete observed signal as a weighted sum of basis functions, which are in fact the eigenvectors of the covariance matrix. The observed pixel vector spectrum may be thought of as a discrete

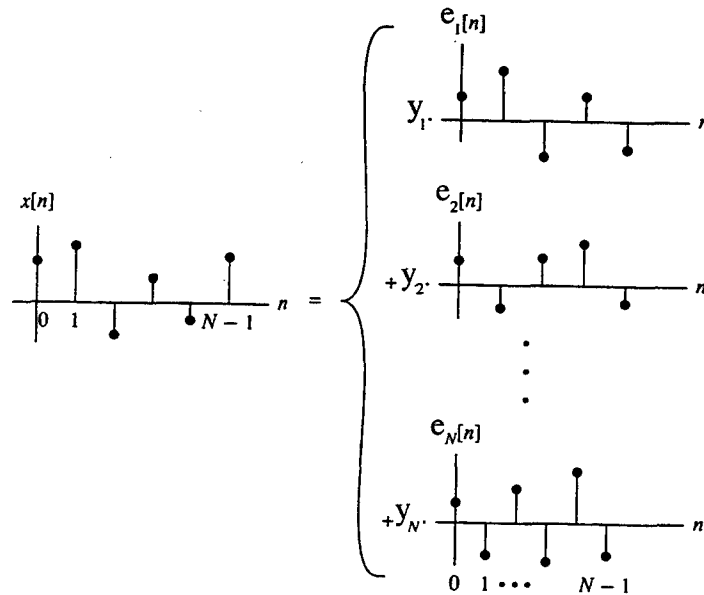


Figure 4.2: The Karhunen-Loeve Expansion in Terms of Discrete Signals.  
After Therrien, 1992, p. 175.

signal, indicated by the square brackets in the notation of Figure 4.2. Whereas in the PCA approach the original variables are weighted by eigenvector components to form principal components, in the DKLT the eigenvector basis functions,  $\{e_1, \dots, e_N\}$ , are weighted by the principal component scores,  $\{y_1, \dots, y_N\}$ , to form a representation of the observation. The DKLT has an optimal representation property in that it is the most efficient representation of the observed random process if the expansion is truncated to use fewer than  $N$  orthonormal basis functions. This makes it very attractive from a compression perspective, and explains the popularity of DKLT as a compression scheme.

Another important property associated with the DKLT is the equivalence between the total variance in the vector  $x$  and the sum of the associated eigenvalues. This property is mathematically stated by the equation

$$\sum_{i=1}^l \sigma_i^2 = \sum_{i=1}^l \lambda_i \quad (4.4)$$

where the  $\sigma_i^2$  are the variances of the original variables, the  $\lambda_i$  are the eigenvalues, which also represent the variances of the transformed variables, and the index  $i$  ranges over all  $l$  bands. This property only holds for the orthonormal vectors which are eigenvectors of  $\Sigma_x$  and not for other orthonormal basis sets of vectors (Kapur, 1989, p. 501). When a representation of a signal is formed by using fewer than  $l$  basis functions, the mean square error (MSE) is a means of quantifying how well the representation corresponds to the original signal by measuring the power of the difference between the representation and original signals. The MSE incurred by truncating the representation is equal to the sum of the eigenvalues of the covariance matrix that were left out of the representation. (Therrien, 1992, p. 179) Conversely, the largest eigenvalues and their corresponding eigenvectors can be used to represent the intrinsic dimensionality of the signal. This corresponds to the number of dimensions that would be needed to represent the signal to some predetermined MSE.

In signal processing applications, the DKLT is a means of compressing data by representing it with a truncated number of eigenvectors. It is also an optimum way of detecting a signal in noise and works particularly well for the detection of narrowband signals. Since a significant portion of the signal energy lies in the direction of the first few eigenvectors, those eigenvectors can be said to define a subspace for the signal and all other eigenvectors define the subspace for the noise. This simple example is the basis

for several high resolution methods of spectral estimation used to detect sinusoids in noise (Scharf, 1991, p. 483).

### 3. Pattern Recognition View

The optimal representation properties of the DKLT were extended to pattern recognition by Watanabe in 1965 (Singh and Harrison, 1985, p. 884). The application of the DKLT for feature extraction is a first step in the pattern recognition process. Figure 4.3 shows the pattern recognition process. The goal of feature extraction is to find a

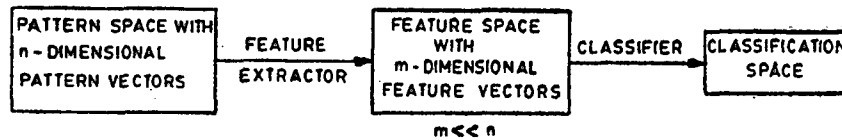


Figure 4.3: Paradigm of the Pattern Recognition Process. From Kapur, 1989, p. 497.

transformation from an  $n$ -dimensional observation space to a smaller  $m$ -dimensional feature space that retains most of the information needed for the next step in pattern recognition. The second step in pattern recognition involves classifying the pixels in an image by using some measure of separability. Feature extraction seeks to maximize the separation between classes in order to make classification easier and more accurate. The mutually uncorrelated coordinate axes (principal components) that define the feature space are called features.

The effectiveness of each feature in terms of representing  $\mathbf{x}$  is determined by the magnitude of its corresponding eigenvalue. There are various criteria for measuring the effectiveness of these features in representing  $\mathbf{x}$ . The MSE is one mentioned above. In addition, the scatter and entropy are criteria that could be used. The scatter is the expected value of the squared distance between elements of two different random vectors of the same random process. The entropy is a measure of the diversity of a distribution. Entropy is defined as:

$$H = -E\{\ln[p(\mathbf{x})]\} \quad (4.5)$$

where  $p(\mathbf{x})$  is the probability density function (pdf) of the random vector and  $E$  is the expectation operator. It is a complicated criterion since knowledge of the pdf of  $\mathbf{x}$  is required. The eigenvector decomposition given by the DKLT turns out to be the transform which maximizes the scatter and entropy of the distributions under

consideration (Fukunaga, 1971, p. 236). The maximization of entropy is equivalent to maximizing variance or uncertainty or information content, and is desirable in a feature extraction context, since the goal is to separate the unique classes in the data. It is important to note that this maximization of entropy occurs in the small number of principal components associated with the intrinsic dimensionality, and not over the entire range of the transformed variables. This is another means of stating that the PC transform concentrates the variance of highly correlated original data in the first few variables of the transformed data.

The topic of entropy is a complicated one, and a little more elaboration is required. A slightly different view would be to minimize the total amount of entropy in order to send less volume of data but retain the original information content, as is done with compression. In contrast to the above view, which maximized the entropy for a portion of the transformed variables, this outlook applies to all of the transformed variables. Ready and Wintz (1973) define the entropy in spectral imagery as:

$$H(\sigma^2) = -\sum_{i=1}^l \rho_i \log \rho_i$$

$$\rho_i = \frac{\sigma_{x_i}^2}{\sum_{j=1}^l \sigma_{x_j}^2} \quad (4.6)$$

where the indices  $i$  and  $j$  range over the  $l$  bands,  $\rho_i$  is the probability defined over the variances, and  $\sigma_{x_i}^2$  is the variance of the  $i^{\text{th}}$  spectral band (Ready and Wintz, 1973, p. 1124). In pattern recognition, structures in the data imply that the system is being constrained, that the amount of uncertainty has decreased, and hence that the entropy is smaller (Kapur, 1989, p. 514). According to Kapur (1989), the key is the uncertainty in the system. A system is completely unstructured, random, or simple if its entropy is the maximum possible. It is said to be completely structured, deterministic, or maximally complex if the entropy is zero. The DKLT minimizes the entropy defined in this fashion, and gives the least objective, most biased, the least uniform, the least random, and most predictable pdf that is consistent with imposed constraints. The DKLT lowers the entropy over the entire range of transformed variables because it has in effect provided structure to the data.

## C. OPERATION

The family of techniques considered in this section are motivated by the principal components transform. Though the techniques all share the common basic roots discussed in the context of multivariate data analysis, the DKLT, and pattern recognition, they are specifically designed for application to spectral imagery analysis.

### 1. Basic Principal Components Analysis (PCA)

PCA uses the eigenvectors of  $\Sigma_x$  to assemble a unitary transformation matrix which, when applied to each pixel vector, transforms the original pixel vector into a new vector with uncorrelated components ordered by variance. The eigenvector components act as weights in the linear combination of the original band brightness values that form the principal components (Richards, 1986, p. 137). The new image associated with each eigenvector is referred to as the principal component image. The principal component images are ordered from largest to smallest in terms of variance, and are revealing in their composition. As Singh and Harrison (1985) point out, it must be kept in mind that the PCA is an exploratory technique that constructs new variables called the principal components (PCs). These new variables are artificial and do not necessarily have a physical meaning, as they represent linear combinations of the observed variables, but cannot themselves be observed directly. In traditional application of PCA, the hope is that the transformation will enhance the contrast of the image to such an extent that objects or areas of interest can be more readily discriminated in the principal component images. Jenson and Waltz (1979) give an analogy which clearly explains the role of PCA in the traditional application. They imagine a tube filled with ping pong balls. Looking at the tube directly from an end, only one ball is apparent, the same way that the original spectral image is highly correlated. Turning the tube sideways, all of the balls become visible (Jenson and Waltz, 1979, p. 341). PCA has the effect of decorrelating the data so that independent sources of spectral features can be discerned.

Though PCA assumes no *a priori* knowledge of the scene, it cannot be applied totally independently of the specific scene. Scene-specific features will dictate the behavior of the PCA. Nevertheless, certain general observations can be made regarding the PCA and an associated physical meaning without any knowledge of the scene. The following two figures seek highlight these observations. Figure 4.4 shows the first 25



25 Principal Component Images using Radiance Covariance — Davis Monthan

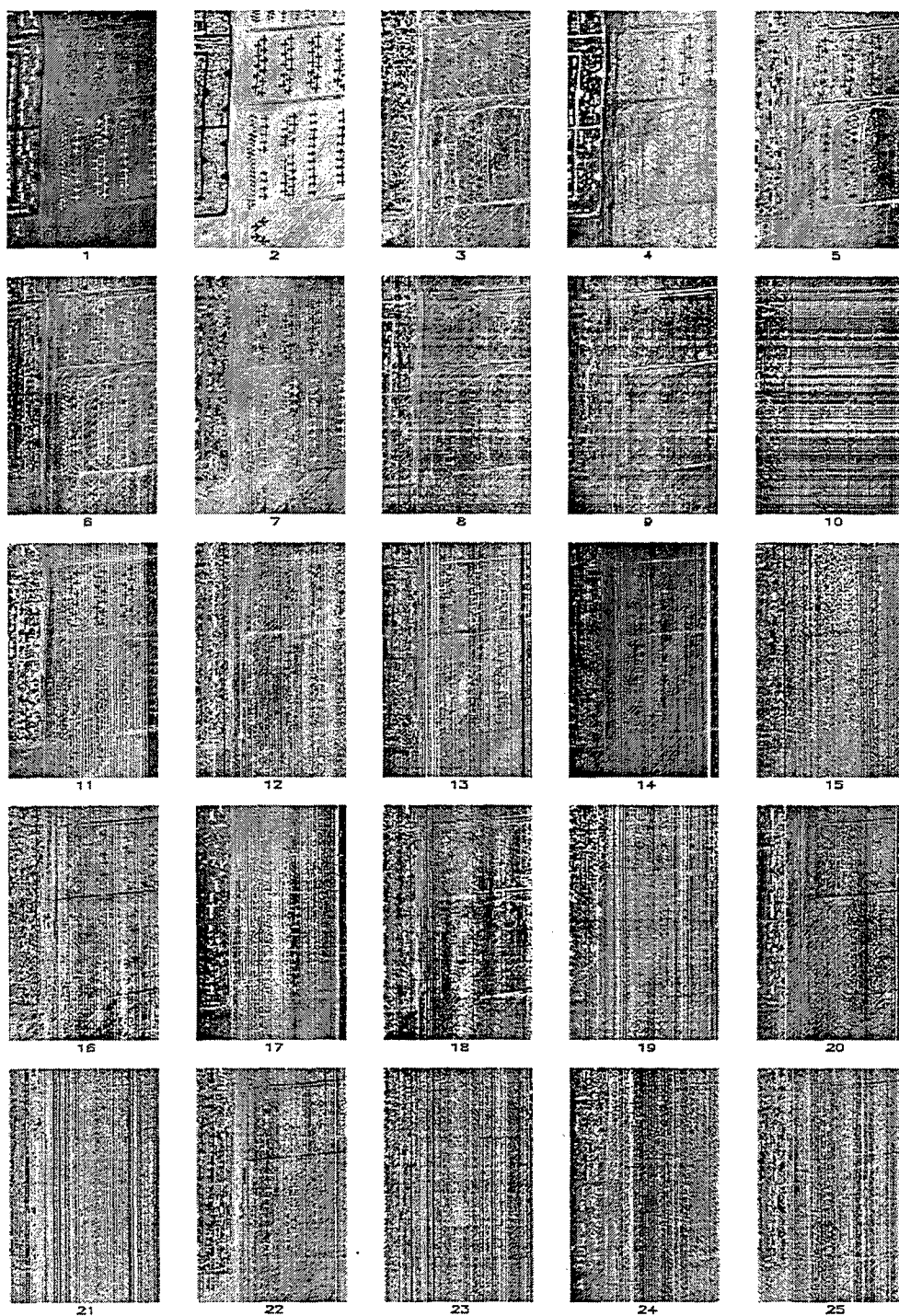


Figure 4.4: First 25 PC Images of Davis Monthan Radiance Scene.

PC images of the Davis Monthan HYDICE scene. A color version of this figure appears in Appendix A. The first principal component image is a typically a representation of the scene average brightness. This is due to the fact that in forming the first principal component image, the first eigenvector has heavily weighted the original bands possessing the most variance. Thus, the first principal component image will have a variance that is larger than that of any single original band image. It is the sum of the overall response level in all original band images. The second principal component image is typically the difference between certain original band images. As the principal component image number increases, the PC image holds less of the data variance. This effect manifests itself as a rough decrease in image quality with increasing PC image number. In Figure 4.4, the fact that the first seven PC images contain relatively clear details of the scene indicates that these PC images together account for the majority of the overall spectral variance in the scene. An interesting point to note when using PCA is that the higher numbered PC images sometimes contain a large amount of local detail. Though it is tempting to dismiss the higher numbered PC images as not containing any useful information because they have low variance, one must keep in mind that the covariance matrix on which PCA is based is a global measure of the variability of the original image (Richards, 1986, p. 138). This implies that small areas of local detail will not appear until higher PC images since they did not make a statistically significant impact on the covariance matrix. Another point that is noteworthy is the issue of SNR. PCA orders PC images based on total variability. It does not differentiate between the variability representing desirable information and the variability representing undesirable noise (Jenson and Waltz, 1979, p. 338). Ready and Wintz (1973) argue that PCA improves the SNR of the spectral image. Their definition of noise is additive white Gaussian noise with a variance of  $\sigma_n^2$ . The SNR of the original image is:

$$(SNR)_x = \frac{\sigma_{x_{\max}}^2}{\sigma_n^2} \quad (4.7)$$

which is the maximum original band variance over the noise variance. The SNR of the PC images is:

$$(SNR)_y = \frac{\lambda_1}{\sigma_n^2} \quad (4.8)$$

which is the largest eigenvalue (or new variance) over the noise variance. Since the first eigenvalue always has a greater variance than any of the original bands, the improvement in SNR is:

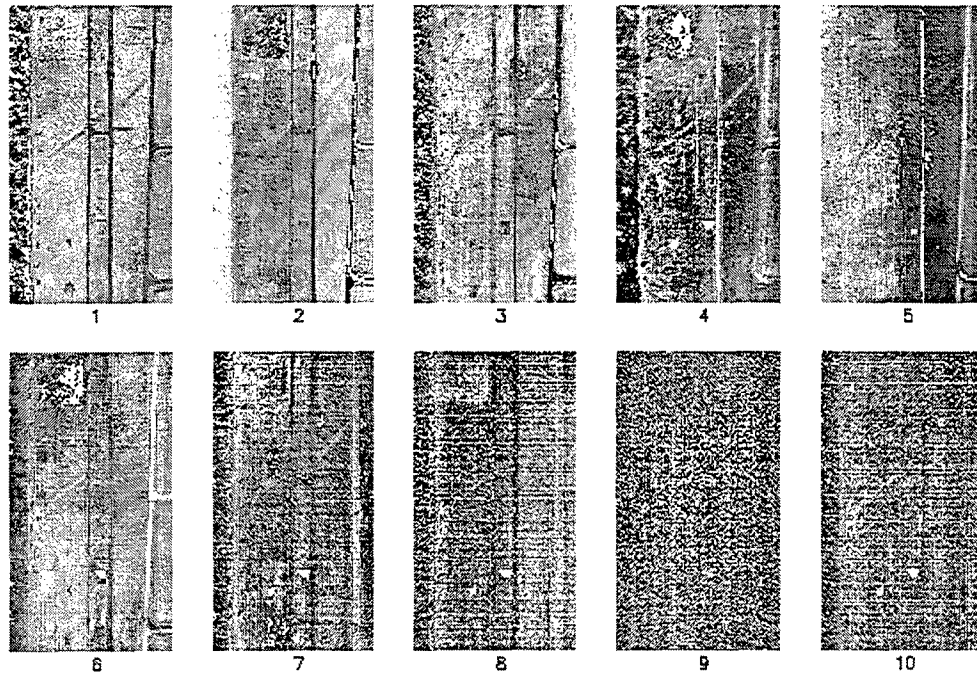
$$\Delta SNR = \frac{(SNR)_y}{(SNR)_x} = \frac{\lambda_1}{\sigma_{x_{max}}^2} \quad (4.9)$$

and will be greater than one. The SNR improvement applies as long as the variance of the eigenvalue exceeds that of the original bands. The diminishing SNR manifests itself in Figure 4.4 as striping patterns that begin to appear at the eighth PC image. Figure 4.5 further accentuates the above observations using the Aberdeen radiance and reflectance images. The first ten PC images are shown for each data set. A color version of Figure 4.5 appears in Appendix A. The same general trends noted for Figure 4.4 appear. The first few PC images offer the greatest amount of contrast. The effects of noise become apparent sooner in decreased image quality with the reflectance data than the radiance data.

A traditional means of presenting PCA images is to form a false color composite image consisting of the first, second, and third PC images as the red, green, and blue colors. Appendix B presents such false color images for the Davis Monthan radiance and Aberdeen radiance and reflectance PC images in Figures B.1, B.2, and B.3. This mode of presentation captures the major sources of spectral variability in one image. The levels of detail and contrast apparent in the composite image are interesting to compare with the original image cube shown in Figure 3.2.

A facet of PCA rarely mentioned in the pertinent literature on PCA is the characterization of the original and PC images using the behavior of the eigenvalues, entropy, and eigenvectors. These attributes form an important part of analyzing the scene information content. In spectral images, the typical trend in the eigenvalue magnitude is that a very small number of eigenvalues have a disproportionately large magnitude compared to the others. The obvious reason for this distinct grouping of eigenvalues is that the data in the original image exhibits a high degree of interband correlation and the magnitude of the eigenvalues reflects the degree of redundancy in the data. (Richards, 1986, p. 137). Phrased another way, the intrinsic dimensionality, which is represented by the number of large eigenvalues of the data, is very small. This is good from a compression view, since the image variance will be accounted for by a very small number of principal components. From an analysis vantage, it does not reveal as much information. If the problem were that of a narrowband signal embedded in noise, then the large eigenvalues would be associated with the signal. In the hyperspectral imagery analysis problem, the spectrum associated with a target is not narrowband, and hence is

10 Aberdeen Principal Component Images using Radiance Covariance



10 Aberdeen Principal Component Images using Reflectance

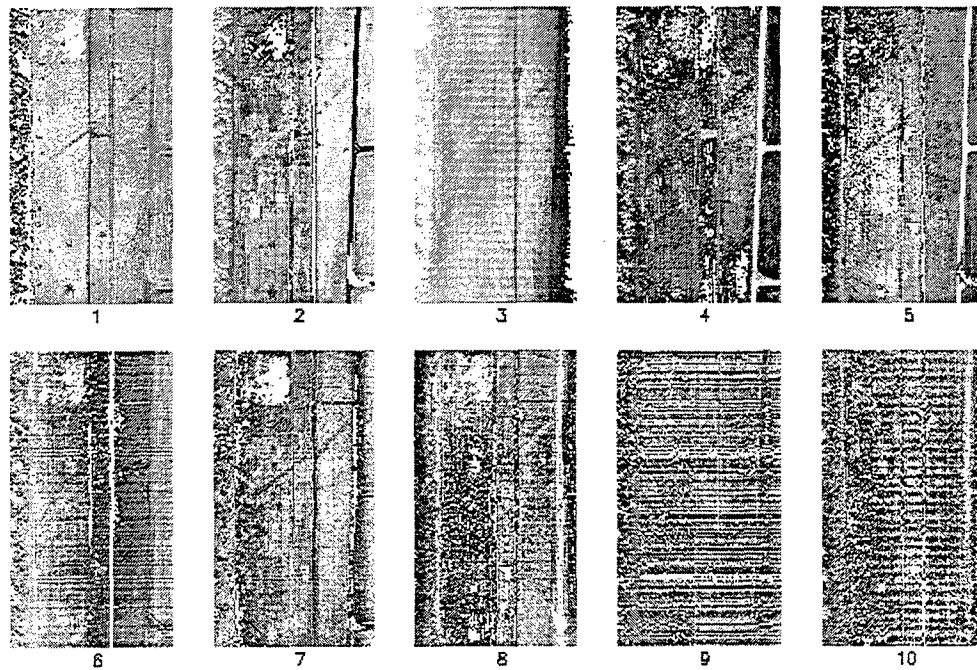


Figure 4.5: First Ten PC Images of Aberdeen Radiance and Reflectance Scenes.

not clearly delineated from the eigenvalues of the background and other interfering signatures. The eigenvalues can be divided into a primary and a secondary set, where the secondary set roughly corresponds to the effects of instrumentation noise (Smith, Johnson, and Adams, 1985, p. C798). The primary set corresponds to the linear combinations of original bands that cause the most variance in the scene. Figure 4.6 illustrates the eigenvalues of the Davis Monthan and Aberdeen radiance images together.

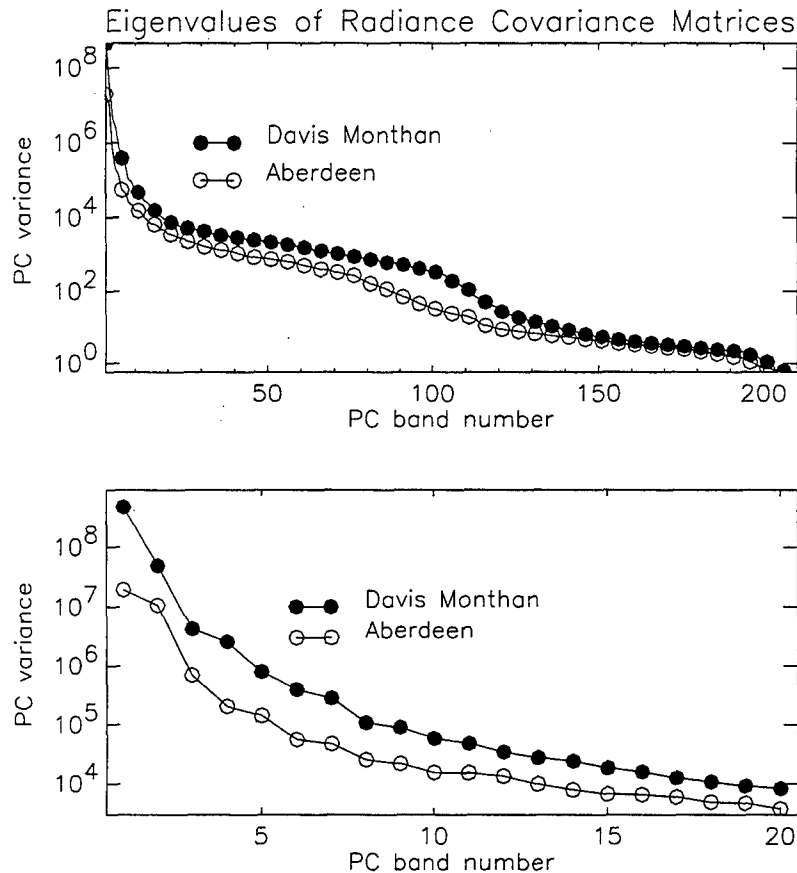


Figure 4.6: Eigenvalue Behavior of the Davis Monthan and Aberdeen Radiance Scene Covariance Matrices.

The y-axis of these plots is logarithmic, and represents the variance of each PC image. The lower plot is a detailed view of the first twenty eigenvalues. The Davis Monthan PC images exhibit slightly higher variances (eigenvalues) than the Aberdeen scene. The quality of the first eight PC images noted in Figure 4.4 corresponds to the steeper initial slope of the detailed eigenvalue plot. Likewise, the first six images of the Aberdeen

radiance PC images in Figure 4.5 are reflected in the steeper slope of the first six eigenvalues of Figure 4.6

Figure 4.7 shows the eigenvalues of the Aberdeen reflectance image. The sharp drop in the slope of the eigenvalues is paralleled by the drop in image quality noted in Figure 4.5 after the second PC image. In general, the HYDICE reflectance eigenvalues are lower in magnitude than those of the radiance.

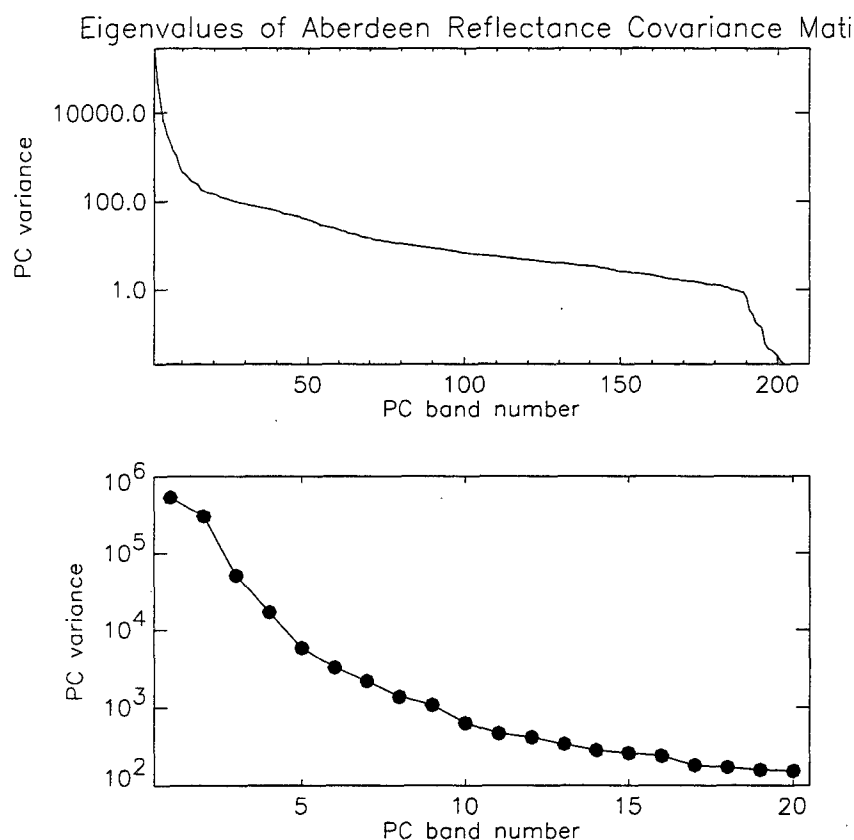


Figure 4.7: Eigenvalue Behavior of the Aberdeen Reflectance Scene Covariance Matrix.

The behavior of the entropy is another attribute of the original scene and the PC images. Ready and Wintz's (1973) definition of entropy found in Equation 4.6 is used to calculate the entropy of the representative scenes. The next three figures seek to demonstrate the concept of entropy by presenting it along with the behavior of the scene variance before and after the PC transform. The comparison with variance is necessary because entropy is defined in terms of variance. All of the plots are on a logarithmic  $x$ - and  $y$ -axes. Figure 4.8 portrays the variance and entropy behavior of the Davis Monthan

original and PC transformed data. The variance behavior of the original and transformed data shows that the original variance is highest in bands 20 to 60, corresponding to the effect of the sun on the variance of radiance data. The bands with high variance for the transformed data are concentrated in the first few bands, showing that the PC transform orders the PC images based on decreasing variance. It is important to note that the variance of the original data is equal to that of the transformed data. This property shows

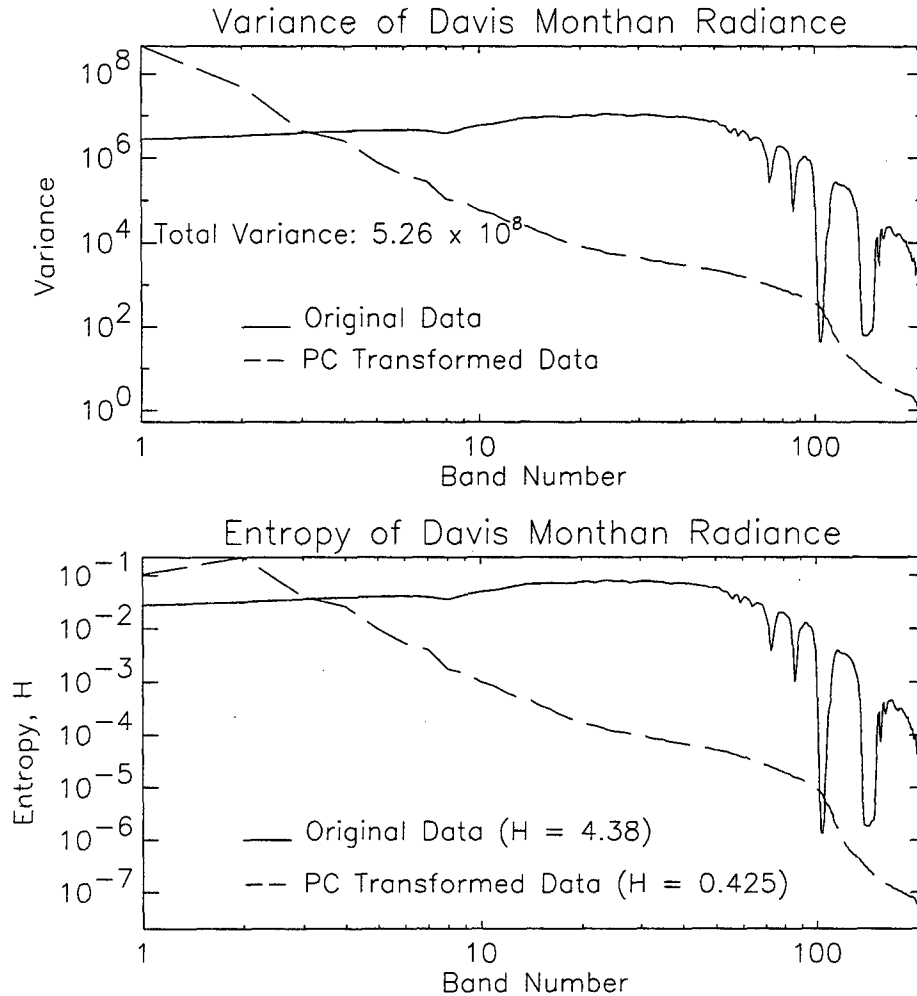


Figure 4.8: Variance and Entropy Behavior of Davis Monthan Radiance Covariance Matrix.

that the PC transformation merely redistributes the concentration of variance in the bands of a spectral image so that the higher variances occur in the first PC bands. While the shape of the entropy curve of the original data resembles that of the variance, the entropy curve of the transformed data has a different shape. The difference is the peak in entropy that occurs at the second band. Another observation is that the total entropy of the scene is not conserved in the PC transformation. The entropy associated with the transformed data is an order of magnitude less than that of the original data. The explanation for this behavior lies in the fact that the PC transformation reduces the entropy because it forms new variables which are linear combinations of the original variables.

Figures 4.9 and 4.10 demonstrate the same general observations noted above for the Aberdeen radiance and reflectance data and their PC transforms. The variance of the

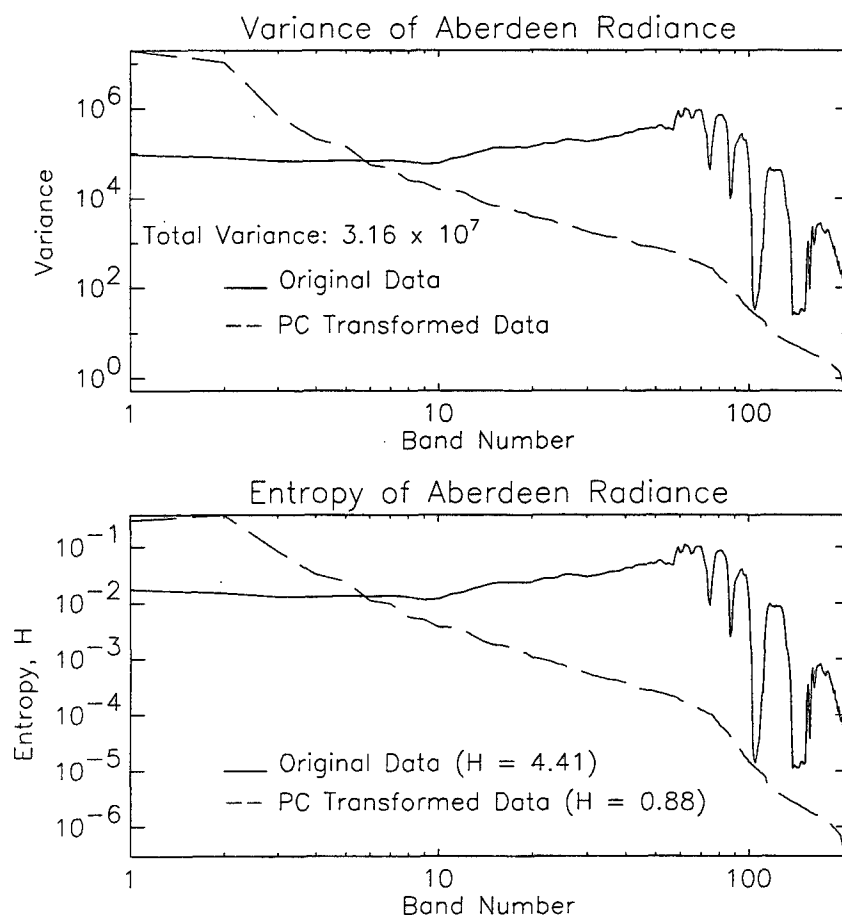


Figure 4.9: Variance and Entropy Behavior of Aberdeen Radiance Covariance Matrix.



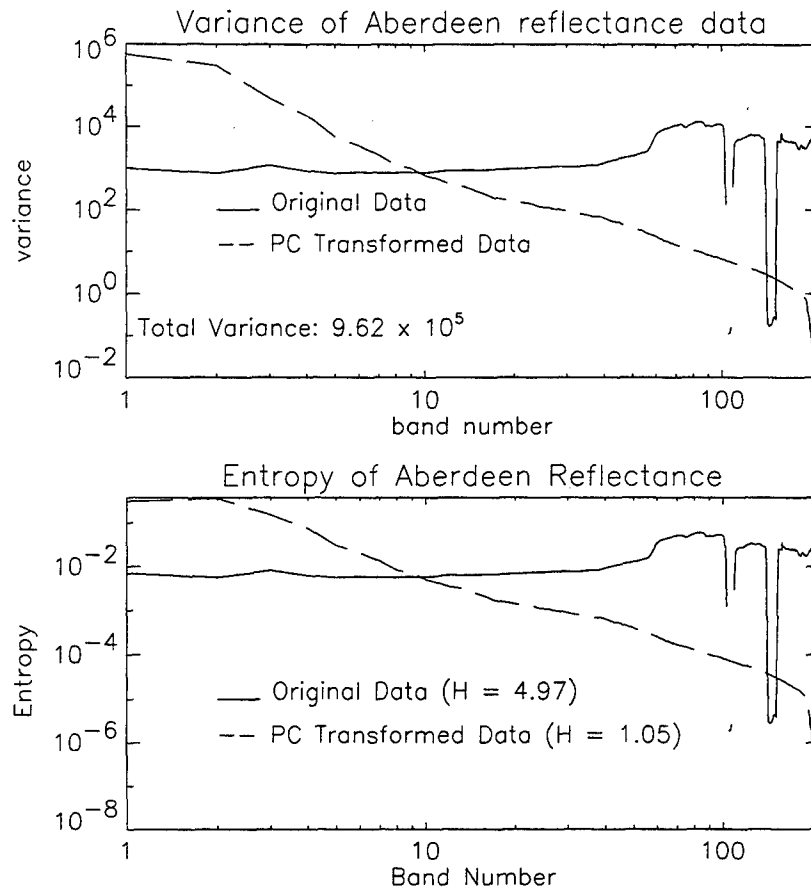


Figure 4.10: Variance and Entropy Behavior of Aberdeen Reflectance Matrix.

original reflectance data has a flatter shape than that of the original radiance data because of the removal of the sun's effect on variance by the conversion to reflectance.

The eigenvector behavior is less clear than that of the eigenvalues. The eigenvectors form the bases of the principal components subspace. Physically, the eigenvectors correspond to the principal independent sources of spectral variation. As such, the wavelengths at which the maxima and minima of the eigenvectors occur account for the wavelengths that contribute the most to a particular independent axis of variation (Smith, Johnson, and Adams, 1985, p. C808). Another interpretation of the eigenvectors is that the eigenvectors act as band pass filters that transform an input observed spectrum into a new spectrum that has fewer data points (Johnson, Smith, and Adams, 1985, p. C808). This interpretation is analogous to the optimum representation property of the DKLT. Figure 4.11 shows two-dimensional color representations of the

eigenvector matrices of the representative HYDICE data sets. The eigenvectors appear in the color plots as row vectors with the first eigenvector across at the bottom of the plot. The yellow and blue colors corresponding to large positive and negative

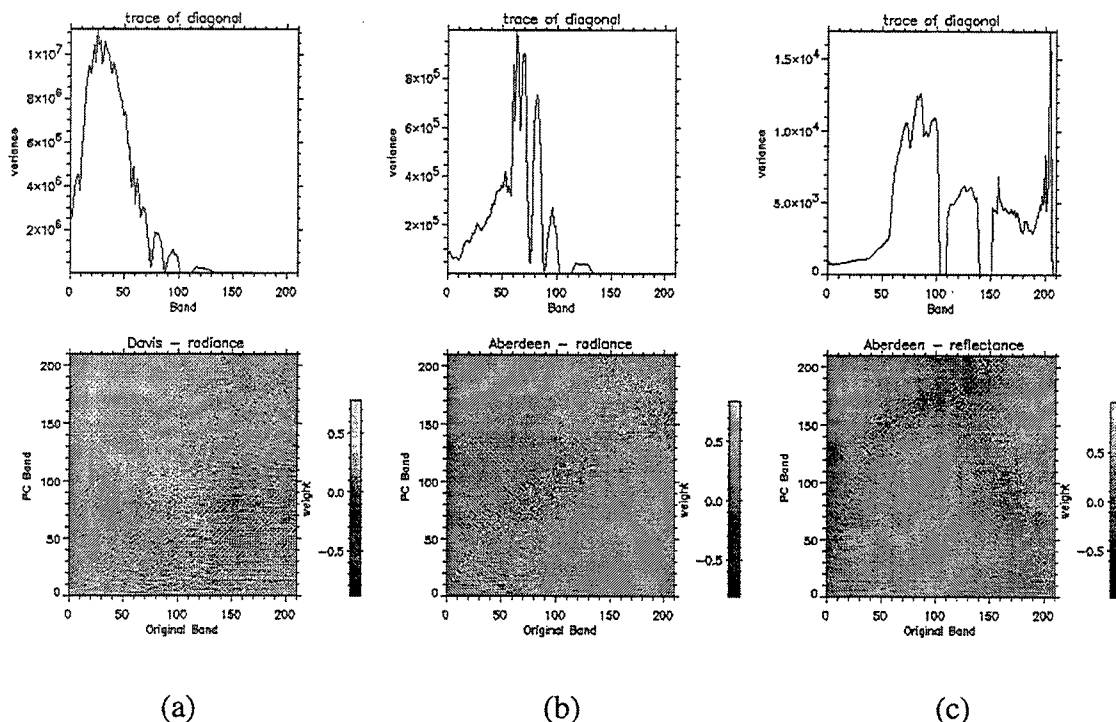


Figure 4.11: Eigenvectors and Traces of the Covariance Matrices of Davis Monthan and Aberdeen Radiance and Reflectance.

values serve to highlight the overall trends in eigenvector behavior. The  $x$ -axes of the eigenvector plots are labeled as the original bands to emphasize the fact that the eigenvector is a sequence of weights which are applied to original bands. In the case of this plot, the weight magnitude corresponds to the color as indicated by the accompanying bar scales. The  $y$ -axes of the plots are labeled as PC bands to emphasize the role of that particular eigenvector in forming the corresponding PC image. Specifically, the  $i^{\text{th}}$  PC image is formed by application of the  $i^{\text{th}}$  eigenvector to the original data. Above each eigenvector plot is a plot of the trace of the associated covariance matrix. The trace of a covariance matrix is the variance, and the PC transform orders the original variables according to their variances. By noting where the variance is high in the plot of the trace and comparing these band numbers with the magnitude of the weights in the corresponding eigenvector band numbers, one can

determine the relative importance placed on those original bands in forming a particular PC image. Further, by noting the relative position of large weights with respect to PC band, one can see the effect of a large variance in the original data manifested as an appearance of large magnitudes in a low numbered PC band. For example, note how the peak in variance of the Davis Monthan data manifests itself as significant weight activity in the early PC bands while the original bands corresponding to the very small variances do not experience significant weight activity until the last PC bands.

Additional insight into the foregoing discussion about weights and relative importance of PC bands is gained by viewing the eigenvector matrix as a surface plot. The following three figures attempt to capture the eigenvector behavior of each of the three data sets. Figure 4.12 shows the eigenvectors for the Davis Monthan data. As in

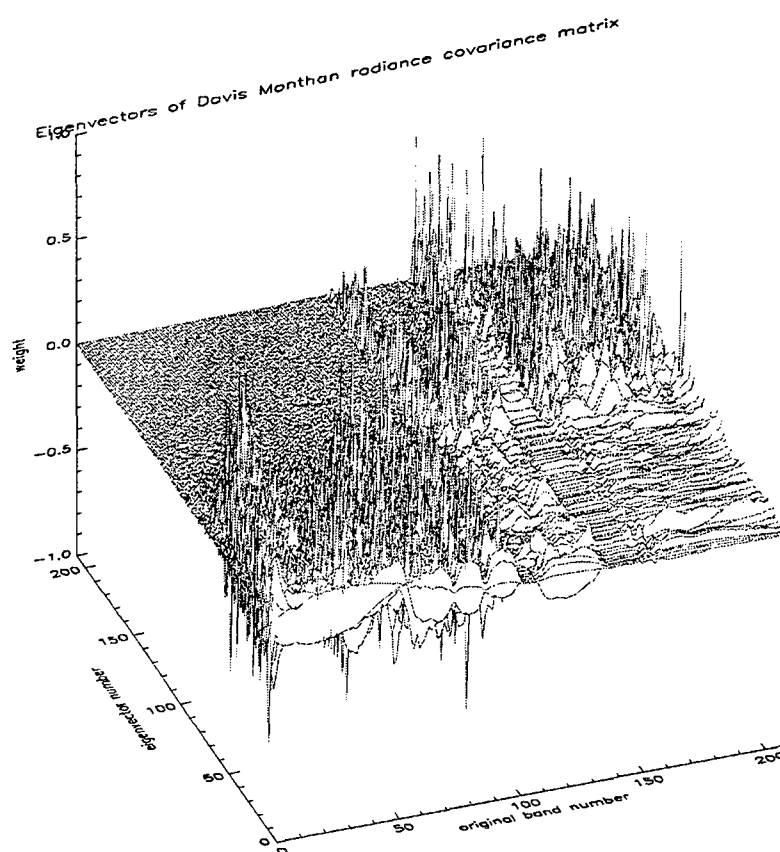


Figure 4.12: Eigenvectors of the Davis Monthan Radiance Covariance Matrix.

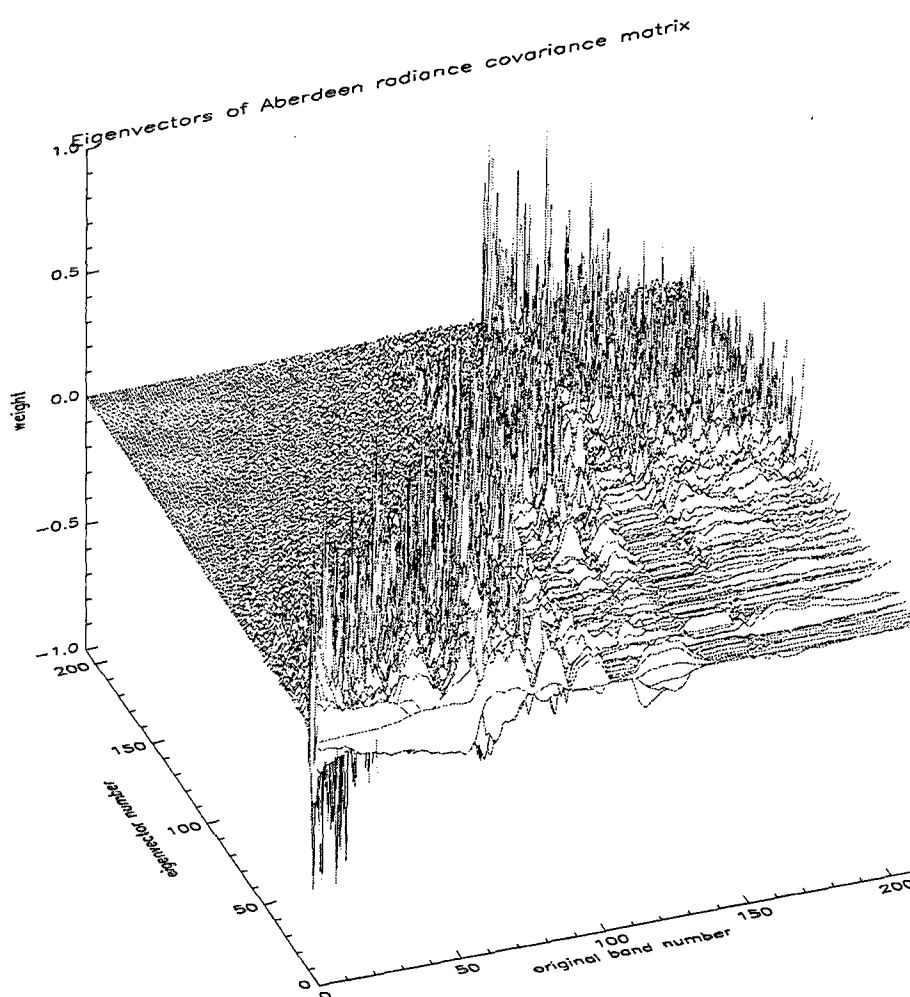


Figure 4.13: Eigenvectors of Aberdeen Radiance Covariance Matrix.

the eigenvector plots of Figure 4.11, the original bands are indicated on the  $x$ -axis and the PC bands or eigenvector numbers are indicated on the  $y$ -axis. The  $z$ -axis represents the weights. Note how the first 100 eigenvectors all heavily weight original bands one to 120 and provide virtually no weight to the remaining original bands. Original bands 120 to 210 are weighted in eigenvectors 100 to 200, which implies that these original bands had small variances. The absorption bands (original bands 100 to 110 and 140 to 150) are weighted heavily in the last ten eigenvectors since the variance in absorption bands is effectively zero. The abrupt checkerboard appearance of this surface plot is contrasted by the more linear appearance of the surface plot in Figure 4.13. Figure 4.13 displays the eigenvectors of the Aberdeen radiance data. There appears to be a roughly

linear relationship between the eigenvector number and the original band number. The placement of major weights in the eigenvectors corresponds to this diagonal line. The implication is that the low numbered original bands receive greater emphasis in the PC transform by virtue of their weighting in the low numbered eigenvectors. This is

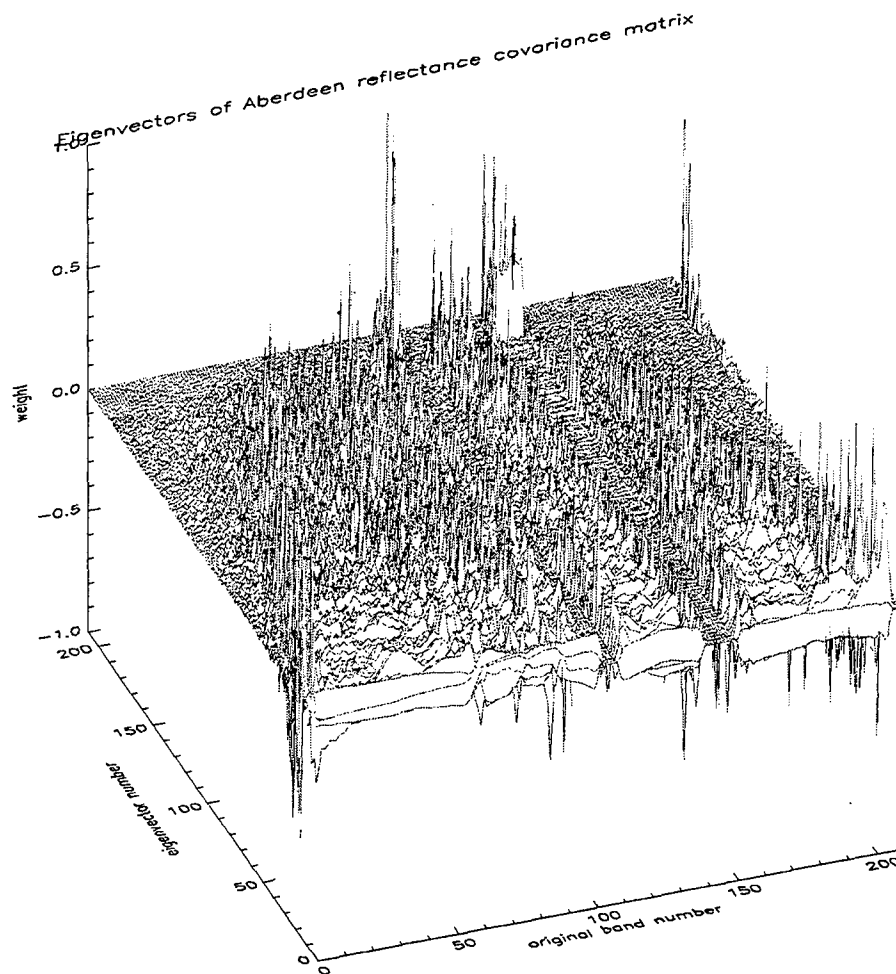


Figure 4.14: Eigenvectors of Aberdeen Reflectance Covariance Matrix.

a statement about how radiance data variance is highly concentrated in the first 100 bands due to the sun's effect. In contrast to this, Figure 4.14 reveals that the eigenvectors of the Aberdeen reflectance data have a more distributed appearance. The effect of the sun on the low numbered original bands has been mitigated in the conversion to reflectance. The patterns observed in Figures 4.12 and 4.13 for the radiance data are missing from this eigenvector plot.

In further examining eigenvector behavior for the three representative HYDICE data sets, the first eight eigenvectors are plotted. The plots are superimposed on a slice across the image hypercube to which those eigenvectors apply. The purpose of this portrayal is to emphasize eigenvector behavior with respect to the variance occurring in the original image bands. The background images in the next three figures are oriented so that bands range horizontally from left to right and spatial samples range vertically. Note the presence of the absorption bands across all samples as dark vertical lines. The previous figures have explained how the eigenvectors of the PC transform tend to emphasize those original bands that contain the most variance with larger weights and inclusion in the low numbered eigenvectors. These three figures, which may be found in their color version in Appendix A, emphasize the point more specifically. Since the first eight eigenvectors are those which are used to generate the PC images with the highest variances, they will tend to weight the original bands with the greatest variances. In the background plots, the amount of variance may roughly be discerned as the amount of change occurring in the colors of a particular band as one ranges over the samples. Figure 4.15 shows a great deal of background image variance in the first 80 bands. Consequently, the eigenvectors show much weighting in this region. The first eigenvector heavily emphasizes bands ten to 40 because these bands show the greatest variance as noted in the variance plot of Figure 4.11. In the background of Figure 4.16, the greatest variance can be seen to occur between bands 30 to 70. Note how the weights of the eight eigenvectors place emphasis on different portions of this region. The eigenvectors in this figure all place no weight on the absorption bands, as seen in Figure 4.13. The majority of weighting in the first eigenvector appears between bands 55 to 70, which corresponds to the area of greatest variance in the original data. The difference in Figures 4.15 and 4.16 makes it clear that the predominant spectra in the Aberdeen scene are those of vegetation. The existence of an infrared "ledge" at band 55 is associated with the spectra of vegetation. Figure 4.17 displays different eigenvector behavior for the reflectance data of the Aberdeen scene. The first eigenvector appears to place the emphasis on bands 50 to 100, but also emphasizes later bands as well. This is indicative of the more uniformly distributed variance of the reflectance data. The reflectance data eigenvectors also seem to place emphasis on the bands in the vicinity of band 200. This corresponds to the spike in variance seen in Figure 4.11 for the reflectance data.

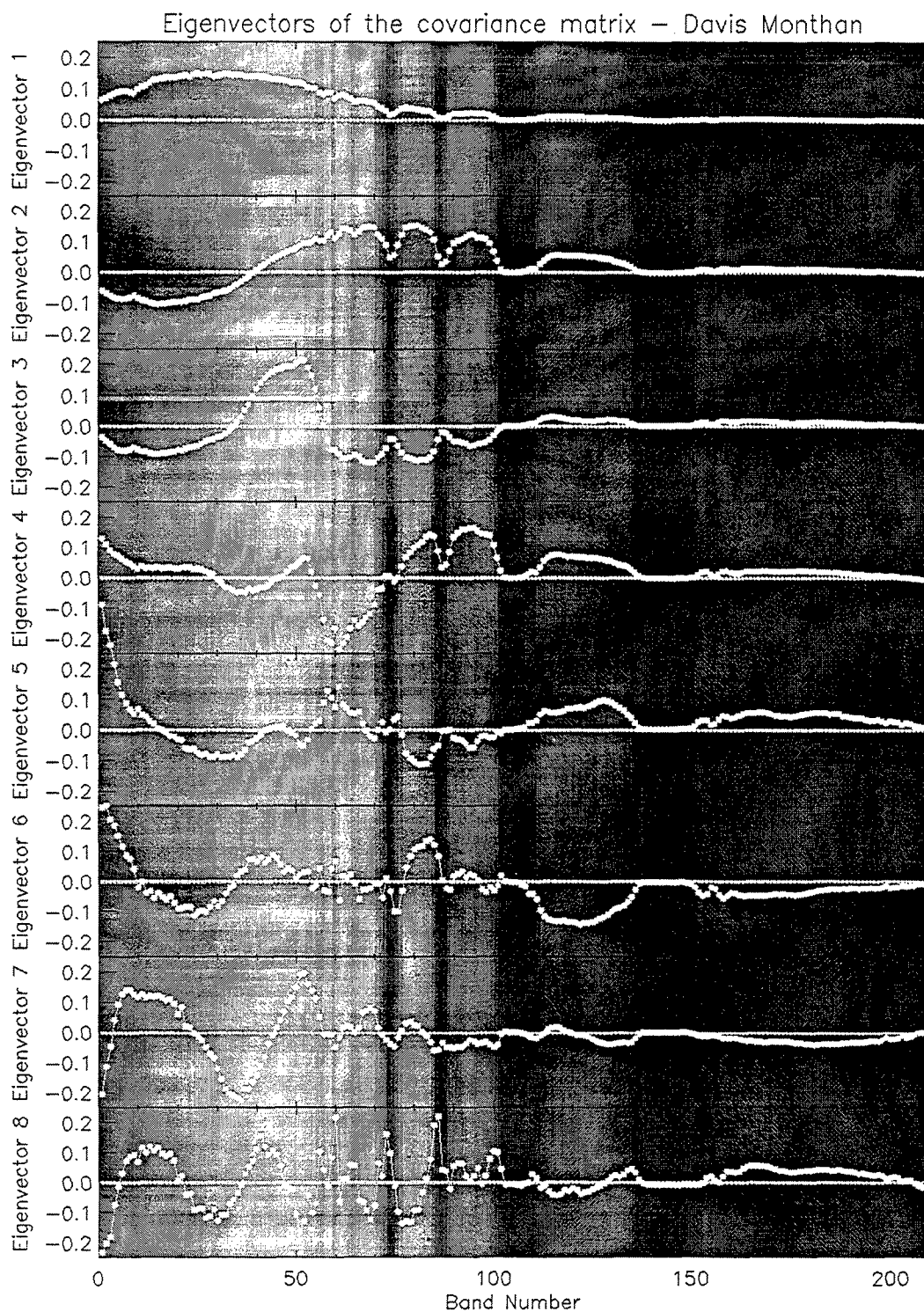


Figure 4.15: First Eight Eigenvectors of Davis Monthan Scene Superimposed on a Random Slice Across the Hypercube.

