ESC-TR-2001-080

Project Report
HTAP-12

Distance Metrics and Band Selection in Hyperspectral Processing with Applications to Material Identification and Spectral Libraries

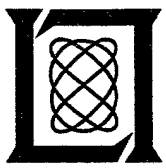
N. Keshava

18 December 2002

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LEXINGTON, MASSACHUSETTS



Prepared for the Department of the Under Secretary of Defense, S&T,
under Air Force Contract F19628-00-C-0002.

This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. The work was sponsored by the Department of the Under Secretary of Defense, S&T, under Air Force Contract F19628-00-C-0002.

The ESC Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Gary Tutungian
Administrative Contracting Officer
Plans and Programs Directorate
Contracted Support Management

Massachusetts Institute of Technology
Lincoln Laboratory

Distance Metrics and Band Selection in Hyperspectral Processing with Applications to Material Identification and Spectral Libraries

N. Keshava
Group 97

Project Report HTAP-12

18 December 2002

Lexington Massachusetts

EXECUTIVE SUMMARY

In this report, we investigate and exploit the properties of distance metrics in hyperspectral processing to achieve superior algorithm performance as well as dimension reduction. Distance metrics are mathematical operators that provide a scalar measure of similarity for two hyperspectral (vector) signals, and they are at the nucleus of many application algorithms. The similarity between two signals, however, can be measured by various means, and different distance metrics offer distinct notions of similarity. Consequently, a thorough understanding of the mathematical and physical properties of distance metrics is crucial to the accurate and efficient processing of hyperspectral data.

After formally introducing the mathematical definitions and properties of distance metrics, we focus on two distance metrics that frequently appear in hyperspectral processing and provide complementary interpretations of distance in high-dimensional space. The Euclidean Minimum Distance (EMD) measures the shortest distance between two spectra, whereas the Spectral Angle Mapper (SAM) measures the angle created by the two spectra. After enumerating their properties and demonstrating how each appears in several detection, classification, and unmixing algorithms, we focus on SAM because of its widespread use, and its unique mathematical and physical properties.

A simple example demonstrates how the angle between two spectra changes as subsets of different bands are retained and omitted. This inherent property of SAM entertains the possibility of increasing the angle between two spectra, and hence their discriminability, by selecting an appropriate subset of available bands. Simple search algorithms are explored to find the contiguous segment(s) of bands that maximize the angle between two spectra, but these approaches are highly sub-optimal, as well as computationally impractical. However, an analytical approach (Band Add-On or BAO) based on a mathematical decomposition of SAM incrementally "builds up" a set of bands that maximize the angle between two spectra. This approach compares very favorably against the results of exhaustively enumerating every angle between two spectra, which is computationally impractical for hyperspectral signals.

The BAO approach is then extended to select bands that increase the angular separation between two classes of spectra, where each class is populated by a set of reference spectra. This scenario strongly parallels the material identification problem, where a small number (< 10) of laboratory reflectance measurements are collected to provide a signature for a material, and the goal is to assign an unknown pixel spectrum measured by a sensor to one of many material classes. Two complementary band selection techniques based on BAO are developed that select bands and template spectra to increase the angular separation between two such classes of spectra. Their ability to discriminate two very similar target classes is tested using laboratory and sensor data collected with the HYDICE sensor. Our experimental results using real data show that using all available bands in an angle-based test misclassifies half the pixels, but band selection succeeds in correctly classifying all

pixels while using only a fraction of the available bands.

This two-class discrimination technique provides the fundamental unit for a multi-class, hierarchical architecture for material identification, which generalizes the standard, linear architecture that sequentially measures the angle between an unknown pixel and every library template spectra using all bands collected by the sensor. The basic kernel of the hierarchical approach is a binary test that compares an unknown pixel to two classes at a time using a set of bands and template spectra unique to the two classes. The class having the greater angle is eliminated from further consideration, and another binary test is performed with the retained class and a new class, using a new set of bands and templates. Employing 10 similar target classes, the results from the hierarchical architecture using two different band and template selection approaches are compared to the linear architecture using all bands. The band selection approaches clearly yield better classification performance than using all bands, while only using a small fraction of bands.

Other benefits of angle-based band selection are also discussed. Statistical target detection algorithms are designed to distinguish desired target pixels from natural background pixels. Examples demonstrate that detection statistics from similar targets, which are common in CC&D environments, are difficult to differentiate from the desired target. Band selection for material identification can provide detection post-processing that mitigates false alarms that arise from pixels that are similar to the desired target spectrum, yet are still different. Further, the examples generated in this report demonstrate that significant improvement in material identification performance results with a dramatic reduction in the number of bands utilized. This form of dimension reduction has the potential to reduce the requirements for future sensors, especially those striving for real-time performance.

ACKNOWLEDGMENTS

The author would like to thank CAPT Frank Garcia at DUSD (S&T), program manager for the Lincoln Laboratory Hyperspectral Technology Assessment Program (HTAP) for supporting this effort. Arun Batra and Peter Boettcher were instrumental in implementing many of the algorithms in Matlab. The author would also like to thank David Marden for providing code and advice on computational issues and Dimitris Manolakis, Dan Dudgeon, and John Kerekes for several meaningful technical discussions. Carolyn Upham also provided valuable assistance with data and ground truth. Finally, the author extends his gratitude to Andy McKellips for providing a LaTeXtemplate for this report.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

Much of the challenge in modern sensing technologies focuses on techniques to efficiently process increasingly vast amounts of data. The flood of data can arise from either temporal measurements of one or a few quantities with rapidly dwindling re-visit times, a whole assortment of parameters estimated at one instant, or from both of these circumstances. Intuitively, an improvement in algorithm performance commensurate to the increase in input data might be anticipated, but no axiom guarantees that such a scaling always holds.

Passive sensing has followed this progression, originating from single, wide-band measurements to strategically placed bands in multispectral processing, to hyperspectral sensing where large intervals of the electromagnetic spectrum are measured in contiguous bins having widths as narrow as 3 nm. The explosion of data is occurring along the axes of spectral and spatial resolution as well as temporal frequency. For example, Figure 1 compares the bands from Landsat 7, which was launched in 1999, to the coverage of hyperspectral sensors in the reflective regime such as AVIRIS (Airborne Visible/Infrared Imaging Spectrometer)and HYDICE (Hyperspectral Digital Imagery Collection Experiment). Accompanying the increased capability of collecting spectral information has been a growing demand to measure quantities of interest in even greater detail than before, as well as to derive altogether new information products. In either case, the potential for extracting useful information from hyperspectral data is immense, but the realization of this goal does not reside solely in the expanding volume of data, but in the techniques employed to process it.

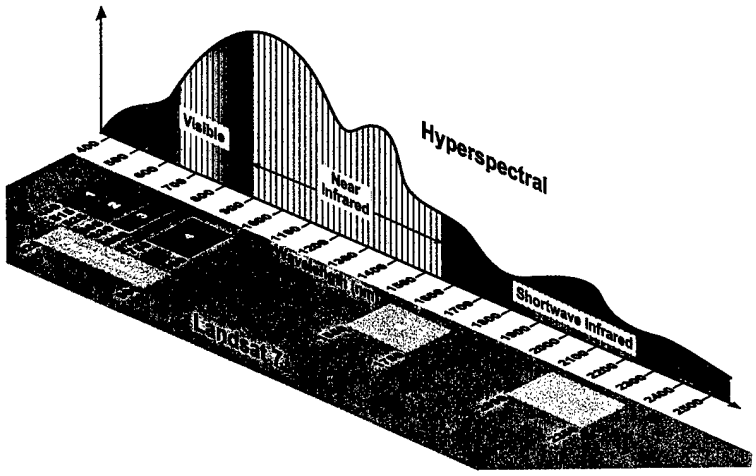


Figure 1. Comparison of hyperspectral and Landsat spectral coverage. Band 6 extends from 10400 nm to 12500 nm. The HYDICE sensor has 210 bands with widths ranging from 3-11 nm. The AVIRIS sensor has 224 bands with widths of 10 nm.

1.1 WHY ARE DISTANCE METRICS IMPORTANT?

Hyperspectral data processing is one of a multitude of scientific endeavors that extrapolates meaningful information from numerical data. Algorithms are developed for useful applications that automate basic tasks that humans are unable to do in a timely fashion. Considering that space-based platforms have become standard implements for persistent civilian and military monitoring of the Earth, the amount of data being constantly collected far exceeds the human resources necessary for processing and analysis. Hence, the development of successful algorithms must achieve two goals. First, an algorithm must yield accurate and verifiable answers. At the same time, however, it must deliver these results using a minimum of data and with maximum computational efficiency.

Hyperspectral applications are varied and have been designed to satisfy different criteria (e.g., least squared error (LSE), maximum likelihood (ML), maximum a posteriori (MAP)) to meet their goals. In doing so, the approaches may utilize different variables and computational kernels. **However, most algorithms share one critical component at their core: an operator known as a distance metric that mathematically quantifies the similarity between two spectra.** For example, in target detection, the comparison occurs between a desired spectral signature and a pixel spectrum collected by a sensor from a scene. The result of the distance metric is compared to a threshold.

Distance metrics should not be confused with the features they compare. Each band in a spectrum is, by itself, a feature, but a distance metric provides the means of comparing two sets of bands. Similarly, when comparing two people, different features can be employed (e.g., height, weight, eye color). Individual features can be compared by their associated metrics to provide discrimination (e.g., *height1 - height2*). However, provided all features can be expressed numerically, the means for comparing two sets (or vectors) of disparate features is not as clear. Consequently, the selection of an appropriate distance metric is crucial. This report explains what distance metrics have been employed for hyperspectral processing, and by virtue of developing a strong mathematical foundation for comparing two physical spectra, we discover opportunities to substantially improve upon the two aforementioned objectives of algorithm design for hyperspectral processing, improved performance and more efficient computation.

1.2 HYPERSPECTRAL ALGORITHMS

The principle end products from the processing of military hyperspectral data are derived from four categories of algorithmic processing: 1) target detection, 2) classification, 3) spectral unmixing, and 4) material identification. For each of these categories, distance metrics provide algorithms the core capability of discriminating one class of signal from another. For instance, in the detection of known targets, a distance metric compares a pixel spectrum collected from a scene by a hyperspectral sensor to a reference, or library, spectrum, and based on the scalar measure of distance and a user-defined threshold, deems the pixel to be either of the same type as the reference spectrum or from a different class. Likewise, unsupervised classification strives to naturally segregate data into distinct classes, and distance metrics enable the comparisons of individual spectra with class centroids. Unmixing also employs a distance metric to estimate the

sub-pixel components that comprise the measured spectrum from a pixel. In this case, squared-error is the quantity that is frequently minimized. Finally, material identification of a pixel is accomplished by a distance metric that compares a received pixel spectrum with a series of template spectra in a spectral library, and assigns the pixel to the material class having the smallest distance.

1.3 CONTRIBUTIONS TO HYPERSPECTRAL PROCESSING

In addition to a structured, analytical explanation of distance metrics used in hyperspectral processing, this report outlines several tangible benefits that arise from exploiting the mathematical and physical properties of the most commonly used distance metric, the Spectral Angle Mapper (SAM), which will be discussed in detail later in this report.

- Band Selection

  The mathematical structure of SAM directly reveals how a subset of hyperspectral bands may be selected to improve the discriminability of two classes of targets. Experiments have shown that a significantly lower number of bands can provide better separability than using every available band from a hyperspectral sensor.

- Improved Material Identification/Classification for CC&D

  As a consequence of selecting bands that increase the capability of distinguishing one class from another, the ability to correctly classify the material composition of a pixel is also enhanced. This has important ramifications for material identification, classification, and detection of CC&D targets.

- Dimension Reduction/Real-Time Processing

  Our results demonstrate that superior performance can be achieved using significantly fewer bands than are available from the sensor. This has the potential of reducing the requirements on sensor design and algorithmic processing. For certain, important applications, band selection may be performed off-line, permitting fast and efficient real-time processing using only a subset of the data collected by the sensor.

1.4 IN THIS REPORT

This project report provides a detailed technical discussion of distance metrics in hyperspectral processing, metric-based band selection, and applications toward material identification, statistical target detection, and dimension reduction. Section 1 motivates the importance of distance metrics as a way of comparing two spectra measured by a sensor. It also discusses how distance metrics are at the core of many common application algorithms, and that their proper use and optimization can significantly improve the performance of important hyperspectral applications.

Section 2 introduces formal, mathematical definitions for quantities that are important to the study of distance metrics. After providing a definition for distance metrics, two distance metrics commonly used in hyperspectral processing, the Spectral Angle Mapper (SAM) and the Euclidean

Minimum Distance (EMD), are introduced and their properties are enumerated. The roles played by SAM and EMD in common hyperspectral applications such as unmixing, classification, and detection are shown.