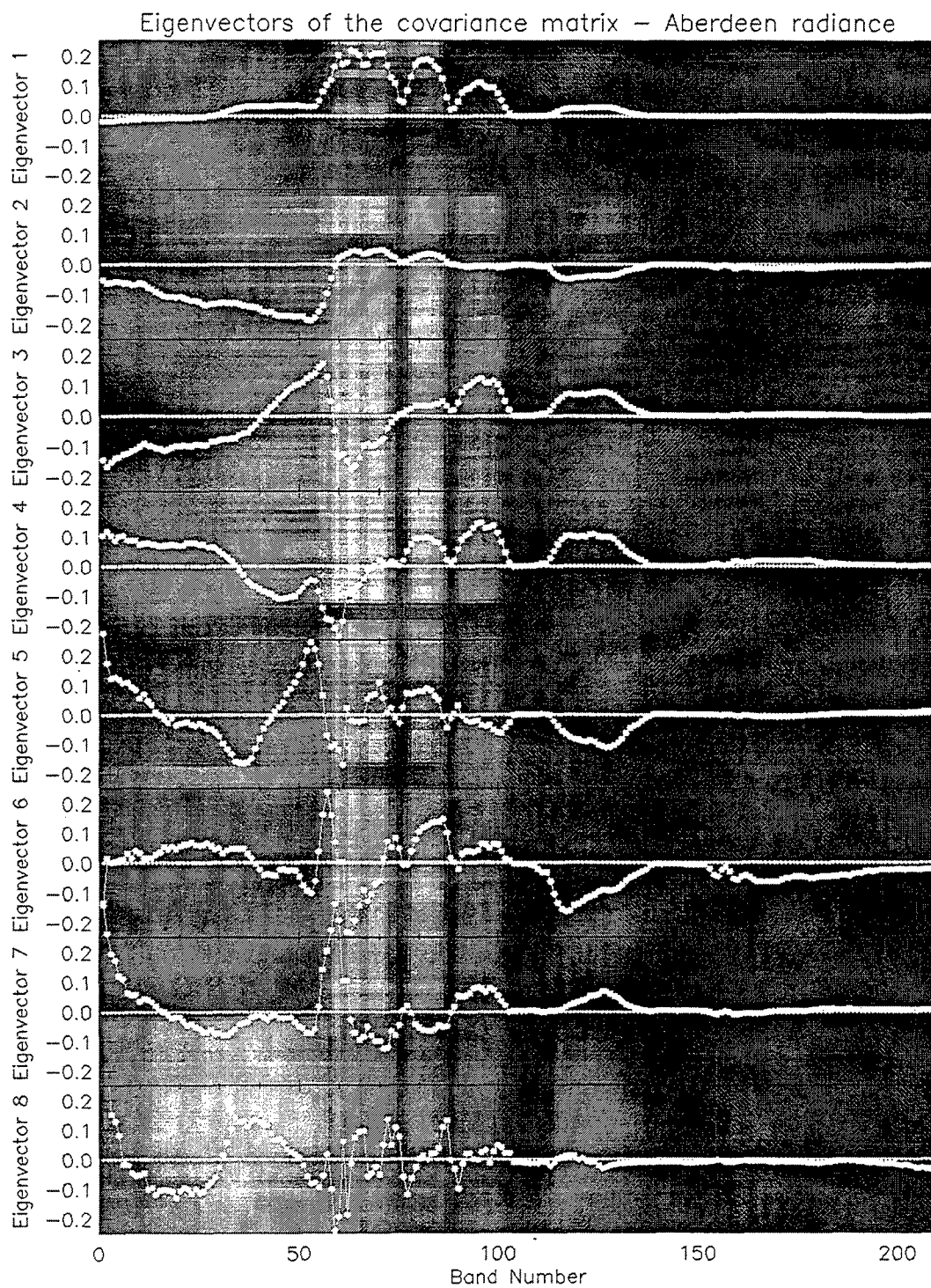


Figure 4.16: First Eight Eigenvectors of Aberdeen Radiance Scene Superimposed on a Random Slice Across the Hypercube.



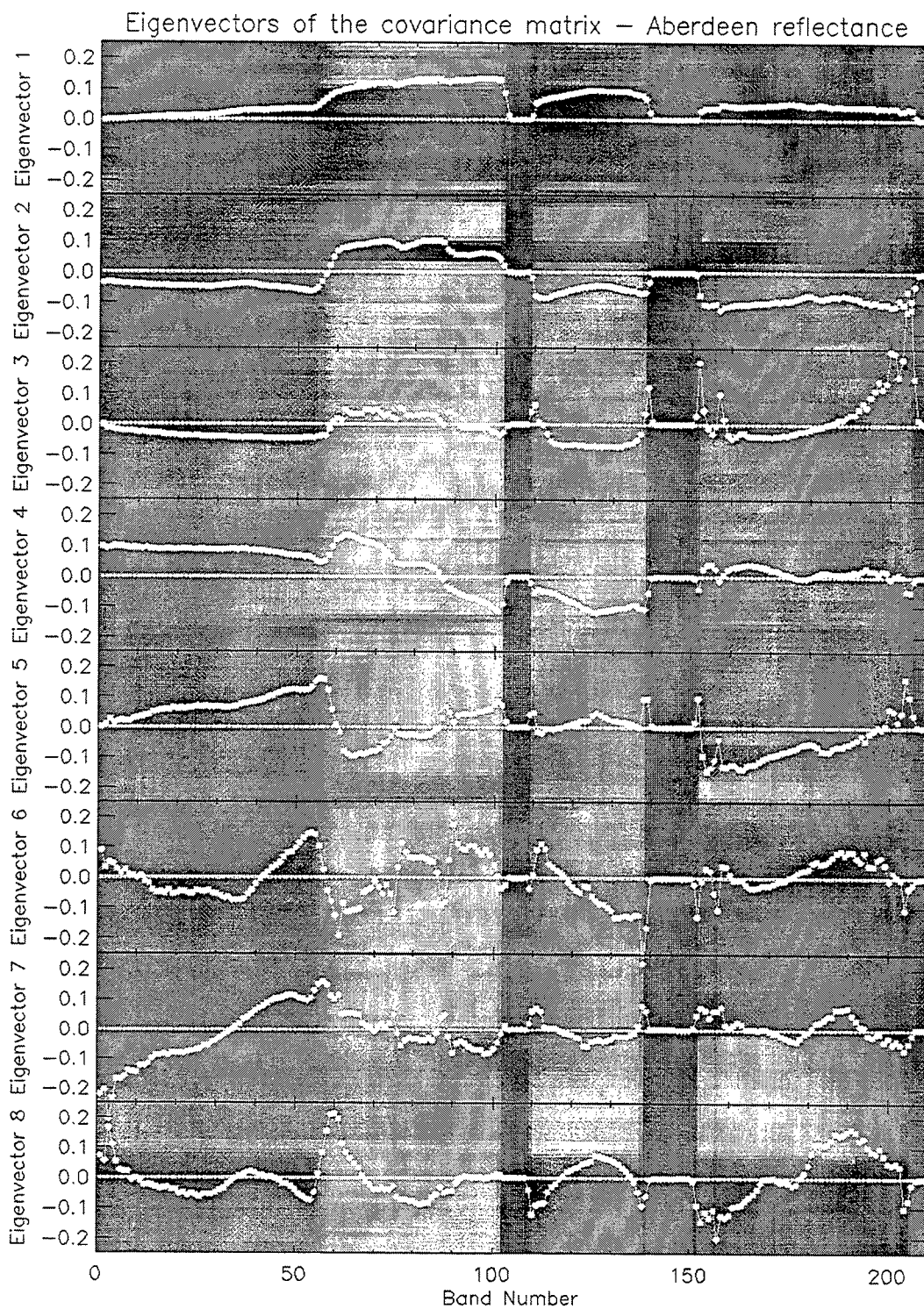


Figure 4.17: First Eight Eigenvectors of Aberdeen Reflectance Scene Superimposed on a Random Slice Across the Hypercube.

The PCA technique has been examined from the perspective of its results and the significance of its inner workings. In general, PCA provides an analysis of the data which guarantees an output set of images ordered by variance. It improves the SNR in the transformation from the original image cube to the PC images. The PC images accentuate spectral regions of high variance. An area of local detail may not be accentuated by a PC image due to its statistical insignificance. The user, in searching for a target of interest, has no control as to the emphasis that will be placed on the target spectrum in the PC transform. Because the variability of the data is scale-dependent, PCA is sensitive to the scaling of the data to which it applied, and as a result, the PCA of radiance data will place more emphasis on the visible bands due to the sun than the PCA of reflectance data. PCA does not differentiate between noise and signal variances, because it operates strictly on the variance of the observed data. As a practical note in the implementation of PCA, the computation of the eigenvectors and eigenvalues of  $\Sigma_x$  is an expensive operation. Specific methods from computational linear algebra such as inverse iteration, QR factorization, and singular value decomposition (SVD) are all applicable in their calculation (Watkins, 1991, p. 251).

## **2. Maximum Noise Fraction (MNF) or Noise Adjusted Principal Components (NAPC) Transform**

In noisy image data, the noise may contribute substantially to a principal components' variance, so that the useful information, or signal, contained in a large eigenvalue may actually be less than that of a smaller eigenvalue (Roger, 1994, p. 1194). Since the PCA is based strictly on constructing new components that maximize the variances of the original bands without regard to signal or noise, it cannot reliably separate the signal and noise components of spectral imagery. The maximum noise fraction (MNF) was introduced by Green, Berman, Switzer, and Craig in 1988 to help solve this basic undesirable feature of the PCA and equivalently derived as the noise adjusted principal components transform (NAPC) transform in 1990 by Lee, Woodyatt, and Berman.

The impetus for the MNF transform was to design a unitary transform that would order PC images based on image quality, commonly measured by signal-to-noise ratio (SNR) (Green, Berman, Switzer, and Craig, 1988, p. 65). The model of the observations is that of a signal and additive noise, as given by the equation:

$$\mathbf{x} = \mathbf{s} + \mathbf{n} \quad (4.10)$$

where the vectors are all  $l$ -band pixel vectors. It is assumed that the signal and noise are uncorrelated, which implies that the second order statistics of the model may be written as:

$$\Sigma_{\mathbf{x}} = \Sigma_{\mathbf{s}} + \Sigma_{\mathbf{n}} \quad (4.11)$$

where the covariance matrices are of dimension  $l \times l$ . The noise fraction of the  $i^{\text{th}}$  band is defined as the ratio of the noise variance to total variance for that band and stated mathematically as:

$$\text{noise fraction} = \frac{\text{VAR}[n_i]}{\text{VAR}[x_i]} \quad (4.12)$$

where  $n_i$  is the  $i^{\text{th}}$  component of the noise vector  $\mathbf{n}$  and  $x_i$  is the  $i^{\text{th}}$  component of the observed pixel vector  $\mathbf{x}$  over all spatial locations in the image. (Green, Berman, Switzer, and Craig, 1988, p. 66). The MNF is the linear transformation which maximizes the noise fraction in the new variables while guaranteeing that the new variables are uncorrelated. The MNF transform is derived in a similar fashion to the principal components transform with the exception that the transformation matrix is built using the transposed eigenvectors of the matrix  $\Sigma_{\mathbf{n}}\Sigma_{\mathbf{x}}^{-1}$  (Green, Berman, Switzer, and Craig, 1988, p. 66).

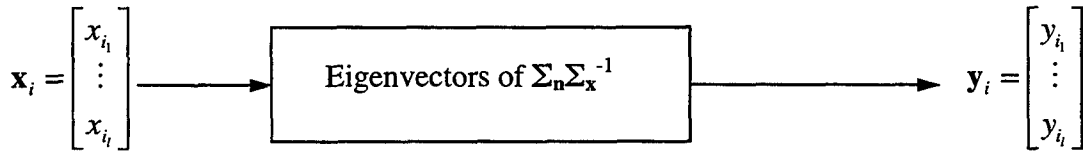


Figure 4.18: The MNF Transform.

Figure 4.18 shows the MNF transform as a linear transformation much like that of Figure 4.1 for the PC transform. The eigenvalues of  $\Sigma_{\mathbf{n}}\Sigma_{\mathbf{x}}^{-1}$  are actually the noise fractions of the corresponding new variables created by the transformation. They are ordered from largest noise fraction to smallest, implying that the image quality increases with component number. The MNF can also be constructed so that the image quality decreases with increasing component image number. This reversed form is called the minimum noise fraction transform. This is the form of the transform that is displayed in this discussion, though it is still referred to as the MNF transform. Figure 4.19 shows the eigenvalues

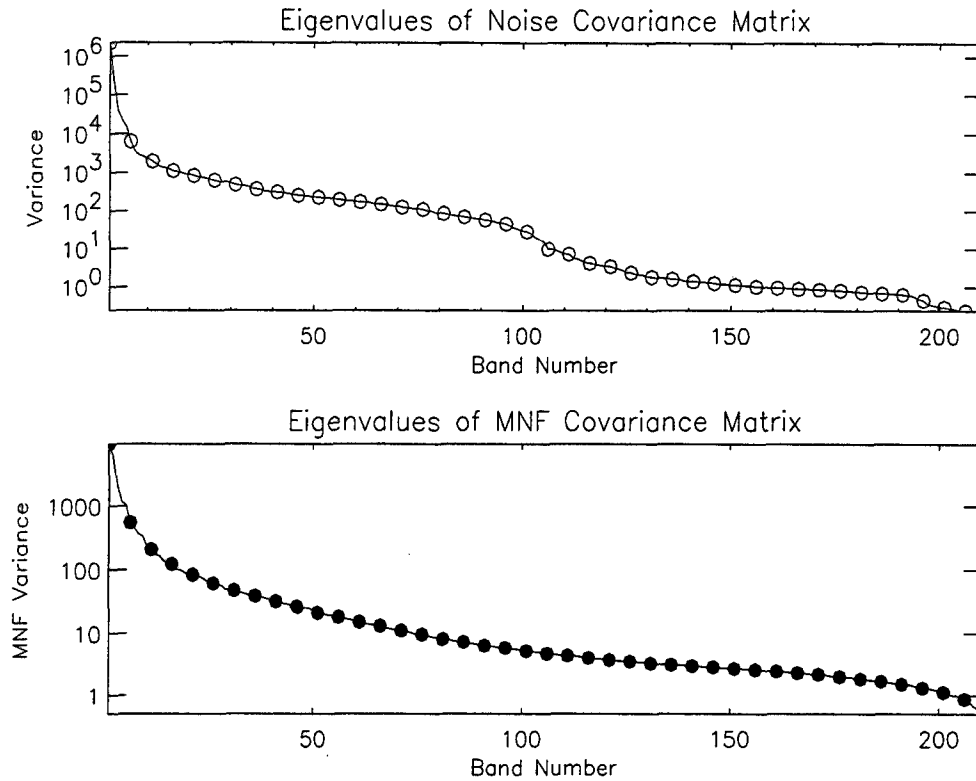


Figure 4.19: Eigenvalues of the Estimated Noise and Noise-whitened Covariance Matrices of the Davis Monthan Scene.

associated with the covariance matrices used in the MNF transform. The eigenvalues in Figure 4.19 are ordered from largest to smallest, implying that the noise fractions are arranged in increasing order in this representation. The solid line and open circles in Figure 4.19 represent the noise fractions of  $\Sigma_n \Sigma_x^{-1}$ , also called the noise-whitened covariance matrix. These eigenvalues are seen to be smaller than those of the noise covariance, represented by the solid line and filled circles. The noise covariance was estimated from a subset of the data corresponding to uniform background. This difference in eigenvalue magnitude is due to the presence of the  $\Sigma_x^{-1}$  term in the noise-whitened covariance matrix.

The MNF transform, unlike the principal components transform, is invariant to scale changes to any band because it depends on the SNR instead of variance to order the PC images (Green, Berman, Switzer, and Craig, 1988, p. 66). The MNF is equivalent to PCA when the noise has equal variance  $\sigma_n^2$  in all bands. This is because the eigenvectors of the matrices  $\sigma_n^2 \Sigma_x^{-1}$  and  $\Sigma_x$  are identical from the properties of unitary transforms.

25 MNF Images using Radiance Covariance — Davis Monthan

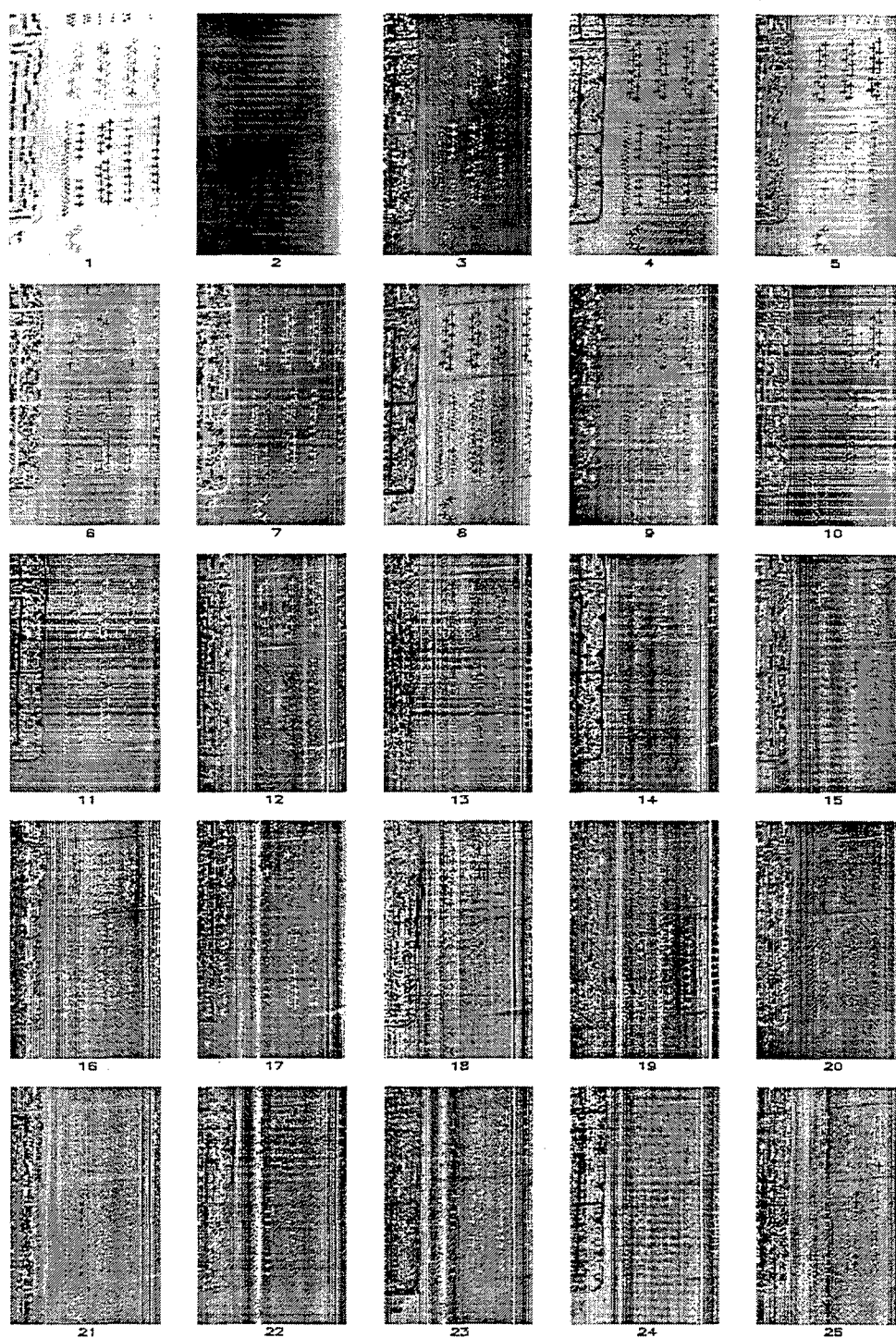


Figure 4.20: First 25 MNF Component Images of the Davis Monthan Scene.

The common case of equal noise in all bands explains the observation that PCA tends to order component images by image quality in most cases (Green, Berman, Switzer, and Craig, 1988, p. 67). To illustrate this point, Figure 4.20 presents the 25 MNF component images for the Davis Monthan radiance scene. A color version of this figure may be found in Appendix A. To determine the advantage of MNF with respect to PCA, comparison of this figure with Figure 4.4 is required. In Figure 4.4, the PC image quality is good until the eighth image. In later PC images, the effects of instrumental noise produce striping. In Figure 4.20, the image quality of the first few MNF images does not appear to have significantly improved over that of the PC images. The effects of noise and striping are less pronounced in the higher MNF images than in the corresponding PC images. The MNF transform in Figure 4.20 seems to arrange higher image quality in the first few images, with the exception of the first two.

The NAPC is based on viewing the MNF as a two-step process. This approach makes it more obvious that an estimate of the noise covariance is required in order to apply this technique. It first transforms the data to a coordinate system in which the noise covariance matrix has been whitened so that it is now the identity matrix, and then applies a principal components transformation (Lee, Woodyatt, and Berman, 1990, p. 295). This technique is equivalent to simultaneously diagonalizing the noise and a signal covariance matrices as was illustrated in Figure 3.15. Figure 4.21 shows a representation of the

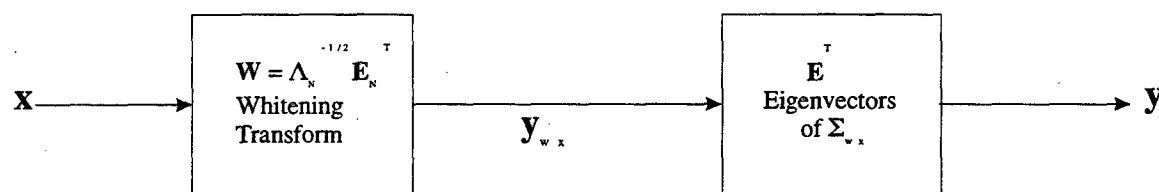


Figure 4.21: The NAPC Transform.

NAPC transform as first a whitening transform and then a PC transform. The whitening transform discussed in Chapter III uses the eigenvalues and eigenvectors of the estimated noise covariance matrix form the whitening operator,  $W$ . Application of  $W$  to the observed pixel vector  $x$  is in effect a unitary transform which produces the noise-whitened pixel vector,  $y_{wx}$ . The eigenvectors of the noise-whitened covariance matrix are then used to rotate  $y_{wx}$  so that it will have uncorrelated components in the new pixel

vector,  $y$ . Chapter VI illustrates the intermediate covariance matrices which result from the whitening transformation.

The NAPC technique is applicable when there is noise that has affected the SNR in certain bands of the data, and its covariance matrix can be estimated. The significance of the NAPC is that the eigenvectors associated with the most significant eigenvalues can truly represent the "signal." By whitening the noise, the noise variance is made equal over all bands, and the effects of noise variance do not create undesired "mixing" with the signal variances (Lee, Woodyatt, and Berman, 1990, p. 299). An important aspect of the NAPC is that it can be implemented using standard principal components software. The authors of the NAPC illustrate its operation using data from the 64-channel Geophysical and Environmental Research (GER) scanner that possessed an instrumental noise artifact that caused a significantly lower SNR in one band. The NAPC successfully allows a separation of noise and signal because the eigenvectors corresponding to the largest eigenvalues display signal effects whereas those associated with the smallest eigenvalues show noise effects (Lee, Woodyatt, and Berman, 1990, p. 298).

Both the MNF and NAPC require a knowledge of the noise covariance matrix. This information may be available from the dark current measurements of the sensor. If it is not available, then it must be estimated. Green, Switzer, Berman, and Craig (1988) propose a method of estimating the covariance structure of the noise in various bands of multispectral imagery directly from the data. Their approach is to select an appropriate spatial filter that will extract the noise portion of the observations using the spatial characteristics of noise and signal by subtracting neighboring pixels. A procedure known as minimum/maximum autocorrelation factors (MAF) was developed by Switzer and Green in 1984 to estimate the noise covariance matrix for certain types of noise by using a characteristic found in most remotely sensed images. This characteristic is that the signal at any point in the image exhibits a high degree of correlation with its spatial neighbors, while noise is only weakly correlated with its spatial neighbors. The covariance structure of the observations and their spatially lagged counterparts is proportional to the estimate of the noise covariance:

$$\Sigma_N \propto \text{COV}[x_i - (x_i + \Delta)] \quad (4.13)$$

where  $x_i$  is the observed pixel vector at spatial location  $i$ ,  $\Delta$  is the spatial lag, and COV denotes calculation of the covariance matrix. This covariance structure of the lagged

observations is a measure of the noise covariance structure,  $\Sigma_N$ , in the sense the two are proportional depending on the amount of spatial correlation present in the image.

### 3. Standardized Principle Components Analysis (SPCA)

A major drawback of using PCA based on the data covariance matrix is the sensitivity of the principal components to the units of measurement used for each variable (Jolliffe, 1986, p. 17). In the context of remotely sensed data, the units of measurement at the sensor from band to band are the same, but unequal SNR in all bands creates a problem which can have a similar deleterious effect on the PCA. The basic problem is the same as that solved by the MNF in that ordering principal components based strictly on variance does not account for the fact that noise in some bands may contribute a significant portion to the variance. The solution to this problem in the standardized principal components analysis (SPCA) technique is to normalize the variances of all bands to be unity. This standardization or normalization of the covariance matrix results in the correlation matrix earlier defined. Singh and Harrison (1985) have argued that the use of  $\mathbf{R}_x$ , or the standardized covariance matrix, is pertinent when it is undesirable to have the relative importance of components be weighted by the individual band signal-to-noise ratios (SNR). The application of SPCA forces each band to contribute equal weight to the analysis since each band has equal variance. Since the original variables have been scaled in the standardization process, and linear transformations are not invariant under such scalings, the eigenvectors and principal components of  $\Sigma_x$  will not be the same as those of  $\mathbf{R}_x$  (Singh and Harrison, 1985, p. 888). Singh and Harrison assert that the eigenvectors of  $\mathbf{R}_x$  are equally sensitive to all bands irrespective of the SNR in the original data, and hence provide an unbiased set of eigenvalues. Figure 4.22 shows the eigenvalue behavior of the Davis Monthan correlation matrix. In comparing the shape of the eigenvalue curves in Figure 4.22 with those of Figure 4.6, it can be noted that the dynamic range of the correlation matrix eigenvalues is smaller by an order of magnitude than that of the covariance matrix eigenvalues. Also, the shape of the correlation matrix eigenvalue curve is somewhat smoother than the covariance matrix eigenvalue curve. Singh and Harrison (1985) state that SPCA actually improves the visual contrast in each successive component image to a greater extent than PCA. Figure 4.23, also found in Appendix A in color, is presented for comparison with Figure 4.4, to further explore this claim. The most obvious aspect of the standardized PC images is the remarkably better



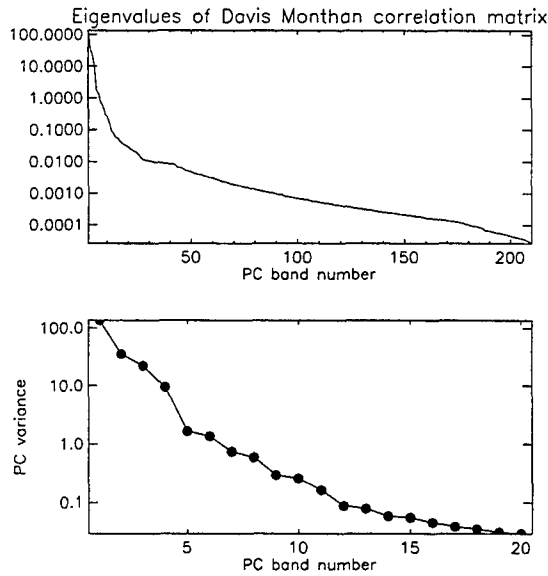


Figure 4.22: Eigenvalue Behavior of the Davis Monthan Correlation Matrix.

image quality than the PC images. While in Figure 4.4 the effects of noise become apparent after the eighth PC image, Figure 4.23 shows no such effects in the 25 standardized PC images. Recalling Ready and Wintz's (1973) argument about SNR improvement using the PC transform, Singh and Harrison (1985) state that the SPCA improves the SNR to a greater extent because the maximum variance, which is the only variance in a correlation matrix, is always one. Thus, the improvement in SNR achieved by SPCA is greater than that which can be gained by PCA as is shown in the following relation:

$$\Delta SNR_{PCA} = \frac{\lambda_{1_{PCA}}}{\sigma_{x_{max}}^2} \leq \Delta SNR_{SPCA} = \frac{\lambda_{1_{PCA}}}{1} \quad (4.14)$$

Note that the eigenvalues in this relation are different because they are associated with different second order statistics. As Figure 4.22 illustrated, the eigenvalues of the correlation matrix are smaller than those of the covariance matrix. The difference implies that the information conveyed by the eigenvalues and eigenvectors of the correlation matrix is not equivalent to that conveyed by the covariance matrix. The normalization of the data by its variance creates new characteristics that are not simple linear scalings.

A further appreciation for the effects of using the correlation matrix is gained by examining the entropy and eigenvector behavior of the SPCA transform and contrasting this with the PCA transform. Figure 4.24 shows the variance and entropy behavior of the SPCA transform for the Davis Monthan data. Comparing this figure with Figure 4.8, the

25 standardized PC Images using Radiance Correlation — Davis Monthan



Figure 4.23: First 25 Standardized PC images of the Davis Monthan Scene Using the Correlation Matrix.

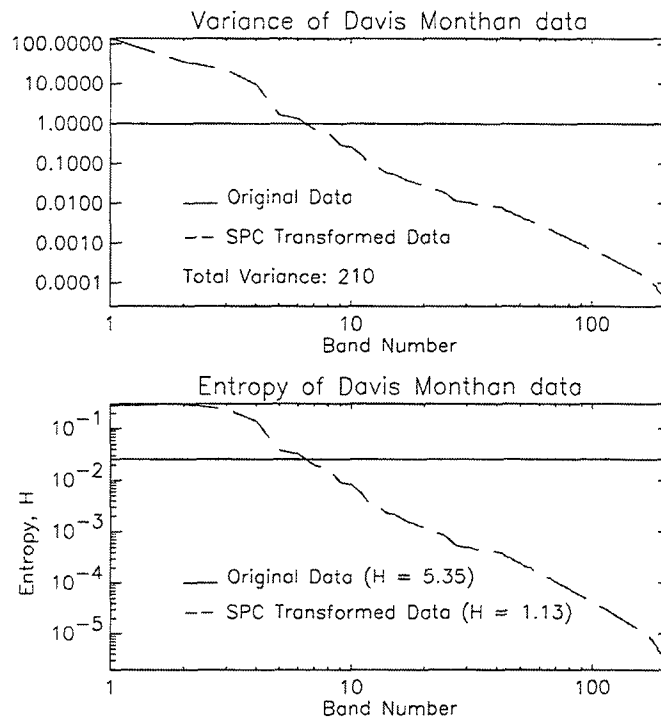


Figure 4.24: Variance and Entropy Behavior of the Davis Moonthan Correlation Matrix.

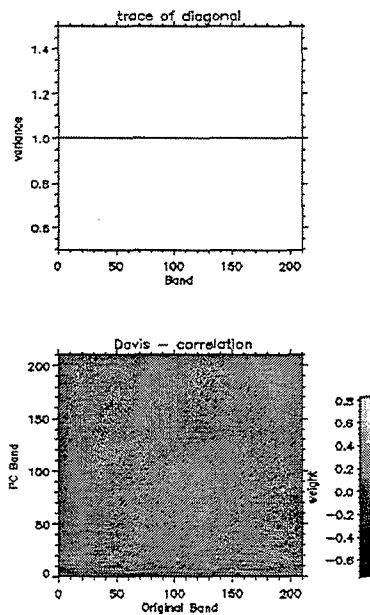


Figure 4.25: Eigenvectors and Trace of the Correlation Matrix of Davis Moonthan Scene.

unit variance and entropy of the original data are immediately apparent. The behavior of the transformed variance and entropy is much like that of the PC transform in Figure 4.8. The behavior of the eigenvectors is shown in Figure 4.25. Recall from Figure 4.11, the eigenvectors used in PCA, that a checkerboard pattern was apparent. This corresponded to a heavy weighting of original bands one to 100 by the low numbered eigenvectors and a weighting of the higher numbered original bands by the higher numbered eigenvectors. In Figure 4.25, also found in Appendix A, the checkerboard pattern is not apparent. The emphasis of the eigenvectors is more distributed. This is further depicted in Figure 4.26, where the surface plot of the eigenvectors is shown. There appears to be no clear trend of

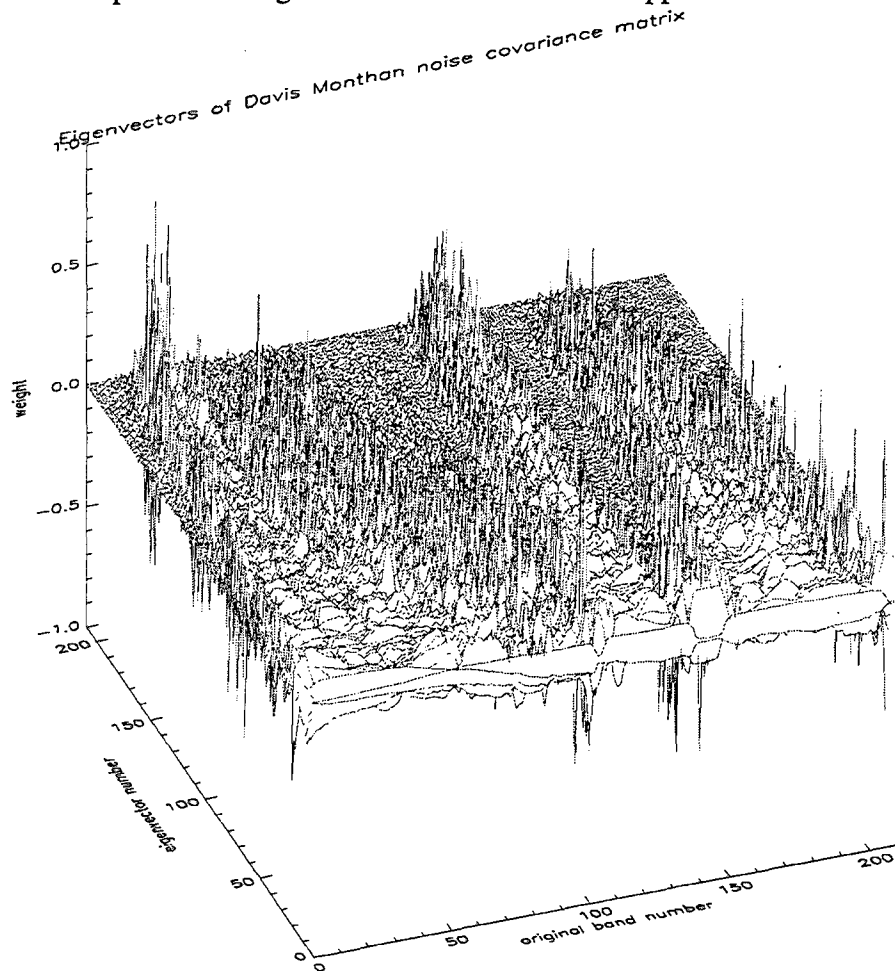


Figure 4.26: Eigenvectors of the Davis Monthan Correlation Matrix.

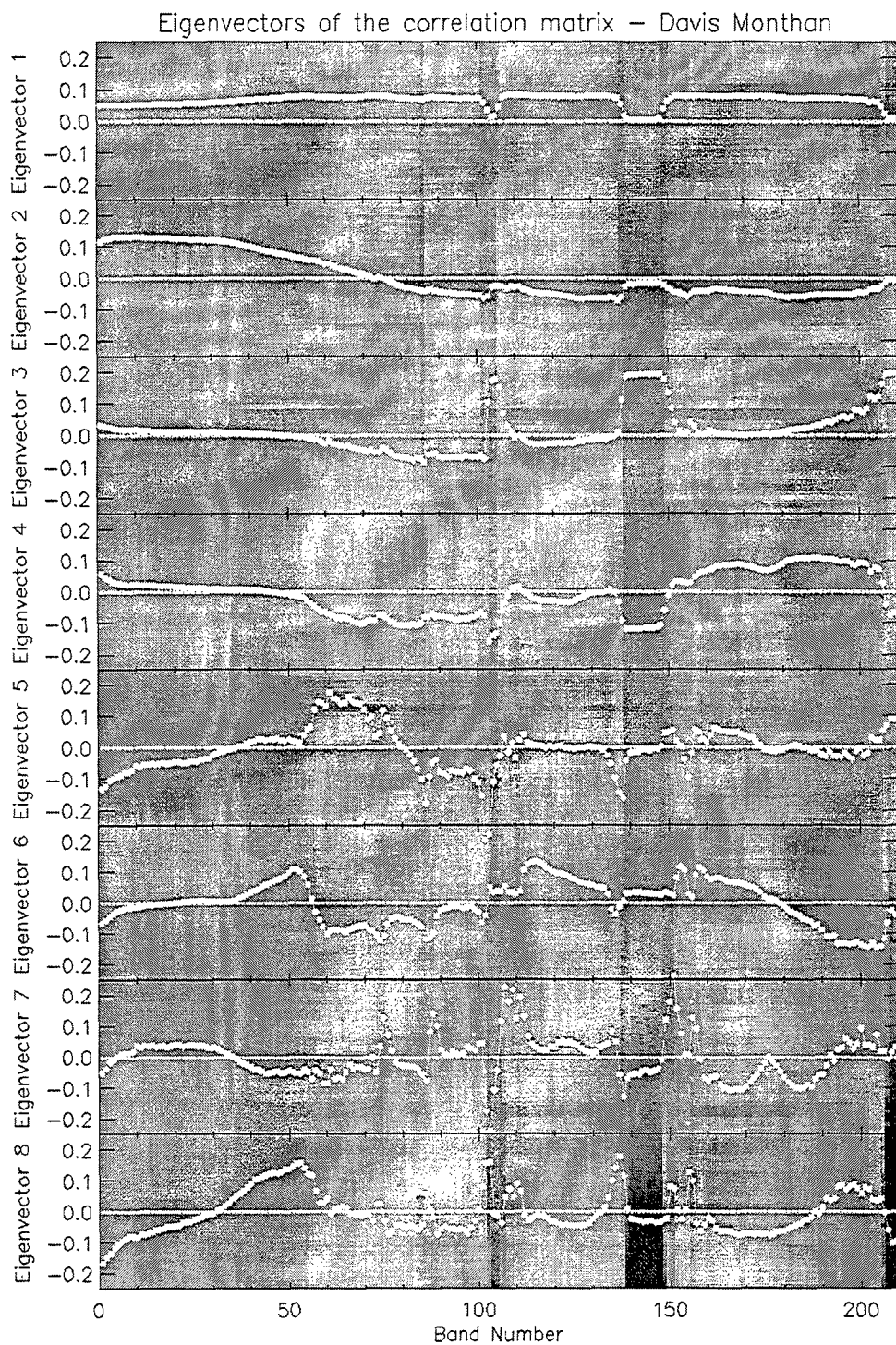


Figure 4.27: First Eight Eigenvectors of Davis Monthan Normalized Scene Superimposed on a Random Slice Across the Hypercube.

eigenvector weights in this surface plot. Several of the eigenvectors place emphasis on the very first and last original bands. The behavior of the first eight eigenvectors presented against a normalized slice of the Davis Monthan hypercube is shown in Figure 4.27. The color version of this figure may be found in Appendix A. The first eigenvector places a uniform weighting on all original bands except the last few and the absorption bands. This underscores the fact that SPCA is totally unbiased in the formation of component images. In quite different behavior than their covariance counterparts, the eigenvectors of the correlation matrix actually weight the absorption bands in the first few eigenvectors. Some of the weights around band 50 are similar to the PCA eigenvectors, but the behavior of the first eight is mostly different than the eigenvectors of the covariance matrix.

The three techniques within the PCA family are based on subtle changes to the basic PCA technique. One means of further understanding the implications of each technique is to show the effect of each on the same image pixel. Figure 4.28 does this by

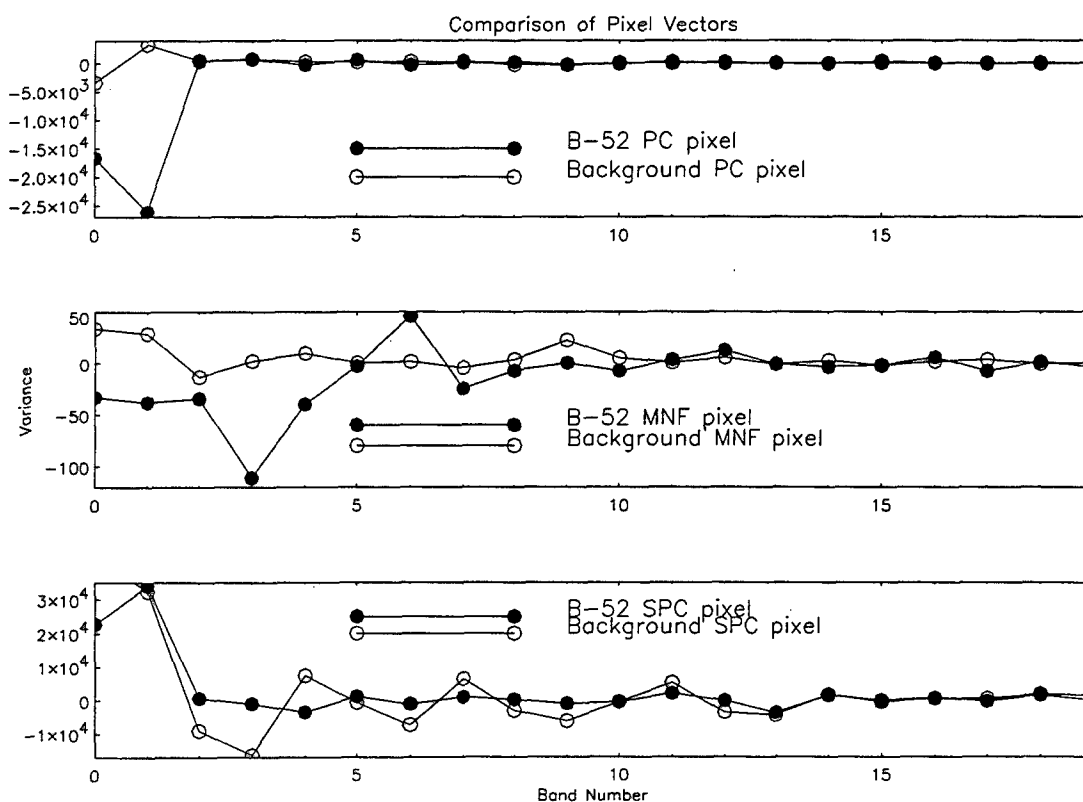


Figure 4.28: Comparison of Pixel Vectors from Component Images of the PCA Family of Techniques.

portraying a B-52 pixel vector and a background pixel vector from the component images of the PCA, MNF, and SPCA techniques. The trend in all techniques is to order the components by decreasing magnitude. Components with large variances appear in the early bands with a sharp decrease in variance by about the tenth band. The SPCA pixel vectors display the greater dynamic range, while those of the MNF display the least. This is reflected in the image quality achieved by each technique. A higher dynamic range equates to better contrast and image quality. In general, the target vector appears to display more oscillation into higher band numbers than the background vector.





## V. THE MATCHED FILTER FAMILY OF TECHNIQUES

### A. DESCRIPTION

Complete *a priori* knowledge of the scene endmembers reduces the problem of detecting a target spectral signature to one very much like that of the matched filter scenario in communications and signal processing. The matched filter developed below is based on problem of detecting a deterministic signal in white noise. The problem is not as simple when dealing with target detection in a hyperspectral image due to the effect of multiple interfering background material signatures as is illustrated in the mixed pixel concept of Figure 2.2. Four hyperspectral imagery analysis techniques that use a matched filter approach are introduced and developed in this chapter. They are the simultaneous diagonalization filter, the orthogonal subspace projection, the least squares orthogonal subspace projection, and the filter vector algorithm. The techniques deal with the problem of multiple interfering signatures in a deterministic fashion derived from the theory of least squares. The significance of these techniques is that given *a priori* information regarding the endmembers of the scene, the target detection problem can be reduced to the matched filter problem.

### B. BACKGROUND DEVELOPMENT

In order to understand the motivation behind this family of techniques, it is necessary to study its origins. They share a common theme that is based on the signal processing idea of detecting a known signal embedded in noise. The starting point in this development is the spatially invariant image sequence, which is a general model that is applicable to any problem which assumes mixed pixels. Next, the theory of least squares is viewed geometrically and in terms of subspaces. Finally, the matched filter is developed from a signal-to-noise ratio (SNR) perspective. These concepts are central to the matched filter family of techniques.

#### 1. Linearly Additive Spatially Invariant Image Sequences

Miller, Farison, and Shin (1992) introduce the idea of image sequences in a context general enough to include multispectral images as a subset. Their definition is

insightful, and is briefly described here. An image sequence is defined as a series of images obtained by varying a property of the imaging system so that the intensity of the image features changes from image to image. The spatial invariance property comes from the fact that all image features are in the same spatial position in each image of the sequence. There are three salient types of these sequences: 1) functional images, where temporal changes within an object are traced by making successive images over time, 2) parametric images, that result by varying some parameter of the imaging device over successive images, and 3) multispectral images, in which successive images are formed by imaging in specific spectral ranges (Miller, Farison, and Shin, 1992, pp. 148-149). The linearly additive property of these sequences is based on the notion that a finite number of image formation processes (endmembers in hyperspectral terminology) contribute linearly and additively to each image of the sequence. If the vector  $s_m$  is defined as the  $m^{\text{th}}$  image formation process (the  $m^{\text{th}}$  endmember), then the observed pixel vector of brightness levels is given by:

$$\mathbf{x} = \sum_m \alpha_m \mathbf{s}_m \quad (5.1)$$

The summation is over all of the image formation processes, and the scalar  $\alpha_m$  describes the relative abundance of the  $m^{\text{th}}$  process (endmember) at a given spatial location (Miller, Farison, and Shin, 1992, p. 149). In multispectral imagery, such a situation occurs when a single pixel of a given image may cover a variety of constituent scene elements due to a large GIFOV. This situation is illustrated using the Davis Monthan scene. The same two pixels that were chosen to show the effects of the PCA family of techniques in Figure 4.28 are shown in Figure 5.1. The target pixel vector corresponds to a pixel in the middle of a B-52 aircraft wing. The other pixel is from the prevalent sandy background. These two pixels represent relatively "pure" pixels because they are not taken from areas where aircraft and background mixing might occur. Equation 5.1 is illustrated in Figure 5.1 by mixing the two pixel vectors equally, so that each component of the relative abundance vector,  $\alpha$ , is 0.5. The definition of Equation 5.1 forms the basis of the mixed pixel problem, which is to detect one of the endmembers over all pixels in the scene. Mixing occurs to some extent over all pixels of the scene. This is partially due to the finite GIFOV and also the natural variability of spectra.

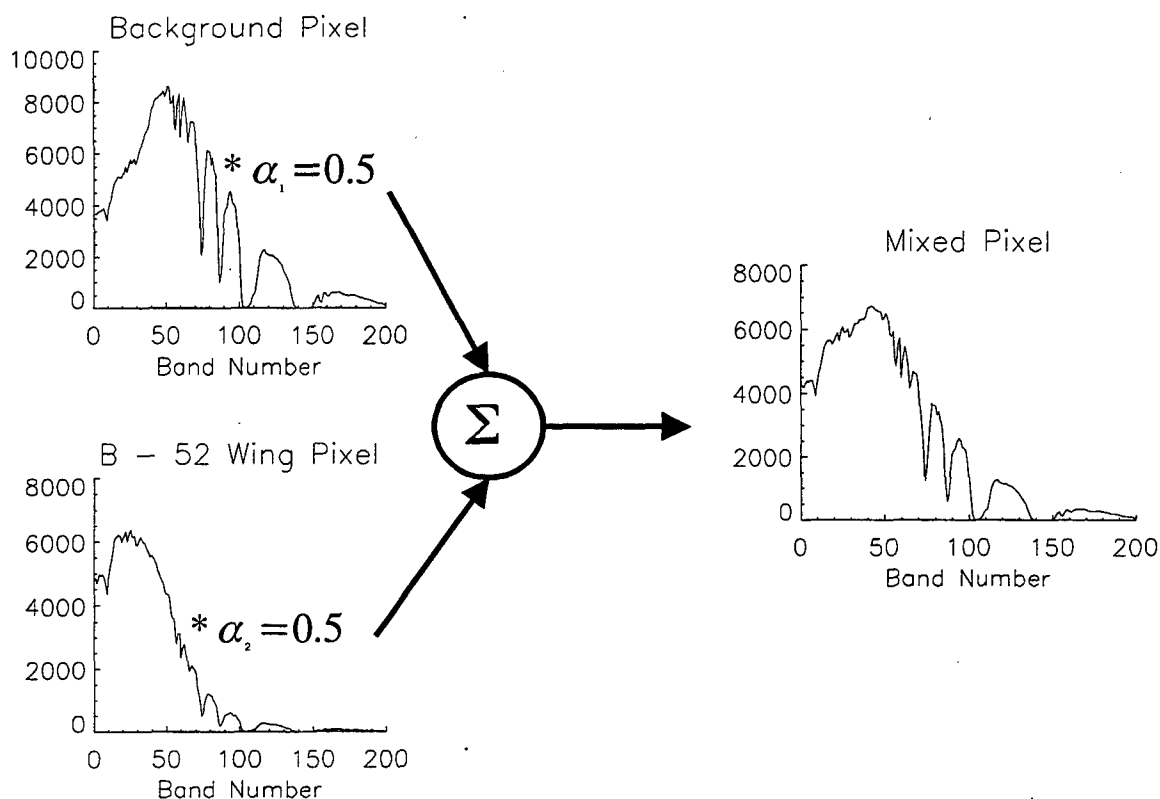


Figure 5.1: Linear Mixing of Target and Background Spectra.

## 2. The Theory of Least Squares: The *A Priori* Model

The least squares problem is one which arises in many scientific investigations when the task is to find the linear function which best “fits” a given set of data points (Watkins, 1991, p. 135). Goodness of fit is determined by the sum of the squares of the residuals, where a residual is defined as the difference between an estimated and a true quantity. In signal processing applications, the problem description is very similar, and the goal is to fit a signal model to the observations in such a manner that the residual error between model and observations is minimized. The major philosophical difference between a least squares approach and a classical statistically-based signal detector or estimator is that least squares works with observed data as opposed to known or estimated statistics (Therrien, 1992, p. 518). Scharf (1991) develops the least squares problem and optimal solution in a manner which lends itself very nicely to application in the matched filter family of hyperspectral imagery analysis techniques. His development

will be followed here, though the original notation has been altered to be consistent with that followed by this study.

Assume that the observations,  $\mathbf{x}$ , consist of a signal component vector,  $\mathbf{s}$ , and a noise component vector,  $\mathbf{n}$ . The vector equation describing this situation consists of the  $l$ -dimensional column vectors:

$$\mathbf{x} = \mathbf{s} + \mathbf{n} \quad (5.2)$$

The key to least squares is the assumption that the signal component can be modeled by a linear equation

$$\mathbf{s} = \mathbf{M}\alpha \quad (5.3)$$

where  $\mathbf{M}$  is an  $l \times p$  matrix describing the dynamics or modes of the signals, and  $\alpha$  is a  $p$ -dimensional vector of unknown parameters (Scharf, 1991, p.360). In the notation paradigm used in this study, Equation 5.3 implies that the signal vector  $\mathbf{s}$  is composed of various proportions of endmembers, contained in matrix  $\mathbf{M}$ . The noise vector  $\mathbf{n}$  can be viewed as the residual or error produced by fitting the model  $\mathbf{M}\alpha$  to the observed data  $\mathbf{x}$ . If we choose to view the matrix  $\mathbf{M}$  as a collection of  $p$   $l$ -dimensional column vectors, as:

$$\mathbf{M} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{m}_1 & \cdots & \mathbf{m}_p \\ \downarrow & & \downarrow \end{bmatrix} \quad (5.4)$$

then the signal vector,  $\mathbf{s}$ , can be represented as a linear combination of the columns of  $\mathbf{M}$ :

$$\mathbf{s} = \sum_{n=1}^p \alpha_n \mathbf{m}_n \quad (5.5)$$

where  $\alpha_n$  is a scalar parameter (Scharf, 1991, p.361). The problem is to find the parameter vector,  $\alpha$ , that fits the model,  $\mathbf{M}\alpha$ , to the observation vector,  $\mathbf{x}$ , in the least squares sense. In the case where the number of measurements (dimensions of the observation vector or bands in spectral imagery),  $l$ , are greater than the number of parameters,  $p$ , an exact fit of the model to the data is not possible, and a least squares fit must be employed in this overdetermined situation.

The least squares solution to the problem is found by minimizing the squared error between  $\mathbf{x}$  and  $\mathbf{M}\alpha$ . This error is formulated as:

$$e^2 = (\mathbf{x} - \mathbf{M}\alpha)^T (\mathbf{x} - \mathbf{M}\alpha) = \mathbf{n}^T \mathbf{n} \quad (5.6)$$

and is minimized by equating the gradient to zero and solving for  $\alpha$ :

$$\frac{\partial}{\partial \alpha} e^2 = 2\mathbf{M}^T (\mathbf{x} - \mathbf{M}\alpha) = 0 \quad (5.7)$$

Solution of Equation 5.7 leads to the optimal least squares solution:

$$\hat{\alpha} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{x} \quad (5.8)$$

(Scharf, 1991, p. 365). This estimate of the parameter vector may be used to estimate the signal vector as:

$$\hat{\mathbf{s}} = \mathbf{M} \hat{\alpha} = \mathbf{P}_M \mathbf{x} \quad (5.9)$$

The  $\mathbf{P}_M$  is termed an orthogonal projector and after a rearrangement and substitution of terms can be defined as:

$$\mathbf{P}_M = \mathbf{M}(\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \quad (5.10)$$

(Scharf, 1991, p. 366). The estimated noise is the difference between the measurements and the estimate of the signal. This relationship is given by:

$$\hat{\mathbf{n}} = \mathbf{x} - \hat{\mathbf{s}} = (\mathbf{I} - \mathbf{P}_M) \mathbf{x} = \mathbf{P}_A \mathbf{x} \quad (5.11)$$

Equation 5. 11 introduces the projector  $\mathbf{P}_A$ , which Haykin (1996) calls the orthogonal complement projector. The significance of the projectors will become clear shortly. When the  $p$  columns of  $\mathbf{M}$  are linearly independent, then only  $l-p$  linearly independent vectors can be orthogonal to them. If these orthogonal vectors are organized into the  $l \times (l-p)$  matrix  $\mathbf{A}$  as:

$$\mathbf{A} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{a}_1 & \cdots & \mathbf{a}_{l-p} \\ \downarrow & & \downarrow \end{bmatrix} \quad (5.12)$$

where each column of  $\mathbf{A}$  is orthogonal to all columns of  $\mathbf{M}$ , then a vector  $\mathbf{u}$  can be represented as:

$$\mathbf{u} = \mathbf{A} \phi \quad (5.13)$$

where  $\phi$  is an  $l-p$  parameter vector (Scharf, 1991, p. 367). Further, the orthogonal complement projector can be written in terms of  $\mathbf{A}$  as

$$\mathbf{P}_A = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \quad (5.14)$$

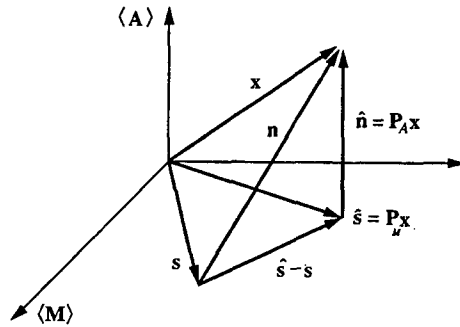


Figure 5.2: Least Squares Illustrated Geometrically. After Scharf, 1991, p. 367.

The best way to describe the above discussion is with a picture. As such, Figure 5.2 illustrates the concepts of projectors and least squares with a geometric representation of subspaces to put the above discussion into perspective. The  $xy$ -plane in Figure 5.2 represents the signal subspace,  $\langle \mathbf{M} \rangle$  which is spanned by the signal system model matrix,  $\mathbf{M}$ . It represents all vectors  $\mathbf{s} = \mathbf{M}\alpha$  that are linear combinations of the columns of  $\mathbf{M}$ . The orthogonal subspace,  $\langle \mathbf{A} \rangle$ , is represented in the figure by the vertical axis. This represents all the vectors  $\mathbf{u}$  that are orthogonal to the columns of  $\mathbf{M}$ . The projectors  $\mathbf{P}_M$  and  $\mathbf{P}_A$  decompose Euclidean space into the signal subspace  $\langle \mathbf{M} \rangle$  and its orthogonal complement subspace  $\langle \mathbf{A} \rangle$ , which implies that any arbitrary vector can be decomposed into the sum of a component projected onto  $\langle \mathbf{M} \rangle$  by  $\mathbf{P}_M$  and a component projected onto  $\langle \mathbf{A} \rangle$  by  $\mathbf{P}_A$ . The decomposition of observation vector  $\mathbf{x}$  is given by:

$$\mathbf{x} = \mathbf{P}_M \mathbf{x} + \mathbf{P}_A \mathbf{x} = \hat{\mathbf{s}} + \hat{\mathbf{n}} \quad (5.15)$$

This orthogonal decomposition is depicted in Figure 5.2 by  $\hat{\mathbf{s}}$ , the estimated signal vector which lies in the subspace spanned by  $\mathbf{M}$ , and the  $\hat{\mathbf{n}}$ , the estimated noise vector that is orthogonal to every vector in the signal subspace. Similarly, the noise vector,  $\mathbf{n}$ , is decomposed into the orthogonal components  $\hat{\mathbf{n}}$  and  $(\hat{\mathbf{s}} - \mathbf{s})$  (Scharf, 1991, p. 368). The principal of least squares is summarized in geometric terms by observing that there is no value of  $\hat{\mathbf{s}}$  in the signal subspace that provides a smaller norm of the estimated error than that generated by the orthogonal decomposition of the observations. This orthogonalization procedure is used in the orthogonal subspace projection (OSP) technique to form the least squares optimal interference rejection operator,  $\mathbf{P}$ .

### 3. The Theory of Least Squares: The *A Posteriori* Model

There is another view of orthogonal projections that is described by Scharf (1991) and utilized in the least squares orthogonal subspace projection technique of Tu, Chen, and Chang (1997). In it, the *a priori* model:

$$\mathbf{x} = \mathbf{M}\alpha + \mathbf{n} \quad (5.16)$$

is replaced by the *a posteriori* model also seen in Equation 5.15:

$$\mathbf{x} = \mathbf{P}_M \mathbf{x} + \mathbf{P}_A \mathbf{x} \quad (5.17)$$

where the projection of observations onto the signal subspace:

$$\hat{\mathbf{s}} = \mathbf{P}_M \mathbf{x} \quad (5.18)$$

produces the fitting error:

$$\hat{\mathbf{n}} = \mathbf{P}_A \mathbf{x} = (\mathbf{x} - \hat{\mathbf{s}}) \quad (5.19)$$

of minimum norm. The *a priori* model is often used in the signal-in-noise detection problem. When the observations are not available at the beginning of processing, as is often the case in real-time signal processing, this model is useful. When the observations are available before processing, as in the processing of a remotely sensed image, the *a posteriori* model can be used to improve the output SNR. The key point of the *a posteriori* model is that if we can predict the signal prior to processing as  $\hat{s}$ , then we only need to process the prediction error,  $\hat{n}$ , which is caused by an inaccurate prediction plus additive noise (Tu, Chen, and Chang , 1997, p. 128). The *a posteriori* model is restated as:

$$\mathbf{x} = \hat{\mathbf{s}} + \hat{\mathbf{n}} = \hat{\mathbf{s}} + (\mathbf{x} - \hat{\mathbf{s}}) \quad (5.20)$$

If a signal can be estimated completely from the observation, then the prediction error  $\hat{n}$  must be orthogonal to the estimated signal, and further it is completely unpredictable and contains no information that could be retrieved from the observation. This is the case with the least squares estimator.

#### 4. The Matched Filter

Determining an optimum finite impulse response filter that maximizes SNR is a fundamental issue in communications theory (Haykin, 1996, p. 2). The solution to this problem employs the generalized eigenvalue problem, and is applicable equally to maximizing the output SNR of a filter which detects a random or a deterministic signal buried in noise. Since our outlook in the matched filter family of techniques is deterministic, only the statistics of the noise are assumed. Therrien's (1992) derivation of the matched filter from the perspective of maximizing the SNR of a known deterministic signal in noise, is followed, with the exception that all signals are real in this case and the notation has been tailored to fit this study.

The input to the filter is the vector  $\mathbf{x}$ , which is convolved with the impulse response of the filter,  $\mathbf{w}$ , to give the output vector  $\mathbf{y}$ . Figure 5.3 illustrates this situation.

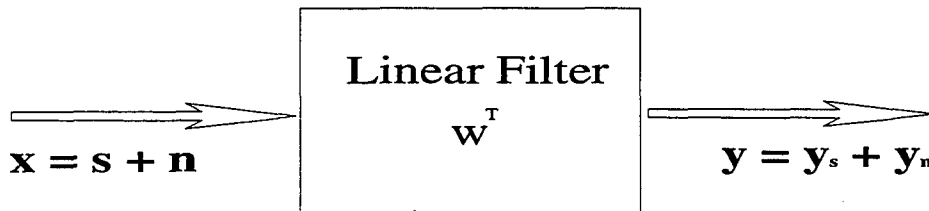


Figure 5.3: Simple Linear Filter Showing Signal and Noise Components.

It is assumed that the filter in question is a linear filter. If the input can be represented as a sum of signal and noise,  $\mathbf{x} = \mathbf{s} + \mathbf{n}$ , then the output can similarly be represented as a sum of response due to the signal and response due to noise,  $\mathbf{y} = \mathbf{y}_s + \mathbf{y}_n$ . The goal is to maximize the SNR, which is defined as the ratio of output signal energy to noise energy, defined as:

$$\text{SNR} = \frac{|\mathbf{y}_s|^2}{|\mathbf{y}_n|^2} = \frac{|\mathbf{w}^T \mathbf{s}|^2}{E\{|\mathbf{w}^T \mathbf{n}|^2\}} = \frac{(\mathbf{w}^T \mathbf{s})(\mathbf{s}^T \mathbf{w})}{E\{(\mathbf{w}^T \mathbf{n})(\mathbf{n}^T \mathbf{w})\}} = \frac{\mathbf{w}^T \mathbf{s} \mathbf{s}^T \mathbf{w}}{\mathbf{w}^T \Phi_n \mathbf{w}} \quad (5.21)$$

$\Phi_n$  is the correlation matrix of the noise, which is formed as a consequence of the statistical description of the noise (Therrien, 1992, p. 242). The correlation matrix in this context is the signal processing version. In order to simplify the process of finding a filter impulse response weight vector,  $\mathbf{w}$ , to maximize the SNR, we constrain the denominator to be equal to unity. Now the SNR may be written as:

$$\text{SNR} = \mathbf{w}^T \mathbf{s} \mathbf{s}^T \mathbf{w} \quad (5.22)$$

The SNR is maximized by using Lagrangian multipliers and setting the gradient with respect to  $\mathbf{w}$  equal to zero as shown below:

$$\begin{aligned} \mathcal{L} &= \mathbf{w}^T \mathbf{s} \mathbf{s}^T \mathbf{w} + \lambda(1 - \mathbf{w}^T \Phi_n \mathbf{w}) \\ \nabla_{\mathbf{w}} \mathcal{L} &= \mathbf{s} \mathbf{s}^T \mathbf{w} - \lambda \Phi_n \mathbf{w} = 0 \Rightarrow \mathbf{s} \mathbf{s}^T \mathbf{w} = \lambda \Phi_n \mathbf{w} \end{aligned} \quad (5.23)$$

The last equation is a generalized eigenvalue problem involving the matrices  $\mathbf{s} \mathbf{s}^T$  and  $\Phi_n$ , and  $\mathbf{w}$  is termed the generalized eigenvector (Therrien, 1992, p. 242). The matrix  $\mathbf{s} \mathbf{s}^T$  has only one linearly independent column, which implies that it has rank one. The generalized eigenvalue equation can be reconfigured by premultiplying both sides by  $\Phi_n^{-1}$  to obtain

$$(\mathbf{s}^T \Phi_n^{-1} \mathbf{s}) \mathbf{w} = \lambda \mathbf{w} \quad (5.24)$$

The matrix on the left hand side multiplying  $\mathbf{w}$  is also of rank one, implying that it has only one nonzero eigenvalue. By equating the left and right sides of the equation, this eigenvalue is seen to be

$$\lambda_{\max} = \mathbf{s}^T \Phi_n^{-1} \mathbf{s} \quad (5.25)$$

and represents the maximum SNR. The eigenvector associated with this eigenvalue is proportional to  $\Phi_n^{-1} \mathbf{s}$  (Therrien, 1992, p. 243). Thus, the matched filter vector,  $\mathbf{w}$ , that maximizes the SNR corresponds to the generalized eigenvector associated with the largest eigenvalue. The matched filter vector is seen to be just a scaled version of the desired signature vector,  $\mathbf{s}$ . The process of deriving the optimal matched filter has also



been referred to as eigenfiltering because of the dependence on eigenanalysis (Haykin, 1996, p. 181).

### C. OPERATION

The simultaneous diagonalization (SD) filter was developed by Miller in 1982 to filter linearly additive spatially invariant image sequences to enhance a desired feature while suppressing undesired features. In the target detection context of hyperspectral imagery, the goal is to produce a single-band image which contains information regarding the abundance of a particular target spectrum in every spatial pixel. While the SD filter of Miller, Farison, and Shin (1992) approaches the problem by deriving an optimal filter vector using an energy ratio and the generalized eigenvalue problem, Harsanyi (1993) obtains equivalent results in the orthogonal subspace projection (OSP) technique by breaking the process into two steps. The first step employs an optimal least squares projection operator to minimize the undesired signature energy, and the second maximizes the SNR to find the optimal filter vector. The OSP technique is actually a special case of the SD filter when the additive noise variance is zero (Harsanyi and Chang, 1994, p. 781). The LSOSP technique improves the OSP output SNR by using a slightly different model. The filter vector algorithm directly applies the ideas of matched filters. All of the techniques are deterministic in their view of the data and require full knowledge of the endmember spectra of the target and the background interfering signatures.

#### 1. The Simultaneous Diagonalization (SD) Filter

The goal of the SD filter is to perform linear filtering on an image sequence to obtain a new image in which the scalar values at each pixel location are represented as:

$$y = \mathbf{w}^T \mathbf{x} = \langle \mathbf{w}, \mathbf{x} \rangle \quad (5.26)$$

The filtering operation is analogous to forming the inner product of each observed pixel vector  $\mathbf{x}$  with the filter vector  $\mathbf{w}$ , which has been optimized to maximize the output SNR of the target spectrum. The magnitude of the gray scale values,  $y$ , in the filtered image corresponds to the pixel vectors that have a significant correlation with the filter vector,  $\mathbf{w}$  (Miller, Farison, and Shin, 1992, p. 150).

The nomenclature that Miller, Farison, and Shin (1992) use in describing the SD filter is recast here in terms of spectral imagery analysis. The vector  $\mathbf{x}$  represents a mixed

pixel that is composed of varying abundances,  $\alpha_m$ , of the endmember spectra,  $s_m$ . There are a total of  $m = 1, \dots, p$  endmembers contributing to the formation of mixed pixels in the hypercube. There are  $l$  spectral bands. The assumption is made that the image is corrupted by white noise with zero mean and zero interpixel correlation, represented by the  $l$ -dimensional vector  $\mathbf{n}$ . The filter vector which is derived is also an  $l$ -dimensional vector (Miller, Farison, and Shin, 1992, p. 150). The observed pixel vector is written in accordance with the spatially invariant image sequence model with noise added:

$$\mathbf{x} = \sum_m \alpha_m \mathbf{s}_m + \mathbf{n} \quad (5.27)$$

The scalar image resulting from filtering this noisy image is explicitly written as the inner product:

$$y = \langle \mathbf{w}, \sum_m \alpha_m \mathbf{s}_m \rangle + \langle \mathbf{w}, \mathbf{n} \rangle = \sum_m \alpha_m \langle \mathbf{w}, \mathbf{s}_m \rangle + \langle \mathbf{w}, \mathbf{n} \rangle \quad (5.28)$$

These representations of the filtered image show that the SD filter is achieving two objectives at the same time: 1) collecting the abundance information about a target endmember spectrum in the image, and 2) suppressing interfering endmember spectra and additive noise (Miller, Farison, and Shin, 1992, p. 150).

Assuming that all endmember spectra are known, let  $\mathbf{d}$  represent the spectrum of the target endmember material and  $\mathbf{U}$  represent a matrix of undesired interfering endmember spectra, the columns of which are individual endmember spectra. Further, assume that the noise has a covariance matrix given by  $\sigma^2 \mathbf{I}$ . The derivation of the optimal filter vector  $\mathbf{w}$  begins with the formation of an energy ratio of desired energy to undesired and noise energies. Energy in this context is taken to be the sum of the squares of the vector components. The energy ratio of desired to undesired is defined as:

$$r_E(\mathbf{w}) = \frac{E_d}{E_u + E_n} = \frac{\langle \mathbf{w}, \mathbf{d} \rangle^2}{\langle \mathbf{w}, \mathbf{U} \rangle^2 + \langle \mathbf{w}, \mathbf{n} \rangle^2} = \frac{\mathbf{w}^T \mathbf{d} \mathbf{d}^T \mathbf{w}}{\mathbf{w}^T (\mathbf{U} \mathbf{U}^T + \sigma^2 \mathbf{I}) \mathbf{w}} = \frac{\mathbf{w}^T \mathbf{A} \mathbf{w}}{\mathbf{w}^T \mathbf{B} \mathbf{w}} \quad (5.29)$$

where energies are expressed as the square of the inner product of a vector or matrix with the filter vector (Miller, Farison, and Shin, 1992, pp. 150-151). The  $l \times l$  matrices  $\mathbf{A}$  and  $\mathbf{B}$  are employed for ease of mathematical manipulation. Miller, Farison, and Shin (1992) observe that the ratio is a generalization of Rayleigh's quotient, and by setting the gradient of  $r_E(\mathbf{w})$  with respect to  $\mathbf{w}$  equal to zero, the energy ratio is transformed into the generalized eigenvector problem

$$\mathbf{A} \mathbf{w} = r_E(\mathbf{w}) \mathbf{B} \mathbf{w} \quad \Rightarrow \quad \mathbf{B}^{-1} \mathbf{A} \mathbf{w} = r_E(\mathbf{w}) \mathbf{w} \quad (5.30)$$

as in Therrien's (1992) derivation of the matched filter. The difference in this derivation is that its "noise" component (the denominator) includes undesired endmember vectors

whose energy must also be minimized. The objective is to obtain the filter  $\mathbf{w}_{\max}$  that will maximize the energy of the target spectrum relative to the multiple interfering endmember spectra and noise, so  $\mathbf{w}_{\max}$  is chosen to be the eigenvector associated with the largest eigenvalue,  $r_E(\mathbf{w}_{\max})$  (Miller, Farison, and Shin, 1992, p. 151). The name of the SD filter stems from the property of the eigenvectors of  $\mathbf{B}^{-1}\mathbf{A}$  to simultaneously diagonalize the matrices  $\mathbf{A}$  and  $\mathbf{B}$ .

The derivation of the optimum filter vector corresponds to that given in the development of the matched filter. The matrix  $\mathbf{B}^{-1}\mathbf{A}$  is singular because the matrix  $\mathbf{d}\mathbf{d}^T$  has rank one. This implies that there is only one nonzero eigenvalue. The nonzero eigenvalue is given by the trace of the matrix  $\mathbf{B}^{-1}\mathbf{A}$  and is represented in expanded form as:

$$\mathbf{d}^T(\mathbf{U}\mathbf{U}^T + \sigma^2\mathbf{I})^{-1}\mathbf{d} \quad (5.31)$$

The filter vector  $\mathbf{x}$  must satisfy the eigenvector equation:

$$\begin{aligned} \mathbf{B}^{-1}\mathbf{A}\mathbf{x} &= \lambda\mathbf{x} \\ (\mathbf{U}\mathbf{U}^T + \sigma^2\mathbf{I})^{-1}\mathbf{d}\mathbf{d}^T\mathbf{w} &= \mathbf{d}^T(\mathbf{U}\mathbf{U}^T + \sigma^2\mathbf{I})^{-1}\mathbf{d}\mathbf{w} \\ &= \mathbf{w}\mathbf{d}^T(\mathbf{U}\mathbf{U}^T + \sigma^2\mathbf{I})^{-1}\mathbf{d} \end{aligned} \quad (5.32)$$

This implies that the eigenvector is the filter vector:

$$\mathbf{w} = \gamma(\mathbf{U}\mathbf{U}^T + \sigma^2\mathbf{I})^{-1}\mathbf{d} \quad (5.33)$$

where  $\gamma$  is a nonzero scalar (Miller, Farison, and Shin, 1992, p. 151). An alternative derivation using a subspace approach is also given by Miller, Farison, and Shin (1992). It incorporates some of the elements of the least squares approach to orthogonal subspaces to arrive at the result that the filter vector is given by:

$$\mathbf{w} = \beta[\mathbf{I} - \mathbf{U}(\mathbf{U}^T\mathbf{U} + \sigma^2\mathbf{I})^{-1}\mathbf{U}^T]\mathbf{d} \quad (5.34)$$

where  $\beta$  is an arbitrary scalar (Miller, Farison, and Shin, 1992, p. 152). Note the similarity of the quantity in brackets to the orthogonal complement projector of Equation 5.11. The filter vector of Equation 5.34 is formed so that it incorporates this projector.

Two special cases of the SD filter are important in that they bring us back to the basic concepts that led to the SD filter. In the case of no interfering spectra,  $\mathbf{U} = \mathbf{0}$ , the optimum filter vector reduces to the result obtained for the matched filter:

$$\mathbf{w} = \beta\mathbf{d} \quad (5.35)$$

where the filter vector is a scaled version of the target spectrum. In the case of no additive noise,  $\sigma^2 = 0$ , the filter vector derived from the subspace interpretation results in the relation:

$$\lim_{\sigma^2 \rightarrow 0} \mathbf{w} = \beta [\mathbf{I} - \mathbf{U}(\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T] \mathbf{d} \quad (5.36)$$

in which the orthogonal complement projector of the least squares approach is evident in the brackets. Thus,  $\mathbf{w}$  in this case is in the direction of the target spectrum which is orthogonal to the undesired processes  $\mathbf{U}$  (Miller, Farison, and Shin, 1992, p. 152).

The common problem in the matched filter family of techniques is to identify the pixels which have nonzero abundances of the target spectrum present. The general nature of this problem lends itself to the investigation of similar problems in many fields. The SD filter was originally developed to process temporal x-ray image sequences which recorded the flow of a contrast medium injected into a vein or artery (Miller, Farison, and Shin, 1992, p. 148). The technique has been extended to processing various types of biomedical imagery. Miller, Farison, and Shin (1992) mention the usefulness of the SD filter to multispectral imagery, but do not illustrate the specific application of the technique. Their demonstration of the SD filter consists of extracting an endmember from a group of two to four endmembers embedded in a simulated image sequence.

This study demonstrates the SD filter as broad-based technique of which the OSP technique is a special case. In order to establish the connection between these two techniques, a 100 x 100 pixel sub-scene of the Davis Monthan image is chosen. Figure 5.4 shows the sub-scene image used for this discussion. It is a monochromatic image formed using band 80 of the Davis Monthan sub-scene, and clearly shows four B-52 aircraft against a fairly uniform background. The image also shows white boxes around the pixels corresponding to four various types of pixel vectors. The desired endmember is chosen as the pixel chosen from the aircraft wing. The other pixels are chosen from an aircraft fuselage, an engine nacelle, and an aircraft nose. These last two pixels represent mixed pixels that have occurred because elements of aircraft skin and ground are mixed within the same spatial area defining a pixel. These pixels appear darker than the background in the band 80 image of Figure 5.4. A color version of Figure 5.4 may be found in Appendix A, where the four pixels are indicated by different box colors. Figure 5.5 shows the spectra associated with the chosen pixels. The logarithm of the pixel brightness values has been taken to accentuate the subtle differences that exist between spectrally pure and mixed pixels over all 210 bands. Additionally, the plots corresponding to each pixel vector have been offset to make it easier to compare spectral details. The pixel vectors chosen from the aircraft wing and fuselage correspond to the ideal pure target pixel vector. Note the similar shape of these two spectra. The pixel

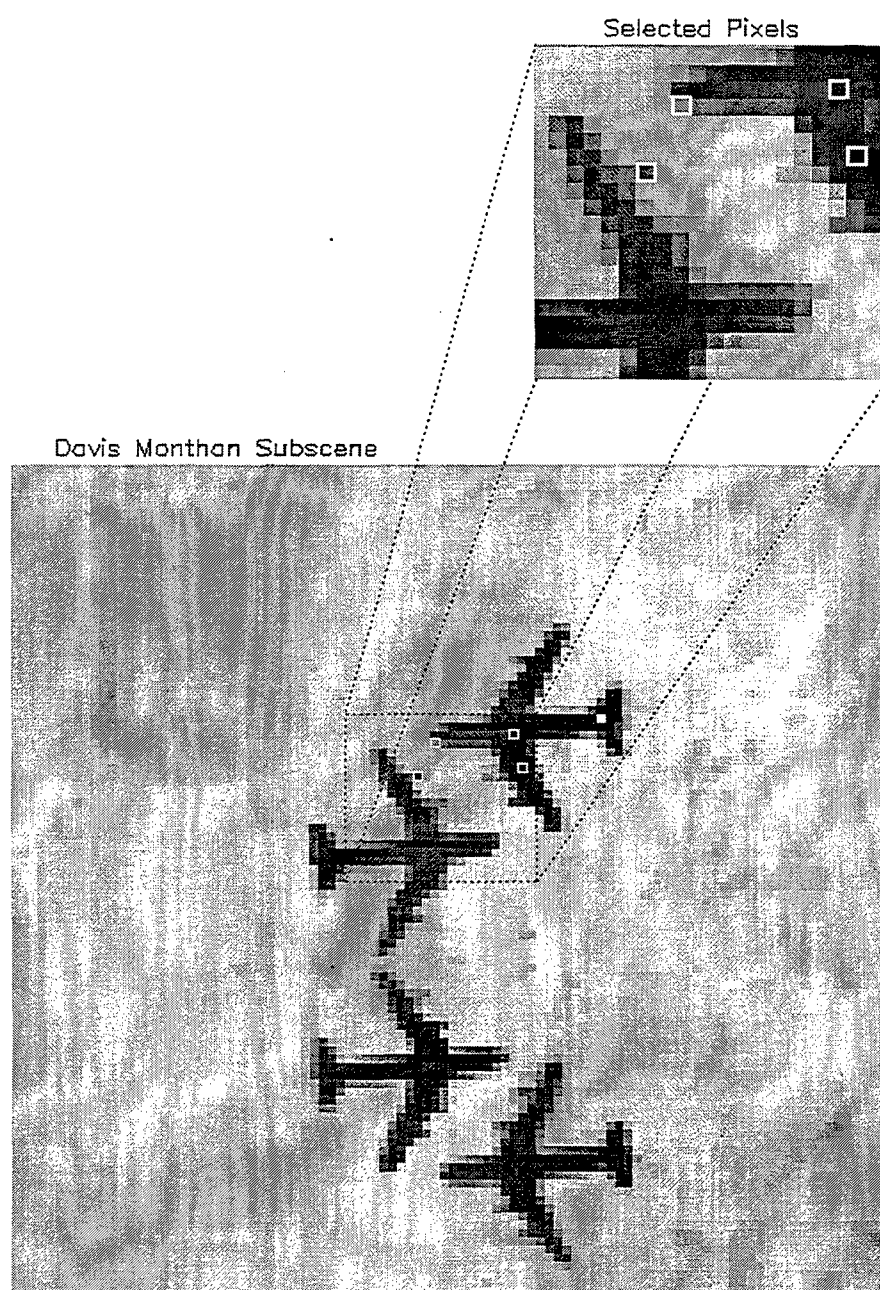


Figure 5.4: Sub-scene of Davis Monthan Image with Boxes Showing Chosen Pixels.

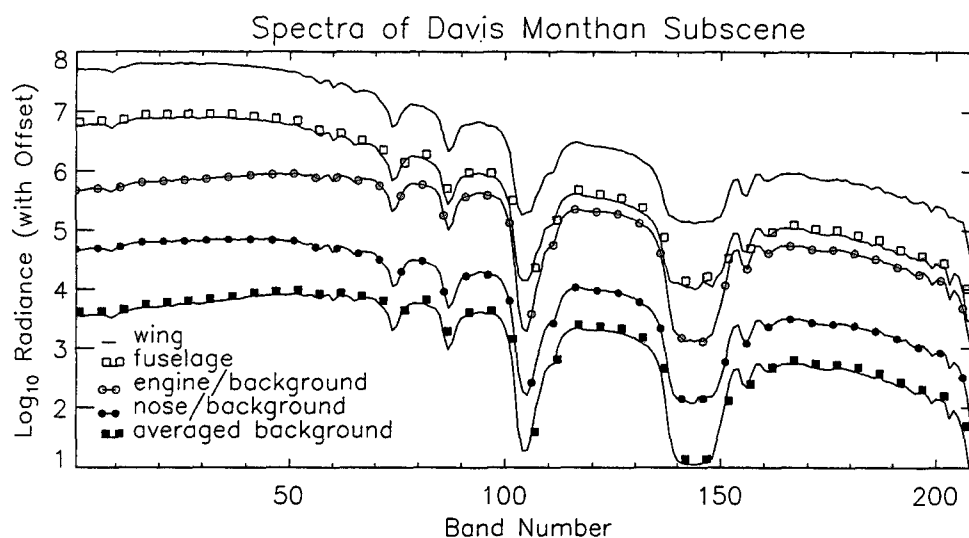


Figure 5.5: Spectra of Selected Davis Monthan Sub-scene Pixel Vectors.

vector corresponding to the averaged background has a positively-sloped shape in the first 70 bands that is distinct from the target spectra. The effect of mixing target and background endmembers is seen clearly in the engine and nose mixed pixels as spectral shapes that assume elements of both endmembers to various degrees. The mixing seen in Figure 5.5 can be compared with the uniform mixing of two endmembers seen in Figure 5.1.

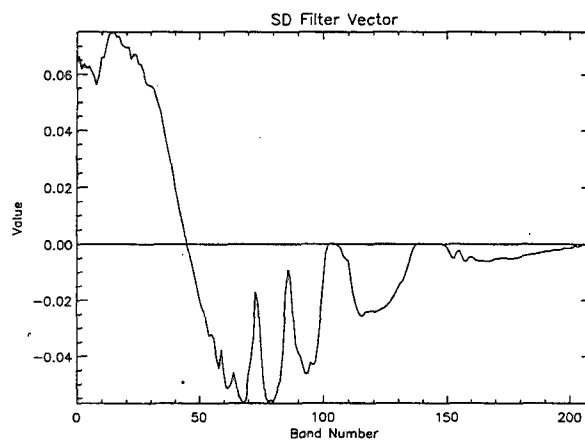


Figure 5.6: The SD Filter Vector.

In order to fully appreciate the operation of the SD filter, a more detailed inspection of the filter vector,  $w$ , is useful. Figure 5.6 plots the filter of Equation 5.36

that was derived using the aircraft wing as the desired pixel vector,  $\mathbf{d}$ , and an averaged background pixel vector as the single undesired endmember,  $\mathbf{u}$ . The scalar  $\beta$  in Equation 5.36 corresponds to the inverse of the Euclidean norm of the desired pixel vector. The significance of this filter vector is that it has been designed to maximize the response to the target spectrum while suppressing the effect of the undesired background spectrum.

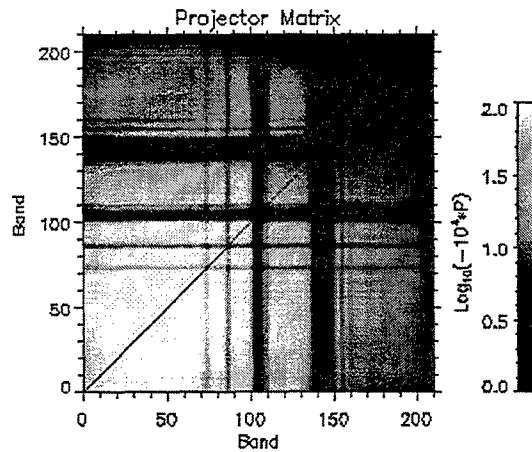


Figure 5.7: The Orthogonal Complement Projector.

A key part of the filter vector is the orthogonal complement projector. This  $l \times l$ -band projector has a structure that is very similar to that of the covariance matrix. This is illustrated in Figure 5.7 and its color version in Appendix A. The reason for this similarity is found in the explicit definition of the orthogonal complement projector found in the brackets of Equation 5.36. The inner product  $\mathbf{U}^T \mathbf{U}$  reduces to a scalar when only one undesired endmember is defined, as in this case. The outer product  $\mathbf{U} \mathbf{U}^T$  is then just the signal processing version of the correlation matrix of the undesired pixel vector. In the case of the Davis Monthan sub-scene, the value of the inner product is roughly a scalar on the order of  $10^9$ , and the outer product scaled by the inner product results in a diagonal matrix with small values, in the range of  $10^{-9}$  to  $10^{-2}$ , with the largest values occurring along the main diagonal. Like the covariance matrices typical of hyperspectral imagery, shown in Figures 3.10 and 3.12, the outer product matrix results in a peak in the bands corresponding to the solar part of the spectrum. When subtracted from the identity matrix, the result is the orthogonal complement projector, which has values very close to

one on the main diagonal and small negative values on the off-diagonal elements. In Figure 5.7, the elements of the projector have been scaled by a factor of  $-10^4$  and the logarithm of these numbers has been shown in the color plot. The effect of the bands in the solar portion of the spectrum is evident as a lighter region.

## 2. Orthogonal Subspace Projection (OSP)

A special case of the SD filter that occurs when the additive noise is assumed to be zero is the orthogonal subspace projection (OSP) technique. The OSP technique is a two-step process, which first applies the least squares orthogonal complement projector and then maximizes the SNR via a matched filter. In OSP, the  $l$ -band mixed pixel observation vector is described by the equation:

$$\mathbf{x} = \mathbf{M}\alpha + \mathbf{n} \quad (5.37)$$

where  $\mathbf{n}$  is the  $l$ -band noise vector assumed to be an independent identically distributed Gaussian process with zero mean and covariance matrix  $\sigma^2 \mathbf{I}$ ,  $\alpha$  is a  $p$ -dimensional vector in which the  $i^{\text{th}}$  element represents the fraction of the  $i^{\text{th}}$  signature present in the observed pixel, and  $\mathbf{M}$  is a  $l \times p$  matrix that represents the spectra of the  $p$  constituent endmembers of the scene:

$$\mathbf{M} = \begin{bmatrix} \uparrow & & \uparrow & \uparrow \\ \mathbf{u}_1 & \cdots & \mathbf{u}_{p-1} & \mathbf{d} \\ \downarrow & & \downarrow & \downarrow \end{bmatrix} \quad (5.38)$$

The  $\mathbf{u}_i$  represent the linearly independent endmembers corresponding to the undesired interfering spectra, and  $\mathbf{d}$  is the target spectrum (Harsanyi and Chang, 1994, p. 780). The result of Equation 5.37 is a combination of Equations 5.2 and 5.3, and serves as an important model for this and subsequent techniques. The observed pixel vector may be written equivalently in a manner that separates the desired and undesired signatures:

$$\mathbf{x} = \mathbf{d}\alpha_d + \mathbf{U}\alpha_u + \mathbf{n} \quad (5.39)$$

or more explicitly in terms of vectors and components:

$$\begin{bmatrix} x_1 \\ \vdots \\ x_l \end{bmatrix} = \begin{bmatrix} d_1 \\ \vdots \\ d_l \end{bmatrix} \alpha_d + \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{u}_1 & \cdots & \mathbf{u}_{p-1} \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \alpha_{u_1} \\ \vdots \\ \alpha_{u_{p-1}} \end{bmatrix} + \begin{bmatrix} n_1 \\ \vdots \\ n_l \end{bmatrix} \quad (5.40)$$

where  $\alpha_d$  is the scalar representing abundance of the desired spectrum in the observed pixel,  $\alpha_u$  is the vector of abundances of the undesired spectra in the observed pixel, and  $\mathbf{U}$  is the  $l \times p-1$  matrix comprised of the undesired spectra.



Rather than attempt to demonstrate OSP on an entire data set using several endmembers, a simplified model is assumed for didactic reasons. It is difficult to gain an intuitive understanding of the OSP technique unless it is simplified to deal with two-dimensional data. The mechanics of OSP are decomposed to their most fundamental level, that of projections of data onto subspaces. The emphasis in this discussion is to make the OSP technique a clear and logical progression of events. The simplifying assumptions are as follows. First, the Davis Monthan sub-scene of Figure 5.4 that includes only two major endmembers is chosen. These endmembers are the B-52 aircraft and background. Second, two bands that provide relatively good discrimination between aircraft and background radiance values are chosen as the two spectral dimensions. Third, the desired endmember is chosen from a pixel of the aircraft wing and the undesired endmember is formed from a averaged background value. The following paragraphs progress through the steps of OSP.

The first step of the OSP technique is to employ the idea of the orthogonal complement projector from the theory of least squares to eliminate the effects of the interfering signatures (Harsanyi, 1993, p. 27). The least squares optimal interference rejection operator is given by the  $l \times l$  matrix:

$$\mathbf{P} = (\mathbf{I} - \mathbf{U}\mathbf{U}^\#) \quad (5.41)$$

where  $\mathbf{U}^\#$  is the Moore-Penrose pseudo inverse of  $\mathbf{U}$ , defined as:

$$\mathbf{U}^\# = (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T \quad (5.42)$$

This is the projector that is shown in Figure 5.7. In OSP,  $\mathbf{P}$  is analogous to  $\mathbf{P}_A$  from the theory of least squares since it projects the observed pixel vectors into a subspace orthogonal to the undesired endmembers. This minimizes the energy associated with the undesired signatures by reducing the contribution of  $\mathbf{U}$  to zero when the operator is applied to the observed pixel vector as is seen in Equation 5.43:

$$\mathbf{P}\mathbf{x} = \mathbf{P}\mathbf{d}\alpha_d + \mathbf{P}\mathbf{n} \quad (5.43)$$

Equation 5.43 is important and requires careful explanation. As such, the same pixel locations chosen in Figure 5.4 are represented again in Figure 5.8 as vectors in a two-dimensional space which has been created by choosing bands 50 and 120. Figure 5.8 portrays all of the sub-scene pixels as points on the plane formed by plotting bands 50 and 120 against each other. The data from these two bands falls into two regions. The upper region, in which the majority of points lie, is located at high radiance values for both bands. These radiance values indicate that the points in this region correspond to background pixels, which is confirmed by noting this behavior for bands 50 and 120 in

Figure 5.6. The other region in Figure 5.8 is an area in which band 50 radiance values are low and band 120 radiance values are significantly lower than the background region. Referring again to Figure 5.6, it is apparent that the target spectra display this behavior.

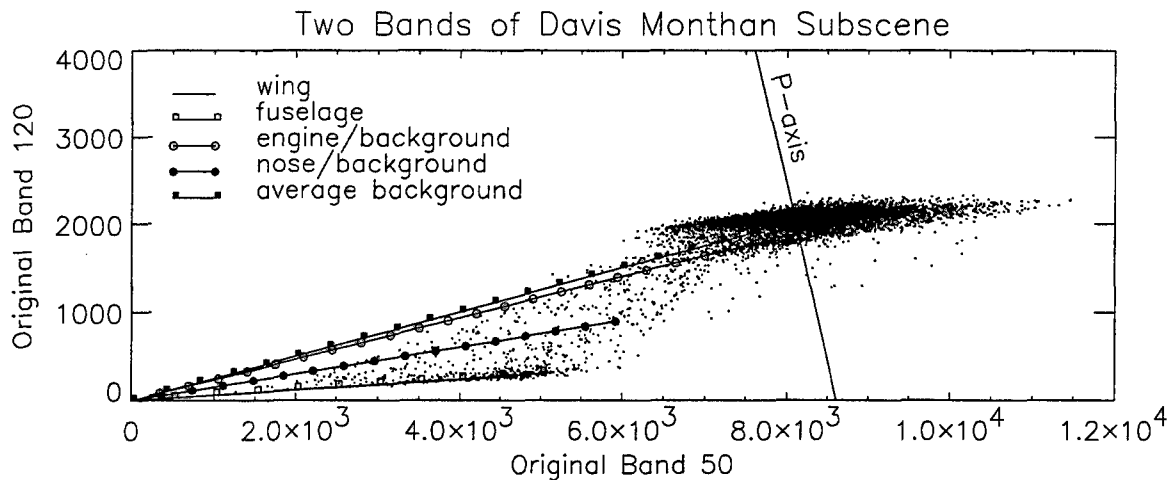


Figure 5.8: Scatter Plot of Davis Monthan Sub-scene Indicating Various Pixel Vector Locations and the Subspace Orthogonal to the Undesired Pixel Vector.

The appearance of these regions in two-dimensional space varies greatly according to the two bands chosen for the scatter plot. In the two-dimensional space of bands 50 and 120, the background region and target region show very strong consistencies between pixels of the same class. The important point is that the different spectral behavior of different types of pixel vectors is manifested in the scatter plot. The vectors emanating from the origin in Figure 5.8 confirm the fact that the upper region is composed of background pixels and the lower region of target pixels. The wing and fuselage pixel vectors are too close to each other to be discriminated as separate in Figure 5.8. A few observations about the five vectors plotted in Figure 5.8 are informative. First, points in spectral space may be represented as vectors. Second, the mixed pixel vectors lie between the extremes represented by the pure target and background vectors. Third, the angles between all of these vectors are relatively small. This confirms the fact that in hyperspectral imagery, the signals are not orthogonal. The vectors could be discriminated easily if they were orthogonal. This is precisely the goal of the first step of the OSP technique. The orthogonal complement projector,  $\mathbf{P}$ , is formed to project the data into a subspace which is orthogonal to the undesired endmembers. This operator is

the same as  $\mathbf{P}_A$  in Figure 5.2, except that  $\mathbf{P}$  is formed to project signals into a space which is orthogonal to the undesired endmembers. In our simplified scenario, the undesired endmember matrix,  $\mathbf{U}$ , has one endmember, which implies that the subspace that is orthogonal to  $\mathbf{U}$  is one-dimensional. The orthogonal subspace is shown in Figure 5.8 as the line labeled "P-axis," perpendicular to the undesired endmember pixel vector. It is on this line that all of the data is projected by the first step of the OSP technique. Its purpose is to cancel the effect of interfering undesired signals, which is readily apparent in the two-dimensional case.

The second step in the development of the OSP technique is to find the  $l \times 1$  filter vector that will maximize the SNR. The filter vector to be derived,  $\mathbf{w}$ , is applied to Equation 5.43:

$$\mathbf{w}^T \mathbf{P} \mathbf{x} = \mathbf{w}^T \mathbf{P} \mathbf{d} \alpha_d + \mathbf{w}^T \mathbf{P} \mathbf{n} \quad (5.44)$$

and results in a scalar quantity. The signal-to-noise energy ratio is formed as:

$$\text{SNR} = \lambda = \frac{\mathbf{w}^T \mathbf{P} \mathbf{d} \alpha_d^2 \mathbf{d}^T \mathbf{P}^T \mathbf{w}}{\mathbf{w}^T \mathbf{P} \mathbf{E} \{ \mathbf{n} \mathbf{n}^T \} \mathbf{P}^T \mathbf{w}} = \frac{\alpha_d^2}{\sigma^2} \frac{\mathbf{w}^T \mathbf{P} \mathbf{d} \mathbf{d}^T \mathbf{P}^T \mathbf{w}}{\mathbf{w}^T \mathbf{P} \mathbf{P}^T \mathbf{w}} \quad (5.45)$$

As in the derivation of the matched filter, maximization of this ratio leads to a generalized eigenvalue problem given in this particular case by:

$$\mathbf{P} \mathbf{d} \mathbf{d}^T \mathbf{P}^T \mathbf{w} = \left( \frac{\alpha_d^2}{\sigma^2} \right) \lambda \mathbf{P} \mathbf{P}^T \mathbf{w} \quad (5.46)$$

The eigenvector which maximizes the SNR is the optimum matched filter vector and is given by

$$\mathbf{w}^T = \mathbf{d}^T \quad (5.47)$$

Figure 5.9 shows the effect of projecting the data onto the orthogonal subspace. The lower plot in the figure is a scatter plot of data in a projected space which results from application of the  $\mathbf{P}$  operator to the original data. The  $\mathbf{P}$  operator is formed using a least squares criterion of optimality to be orthogonal to the undesired pixel vector. The result of projecting the data onto  $\mathbf{P}$  is that the pixels corresponding to the background have been reduced to very small magnitudes, as is evidenced by their location along the origin in the projected data scatter plot of Figure 5.9. The target pixel vectors and mixed pixel vectors retain their positions away from the background, but the wing and fuselage pixel vector locations have changed relative locations, with the fuselage pixel actually showing a greater separation from the origin in the projected data space whereas the wing pixel has a slightly larger angle away from the background in the original data space. The

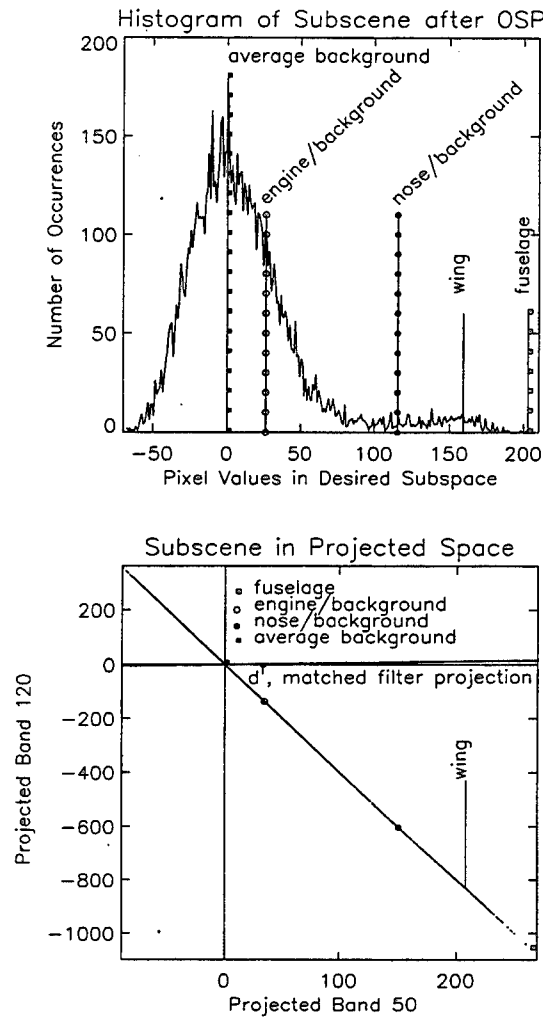


Figure 5.9: Davis Monthan Sub-scene Scatter Plot in Projected Space and Histogram of Final OSP Image.

distribution of the background pixel vectors with respect to the target pixel vectors in the original data space determines the appearance of the projected data. Recall that in Figure 5.8, the background region data scatter plot is roughly elliptical in shape. The major axis of this ellipse coincides with the averaged background pixel vector. The minor axis of this ellipse is relatively small, and when the data is projected onto the subspace orthogonal to the background by  $\mathbf{P}$ , the result is that the ellipse collapses onto a line whose length is the same as the magnitude of the minor axis. Also, the target data region appears as a separate region away from the background region along this one-dimensional projected subspace. The situation would be different if the distribution of data in the original two-dimensional space did not separate the target and background regions when

projected by  $\mathbf{P}$ . For instance, if the background region was more circular rather than elliptical, and the circle radius was wide enough, then when the data was projected onto the one-dimensional subspace orthogonal to the average undesired pixel vector (the center of the circle), the target would be lost in the clutter of the background. Recalling the results of eigendecomposition of spectral data, the eigenvalues of the covariance matrix of the background of the more elliptical case would show a large and a small value, whereas the more circular background distribution would show more equal eigenvalues. The eigenvalues and the associated eigenvectors represent the natural bases of the data. The  $\mathbf{P}$  operator could be formed from using an eigenvector associated with the major axis of the background ellipse to ensure that the projector  $\mathbf{P}$  is truly orthogonal to the undesired endmember. In the OSP technique, however, it is assumed that the undesired endmember is known and the use of eigenvectors of the background covariance matrix is discussed in the LPD technique in Chapter VI. We will retain the assumption that the average background pixel is the known background endmember.

The scatter plot of Figure 5.9 also shows the vector which represents the matched filter portion of OSP. This serves as a second projection for the data in which SNR is maximized. The result of projecting the projected data again is shown in the upper figure of Figure 5.9 as a histogram. This is a histogram of the scalar image associated with the output of OSP. The magnitude assigned to each pixel corresponds to the amount of the target endmember contained in each image pixel. The histogram shows that a large number of pixels have very small values, associated with the background. A smaller number of pixels have a significantly larger magnitude, which represent the target pixel vectors. The locations corresponding to the five pixel vectors of interest are also indicated in the histogram of Figure 5.9. The goal of the last step of OSP is to take the data which has already been projected into a subspace orthogonal to undesired signals and project it into another subspace which emphasizes the target signal.

The last step of OSP seeks to maximize the SNR of target to background. The matched filter is the optimal means of maximizing the SNR, as noted in the derivation of the matched filter. In order to demonstrate the concept of SNR maximization, Figure 5.10 shows a situation which projects onto a pixel vector other than the desired target pixel vector. The lower plot in Figure 5.10 is identical to that of Figure 5.9 except for the pixel vector that represents the final projector. The pixel vector used corresponds to that of the engine and background mixed pixel. In Figure 5.10, note that this matched filter vector has a different angle with respect to the horizontal than that of the matched filter

vector used Figure 5.9. As a result, the histogram of the final OSP image shows values that are smaller than the case where the desired target pixel vector is used. If SNR in this case is defined as the ratio of the magnitude of the wing pixel vector to the value of one standard deviation added to the average background pixel magnitude, then the SNR for

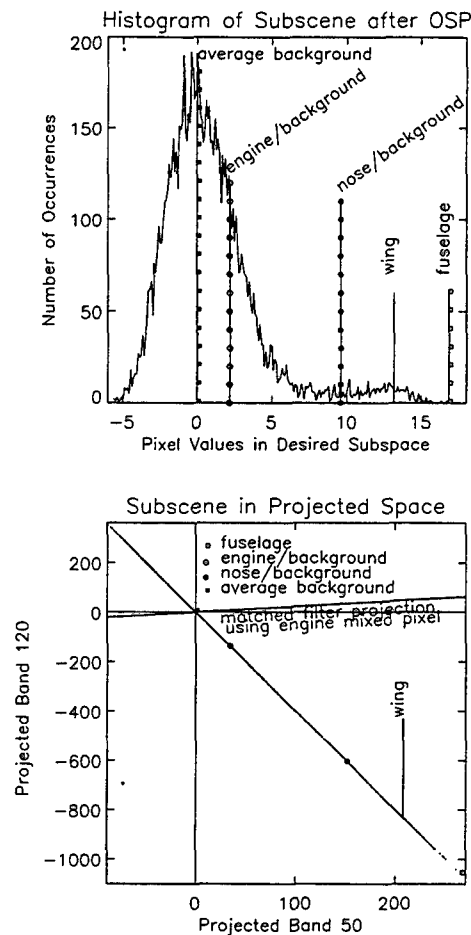


Figure 5.10: Davis Monthan Sub-scene Scatter Plot in Projected Space and Histogram of the Final OSP Image Using Mixed a Pixel Vector as the Matched Filter.

the case in Figure 5.9 is 6.46 and the SNR for Figure 5.10 is 5.82. The decrease in SNR is not large in this case. It is clear that the use of the target pixel vector as a matched filter improves the performance of the OSP in emphasizing target signal from background. Since target and background are being differentiated, the SNR could also be referred to as the signal-to-clutter ratio, or SCR.

Harsanyi and Chang (1994) combine the orthogonal projection operator  $\mathbf{P}$  and the matched filter vector into a single classification operator given by the  $1 \times l$  vector:

$$\mathbf{w}_{OSP}^T = \mathbf{d}^T \mathbf{P} \quad (5.48)$$

When the operator  $\mathbf{w}_{OSP}$  is applied to a hypercube, each  $l \times 1$  pixel vector is reduced to a scalar which is a measure of the presence of the signal of interest (Harsanyi, 1993, p. 29). The final results of OSP are scalar images whose pixel values correspond to the relative existence of a target material.

The use of two-dimensional vectors is helpful for visualizing the workings of OSP. When all 210 bands of the HYDICE image are used, the differentiation between various pixel vectors is improved. The discussion of OSP is continued from the two-

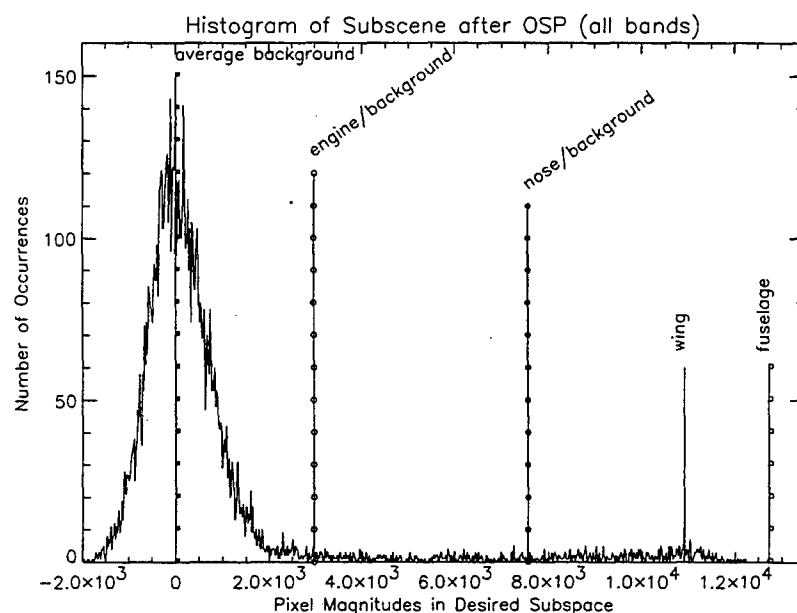


Figure 5.11: Histogram of the Davis Monthan Sub-scene OSP Output Image.

band case to all 210 bands with Figure 5.11 showing the resulting histogram of the output image. The SCR calculated using the same criteria as the two-band case yields a SCR of 14.82. This is an improvement over the two-band scenario. The reason for such better performance is that the additional bands assist in the discrimination that can be achieved and allow a more accurate projection operator to be formed. The image resulting from using OSP with all bands is shown in Figure 5.12. The image has been thresholded around the target region brightness values to better contrast detail within the targets. The color version of this figure is in Appendix A. It shows the different pixel values that

result after OSP. Note how the pixels corresponding to the original chosen fuselage and wing pixels show the highest values, while the mixed pixels around the aircraft show the

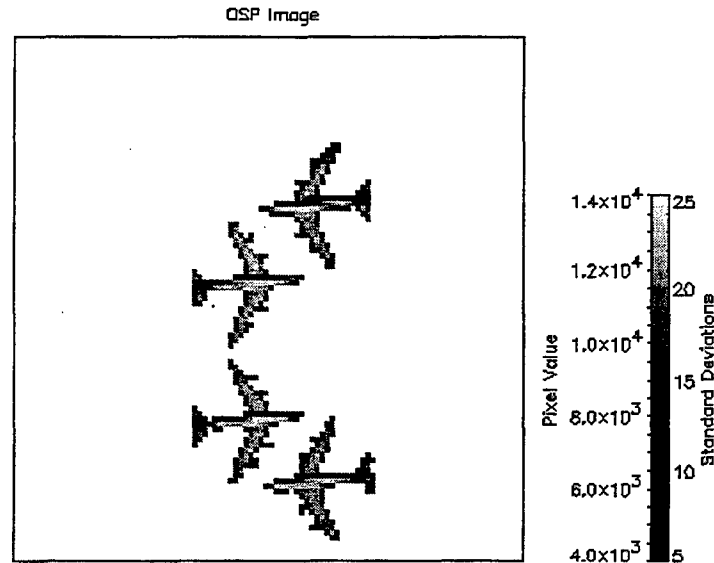


Figure 5.12: Davis Monthan Sub-scene OSP Output Image.

lowest. The color bar also indicates the brightness values in terms of standard deviations away from the center of the center of the Gaussian distribution representing the background in Figure 5.11.

The final step in the discussion of the OSP technique uses the entire Davis Monthan scene. The same B-52 wing pixel acts as the desired pixel vector. In the case of the entire scene, a total of seven undesired endmembers are chosen to form the  $U$  matrix. These include other types of aircraft and elements from the background. Figure 5.13 shows the histogram associated with the output image in this case. The fact that the road pixels are greater in magnitude than the target pixels suggests that the OSP technique did not have a complete enough characterization of the background endmembers to emphasize the target over the background. The SCR also reflects this fact. The center of the Gaussian curve corresponds to sand, which is intuitively pleasing, since we know that sand dominates this scene background. The output image is presented in Figure 5.14 and



its color version in Appendix A. The image has been thresholded to emphasize the B-52 pixels, but several undesired pixels corresponding to buildings and roads have higher

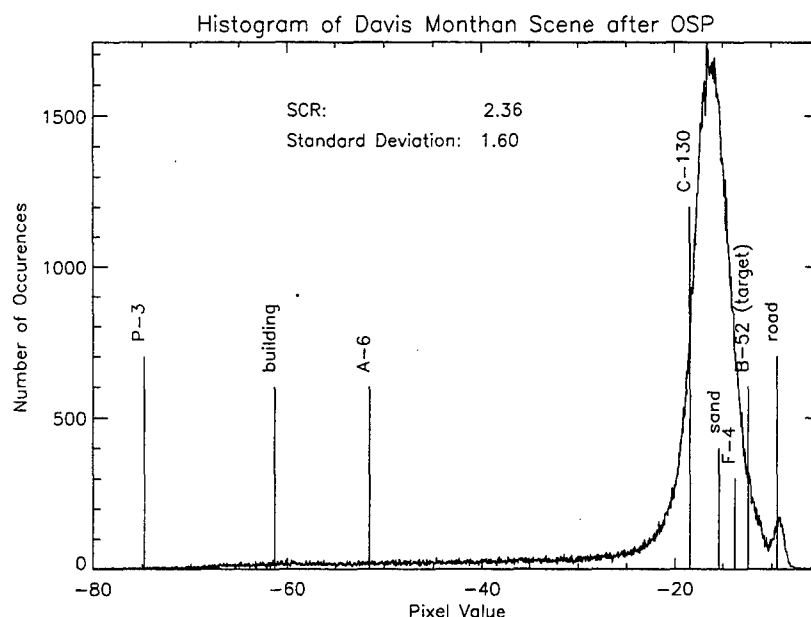


Figure 5.13: Histogram of Output OSP Image for Davis Monthan Scene.

magnitudes than the target pixels.

This series of simplified experiments with HYDICE data clarifies the operation of the OSP technique. The concepts were simplified by using two dimensions for illustration purposes, and choosing an image subset in which only two endmembers existed. The results show three important points. First, that OSP operates by minimizing predefined deterministic undesired endmembers through the projection of the data onto a subspace which is orthogonal to these undesired elements. Second, the second step of OSP is the matched filtering operation, in which the desired pixel vector provides the optimal response. Third, OSP discrimination is improved when more spectral bands are used to form the associated pixel vectors. The results of this experiment apply to any spectral imagery in which the mixed pixel model is assumed. Thus, the results of Harsanyi's (1993) work using AVIRIS data have been illustrated using HYDICE data, which has mixed pixels, but not on the large scale of AVIRIS data. With the OSP technique, Harsanyi (1993) slightly modifies the SD filter for application to hyperspectral imagery in which mixed pixels are assumed. He demonstrates the performance of the

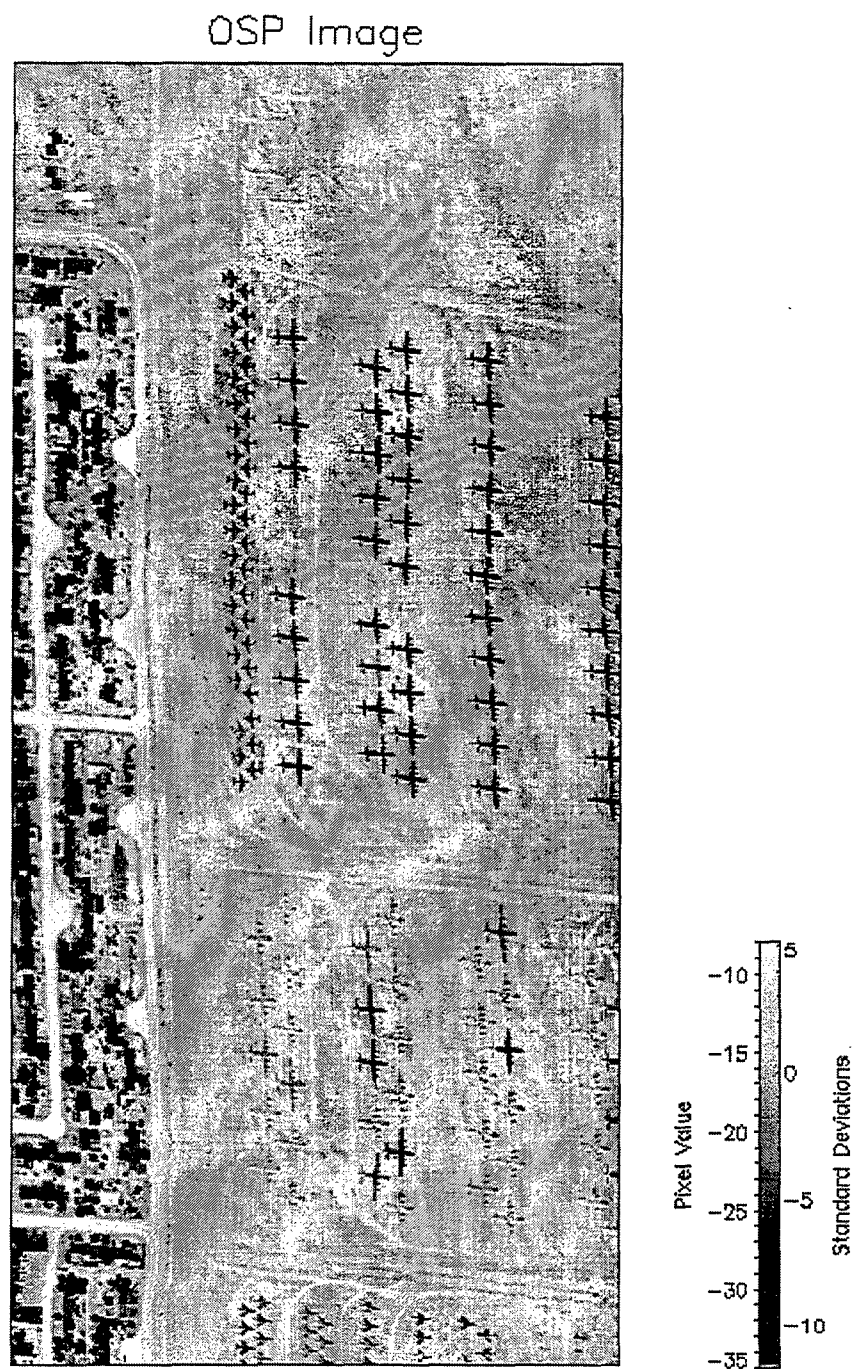


Figure 5.14: Davis Monthan OSP Output Image.

OSP technique by applying it first to a simulated hyperspectral data set generated by linearly mixing three different endmembers, one of which is identified as the target spectrum. The results show that the OSP technique identifies a target spectrum when mixed in greater than 5% abundance in a pixel. Next, Harsanyi uses an AVIRIS scene of the Lunar Crater Volcanic Field in Northern Nye County, Nevada, to test the OSP technique. Five mineral endmembers are extracted directly from the image, and OSP is used to generate five component images with each image showing the relative abundance of one endmember in the scene. The results indicate that OSP accurately detects the presence of various mineral deposits, and compares favorably with ground truth (Harsanyi, 1993, p. 37).