In Section 3, the idea of band selection to improve algorithm performance is discussed. Knowledge of the spectral intervals where phenomenology is observable has been the principle form of band selection. However, while this approach identifies the appropriate spectral interval, it does not optimize the performance of mathematical algorithms that may ultimately process the data. The complementary concept of selecting bands to mathematically optimize a distance metric, namely SAM, is introduced, with the goal of increasing the angular separation between two spectra to improve the performance of algorithms based on that metric. Simple examples demonstrate that the sub-angles measured between two spectra can vary as different bands are selectively retained and omitted. In Section 3.3 and Section 3.4, primitive search algorithms are investigated to find contiguous segments of bands that increase the sub-angle between two spectra. These methods, however, are grossly inefficient. In Section 3.5 an analytical decomposition of SAM provides the foundation for a band selection algorithm that rapidly selects bands that maximize the angular separation between two spectra. A graphic description of the Band Add-On (BAO) technique is provided with examples. The results of the different band selection algorithms are compared to answers determined by exhaustive evaluation of all possible sub-angles, conclusively demonstrating that the BAO technique is able to rapidly find sub-angles that exceed, sometimes significantly, the angle created using all available bands. The sensitivities of the BAO approach are discussed.

Section 4 extends the BAO concept to increase the angular separability between two classes of spectra, where each class may consist of several sample spectra. The strong parallelism of this scenario to the material identification problem in hyperspectral processing is noted, where the goal is to classify an unknown reflectance spectrum into one of several library classes that are each defined by multiple laboratory reference spectra. Two complementary philosophies for selecting bands, both based on the BAO approach, are proposed. The Average Distance Method (ADM) selects bands to maximize the average angular distance between the spectra in the two classes. The Minimum Distance Method (MDM) selects bands with the goal of maximizing the angle between the members of each class that are the most similar, thus maximizing the worst-case angle between the two classes.

The task of accurately discriminating between two similar target classes is explored in Section 4.4 with the motivation that many CC&D targets possess very similar spectra. Real laboratory reflectance measurements as well as sensor data collected by the HYDICE sensor are used. The traditional approach that uses all available bands (in this case there are 145 available bands) fails to correctly discriminate the pixels. However, MDM correctly classifies pixels from both classes, while only using 20 bands.

In Section 5, the methods used to increase the angular separability and classification performance for two classes is extended to a hierarchical architecture suitable for material identification with an unlimited number of spectral classes. A classification experiment having ten similar classes is conducted and the results from band selection using ADM and MDM are compared to the results generated by using all available bands. The results demonstrate that ADM and MDM provide superior classification performance using significantly fewer bands.

In Section 6, the applicability of the procedures developed in Section 5 and Section 4 are demonstrated toward the task of mitigating false alarms in statistical target detection. Further, the benefits of band selection are discussed in the context of reducing the dimension of hyperspectral data.

2. DISTANCE METRICS

Hyperspectral sensing derives a strong foundation from the passive observation of physical phenomena that are active in the area being imaged. More often than not, algorithms for processing hyperspectral data are essentially attempting to unravel, or unwrap, the parameters of these phenomena from data and manipulate them in a way that results in acceptable performance. The physical parameters of interest, however, have been corrupted by numerous sources of interference (e.g., atmospheric, sensor) and distorted by limitations in viewing (e.g., spectral and spatial resolution, adjacency, quantization, focal plane defects). The idea of "looking backwards" from the data to "see" the phenomenology is not new and has been studied in the form of inverse problems for many years [42, 43].

In this section, we present analytical explanations of two distance metrics that are commonly used in hyperspectral processing [23]. In essence, a *distance metric* is a mathematical operator that conveys how similar two (possibly vector-valued) members of a set are with a single, scalar value, based on a notion of similarity. Different metrics employ alternative notions of similarity, and, consequently, each metric uniquely translates the phenomenology. In this sense, a metric is well-suited to a problem when it is *matched* to, and exposes, the aspect of the underlying physics that the application algorithm seeks to exploit. Before exploring this further, however, we present some mathematical definitions that provide a theoretical context for distance metrics.

2.1 MATHEMATICAL PRELIMINARIES

In order to discuss the properties of distance metrics in hyperspectral processing, we first present definitions that provide a foundation for further discussion and analysis. While a strong mathematical background is not a prerequisite for comprehending the results of this report, these concepts formalize the arguments. Rigorous discussions of these arguments may be found in any mathematical text on real analysis [31].

Definition: Linear Space. *A linear space (or a vector space) consists of a set Ω, a field F, and two functions $+ : \Omega \times \Omega \to \Omega$ and $\cdot : F \times \Omega \to \Omega$, where we denote $+(\mathbf{x}, \mathbf{y})$ by $\mathbf{x} + \mathbf{y}$ and $\cdot(\alpha, \mathbf{x})$ by $\alpha\mathbf{x}$, such that the following conditions are satisfied for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \Omega$ and $\alpha, \beta \in F$:*

a) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.

b) $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$.

c) There exists a $0 \in \Omega$ *such that* $\mathbf{x} + 0 = \mathbf{x}$.

d) There exists $-\mathbf{x} \in \Omega$ *such that* $\mathbf{x} + (-\mathbf{x}) = 0$.

e) $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$.

f) $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$.

g) $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \alpha\mathbf{y}$.

h) $1 \cdot \mathbf{x} = \mathbf{x}$.

Here, the field, F is called the field of scalars, + vector addition, and · scalar multiplication.

For hyperspectral sensing, the set, Ω, corresponds to the M-dimensional vector space of real numbers, heretofore denoted as $\Re^M$ (where M is the number of spectral bands), that describes the vector of entries (reflectance or radiance) from one pixel. In fact, reflectance and radiance values are inherently non-negative real numbers, but subsequent manipulations can lead to values from the entire real number line. The properties confirm for two vectors, $\mathbf{x}, \mathbf{y} \in \Re^M$ the commutative, associative, and distributive properties of addition as well as multiplication by a scalar, and the existence of additive inverses and identities. The next definition demonstrates the notion of length for a vector in $\Re^M$.

Definition: Norm, Normed Space. *Let Ω be a linear space. A function $||\ ||: \Omega \rightarrow \Re$, whose value at $\mathbf{x}$ is written as $||\mathbf{x}||$, is said to be a* **norm** *on Ω if it satisfies the following conditions for all $\mathbf{x}, \mathbf{y} \in \Omega$ and $\alpha \in F$.*

a) $||\mathbf{x}|| \geq 0$, *with equality if and only if* $\mathbf{x} = 0$.

b) $||\alpha\mathbf{x}|| = |\alpha|||\mathbf{x}||$.

c) $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$.

If $||\ ||$ is a norm on Ω, then the pair $(\Omega, ||\ ||)$ is called a **normed space.**

A linear space that is normed conforms to properties that guarantee that the norm (or length) of a vector, $\mathbf{x}$, may only be zero if $\mathbf{x} = 0$ and that scalar multiplication of $\mathbf{x}$ may travel outside the norm operator. The third condition is an obvious property of vector geometry that guarantees the sum of the lengths of two vectors is greater than or equal to the length of the sum. Equality holds when the two vectors are parallel.

Finally, a distance metric that is induced by a norm provides a notion of closeness or similarity for two members of $\Re^M$. The definition is as follows.

Definition: Distance Metric, Metric Space. *Let Ω be a set. A function $d : \Omega \times \Omega \rightarrow \Re$ is said to be a* **distance metric** *on Ω if it satisfies the following conditions for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \Omega$:*

a) $d(\mathbf{x}, \mathbf{y}) \geq 0$, *with equality if and only if* $\mathbf{x} = \mathbf{y}$.

b) $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$.

c) $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$.

The first condition for a distance metric is intuitive. All distances must be non-negative. The second condition requires the operator to yield the same value independent of the order of the operands. Finally, the third, and perhaps most important property of normed spaces is given by the triangle inequality. Essentially, the distance between two points in a normed space must be less

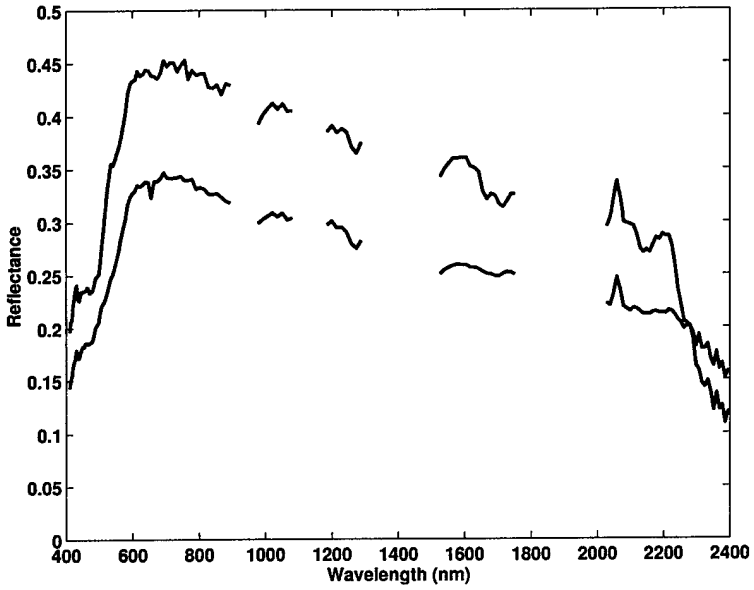


Figure 2. Spectra for two vehicles (water vapor bands removed).

than or equal to the sum of the distances between the first point and an intermediate point and the second point and the intermediate point. Equality is only met by choosing the intermediate point to reside on the line connecting the two points. This characteristic is an important property that will be highlighted in subsequent sections.

2.2 TWO DISTANCE METRICS

Consider the reflectance spectra for two different targets in Figure 2. Clearly, they are visually different in both shape and amplitude. The notion of similarity shared by the spectra can be measured differently, depending on the metric that is used to compare them. In this section, we discuss the two most prominent distance metrics in hyperspectral processing: the Spectral Angle Mapper (SAM) and the Euclidean Minimum Distance (EMD). Each metric provides a unique measure of distance from two complementary viewpoints of Euclidean geometry.

2.2.1 An Important Caveat About Hyperspectral Data

It is important to note that although the two spectra in Figure 2 are plotted on a two-dimensional plane, with reflectance as a function of wavelength, the two metrics do not perform their calculation in this plane. Instead of interpreting a reflectance spectrum as samples from a continuous parametric curve, $r(\lambda_i), i = 1, \ldots, M$, where M is the number of bands, SAM and EMD interpret the spectral reflectance values as coordinates for a vector in a high-dimensional space, $r(\lambda_1, \ldots, \lambda_M)$. For simplicity, the discussion of the metrics will refer to notional diagrams in three-dimensional space (see Figure 3). However, it is important to understand that SAM and

Metric	SAM	EMD
Equation	$\theta(\mathbf{x},\mathbf{y}) = \arccos(\frac{<\mathbf{x},\mathbf{y}>}{\|\mathbf{x}\|\|\mathbf{y}\|})$	$\Delta(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|$
Values	$0 \leq \theta \leq \frac{\pi}{2}$	$0 \leq \Delta < \infty$
Invariance	Multiplicative scaling	Rotational
Additivity	No	Yes
Monotonicity	No	Yes

TABLE 1. Summary of properties of SAM and EMD

EMD perform their comparison in a high-dimensional environment having as many dimensions as spectral bands and representations such as Figure 2 are convenient for illustration purposes. Metrics that compare spectra as parametric curves ($r(\lambda_i)$) are a different area of spectral analysis, which are not considered in this report.

2.2.2 Spectral Angle Mapper (SAM)

Figure 3(a) depicts a pair of three-dimensional spectra and indicates the angle, θ, created by them that SAM quantifies. For two M-dimensional spectra, $\mathbf{x}$ and $\mathbf{y}$, θ is given by the following analytical expression:

$$\theta(\mathbf{x},\mathbf{y}) = \arccos(\frac{<\mathbf{x},\mathbf{y}>}{\|\mathbf{x}\|\|\mathbf{y}\|}), \qquad 0 \leq \theta \leq \frac{\pi}{2} \tag{1}$$

where $<\cdot,\cdot>$ is the dot product operator, and $\|\cdot\|$ is the 2-norm which may be written using the dot product operator as $\sqrt{<\cdot,\cdot>}$ [15]. From its mathematical definition in (1), SAM possesses unique properties that distinguish it from EMD. These are enumerated here.

Invariance to Multiplicative Scaling: The angle measured by SAM is invariant to multiplication of $\mathbf{x}$ and $\mathbf{y}$ by scalars, $a, b \in \Re$:

$$\begin{aligned} \theta(a\mathbf{x}, b\mathbf{y}) &= \arccos(\frac{<a\mathbf{x}, b\mathbf{y}>}{\|a\mathbf{x}\|\|b\mathbf{y}\|}) \\ &= \arccos(\frac{ab<\mathbf{x},\mathbf{y}>}{ab\|\mathbf{x}\|\|\mathbf{y}\|}) \\ &= \theta(\mathbf{x},\mathbf{y}). \end{aligned} \tag{2}$$

This property is apparent by examination of Figure 3(a). Multiplication of a vector by a scalar simply increases its extent in a particular direction, but it does not alter the angle it creates with another vector.