### 3. Least Squares Orthogonal Subspace Projection (LSOSP)

A variant of OSP, based on the *a posteriori* model of the observations, improves the performance of OSP by reducing the effects of noise. The least squares orthogonal subspace projection (LSOSP) technique developed by Tu, Chen, and Chang (1997), optimally estimates the desired signal abundance by minimizing a least squares error between the least squares estimate and the observations. The first step of the technique entails decomposing the observation space into a signature and a noise space and projecting the observations into the signature space. The second step of the process uses OSP to eliminate the undesired signatures in the signature space. Since the first step of LSOSP reduces the noise in the observations, it improves the ability of OSP to distinguish minority target spectra from the background. Whereas OSP uses the *a priori* model of observations, the LSOSP employs the *a posteriori* model. The following description of LSOSP details the first step of the technique.

The LSOSP uses least squared error estimation to convert the *a priori* model to an *a posteriori* model. As Settle (1996) observed, the optimal estimate of the abundance vector  $\alpha$  described in OSP is produced by a least squares estimate and is formulated as:

$$\hat{\alpha} = M^*x \quad (5.49)$$

where the notation used is identical to that developed in the OSP approach. The *a priori* model can be rewritten as:

$$x = M\alpha + n = M\hat{\alpha} + \hat{n} \quad (5.50)$$

where:

$$\hat{n} = x - M\hat{\alpha} = M(\alpha - \hat{\alpha}) + n \quad (5.51)$$

represents the estimation error term. The projector

$$\mathbf{P}_M = \mathbf{M}\mathbf{M}^* \quad (5.52)$$

projects the observation  $\mathbf{x}$  onto the signature space, and the projector

$$\mathbf{P}_A = (\mathbf{I} - \mathbf{P}_M) \quad (5.53)$$

projects the observations into the noise space  $\langle \mathbf{A} \rangle$ . Recall from Figure 5.2 that the entire signal signature space is designated as  $\langle \mathbf{M} \rangle$  and is comprised of the undesired subspace,  $\langle \mathbf{U} \rangle$  and the desired subspace,  $\langle \mathbf{d} \rangle$ . When the projector  $\mathbf{P}_M$  is applied to the observation vector,  $\mathbf{x}$ , the result is  $\hat{\mathbf{s}}$ , the optimal least squares estimate of the signal. Assuming that the noise  $\mathbf{n}$  is Gaussian,  $N(0, \sigma^2 \mathbf{I})$ , this implies that the observed vector will likewise be normally distributed,  $N(\mathbf{M}\alpha, \sigma^2 \mathbf{I})$ . Since the least squares estimates of signal and noise are linear transformations of the observation vector, they are also Gaussian random vectors  $\hat{\mathbf{s}}: N(\mathbf{s}, \sigma^2 \mathbf{P}_M)$ , and  $\hat{\mathbf{n}}: N(0, \sigma^2 \mathbf{P}_A)$ . Defining the SNR of a random vector as the mean squared over the variance, the following SNRs apply to the *a priori* (pr) and *a posteriori* (ps) models and their ratio:

$$\begin{aligned} \text{SNR}_{pr} &= \frac{E\{\mathbf{s}\}^T E\{\mathbf{s}\}}{\text{VAR}[\mathbf{s}]} = \frac{[\mathbf{M}\alpha]^T [\mathbf{M}\alpha]}{\text{tr}[\sigma^2 \mathbf{I}]} = \frac{[\mathbf{M}\alpha]^T [\mathbf{M}\alpha]}{\sigma^2 l} \\ \text{SNR}_{ps} &= \frac{E\{\hat{\mathbf{s}}\}^T E\{\hat{\mathbf{s}}\}}{\text{VAR}[\hat{\mathbf{s}}]} = \frac{[\mathbf{M}\alpha]^T [\mathbf{M}\alpha]}{\text{tr}[\sigma^2 \mathbf{P}_M]} = \frac{[\mathbf{M}\alpha]^T [\mathbf{M}\alpha]}{\sigma^2 p} \\ \frac{\text{SNR}_{ps}}{\text{SNR}_{pr}} &= \frac{l}{p} \end{aligned} \quad (5.54)$$

Note that  $p$  is the dimension of the signature space  $\langle \mathbf{M} \rangle$ , which is equivalent to the number of image endmembers, and  $l$  is the dimension of the observation space, which is expressed as the direct sum of subspaces:

$$\langle \mathbf{M} \rangle \oplus \langle \mathbf{A} \rangle \quad (5.55)$$

Since in most hyperspectral images the dimension of the observation space, or the number of bands, is greater than the number of image endmembers, the ratio of SNRs shows that the *a posteriori* model formulation and LSOSP result in a SNR improvement over the *a priori* model and OSP. The remainder of the LSOSP technique is identical to OSP in finding the eigenvector which maximizes the SNR and applying this as the filter vector.

Tu, Chen, and Chang (1997) use LSOSP to improve the ability of OSP to detect target spectra resident in mixed pixels in very small abundances. This improvement is a

result of the higher SNR afforded by LSOSP. They create simulated hyperspectral images using varying combinations of three endmember spectra collected by the 60-band field spectrometer system (FSS). The results show that target spectra in abundances smaller than the 5% can be detected (Tu, Chen, and Chang, 1997, p. 134). Figure 5.15 shows the detection improvement that is afforded by the LSOSP approach in a low SNR situation. LSOSP is able to detect the target pixels in substantially smaller abundances than OSP.

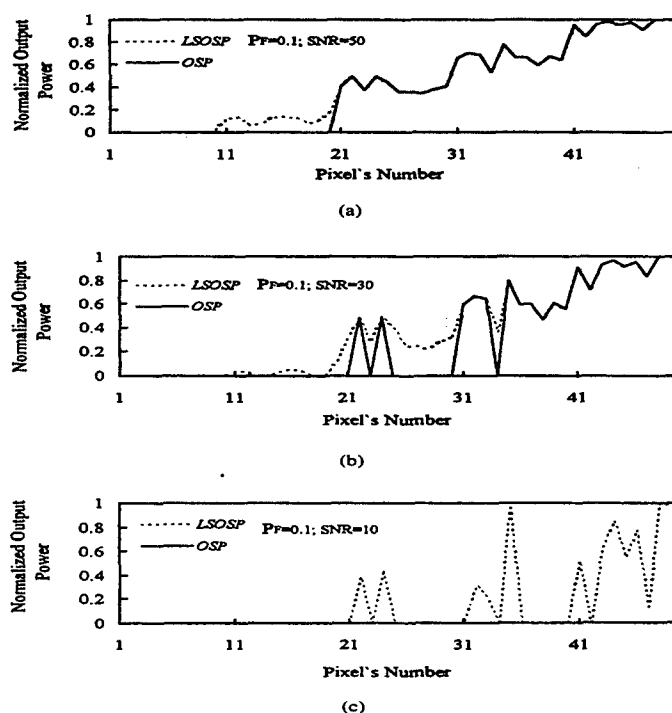


Figure 5.15: Comparison of LSOSP and OSP Simulation Results for Three SNRs.  
From Tu, Chen, and Chang, 1997, p. 137.

#### 4. The Filter Vector Algorithm (FVA)

The goal of this algorithm is to demix the scene, which is comprised of mixed pixels. The linear mixture model is assumed. The purpose of the demixing process is to find the relative abundance of each endmember in a pixel. A set of matched filters called filter vectors is constructed to demix the entire scene using linear vector spaces and orthogonal projections. The relative abundance of each endmember in a pixel is found by taking the inner product of the filter vector with the observed pixel vector. The filter vector must satisfy some requirements. First, it must be orthogonal to all endmembers in

the scene except for the one for which it is trained. Assuming that  $\mathbf{w}_n$  is the filter vector and  $\mathbf{s}_m$  is the endmember of interest, this orthogonality is stated as:

$$\mathbf{w}_n \mathbf{s}_m = \delta_{nm} \quad (5.56)$$

Second, the filter vector is zero mean and is uncorrelated with the noise random vector so that on average,

$$\mathbf{w}_n \mathbf{n} \cong 0 \quad (5.57)$$

where  $\mathbf{n}$  is the noise vector. Third, the inner product of the filter vector with itself is a minimum, so that the residual value of  $\mathbf{w}_n \mathbf{n}$  is minimized (Bowles, Palmadesso, Antoniadis, Baumbach, and Rickard, 1995, p. 150). These requirements ensure that when the filter vector is applied to the observed pixel vector,  $\mathbf{x}$ , it will yield a scalar corresponding to the abundance of the endmember of interest in that pixel. This is stated in the nomenclature of the SD filter as:

$$\mathbf{w}_n \bullet \mathbf{x} = \mathbf{w}_n \bullet \sum_m \alpha_m \mathbf{s}_m + \mathbf{w}_n \bullet \mathbf{n} \cong \alpha_n \quad (5.58)$$

where  $\alpha_n$  represents the abundance of the  $n^{\text{th}}$  endmember at a particular spatial location and  $\mathbf{w}_n$  is the filter vector designed to maximize the  $n^{\text{th}}$  endmember. The derivation of the filter vector is obtained by solving a constrained minimization problem using the calculus of variations (Bowles, Palmadesso, Antoniadis, Baumbach, and Rickard, 1995, p. 150). This is equivalent to the derivation of the matched filter described earlier.

The FVA is applied by Bowles, Palmadesso, Antoniadis, Baumbach, and Rickard (1995) to 1000-band hyperspectral data collected by the Naval Research Laboratory's Portable Hyperspectral Images for Low Light Spectroscopy (PHILLS) instrument flown on a P-3 aircraft. The scene consists of a beach in the Florida Keys, and FVA is used to detect endmembers corresponding to water and sand. The FVA is also demonstrated using a synthetic data set. The top plot in Figure 5.16 shows a synthetic spectrum created by mixing 100 known endmembers and adding noise. The lower plot shows the filter vector that was derived to detect this spectrum plotted together with the target spectrum. The peaks in the filter vector correspond to those in the target spectrum. Negative values of the filter vector suppress the contributions of other interfering spectra (Bowles, Palmadesso, Antoniadis, Baumbach, and Rickard, 1995, p. 151).

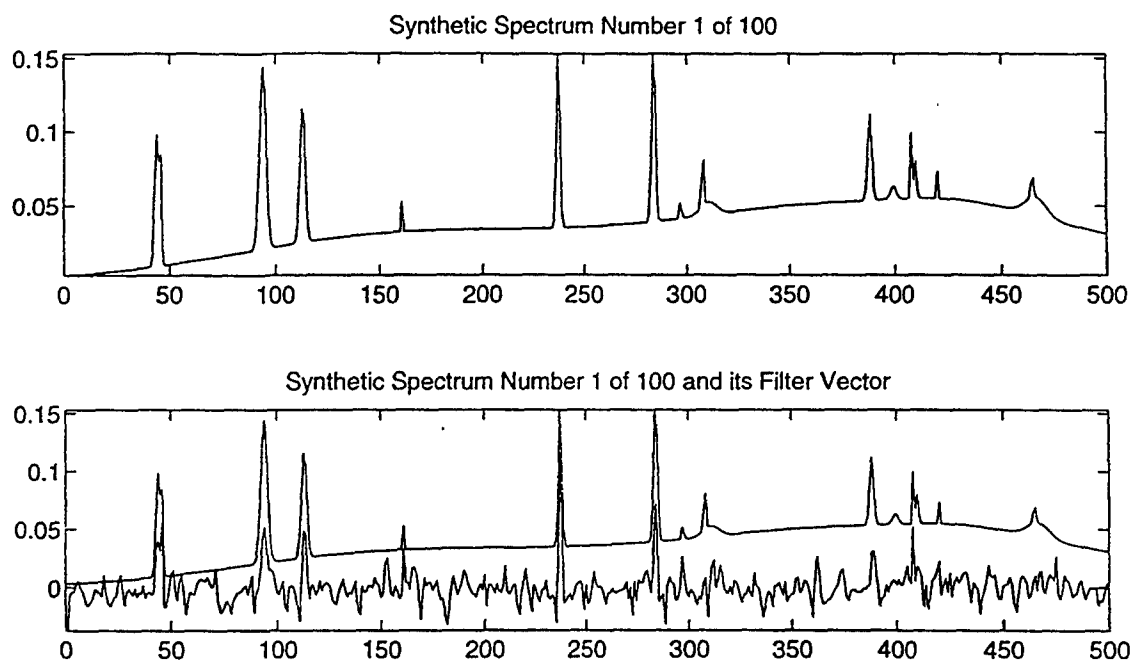


Figure 5.16: FVA Results for a Simulated Spectrum.  
From Bowles, Palmadesso, Antoniades, Baumbach, and Rickard, 1995, p. 154 .





## **VI. THE UNKNOWN BACKGROUND FAMILY OF TECHNIQUES**

### **A. DESCRIPTION**

The matched filter scenario is made more challenging when the only *a priori* knowledge is the target spectrum. Using an analogy from array signal processing, the endmembers which constitute the unknown background are the interfering signals which must be suppressed. Without prior knowledge of the constituent background scene endmembers, however, the interfering signals must be estimated from the statistics of the scene. This scenario returns to the ideas underlying the principal components transform, which relies on the covariance matrix of the observations. In the low probability detection (LPD) and the constrained energy minimization (CEM) techniques both introduced by Harsanyi (1993), the problem of estimating endmembers from the background is addressed as a first step towards reducing the amount of *a priori* knowledge one needs to perform target detection in a mixed pixel hyperspectral image. In a different approach, the adaptive multidimensional matched filter described by Stocker, Reed, and Yu (1990) solves the problem of a known target in an unknown background using ideas from statistical signal processing.

### **B. BACKGROUND DEVELOPMENT**

The development of techniques that work with the target detection problem in unknown backgrounds is based on ideas from the statistical signal processing and array processing communities. The background issues that support the framework of these techniques are grouped according to the technique that they support. The intrinsic dimensionality and information theoretic criteria apply to both the LPD and CEM techniques, while the idea of beamforming is the basis for the CEM technique.

#### **1. Determining the Intrinsic Dimensionality Based on Information Theoretic Criteria**

One of the fundamental parameters when inferring information from the covariance matrix of data is the intrinsic dimensionality of the data. In the eigendecomposition of a covariance matrix, the intrinsic dimensionality is the

number of eigenvectors which are needed to adequately represent the original signal. Recalling the results of PCA in characterizing the eigenvalues of the covariance matrix, this number can be deduced by noting the number of eigenvalues with significantly larger magnitudes than the others. In principle, simply counting the number of significant eigenvalues should be an easy task because the smallest eigenvalues all are of magnitude roughly equal to the variance of the additive noise. In practice, when the covariance matrix is estimated from a finite number of points, the resulting eigenvalues are different with probability one, and it is difficult to estimate the number of significant eigenvalues (Wax and Kailath, 1985, p. 388). In Wax and Kailath (1985) the determination of the intrinsic dimensionality is pursued by modeling a vector of observations as a superposition of signals embedded in additive noise. Information theoretic criteria are then used to objectively determine the number of signals from the information inherent in the covariance matrix. We will follow Wax and Kailath's (1985) problem development, but we shall assume that the signals in question are real vice complex, and that the vectors are not functions of time. Also, the notation used here is consistent with that of the linear mixing model. Their development starts with the model of the observed signal:

$$\mathbf{x} = \mathbf{M}\alpha + \mathbf{n} \quad (6.1)$$

where  $\mathbf{x}$  is the  $l \times 1$  vector of observations,  $\alpha$  is a  $p \times 1$  vector of scalars associated with each of the  $p$  signals,  $\mathbf{n}$  is the  $l \times 1$  noise vector which is assumed to be Gaussian with parameters  $N(0, \sigma^2 \mathbf{I})$ , and  $\mathbf{M}$  is the  $l \times p$  matrix of signals:

$$\mathbf{M} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{m}_1 & \cdots & \mathbf{m}_p \\ \downarrow & & \downarrow \end{bmatrix} \quad (6.2)$$

Each of the  $p$  signals in  $\mathbf{M}$  is a  $l \times 1$  column vector. Since the noise is assumed to be zero mean and independent of the signals, the covariance matrix of the observations is given by:

$$\Sigma_{\mathbf{x}} = \Sigma_s + \sigma^2 \mathbf{I} \quad (6.3)$$

$\Sigma_s$  represents the covariance matrix of the signals, which is formed using the expression:

$$\Sigma_s = \mathbf{M} \mathbf{E}\{\alpha\alpha^T\} \mathbf{M}^T \quad (6.4)$$

(Wax and Kailath, 1985, p. 388). All  $p$  columns of  $\mathbf{M}$  are linearly independent, implying that the rank of  $\Sigma_s$  will be  $p$ . The number of signals,  $p$ , is the quantity that must be estimated, and is renamed as the unknown variable  $k$ .

Information theoretic criteria address the general problem posed in the following statement, "Given a set of  $N$  observations,  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ , and a family of models, select the model that best fits the data." The models are analogous to a parameterized family of the probability density functions (pdfs) represented by  $f(\mathbf{X}|\theta)$ , where  $\theta$  is the vector of parameters. Since the model can be used to encode the observed data, Wax and Kailath (1985) use Rissanen's (1978) approach of selecting the model which represents the minimum code length to find the model which best fits the data. This minimum description length (MDL) criterion is stated as:

$$\text{MDL} = -\log f(\mathbf{X}|\hat{\theta}) + \frac{1}{2}k \log N \quad (6.5)$$

where  $\hat{\theta}$  is the maximum likelihood estimate of the parameter vector  $\theta$ , and  $k$  is the number of free adjusted parameters in  $\theta$ . The first term represents the log-likelihood of the maximum likelihood estimator of the parameters of the model, and the second term is a bias correction term (Wax and Kailath, 1985, p. 388).

Implementing the above definition begins with the spectral representation theorem, which is used to decompose the correlation matrix of the data into eigenvalues and eigenvectors. The parameter vector,  $\theta$ , is formed by concatenating all of the eigenvalues, the noise variance, and all eigenvectors into a vector of length  $(k+1+pk)$ . The joint pdf of the observations is represented as the product of the independent Gaussian random vectors, and the log-likelihood function of the parameter vector is formed using the sample covariance matrix. The maximum likelihood estimates of the components of the parameter vector are substituted into this log-likelihood expression adjusted with the bias correction term using the number of free adjusted parameters. The end result of the MDL criterion implementation is stated as:

$$\text{MDL}(k) = -\log \left( \frac{\prod_{i=k+1}^l \lambda_i^{1/(l-k)}}{\frac{1}{l-k} \sum_{i=k+1}^l \lambda_i} \right)^{(l-k)N} + \frac{1}{2}k[2l - (k+1)] \log N \quad (6.6)$$

where the  $\lambda_i$  are the maximum likelihood estimates of the eigenvalues of the sample correlation matrix (Harsanyi, Farrand, Hejl, and Chang, 1994, p. 270). The first term is the ratio of the geometric mean to the arithmetic mean of the  $l-k$

smallest eigenvalues of the sample correlation matrix. The number of signals is determined as the value of  $k \in \{0, 1, \dots, p-1\}$  for which the MDL is minimized.

## 2. Beamforming

In signal processing, a requirement that commonly occurs is to design a filter that minimizes the average output power of the filter while constraining the filter response to signals of a specific frequency to remain constant. Haykin (1996) presents the spatial analogue of this constrained optimization problem in the process called beamforming by the array processing community. The current development will follow Haykin's (1996) presentation of the topic. Figure 6.1 illustrates the basic model used in depicting the beamforming concept. It

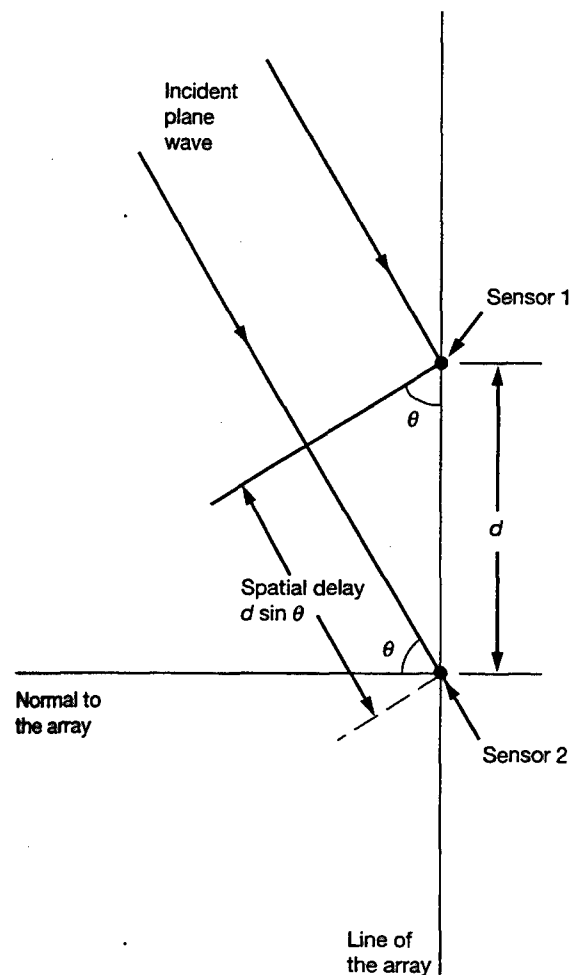


Figure 6.1: Uniform Linear Array. From Haykin, 1996, p. 63.

represents a linear array of uniformly spaced antenna elements. The plane wave impinges on the array along the direction  $\theta$  with respect to the vertical measured at the reference antenna element. The direction of arrival for our applications is translated into the electrical angle  $\phi$ , where

$$\phi = \frac{2\pi d}{\lambda} \sin \theta \quad (6.7)$$

The variable  $d$  represents the distance between antenna elements and  $\lambda$  is the wavelength of the impinging wave. The goal is to find the optimum set of weights represented by the  $l$ -dimensional vector  $\mathbf{w}_o$ , with elements  $w_k$ , that minimize the mean square value of the beamformer output subject to the linear constraint:

$$\sum_{k=0}^{l-1} w_k^* e^{-jk\phi} = g \quad (6.8)$$

where  $\phi$  is a prescribed value of the electrical angle and  $g$  is a complex valued gain (Haykin, 1996, p. 222). This linear constraint is important, as it preserves the signal of interest. The asterisk indicates complex conjugation. The output of the beamformer is given by:

$$y[n] = x[n] \sum_{k=0}^{l-1} w_k^* e^{-jk\phi} \quad (6.9)$$

where the  $x[n]$  is the electrical signal picked up by an antenna element. The notation of the index  $n$  refers to fact that the signals in this scenario are discrete sequences.

The method of Lagrange multipliers is employed to solve the constrained optimization problem using a real valued cost function,  $J$ , which combines the output power minimization and linear constraint into a single expression. The gradient of the cost function with respect to the weight vector elements is formed and set equal to zero for the minimization. The resulting condition for the optimality of the beamformer is recast in terms of vectors as:

$$\Phi_x \mathbf{w}_o = \frac{-\lambda^*}{2} \mathbf{s}(\phi) \quad (6.10)$$

where  $\Phi_x$  is the  $l \times l$  correlation matrix of the observed signal,  $\mathbf{w}_o$  is the optimum weight vector of the constrained beamformer,  $\lambda^*$  is the complex Lagrange multiplier, and  $\mathbf{s}(\phi)$  is the  $l \times 1$  steering vector represented as:

$$\mathbf{s}(\phi) = \begin{bmatrix} 1 \\ e^{-j\phi} \\ \vdots \\ e^{-j(l-1)\phi} \end{bmatrix} \quad (6.11)$$

(Haykin, 1996, p. 224). Using the linear constraint of Equation 6.8 to solve for the complex Lagrange multiplier and substituting it in the optimality condition of Equation 6.10, we arrive at an expression for the optimum weight vector:

$$\mathbf{w}_o = \frac{g^* \Phi_x^{-1} \mathbf{s}(\phi)}{\mathbf{s}^{*T}(\phi) \Phi_x^{-1} \mathbf{s}(\phi)} \quad (6.12)$$

(Haykin, 1996, p. 225). The significance of the beamformer is that signals incident on the array along directions other than  $\phi$  are attenuated by virtue of the minimization of the output power subject to the linear constraint.

In the special case of  $g = 1$ , the beamformer is constrained to produce a distortionless response along the electric angle  $\phi$ , and is termed a minimum variance distortionless response beamformer (MVDR). The optimum weight vector is then:

$$\mathbf{w}_o = \frac{\Phi_x^{-1} \mathbf{s}(\phi)}{\mathbf{s}^{*T}(\phi) \Phi_x^{-1} \mathbf{s}(\phi)} \quad (6.13)$$

and the minimum mean square value is expressed in terms of the cost function:

$$J_{\min} = \mathbf{w}_o^{*T} \Phi_x \mathbf{w}_o = \frac{1}{\mathbf{s}^{*T}(\phi) \Phi_x^{-1} \mathbf{s}(\phi)} \quad (6.14)$$

(Haykin, 1996, p. 226). Equation 6.14 also represents the average power or variance of the beamformer output, and is an estimate of the variance of the signal impinging on the array along the direction corresponding to  $\phi$ . The optimum beamformer passes the target signal with unit response while simultaneously minimizing the total output variance.

In a broader context, the beamforming method is a non-parametric method of developing a spatial filter that can form a reception beam in a prescribed direction. It can also be viewed as a non-parametric method of spectral estimation with specific application to the direction-of-arrival problem (Stoica and Moses, 1996, p. 311). Non-parametric refers to the fact that the method makes no assumptions about the statistics of the covariance matrix of the data.

## C. OPERATION

The three techniques presented in this section are similar in that none assume *a priori* knowledge of the endmembers comprising the background. They differ in their origins and conceptual framework. The LPD technique is based on the application of the image covariance matrix eigenvectors to the task of suppressing the scene background endmembers. The CEM technique applies the beamforming methodology to minimize the total response of the target detection operator while constraining it to respond only to target pixels. The approaches lead to different techniques, which apply to different situations. The LPD and CEM techniques were introduced by Harsanyi (1993), and this study follows the original development of the techniques found in Harsanyi's (1993) work. The adaptive multidimensional matched filter improves the idea of a matched filter by forming a likelihood test that depends on the second order statistics of the unknown background to determine the presence or absence of a target.

### 1. Low Probability of Detection (LPD)

The basic premise of the LPD technique is that the contribution of unknown undesired signatures can be estimated directly from the data and eliminated using an orthogonal complement projector operator (Harsanyi and Farrand, 1995, p. 1566). The important assumption made by this technique is that the signature of interest occurs in the image with a low probability. This implies that the target is only present in a small number of image pixels, so that the abundance of the target material,  $\alpha_d$  in Equation 5.34, can be set to zero for almost all of the image pixels (Harsanyi, Farrand, and Chang, 1994, p. 5). Harsanyi (1993) further assumes that the target spectrum occurs at subpixel levels, and the signatures of the unknown naturally occurring background endmembers dominate the observed pixel vectors.

The technique begins with modification of the linear mixture model of the OSP technique to account for the very small number of target-bearing pixels. This modification is made by setting  $\alpha_d = 0$  in Equation 5.34. The resulting expression for the observed  $l \times 1$  pixel vector  $\mathbf{x}$  is now:

$$\mathbf{x} = \mathbf{U}\alpha_u + \mathbf{n} \quad (6.15)$$

where the matrix  $\mathbf{U}$  is the  $l \times p-1$  matrix whose columns are the spectra of the  $p-1$  unknown background endmembers,  $\alpha_u$  is the  $p-1 \times 1$  vector representing the relative abundances of the undesired endmembers, and  $\mathbf{n}$  is the  $l \times 1$  noise vector,

which is not necessarily white. The intent of the LPD technique is to 1) model the unknown background spectra,  $\mathbf{U}$ , as a linear combination of eigenvectors of the sample covariance matrix of the image, and 2) derive an orthogonal subspace projector that minimizes the effects of  $\mathbf{U}$  while maximizing the target-to-background ratio (Harsanyi, 1993, p. 59). This ratio is also termed the signal-to-clutter ratio (SCR). The image background may be modeled using the eigenvectors of the sample covariance matrix because the assumption is that the target-bearing pixels are not statistically significant. This key assumption of the LPD technique implies that the statistics of the scene are essentially the statistics of the background.

Since most of the observation vectors,  $\mathbf{x}_i$ , are linear combinations of the  $p-1$  independent undesired endmembers,  $\{\mathbf{u}_1, \dots, \mathbf{u}_{p-1}\}$ , the first  $p-1$  covariance matrix eigenvectors account for most of the image variance (Harsanyi, 1993, p. 60). This result follows from the optimal representation property of the DKLT. Though the eigenvectors are not the endmembers, they can be used to account for the majority of spectral variation in the image (Harsanyi and Farrand, 1995, p. 1571). The  $l \times p-1$  matrix representing the most significant covariance matrix eigenvectors is given by:

$$\tilde{\mathbf{E}} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{e}_1 & \cdots & \mathbf{e}_{p-1} \\ \downarrow & & \downarrow \end{bmatrix} \quad (6.16)$$

The number of significant eigenvectors,  $p-1$ , is unknown, and must be estimated using the MDL information theoretic criterion. Following this estimation, the LPD technique closely resembles the OSP technique. The orthogonal subspace projection operator is formed using the above eigendecomposition-based estimate of the background endmembers,  $\tilde{\mathbf{E}}$ , as:

$$\tilde{\mathbf{P}} = (\mathbf{I} - \tilde{\mathbf{E}}\tilde{\mathbf{E}}^\#) \quad (6.17)$$

with the  $\#$  denoting the pseudoinverse operation. As with the OSP technique  $\mathbf{P}$  operator, the LPD projection operator,  $\tilde{\mathbf{P}}$ , serves to cancel the effect of the interfering undesired signatures. Finally, the target detection operator that is applied to each observation pixel vector is formed as:

$$\mathbf{w}_{LPD}^T = \mathbf{d}^T \tilde{\mathbf{P}} \quad (6.18)$$

As in the OSP technique, the  $l \times 1$   $\mathbf{d}$  vector represents the desired endmember. The result of applying  $\mathbf{w}_{LPD}^T$  to the image hypercube is a scalar image in which the pixel magnitudes represent the relative presence of the target material.



The steps involved in estimating the intrinsic dimensionality are discussed and illustrated below in a sequential manner using the same 100 x 100 pixel Davis Monthan sub-scene as in the OSP technique. As a first step to improve the determination of the intrinsic dimensionality of the data, the additive noise term,  $\mathbf{n}$ , in Equation 6.15 must be whitened. This provides a better separation of signal and noise subspaces in the MDL calculation (Harsanyi, Hejl, Farrand, and Chang, 1994, p. 270). Since the noise and endmember pixel vectors of Equation 6.15 are assumed to be independent, the covariance matrix of the observations may be written as:

$$\Sigma_{\mathbf{x}} = \Sigma_{\mathbf{s}} + \Sigma_{\mathbf{n}} \quad (6.19)$$

where  $\Sigma_{\mathbf{s}}$  represents the covariance matrix of the known signal endmembers. Note that signal in this context includes both desired and undesired endmembers, but under the low probability assumption it contains only undesired endmembers.  $\Sigma_{\mathbf{n}}$  represents the covariance matrix of the additive noise. This matrix must be estimated from the data using the same method of shift differences as used in the MNF technique. The noise covariance matrix may be decomposed by a unitary transformation into:

$$\Lambda_{\mathbf{n}} = \mathbf{E}_{\mathbf{n}} \Sigma_{\mathbf{n}} \mathbf{E}_{\mathbf{n}}^T \quad (6.20)$$

where  $\Lambda_{\mathbf{n}}$  is the diagonal matrix of eigenvalues and  $\mathbf{E}_{\mathbf{n}}$  is the matrix of eigenvectors of  $\Sigma_{\mathbf{n}}$ . This eigendecomposition of  $\Sigma_{\mathbf{n}}$  is used to form the whitening transform:

$$\mathbf{W} = \mathbf{D}_{\mathbf{n}}^{-1/2} \mathbf{E}_{\mathbf{n}} \quad (6.21)$$

which is applied to the sample covariance matrix to produce the following whitened signal covariance matrix:

$$\Sigma_{\mathbf{wx}} = \mathbf{W} \Sigma_{\mathbf{x}} \mathbf{W}^T = \mathbf{W} \Sigma_{\mathbf{s}} \mathbf{W}^T + \mathbf{I} \quad (6.22)$$

Figure 6.2 shows the results of applying the whitening transformation to the noise covariance matrix. The noise covariance matrix is very similar in appearance to typical HYDICE radiance covariance matrices, with a region of high variance occurring in the solar portion of the spectrum. The whitening transform converts the noise covariance matrix into the identity matrix shown in the right side of Figure 6.2. Figure 6.3 shows the effect of the whitening transform on the observed data covariance matrix. The resulting noise-whitened signal matrix bears no resemblance to typical HYDICE covariance matrices. The noise-whitened signal matrix has relatively small variances uniformly distributed over all bands. The absorption bands are notable features in addition to the diagonal elements which are evident in Figure 6.3. The important point regarding this

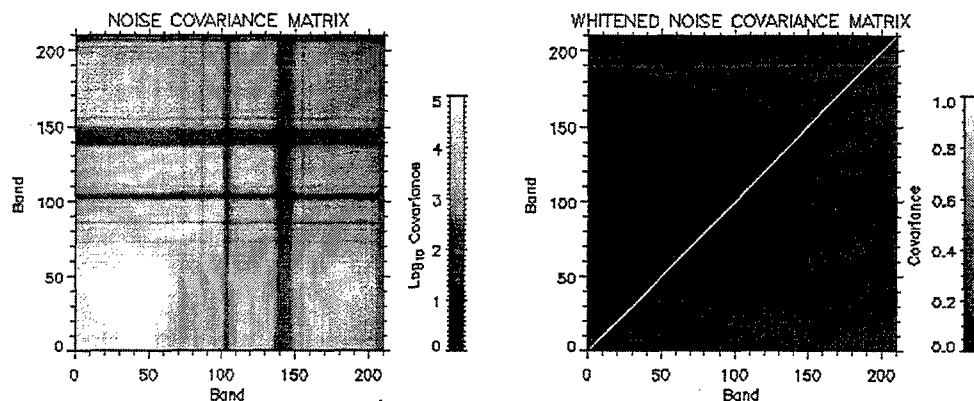


Figure 6.2: Whitening of the Davis Monthan Sub-scene Noise Covariance Matrix.

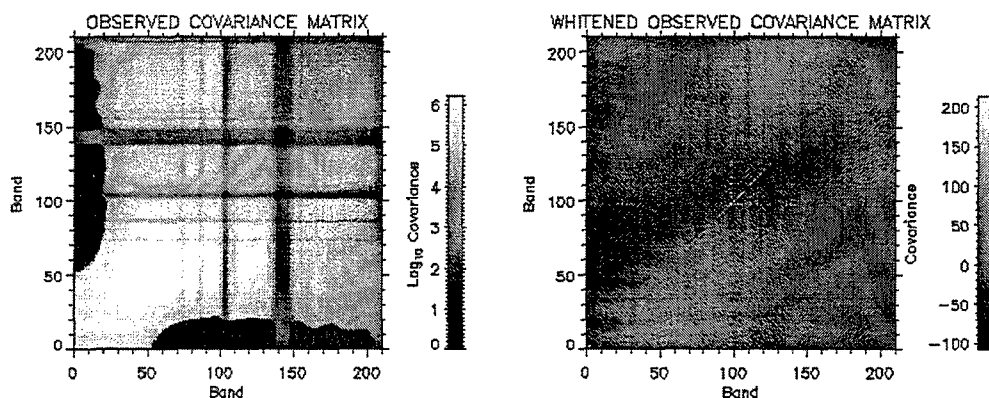


Figure 6.3: Whitening of the Observed Davis Monthan Sub-scene Data Covariance Matrix.

matrix is that the effects of noise have been mitigated, so that the noise-whitened covariance matrix can be used as a basis for determining the intrinsic dimensionality of the data. Both Figures 6.2 and 6.3 are presented in Appendix A as color plots, in which much of the finer structure of the matrices is accentuated.

The MDL criterion is applied to the noise-whitened signal covariance matrix, and the value of  $k$  which minimizes the MDL expression of Equation 6.6 is a strongly consistent estimate of the intrinsic dimensionality of the image background (Harsanyi, Hejl, Farrand, and Chang, 1994, p. 270). In this study, the

application of the MDL criterion to HYDICE data did not yield a minimum value that could be construed as the intrinsic dimensionality. Rather, Figure 6.4 resulted from MDL( $k$ ) plotted against various values of  $k$ . This figure shows a monotonically decreasing behavior in the MDL criterion. The MDL criterion was

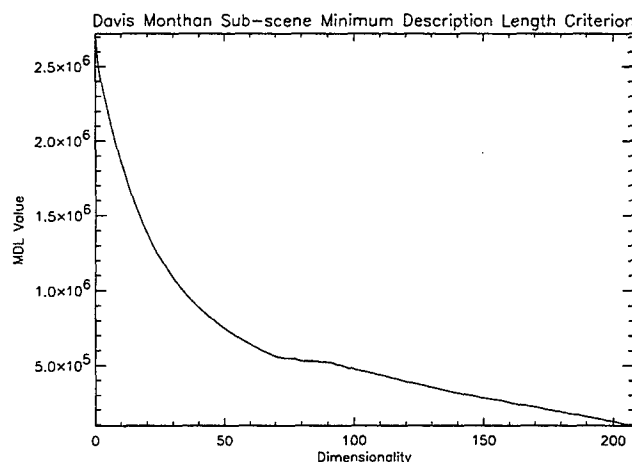


Figure 6.4: MDL Criterion Applied to Davis Monthan Subscene.

applied to Landsat data, which has more established noise characteristics than HYDICE data, in hopes that a minimum MDL value would be found. The Landsat data produced the same decreasing behavior as the HYDICE data. These results do not correspond with the predicted behavior described by Wax and Kailath (1985) nor to the applied results of Harsanyi, Hejl, Farrand, and Chang (1994) with AVIRIS data. For the purposes of continuing with the LPD technique, an intrinsic dimensionality of between one and ten is assumed. This is based on the general behavior of the eigenvalues of HYDICE imagery, which fall abruptly in magnitude after roughly the first ten eigenvalues.

The following paragraphs describe the operation of the LPD technique under various conditions. First, the Davis Monthan sub-scene is used as a simple test case for LPD performance. This sub-scene only contains one type of aircraft and a uniform background. The fact that the aircraft in the sub-scene do not occur in sub-pixel quantities challenges the basic low probability assumption of the LPD technique. Second, the effect of assuming various intrinsic dimensionalities to form the orthogonal subspace projection operator,  $\tilde{\mathbf{P}}$ , are investigated using the entire Davis Monthan scene.

The Davis Monthan sub-scene shown in Figure 5.4 is simple because it only contains two endmembers. Although the aircraft endmember appears in a

substantial number of pixels, the LPD technique is applied as a test of performance when the low target probability assumption is not satisfied. The  $\tilde{\mathbf{P}}$  operator is formed using the first eigenvector of the scene covariance matrix. The desired pixel vector,  $\mathbf{d}$ , is chosen from a wing of an aircraft, as in the OSP technique. The application of  $\mathbf{w}_{LPD}$  to the sub-scene results in a scalar image whose histogram is shown in Figure 6.5. This figure appears very much like that

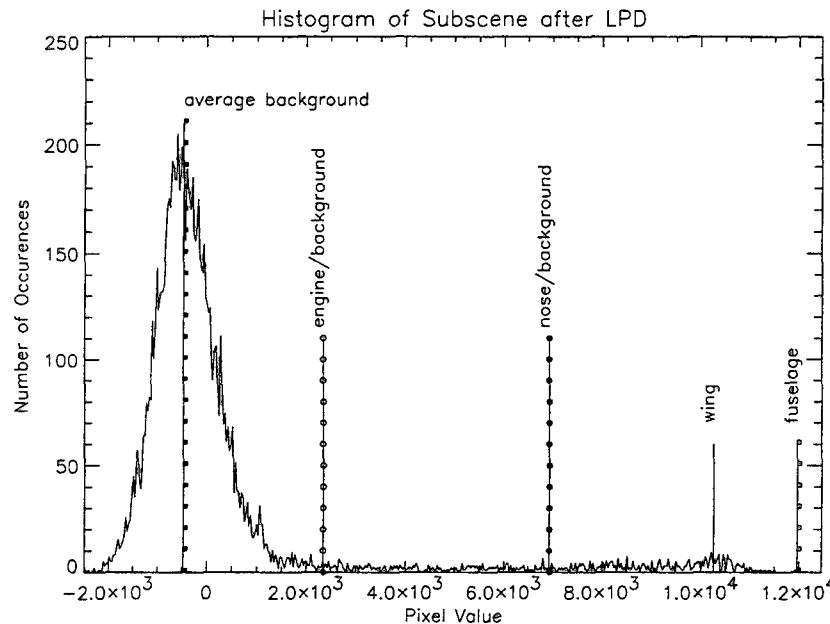


Figure 6.5: Histogram of Davis Monthan Sub-scene LPD Output.

of Figure 5.11, the result of the OSP technique. The relative position of the selected pixel vectors is the roughly the same. The LPD technique has successfully separated the target from the background and achieved a higher SCR than the OSP technique. SCR is defined here, as in the OSP technique, as the ratio of the target pixel magnitude to one standard deviation away from the center of the background distribution. The brightness values of the scalar image produced by the LPD technique are thresholded to produce Figure 6.6. This image also looks very similar to the corresponding OSP output of Figure 5.12. Here, though, the LPD algorithm achieves a better degree of target segregation as

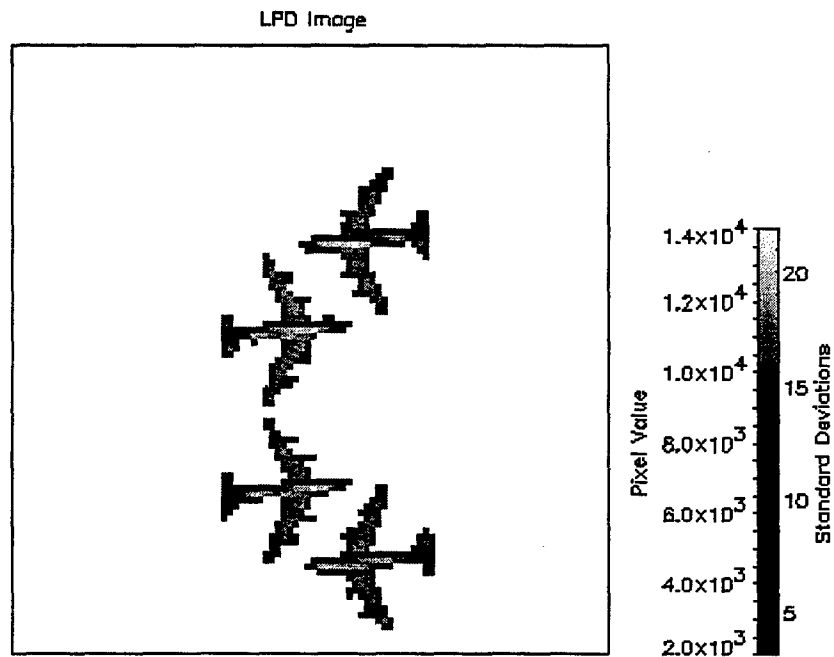


Figure 6.6: Davis Monthan Sub-scene LPD Output.

depicted in the bar scale which shows pixel magnitudes in absolute terms and in terms of number of standard deviations away from the center of the background distribution. A color version of this figure may be found in Appendix A. The simple Davis Monthan sub-scene has shown that the LPD technique actually performs slightly better than the OSP technique even when the low target pixel probability assumption is not met.

The above example was conducted using the first eigenvector to construct the  $\tilde{\mathbf{P}}$  operator. The next series of images and histograms investigates the performance of the LPD technique when one eigenvector and then the first five eigenvectors are used to form  $\tilde{\mathbf{P}}$ . The entire Davis Monthan scene serves as the data for the LPD technique. The pixel vector chosen from the B-52 wing is used as the desired pixel vector. Figure 6.7 shows the histogram of the output image when the first eigenvector is used to form the projection operator. The B-52 pixel displays the highest scalar value and is well-differentiated from the background

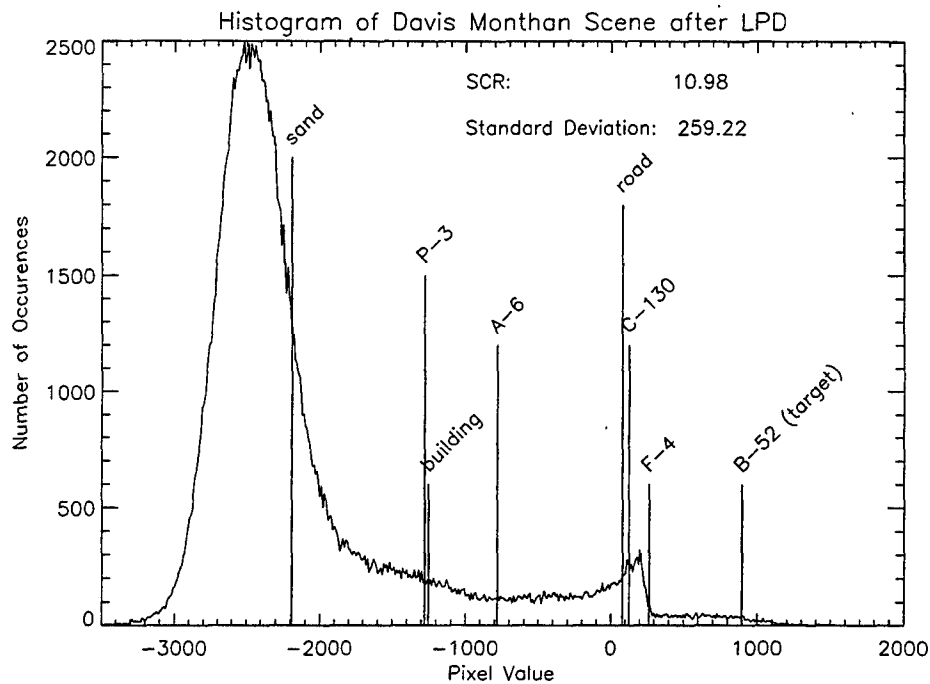


Figure 6.7: Histogram of LPD Scalar Image Using the First Eigenvector.

clutter and other type aircraft. The figure reaffirms the intuitive notion that the majority of the scene is composed of a sandy background, which is represented by the major peak in the histogram. The image that this histogram represents is presented in Figure 6.8. The image brightness values have been thresholded in a manner that accentuates the B-52 target aircraft. These appear prominently in Figure 6.8. The color version of Figure 6.8 in Appendix A shows that other aircraft actually have similar values, namely the C-130 aircraft. This might be attributable to the fact that the paint used on these aircraft is similar to that used on the B-52s. The application of thresholding was enabled by the good SCR that the LPD technique achieved as evident in Figure 6.7. It is interesting to note that the B-52 aircraft pixels are actually a minority element of the scene, though they still do not meet the strict requirements of Harsanyi (1993) to occur on a subpixel level. The case where  $\bar{\mathbf{P}}$  is formed using the first five eigenvectors of the covariance matrix would seemingly improve the performance of the LPD algorithm in theory, since more eigenvectors provide a better representation of the scene variability. Five eigenvectors also correspond more closely with the expected intrinsic dimensionality for a hyperspectral scene than does one eigenvector. The results of using the first five eigenvectors are seen in Figure 6.9,

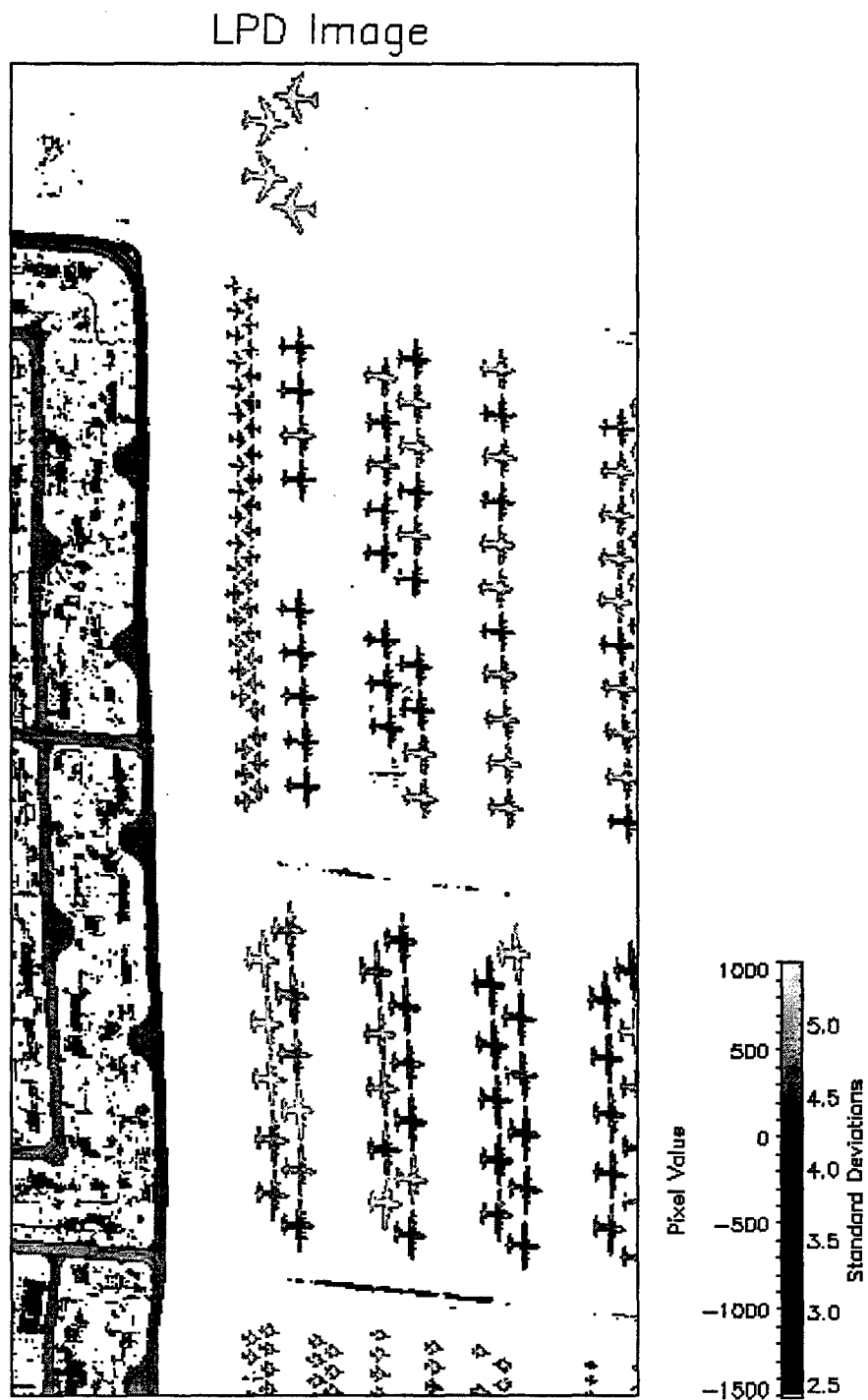


Figure 6.8: Davis Monthan LPD Output Image Using the First Eigenvector.

which shows the histogram for the output LPD scalar image. The large Gaussian distribution has subsumed the pixels that were outliers in Figure 6.7. The SCR has decreased by an order of magnitude from the case using the first eigenvector. The scalar image associated with Figure 6.9 reflects the difficulty in differentiating targets. The closeness of pixel brightness values in magnitude is apparent in the image as objects such as roads and background that have large magnitudes, as do the targets. Thresholding the values does little to accentuate the pixels of interest. The situation does not improve as more eigenvectors are included in the formation of the projection operator. The trend as more eigenvectors are included is that the histogram of the scalar image continues to appear Gaussian and target pixels cannot be distinguished.

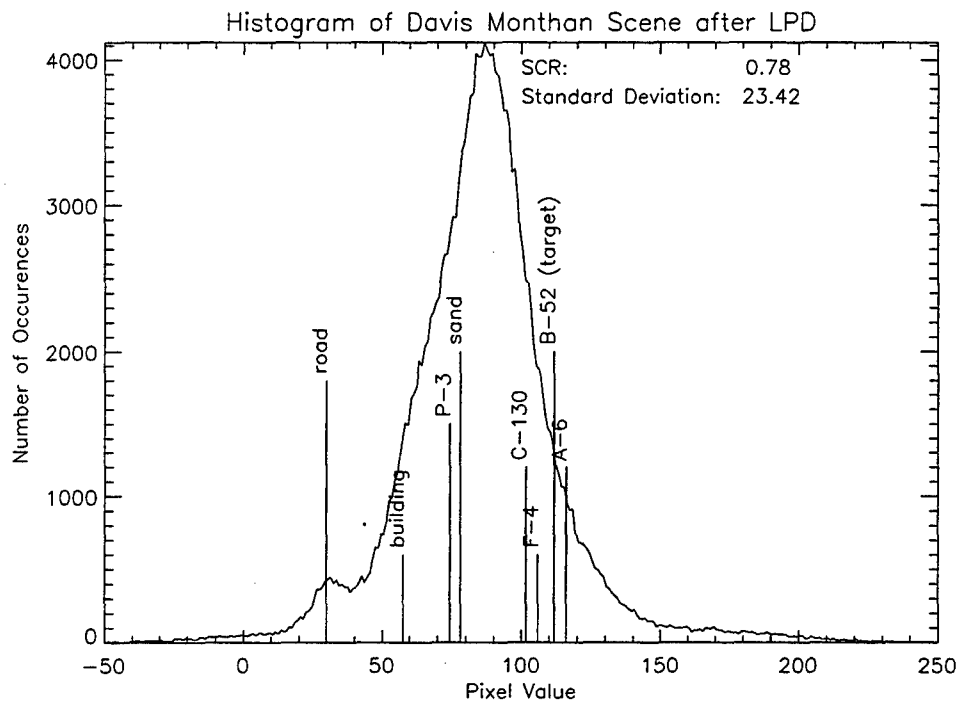


Figure 6.9: Histogram of LPD Scalar Image Using the First Five Eigenvectors.

As a means of trying to understand the dynamics of the LPD technique, it is informative to examine the projector matrices formed in the previous two cases. Figure 6.10 presents the projector matrix formed by using the first eigenvector on the left and the first projector matrix formed with the first five eigenvectors on the right. The  $\tilde{\mathbf{P}}$  matrix created with one eigenvector has a more uniform appearance than that created with five eigenvectors. This observation is put in perspective by recalling from Figure 4.15 that the first eigenvector of the Davis Monthan scene has all positive values and a relatively smooth shape, whereas the subsequent



eigenvectors are more oscillatory in nature and possess negative values. As in Figure 5.7, the elements of  $\tilde{\mathbf{P}}$  have been scaled by a factor of  $-10^4$  and the logarithm of these numbers has been shown in Figure 6.10. The ones on the diagonal appear as black. The color version of this figure may be found in Appendix A. The  $\tilde{\mathbf{P}}$  matrix corresponding to the first eigenvector looks very similar to that of the OSP operator. The implication is that the first eigenvector captures the essence of the background in a manner that is very similar to when

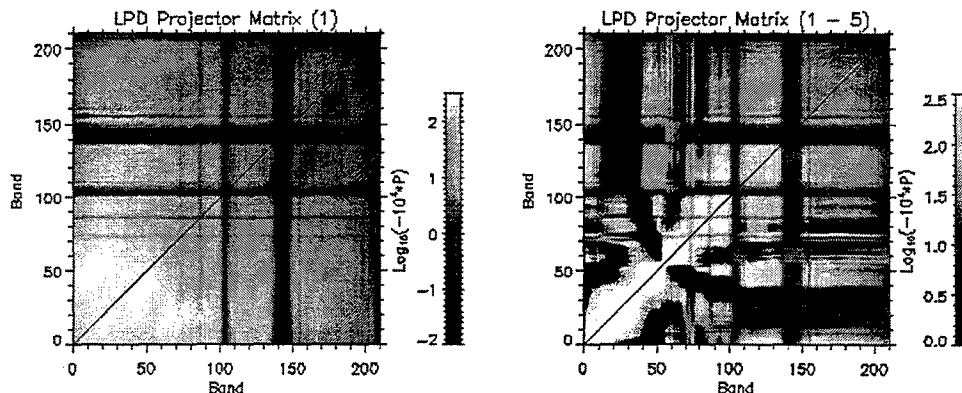


Figure 6.10: LPD Projector Matrices Created with the First Eigenvector and the First Five Eigenvectors.

the background is known. The first eigenvector acts as a scene average, and this scene is dominated by the background, so the use of one eigenvector has produced optimal results.