What impact does this invariance have on hyperspectral processing? Although all objects have a distinct reflectance spectrum, the recovery of accurate reflectance estimates from hyperspectral measurements can be complicated by numerous factors. Atmospheric compensation is

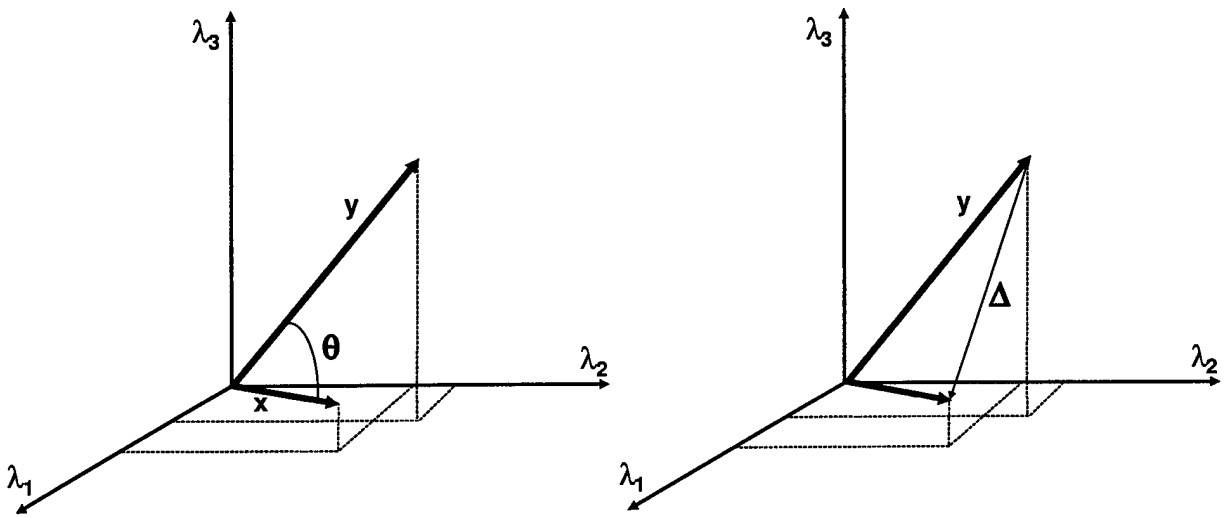


Figure 3. (a) Spectral Angle Mapper (SAM), θ, (b) Euclidean Minimum Distance (EMD), Δ.

the procedure that is applied to the radiance measurements collected by a hyperspectral sensor to recover the intrinsic reflectance values for each pixel in a scene. However, reflectance estimates that are recovered using atmospheric compensation algorithms such as ATREM can only estimate spectra within a multiplicative constant. In order to obtain the real reflectance value, knowledge of the terrain slopes and aspect angles with respect to the sensor must be known so that the effective surface area seen by a sensor within a pixel is known exactly [13].

This type of uncertainty, or variability, can be seen in Figure 4. Plotted in blue are the spectra of pixels taken from the same vehicle in a single scene when imaged by the HYDICE sensor. Plotted in red are a collection of reference measurements taken from the same target at close range by a hand-held spectroradiometer. There is a significant variation in both, but the reflectance spectra seen by the HYDICE sensor show a distinct variability that resembles an unknown multiplicative scaling. The variability evident in Figure 4 can sabotage an automated recognition system unless it is designed to be invariant to such behavior. Thus, if SAM is used to classify an unknown pixel as belonging to one of many reference classes, its invariance to multiplicative (or near-multiplicative) scaling is a considerable benefit in light of the real-world behavior of hyperspectral signals.

Non-Additivity: Another important property possessed by SAM is that it is a *non-additive* distance metric. To explore this subject, we introduce a useful definition [9].

Definition: Non-Additive Distance Metric. *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be two length-M vectors in* $\Re^M$. *Let the elements of* $\mathbf{x}$ *and* $\mathbf{y}$ *be partitioned in such a way that* $\mathbf{x} = [\mathbf{x}_a \ \mathbf{x}_b]$ *and* $\mathbf{y} = [\mathbf{y}_a \ \mathbf{y}_b]$, *where* $M = a + b$ *and* $\mathbf{x}_a, \mathbf{y}_a \in \Re^a$ *and* $\mathbf{x}_b, \mathbf{y}_b \in \Re^b$. *Then, a distance metric,* $d(\cdot, \cdot)$, *is non-additive when* $d(\mathbf{x}, \mathbf{y}) \neq d(\mathbf{x}_a, \mathbf{y}_a) + d(\mathbf{x}_b, \mathbf{y}_b)$.

SAM is a non-additive distance metric. This may be demonstrated by using the same termi-

nology from the definition stated above. Letting $\theta(\mathbf{x},\mathbf{y})$ correspond to the SAM angle using all the elements in $\mathbf{x}$ and $\mathbf{y}$, we can re-express (1) in terms of the angle between $\mathbf{x}_a$ and $\mathbf{y}_a$, θ_a, and $\mathbf{x}_b$ and $\mathbf{y}_b$, θ_b.

$$\begin{aligned}
\cos\theta(\mathbf{x},\mathbf{y}) &= \frac{<\mathbf{x},\mathbf{y}>}{||\mathbf{x}||\,||\mathbf{y}||} \\
&= \frac{<\mathbf{x}_a,\mathbf{y}_a> + <\mathbf{x}_b,\mathbf{y}_b>}{\sqrt{||\mathbf{x}_a||^2+||\mathbf{x}_b||^2}\sqrt{||\mathbf{y}_a||^2+||\mathbf{y}_b||^2}} \\
&= \frac{<\mathbf{x}_a,\mathbf{y}_a>}{||\mathbf{x}_a||||\mathbf{y}_a||}\frac{1+\frac{<\mathbf{x}_b,\mathbf{y}_b>}{<\mathbf{x}_a,\mathbf{y}_a>}}{\sqrt{1+\frac{||\mathbf{x}_b||^2}{||\mathbf{x}_a||^2}}\sqrt{1+\frac{||\mathbf{y}_b||^2}{||\mathbf{y}_a||^2}}} \\
&= \cos\theta_a\frac{1+\frac{<\mathbf{x}_b,\mathbf{y}_b>}{<\mathbf{x}_a,\mathbf{y}_a>}}{\sqrt{1+\frac{||\mathbf{x}_b||^2}{||\mathbf{x}_a||^2}}\sqrt{1+\frac{||\mathbf{y}_b||^2}{||\mathbf{y}_a||^2}}} \qquad (3)
\end{aligned}$$

The consequence of this expansion is that $\cos(\theta)$ is expressible as a function of $\cos(\theta_a)$ and a multiplicative factor which is a function of $(\mathbf{x}_a,\mathbf{y}_a)$ as well as $(\mathbf{x}_b,\mathbf{y}_b)$. Clearly, $\theta(\mathbf{x},\mathbf{y}) \neq \theta(\mathbf{x}_a,\mathbf{y}_a)+\theta(\mathbf{x}_b,\mathbf{y}_b)$, and, therefore, SAM is a non-additive distance metric.

Non-Monotonicity: Another property of SAM is that it is a *non-monotonic* distance metric. We again introduce another definition that is applicable to distance metrics.

Definition: Monotonic Distance Metric. *A distance metric, $d(\cdot,\cdot)$, is monotonic if its value must increase monotonically as the dimension of its operands, $\mathbf{x}$ and $\mathbf{y}$, increase.*

By examining (3), it is clear that the right term on the right-hand side may be greater or less than one, depending on the values in x_a, y_a, x_b, y_b. Thus, the addition of more spectral bands does not always guarantee an increase in angular separability. Hence, SAM is a non-monotonic distance metric. In conjunction with the fact that SAM is also non-additive, the fact that SAM is non-monotonic will also be exploited for band selection in later sections of this report.

2.2.3 Euclidean Minimum Distance (EMD)

In contrast, Figure 3(b) shows that EMD measures the shortest distance between two vectors, $\mathbf{x},\mathbf{y} \in \Re^M$, and is defined as

$$\begin{aligned}
\Delta(\mathbf{x},\mathbf{y}) &= \|\mathbf{x}-\mathbf{y}\| \\
&= \sqrt{\sum_{i=1}^{M}(x_i-y_i)^2}. \qquad (4)
\end{aligned}$$

From the definition of EMD in (4), EMD possesses properties that make it distinct from SAM.

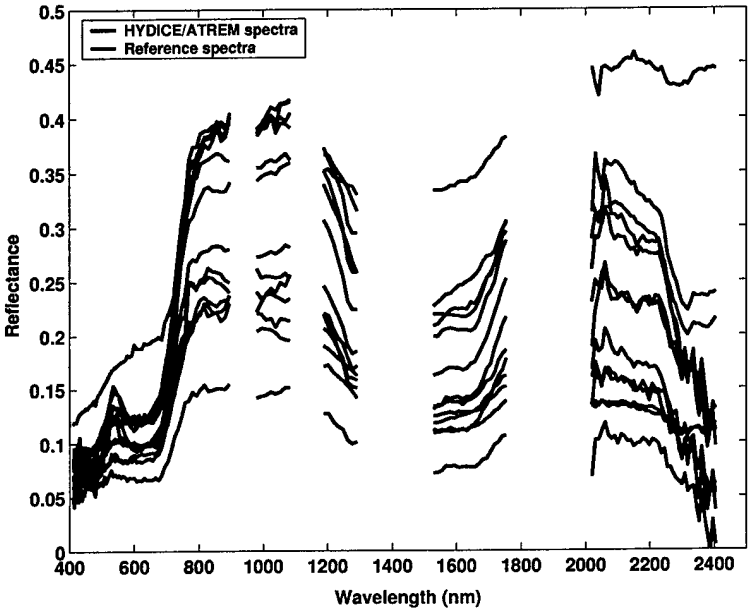


Figure 4. Spectra of pixels derived from the same vehicle in one scene imaged by the HYDICE sensor (blue). Reference measurements made of the same target by a hand-held spectroradiometer (red).

Invariance to Unitary Coordinate Transformation: Given a unitary $M \times M$ matrix, U_M, $\Delta(\mathbf{Ux}, \mathbf{Uy}) = \Delta(\mathbf{x}, \mathbf{y})$.

When does this invariance become useful in hyperspectral processing? Coordinate transformations in the spectral domain do not occur naturally in normal hyperspectral imaging environments. Moreover, a unitary transformation of coordinates can lead to negative values in the transformed domain ($\mathbf{Ux}$), which is impossible for reflectance and radiance values.

Additivity: Although Δ is not an additive distance metric, Δ^2 is an additive cost metric. Let $\mathbf{x}$ and $\mathbf{y}$ be two vectors in $\Re^M$. Let the elements of $\mathbf{x}$ and $\mathbf{y}$ be partitioned in such a way that $\mathbf{x} = [\mathbf{x}_a \ \ \mathbf{x}_b]$ and $\mathbf{y} = [\mathbf{y}_a \ \ \mathbf{y}_b]$, where $M = a + b$ and $\mathbf{x}_a, \mathbf{y}_a \in \Re^a$ and $\mathbf{x}_b, \mathbf{y}_b \in \Re^b$. Then Δ^2 can be decomposed as

$$\Delta^2(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{M} (x_i - y_i)^2 \tag{5}$$

$$= \sum_{i \in a} (x_i - y_i)^2 + \sum_{i \in b} (x_i - y_i)^2 \tag{6}$$

$$= \Delta^2(\mathbf{x}_a, \mathbf{y}_a) + \Delta^2(\mathbf{x}_b, \mathbf{y}_b). \tag{7}$$

Monotonicity: By examining (4) it is evident that an addition of bands to $\mathbf{x}$ and $\mathbf{y}$ cannot decrease EMD. In other words, additional, non-zero spectral bands necessarily lead to an increase in the distance metric. Therefore, EMD is monotonic.

2.3 SAM AND EMD IN HYPERSPECTRAL PROCESSING

Most algorithms for detection, classification, and unmixing utilize SAM or EMD as the metric that compares two spectra. Previous efforts to develop hierarchical taxonomies of algorithms for hyperspectral processing demonstrate that the metric utilized by an algorithm is a prominent feature that discriminates one class of algorithms from another [24, 29]. In this section we discuss several important application algorithms in hyperspectral processing and the distance metrics upon which they are built.

Before proceeding, we review a model that is frequently used to describe the synthesis of a single pixel from distinct endmembers in the scene. The equation for the linear mixing model (LMM) is given by:

$$\mathbf{x} = \sum_{k=1}^{P} a_k \mathbf{s}_k + \mathbf{w} = \mathbf{S}\mathbf{a} + \mathbf{w} \tag{8}$$

where $\mathbf{x}$ is the $M \times 1$ received pixel spectrum vector, $\mathbf{s}_k$ is the k-th $M \times 1$ column of $\mathbf{S}$, $\mathbf{a}$ is the $P \times 1$ fractional abundance vector, $\mathbf{w}$ is the $M \times 1$ additive observation noise vector, $\mathbf{S}$ is the $M \times P$ matrix of endmembers whose columns are $\mathbf{s}_k$, M is the number of spectral bands, and P is the number of endmembers. The two constraints imposed on $\mathbf{a}$ are (1) $\sum_{i=1}^{P} a_i = 1$, and (2) $a_i \geq 0, i = 1, \ldots, P$.

2.3.1 Spectral Unmixing

In spectral unmixing the objective is to estimate endmembers and abundances from a mixed pixel. The procedure consists of three steps:

Dimension reduction: Reduce the dimension of the data in the scene. This step is optional and is only invoked by some algorithms to reduce the computational load of subsequent steps.

Endmember determination: Estimate the distinct spectra, or endmembers, that constitute the mixed pixels in the scene.

Inversion: Estimate the fractional abundances of each mixed pixel from its spectrum and the endmember spectra.

There are numerous algorithms in the literature that perform one or more of these stages. For dimension reduction, principal component analysis (PCA) applies an eigendecomposition to the covariance of a set of pixels to identify the orthogonal axes in $\Re^M$ where most of the energy in the pixels resides. A key property of the resulting eigenvector and eigenvalue pairs, $\{\mathbf{u}_i.\lambda_i\}, i = 1, \ldots, M$, is that they are ordered ($i = 1, \ldots, M$) so that any truncated subset provides the best average

approximation of a pixel having the same statistics $(\mu, \mathbf{\Gamma})$, and this approximation is measured by EMD. Let $\hat{\mathbf{x}}$ be a Q-term $(Q < M)$ approximation to $\mathbf{x}$ given by

$$\hat{\mathbf{x}} = \mu + \sum_{i=1}^{Q} \mathbf{u}_i < \mathbf{x} - \mu, \mathbf{u}_i > . \tag{9}$$

Then, the average error in the approximation, $E[||\mathbf{x}-\hat{\mathbf{x}}||]^2$, is minimized over all possible collections of Q vectors. Hence, PCA possesses properties directly linked to the optimization of EMD [44].

Another method of dimension reduction that has been developed for use with real-time hyperspectral data collection platforms is part of the Naval Research Laboratory's ORASIS (Optical Real-time Adaptive Spectral Identification System), which is a series of hyperspectral processing modules [7, 8]. In the Exemplar Selector Module (ESM), when a new pixel is collected from the scene by the sensor, its spectra is compared to the existing set of exemplars (the first pixel in a scene automatically becomes the first exemplar). This comparison is performed by SAM, and if the new pixel exceeds an angular threshold with every exemplar, it is added to the collection. This set of exemplars is periodically orthogonalized to yield a basis which is used to reduce the dimension of the data before it undergoes further analysis. Similar to PCA, where EMD plays a prominent role, SAM is the distance metric that is used to regulate the admission of pixels into the set of exemplars.

Furthermore, we can identify an important class of unmixing algorithms where EMD plays a prominent role. Many inversion algorithms estimate abundances that minimize a least-squares criterion [24, 28]. Given a mixed pixel, $\mathbf{S}$, and a set of endmembers organized in a matrix, $\mathbf{S}$, least-squares-based inversion algorithms estimate a set of abundances, $\hat{\mathbf{a}}$, that minimize $||\mathbf{x} - \mathbf{S}\hat{\mathbf{a}}||$. Once again, EMD plays the role of comparing the estimated pixel, $\mathbf{S}\hat{\mathbf{a}}$, with the original pixel, $\mathbf{x}$.

2.3.2 Classification

A frequently used tool for classifying pixels in a scene into distinct, homogeneous classes is clustering. For example, algorithms based on K-means clustering identify natural partitions in data based on distinct statistical behavior. The fundamental instrument for assigning and re-assigning pixels to classes is the measurement of pairwise similarity between pixels and the centroids of each class. This measurement is frequently performed by EMD, and numerous variations on this form of clustering have been attempted, all using a distance metric based on EMD [6, 11].

2.3.3 Statistical Target Detection

Finally, we can examine a large suite of statistical detection algorithms that have been proposed to detect targets. These algorithms detect targets occurring with low probability amid background and provide the basis for estimates of probabilities of detection (P_D) and false alarm (P_{FA}) as a function of relevant operating parameters (e.g., number of bands, SNR, etc.). The mathematical expressions for the detectors originate from formulations of a binary hypothesis test for a single pixel, where one hypothesis, $H0$, assumes no target is present, and the other hypothesis,

$H1$, supposes a target exists,

$$\begin{aligned} H0: \quad \mathbf{x} &= \mathbf{S}_b\mathbf{a}_b + \mathbf{w} \\ &= \mathbf{v} \end{aligned} \tag{10}$$

$$\begin{aligned} H1: \quad \mathbf{x} &= \mathbf{s}_t + \mathbf{S}_b\mathbf{a}_b + \mathbf{w} \\ &= \mathbf{s}_t + \mathbf{v}. \end{aligned} \tag{11}$$

Here, $\mathbf{S}_b$ is a matrix of background endmembers, $\mathbf{a}_b$ is the vector of background abundances, and $\mathbf{s}_t$ is the desired target spectrum. The additive noise is given by $\mathbf{w}$ and in most cases is assumed to be Gaussian.

Three common detectors for unstructured backgrounds [29] are the Generalized Likelihood Ratio Test (GLRT) [21], the Adaptive Coherence Estimator (ACE) [26, 27], and the Adaptive Matched Filter (AMF) [34]. For a desired target signature, $\mathbf{s}_t$, the GLRT is given by

$$T_{GLRT}(\mathbf{x}) = \frac{|\mathbf{s}_t^T\hat{\mathbf{\Gamma}}_v^{-1}\mathbf{x}|^2}{(\mathbf{s}_t^T\hat{\mathbf{\Gamma}}_v^{-1}\mathbf{s}_t)(1+\mathbf{x}^T\hat{\mathbf{\Gamma}}_v^{-1}\mathbf{x})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{12}$$

Here, $\hat{\mathbf{\Gamma}}_v^{-1}$ is the inverse of the non-normalized estimated covariance from N pixels that are presumed to be target-free and zero-mean.

$$\hat{\mathbf{\Gamma}}_v = \sum_{n=1}^{N} \mathbf{x}(n)\mathbf{x}(n)^T. \tag{13}$$

If we perform a symmetric factorization, we can write $\mathbf{\Gamma}_v^{-1} = \mathbf{W}^T\mathbf{W}$ and rewrite (12) as

$$\begin{aligned} T_{GLRT}(\mathbf{x}) &= \frac{|\mathbf{s}_t^T\mathbf{W}^T\mathbf{W}\mathbf{x}|^2}{(\mathbf{s}_t^T\mathbf{W}^T\mathbf{W}\mathbf{s}_t)(1+\mathbf{x}^T\mathbf{W}^T\mathbf{W}\mathbf{x})} \\ &= \frac{|<\mathbf{W}\mathbf{s}_t, \mathbf{W}\mathbf{x}>|^2}{(<\mathbf{W}\mathbf{s}_t, \mathbf{W}\mathbf{s}_t>)(1+<\mathbf{W}\mathbf{x}, \mathbf{W}\mathbf{x}>)}. \end{aligned} \tag{14}$$

We can see from (14) that the GLRT closely resembles the form of the mathematical definition of $\cos^2\theta$ in (1). The notable difference is that the spectra being compared, $\mathbf{x}$ and $\mathbf{s}_t$, are transformed by the $M \times M$ matrix, $\mathbf{W}$. The other difference is the second term in the denominator, which contains a one in addition to the inner product. This accounts for the fact that the estimate of the covariance that gives rise to $\mathbf{W}$ increases in accuracy as $N \to \infty$. If $N \to \infty$, this term approaches unity, and the GLRT reduces to

$$T_{AMF}(\mathbf{x}) = \frac{|\mathbf{s}_t^T\hat{\mathbf{\Gamma}}_v^{-1}\mathbf{x}|^2}{(\mathbf{s}_t^T\hat{\mathbf{\Gamma}}_v^{-1}\mathbf{s}_t)}. \tag{15}$$

Under the condition that $N \to \infty$, the expression in (15) yields a detector known as the Adaptive Matched Filter (AMF). If, however, N is relatively small, the inner product in the second term of

the denominator in (14) will dominate, resulting in the following detector:

$$T_{ACE}(\mathbf{x}) = \frac{|<\mathbf{W}\mathbf{s}_t, \mathbf{W}\mathbf{x}>|^2}{(<\mathbf{W}\mathbf{s}_t, \mathbf{W}\mathbf{s}_t>)(<\mathbf{W}\mathbf{x}, \mathbf{W}\mathbf{x}>)}. \tag{16}$$

This detector is known as the Adaptive Coherence/Cosine Estimator (ACE), and without unity in the denominator of (12), (16) is clearly the squared-cosine of the angle created by whitened versions of $\mathbf{s}_t$ and $\mathbf{x}$. Further, in the unlikely case that the background covariance is uncorrelated, i.e., $\hat{\mathbf{\Gamma}}_v = \mathbf{I}$, then $\mathbf{W} = \mathbf{I}$, and (16) simplifies to a pure measure of the squared-cosine,

$$T_{SAM}(\mathbf{x}) = \frac{|<\mathbf{s}_t, \mathbf{x}>|^2}{<\mathbf{s}_t, \mathbf{s}_t><\mathbf{x}, \mathbf{x}>}. \tag{17}$$

Hence, many of the common statistical detectors used for target detection in hyperspectral processing are built from the measurement of the angle between a test pixel and a reference signature spectrum. The presence of the estimated background covariance induces a coordinate stretching and rotation, but the essential measurement is still angular. Furthermore, the range of values produced by $T_{GLRT}(\mathbf{x}), T_{ACE}(\mathbf{x})$, and $T_{SAM}(\mathbf{x})$ is identical, $[0, 1]$.

2.4 OTHER DISTANCE METRICS

We have introduced concepts from mathematics for the purpose of specifically defining what properties a distance metric should have. For hyperspectral processing, two metrics that obey these properties have been been borrowed from Euclidean geometry, SAM and EMD. Both have relatively intuitive interpretations, and we have shown that these metrics provide the cornerstone for a large number of algorithms in hyperspectral processing.

The obvious question is whether there are other functions that are useful for comparing two spectra in hyperspectral processing? The answer is "yes", but they have not gained as much acceptance and credibility as SAM or EMD. This may be because the function is not as intuitive, or because it does not have any physically meaningful properties, as SAM does. In other cases, distance measures have been proposed, but they do not meet the criteria for a metric. For instance, the I-divergence compares two non-negative deterministic functions and yields a non-negative measure of dissimilarity having a value of zero when the functions are identical [40]. For two non-negative vectors of length M, $\mathbf{x}$ and $\mathbf{y}$, the I-divergence is defined by

$$I(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{M} x_i \ln\left[\frac{x_i}{y_i}\right] - \sum_{i=1}^{M}(x_i - y_i). \tag{18}$$

Simple substitution reveals that $I(\mathbf{x}, \mathbf{y}) \neq I(\mathbf{y}, \mathbf{x})$ and hence, I is not a distance metric. However, this fact alone does not disqualify it from being useful. The three properties of a distance metric are desirable, but not absolutely necessary.

While distance metrics provide the foundation for mathematically establishing the distance between two spectra, there are also numerous statistical measures that quantify the distance between two classes of spectra, when the intra-class variability is expressed using a covariance. In

their own way, these statistical measures of distance between two classes assign a scalar measure of similarity to two classes. Numerous distance measure have been exploited in all areas of signal processing, pattern recognition, and cognitive science [4]. For example, the Bhattacharyya coefficient [20], ρ, is often cited as a measure of the similarity between two Gaussian classes, $\mathbf{X}_1$ and $\mathbf{X}_2$, each defined by means and covariances, (μ_1, Σ_1) and (μ_2, Σ_2), respectively. It is a special case of the Chernoff measures that provide upper and lower bounds on the probability of error when classifying signals originating with equal probability from both classes:

$$\rho(\mathbf{X}_1, \mathbf{X}_2) = e^{-B} \tag{19}$$

$$B = \frac{1}{8}(\mu_1 - \mu_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1}(\mu_1 - \mu_2) + \frac{1}{2}\ln\frac{\left|\frac{\Sigma_1+\Sigma_2}{2}\right|}{\sqrt{|\Sigma_1||\Sigma_2|}}. \tag{20}$$

From (19), $0 < \rho \leq 1$. Since $\rho = 1$ when $\mathbf{X}_1 = \mathbf{X}_2$, the Bhattacharyya coefficient does not satisfy the triangle inequality, but $\sqrt{1-\rho}$ does. Although there is a similar geometric logic to statistical distance measures as there is to distance metrics for deterministic signals, we will not discuss statistical measures of distance in this report.

3. METRIC-BASED BAND SELECTION

We begin this section by discussing how a priori knowledge of phenomenology has influenced the design requirements for current sensors. Hyperspectral sensors are designed to collect measurements in intervals where exploitable phenomenology exists. For example, sensors collecting data for ocean color remote sensing do not collect data beyond approximately 800nm because there is little, if any, reflected light at higher wavelengths. Likewise, similar arguments for the detection of mixed gases and camouflaged targets have driven the spectral requirements for sensors (e.g., lowest wavelength, highest wavelength, spectral resolution). Motivated by this knowledge of physics, the entire spectral range of collected data is often processed by algorithms with little, if any, consideration to the mathematical properties of the algorithms.