The inability of the LPD technique to successfully differentiate aircraft in the Davis Monthan Scene is a result of attempting to apply it in a circumstance for which it was not designed. The aircraft in the Davis Monthan are not minority elements. Forming the LPD projector based on only the first eigenvector produces better results because the first eigenvector captures the scene average elements. Inclusion of more eigenvectors does not characterize the background any better, instead it has converse effect and further characterizes the target. A greater understanding of this phenomenon is afforded if we look at the classification operator,  $\mathbf{w}_{LPD}$ , that is created in the case of including one through ten eigenvectors in the formation of the projection operator. The operators are 210-element vectors. In order to demonstrate a more successful application of the LPD technique, the Aberdeen HYDICE reflectance scene is used. This scene has

targets which are minority elements of the scene. Figure 6.11 shows the progression in  $w_{LPD}$  operators and image appearance as the number of eigenvectors included in the operator changes from one to ten starting from the top. Note that the entire scene as shown in figure B.1 was used in processing. The left plots show the LPD operator behavior and the right plots show the changing appearance of a small subset of the image which contains the target

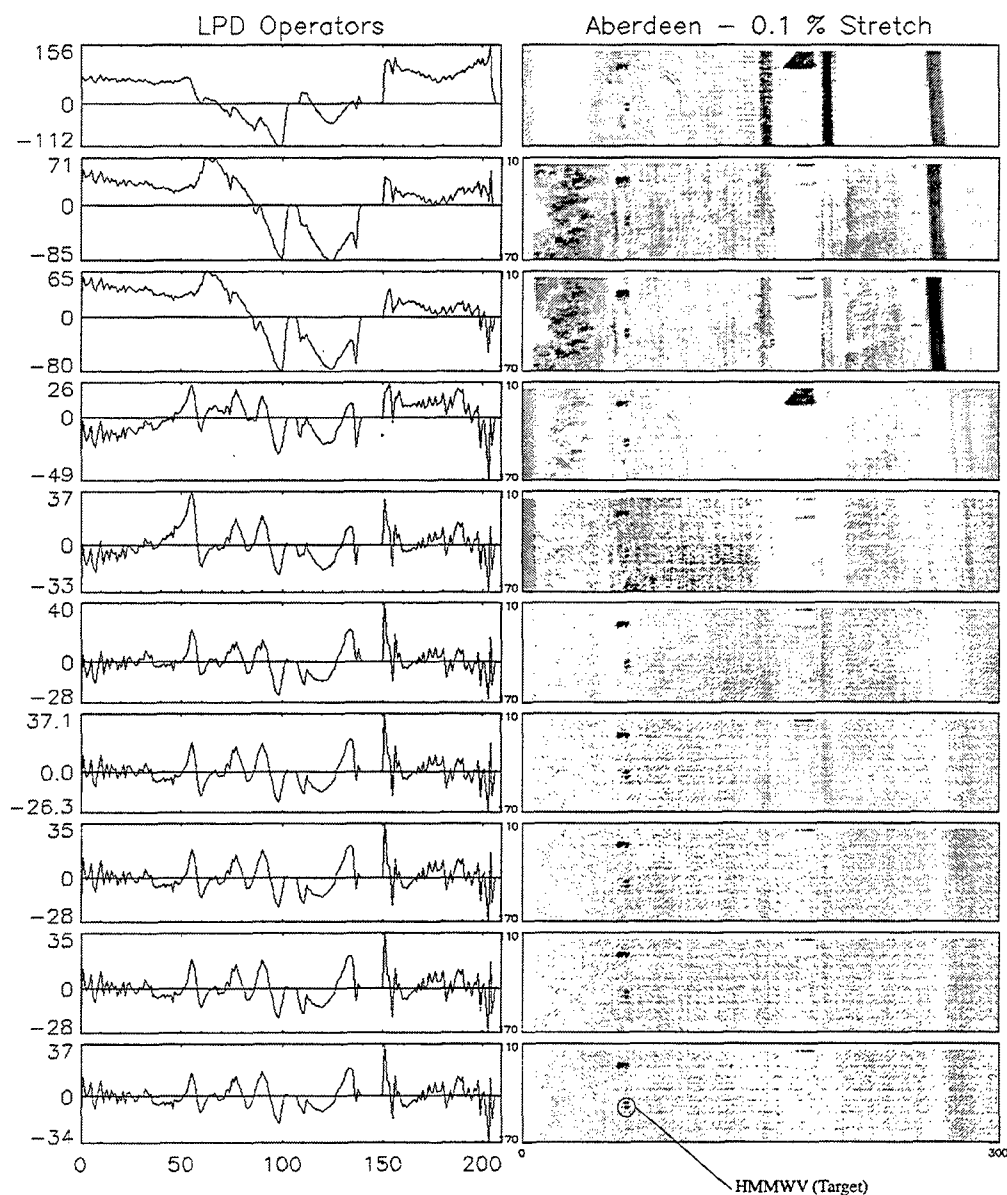
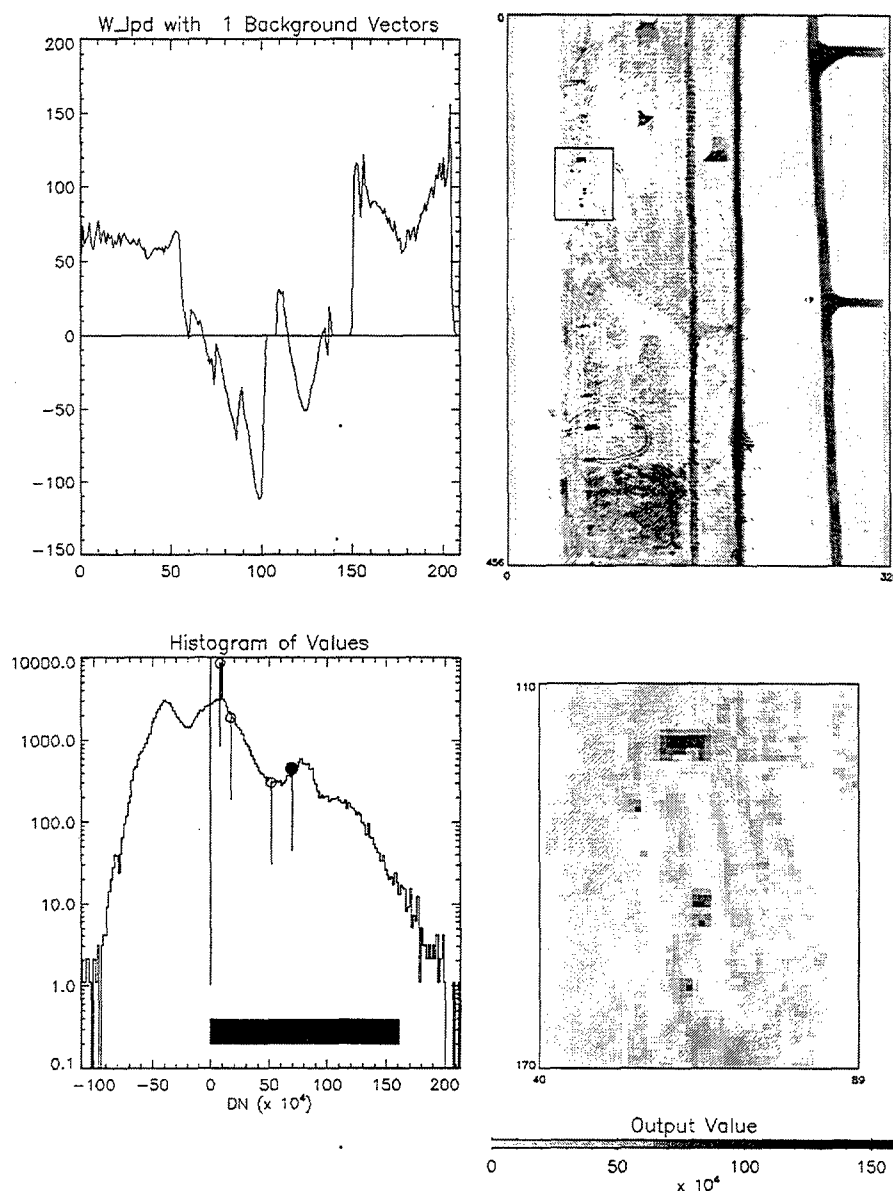


Figure 6.11: LPD Operators and Output Images for the First Ten Eigenvectors.

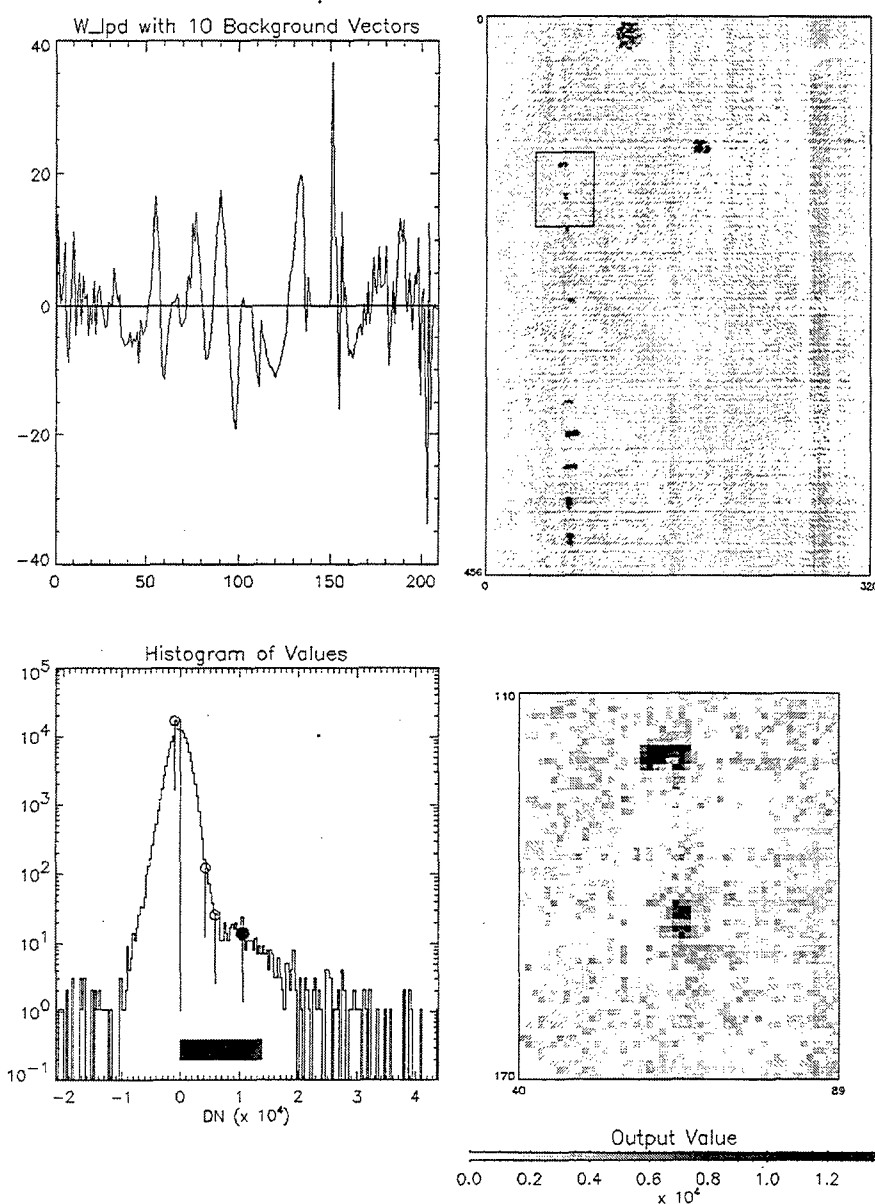
pixel. The target in this case is a HMMWV painted with a woodland camouflage scheme, and is the smaller group of dark pixels in the left third of the image. The obvious trend in image target differentiation is that it improves as more eigenvectors are included. The operator associated with the first eigenvector case shows the most dynamic range. The result is apparent in the output images produced with these operators. The histograms of the LPD output images in Figure 6.11 have undergone a 0.1% "histogram stretch" for better contrast. To



29-May-1997 10:11:47.00

Figure 6.12: Aberdeen Scene LPD Output Image Using One Eigenvector.

emphasize the difference between the LPD technique application to the Davis Monthan and the Aberdeen scenes, Figure 6.12 shows the entire Aberdeen output scene along with the associated histogram and LPD operator. The dark circle on the histogram indicates the target pixel, and the dark bar on the histogram indicates the dynamic range chosen for thresholding of the output image. Figure 6.13 shows the results of LPD on the same scene using the first ten eigenvectors. The background has been suppressed to a far greater extent than that of the first



29-May-1997 10:12:10.00

Figure 6.13: Aberdeen Scene LPD Output Image Using Ten Eigenvectors.

eigenvector output image. The opposite situation of Figures 6.7 and 6.8 for the Davis Monthan scene is seen in Figures 6.12 and 6.13 for the Aberdeen scene. In the case of the Aberdeen scene, more than one eigenvector is needed to adequately characterize the background and construct an effective LPD operator. The primary cause of this observed effect is that the Aberdeen scene contains the target as a minority element, whereas the Davis Monthan scene contains many similar aircraft target pixel vectors.

The above results demonstrate the behavior of the LPD technique using HYDICE imagery. In a broader context, the LPD technique is useful for detecting objects occurring as isolated minority elements of a scene. Harsanyi (1993) points out several specific applications of LPD such as: the detection of isolated vegetation, subpixel rock outcroppings, poorly exposed geological features, toxic materials in landfills, low level pollutants in waterways, scarce mineral deposits, and man-made objects in naturally occurring backgrounds. All of these scenarios involve spectral signatures of materials of interest occurring with a low probability throughout the scene (Harsanyi, 1993, p. 56). Harsanyi (1993) validates the LPD technique using simulations of a sparse vegetation detection problem. He simulates AVIRIS data by mixing the known spectra of welded tuff and basalt with varying amounts of the black brush spectrum in a few pixels of a simulated scene. The result of the LPD technique is detection of the pixels containing black brush down to 10% abundance (Harsanyi, 1993, p. 69). Harsanyi (1993) further demonstrates the utility of the LPD technique by detecting a large canvas tarp in an AVIRIS scene of Mono Lake, California. More insight into the applicability of the LPD technique comes as a result of understanding some of its limitations. Harsanyi and Farrand (1995) observe that a drawback of LPD is that the target material cannot be in sufficient abundance to be included as a spectral component of any of the primary eigenvectors of the covariance matrix of the observations. They further note that materials of low abundance which are not the target spectrum can be erroneously detected as target material because the LPD has been designed to suppress only the majority endmembers as identified by the significant eigenvectors. With these limitations in mind, the LPD technique can be applied in a nearly automatic manner, with only knowledge of the target spectrum required.

## 2. Constrained Energy Minimization (CEM)

The CEM technique makes the same assumptions as the LPD technique except that of requiring a low probability of target material abundance in the image. The CEM technique relaxes the constraint of LPD that target pixels appear with low probability, and also applies to pure and mixed pixels. The CEM technique, based on the MVDR concept, seeks to create a linear operator which minimizes the total energy in the scene while constraining the response of the signature of interest to a constant level (Harsanyi, 1993, p. 82).

The goal of the CEM technique is to find the weight vector  $\mathbf{w}_{CEM}$  which, when applied to each observation pixel vector, produces a scalar,  $y_i$ . The scalar  $y_i$  is a weighted sum of the responses at each of the spectral bands within the observed pixel vector. The  $l \times 1$  weight vector  $\mathbf{w}$  is given by:

$$\Phi_x = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T \quad (6.23)$$

and the output of the combining process at each pixel is:

$$y_i = \mathbf{x}_i \mathbf{w}_{CEM}^T \quad (6.24)$$

Figure 6.14 graphically depicts the effect of the weight vector on an observed pixel vector.

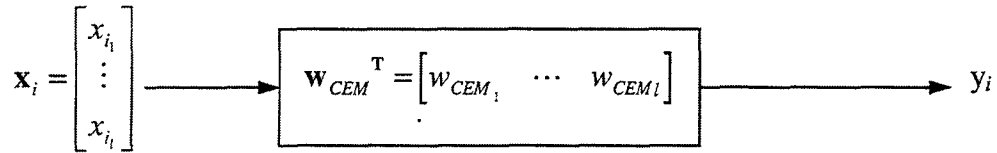


Figure 6.14: Effect of the CEM Weight Vector Components on the Observed Pixel Vector Components.

The derivation of a  $\mathbf{w}_{CEM}$  is driven by two conditions. The first requires that the energy at each output image pixel,

$$E = \sum_{i=1}^q y_i^2 \quad (6.25)$$

is minimized. The second requires that the result of applying the operator to the pixel vector of interest,  $\mathbf{d}$ , be unity, or

$$\mathbf{w}_{CEM}^T \mathbf{d} = 1 \quad (6.26)$$

The solution to the minimization problem is the same as the MVDR beamformer. The CEM operator is defined as:

$$\mathbf{w}_{CEM} = \frac{\Phi_x^{-1} \mathbf{d}}{\mathbf{d}^T \Phi_x^{-1} \mathbf{d}} \quad (6.27)$$

The CEM operator in this case differs from the MVDR beamformer because it uses the known desired pixel vector,  $\mathbf{d}$ , instead of the complex steering vector,  $\mathbf{s}$ , which represents the frequency response for a particular electric angle. The CEM operator defined in Equation 6.27 uses the signal processing definition of the correlation matrix because this is the form used in deriving the MVDR. The difference between the covariance and correlation matrices is that the covariance matrix is formed by removing the mean of the vectors. The sample correlation matrix is also formed as an outer product, but the mean of the vectors is not removed. Assuming that there are  $N$  observation pixel vectors,  $(\mathbf{x}_1, \dots, \mathbf{x}_k, \dots, \mathbf{x}_N)$  in the image, the sample correlation matrix is formed as:

$$\Phi_x = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T \quad (6.28)$$

For the remainder of this chapter, the “correlation” matrix refers to the signal processing version vice the remote sensing version, as defined in the Chapter III of this study.

The following paragraphs investigate the effect of using both the signal processing correlation and covariance matrices to form the CEM operator. The CEM operator formed with the correlation matrix is applied to the Davis Monthan sub-scene. The CEM operator formed with the covariance matrix is applied to the entire Davis Monthan scene. Two types of aircraft are chosen as desired endmembers in the whole scene and the performance of the CEM technique in detecting each type is noted.

The implementation of the CEM technique must take the numerical behavior of the second order statistical matrices into account. In hyperspectral imagery covariance or correlation matrices, the ratio of the largest to the smallest eigenvalue is always very large. This is a comment on the high spectral redundancy prevalent in this type of data. The ratio of largest to smallest eigenvalues is referred to as the condition number. The condition number indicates the degree of accuracy that may be expected in numerical calculation of a matrix inverse (Golub and Van Loan, 1983, p. 26). The large condition number of  $\Phi_x$  or  $\Sigma_x$  for hyperspectral imagery leads to a highly inaccurate inverse if standard inversion methods are used (Harsanyi, 1993, p. 87). The correlation or covariance matrix is said to be ill-conditioned in this case. The CEM technique

requires inversion of the correlation or covariance matrix, so an alternative that alleviates the problem of ill-conditioning is to use an approach based on the eigenstructure of  $\Phi_x$  or  $\Sigma_x$ . This approach takes advantage of two properties of the unitary transform, which will be demonstrated with the correlation matrix but are equally applicable to the covariance matrix. First,  $\Phi_x$  and  $\Phi_x^{-1}$  have the same eigenvectors. Second,  $\Phi_x$  and  $\Phi_x^{-1}$  have eigenvalues that are reciprocals of each other. The end result is that  $\Phi_x^{-1}$  may be decomposed as:

$$\Phi_x^{-1} = \mathbf{E}\mathbf{\Lambda}^{-1}\mathbf{E}^T \quad (6.29)$$

where  $\mathbf{E}$  is the matrix of correlation matrix eigenvectors packed into the columns and  $\mathbf{\Lambda}^{-1}$  is the diagonal matrix whose diagonal elements are the reciprocals of the eigenvalues of the correlation matrix. Additionally, the fact that only the first few eigenvectors of the sample correlation matrix contribute significantly to the total energy of a hyperspectral scene allows a good estimate of  $\Phi_x^{-1}$  using only the eigenvectors and eigenvalues of  $\Phi_x$  which correspond to the intrinsic dimensionality (Harsanyi, 1993, p. 88). The resulting estimate of the inverse of the sample correlation matrix may be written in terms of the first  $p-1$  eigenvectors and eigenvalues as

$$\hat{\Phi}_x^{-1} = \tilde{\mathbf{E}}\tilde{\mathbf{\Lambda}}^{-1}\tilde{\mathbf{E}}^T \quad (6.30)$$

where  $\tilde{\mathbf{E}}$  is the matrix of the first  $p-1$  eigenvectors defined in the previous LPD section and  $\tilde{\mathbf{\Lambda}}^{-1}$  is the diagonal matrix containing the reciprocals of the first  $p-1$  eigenvalues of  $\Phi_x$  (Harsanyi, 1993, p. 89). The intrinsic dimensionality is derived using the MDL criterion, or in this case has been determined from the eigenvalue magnitudes to be ten.

The CEM technique is applied to the Davis Monthan sub-scene using the correlation matrix to form the CEM operator. The scalar image that results is very much like those formed by the LPD and OSP techniques with an important exception. In the CEM output image, the target material is assigned values very near to unity. This corresponds to the intended effect of the CEM operator as seen in Equation 6.26. Figure 6.15 presents the histogram of this output image and shows how the target pixel vector has been assigned a value of unity in the CEM output image. The desired pixel vector was chosen from the wing of an aircraft. The wing pixel and the fuselage pixel appear in the opposite relative positions in the CEM output histogram than in the OSP and LPD output histograms. The CEM output preserves the target pixel vector as the highest magnitude in this case. The image associated with Figure 6.15 is presented in Figure 6.16. As with



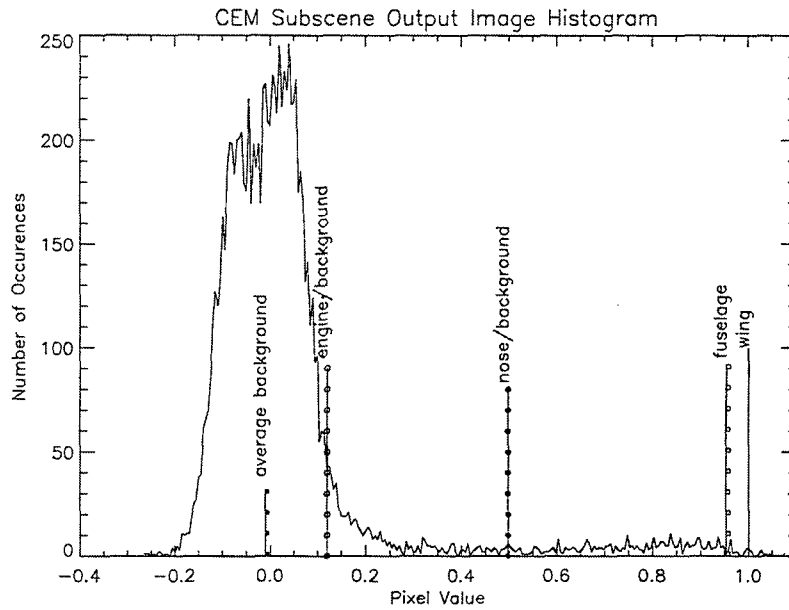


Figure 6.15: Histogram of Davis Monthan Sub-scene CEM Output.

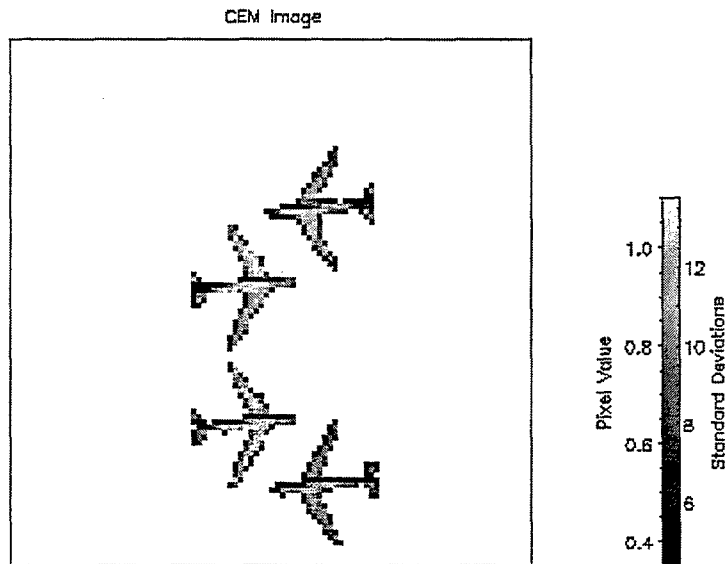


Figure 6.16: Davis Monthan Sub-scene CEM Output.

the OSP and LPD output images, the scene has been thresholded to accentuate the target pixels. The large number of standard deviations away from the center of

the background distribution is due to the small standard deviation of the background distribution. The color version of Figure 6.16 may be found in Appendix A.

The entire Davis Monthan scene is used to demonstrate the ability of the CEM technique to distinguish various types of targets within the scene. The covariance matrix of the data is used in this case, and the results show that the CEM output behaves similarly to the case in which the correlation matrix was used. Figure 6.17 shows the histogram of the output image in which a P-3 aircraft pixel was chosen as the desired pixel vector. The CEM technique displays the

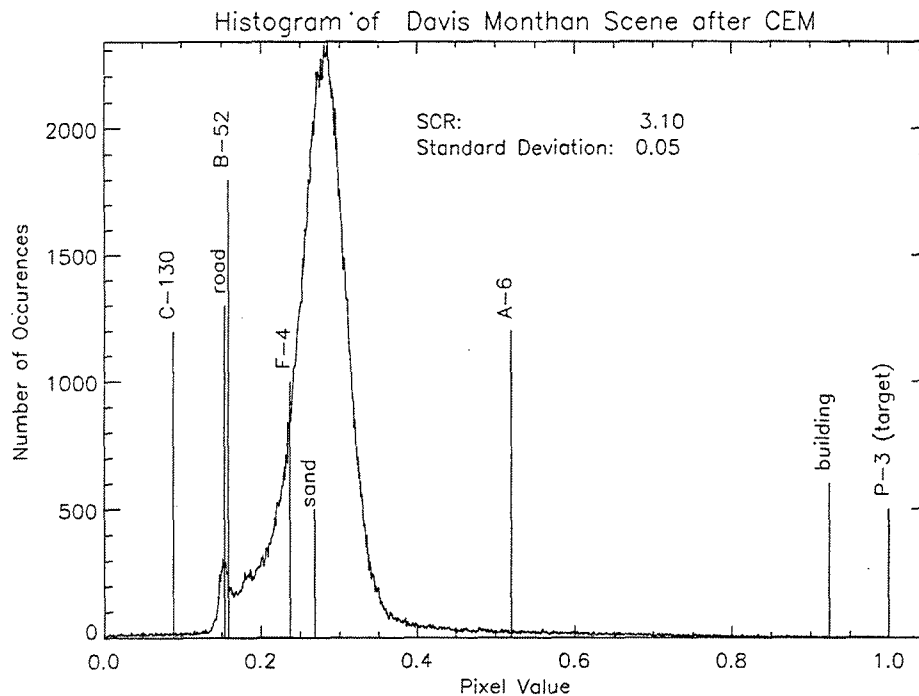


Figure 6.17: Histogram of CEM Scalar Image Using P-3 Pixel Vector as the Target.

quality of ordering the magnitudes of the output image pixels in such a fashion as to maximize the target. The image associated with this histogram is shown in figure 6.18. The color version may be found in Appendix A. The most notable feature of this image is that the P-3 aircraft have been successfully extracted from

# CEM Image

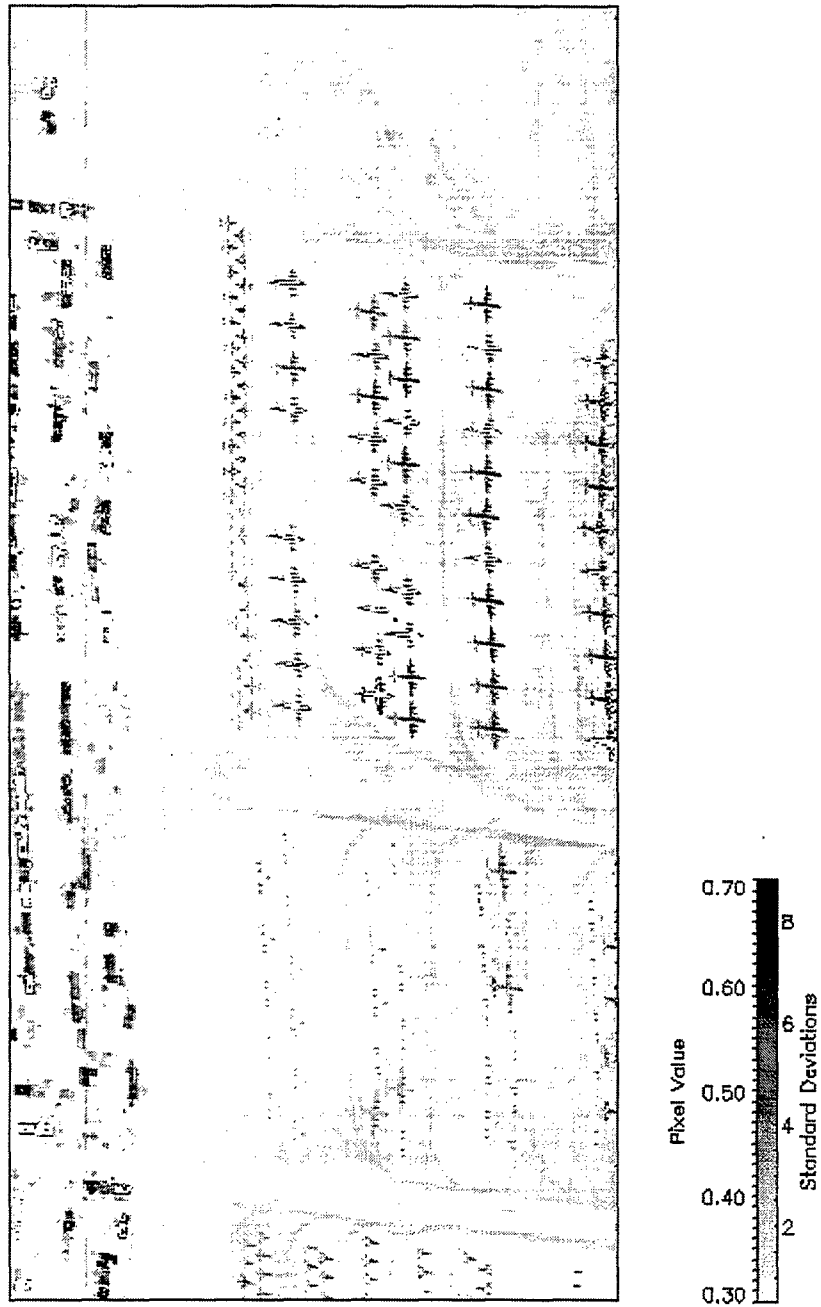


Figure 6.18: Davis Monthan CEM Output Image Using P-3 Pixel Vector as the Target.

the background. In most instances, the fuselages of the aircraft have higher magnitudes than the other parts of their bodies. Several buildings are also present with high enough magnitudes to be included in the thresholded image.

The CEM technique was also applied to the B-52 aircraft in the scene. Figure 6.19 shows the histogram of the resulting output image. The differentiation of the target from the background is not as evident in this case as in the P-3 target case. The target pixel is not the highest in magnitude, but its value is unity. The SCR has decreased in this case. The lack of clear distinction between targets is readily apparent in Figure 6.19. It reveals that though the B-52 pixels are all uniformly high, there are also several other areas which have high values in the scene, such as other aircraft types.

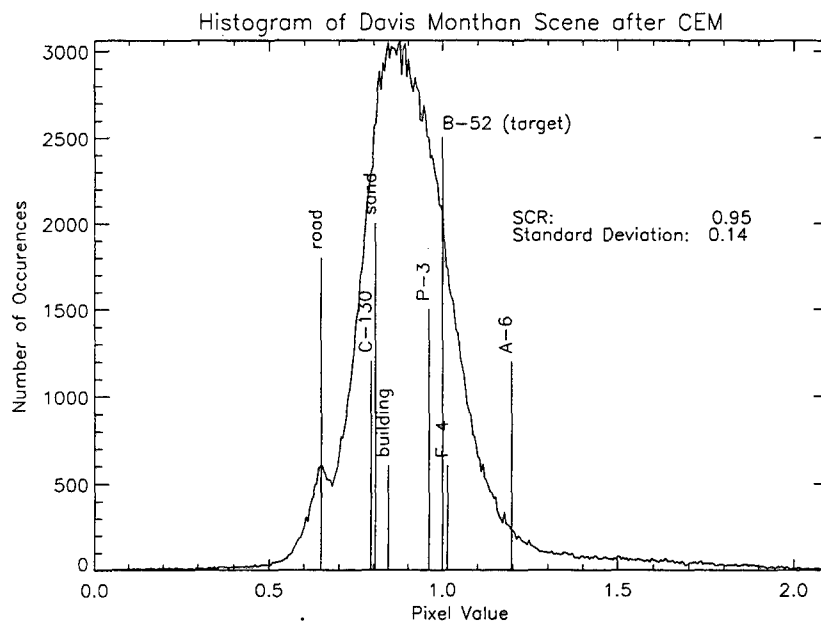


Figure 6.19: Histogram of CEM Scalar Image Using B-52 Pixel Vector as the Target.

The preceding results demonstrate that the CEM technique behaves similarly using the covariance or correlation matrices. It also has the desirable quality of assigning the target pixel vector values at or near unity in the output image. Harsanyi, Farrand, and Chang (1994) suggest that the CEM technique is well suited for the detection of distributed subpixel targets. Their examples include the detection of vegetation and rocks, and soils partially covered by vegetation. The target spectrum in this case occurs over numerous pixels. Their experimental evaluation of the CEM technique involves the detection of various image endmembers known to occur within an AVIRIS image of the Lunar Crater

Volcanic Field, Nevada. The CEM output detection of pixels containing red oxidized basaltic cinders, rhyolite, playa, and vegetation corresponds to ground truth maps of the same area.

As a final comment on the CEM technique, it is useful to examine the  $w_{CEM}$  operator developed for cases where the target pixel vectors are different. Figure 6.20 presents the plot of the  $w_{CEM}$  operator associated with the B-52, P-3,

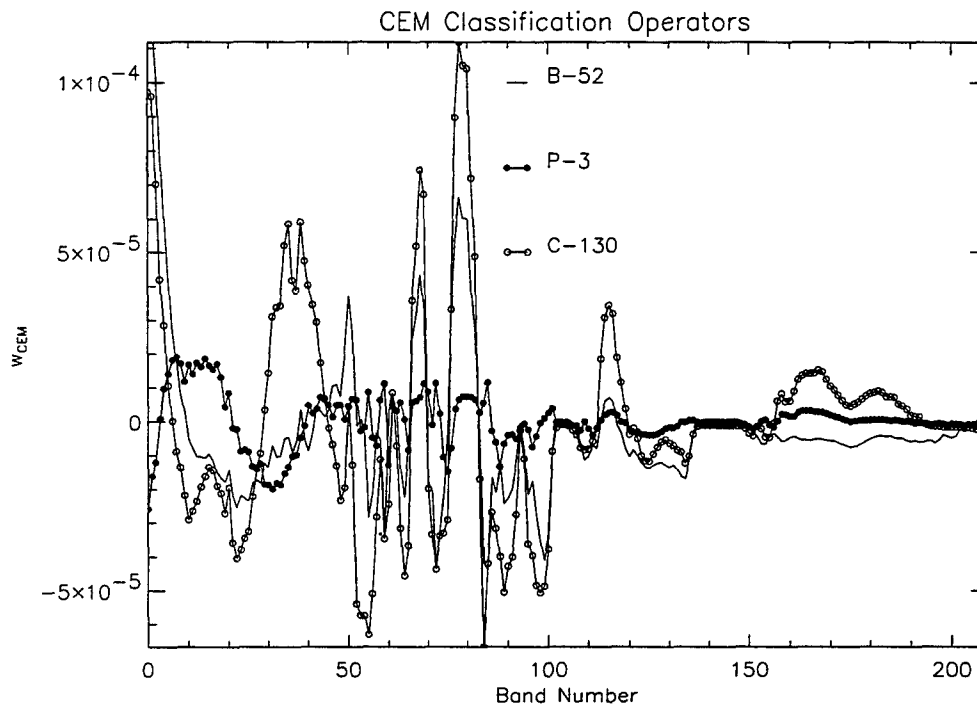


Figure 6.20: Comparison of the CEM operator for Three Different Target Pixel Vectors.

and C-130 aircraft target pixel vectors. The CEM operator associated with the P-3 shows less variability than those associated with the C-130 and the B-52. This behavior unexpectedly results in better target differentiation in the output images. The CEM operator in this case only depends on the behavior of the target pixel vectors, since the inverse of the covariance matrix is the same for all three. Thus, the results of Figure 6.20 are in essence a statement about the subtle differences in various aircraft spectra.

### 3. Adaptive Multidimensional Matched Filter

The matched filter can also be derived from a hypothesis test approach, which is more commonly associated with statistically-based classification. Stocker, Reed, and Yu (1990) use techniques from adaptive signal processing which exploit spatial and spectral differences between a target and the background. They apply the techniques to the problem of multispectral infrared imagery target enhancement. The goal is to test the data for the presence of a signal of known spatial shape and spectral signature. The image is partitioned into  $N$ -pixel subframes. It is assumed that there are  $l$  bands and that the pixel vector  $\mathbf{x}_i$  represents the  $l$  spectral observations from the  $i^{\text{th}}$  pixel of the subframe. It is further assumed that the target shape in each band can be described as:

$$\mathbf{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_N \end{bmatrix} \quad (6.31)$$

and that the shape vector satisfies the normalization  $\mathbf{s}^T \mathbf{s} = 1$ . The spectral signature of the target is described as:

$$\mathbf{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_l \end{bmatrix} \quad (6.32)$$

(Stocker, Reed, and Yu, 1990, p. 219). Both  $\mathbf{s}$  and  $\mathbf{d}$  are known *a priori*. The optimal detector for the target is derived from the joint probability distribution of the spectral observations using signal present and signal absent hypotheses (Stocker, Reed, and Yu, 1990, p. 219). The target may be viewed as an additive signal intensity pattern to the observations, given by  $s_i \mathbf{d}$  for the pixel vector  $\mathbf{x}_i$ . Since zero mean data is assumed in each subframe, the presence of the additive signal will only affect the mean of the data, as  $E\{\mathbf{x}_i\} = s_i \mathbf{d}$ , and the parametric form of the distribution will depend only on the background scene statistics. Thus, spectral images can be modeled as nonstationary Gaussian random processes with rapidly varying spectral mean and more slowly varying covariance functions (Stocker, Reed, and Yu, 1990, p. 219).

Each band of the image is prefiltered so that the local mean of each subframe is removed. This allows the background random process in each subframe to be approximated by zero mean locally stationary Gaussian statistics. The spectral observations in each subframe,  $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  can be modeled as independent Gaussian random vectors with zero mean and a common spectral

covariance matrix,  $\Sigma_x$  (Stocker, Reed, and Yu, 1990, p. 220). The pdf for the signal absent hypothesis is:

$$p_0(x_1, \dots, x_N) = (2\pi)^{-Nl/2} |\Sigma_x|^{-N/2} e^{-\frac{1}{2} \sum_{i=1}^N x_i^T \Sigma_x^{-1} x_i} \quad (6.33)$$

and the pdf for the signal present hypothesis is:

$$p_1(x_1, \dots, x_N) = (2\pi)^{-Nl/2} |\Sigma_x|^{-N/2} e^{-\frac{1}{2} \sum_{i=1}^N (x_i - s_i d)^T \Sigma_x^{-1} (x_i - s_i d)} \quad (6.34)$$

(Stocker, Reed, and Yu, 1990, p. 220) A likelihood ratio test formed from these two hypotheses leads to the optimal detector which is a linear filter that can be expressed in terms of the data as:

$$y(\mathbf{X}) = \mathbf{d}^T \Sigma_x^{-1} \mathbf{X} \mathbf{s} \quad (6.35)$$

where  $\mathbf{X}$  is a  $l \times N$  matrix formed by packing all of the pixel vectors  $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  into the columns of  $\mathbf{X}$ . The linear filter can be expressed in terms of its components as:

$$y_i = (\mathbf{d}^T \Sigma_x^{-1} \mathbf{x}_i)(s) \quad (6.36)$$

The parentheses emphasize that  $y_i$  may be interpreted as an optimum weighted combination of spectral samples in each subframe followed by a spatial matched filter which is matched to the target shape,  $s$  (Stocker, Reed, and Yu, 1990, p. 221). The optimum weighted combination arises from the fact that the covariance matrix may be decomposed using a unitary transformation, so that the linear filter may be rewritten as:

$$y_i = (\mathbf{d}^T \mathbf{E} \mathbf{\Lambda}^{-1})(\mathbf{E}^T \mathbf{x}_i) s \quad (6.37)$$

with  $\mathbf{Q}$  being the matrix of eigenvectors, and  $\mathbf{\Lambda}$  representing the matrix of eigenvalues. The term in the first parentheses serves as a weight vector which gives the optimum weighted combination of principal components for the term in the second parentheses, which is the projection of the original data onto the eigenvectors of  $\Sigma_x$ . The end result of this matched filtering is to maximize the signal-to-clutter ratio (SCR) and allow easier detection of the target.

The application of the multidimensional matched filter to an unknown background is accomplished by estimating the covariance matrix for each subframe using the zero mean data matrix outer product as:

$$\hat{\Sigma}_x = \frac{1}{N} \mathbf{X} \mathbf{X}^T \quad (6.38)$$

(Yu, Reed, and Stocker, 1993, p. 2463). The clutter adaptive detector is formed using a generalized likelihood ratio test. The estimate of the background

covariance replaces the known covariance matrix, and the clutter adaptive test may be written as:

$$\frac{(\mathbf{d}^T \hat{\Sigma}_x^{-1} \mathbf{Xs})^2}{(\mathbf{d}^T \hat{\Sigma}_x^{-1} \mathbf{d})(1 - \frac{1}{N} \mathbf{s}^T \mathbf{X}^T \hat{\Sigma}_x^{-1} \mathbf{Xs})} \quad (6.39)$$

If the ratio is larger than a specified threshold, then the target is present (Yu, Reed, and Stocker, 1993, p. 2463). This approach is adaptive in that the above likelihood test changes according to the background statistics estimated for each subframe.

Yu, Reed, and Stocker (1993) demonstrate the effectiveness of the adaptive multiband matched filter using data collected by the six-band Thermal Infrared Multispectral Scanner (TIMS) instrument of Adelaide, Australia. This instrument has ten meter ground resolution. The adaptive matched filter is applied to the scene and extracts certain man-made features such as homes and roads.



## **VII. THE LIMITED IMAGE ENDMEMBERS FAMILY OF TECHNIQUES**

### **A. DESCRIPTION**

Though *a priori* knowledge of the image endmembers may be unavailable, a library of reference spectra is readily available in many cases. The reference spectra have been collected by laboratory spectral analysis of materials or previous remotely sensed observations of ground truth. Three techniques exploit this knowledge of spectral libraries. They are based upon comparing observed pixel vectors with spectra in the library and using an objective criterion to decide on the constituent endmembers of the observed pixel vectors. Two techniques assume the mixed pixel problem and attempt to unmix the observations into constituent endmembers. These are the endmember identification and the partial unmixing techniques. The third technique, the spectral angle mapper (SAM), does not assume mixed pixels (Harsanyi, 1993, p. 11). This distinction is important to keep in mind when deciding when and how a specific technique should be applied. All of these techniques seek to classify observed pixel vectors by using a reference library. In other cases, a limited amount of information regarding the abundance of a particular endmember is available via ground truth. A fourth technique based on the singular value decomposition capitalizes upon this information to form an operator that can be applied to future images in which no ground truth is available.

### **B. BACKGROUND DEVELOPMENT**

The techniques discussed in this section have their origins in diverse areas. The multiple signal classification (MUSIC) approach has its roots in the field of array processing and high resolution spectral estimation. The partial unmixing technique of endmember identification has the concepts of convex geometry as its foundation. The SAM technique is very much like the signal processing concept of the correlation detector for signal detection. The singular value decomposition (SVD) is a powerful tool from linear algebra that can decompose a matrix in a manner similar to eigendecomposition. The SVD technique is computationally efficient and reveals information about the

structure of the data. Though they have different roots, the techniques all strive to use available information to the fullest extent.

## 1. Multiple Signal Classification (MUSIC) Technique

Schmidt originally developed the MUSIC technique to determine the parameters of multiple wavefronts arriving at an antenna array (Schmidt, 1986, p. 276). His development is followed here with the notation altered to correspond with that used throughout this study in describing the hyperspectral problem. The model for the observations of the signals received at the antenna elements is given by:

$$\mathbf{x} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{n} \quad (7.1)$$

where  $\mathbf{x}$  is the  $l \times 1$  observation vector,  $\mathbf{M}$  is a  $l \times p$  matrix of known functions of the signal arrival angles and array element locations which may be written as:

$$\mathbf{M} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{m}(\theta_1) & \cdots & \mathbf{m}(\theta_p) \\ \downarrow & & \downarrow \end{bmatrix} \quad (7.2)$$

The columns of  $\mathbf{M}$  are called the mode vectors, and represent the array responses to a particular direction of arrival of a signal. There are  $p$  incident signals and  $l$  array elements, and it is assumed that  $l > p$ . The  $p \times 1$  vector  $\boldsymbol{\alpha}$  represents the amplitude and

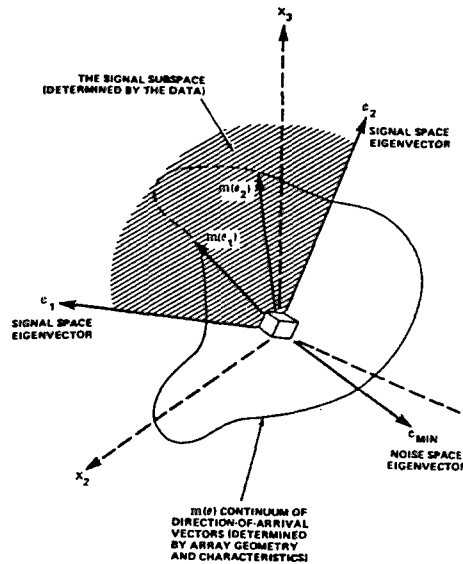


Figure 7.1: Geometric Representation of the Observed Signal Model with Three Antenna Elements. After Schmidt, 1986, p. 277.

phase of the incident signal at an arbitrary reference point, and  $\mathbf{n}$  is the  $l \times 1$  noise vector (Schmidt, 1986, p. 276). The model of Equation 7.1 is expressing  $\mathbf{x}$  as a linear combination of mode vectors, with the elements of  $\alpha$  representing the coefficients of combination. A geometric view of the situation puts the problem into perspective. Figure 7.1 shows that the linearly independent columns of  $\mathbf{M}$  determine the subspace within which the observations must exist. The vectors  $\mathbf{m}(\theta_i)$  represent all possible choices of incident mode vectors, and are depicted by the continuum in  $l$ -dimensional space in Figure 7.1. The signal subspace is spanned or defined by the first two eigenvectors of the covariance matrix of the observations, and is denoted by the shaded lines. The eigenvector associated with the smallest eigenvalue defines the noise subspace in three-dimensional case of Figure 7.1. The direction of arrival estimation problem for several incident wavefronts consists of locating the intersections of  $\mathbf{m}(\theta_i)$  with the signal subspace (Schmidt, 1986, p. 277).

The MUSIC algorithm begins with the covariance matrix model of the data, which is obtained by applying the definition of covariance as a statistical expectation to the data vector:

$$\Sigma_{\mathbf{x}} = E\{\mathbf{x}\mathbf{x}^{*T}\} = \mathbf{M}E\{\alpha\alpha^{*T}\}\mathbf{M}^{*T} + E\{\mathbf{n}\mathbf{n}^{*T}\} = \mathbf{M}\Sigma_s\mathbf{M}^{*T} + \sigma^2\mathbf{I} \quad (7.3)$$

$\Sigma_{\mathbf{x}}$  is a  $l \times l$  covariance matrix, where the results assume that the signal and noise random vectors are uncorrelated. The complex conjugate transpose operation is denoted by the  $^{*T}$  and is required since the MUSIC problem is formulated to deal with complex signals. The  $p \times p$  matrix  $\Sigma_s$  is diagonal if all of the elements of  $\alpha$  are uncorrelated. The  $\sigma^2\mathbf{I}$  matrix assumes that the additive noise in the problem is white with variance  $\sigma^2$  (Schmidt, 1986, p. 277). The eigenstructure of  $\Sigma_{\mathbf{x}}$  contains complete information on the frequencies  $\{\omega_1, \dots, \omega_k, \dots, \omega_l\}$ , which is the parameter of interest in determining the direction of arrival of a particular signal. The eigendecomposition of  $\Sigma_{\mathbf{x}}$  yields eigenvalues which can be divided into two groups based on their magnitudes. Since the number of linearly independent columns or rank of the matrix  $\mathbf{M}\Sigma_s\mathbf{M}^{*T}$  is  $p$ , the implication is that this matrix has  $p$  strictly positive eigenvalues with the remaining  $l-p$  eigenvalues equal to zero. Accounting for the constant variance of the noise, which adds  $\sigma^2$  to all eigenvalue magnitudes, the magnitude of eigenvalues can be summarized as:

$$\begin{aligned} \lambda_k &> \sigma^2, \quad k = 1, \dots, p \\ \lambda_k &= \sigma^2, \quad k = p+1, \dots, l \end{aligned} \quad (7.4)$$

Consequently, the eigenvectors associated with the  $p$  eigenvalues greater than  $\sigma^2$  are packed into a matrix  $\mathbf{S}$  which defines the signal subspace, and the eigenvectors associated with the  $l-p$  eigenvalues equal to  $\sigma^2$  are packed into a matrix  $\mathbf{G}$  which defines the noise subspace as shown below:

$$\mathbf{E}_s = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{e}_{s_1} & \cdots & \mathbf{e}_{s_p} \\ \downarrow & & \downarrow \end{bmatrix} \quad \mathbf{E}_n = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{e}_{n_1} & \cdots & \mathbf{e}_{n_{l-p}} \\ \downarrow & & \downarrow \end{bmatrix} \quad (7.5)$$

(Stoica and Moses, 1996, p. 206). The MUSIC algorithm involves the projection of the signal onto the noise subspace, and is developed using the  $\mathbf{E}_n$  matrix. The application of the noise subspace eigenvectors to the covariance matrix results in the following equalities:

$$\Sigma_x \mathbf{E}_n = \mathbf{M} \Sigma_s \mathbf{M}^{*T} \mathbf{E}_n + \sigma^2 \mathbf{E}_n = \sigma^2 \mathbf{E}_n \quad (7.6)$$

This equation implies that  $\mathbf{M} \Sigma_s \mathbf{M}^{*T} \mathbf{E}_n = \mathbf{0}$ , and since  $\mathbf{M} \Sigma_s$  has full column rank ( $p$  linearly independent columns), it follows that  $\mathbf{M}^{*T} \mathbf{E}_n = \mathbf{0}$  (Stoica and Moses, 1996, p. 207). The orthogonal columns  $\mathbf{E}_n$  of belong to the null space of  $\mathbf{M}^{*T}$ , and can be used to form a projector onto the noise subspace. Using linear algebra concepts from the theory of least squares, the projection matrix onto the noise subspace is formed as:

$$\mathbf{P}_{\text{noise}} = \mathbf{E}_n (\mathbf{E}_n^{*T} \mathbf{E}_n)^{-1} \mathbf{E}_n^{*T} = \mathbf{E}_n \mathbf{E}_n^{*T} \quad (7.7)$$

where the second equality is true because  $\mathbf{E}_n$  is an orthonormal unitary matrix (Therrien, 1992, p. 623). The important result of this derivation is that the true frequencies associated with the direction of arrival of signals are the true solutions to the equation:

$$\mathbf{m}^{*T}(\theta) \mathbf{P}_{\text{noise}} \mathbf{m}(\theta) = \mathbf{0} \quad (7.8)$$

(Stoica and Moses, 1996, p. 208). The noise subspace is orthogonal to the signal subspace, implying that the noise subspace projector,  $\mathbf{P}_{\text{noise}}$ , nulls incoming signals while allowing incoming noise to pass. By noting at which frequency the nulls occurred, the frequency or direction of arrival for a signal is estimated. The MUSIC algorithm uses the eigenvectors of the data covariance matrix to construct an orthogonal projector. It is worth noting that this algorithm is designed to detect the direction of arrival and spectral content of orthogonal signals such as complex sinusoids.

## 2. Convex Sets

Convex sets are based on concepts from linear algebra and geometry, and have a history of application to mathematical optimization problems (Lay, 1982, p. vii). A

definition of points in a convex set is that they are positive, unit-sum linear combinations of a fixed set of points (Boardman, 1995, p. 15). The fixed set of points define the vertices of a convex hull, which is the intersection of all the convex sets containing the particular convex set of interest (Lay, 1982, p. 11). A flat is defined as a translate or mapping of a linear subspace (Lay, 1982, p. 12). This is basically a projection of an  $l$ -dimensional cloud onto a subspace. An  $l$ -simplex is the simplest geometric figure that has no redundancy in representing a set of data points (Boardman, 1995, p. 17). A fledge is a flat that is on the edge of a data cloud (Boardman, 1995, p. 17). While these concepts may seem vague, their application to high dimensional hyperspectral data in the partial unmixing technique clarifies their utility. The reader is referred to Lay (1982) for a complete discussion of these concepts.

### 3. The Correlation Detector

The detection of deterministic signals in noise is a classic problem in signal processing. The problem is formulated as an observed sequence of the form:

$$x[n] = s[n] + n[n] \quad 0 \leq n \leq N-1 \quad (7.9)$$

where  $s[n]$  is a deterministic sequence and  $n[n]$  is added noise (Therrien, 1992, p. 4). If the noise is white, where the samples are uncorrelated random variables, then the optimum way to detect the signal is a correlation detector. Figure 7.2 illustrates the

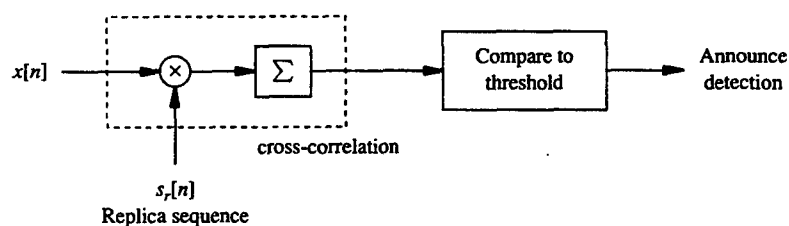


Figure 7.2: Correlation Detector. From Therrien, 1992, p. 5.

detection process, where the replica sequence  $s_r[n]$  is the same as  $s[n]$ . The replica sequence is multiplied by the input sequence,  $x[n]$ , summed, and compared to a threshold to determine whether or not the deterministic signal is present in the input sequence. The process of multiplication and summation of the two sequences is known as cross-correlation (Therrien, 1992, p. 5). This simple detector assumes *a priori* knowledge of a particular signal that one wishes to detect. One shortfall in the concept is that the input

sequence is assumed to be one signal buried in noise. If several interfering signatures were present that were correlated with the signal of interest, the correlation detector would be unable to successfully announce its presence (Harsanyi, 1993, p. 11).

#### 4. Singular Value Decomposition (SVD)

The SVD is a powerful tool that enables eigenvalues and eigenvectors to be found with better numerical precision and allows decomposition of all types of matrices, not just symmetric ones (Therrien, 1992, p. 54). The best way to understand the operation of the SVD is to view it in terms of vector subspaces. Van Der Veen, Deprettere, and Swindlehurst (1993) review the subject of the SVD in subspace based spectral estimation, and their background description of the SVD is given here in notation that is consistent the notation of spectral imagery analysis developed in this study. Assuming that one begins with real-valued data in the form of a  $l$ -band  $\times$   $N$ -sample matrix  $\mathbf{X}$ , it is desirable to know the number of linearly independent columns of  $\mathbf{X}$ . If the number of linearly independent columns of  $\mathbf{X}$  is  $p$ , then  $p$  is referred to as the dimension of the column space of  $\mathbf{X}$ . If  $p = l$ , then  $\mathbf{X}$  is said to have full rank, and is rank-deficient if  $p < l$ . Euclidean  $l$ -dimensional space can be completely described or spanned by the columns of any unitary  $l \times l$  square matrix, which form an orthonormal basis. A  $l \times l$  unitary matrix  $\mathbf{U}$  can be chosen so that the  $p$ -dimensional column space of  $\mathbf{X}$  is spanned by a subset represented by the first  $p$  columns of  $\mathbf{U}$ . This  $l \times p$  subspace is called  $\hat{\mathbf{U}}$  and its  $l \times l-p$  orthogonal complement is called  $\hat{\mathbf{U}}^\perp$ , both of which are shown below as parts of the matrix  $\mathbf{U}$ :

$$\mathbf{U} = [\hat{\mathbf{U}} \quad \hat{\mathbf{U}}^\perp] \quad (7.10)$$

where the dimensions of the partitioned matrix  $\mathbf{U}$  are  $l \times l$ . The  $\wedge$  in this context does not refer to an estimate, but rather a smaller-sized component matrix. The fact that  $\mathbf{U}$  is a unitary matrix leads to several properties of its component matrices:

$$\begin{aligned} \hat{\mathbf{U}}^T \hat{\mathbf{U}} &= \mathbf{I}_p \\ \hat{\mathbf{U}}^T \hat{\mathbf{U}}^\perp &= 0 \\ (\hat{\mathbf{U}}^\perp)^T \hat{\mathbf{U}}^\perp &= \mathbf{I}_{l-p} \\ \hat{\mathbf{U}} \hat{\mathbf{U}}^T + \hat{\mathbf{U}}^\perp (\hat{\mathbf{U}}^\perp)^T &= \mathbf{P}_C + \mathbf{P}_C^\perp = \mathbf{I}_l \end{aligned} \quad (7.11)$$

where the subscripts to the identity matrices denote their dimension, and the  $\mathbf{P}_C$  and  $\mathbf{P}_C^\perp$  denote the orthogonal projector matrices onto the column space of  $\mathbf{X}$  and its orthogonal

compliment, respectively (Van Der Veen, Deprettere, and Swindlehurst, 1993, p. 1289). These results were seen in the Chapter V discussion of the theory of least squares, but were not stated explicitly. The important observation is that any vector  $\mathbf{x}$  in  $l$ -space can be decomposed into two mutually orthogonal vectors that belong to the spaces spanned by the columns of  $\hat{\mathbf{U}}$  and  $\hat{\mathbf{U}}^\perp$ . The  $N \times N$  unitary matrix,  $\mathbf{V}$ , may be decomposed similarly to yield:

$$\mathbf{V} = [\hat{\mathbf{V}} \mid \hat{\mathbf{V}}^\perp] \quad (7.12)$$

The dimension of the matrix component  $\hat{\mathbf{V}}$  are  $N \times p$  and the dimensions of  $\hat{\mathbf{V}}^\perp$  are  $N \times N-p$ .