In Section 2, we introduced two distance metrics, SAM and EMD, and enumerated their properties, highlighting the places where they are similar and different. In contrast, we consider an opposite methodology for processing hyperspectral data. Recalling that most hyperspectral algorithms are based on one of two metrics, we investigate the idea of processing only specific subsets of bands collected by a sensor (and hence, rejecting the remainder), based purely on the optimization of the distance metric being utilized. As a benchmark, we compare the performance of the subset of bands to identical processing performed using every available band.

3.1 PHYSICS-BASED BAND SELECTION

As discussed in Section 1.2, virtually all applications capitalize on exposing some type of contrast that exists between (at least) two classes of signals. Knowledge of where contrast resides, spectrally, and what spectral resolution is necessary to reveal it, can drive the design specifications of sensors. For example, ocean color remote sensing collects measurements in spectral regions where the relevant physical processes are observable. The Sea-viewing Wide Field-of-view Sensor (SeaWiFS) sensor developed by NASA and launched in 1997 was designed to provide quantitative data on global ocean bio-optical properties to the Earth science community [18]. The sensor has eight channels, spanning from 402nm to 885nm with channel widths ranging between 20nm and 40nm. The bands do not provide complete coverage between 402nm to 885nm, and the bands do not overlap spectrally. They were chosen to capitalize on specific intervals of spectral activity due to pigments whose relative quantities can be correlated with the presence of phytoplankton.

Similarly, the recognizable spectral features of gaseous effluents in mixed gases are present principally in the longwave infrared, and because telltale absorption bands are relatively narrow, longwave sensors collect data in bands that are sufficiently narrow to allow accurate chemical "fingerprinting." As in the case of ocean sensors, specific knowledge of the phenomenology directly impacts the sensor requirements. Figure 5 demonstrates how disparate intervals of the electromagnetic spectrum are exploited for different applications.

In many cases, however, where the goal is to distinguish objects having different spectral properties, such specific physical knowledge is unavailable, although, visual inspection can easily see that the spectra are dissimilar. Figure 2 demonstrates this scenario. In such a situation, the

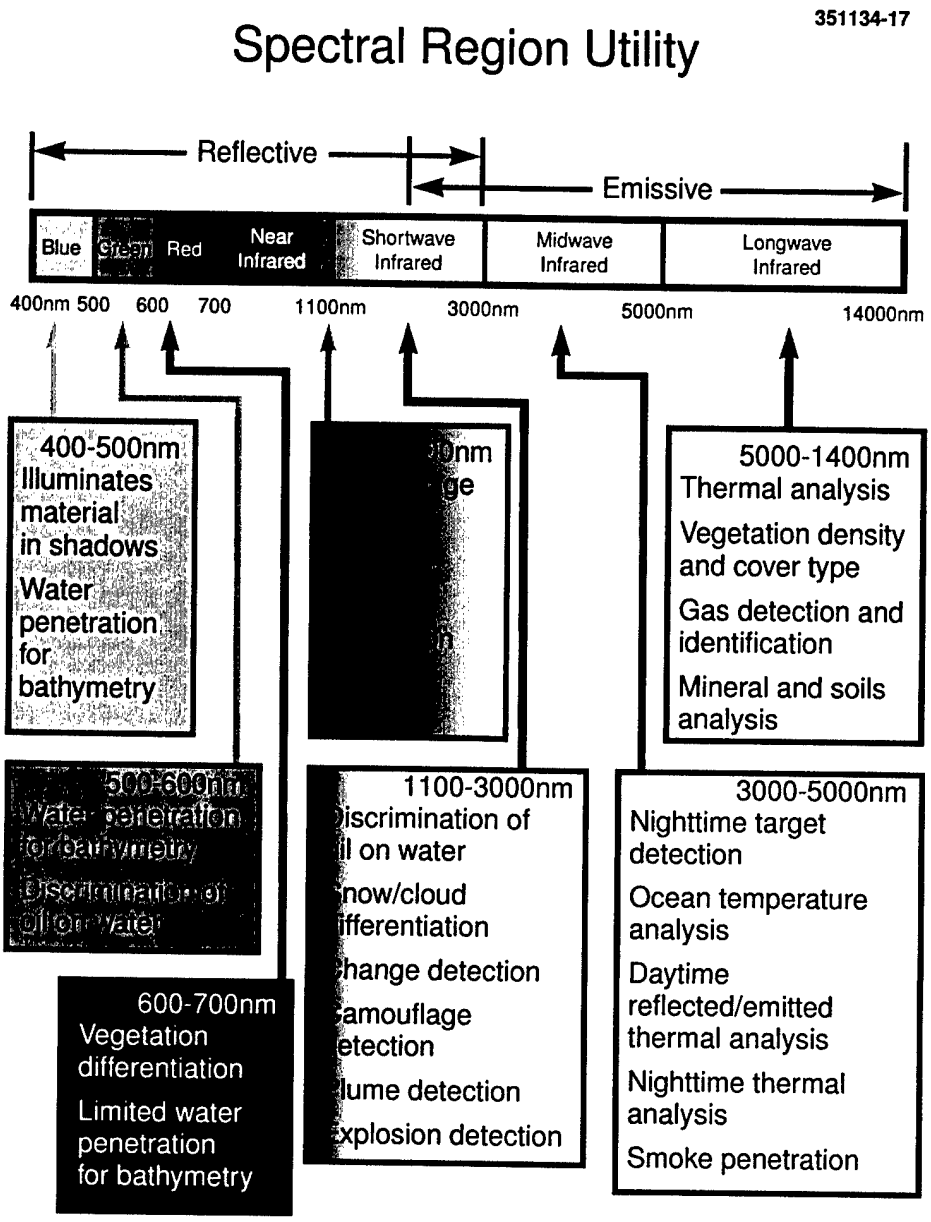


Figure 5. Hyperspectral applications and their associated spectral interval (courtesy SITAC).

correct choice of a distance metric is vital for several reasons. First, the metric determines what feature shared by the two spectra is to be scrutinized and quantified. If the distance metric is unable to capture the distinguishing feature, then the two spectra will not be distinguished. Conversely, the distance metric should ignore features that would incorrectly register a large distance if both spectra are from the same class. The latter case can be exemplified by SAM, which is invariant to multiplicative scaling, and can commonly arise in normal imaging scenarios.

3.2 BAND SELECTION TO INDUCE PHENOMENOLOGY

In contrast to physics-driven band selection, we now investigate selecting bands for processing based on their ability to optimize a cost function, namely the distance metrics discussed in Section 2, SAM and EMD. This approach works with the understanding that band centers and widths have already been determined by the sensor design. The goal is to optimize the distance metric that distinguishes two spectra by selecting only a subset of the available bands. As such, the distance metric is the foundation for identifying contrast that distinguishes two classes of signals. The greater the value of the distance metric, the more contrast that exists to be exploited.

We can consider a situation where this is useful. A binary test using a distance metric, $d(\cdot,\cdot)$, compares an unclassified pixel spectrum, $\mathbf{r}$, with template spectra representing two classes $\mathbf{t}_i, i = 1,2$, to determine which spectrum it more closely resembles. If, for example, $\mathbf{r}$ is from the first class, and there is no noise, $\mathbf{r}$ will exactly match one template, i.e., $d(\mathbf{r},\mathbf{t}_1) = 0$. However, in realistic scenarios, noise will prevent an exact match from occurring, and the distance between $\mathbf{r}$ and both template spectra will invariably be non-zero, i.e., $d(\mathbf{r},\mathbf{t}_1) > 0, d(\mathbf{r},\mathbf{t}_2) > 0$. However, the greater the contrast between $\mathbf{t}_1$ and $\mathbf{t}_2$, the more assurance that the binary test will not be corrupted by noise. Thus, optimizing the distance metric to yield greater contrast creates more robustness to distortions, and better application performance.

In Table 1, the properties of SAM and EMD are summarized and compared side-by-side. We can conclude, based on the monotonicity of EMD discussed in Section 2.2.3, that the contrast between two signals increases with the number of bands. Furthermore, the additivity of EMD confirms that the amount of contrast between two signals increases with additional bands independent of the value of other bands. In short, the greatest contrast between two spectra is necessarily achieved by using every band collected by the sensor.

SAM, however, is neither monotonic nor additive. This combination of properties indicate that the value of SAM, and hence the contrast derived from it, does not necessarily increase as more bands are added. In fact, by examining (3), it is clear that the value of SAM can either decrease or increase as more bands are added. This can be clearly demonstrated in the simple example of two length-3 vectors, $\mathbf{x}$ and $\mathbf{y}$, that are plotted in Figure 6 and given by

$$\mathbf{x} = [\,1\ \ 3\ \ 0\,],\ \ \mathbf{y} = [\,0\ \ 2\ \ 1\,]. \tag{21}$$

Using all three bands, $\theta(\mathbf{x},\mathbf{y}) = 31.95^{\circ}$. If however, we exclude the second band in both vectors, the two-element vectors become orthogonal, and the angle immediately goes to 90°. Table 2 summarizes the value of the resultant angle for all possible combinations of bands in (21).

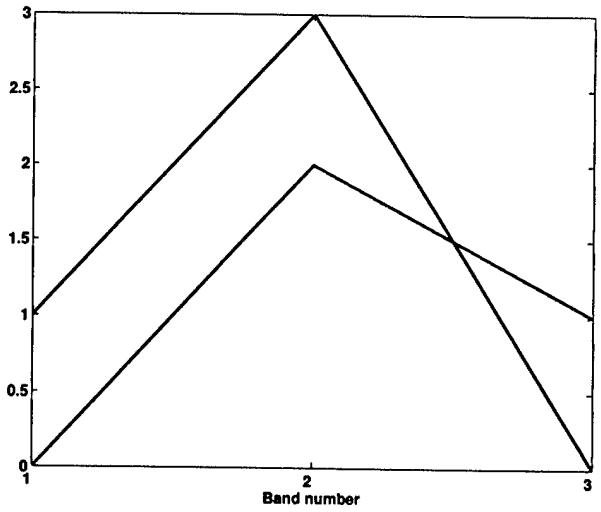


Figure 6. Plot of two three-dimensional spectra, $\mathbf{x} = [1\ 3\ 0]$ *and* $\mathbf{y} = [0\ 2\ 1]$.

Bands	Angle
All	31.95^{o}
{1,2}	18.43
{1,3}	90.00
{2,3}	26.57

TABLE 2. SAM angle value for all possible angles created by the vectors $\mathbf{x} = [1\ 3\ 0]$ *and* $\mathbf{y} = [0\ 2\ 1]$.

There are several implications that can be derived from the simple example summarized in Table 2.

- Different subsets of bands yield different angular values.
- Using all available bands does not necessarily provide the largest angular separation between two spectra.
- Spectral features that yield high angular separability are not immediately obvious from plots, such as Figure 6.

We will refer to an angle that is created from a subset of bands as a *sub-angle*, and we will refer to the angle created using all available bands as the *complete angle*. Thus the greatest sub-angle for $\mathbf{x}$ and $\mathbf{y}$ in Figure 6 is defined by bands 1 and 3.

Finally, the remainder of this report focuses on methods for band selection that optimize SAM. In addition to having a useful invariance to multiplicative scaling, SAM is the most widely used distance metric in hyperspectral processing. Like SAM, EMD provides a useful interpretation for distance, but its utility is limited for practical comparisons of spectra. The properties of non-monotonicity and non-additivity lead to a more complex mathematical interpretation, but the benefits will be shown to be worthwhile.

3.2.1 The Mathematics of M-Dimensions

In Section 3.2, we learned that sub-angles may exist that are larger than the complete angle between two spectra, possibly providing greater contrast and separability than the complete angle. The key to exploiting the capabilities described in the simple example in Figure 6 and Table 2 is to have a thorough understanding of the behavior of vector signals, or spectra, in high dimensions, or hyperspace. Despite the fact that most properties of angles and geometric surfaces in higher dimensions are straightforward extensions of concepts in two and three dimensions, the concepts are difficult, if not impossible, to visualize. Moreover, employing mathematical notation necessary to maintain clear and unambiguous bookkeeping is not simple. Nevertheless, the mathematics of high dimensions have been explored by several mathematicians and statisticians [22, 41].

3.2.2 Hyperspectral Data in High Dimensions

In Section 2.2.1 we stated an important caveat about the difference between interpreting a spectrum as a two-dimensional plot of radiance or reflectance, indexed by wavelength, versus a vector located by axes in a high-dimensional space by each reflectance or radiance value. We are exclusively retaining this latter interpretation of a spectrum for *all* subsequent calculations and results. For a pair of three-dimensional spectra, as given in Figure 6, there were three sub-angles consisting of two bands, and one complete angle consisting of three bands. Thus, the total number of possible angles is four.

Before generalizing the number of possible angles for a pair of M-dimensional signals, we should provide another definition. We will refer to a sub-angle created by k bands as a k-angle. For

M	1-angles	2-angles	5-angles	10-angles	Total number of sub-angles
10	10	45	252	1	1013
20	20	190	15504	184756	1048555
50	50	1225	2118760	1.027×10^{10}	1.126×10^{15}
100	100	4950	75287520	1.731×10^{13}	1.268×10^{30}
150	150	11175	591600030	1.170×10^{15}	1.427×10^{42}
200	200	19900	2.536×10^{9}	2.245×10^{16}	1.607×10^{60}
500	500	124750	2.552×10^{11}	2.456×10^{20}	3.273×10^{150}
1000	100	499500	8.25×10^{12}	2.634×10^{23}	1.072×10^{301}

TABLE 3. Number of k-angles and total number of sub-angles for different values of M.

a sub-angle to exist, there must be a minimum of two bands. Hence, provided two M-dimensional signals, $\mathbf{x}$ and $\mathbf{y}$, the total number of sub-angles is the number of unique, length-M, binomial sequences [35] (2^M) minus the number of sequences having 0 or 1 selected bands ($M+1$). The total number of unique k-angles, where $k = 2, \ldots, M$, is $\binom{M}{k} = \frac{M!}{(M-k)!k!}$. Consequently, the number of sub-angles, $l(M)$, (including the complete angle) is

$$l(M) = 2^M - (M+1). \tag{22}$$

Table 3 tabulates the total number of k-angles and sub-angles for different values of k and M based on (22). A typical hyperspectral sensor, such as HYDICE, has 210 total bands providing coverage from 400nm – 2500nm. Approximately 145 bands remain after bands corrupted by water vapor absorption are discarded. From Table 3, the number of sub-angles is approximately 10^{42}.

3.2.3 Strategies for Band Selection

The expression in (22) and the results in Table 3 demonstrate that the number of sub-angles between two spectra of average length is humongous, larger than would be rational for an exhaustive search of every combination of bands to find the biggest sub-angle. Nor, is there any known analytical solution for identifying the biggest sub-angle from two spectra. The alternative is to select bands sub-optimally, yet with the hope that the resulting sub-angle provides a superior capability to discriminate two spectra under realistically imperfect conditions.

We will discuss several types of band selection algorithms that maximize the angular separation between two spectra by exploiting the mathematical structure of SAM. No approach is guaranteed to identify the biggest possible angle, but the performance of sub-optimal band selection will be compared to the logical benchmark, which is the performance obtained using all bands (the complete angle). Only this will determine whether any form of band selection, optimal or sub-optimal, actually yields superior algorithm performance.

Inasmuch that band selection algorithms we investigate do not formally incorporate physical phenomenology, through the optimization of the distance metric they selectively *induce* phenomenology into the mathematical analysis by the bands they choose. Quite often the band selec-

tion will appear counter-intuitive, defying what would *appear* logical from two dimensional plots, such as Figure 2. The caveat of Section 2.2.1 should be remembered here. Thus, seen as an interface between the physics and mathematics, the optimization of distance metrics indirectly introduces physical aspects of the problem into the optimization of algorithm performance.

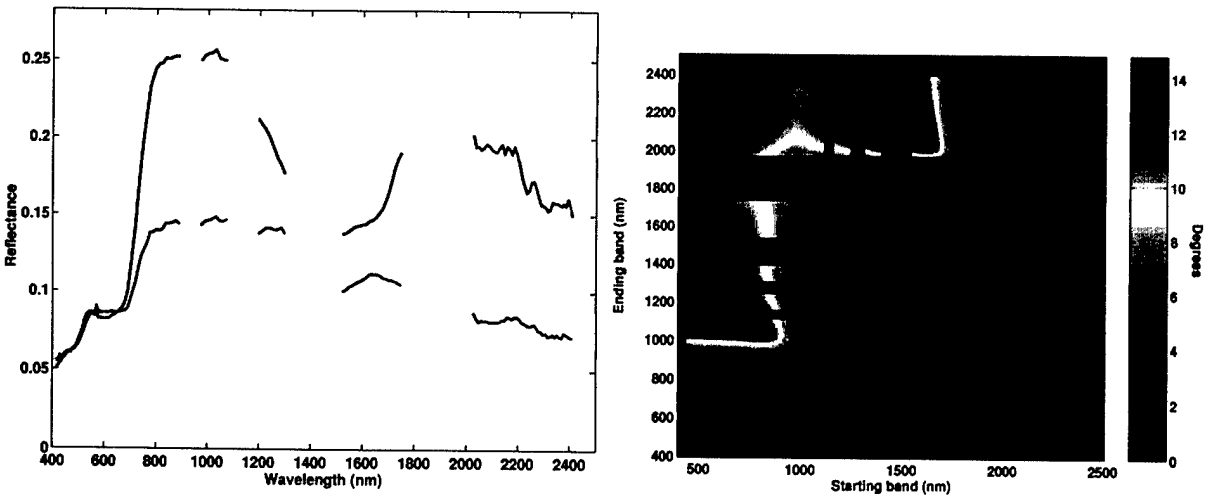
Before hyperspectral sensors existed, the challenge of band selection was posed for applications related to multispectral sensing where the bands are neither contiguous nor possess spectral resolution equivalent to that of hyperspectral sensors [32, 33, 38]. Nor was SAM employed as a distance metric. Several efforts to analyze the information content in multispectral data focused on combinatoric analysis of all possible band combinations [46]. It is understandable that these approaches focus less on streamlined search techniques and more on data analysis, since the number of bands is typically less than ten, and the physical information is relatively coarse and sparse. Having significantly more bands with greater resolution, hyperspectral data poses a more formidable challenge for which the exhaustive verification of these methods becomes impractical. Numerous methods of band selection have been proposed for hyperspectral data to achieve different ends with different measures of performance [3, 10, 16, 19, 39].

3.3 SINGLE CONTIGUOUS SEGMENTS

Our objective is to determine which bands maximize the SAM angle for two M-dimensional vectors, $\mathbf{x}$ and $\mathbf{y}$. The first approach for selecting bands somewhat mimics the physics-based perspective discussed in Section 3.1. We investigate the sub-angles generated by a contiguous segment of bands. Thus, this approach is an exhaustive, two-dimensional search for the starting wavelength and ending wavelength that demarcates a contiguous segment of spectral measurements that are a subset of the total measurements collected by the sensor. The method does not afford considering all $l(M) = 2^M - (M + 1)$ possible band combinations, rather, it investigates a significantly lower number, $(M - 1)(M - 2)/2$ possible solutions (for $M = 145$, this is 10296 solutions). Yet, the graphical insight it provides is a useful source of comparison for more sophisticated methods.

Let $\mathbf{x}$ and $\mathbf{y}$ be the spectra in Figure 7(a). The sub-angles produced for all valid pairs of starting and ending band pairs is illustrated in the two-dimensional map in Figure 7(b). At locations where the map is red, high sub-angles exist for contiguous segments of bands beginning at the starting wavelength and concluding at the ending wavelength. The largest sub-angle is 14.82^{o} and occurs over the interval [1269nm, 2184nm]. There are 39 bands in this interval. The smallest sub-angle is 0.002^{o} and occurs in the spectral interval [848nm, 862nm]. There are only 2 bands in this interval. The complete angle for the two spectra is 14.57^{o} and utilizes all 144 bands. Table 4 summarizes these results.

What does the increase in angular separation offer? The sub-angle created by the selected bands are a feature that should yield greater angular separation between two target classes. Given a test pixel that belongs to one of two classes, it may be compared to the reference spectrum from each class, using SAM, and assigned to the class generating the smallest angle. In reality, the unknown pixel will not identically match either reference spectrum due to numerous sources of noise and interference. However, increasing the angular separation between two classes can minimize the opportunity for misclassification when noise is present.

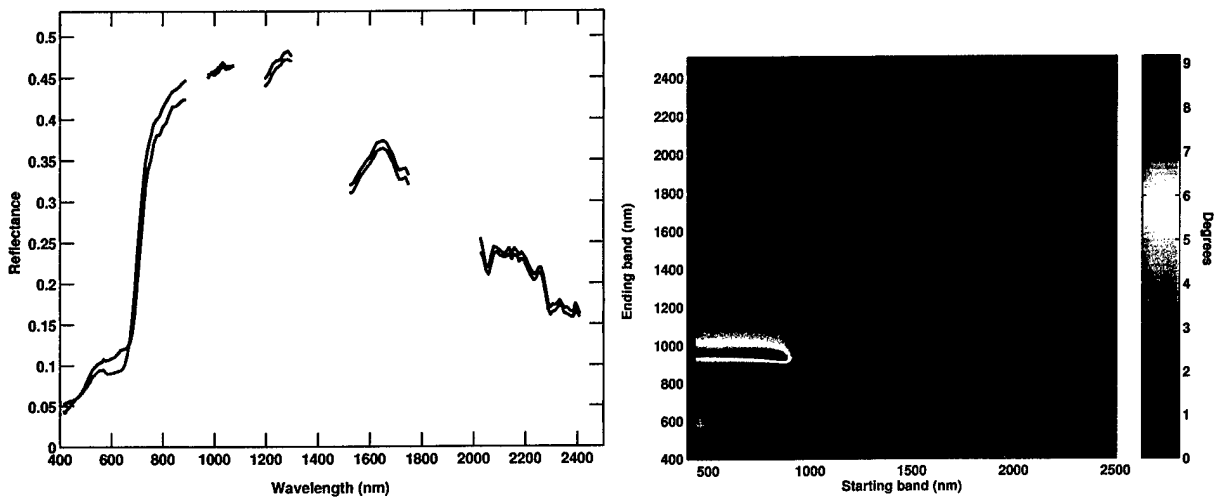


(a) (b)

Figure 7. (a) Plot of two spectra; (b) Two-dimensional contour map of sub-angles formed from all valid starting and ending band pairs.

	Starting wavelength (nm)	Ending wavelength (nm)	Number of bands	θ (^{o})
Largest sub-angle	1269	2184	39	14.82
Complete angle	412	2409	144	14.57
Smallest sub-angle	848	862	2	0.0002

TABLE 4. Largest sub-angle, complete angle, and smallest sub-angle for the two spectra in Figure 7(a) using one contiguous segment.



(a) (b)

Figure 8. (a) Plot of two green fabric spectra; (b) Two-dimensional contour map of sub-angles formed from all valid starting and ending band pairs.

	Starting wavelength (nm)	Ending wavelength (nm)	Number of bands	θ (^{o})
Largest sub-angle	596	699	12	9.18
Complete angle	412	2409	144	2.73
Smallest sub-angle	1561	1574	2	0.0060

TABLE 5. Largest sub-angle, complete angle, and smallest sub-angle for the two fabric spectra in Figure 8(a) using one contiguous segment.

The same procedure documented in Figure 7 and Table 4 is repeated for two fabrics which are different shades of green, as can be visually discerned by the difference in the two spectra in Figure 8. The results are tabulated in Table 5. The set of contiguous bands that maximize SAM span the interval from [596nm, 699nm]. The contour map in Figure 7(b) dramatically illustrates what segments yield a high sub-angle. Not surprisingly, by examining the spectra in Figure 8(a), there is a recognizable difference in the spectra near this interval that accounts for the slight difference in pigmentation.

3.4 MULTIPLE CONTIGUOUS SEGMENTS

Using the exhaustive search for a single contiguous segment as a foundation, we can extend the method of band selection to allow multiple, non-overlapping segments. The corresponding search, however, is no longer two-dimensional, as it was in Section 3.3. For each additional permitted

segment having a starting and ending wavelength, the degree of the search increases by two. This increases the number of admissible solutions and offers more flexibility in exploiting different parts of the spectrum. Unlike the single segment search, where a segment was required to have at least two bands, multiple intervals permit segments to consist of a single band. Unfortunately, with at least four degrees of freedom, a map cannot display the spectral intervals of interest. Moreover, the search for more than three segments using hyperspectral data of typical lengths becomes computationally unfeasible.

Using the same pair of spectra in Figure 7(a), Figure 9 illustrates the band selection, and Table 6 compares the results of the exhaustive search for one, two, and three non-overlapping contiguous segments. Similarly, using the same pair of spectra in Figure 8(a), Figure 10 illustrates the band selection, and Table 7 compares the results of the exhaustive search for one, two, and three non-overlapping, contiguous segments.

3.5 BAND ADD-ON (BAO)

In this section, we derive another algorithm for band selection that overcomes the limitations encountered when searching for contiguous segments of bands. We derive this algorithm directly from the mathematical definition for SAM, starting with the expansion in (3):

$$\cos\theta(\mathbf{x},\mathbf{y}) \;=\; \cos\theta_a \frac{1+\frac{<\mathbf{x}_b,\mathbf{y}_b>}{<\mathbf{x}_a,\mathbf{y}_a>}}{\sqrt{1+\frac{||\mathbf{x}_b||^2}{||\mathbf{x}_a||^2}}\sqrt{1+\frac{||\mathbf{y}_b||^2}{||\mathbf{y}_a||^2}}}. \tag{23}$$

As before, $\mathbf{x}$ and $\mathbf{y}$ are two length-M vectors in $\Re^M$, the elements of $\mathbf{x}$ and $\mathbf{y}$ are partitioned such that $\mathbf{x} = [\mathbf{x}_a \;\; \mathbf{x}_b]$ and $\mathbf{y} = [\mathbf{y}_a \;\; \mathbf{y}_b]$, where $M = a + b$ and $\mathbf{x}_a, \mathbf{y}_a \in \Re^a$ and $\mathbf{x}_b, \mathbf{y}_b \in \Re^b$. We will now exploit the right-most factor in (23).