The SVD of the  $l \times N$  data matrix  $\mathbf{X}$  is defined in terms of the above unitary matrices as:

$$\mathbf{X} = [\hat{\mathbf{U}} \mid \hat{\mathbf{U}}^\perp] \mathbf{S} \begin{bmatrix} \hat{\mathbf{V}}^T \\ (\hat{\mathbf{V}}^\perp)^T \end{bmatrix} = \mathbf{U} \mathbf{S} \mathbf{V}^T \quad (7.13)$$

where  $\mathbf{S}$  is a  $l \times N$  matrix containing the singular values of  $\mathbf{X}$ . The singular values are positive number ordered so that  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq \sigma_{p+1} = \dots = \sigma_l = 0$  and correspond to the square roots of the eigenvalues. Only  $p$  singular values are nonzero, and the corresponding  $p$  columns of  $\hat{\mathbf{U}}$  are called the left singular vectors of  $\mathbf{X}$ . The  $p$  columns of  $\hat{\mathbf{V}}$  are called the right singular vectors of  $\mathbf{X}$ . The SVD can be written in terms of these smaller matrices in what is termed the economy size or reduced rank SVD as:

$$\mathbf{X} = \hat{\mathbf{U}} \hat{\mathbf{S}} \hat{\mathbf{V}}^T = \begin{bmatrix} \uparrow & & \uparrow \\ \hat{\mathbf{u}}_1 & \dots & \hat{\mathbf{u}}_p \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{bmatrix} \begin{bmatrix} \leftarrow \hat{\mathbf{v}}_1^T \rightarrow \\ \vdots \\ \leftarrow \hat{\mathbf{v}}_p^T \rightarrow \end{bmatrix} \quad (7.14)$$

(Van Der Veen, Deprettere, and Swindlehurst, 1993, p. 1289). This decomposition clearly shows that  $\mathbf{X}$  is a rank  $p$  matrix since it is composed of rank  $p$  matrices. It is also useful because it represents the original data matrix with fewer dimensions. The SVD of  $\mathbf{X}$  can be best explained by explicitly stating the steps involved in the mapping of a vector  $\mathbf{a}$  in  $N$ -space to a vectors  $\mathbf{b}$  in  $l$ -space as:

$$\mathbf{b} = \mathbf{X} \mathbf{a} = \mathbf{U} \mathbf{S} \mathbf{V}^T \mathbf{a} \quad (7.15)$$

Vector  $\mathbf{a}$  is rotated in  $N$ -space by  $\mathbf{V}^T$ , then scaled by the entries of  $\mathbf{S}$ , with  $l-p$  components projected to zero by the zero singular values, and finally rotated in  $l$ -space by  $\mathbf{U}$  to give  $\mathbf{b}$  (Van Der Veen, Deprettere, and Swindlehurst, 1993, p. 1289).

## C. OPERATION

The techniques considered in this section seek to determine the endmembers resident in an image by comparison with known spectra. The mechanics of how they work is outlined below.

### 1. MUSIC-Based Endmember Identification

Harsanyi, Farrand, Hejil, and Chang (1994) introduce a modification of the MUSIC method of spectral estimation to remote sensing applications. The advantage of using this method is that the number and identity of spectral signatures in an image can be determined without finding spectrally "pure" pixels in the scene. In assuming the mixed pixel problem and assuming no *a priori* knowledge of specific scene endmembers, one is obligated to use a library of known pure spectra that can be compared to candidates drawn directly from the scene. The basis of the MUSIC approach is the eigendecomposition of the covariance matrix into orthogonal matrices, one of which is used to construct a noise subspace projector. As in the MUSIC algorithm of spectral estimation, the noise subspace projector nulls those reference spectra which correspond to signatures found in the signal subspace of the scene.

The first step of the MUSIC approach is to use the noise-whitened covariance matrix to determine the number of distinct spectral signatures based on the MDL information theoretic criterion (Harsanyi, Farrand, Hejil, and Chang, 1994, p. 269). These concepts are fully developed in Chapter VI, and are viewed here in terms of their results.

The second step is to use the principal eigenvectors of the noise-whitened covariance matrix to form a subspace that is orthogonal to all possible linear combinations of spectral signatures in the scene, and then to use this noise subspace to form an orthogonal subspace projector (Harsanyi, Farrand, Hejil, and Chang, 1994, p. 271). The noise-whitened covariance matrix can be decomposed by unitary transform into:

$$\Sigma_{wx} = E \Lambda E^T = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{e}_1 & \cdots & \mathbf{e}_l \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_l \end{bmatrix} \begin{bmatrix} \leftarrow & \mathbf{e}_1^T & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{e}_l^T & \rightarrow \end{bmatrix} \quad (7.16)$$



where the number of spectral bands is  $l$ , and  $\lambda_i$  and  $\mathbf{e}_i$  are the  $i^{\text{th}}$  eigenvalue and eigenvector, respectively, of the noise-whitened covariance matrix. Note that all data is real. Assuming that the intrinsic dimensionality of the data is determined to be  $p-1$  by the MDL criterion, the eigenvector matrix  $\mathbf{E}$  can be partitioned into a signal subspace consisting of the eigenvectors associated with the  $p-1$  significant eigenvalues and a noise subspace consisting of the remaining  $l-p+1$  eigenvectors. The matrix of principal eigenvectors is called the signal subspace and is denoted as:

$$\tilde{\mathbf{E}}_s = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{e}_{s_1} & \cdots & \mathbf{e}_{s_{p-1}} \\ \downarrow & & \downarrow \end{bmatrix} \quad (7.17)$$

The principal eigenvectors represent the majority of the information regarding the endmembers of the scene, as demonstrated by the optimal representation property of the DKLT. This implies that any linear combination of the endmembers is represented by some linear combination of the principal eigenvectors (Harsanyi, Farrand, Hejil, and Chang, 1994, p. 271). The principal eigenvector matrix is used to form an operator which will project onto the subspace orthogonal to the signal subspace. The projection operator is the optimal least squares operator of the LPD technique expressed here as:

$$\mathbf{P} = (\mathbf{I} - \tilde{\mathbf{E}}_s \tilde{\mathbf{E}}_s^{\#}) \quad (7.18)$$

where the  $\#$  denotes the pseudo-inverse operation.

The final step of the MUSIC approach is to apply the noise subspace projection operator to a spectral library in order to identify those endmembers in the image which are closest to the library endmember spectra. Assuming that the  $i^{\text{th}}$  spectral library signature is given by the zero-mean vector  $\mathbf{r}_i$ , the endmember identification operator can be formulated as:

$$S(\mathbf{r}_i) = \mathbf{r}_i^T \mathbf{W} \mathbf{P} \mathbf{W}^T \mathbf{r}_i \quad (7.19)$$

(Harsanyi, Farrand, Hejil, and Chang, 1994). The whitening operator  $\mathbf{W}$  is derived from the eigenvectors and eigenvalues of the noise covariance matrix, as shown in Equation 6.21. It is applied to the estimated data covariance matrix in order to provide a noise-whitened estimate. The endmember identification operator  $S(\mathbf{r}_i)$  is applied to every reference spectrum, producing a scalar. The elements of the spectral library for which  $S(\mathbf{r}_i)$  is minimized are the spectral signatures closest to the endmembers in the scene. The significance of this technique is that the second order statistics of the observed data allow an objective means of determining image endmember identity.

The MUSIC technique approaches the problem of finding pixels containing target spectra by exploiting the statistics of the covariance matrix to derive an operator which is orthogonal to the subspace of the signals in the image. The rationale is that when this operator is applied to a library of reference spectra, the orthogonal subspace projection operator will minimize those spectra actually contained in the scene. This implies that the MUSIC approach could identify all of scene endmembers given an exhaustive spectral library. In the context of target detection, the problem is simplified in that the reference library need only contain targets of interest instead of all materials. Harsanyi, Farrand, Hejil, and Chang (1994) validate the technique by running it on five hundred simulated AVIRIS data pixels containing three endmembers. Their results show that the correct endmembers were identified by the MUSIC algorithm. The factors in this method which determine the accuracy of the detection center around the estimation of the covariance matrix, the determination of the intrinsic dimensionality of the data, and in the completeness of the spectral library in accounting for natural variability found in different spectra of the same type of target.

## **2. Partial Unmixing**

The traditional application of PCA to multispectral imagery does not account for the fact that the radiance observed from each pixel might be a mixture of spectra from different materials. The classification schemes that generally follow traditional PCA attempt to produce a crisp absence/presence decision for each pixel, ignoring the fact that the pixel may contain a mixture of spectra (Settle, 1996, p. 1045). A technique introduced by Smith, Johnson, and Adams (1985) deals with the mixed pixel issue in the context of determining the mineral types and relative abundances in planetary multispectral observations. The goal of their approach is to reduce the dimensionality of the observations to the total number of parameters that influence the measurements and to identify the parameters on which the spectral reflectance is functionally dependent (Smith, Johnson, and Adams, 1985, p. C797). The assumption is that reference laboratory spectra of the endmembers are available.

Briefly, the technique described by Smith, Johnson, and Adams (1985) is described as first forming a model of endmember mixing, and then applying PCA to determine if the observed spectral variance corresponds to the model. The intrinsic dimensionality of data corresponds to the number of parameters that influence the model.

The observed spectra are then projected onto the principal axes of variation corresponding to the most significant eigenvectors. The vertices of the shape described by this projection are the endmembers, and the abundances of mixed materials are estimated by forming ratios of the mixture to the total volume of the projection (Smith, Johnson, and Adams 1985, p. C798). This provides a systematic method with which to infer how the spectra contributing to the scene are related.

Boardman, Kruse, and Green (1995) generalize the above approach to accommodate the high dimensionality of hyperspectral imagery with the technique of partial unmixing. The partial unmixing technique assumes the mixed pixel problem. It compares the purest pixels in a scene to reference target spectra. The high purity pixels that do not closely match a target spectrum are used to determine a subspace that describes the scene background (Boardman, Kruse, Green, 1995, p. 24). The background subspace can then be used to determine projection vectors for the target subspace that serve to isolate target spectra in the image. The key to this approach is the selection of spectrally pure pixels. This process is facilitated by the MNF or NAPC transform, which reveals the intrinsic dimensionality of the data. The intrinsic dimensionality of the data can be represented by convex sets as the vertices of a convex hull in  $l$ -space (Boardman, Kruse, Green, 1995, p. 23).

Convex sets represent hyperspectral images as a collection of points in  $l$ -dimensional space, where each spectral channel corresponds to an axis of the space (Boardman, 1995, p. 14). The shape of the data represented in  $l$ -space and the patterns inherent within it help one to better understand the spectral information in the data. Boardman (1995) equates the linear mixed pixel problem to a convex set of points in  $l$ -space. The goal is to extract information regarding the presence of target spectra. In the convex set model, the mixed pixel spectra are represented as unit-sum linear combinations of the pure endmember spectra. The pure endmembers determine the vertices of the convex hull, whereas the mixed pixels are located at points inside the hull. The inherent dimensionality of the data is determined by finding the lowest dimensional subspace, or flat, that spans or represents all of the signals in the data, excluding the noise. The MNF or NAPC transform is used to obtain the intrinsic dimensionality through an eigenanalysis. The simplest geometric figure that can conform to the number of dimensions thus determined is termed an  $l$ -simplex. Some of the shapes that simplices can assume are depicted in Figure 7.3. The  $l+1$  pure endmember spectra form the vertices of the  $l$ -simplex. The purest pixels are then located in the data by an iterative projection

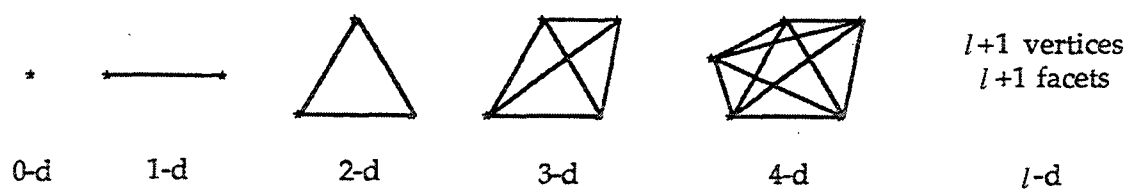


Figure 7.3: Mixing Simplices from Zero to  $l$  Dimensions. After Boardman, 1995, p. 17.

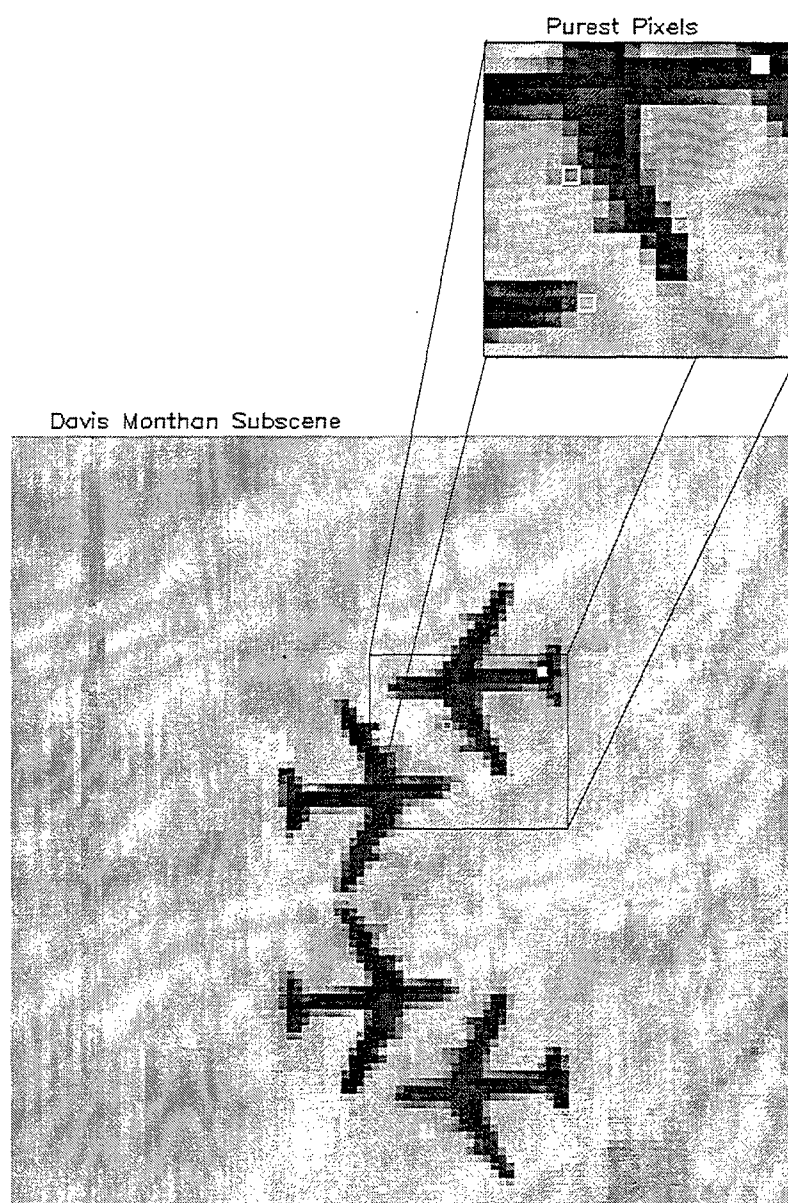
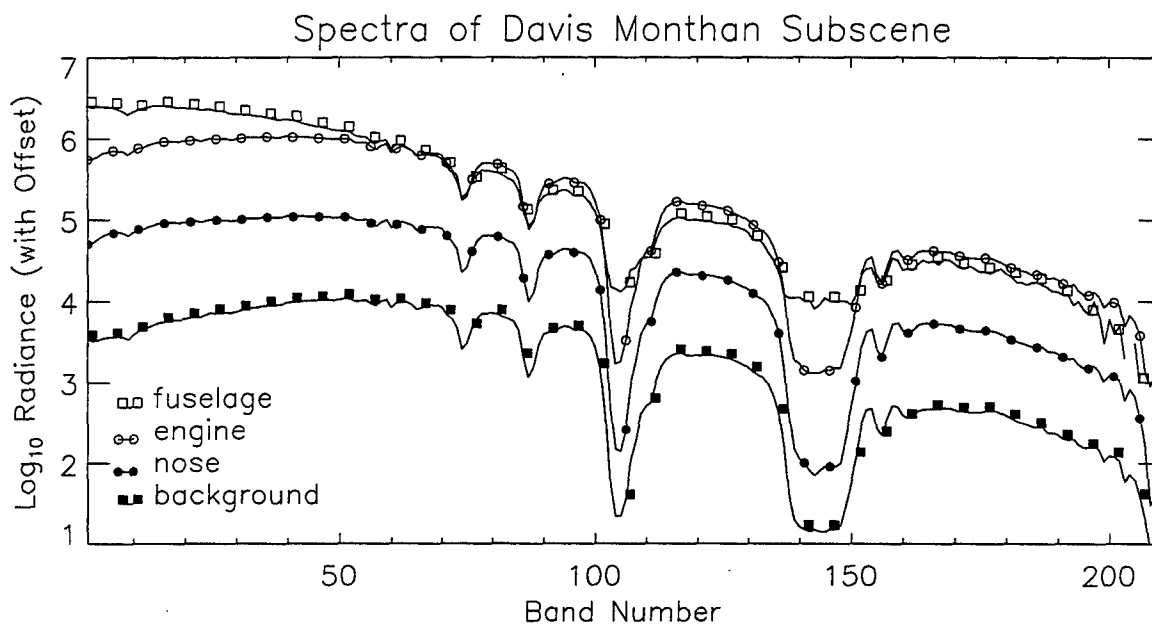


Figure 7.4: Purest Pixels in the Davis Monthan Sub-scene.

to determine which pixel are within a threshold of being on the convex hull of the data cloud.

The purest pixels of the Davis Monthan sub-scene are shown in Figure 7.4. These were located using the MNF output image and the iterative projection scheme described above. The white boxes indicate the four purest pixels in the image that exceeded a threshold of ten after 100 projection iterations. It is interesting to note that these pixels correspond to spectra that appear to be mixed pixels since three of them occur on the aircraft edges. The fourth purest pixel corresponds to the background. The color version of Figure 7.4 may be found in Appendix A. Figure 7.5 shows the spectra of these pixels. The spectra are presented in a manner similar to that of Figure 5.5, using the logarithm to accentuate detail and an offset for clarity. In contrast to Figure 5.5, where the spectra



were picked randomly, Figure 7.5 shows spectra that represent the image endmembers. The spectra in these figures look remarkably similar, the most pronounced difference being between the appearance of the fuselage and nose spectra. The important observation is that the mixed pixels of the scene are actually identified as endmembers by virtue of their unique spectral character.

With the purest pixels of the image located, the purest pixels are matched to reference target spectra. Those that match are set aside, and those that do not are used to

form a subspace called a fledge, which can be visualized as a subspace of the  $l$ -simplex which is missing the target endmember. Using the subspace complementary to the target as a projector has the effect of minimizing the effects of background while emphasizing those pixels which contain the target spectrum.

The partial unmixing technique of Boardman, Kruse, and Green (1995) is a viable means of mapping target spectra. The technique is based on determining the intrinsic dimensionality of the data using the MNF or NAPC, determining the spectrally pure endmembers of the scene, identifying target endmembers using a reference library, and projecting the data onto a subspace orthogonal to the background endmembers. As with all of the techniques which use the covariance matrix, the effectiveness of the subsequent projections and transformations is predicated on the goodness of the estimate of the covariance matrix. The ability of the convex set methods to determine the spectrally pure pixels is an issue, though this study has not addressed the mechanics of the convex set methods in detail. The completeness of the reference library is a further factor that affects the accuracy of the target material mapping. Boardman, Kruse, and Green (1995) apply the partial unmixing technique to AVIRIS data with the intent of mapping carbonate minerals in the North Grapevine Mountains of California and Nevada. Figure 7.6 shows their results. The scatter plots represent the optimal projection of the data which includes

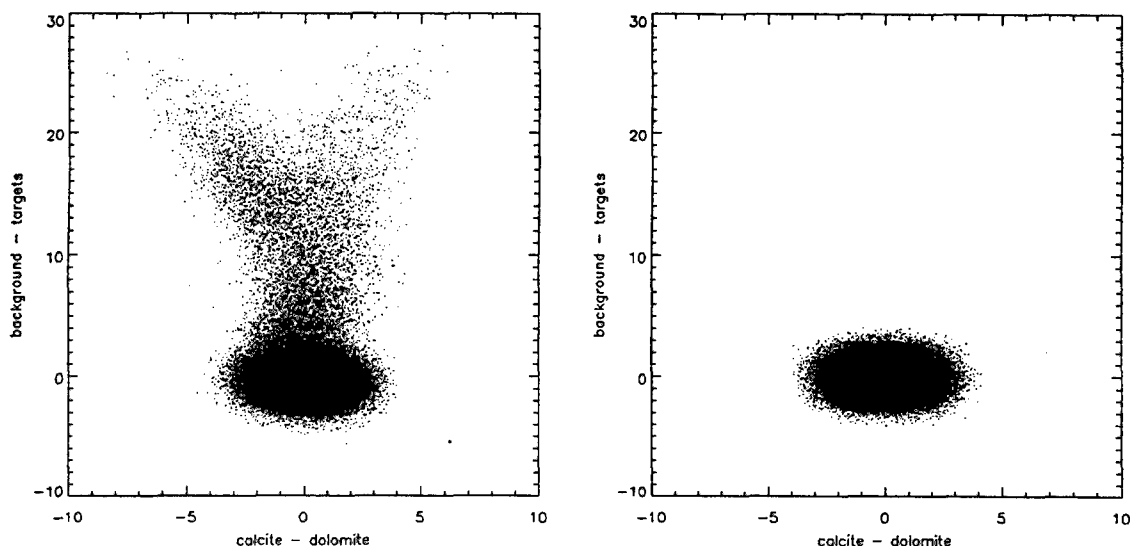


Figure 7.6: Results of Partial Unmixing.  
From Boardman, Kruse, and Green, 1995, p. 26.

the target in the left plot, and the target-free scene in the right plot. The corresponding shape in the center of the two figures shows that the background is being consistently projected by the operator determined using partial unmixing. This implies that the target spectra can be consistently detected in the images.

### 3. Spectral Angle Mapper (SAM)

The SAM technique is a means of determining spectral similarity between a reference spectrum,  $\mathbf{u}$ , and the spectra found at a pixel of the image,  $\mathbf{x}$ . The operation is very similar in concept to the idea of a correlation detector in that an inner product of two  $l$ -dimensional vectors is formed to note the similarity of the vectors to each other in  $l$ -space. The angular difference in radians between two spectra is illustrated by Yuhas, Goetz, and Boardman (1992) as:

$$\cos^{-1}\left(\frac{\mathbf{x} \bullet \mathbf{u}}{\|\mathbf{x}\| \|\mathbf{u}\|}\right) = \cos^{-1}\left(\frac{\sum_{i=1}^l x_i u_i}{\sqrt{\sum_{i=1}^l x_i^2} \sqrt{\sum_{i=1}^l u_i^2}}\right) \quad (7.20)$$

where the Euclidean norms,  $\|\cdot\|$ , provide a normalization of the vectors so that the relative amplitude difference in the two vectors is not a factor (Price, 1994, p. 183). Lower angular values indicate a better match between the reference and test spectra. The application of the SAM method further requires that the vectors to be compared have the same origin in  $l$ -space. This implies that any additive bias induced by instrumental or atmospheric effects must be removed (Yuhas, Goetz, and Boardman, 1992, p. 148). Note that no assumption is made about the compositional nature of the observed spectra. The angular comparison in SAM deals with the gross characteristics of the spectral vectors, and is not concerned with the problem of unmixing the spectrum into constituent endmembers.

Typically, the SAM technique is applied using reference endmember spectra. In this study, the B-52 wing used in the previous chapters is used as a reference endmember. The SAM technique is then applied to the entire Davis Monthan scene. The results of this application are seen in Figure 7.7 and the color version in Appendix A. The pixel values of the image are a measure of the closeness between the observed pixel vector and the reference endmember spectrum. Note how small these values appear. This is an inherent property of hyperspectral imagery and is illustrated in the two-dimensional pixel

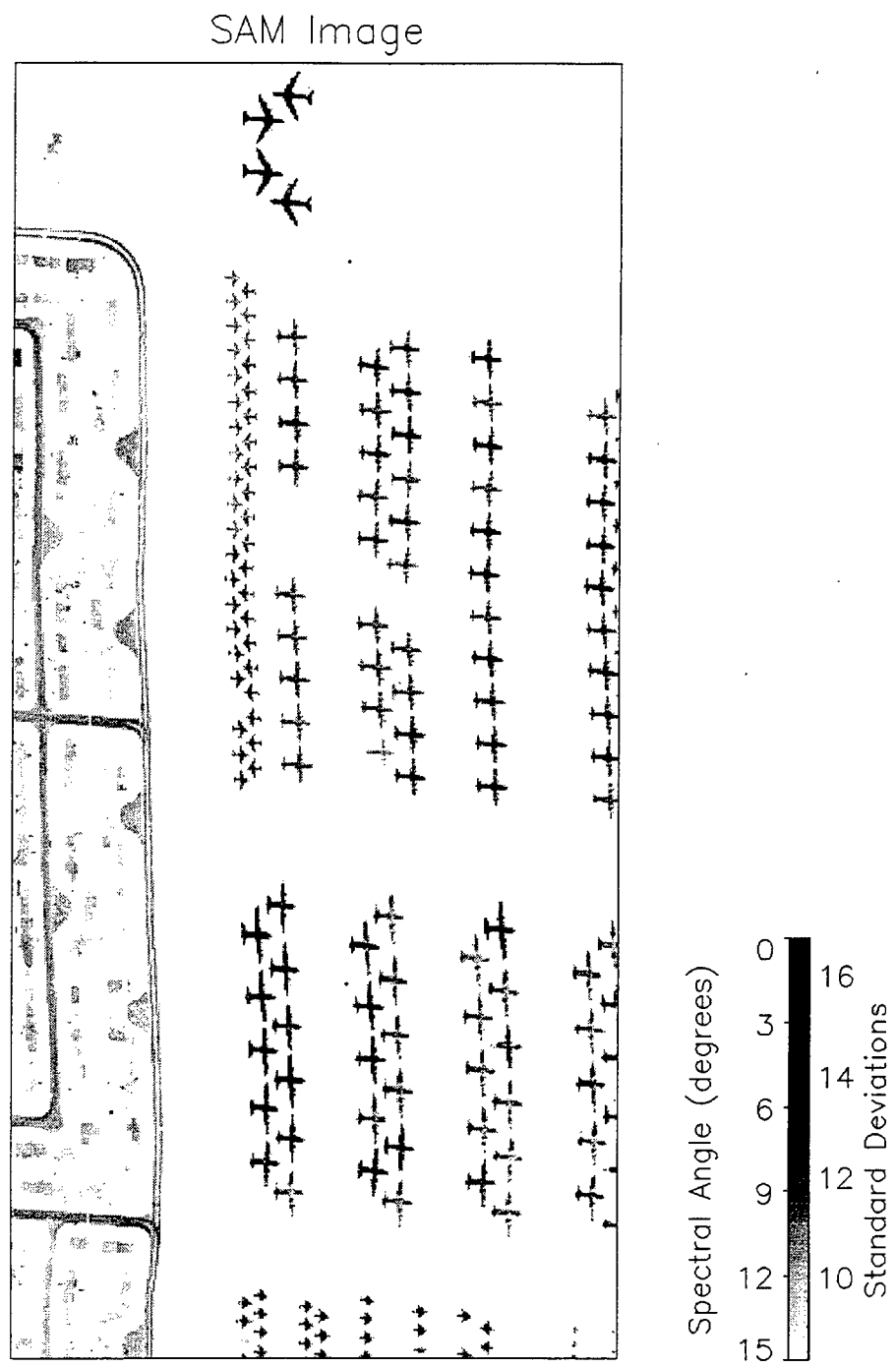


Figure 7.7: SAM Output for Davis Monthan Scene.



vectors of Figure 5.8. The spectral angle in Figure 7.7 is in degrees. Smaller angles indicate a closer degree of fit to the reference spectrum. The SAM output of Figure 7.7 shows that the B-52 aircraft have been accentuated, but that the C-130 aircraft and several buildings have also been assigned high values. The similarity of hyperspectral pixel vectors is the major contributing factor in this figure. The histogram of this image is shown in Figure 7.8. The annotation of the pixel types is the same as that found in the analysis of the LPD and OSP techniques. The x-axis of the plot is in units of degrees, to accentuate the significance of the spectral angle. Note how the target pixel has assumed

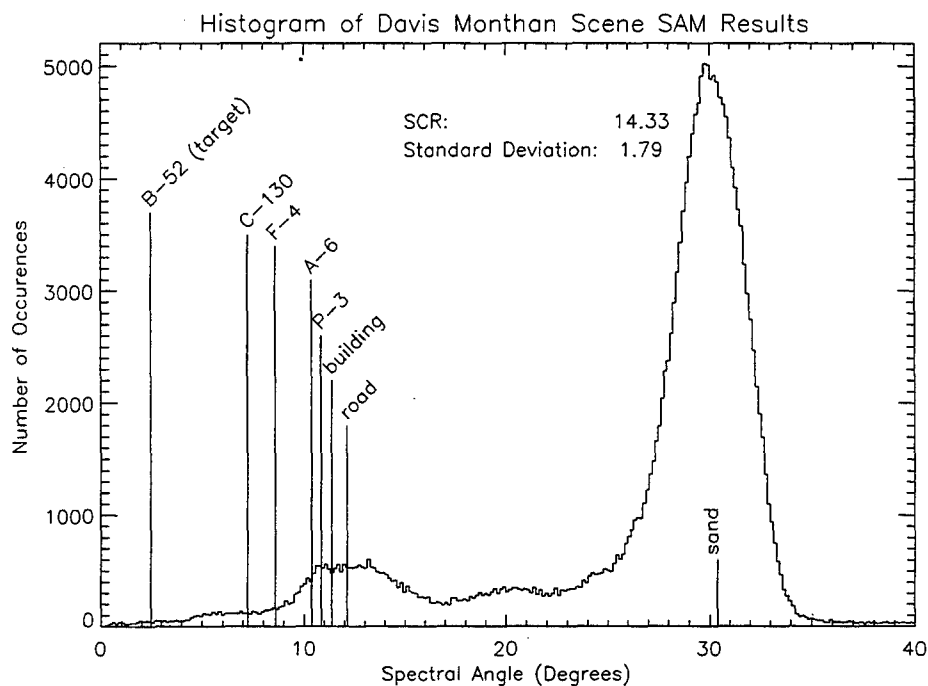


Figure 7.8: Histogram of the Davis Monthan SAM Image.

a value of zero degrees, indicating a perfect match to the reference spectrum. Note also how all of the man-made objects in the scene have been assigned relatively low spectral angles, while the natural background is significantly higher.

The SAM technique is a straightforward way of finding pixels with spectra that are predominately similar on an element-by-element basis to a reference spectrum. It is an innately deterministic approach. The assumptions that it makes are that the spectrum of interest dominates the pixel to such an extent that it will provide a good match with a pure spectrum from a reference library. The deterministic outlook of this approach

overlooks any natural variability that may occur from species to species. The threshold of angular separation which will determine a successful match is the only parameter that can be controlled to alleviate this problem. If the angular separation parameter is made too wide, then the method could be erroneously detect spectra that have been distorted by noise to such an extent that they appear to be targets. Yuhas, Goetz, and Boardman (1992) illustrate the effectiveness of the technique by showing that SAM provides excellent discrimination of endmembers in an AVIRIS scene acquired over the high plains east of Greeley, Colorado. The mixed pixel issue did not seem to confuse the SAM technique in this scene. Perhaps this is because the scene was a relatively simple scene and the objective was not the discrimination of subpixel target spectra. The above observations would imply that the SAM technique is well suited to large scale land use classification.

There is an interesting insight which is gained from the application of SAM to the Davis Monthan scene and sub-scene. We have examined a two-dimensional scatter plot of the Davis Monthan sub-scene in Figure 5.8. In that figure the two-dimensional spectral angle of the target pixels was seen to be different than that of the background. The effect of SAM is to characterize this angle using all 210 spectral dimensions. The result is that SAM gives more accurate assessment of similarity of spectra than can two dimensions. It also provides a means of determining how well the two dimensions represent the inherent nature of 210-band data. Figure B.4 presents the scatter plot of Figure 5.8 with minor changes to the scales of the axes. The color code corresponds to spectral angle produced by the SAM technique. The B-52 wing is used as the reference spectrum. The grouping of colors confirms that the shapes of the scatter plot of bands 50 and 120 accurately represents the actual spectral classes as identified by SAM. The red corresponds to a low spectral angle or a high degree of similarity between that group of pixels and the reference spectrum. Figure B.5 presents the same two bands plotted against each other for the entire scene. The colors and the shape of the scatter plot are not as easily explained in this case. As with Figure B.4, the group of target pixels (red) still appears at the bottom of the distribution, and a strong concentration of background pixels (black) appears at the top of the distribution. The linear shape of the coloration and the blending of colors complicates the interpretation of this figure. The conclusion that may be drawn from this figure is that two bands are not adequate to discriminate the different classes that exist in the entire Davis Monthan scene, whereas they could do so for the sub-scene because of its simplicity.

#### 4. Unmixing Via SVD

The use of the SVD as a tool in the analysis of hyperspectral imagery is introduced by Danaher and O'Mongain (1992) and further elaborated upon by Herries, Seliege, and Danaher (1996). In their problem development, the goal is to develop an algorithm that maps the  $N$ -sample  $\times$   $l$ -band data matrix  $\mathbf{X}^T$  to a vector  $\mathbf{a}$  representing the abundance of an environmental parameter of interest over the image. The data matrix  $\mathbf{X}^T$  consists of  $N$  pixel vectors each of  $l$  bands arranged in the  $N \times l$  matrix:

$$\mathbf{X}^T = \begin{bmatrix} \leftarrow & \mathbf{x}_1^T & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{x}_N^T & \rightarrow \end{bmatrix} \quad (7.21)$$

The goal is to estimate the relative abundance of the target material in the data. This goal is embodied by the environmental parameter vector  $\mathbf{a}$ . The vector  $\mathbf{a}$  is an  $N \times 1$  vector, with each scalar component corresponding the strength of the target material in each row of  $\mathbf{X}^T$ . A model of the data is developed which incorporates the  $N \times l$  matrices corresponding to target component,  $\mathbf{T}$ , background,  $\mathbf{B}$ , and additive noise,  $\mathbf{N}$ . The model appears as:

$$\mathbf{X}^T = \mathbf{T} + \mathbf{B} + \mathbf{N} \quad (7.22)$$

Danaher and O'Mongain (1992) note that the rank of  $\mathbf{X}^T$  is greater than that of  $\mathbf{T}$  or  $\mathbf{B}$ , and that  $\mathbf{T}$  belongs in a subspace of dimension  $e$ , while  $\mathbf{B}$  belongs in a subspace of dimension  $f$  (Danaher and O'Mongain, 1992, p. 1771). For detection of a target, the requirement is that some component of  $\mathbf{T}$ , called the key vector,  $\mathbf{w}$ , be orthogonal to the background subspace. The signal strength is then estimated by finding the component of each row of  $\mathbf{X}^T$  in the direction of  $\mathbf{w}$ . This is formulated mathematically by applying the key vector to the data model and noting that the key vector nulls the background matrix since it is designed to be orthogonal to the background subspace:

$$\hat{\mathbf{a}} = \mathbf{X}^T \mathbf{w} = \mathbf{T} \mathbf{w} + \mathbf{B} \mathbf{w} + \mathbf{N} \mathbf{w} = \mathbf{T} \mathbf{w} + \mathbf{N} \mathbf{w} \quad (7.23)$$

It is very unlikely that one will be able to isolate pure pixels representative of the  $\mathbf{T}$  and  $\mathbf{B}$  matrices. As an alternative approach, ground truth available in one instance is used to develop a key vector using the SVD. This key vector can then be applied to subsequent scenes in which no ground truth is available (Danaher and O'Mongain, 1992, p. 1772). The derivation of the key vector,  $\mathbf{w}$ , begins with the SVD of the data matrix:

$$\mathbf{X}^T = \mathbf{U} \mathbf{S} \mathbf{V} \quad (7.24)$$

The unique nature of the SVD is revealed in the decomposition.  $\mathbf{U}$  is an  $N \times N$  matrix composed of the column eigenvectors of  $\mathbf{X}^T\mathbf{X}$ ,  $\mathbf{V}$  is an  $l \times l$  matrix whose rows are the eigenvectors of  $\mathbf{X}\mathbf{X}^T$ , and  $\mathbf{S}$  is an  $N \times l$  matrix that contains the  $l \times l$  diagonal matrix  $\mathbf{D}$  with elements corresponding to the square roots of the eigenvalues of  $\mathbf{X}\mathbf{X}^T$ . These matrices are represented in their matrix form for clarity:

$$\mathbf{U} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{u}_1 & \cdots & \mathbf{u}_N \\ \downarrow & & \downarrow \end{bmatrix} \quad \mathbf{V}^T = \begin{bmatrix} \leftarrow & \mathbf{v}_1^T & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{v}_l^T & \rightarrow \end{bmatrix} \quad \mathbf{S} = \begin{bmatrix} \mathbf{D} \\ \mathbf{0} \end{bmatrix} \quad \mathbf{D} = \begin{bmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_l} \end{bmatrix} \quad (7.25)$$

The computational burden of working with  $N$ -sized matrices can be overcome by using the economy size SVD in which the rank of the  $\mathbf{S}$  and  $\mathbf{B}$  matrices are used in place of the original number of bands. This smaller number,  $p$ , represents the intrinsic dimensionality of the data. Danaher and O'Mongain (1992) point out that the choice of  $p$  is a compromise between too small a value, which provides insufficient enhancement over the background and too large a value, which makes the process susceptible to noise effects. In the reduced-rank SVD version, the dimensions of the matrix  $\hat{\mathbf{U}}$  are  $N \times p$ , of matrix  $\hat{\mathbf{S}}$  are  $p \times p$ , and of matrix  $\hat{\mathbf{V}}^T$  are  $p \times l$ . The key spectrum is determined by using the SVD and solving for  $\mathbf{w}$  using the initially known ground truth environmental parameter,  $\mathbf{a}$ :

$$\mathbf{a} = \mathbf{X}^T\mathbf{w} = \hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}}^T\mathbf{w} \Rightarrow \mathbf{w} = \hat{\mathbf{V}}^{-1}\hat{\mathbf{S}}^{-1}\hat{\mathbf{U}}^{-1}\mathbf{a} \quad (7.26)$$

This operation is not computationally intensive since the rank has been reduced, the inverses of unitary matrices  $\hat{\mathbf{U}}$  and  $\hat{\mathbf{V}}^T$  are simply their transposes, and the inverse of the diagonal matrix  $\hat{\mathbf{S}}$  is the reciprocal of the diagonal elements. The key vector thus formed is used to estimate the environmental parameter vector in other scenes.

The SVD key vector analysis approach is different from the previous three methods in that it does not rely on reference spectra, but it does require a one-time ground truth image containing target spectra of known abundances from which to develop the key vector. The technique employs the SVD as a computationally efficient means of inverting the unmixing problem to solve for the abundance vector of a target material over the image. Danaher and O'Mongain (1992) test the technique on a 50-band simulation of 100 spectra containing the target material in varying abundances and at various SNRs. Their results show that the target spectra were detected to within 97% agreement with the true abundances. Herries, Selige, and Danaher (1996) apply the SVD key vector analysis technique to spectral imagery of a farm 40 km north of Munich,

Germany. Their focus is on land cover classification. Their results in classifying seven various classes of land cover are in good agreement with ground truth values. The limitations of this approach are the dependence on specific scene ground truth to develop a key vector which is purported to be applicable to different scenes.



## VIII. SUMMARY

The purpose of this study was to organize and present the many existing spectral imagery analysis techniques. This larger context was created by viewing analysis techniques as members of broad strategies. The common themes in the theory and application of the techniques led to the development of a hierarchy of five analysis strategies: 1) linear transformation and projection, 2) classification, 3) linear prediction, 4) optimal band selection, and 5) multiresolution analysis. The elements used to categorize the techniques into strategies were the assumptions made about: 1) the pixel mixing model, 2) the statistical nature of the data, 3) the homogeneity of the scene, and 4) the *a priori* information. Having established a conceptual framework, a number of techniques in each of the five strategies were briefly presented along with pertinent references to more detailed descriptions. A review of historical perspectives and imagery analysis paradigms was given to highlight the unique nature and analysis requirements of hyperspectral imagery.

The focus of this study was on the techniques found within the linear transformation and projection strategy. It was observed that this strategy had many parallels with related ideas from the signal processing community. The *modus operandi* in creating a taxonomy for the techniques within this strategy was the amount of *a priori* information available at the start of the analysis. Four major classes of *a priori* information emerged: 1) none available, 2) complete knowledge of all image endmembers available, 3) knowledge of only the target endmember, and 4) only a reference endmember library or one instance of ground truth available. These categories of available information were used to classify the techniques into four families within the linear transformation and projection strategy.

Before describing the four families of techniques within the linear transformation and projection strategy, some key definitions were addressed in Chapter III. First, the concept of spectral imagery was presented using illustrative multispectral and hyperspectral data sets. The important concept the pixel vector was defined. Second, statistical characteristics of the data were defined and illustrated. Three specific statistical measures of data were discussed: 1) the mean, 2) the covariance matrix, and 3) the correlation matrix. These concepts were graphically depicted using Landsat scatter plots and histograms along with images of the HYDICE covariance and correlation matrices for two different scenes. Third, important concepts from linear algebra and signal

processing were depicted with two-dimensional analogs that could be easily extended to the hundreds of bands found in hyperspectral imagery. These concepts included linear transformations of random vectors, eigenvectors and eigenvalues, unitary transforms, and simultaneous diagonalization of covariance matrices.

Chapter IV discussed the first family of techniques in the linear transformation and projection strategy. These techniques addressed the problem of no *a priori* information about the scene. This family was called the principal components analysis (PCA) family because all of the techniques found in it were based on the important multivariate data analysis method of PCA. Background development for these techniques consisted of exploring different scientific discipline viewpoints of PCA. The multivariate data analysis view derived PCA, the signal processing view saw it in terms of the discrete Karhunen-Loeve transform (DKLT), and the pattern recognition view addressed the criterion of entropy. Three techniques were examined in the context of application to spectral imagery analysis in this chapter. They differ primarily in the weighting given to the variances found in the "raw" data. The basic PCA technique was applied to three different types of hyperspectral scenes. The various facets of the technique were studied using component image appearance; signal-to-noise ratio (SNR) improvement; variance, eigenvalue, and entropy behavior; and eigenvector patterns. The second technique discussed was the maximum noise fraction (MNF), also known as the noise adjusted principle components transform (NAPC). The illustration of the technique was conducted on the HYDICE scene of the Davis Monthan Aerospace Maintenance and Regeneration Center in Tucson, Arizona. The eigenvalues and component images associated with the reversed order MNF, termed the minimum noise fraction transform, were depicted and discussed. The results did not correspond to the expected improvement in image quality over the PCA component images, perhaps due to residual highly structured instrumental noise artifacts. The third technique was the standardized principal components analysis (SPCA) technique. This technique sought to improve on the SNR of the PCA by using the standardized data covariance matrix correlation coefficients. The Davis Monthan scene was analyzed using this technique, and component images were displayed along with eigenvalues, variance, entropy, and eigenvectors. This technique produced an apparent improvement in component image quality over the PCA images. The next step would have been to compare the results of classification schemes based on these transformed data sets, for example, a maximum likelihood classifier.



Chapter V addressed the matched filter family of techniques, which borrowed its name from the similar problem in communications theory. In this family, the assumption was that complete knowledge of image endmembers existed at the start of processing. The background development for this family of techniques began by presenting the concept of linearly additive spatially invariant image sequences, which established the basic model used to pose the linear unmixing problem. Two different models from the theory of least squares were then presented. These were the *a priori* and the *a posteriori* models. Finally, the matched filter was derived using a SNR maximization criterion. Four analysis techniques were described in this chapter. The first was the simultaneous diagonalization (SD) filter. This technique used an optimal filter vector to enhance a desired feature of the image while suppressing undesirable features. It was derived using an output energy ratio and a generalized eigenvalue problem. The results of the original developers of the technique were briefly discussed. The second technique was the orthogonal subspace projection (OSP) technique, which was seen as a special case of the SD filter technique. This technique was derived using the theory of least squares and the optimal matched filter. The OSP technique was applied to the a two-dimensional spectral subset of a small region of the Davis Monthan image. This simplified example was used to show the steps involved in the OSP technique. The original data scatter plot along with the pixel vectors of interest were shown. The effect of the OSP operator on the data was depicted as a two-step operation which first projected the data into a subspace orthogonal to the known background, and then used the known target signal as a matched filter to maximize the signal-to-clutter ratio (SCR). These steps were shown using scatter plots, superimposed one-dimensional subspaces, histograms, and output scalar images. The OSP technique was then applied to the entire scene in an attempt to extract the B-52 aircraft target of interest. The third technique was the least squares orthogonal subspace projection (LSOSP). This technique employed the *a posteriori* least squares model and improved the ability of the OSP technique to distinguish minority spectra from the background. This improvement in SNR was derived. The results of applying this technique to simulated FSS data were briefly discussed. The fourth technique was the filter vector algorithm (FVA). This technique was presented as set of matched filters intended to demix the scene into abundances of the constituent endmembers. The results of applying FVA to data from the PHILLS instrument were briefly discussed.

Chapter VI dealt with the unknown background family of techniques in which the only the target endmember was known. This family relied on an eigendecomposition of

the data covariance matrix to estimate the background endmembers. The statistics of the scene were the important factor in this family of techniques as in the PCA family. The background development of this family addressed two topics which motivated the specific techniques. One was the determination of the intrinsic dimensionality of the data using the minimum description length (MDL) information theoretic criterion. The other was the array processing idea of beamforming, along with a derivation of the minimum variance distortionless response (MVDR) beamformer. There were three techniques in this family. The first was the low probability detection (LPD) technique. The major assumption in this technique involved the minority presence of the target endmember in the scene. The technique was shown to be identical to the OSP technique, with the exception being that the background endmembers were estimated vice known quantities. The determination of the intrinsic dimensionality of the data was attempted using the MDL criterion on the Davis Monthan sub-scene containing four B-52 aircraft. The first step in this process was the noise-whitening of the data for a better MDL estimate. This process was demonstrated with intermediate steps shown in the covariance matrices. The MDL criterion was then applied to the noise-whitened data. The results of this application did not reveal the expected minimum value in the MDL criterion. The observed behavior was monotonically decreasing MDL values. The intrinsic dimensionality was assumed to be less than ten for the remainder of the Chapter. The LPD technique was applied to the sub-scene, and the resulting image and histogram showed a good SCR and clearly distinguishable targets. The LPD technique was then applied to the entire Davis Monthan scene using the first eigenvector to form the LPD classification operator in one case and the first five eigenvectors in another case. The results showed that the operator using the first eigenvector produced a higher SCR than that constructed using the first five. Further examination of the LPD projector matrices and LPD classification operators revealed the differences caused by varying the number of eigenvectors used to estimate the background effects. These results did not correspond to the expected result of an optimum SCR when the number of eigenvectors equaled the intrinsic dimensionality. Comparison with LPD applied to the Aberdeen scene revealed that the discrepancy may have been caused by the failure of the targets used in the Davis Monthan scene to meet the minority pixel requirement of LPD. The second technique in this family was the constrained energy minimization (CEM) technique. The CEM operator was derived using the signal processing version of the correlation matrix, and the subsequent application to data sets showed that the same results occurred for the correlation and

covariance matrices. The CEM operator was applied to the Davis-Monthan B-52 sub-scene, producing a higher SCR than OSP or LPD. The CEM operator was then applied to the entire scene using two different target spectra. The case of the P-3 aircraft target spectrum produced a good SCR and clearly differentiated the P-3 aircraft in the image with a minor number of false alarms. The case of the B-52 aircraft target spectrum did not produce the same clean image results as the P-3 case, and the SCR was smaller by an order of magnitude. The CEM operators formed by using three different types of target aircraft provided insight into the behavior of the above two cases. The third technique in this family was that of the adaptive multidimensional matched filter. This technique was derived from a hypothesis test approach, and assumed that the target spectrum and spatial shape were known. Results of application of this technique to TIMS data were briefly discussed.

Chapter VII detailed the techniques in the family which assumed only knowledge of a reference spectral library or one instance of ground truth. The background development for this family consisted of four different areas. The first area described the multiple signal classification (MUSIC) method of direction-of-arrival estimation in array processing. The second area discussed convex set theory. The third area related the signal processing concept of the correlation detector. The fourth area described the linear algebra tool of the singular value decomposition (SVD) in terms of vector subspaces. The four techniques found in this family were presented and related with the four background areas. The first technique was the MUSIC-based endmember identification. An endmember identification operator was derived using the noise-whitened covariance matrix and a projection operator similar to that used in the LPD technique. The result of applying this technique to simulated AVIRIS data were briefly discussed. The second technique was that of partial unmixing. This technique was shown to be connected with eigenanalysis. It consisted of locating the pixels in the image that most nearly represented the pure endmembers of the scene. This portion of the partial unmixing technique was applied to the Davis Monthan B-52 aircraft sub-scene in an effort to identify the pixels most resembling endmembers. The four purest pixel vectors were located and depicted as spectra. The remainder of the partial unmixing technique consisted of matching the pure pixel vectors with reference library spectra constructing a projector that minimized the effects of the background. Results of this technique were briefly shown using AVIRIS data. The third technique in this family was the spectral angle mapper (SAM). The SAM technique was defined and then applied to the entire Davis Monthan scene. The

endmember for the B-52 aircraft wing was used as the reference pixel vector with which to compare all other scene pixel vectors. The result was shown as an image and a histogram, and had a good SCR, though several objects such as buildings were more prominent than the target. The fourth technique was unmixing via SVD. A key vector based on one instance of known ground truth target abundance was derived. The economy sized SVD corresponding to the intrinsic dimensionality of the data was constructed and allowed for the solution of the linear equation representing the model of the observations.

## IX. CONCLUSIONS

A broad range of techniques were reviewed and characterized from a signal processing perspective. Ideas from the signal processing community are well suited to hyperspectral imagery analysis because of the data dimensionality and the challenge posed by mixed pixels.

Aside from the wide perspective offered by creating a hierarchy of strategies, the intent of this study has been to closely examine the theory, operation, and results of the linear transformation and projection strategy. This enables the user to clearly understand the available tools in spectral imagery analysis. The linear transform and projection strategy parallels the signal processing problem of multiple signal identification. It relies heavily upon the unitary transform and the diagonalization of the data covariance matrix. The linear transform and projection strategy was divided according to *a priori* knowledge, since this is how the image analyst would approach the problem of target detection.

The PCA family of techniques is exploratory in its nature in that nothing is known about the scene. Insight is gained from the theory and results of PCA because this technique serves as the cornerstone for the majority of other spectral analysis techniques. The application of PCA to spectral images began in the early 1970's with the advent of airborne and satellite multispectral sensors. Since the introduction of multispectral remote sensors, the trend has been towards higher spectral resolution and more bands. The commensurate increase in the amount of data motivated the PCA from a data compression viewpoint. The optimal representation properties of the DKLT make the PCA an attractive compression scheme for data transmission and storage. The other major incentive for use of PCA in multispectral imagery analysis was in environmental, agricultural, and geologic quantitative studies. By its very nature, multispectral data has relatively low spatial and spectral resolution, and lends itself well to large ground area analysis. Since there are only a small number of bands (less than ten) to work with, PCA is a good technique to simplify the process of classifying pixels into groups with similar spectral characteristics. In this context, PCA is very similar to the feature extraction application in pattern recognition, the difference being that pattern recognition is a spatial two-dimensional problem, while multispectral images include the third spectral dimension. The application of PCA to hyperspectral imagery has traditionally been viewed from the same classification problem perspective as multispectral imagery.

PCA is a straightforward means of describing the variance in spectral imagery. The traditional application to multispectral imagery has the objective of separating the spectral classes within the data set to make classification more accurate. This separation is entirely based on the second order statistics of the data. Applied in the traditional sense to hyperspectral imagery, it is computationally expensive and has no clearly interpretable results since the orthogonal axes of statistical variance do not have a well defined equivalence in physical terms of observed spectra. The principal components transform does not emphasize a particular class of spectra or individual spectral signatures. The inherent strength of PCA as a multivariate data analysis technique should not be diminished by the shortcomings of attempting to use PCA in a traditional sense for hyperspectral imagery. Rather, it would seem logical to search for modifications to the traditional PCA application that would be better suited to hyperspectral imagery. Specific examples of such modifications are the noise adjusted principal components transform and the use of standardized principal components. These seek to improve SNR in the component images by using second order statistics matrices which have been modified to mitigate the effects of additive noise. The MNF has used a whitening transform to account for the additive noise. The SPCA has standardized the second order statistics so that variance magnitude is no longer the important factor.

An appreciation for the behavior of PCA in generic types of scenes such as urban or rural is required. This appreciation is gained by following the eigenvector, eigenvalue, and entropy behavior of the data covariance matrix. Otherwise, the effects of PCA cannot be totally understood. The PCA family of techniques makes no assumptions regarding the mixed pixel problem. If a target is subpixel in size, the PCA techniques will enhance it only if it is statistically significant throughout the scene. The global nature of the PCA family of techniques makes them more attractive as a preprocessing step than as a means of target detection in and of themselves.

All of the other families of techniques in the linear transform and projection strategy are built around the assumption of a linear model of some sort. The matched filter family assumes complete image endmember knowledge. Orthogonal complement projectors are constructed using the theory of least squares to fill the void that exists in the analogy between signal processing models and spectral imagery. Signal processing models assume orthogonal signals. The step of orthogonalizing the observed pixel vectors is an important one in that it allows a more direct application of matched filtering ideas.

The primary utility of the OSP and LSOSP techniques as enumerated by their authors is in earth remote sensing. The ability to determine the abundance of surface materials is an important facet of geological exploration. In constructing the model of the image, all of the techniques assume that a relatively small number of endmembers comprise the scene. The applicability to the problem of detecting a small target such as a vehicle on a subpixel scale appears to be within the capabilities of these techniques. While this may be true in relatively simple settings, such as a desert environment, the complexity of an urban scene would definitely require a greater number of endmembers to be identified, making the target detection problem more challenging. This is a problem which has not been thoroughly researched, but it would appear that these matched filter family of techniques are a good choice for the application since they assume so much *a priori* knowledge.

The unknown background family of techniques relied on an eigendecomposition to infer the nature of the background. The LPD technique makes the important assumption that the target is a minority element of the scene. The assumption of minority target was tested in the HYDICE Aberdeen and Davis Monthan scenes. The results validate the fact that LPD works optimally in a relatively uniform background when the target is a minority element. The CEM technique provides a more flexible tool than the LPD since the target need not be a minority element of the scene.

The limited endmember family of techniques returns to the ideas found in the PCA family. The key difference between partial unmixing and the traditional multispectral application of PCA is that a model is being evaluated using statistical analysis instead of statistical analysis being used independently on data. In this sense, this technique is closer to Hotelling's (1933) motivation for determining the independent sources of variation within an experiment. This technique's applicability to target detection in hyperspectral imagery analysis is that the mixed pixel problem is considered on a simple level. The mixed pixel problem is a concern when one is searching for a target with a spectrum that is on a subpixel scale. Although reference spectra must be known, the fact that the endmembers can be located by projection onto specified axes of principal variation is an important observation. In one sense, this technique ties physically observable parameters together with the principal components. The innovative approach of the unmixing using the SVD is a means of using all available knowledge of the scene in an efficient manner. The opportunity exists for exploration of the full power of this technique. The SAM technique is a simple and highly effective means of

characterizing pixel vectors using a deterministic view. It performs better than any of the other techniques examined in the task of extracting man-made objects from the scene. The accuracy and applicability of the reference library is the most important consideration.

The interaction of several factors alters the performance of the techniques. These factors must be addressed before application of the technique to a specific problem. One factor is the validity of the model used to construct an optimal detection operator. A model is not appropriate if there is no *a priori* information. A second factor is to consider the type of statistics to be employed. The use of the covariance matrix, its standardized version, or the correlation matrix will yield different eigenvectors and hence a different transform. The implications of these second order statistics should be considered. The type, abundance, and relative proportions of background and target is also a deciding factor in choosing the appropriate technique. The use of radiance or reflectance data is a factor, depending on whether or not the solar effect is desired in the end product. The decision of how to view the data determines the technique and is in large part determined by prior knowledge of the scene. A deterministic view chooses to ignore the natural variability of the data, but is a good decision if extensive reference spectra are available.

Finally, in the test applications conducted in this study, the performance levels of the techniques in the task of target detection were characterized. The OSP technique achieved a good SCR (~2.5), but required extensive *a priori* endmember knowledge. Although the OSP technique would have performed better had the identification of endmembers been more detailed, such an extensive knowledge of image endmembers is rarely available in real applications. The LPD technique demonstrated a powerful solution to the lack of *a priori* knowledge by applying the power of the unitary transform to estimate the contribution of the image background. Results indicate that a high SCR (~11) with LPD is achieved when the target is a minority element of the scene. The SAM technique provided the best differentiation of target from background and the highest SCR (~14) of all techniques evaluated. Its simplicity and small requirement for *a priori* information make it an attractive option as a real-time analysis tool.