3.5.1 The Geometry of β

In (23), $\cos\theta_a$ is the cosine of the angle created by $\mathbf{x}_a$ and $\mathbf{y}_a$:

$$\cos\theta_a = \frac{<\mathbf{x}_a,\mathbf{y}_a>}{||\mathbf{x}_a||\;||\mathbf{y}_a||}. \tag{24}$$

Then, $\cos\theta(\mathbf{x},\mathbf{y})$ can be decomposed as a function of the angle created by $\mathbf{x}_a$ and $\mathbf{y}_a$ scaled by another factor which involves the bands in $\mathbf{x}_b$ and $\mathbf{y}_b$ as well as $\mathbf{x}_a$ and $\mathbf{y}_a$. We will call this factor $\beta(\mathbf{x}_a,\mathbf{y}_a;\mathbf{x}_b,\mathbf{y}_b)$, or just β:

$$\beta(\mathbf{x}_a,\mathbf{y}_a;\mathbf{x}_b,\mathbf{y}_b) = \frac{1+\frac{<\mathbf{x}_b,\mathbf{y}_b>}{<\mathbf{x}_a,\mathbf{y}_a>}}{\sqrt{1+\frac{||\mathbf{x}_b||^2}{||\mathbf{x}_a||^2}}\sqrt{1+\frac{||\mathbf{y}_b||^2}{||\mathbf{y}_a||^2}}}. \tag{25}$$

The terms in β, however, can be further quantified. The first term in the denominator is the secant of the angle created by $\mathbf{x}_a$ and $\mathbf{x}$, $\sec\theta_{x,ab}$, and the second term is the comparable term for

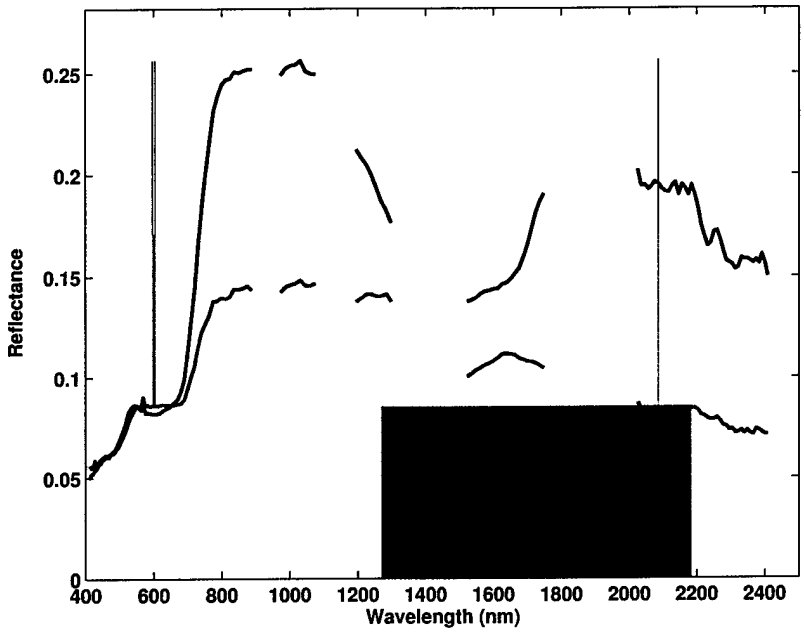


Figure 9. Single segment bands (green), double segment bands (cyan), triple segment bands (magenta) for spectra in Figure 7(a).

	Interval (nm)	Number of bands	θ (^{o})
Single segment	[1269, 2184]	39	14.82
Double segment	[596, 603] [2086, 2086]	3	25.55
Triple segment	[596, 596] [603, 603] [2086, 2086]	3	25.55
Complete angle	[412, 2409]	144	14.57

TABLE 6. Largest sub-angle and complete angle for the two spectra in Figure 7(a) using one, two, and three contiguous segments.

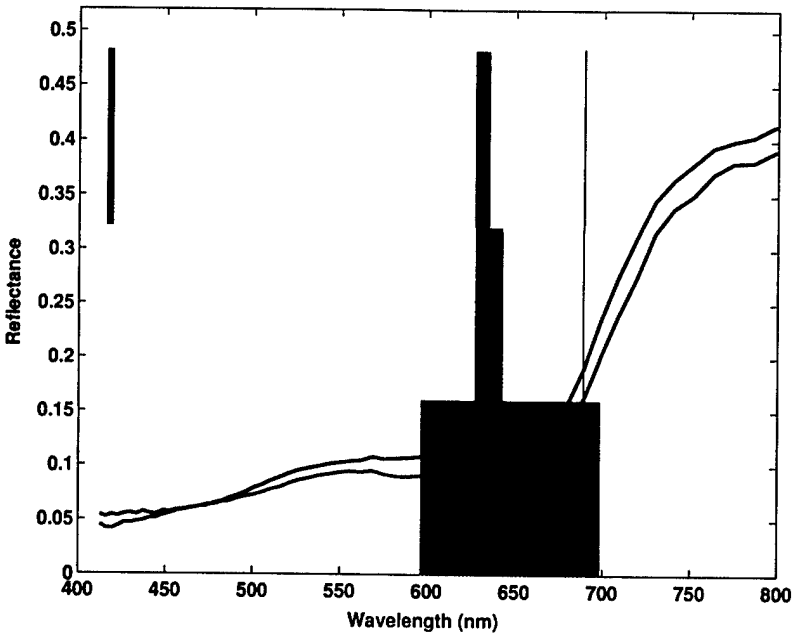


Figure 10. Single segment bands (green), double segment bands (cyan), triple segment bands (magenta) for spectra in Figure 8(a).

	Interval (nm)	Number of bands	θ (^{o})
Single segment	[596, 699]	12	9.18
Double segment	[627, 643] [689, 689]	4	10.80
Triple segment	[416, 419] [627, 635] [689, 689]	5	10.95
Complete angle	[412, 2409]	144	2.73

TABLE 7. Largest sub-angle and complete angle for the two spectra in Figure 10(a) using one, two, and three contiguous segments.

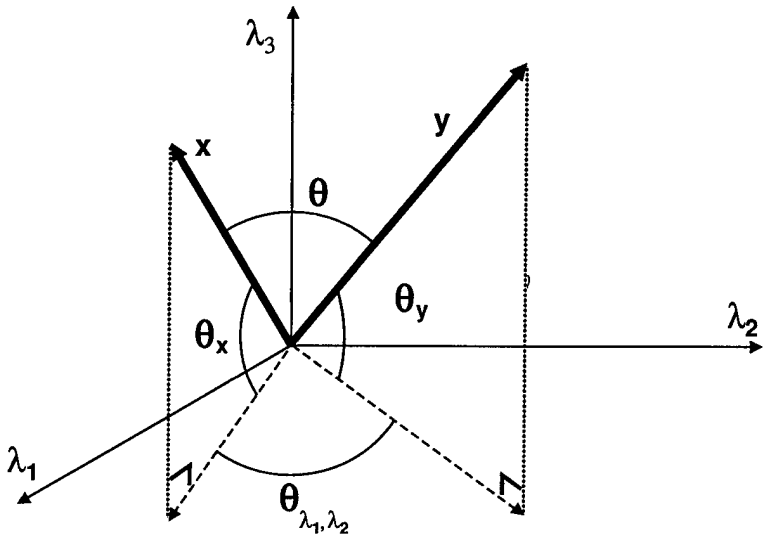


Figure 11. Relationship of sub-angles that comprise the complete angle between two spectra.

$\mathbf{y}$, $\sec\theta_{y,ab}$. Thus, β can be rewritten as

$$\beta(\mathbf{x}_a, \mathbf{y}_a; \mathbf{x}_b, \mathbf{y}_b) = \cos\theta_{x,ab} \cos\theta_{y,ab}(1 + \frac{< \mathbf{x}_b, \mathbf{y}_b >}{< \mathbf{x}_a, \mathbf{y}_a >}). \tag{26}$$

This revised expression for β demonstrates that, given $\mathbf{x}_a$ and $\mathbf{y}_a$, the addition of $\mathbf{x}_b$ and $\mathbf{y}_b$ changes $\cos\theta_a$ by the multiplicative terms in (26). Two of those terms, $\cos\theta_{x,ab}$ and $\cos\theta_{y,ab}$, are necessarily less than one, and separately measure the angular changes in $\mathbf{x}$ and $\mathbf{y}$. The third term is necessarily greater than one (hyperspectral signals are always non-negative) and interrelates the sets of previous and new values in $\mathbf{x}$ and $\mathbf{y}$. Thus, $\cos\theta$ may be greater than, less than, or equal to $\cos\theta_a$. Values of $\beta > 1$ will decrease the resulting angle, whereas values of $\beta < 1$ will perform the opposite. Using three dimensions, these relations are illustrated in Figure 11.

Given two spectra, $\mathbf{x}$ and $\mathbf{y}$, as well as a subset of their bands that serve as a starting point, one or more bands may be selected incrementally from the unused bands in $\mathbf{x}$ and $\mathbf{y}$ and appended to the existing set. Unused band(s) may be ranked by their associated value of β and the band(s) having the lowest value of β is added to the subset of selected bands. Then, $\cos\theta_a$ is re-evaluated with the new band and new values of β are calculated for the remaining unused bands. The process may be repeated iteratively, until a stopping condition is met. One logical criterion is when no remaining bands exist that yield a $\beta < 1$. This is equivalent to adding bands having $\beta < 1$ (i.e., that increase the angle between $\mathbf{x}$ and $\mathbf{y}$) until no bands remain having $\beta < 1$.

3.5.2 Initial Subset of Bands

We will refer to the initial subset of bands for $\mathbf{x}$ and $\mathbf{y}$ as $\mathbf{x}(\mathbf{B}_1)$ and $\mathbf{y}(\mathbf{B}_1)$, where $\mathbf{B}_1$ is a vector containing the associated band numbers. The initial subset of bands, $\mathbf{B}_1$, that begins the procedure can be chosen to meet different requirements. To gain insight on a proper choice, we

can examine the mathematical structure of β in (3). After selecting an initial subset of bands, we would like β to be as small as possible to broaden the initial angle. For this to occur, $\mathbf{B}_1$ should be selected to yield the following properties:

1. $< \mathbf{x}(\mathbf{B}_1), \mathbf{y}(\mathbf{B}_1) >$ should be as large as possible to minimize the numerator of β.
2. $||\mathbf{x}(\mathbf{B}_1)||^2$ and $||\mathbf{y}(\mathbf{B}_1)||^2$ should be as small as possible maximize the numerator of β.

Although, these two requirements are in conflict, many different criteria can be constructed based on some combination of both restrictions. This process is greatly simplified by limiting the size of $\mathbf{B}_1$ to two bands, i.e., the starting angle is a 2-angle. Repeated experiments revealed two starting conditions that yield results that are both useful, as well as, instructive. The first criterion exhaustively identifies the pair of bands that yield the *smallest* 2-angle for $\mathbf{x}$ and $\mathbf{y}$, heretofore referred to as BAO-MIN. The second criterion performs the opposite procedure, identifying the *biggest* 2-angle, heretofore referred to as BAO-MAX.

Based on this formulation, one possible algorithm for incrementally selecting bands can be listed as follows:

Step 1: Select a pair of starting bands. BAO-MIN and BAO-MAX are alternatives.

Step 2: Calculate β for each of the remaining bands.

Step 3: Of the bands having $\beta < 1$, select the band having the lowest β and add it to the set of selected bands. If no band has a $\beta < 1$, then quit. Otherwise go to Step 2.

The flowchart for selecting bands with BAO appears in Figure 12.

3.5.3 Experiments with Band Add-On

In order to demonstrate the BAO band selection technique discussed in Section 3.5, we again consider two spectra, plotted in Figure 13(a). Alongside, we also show the map of all possible 2-angles derived from these spectra in Figure 13(b). Although the map looks somewhat similar to the maps generated during the search for single contiguous segments in Section 3.3, they are very different. The maps in Section 3.3 reveal the SAM values for the bands *enclosed* by a starting and ending wavelength, whereas, the map in Figure 13(b) calculates the 2-angle using all unique pairs of wavelengths.

An analysis of the map in Figure 13(b) reveals that the maximum 2-angle is 29.98^{o} and occurs at $(1632\mathrm{nm}, 2051\mathrm{nm})$. The minimum 2-angle is 0.00012^{o} and occurs at $(684\mathrm{nm}, 757\mathrm{nm})$. Using either of these pair of starting bands, we can proceed to incrementally add bands that maximize the angular separation between the two spectra.