## APPENDIX A. COLOR FIGURES

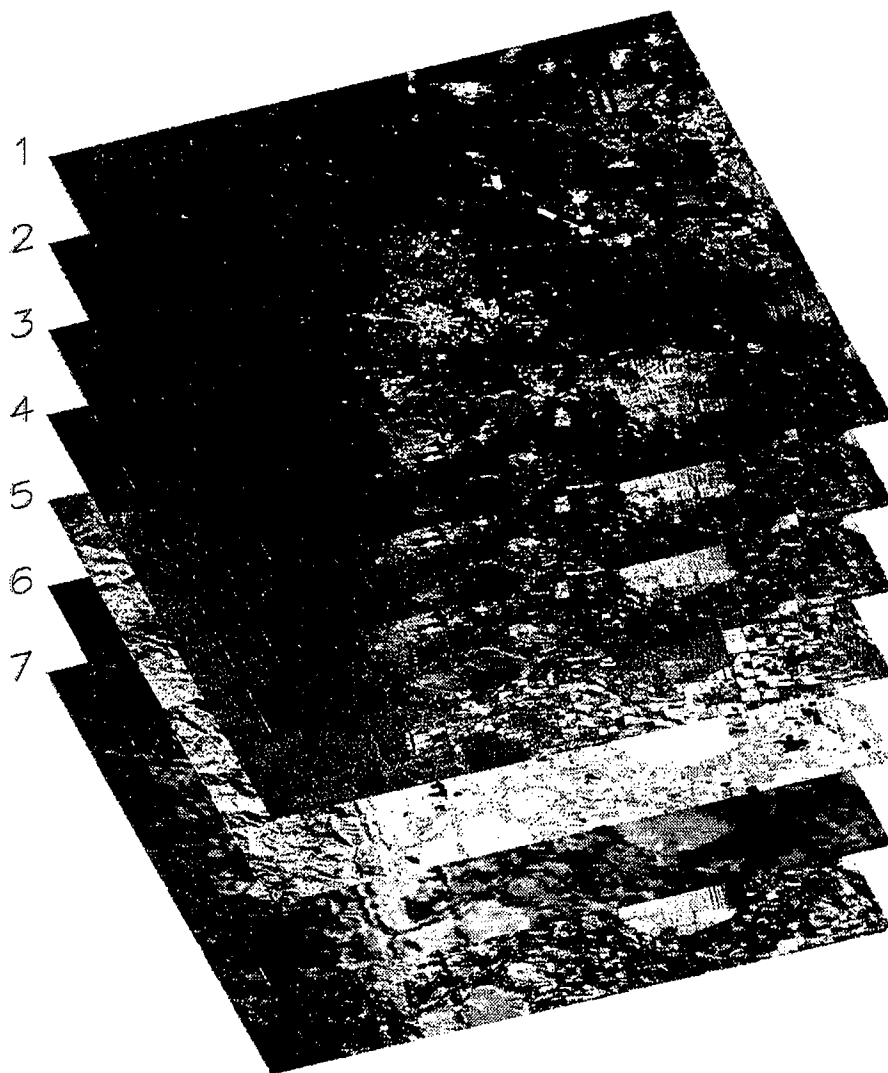


Figure 3.1: A Typical Multispectral Image Produced by Landsat TM.

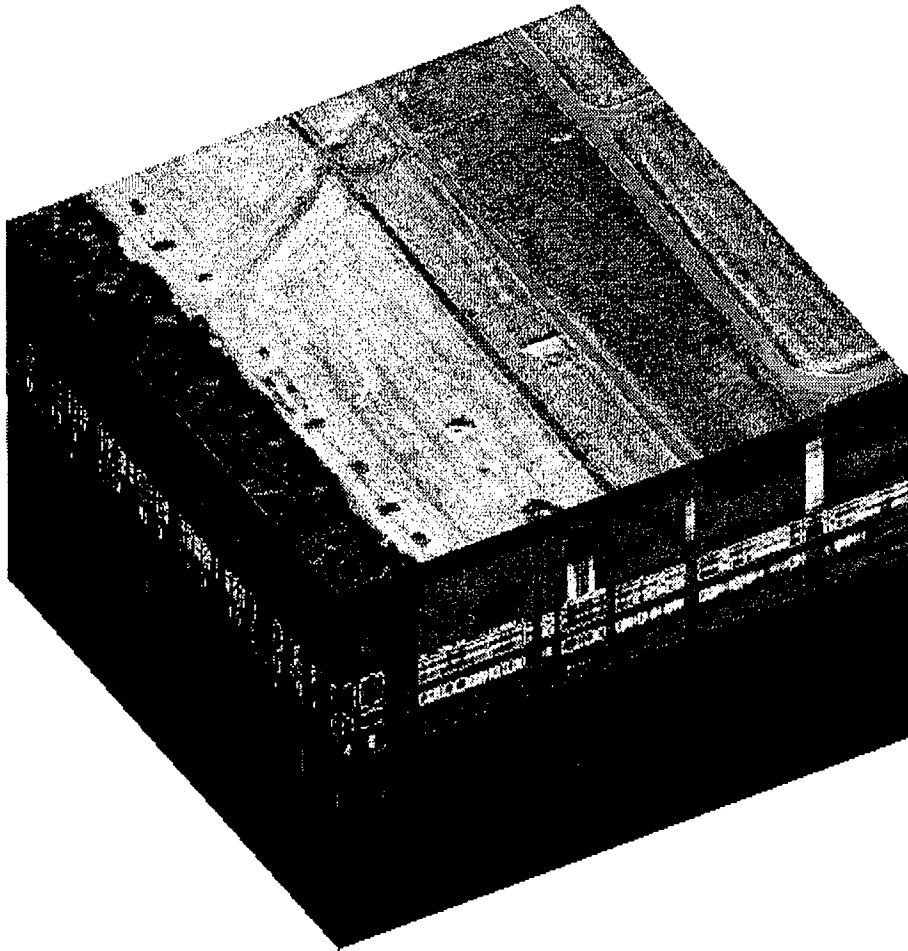


Figure 3.2: A Typical Hyperspectral Image Cube.

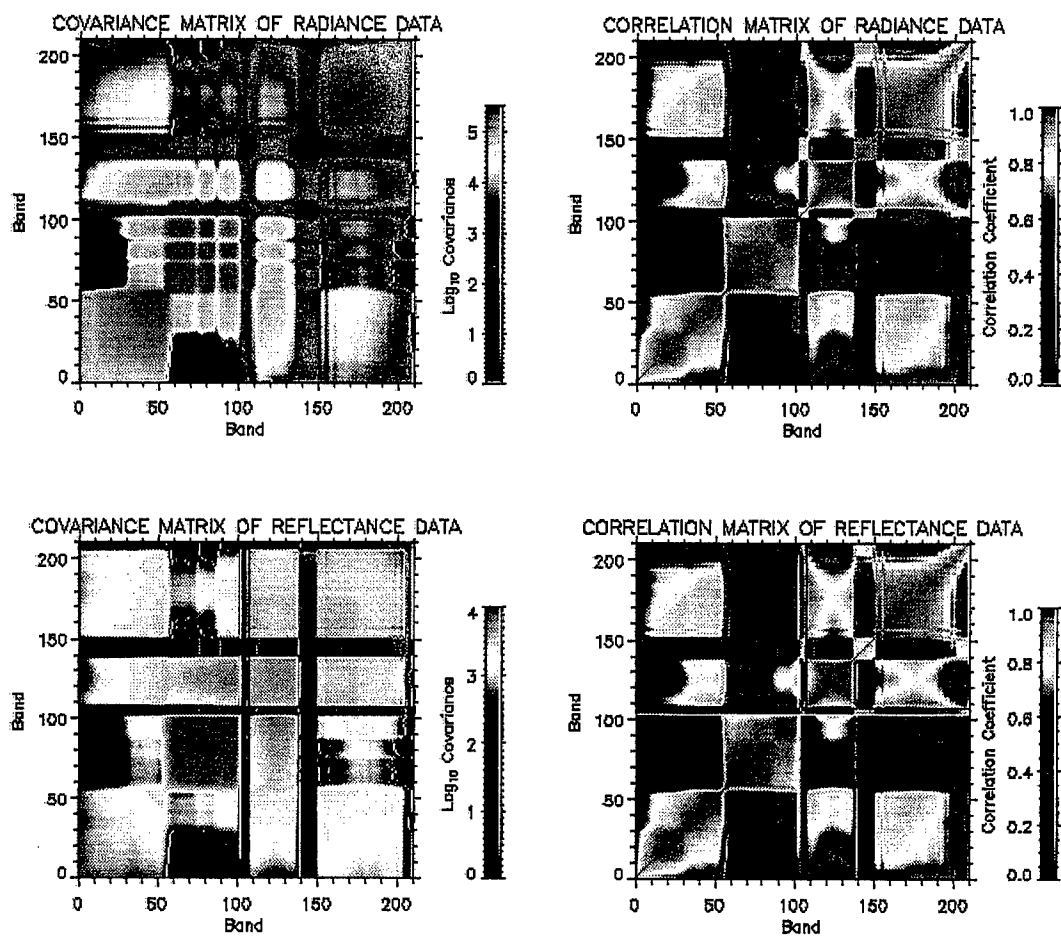


Figure 3.10: Second Order Statistics of the HYDICE Aberdeen Scene.

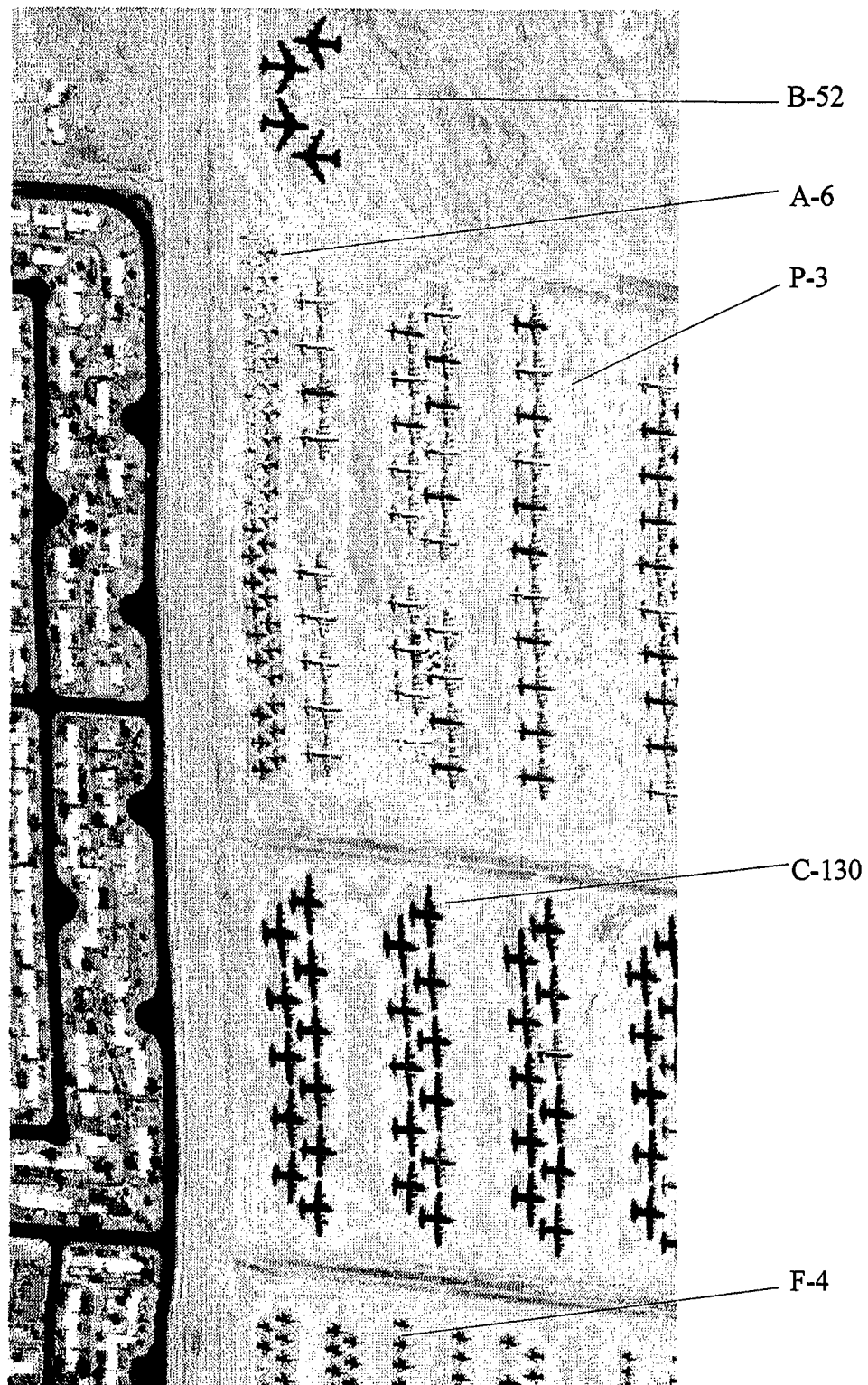


Figure 3.11: HYDICE Scene of Davis Monthan Air Force Base.

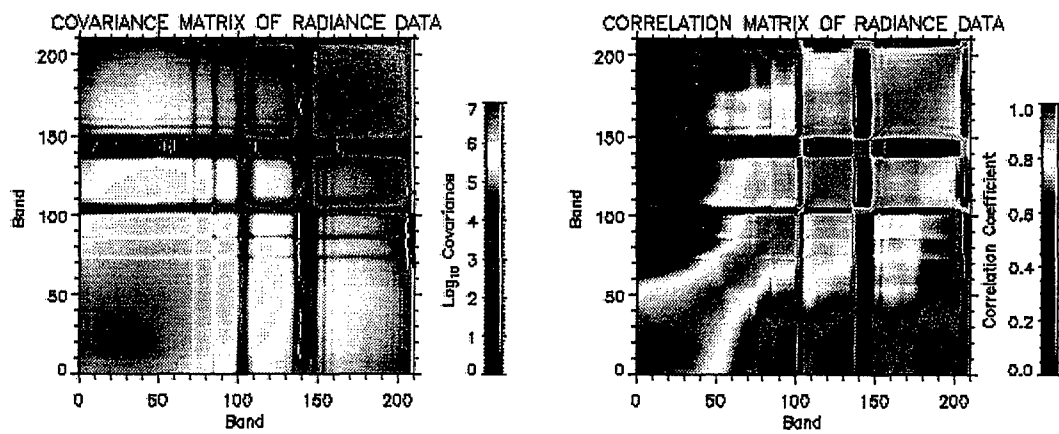


Figure 3.12: Davis- Monthan Radiance Covariance and Correlation Matrices.

25 Principal Component Images using Radiance Covariance — Davis Monthan

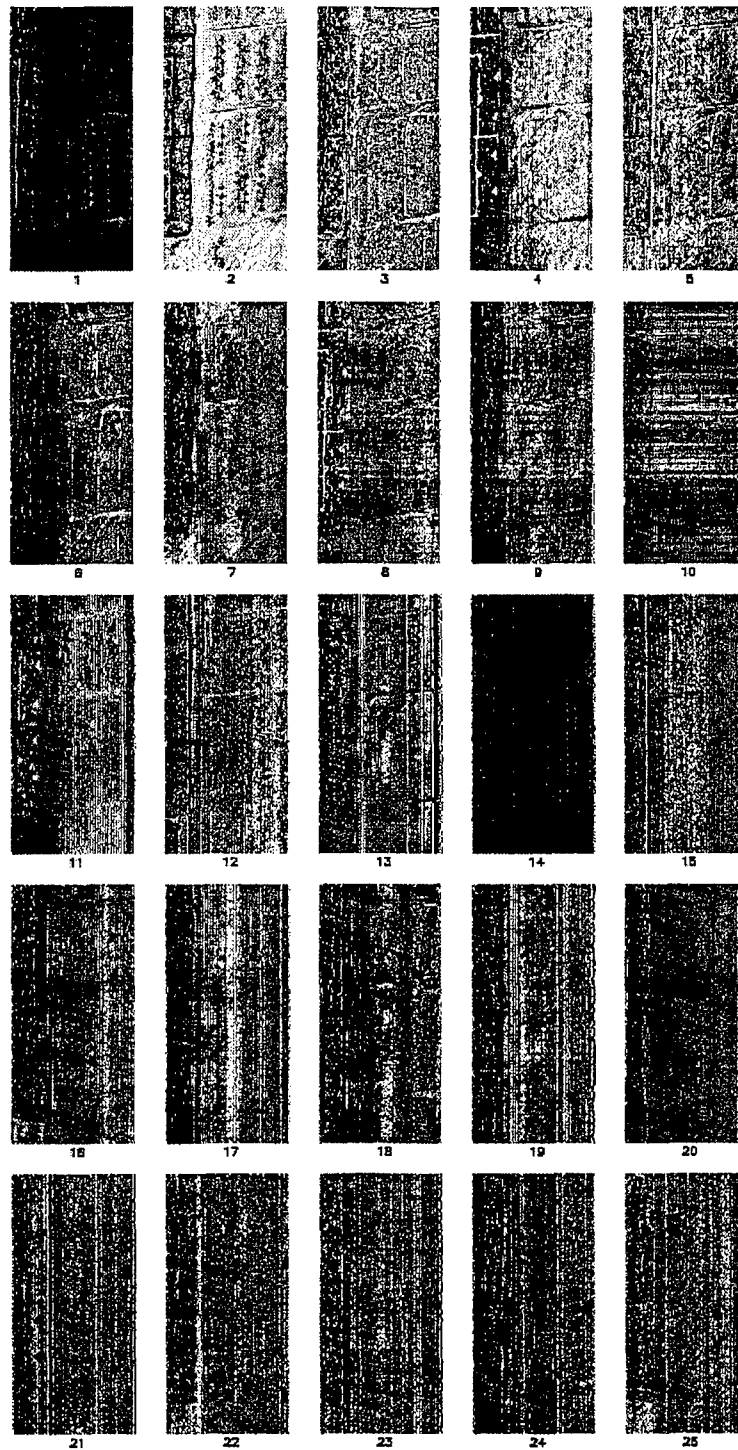
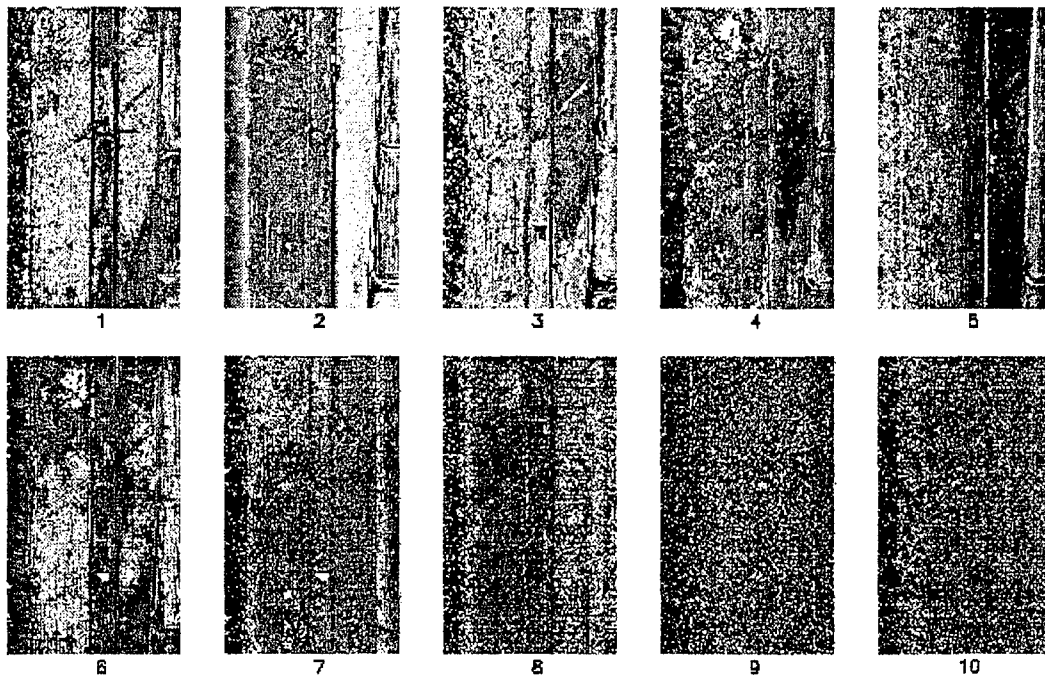


Figure 4.4: First 25 PC Images of Davis Monthan Radiance Scene.

10 Aberdeen Principal Component Images using Radiance Covariance



10 Aberdeen Principal Component Images using Reflectance

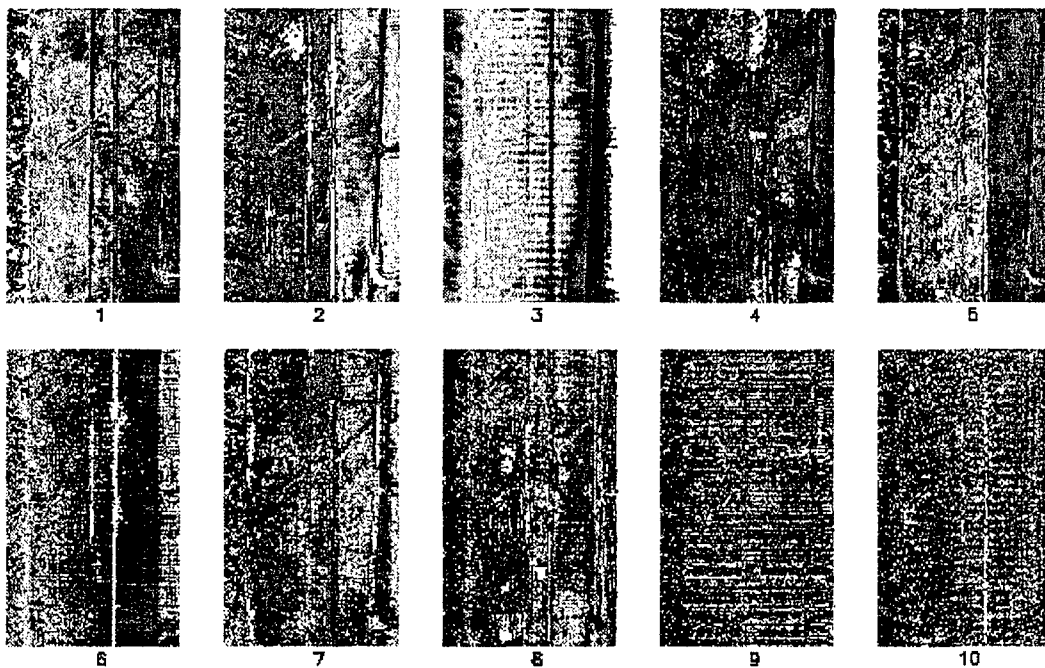


Figure 4.5: First Ten PC Images of Aberdeen Radiance and Reflectance Scenes.

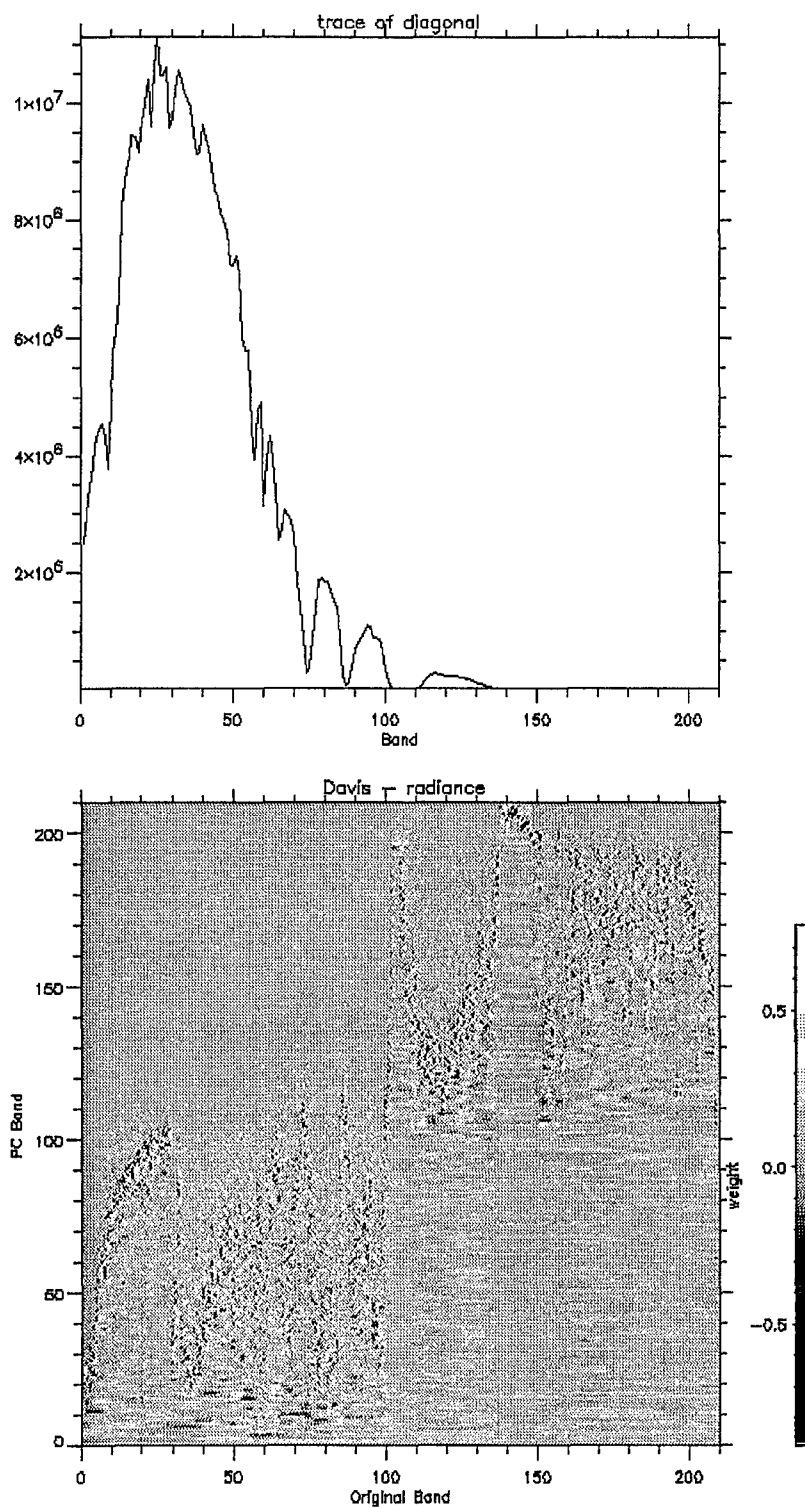


Figure 4.11(a): Eigenvectors and Trace of the Covariance Matrix of Davis Monthan Scene.



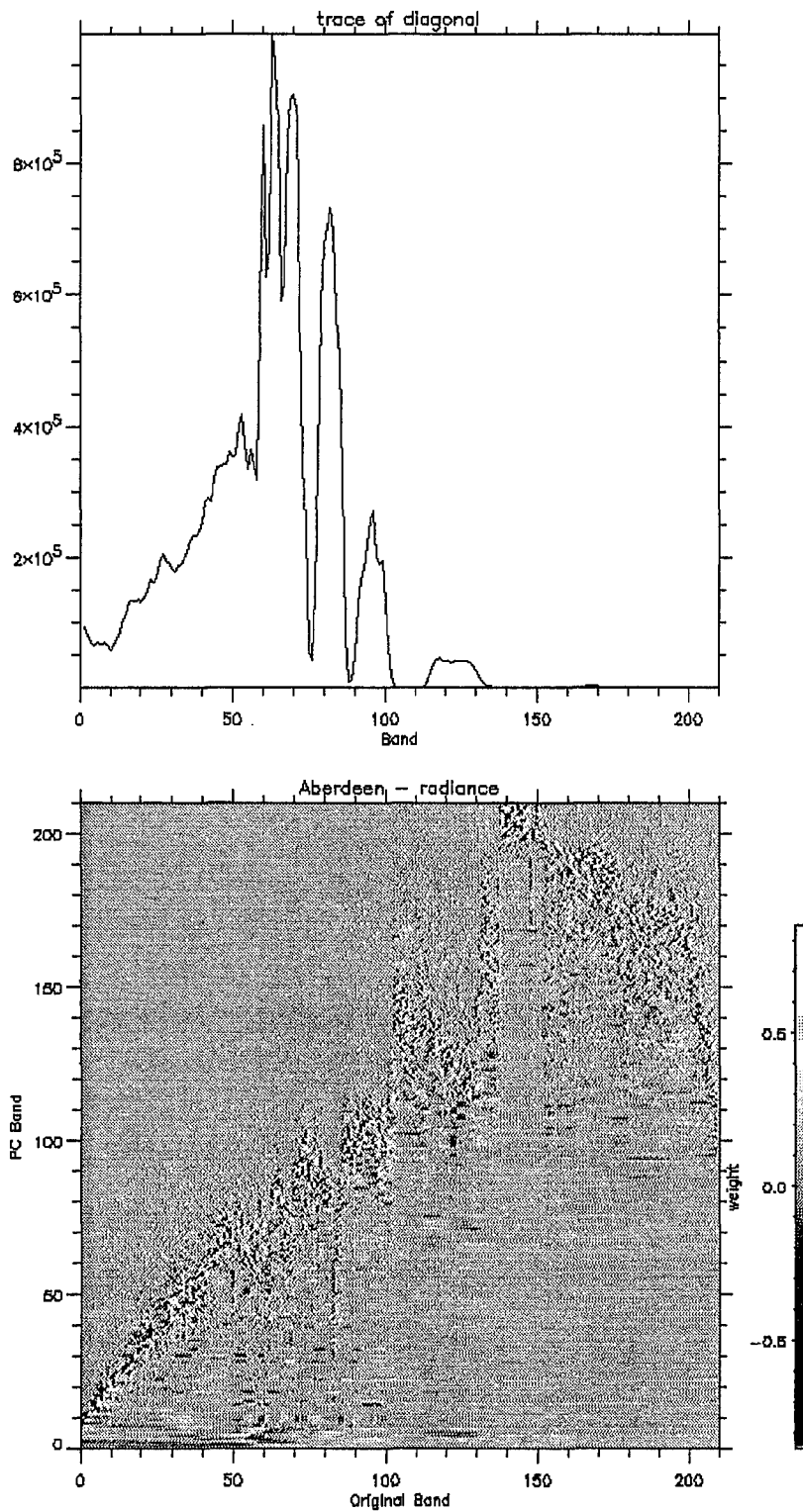


Figure 4.11(b): Eigenvectors and Trace of the Covariance Matrix of Aberdeen Radiance Scene.

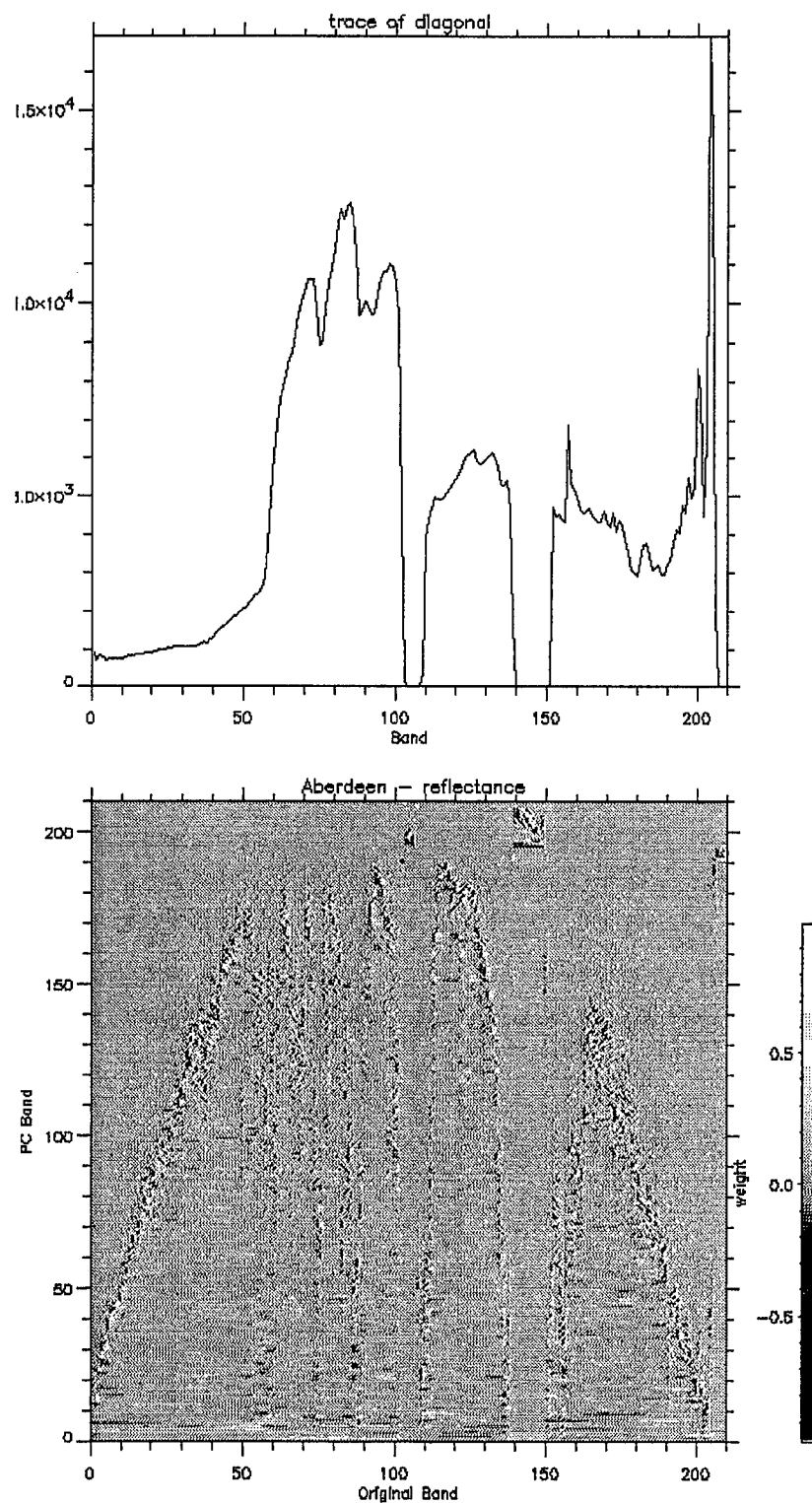


Figure 4.11(c): Eigenvectors and Trace of the Covariance Matrix of Aberdeen Reflectance Scene.

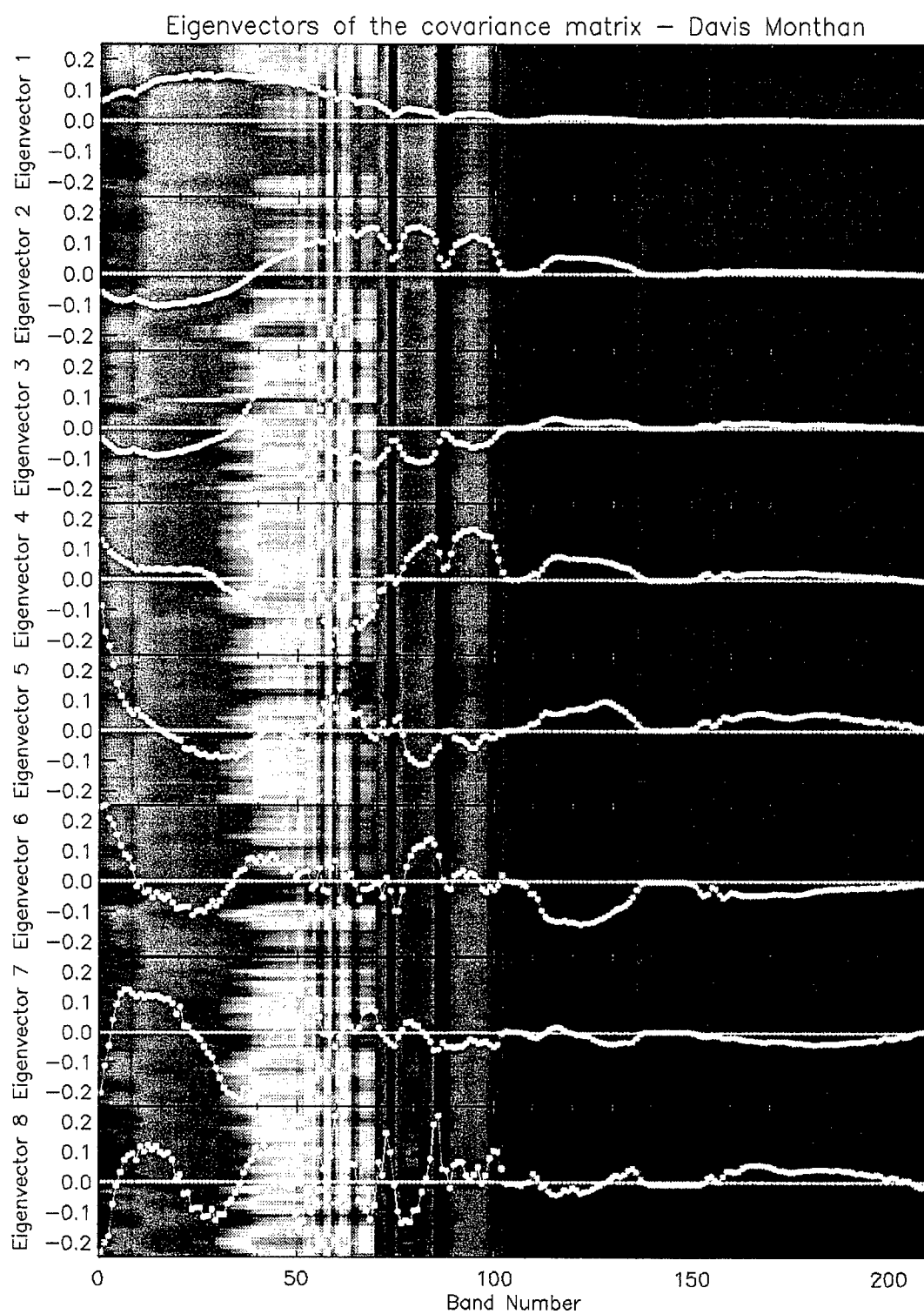


Figure 4.15: First Eight Eigenvectors of Davis Monthan Scene Superimposed on a Random Slice Across the Hypercube.

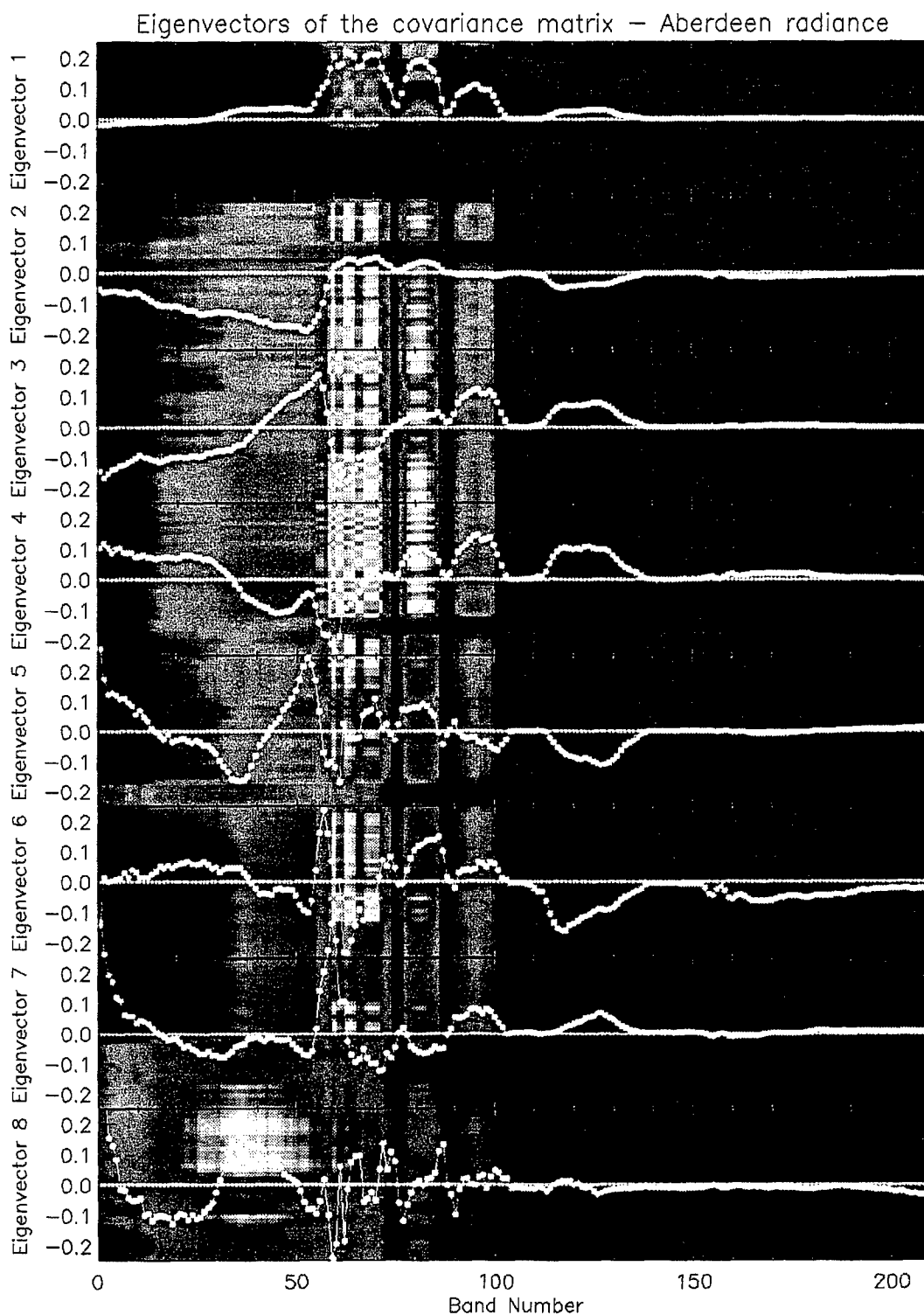


Figure 4.16: First Eight Eigenvectors of Aberdeen Radiance Scene Superimposed on a Random Slice Across the Hypercube.

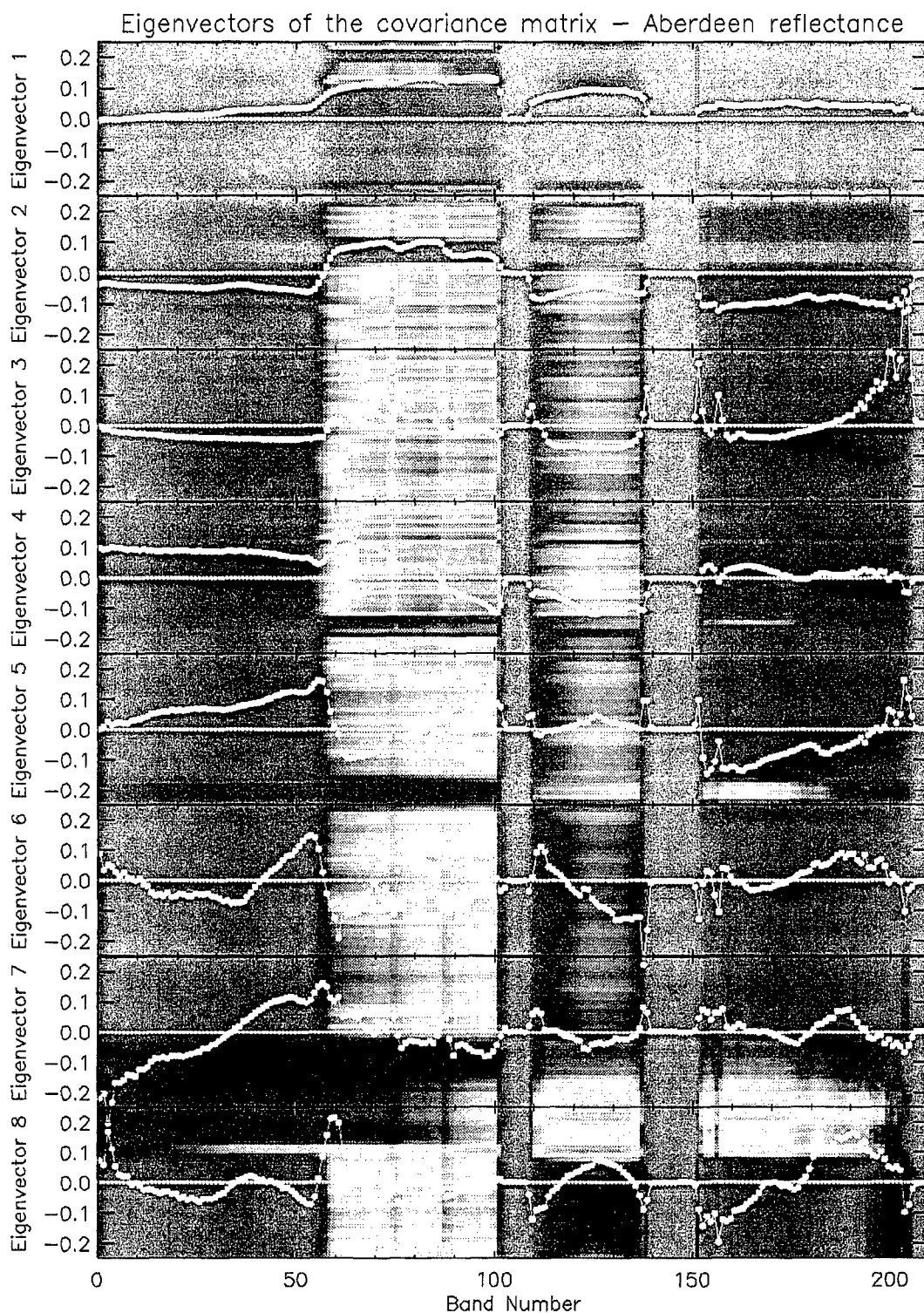


Figure 4.17: First Eight Eigenvectors of Aberdeen Reflectance Scene Superimposed on a Random Slice Across the Hypercube.

25 MNF Images using Radiance Covariance — Davis Monthan

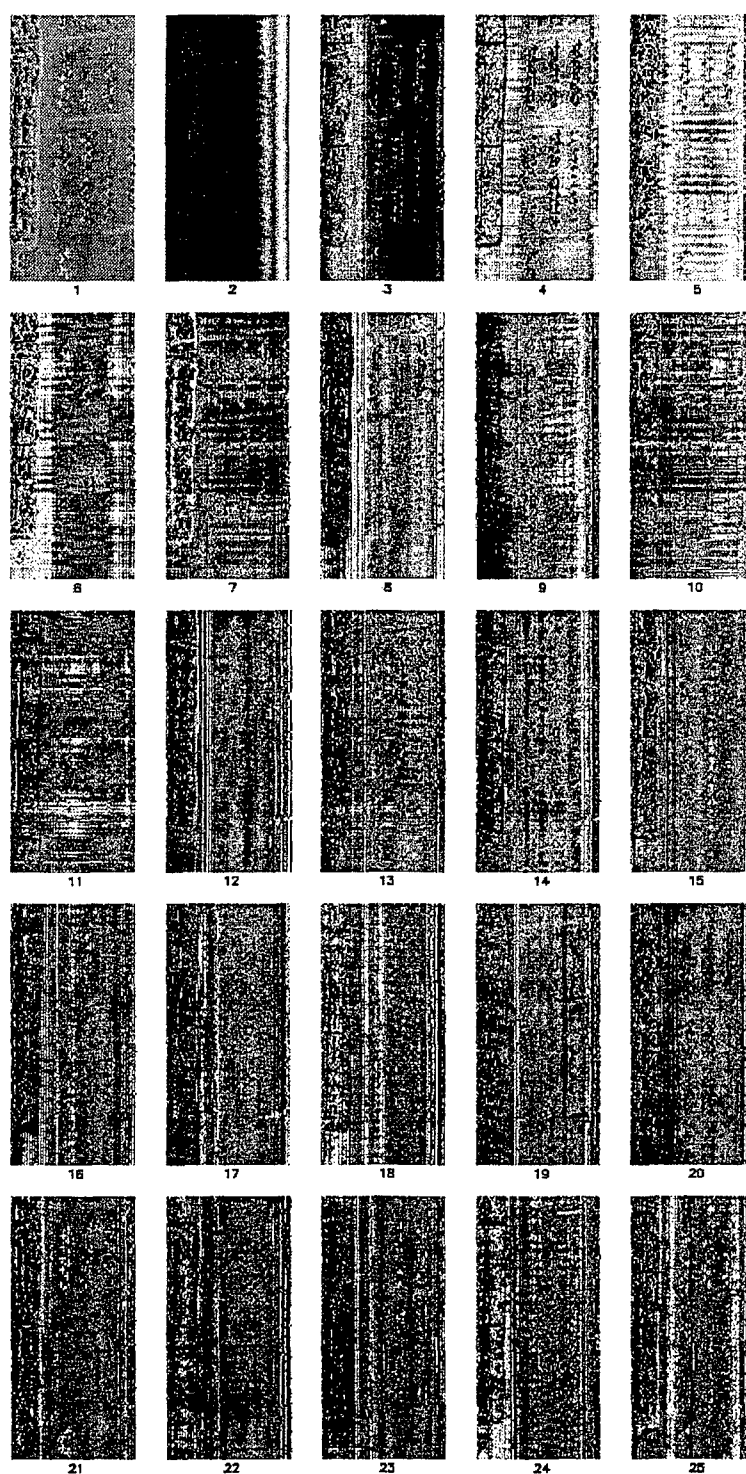


Figure 4.20: First 25 MNF Component Images of the Davis Monthan Scene.

25 standardized PC Images using Radiance Correlation — Davis Monthan

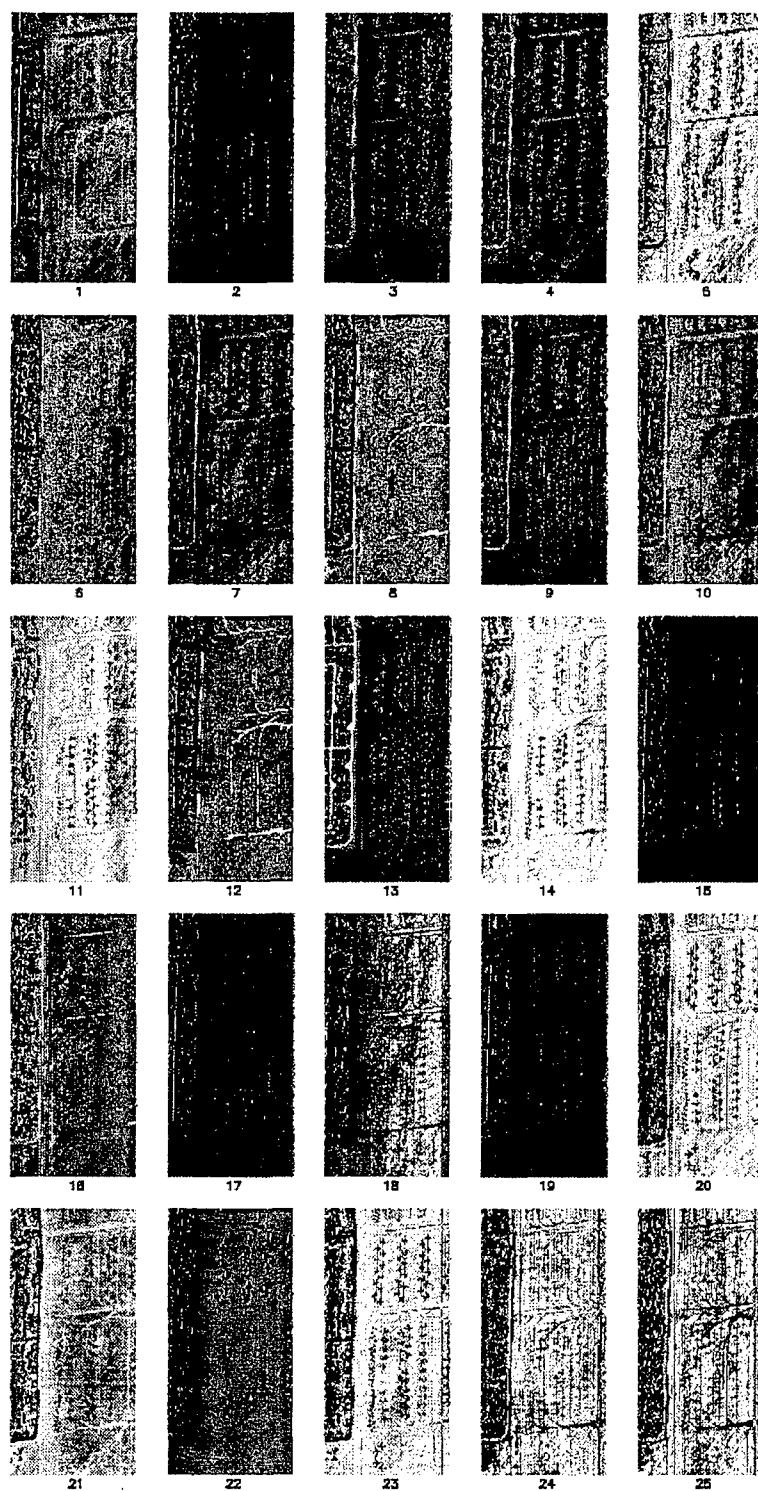


Figure 4.23: First 25 Standardized PC images of the Davis Monthan Scene.

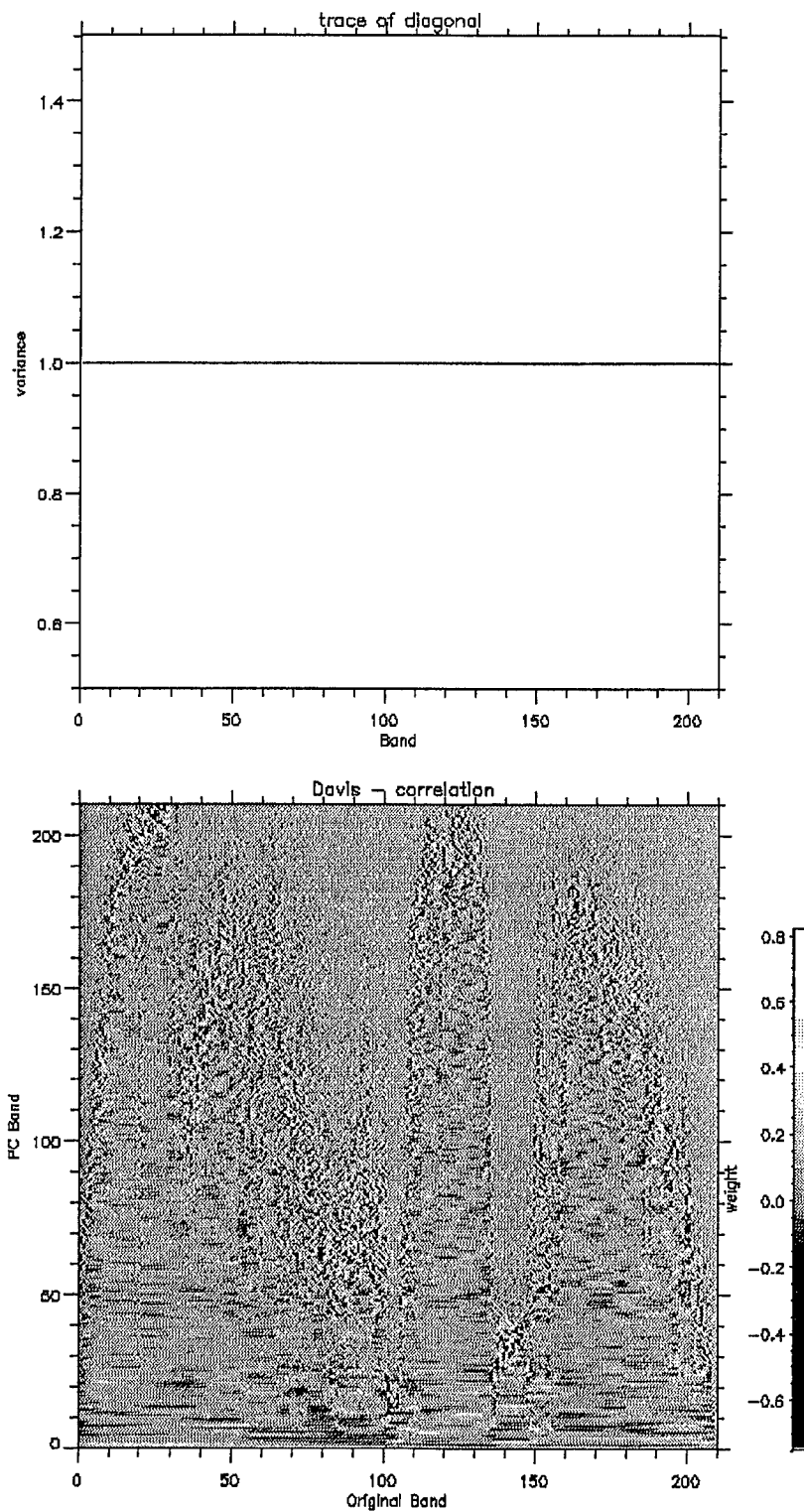


Figure 4.25: Eigenvectors and Trace of the Correlation Matrix of Davis Moon Scene.



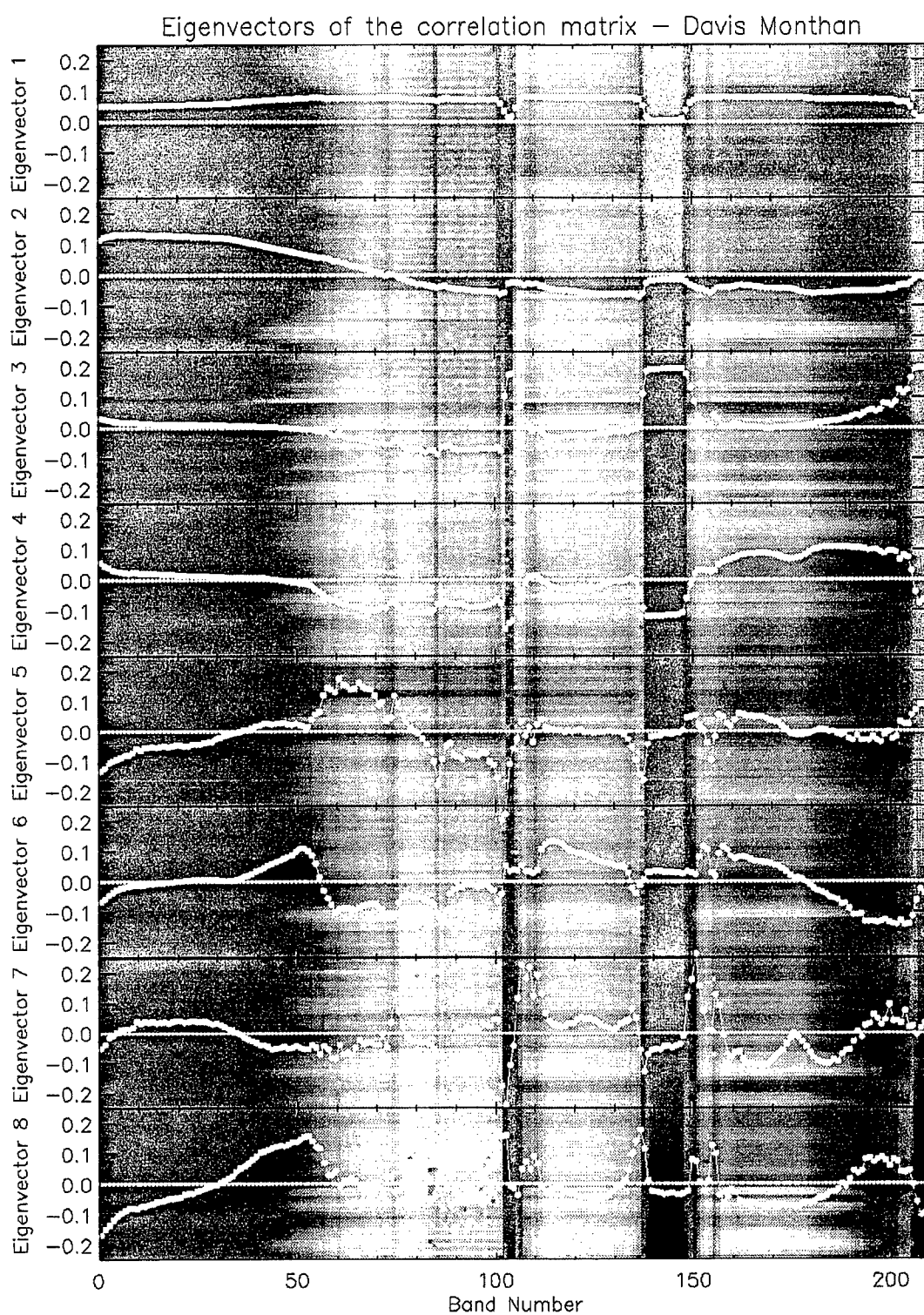


Figure 4.27: First Eight Eigenvectors of Davis Monthan Normalized Scene Superimposed on a Random Slice Across the Hypercube.

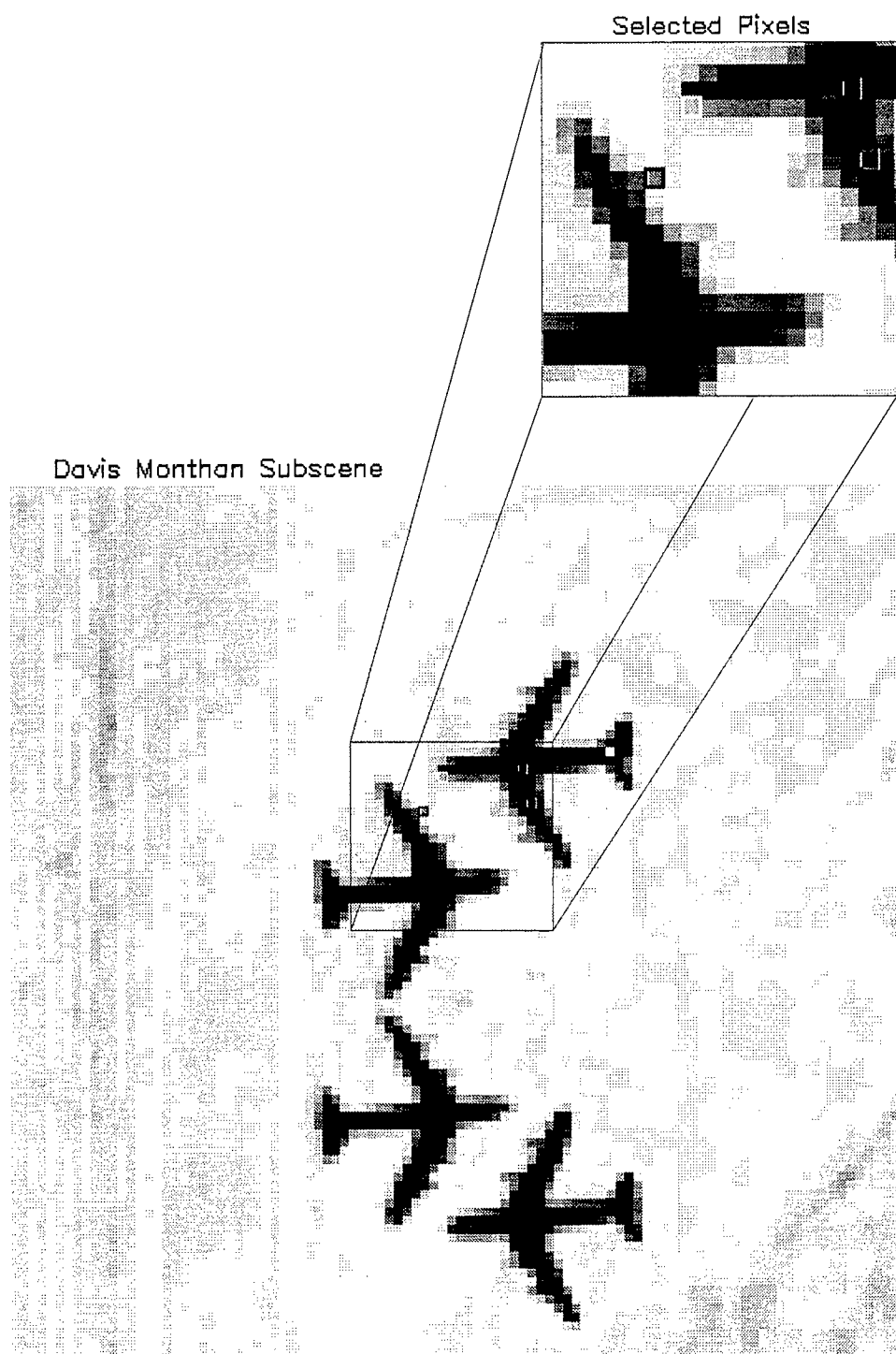


Figure 5.4: Sub-scene of Davis Monthan Image with Boxes Showing Chosen Pixels.

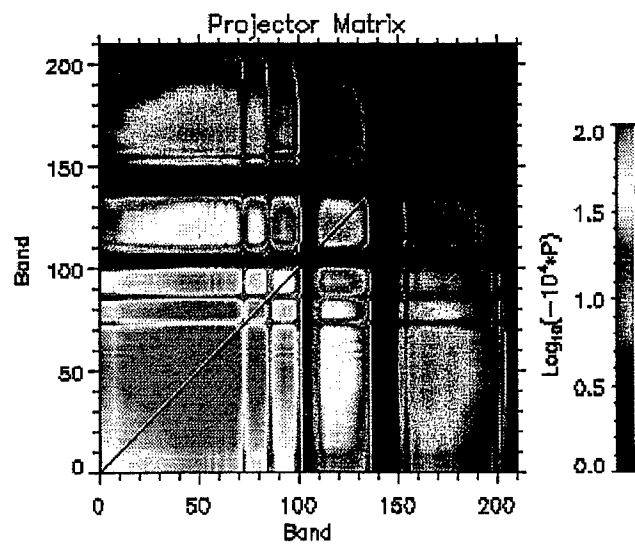


Figure 5.7: The Orthogonal Complement Projector.

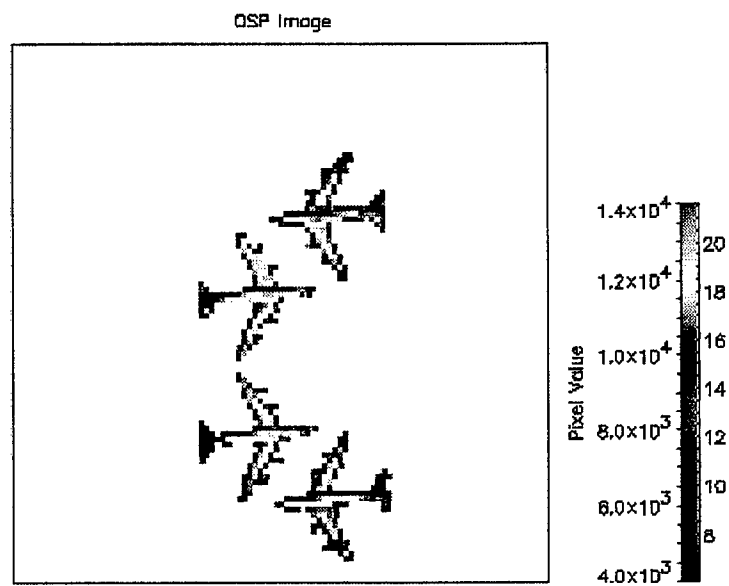


Figure 5.12: Davis Monthan Sub-scene OSP Output Image.

OSP Image

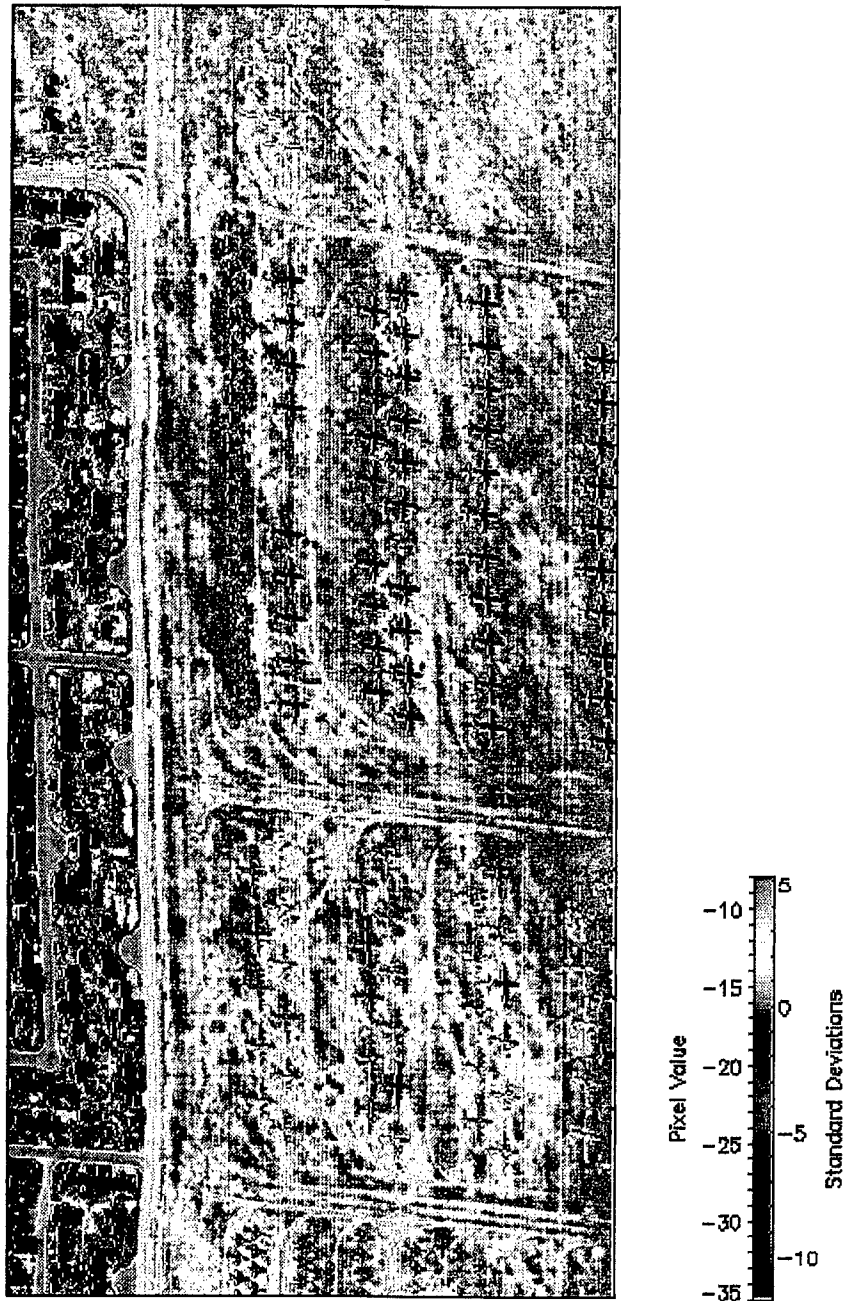


Figure 5.14: Davis Monthan OSP Output Image.

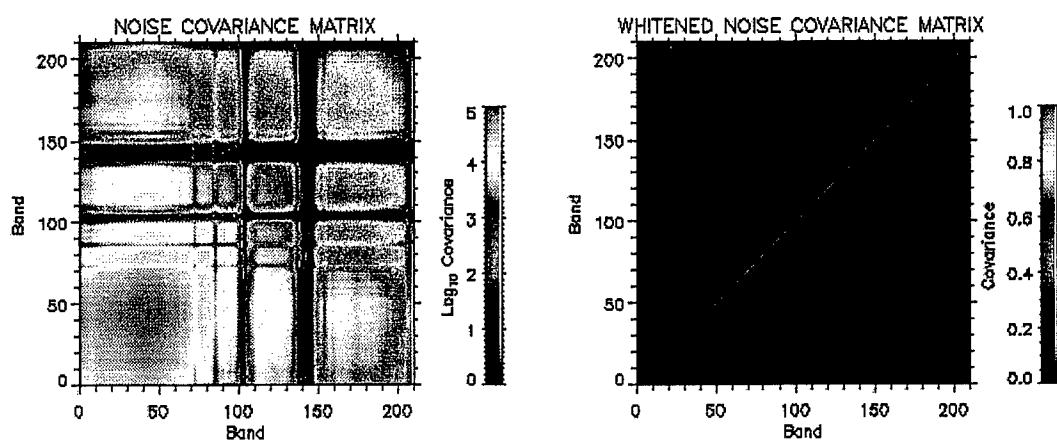


Figure 6.2: Whitening of the Davis Monthan Sub-scene Noise Covariance Matrix.

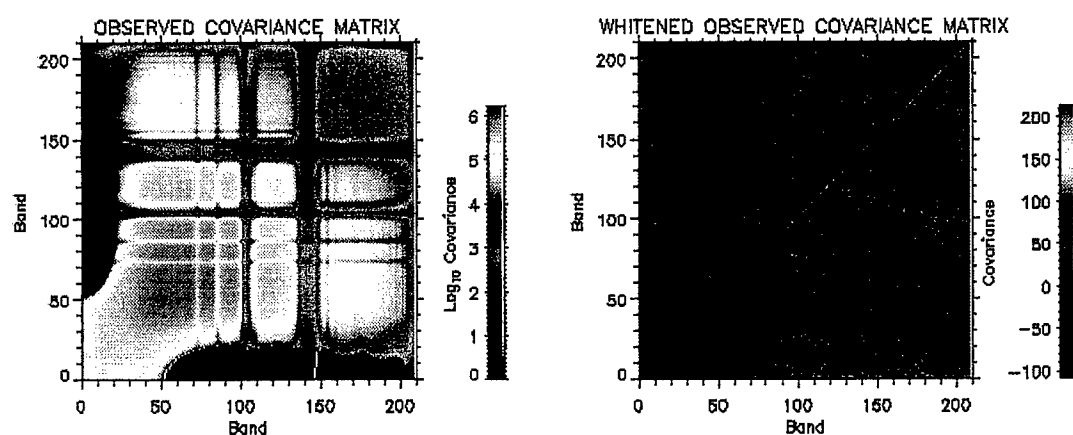


Figure 6.3: Whitening of the Davis Monthan Sub-scene Data Covariance Matrix.

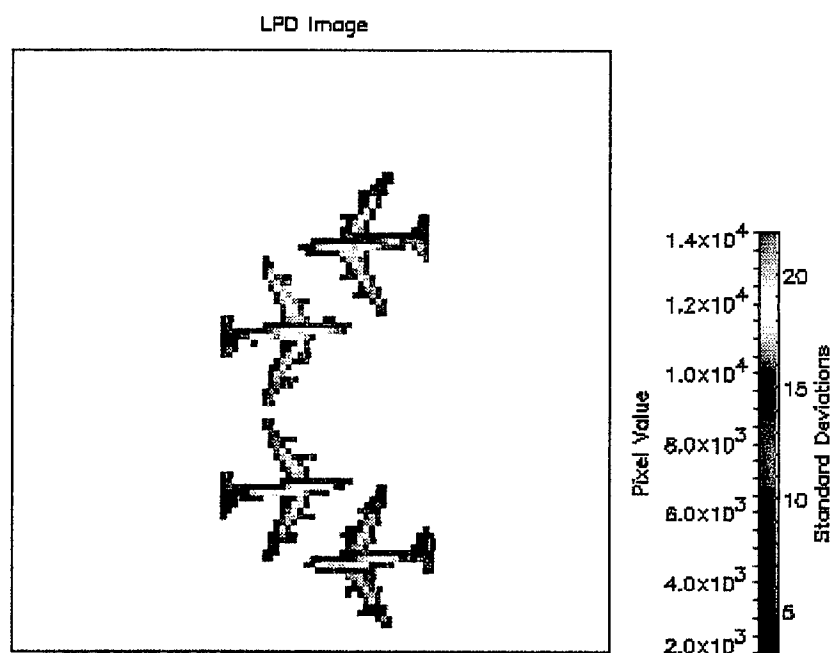


Figure 6.6: Davis Monthan Sub-scene LPD Output

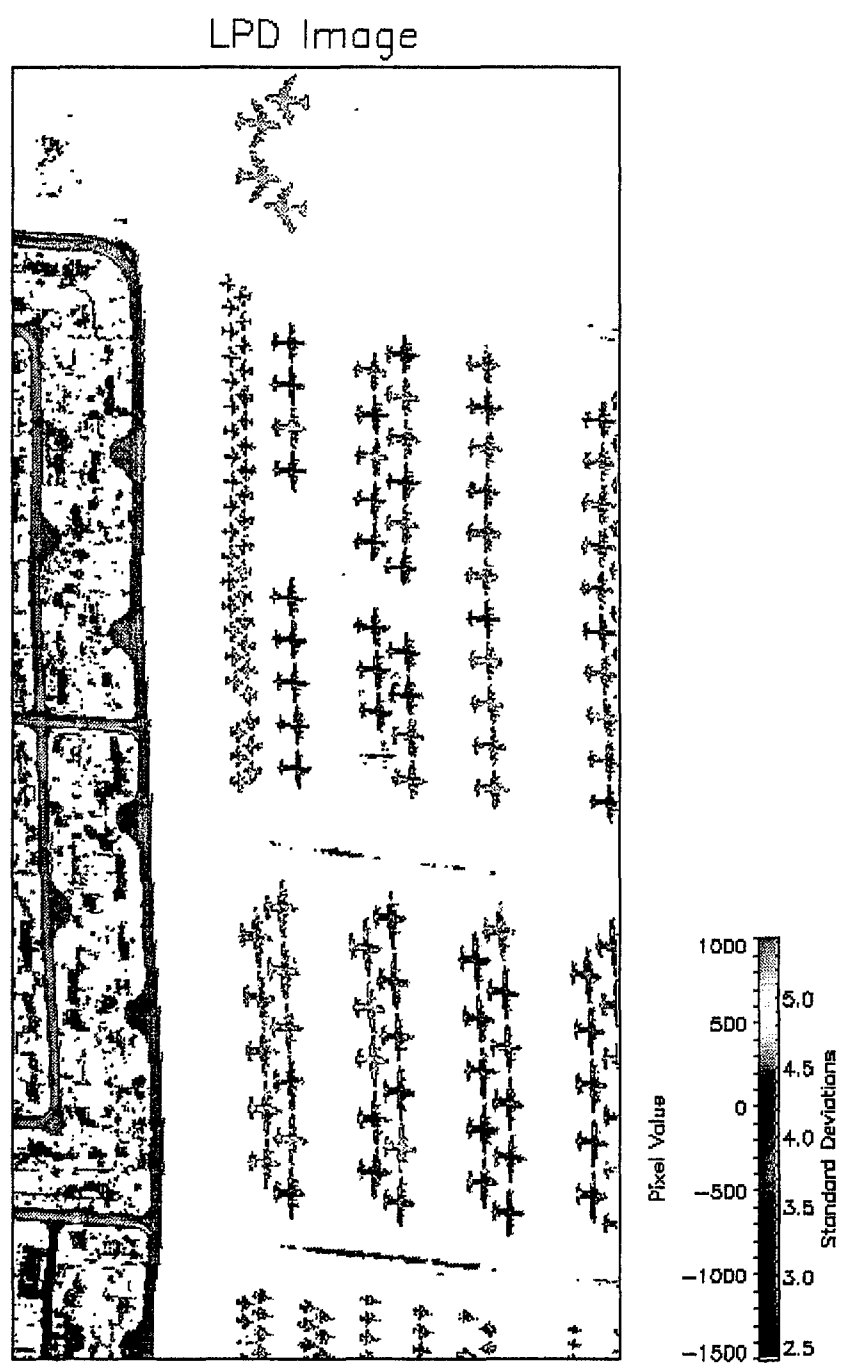


Figure 6.8: Davis Monthan LPD Output Image Using the First Eigenvector.

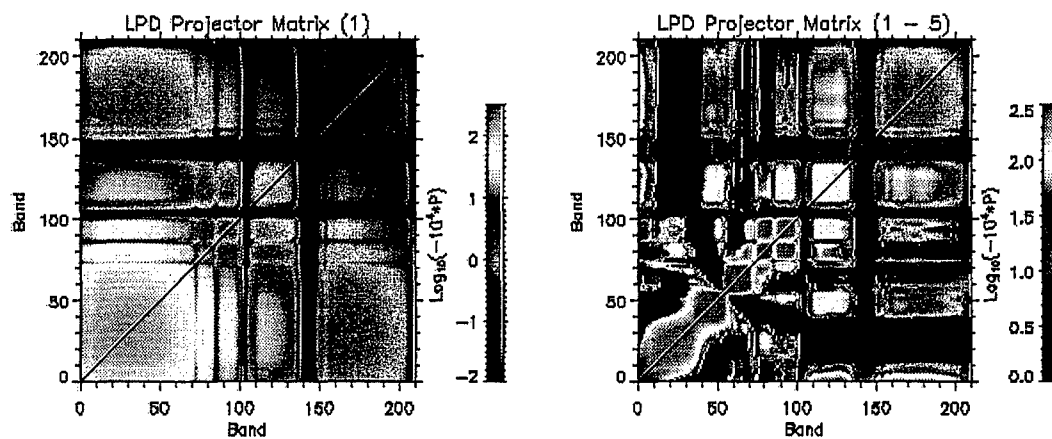


Figure 6.10: LPD Projector Matrices Created with the First Eigenvector and the First Five Eigenvectors for the Davis Monthan Sub-scene.

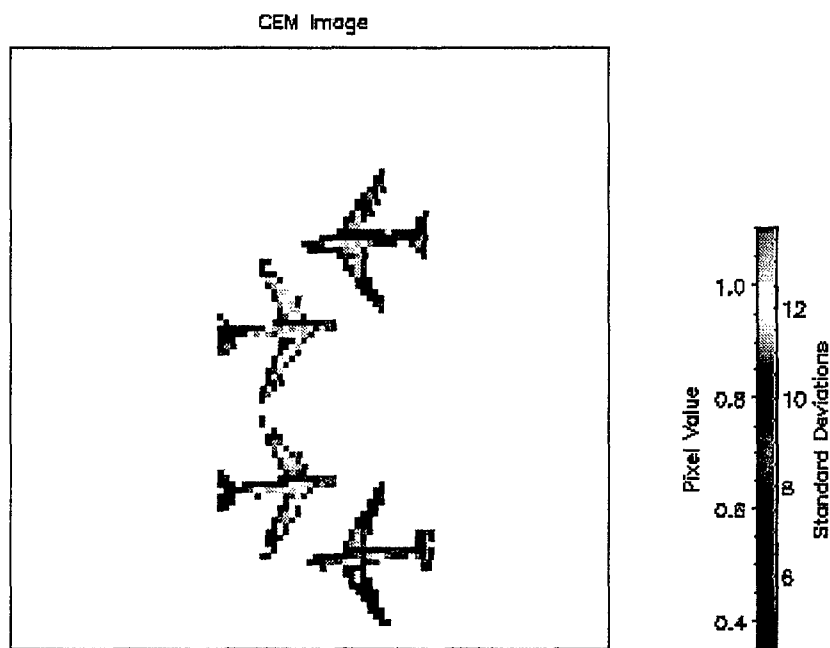


Figure 6.16: Davis Monthan Sub-scene CEM Output.



CEM Image

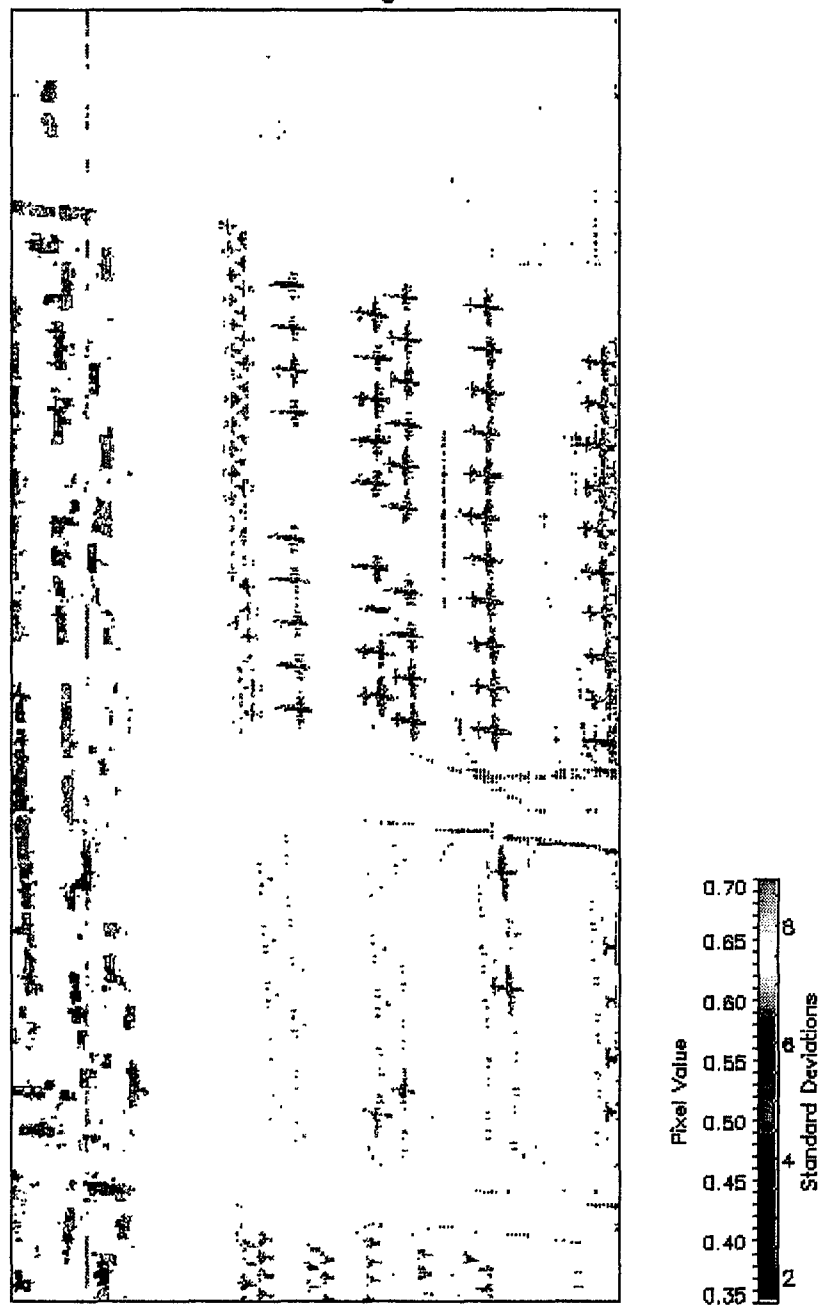


Figure 6.17: Davis Monthan CEM Output Image Using P-3 Pixel Vector as the Target.

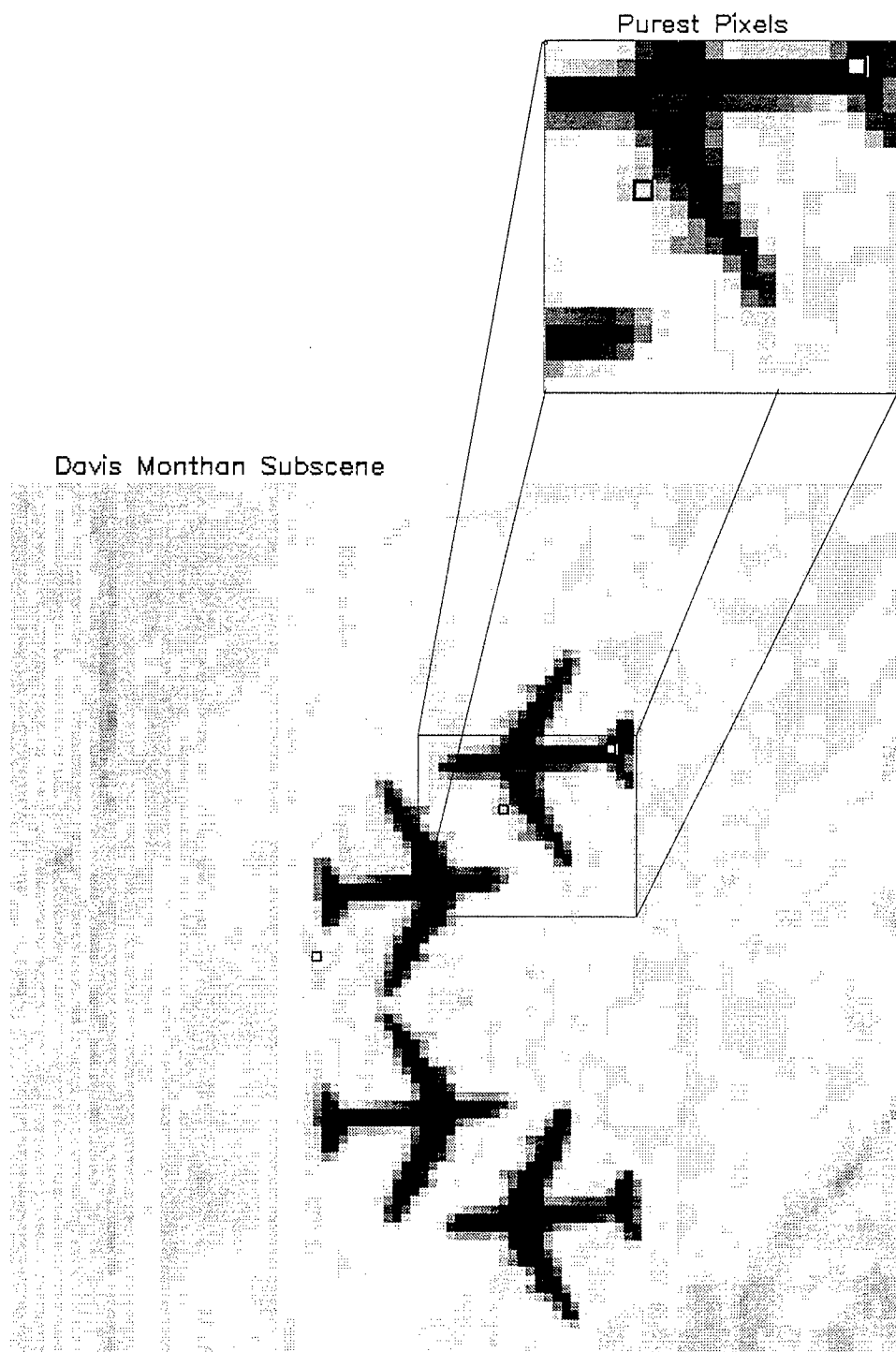


Figure 7.4: Purest Pixels in the Davis Monthan Sub-scene.

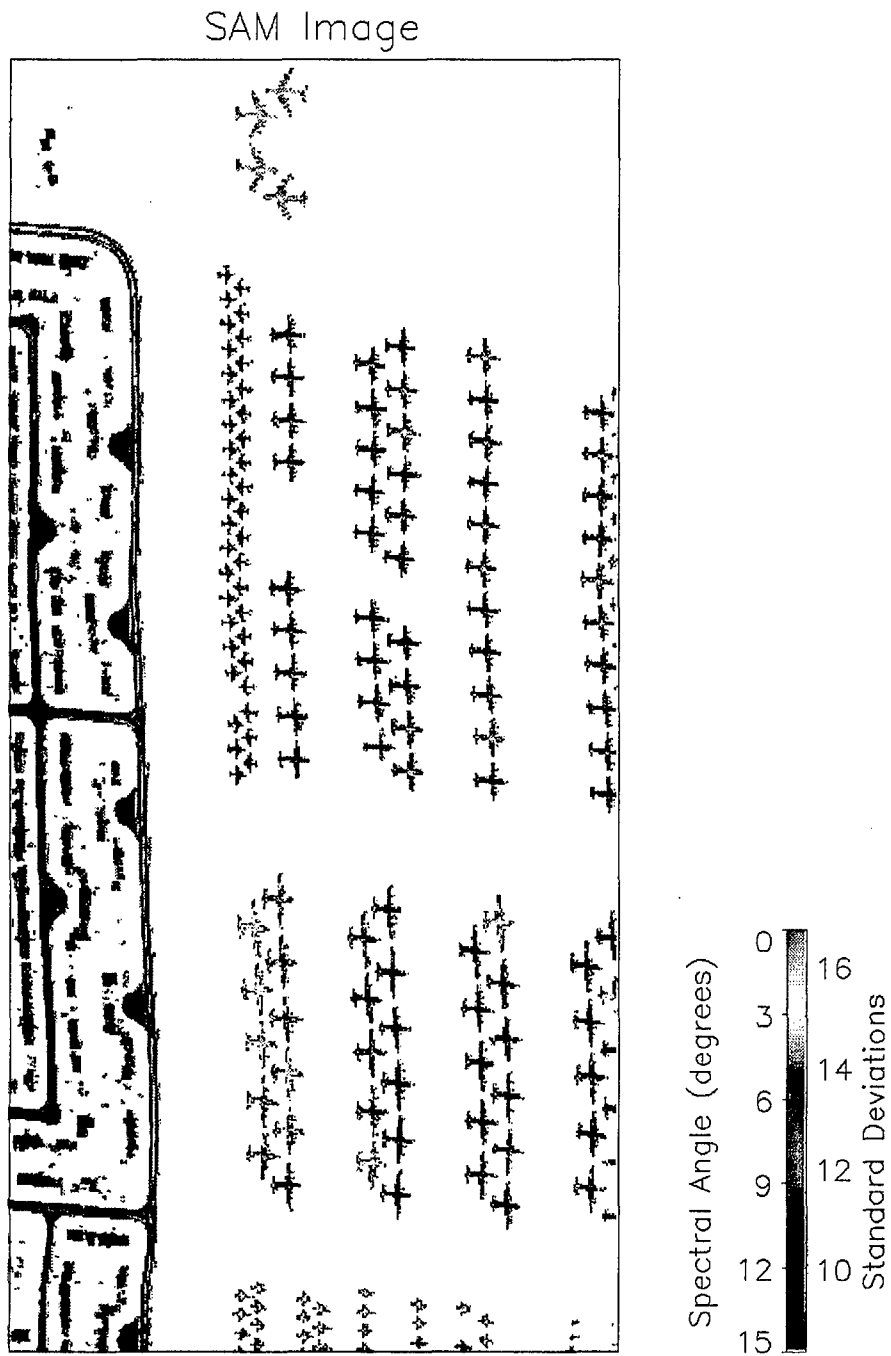


Figure 7.7: SAM Output for Davis Monthan Scene.

## APPENDIX B. SPECIAL COLOR FIGURES



Figure B.1: False Color PC Image of Davis Monthan Scene.



Figure B.2: False Color PC Image of Aberdeen Radiance Scene.



Figure B.3: False Color PC Image of Aberdeen Reflectance Scene.

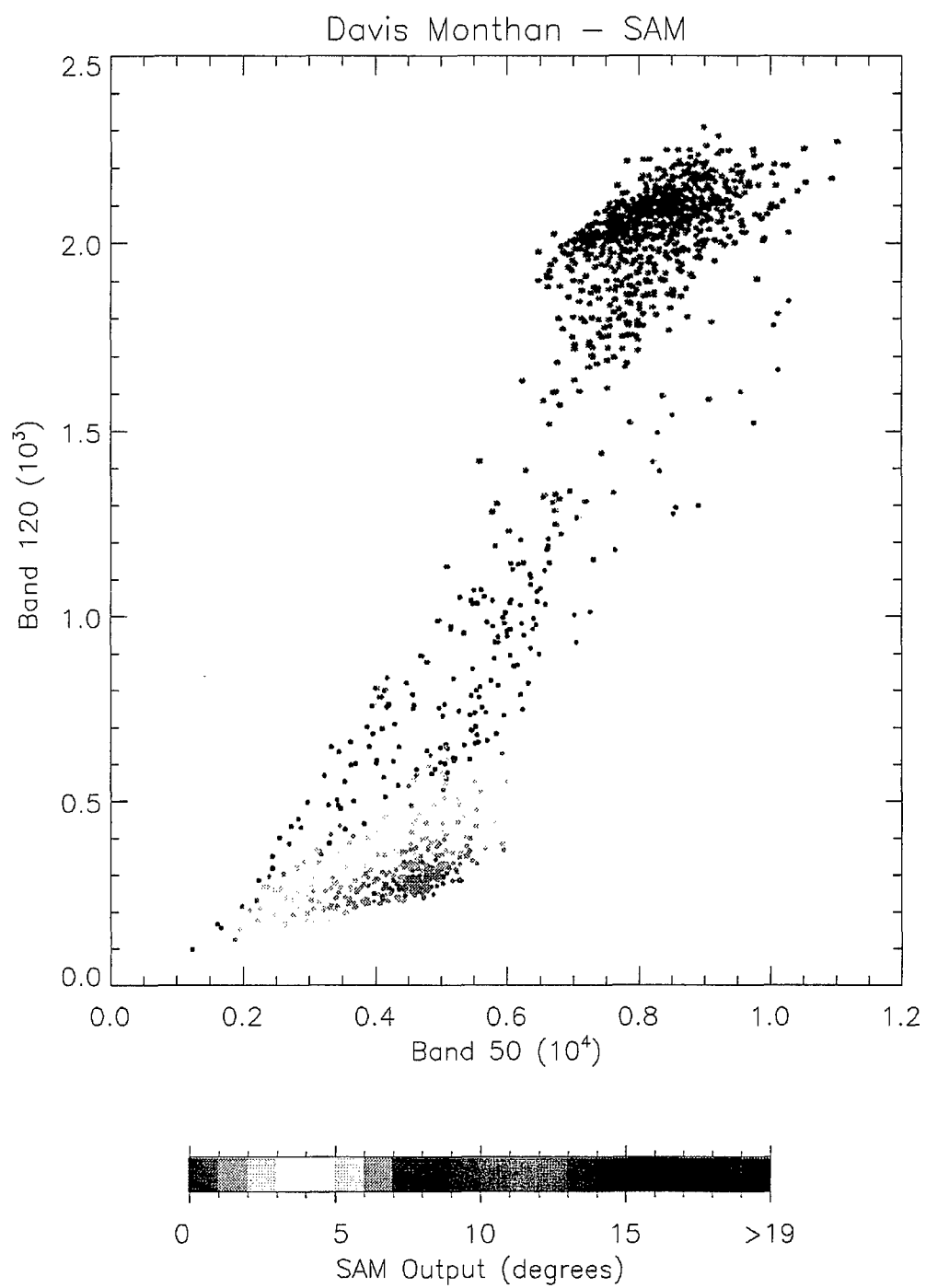


Figure B.4: SAM Output Superimposed on Scatter Plot of Davis Monthan Sub-scene.

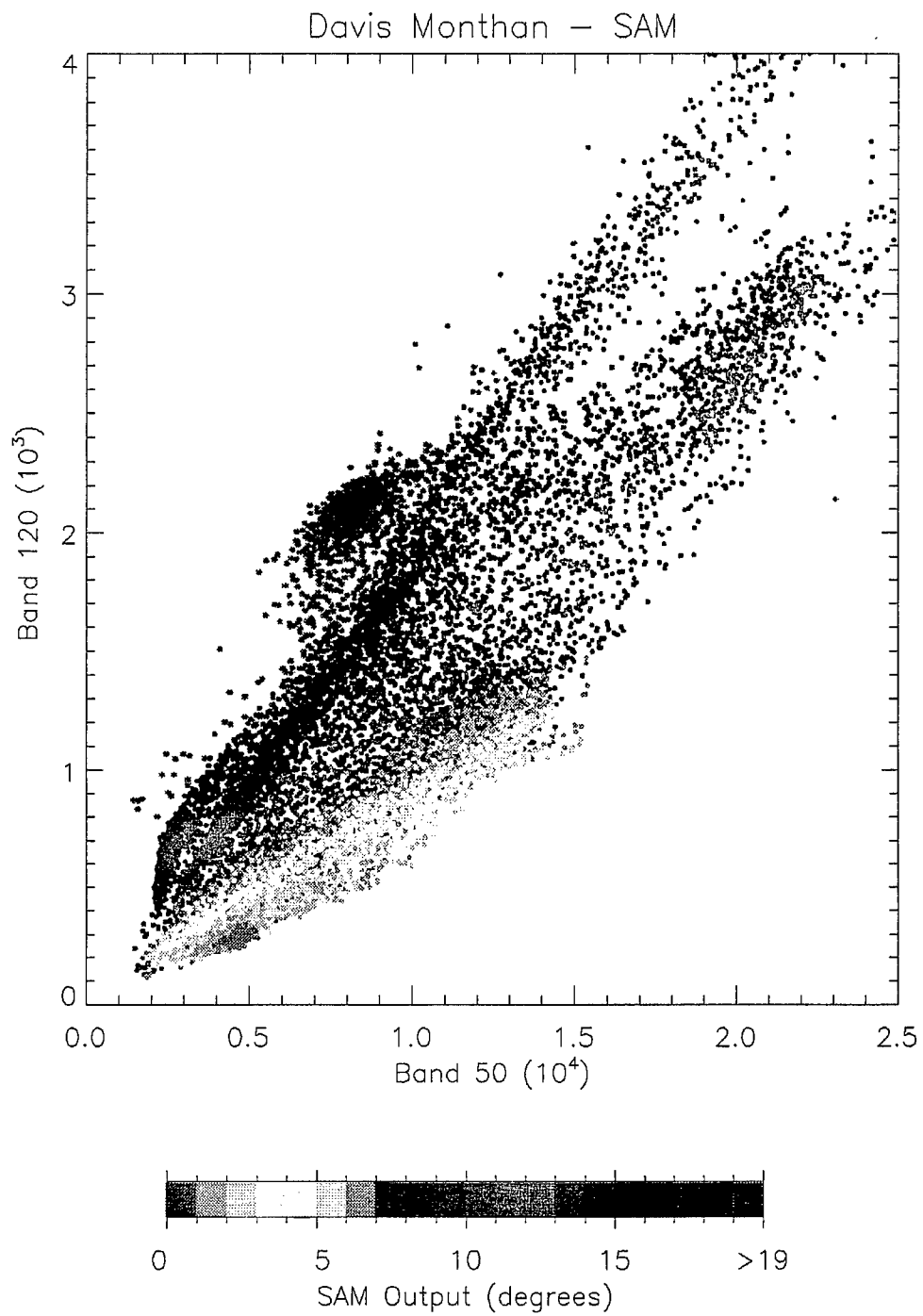


Figure B.5: SAM Output Superimposed on Scatter Plot of Davis Monthan Scene.



## LIST OF REFERENCES

- Adams, J. B., Smith, M. O., and Gillespie, A. R., "Simple Models for Complex Natural Surfaces: A Strategy for the Hyperspectral Era of Remote Sensing," *Proceedings of the International Geoscience and Remote Sensing '89/12<sup>th</sup> Canadian Symposium on Remote Sensing*, pp. 1-9, July, 1989.
- Anderson, T. W., *An Introduction to Multivariate Statistical Analysis*, Second Edition, John Wiley & Sons, New York, NY, 1984.
- Boardman, J. W., "Analysis, Understanding, and Visualization of Hyperspectral Data as Convex Sets in  $N$ -Space," *SPIE Proceedings*, vol. 2480, pp. 14-22, April 17-18, 1995.
- Boardman, J. W., Kruse, F. A., and Green, R. O., "Mapping Target Signatures Via Partial Unmixing of AVIRIS Data," *Summaries of the Fifth Annual JPL Airborne Earth Science Workshop*, vol. 1, pp. 23-26, January 23-26, 1995.
- Bowles, J., Palmadesso, P., Antoniadis, J., Baumbach, M., and Rickard, L. J., "Use of Filter Vectors in Hyperspectral Data Analysis," *SPIE Proceedings*, vol. 2553, 1995.
- Brower, B. V., Haddock, D. H., Reitz, J. P., and Schott, J., "Spectrally and Spatially Adaptive Hyperspectral Data Compression," *SPIE Proceedings*, vol. 2821, August, 1996.
- Burt, P. J., "A Gradient Pyramid Basis for Pattern Selective Image Fusion," *Proceedings of the Society for Information Display Conference*, 1992.
- Cloutis, E. A., "Hyperspectral Geological Remote Sensing: Evaluation of Analytical Techniques," *International Journal of Remote Sensing*, vol. 17, no. 12, pp. 2215-2242, 1996.
- Danaher, S. and O'Mongain, E., "Singular Value Decomposition in Multispectral Radiometry," *International Journal of Remote Sensing*, vol. 13, no. 9, pp. 1771-1777, 1992.
- Farrand, W. H. and Harsanyi, J. C., "Discrimination of Poorly Exposed Lithologies in Imaging Spectrometer Data," *Journal of Geophysical Research*, vol. 100, no. E1, pp. 1565-1578, January 25, 1995.
- Fukunaga, K., *Introduction to Statistical Pattern Recognition*, Academic Press, New York, NY, 1972.

Golub, G. H., and Van Loan, C. F., *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 1983.

Gorman, J. D., Subotic, N. S., and Thelen, B. J., "Robust Material Identification in Hyperspectral Data via Multiresolution Wavelet Analysis," *1995 International Conference on Acoustics, Speech, and Signal Processing Proceedings*, vol. 5, pp. 2805-2808, May 9-12, 1995.

Green, A. A., Berman, M., Switzer, P., and Craig, M. D., "A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no.1, pp. 65-74, January, 1988.

Harsanyi, J. C., *Detection and Classification of Subpixel Spectral Signatures in Hyperspectral Image Sequences*, Ph. D. Dissertation, University of Maryland, College Park, MD, 1993.

Harsanyi, J. C., Farrand, W. H., and Chang, C., "Detection of Subpixel Signatures in Hyperspectral Image Sequences," *ASPRS '94*, pp. 1-15, 1994.

Harsanyi, J. C., and Chang, C., "Hyperspectral Imagery Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no.4, pp. 779-785, July, 1994.

Harsanyi, J. C., Farrand, W. H., Hejl, J. M., and Chang, C., "Automatic Identification of Spectral Endmembers in Hyperspectral Image Sequences," *ISSSR '94*, vol. 1, pp. 267-277, July, 1994.

Haykin, S., *Adaptive Filter Theory*, Third Edition, Prentice Hall, Upper Saddle River, NJ, 1996.

Herries, G., Selige, T., Danaher, S., "Singular Value Decomposition in Applied Remote Sensing," *IEE Colloquium on Image Processing for Remote Sensing*, vol. 5, pp. 1-6, February 13, 1996.

Hoffbeck, J. P., and Landgrebe, D. A., "Classification of Remote Sensing Images Having High Spectral Resolution," *Remote Sensing of the Environment*, vol. 57, pp. 119-126, 1996.

Hotelling, H., "Analysis of Complex Statistical Variables into Principle Components," *The Journal of Educational Psychology*, vol. 24, pp. 417-441 and 498-520, September and October, 1933.

Jaynes, E. T., "On the Rationale of Maximum-Entropy Methods," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 939-952, September, 1982.

Jenson, S. K. and Waltz, F. A., "Principal Components Analysis and Canonical Analysis in Remote Sensing," *Proceedings of the American Society of Photogrammetry, 45<sup>th</sup> Annual Meeting*, pp. 337-348, 1979.

Johnson, P. E., Smith, M. O., and Adams, J. B., "Quantitative Analysis of Reflectance Spectra with Principal Components Analysis," *Proceedings of the 15<sup>th</sup> Lunar and Planetary Science Conference, Part 2, Journal of Geophysical Research*, vol. 90 supplement, pp. C805-C810, February, 1985.

Jolliffe, I. T., *Principal Component Analysis*, Springer-Verlag, New York, NY, 1986.

Kapur, J. N., *Maximum-Entropy Models in Science and Engineering*, John Wiley & Sons, New York, NY, 1989.

Lay, S. R., *Convex Sets and Their Applications*, John Wiley & Sons, New York, NY, 1982.

Lee, J. B., Woodyatt, A. S., and Berman, M., "Enhancement of High Spectral Resolution Remote-Sensing Data by a Noise-Adjusted Principal Components Transform," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no.3, pp. 295-304, May, 1990.

Leon-Garcia, A., *Probability and Random Processes for Electrical Engineering*, Second Edition, Addison-Wesley, Reading, MA, 1994.

Mather, P. M., *Computer Processing of Remotely-Sensed Images*, John Wiley & Sons, New York, NY, 1987.

Miller, J. W. V., Farison, J. B., and Shin, Y., "Spatially Invariant Image Sequences," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 148-161, April, 1992.

Moon, T. and Merenyi, E., "Classification of Hyperspectral Images Using Wavelet Transforms and Neural Networks," *SPIE Proceedings*, vol. 2569, pp. 725-731, 1995.

Nedeljkovic, V. and Pendock, N., "Finding Spectral Anomalies in Hyperspectral Imagery Using Nonparametric Probability Density Estimation Techniques," *Eleventh Thematic Conference and Workshops on Applied Geologic Remote Sensing*, vol. I, pp. 197-204, February 27-29, 1996.

Palmadesso, P., Antoniadis, J., Baumbach, M., Bowles, J., and Rickard, L. J., "Use of Filter Vectors and Fast Convex Set Methods in Hyperspectral Analysis," *Proceedings of the ISSSR Conference in Melbourne, Australia*, 1995.

Price, J. C., "How Unique Are Spectral Signatures?" *Remote Sensing of the Environment*, vol. 49, pp. 181-186, 1994.

Rao, A. K. and Bhargava, S., "Multispectral Data Compression Using Bidirectional Interband Prediction," *IEEE Transactions on Geoscience and Remote Sensing*, vol.34, no. 2, pp. 385-397, March, 1996.

Ready, P. J. and Wintz, P. A., "Information Extraction, SNR Improvement, and Data Compression in Multispectral Imagery," *IEEE Transactions on Communications*, vol. COM-21, no. 10, pp. 1123-1130, October, 1973.

Research Systems, Inc., Boulder, CO, *Interactive Data Language User's Guide*, Version Four, March, 1995.

Richards, J. A., *Remote Sensing Digital Image Analysis*, Springer-Verlag, New York, NY, 1986.

Rinker, J. M., "Hyperspectral Imagery - What Is It? - What Can It Do?" *USACE Seventh Remote Sensing Symposium*, May 7-9, 1990.

Rissanen, J., "Modeling by Shortest Data Description," *Automatica*, vol. 14, pp. 465-471, 1978.

Roger, R. E., "A Faster Way to Compute the Noise-Adjusted Principal Components Transform Matrix," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no.6, pp. 1194-1196, November, 1994.

Roger, R. E., and Cavenor, M. C., "Lossless Compression of AVIRIS Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol.5, no. 5, pp. 713-719, May, 1996.

Scharf, L. L., *Statistical Signal Processing: Detection, Estimation, and Time-Series Analysis*, Addison-Wesley, Reading, MA, 1991.

Schmidt, R. O., "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, no. 3, pp. 276-280, March, 1986.

Searle, S. R., *Matrix Algebra Useful for Statistics*, John Wiley & Sons, New York, NY, 1982.

Settle, J. J., "On the Relationship Between Spectral Unmixing and Subspace Projection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 4, pp. 1045-1046, July, 1996.

Singh, A. and Harrison, A., "Standardized Principal Components," *International Journal of Remote Sensing*, vol. 6, no. 6, pp. 883-896, 1985.

Smith, M. O., Johnson, P. E., and Adams, J. B., "Quantitative Determination of Mineral Types and Abundances from Reflectance Spectra Using Principal Components Analysis," *Proceedings of the 15<sup>th</sup> Lunar and Planetary Science Conference, Part 2, Journal of Geophysical Research*, vol. 90 supplement, pp. C797-C804, February, 1985.

Solberg, R. and Egeland, T., "Automatic Feature Selection in Hyperspectral Satellite Imagery," *IGARSS '93: A Better Understanding of the Earth Environment*, vol. 2, pp. 472-475, August 18-21, 1993.

Stocker, A. D., Reed, I. S., and Yu, X., "Multi-Dimensional Signal Processing for Electro-Optical Target Detection," *SPIE Proceedings of Signal and Data Processing of Small Targets 1990*, vol. 1305, pp. 218-231, April, 1990.

Stoica, P. and Moses, R., *Introduction to Spectral Analysis*, Prentice Hall, Englewood Cliffs, NJ, 1996.

Swain, P. H. and Davis, S. M., *Remote Sensing: The Quantitative Approach*, McGraw-Hill, New York, NY, 1978.

Therrien, C. W., Quatieri, T. F., and Dudgeon, D. E., "Statistical Model-Based Algorithms for Image Analysis," *Proceedings of the IEEE*, vol. 74, no. 4, pp. 532-551, April, 1986.

Therrien, C. W., *Decision, Estimation, and Classification: An Introduction to Pattern Recognition and Related Topics*, John Wiley & Sons, New York, NY, 1989.

Therrien, C. W., *Discrete Random Signals and Statistical Signal Processing*, Prentice Hall, New York, NY, 1992.

Tu, T., Chen, C., and Chang, C., "A *Posteriori* Least Squares Orthogonal Subspace Projection Approach to Desired Signature Extraction and Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 1, pp. 127-139, January, 1997.

Van Der Veen, A., Deprettere, E. F., and Swindlehurst, A. L., "Subspace-Based Signal Analysis Using Singular Value Decomposition," *Proceedings of the IEEE*, vol. 81, no. 9, pp. 1277-1308, September, 1993.

Vane, G. and Goetz, A. F. H., "Terrestrial Imaging Spectroscopy," *Remote Sensing of the Environment*, vol. 24, pp. 1-29, 1988.

Wang, J., Zhang, K., and Tang, S., "Spectral and Spatial Decorrelation of Landsat-TM Data for Lossless Compression," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 5, pp. 1277-1285, September, 1995.

Watkins, D. S., *Fundamentals of Matrix Computations*, John Wiley & Sons, New York, NY, 1991

Wax, M. and Kailath, T., "Detection of Signals by Information Theoretic Criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, no. 2, pp. 387-392, April, 1985.

Wilson, T. A., Rogers, S. K., and Meyers, L. R., "Perceptual-Based Hyperspectral Image Fusion Using Multiresolution Analysis," *Optical Engineering*, vol. 34, no. 11, pp. 3154-3164, November, 1995.

Winter, E., Ed., *DARPA/TTO Hyperspectral Mine Detection and Phenomenology Report*, March 25, 1995.

Yu, X., Reed, I. S., and Stocker, A. D., "Comparative Performance Analysis of Adaptive Multispectral Detectors," *IEEE Transactions on Signal Processing*, vol. 41, no. 8, pp. 2639-2656, August, 1993.

Yuhas, R. H., Goetz, A. F. H., Boardman, J. W., "Discriminating Among Semi-Arid Landscape Endmembers Using the Spectral Angle Mapper (SAM) Algorithm," *Summaries of the Third Annual JPL Airborne Geoscience Workshop*, vol. 1, pp. 147-149, June 1-5, 1992.

## INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center ..... 2  
 8725 John J. Kingman Rd., STE 0944  
 Ft. Belvoir, VA 22060-6218
  
2. Dudley Knox Library ..... 2  
 Naval Postgraduate School  
 411 Dyer Rd.  
 Monterey, CA 93943-5101
  
3. Director, Training and Education ..... 1  
 MCCDC, Code C46  
 1019 Elliot Road  
 Quantico, VA 22134-5027
  
4. Director, Marine Corps Research Center ..... 2  
 MCCDC, Code C40RC  
 2040 Broadway Street  
 Quantico, VA 22134-5107
  
5. Director, Studies and Analysis Division ..... 1  
 MCCDC, Code C45  
 3300 Russell Road  
 Quantico, VA 22134-5130
  
6. Marine Corps Representative ..... 1  
 Naval Postgraduate School  
 Code 037, Bldg. 234, HA-220  
 699 Dyer Road  
 Monterey, CA 93940
  
7. Marine Corps Tactical Systems Support Activity ..... 1  
 Technical Advisory Branch  
 Attn: Maj J. C. Cummiskey  
 Box 555171  
 Camp Pendleton, CA 92055-5080
  
8. Charirman, Code EC ..... 1  
 Department of Electrical and Computer Engineering  
 833 Dyer Road, Room 437  
 Naval Postgraduate School  
 Monterey, CA 93943-5121

9. Dr. Richard C. Olsen, Code PH/Os .....2  
Department of Physics  
Naval Postgraduate School  
Monterey, CA 93943-5117
10. Dr. Roberto Cristi, Code EC/Cx..... 1  
Department of Electrical and Computer Engineering  
833 Dyer Road, Room 437  
Naval Postgraduate School  
Monterey, CA 93943-5121
11. Captain Marcus S. Stefanou, USMC..... 1  
1040 Myerstown Road  
Gardners, PA 17324