We first choose as our starting bands the pair that have the minimum 2-angle (BAO-MIN). The next step requires a calculation of β for all remaining bands. To demonstrate this step, we graphically illustrate the principles behind the band selection. Any band may be chosen, regardless

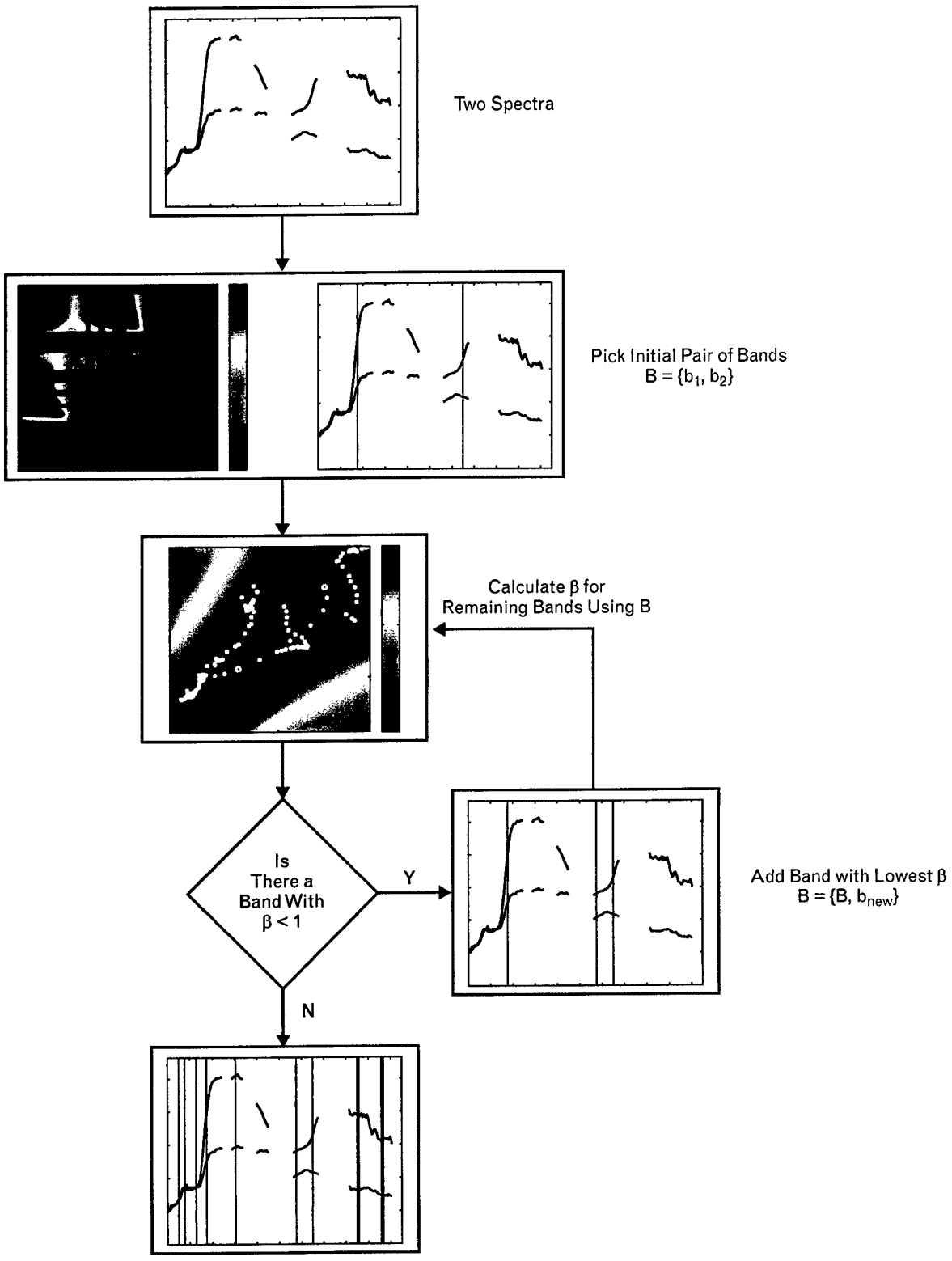


Figure 12. Flowchart for Band Add-On (BAO) algorithm to select bands that maximize the angular separation between two spectra.

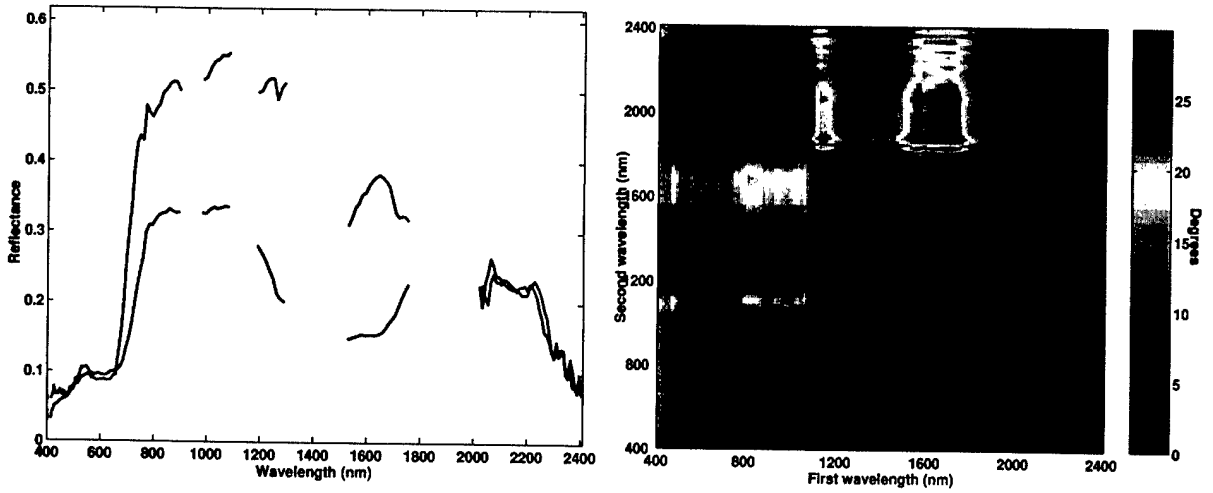


Figure 13. (a) Plot of two spectra; (b) Contour map of all possible 2-angles.

of its proximity to the starting bands. As a consequence, for each band in the two spectra (heretofore called $\mathbf{x}$ and $\mathbf{y}$), we can plot the corresponding reflectance values for the two spectra. This is depicted in Figure 14. Here, bands that have already been selected are plotted with an "O", and the remaining bands are plotted with an "X".

The range of reflectance values for $\mathbf{x}$ and $\mathbf{y}$ define the breadth of the axes in Figure 14. We can add shading to indicate the value of β at each point in this two-dimensional domain to indicate how different pairs of band values increase or decrease the overall angular separation of the two spectra. It is important to note that the shading depends on the initial starting bands used in the expression for β in (25). The case of starting bands with a minimum angle (BAO-MIN) is conveyed in Figure 15(a), and every available band induces a value of $\beta < 1$. As a consequence, any band that is added to the starting pair of bands will necessarily increase the overall angle. This result is not surprising, since BAO-MIN starts with the smallest possible angle. In Figure 15(b) the situation using the starting pair of bands having the largest 2-angle (BAO-MAX) is conveyed. A black contour indicates the band values for $\mathbf{x}$ and $\mathbf{y}$ that lead to $\beta = 1$, identifying the region within which $\beta > 1$. A significant number of bands lead to values of $\beta > 1$, and hence are not candidates to be added to the starting band set.

The shaded scatterplots in Figures 15(a) and 15(b) demonstrate which bands have $\beta < 1$, but do not indicate what wavelengths are associated with different values of β. Figure 16(a) does this for BAO-MIN, coloring band values for the two spectra by their associated value of β. Bands having $\beta > 1$ are colored black. Figure 16(a) shows that the starting band pair for BAO-MIN results in nearly all remaining bands having $\beta < 1$. Hence these bands are viable candidates to increase the angle between $\mathbf{x}$ and $\mathbf{y}$. Figure 16(b) illustrates the same for BAO-MAX, and, only bands in the blue and green part of the visible contribute to low values of β. Table 8 compares the results obtained so far.

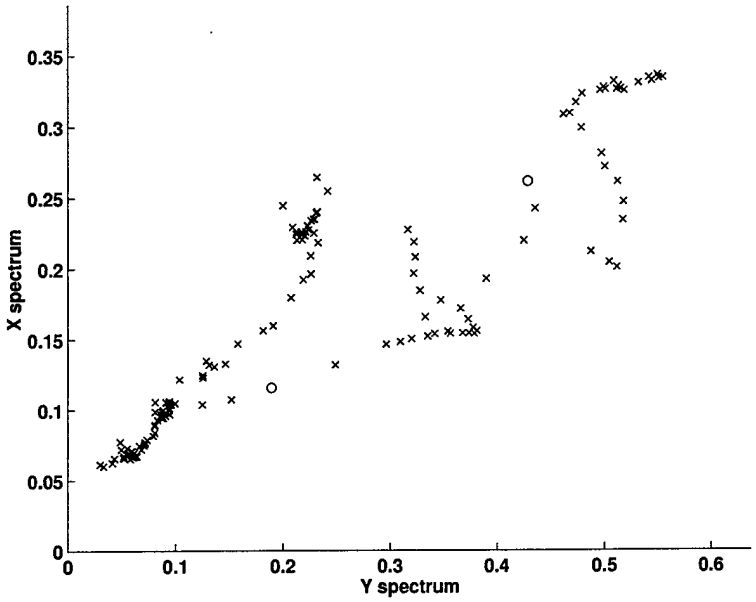


Figure 14. Scatterplot of band values for the two spectra in Figure 13(a).

The illustrations in Figure 15 and Figure 16 provide a graphic illustration of the mathematical principle for adding bands. Bands that possess an associated value of $\beta \leq 1$ will increase the angular separation. For both BAO-MIN and BAO-MAX there are multiple candidate bands with $\beta \leq 1$ that may be added to the initial set with $\beta \leq 1$. In such a case, different criteria could be applied to perform the selection. We will choose here to select the candidate band having the *lowest* β. From Table 8, the new SAM angle for $\mathbf{x}$ and $\mathbf{y}$ having three bands can easily be calculated using (23). For BAO-MIN, the band having the lowest β (0.954) occurs at 2051nm, and the resulting 3-angle using these three bands is 17.42^{o}. For BAO-MAX, the band having the lowest β (0.993) occurs at 411nm, and the resulting 3-angle is 30.70^{o}.

We can now repeat the procedure of evaluating β for the remaining bands using the two initial bands, as well as the first selected band. The corresponding scatterplots for BAO-MIN and BAO-MAX appear in Figure 17 and demonstrate that fewer available bands have a corresponding $\beta \leq 1$. The associated shaded plots of the spectra in Figure 18 illustrate the wavelengths at which these bands occur. Table 9 summarizes the important quantities for this iteration.

This iterative procedure can be repeated until no bands satisfy $\beta \leq 1$. Any available band that is added to the selected bands will only decrease the overall angle between $\mathbf{x}$ and $\mathbf{y}$. Compared to the complete angle of 16.71^{o}, BAO-MIN increased the angle between $\mathbf{x}$ and $\mathbf{y}$ to 27.34^{o} using 37 bands, and BAO-MAX increased the angle to 32.10^{o} using 8 bands. Figure 19 illustrate the bands that were chosen by BAO-MIN and BAO-MAX with respect to the original two spectra, and the final numerical results are summarized in Table 10.

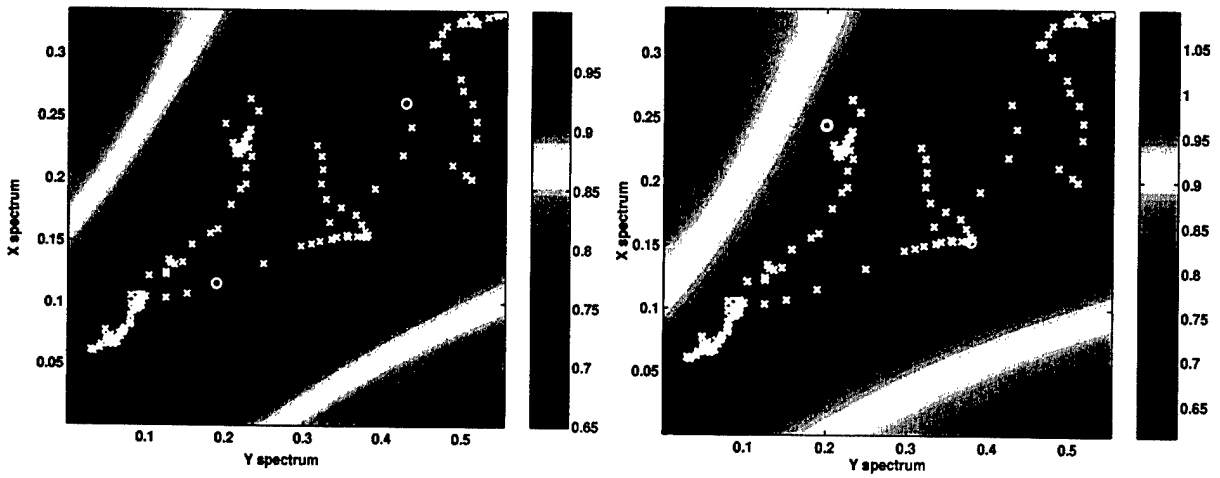


Figure 15. Scatterplots of band values for the two spectra in Figure 13 during the first iteration of BAO when using (a) minimum 2-angle (BAO-MIN) as starting bands; (b) maximum 2-angle (BAO-MAX) as starting bands. The color shading indicates the associated value of β, and the black line corresponds to values where $\beta = 1$.

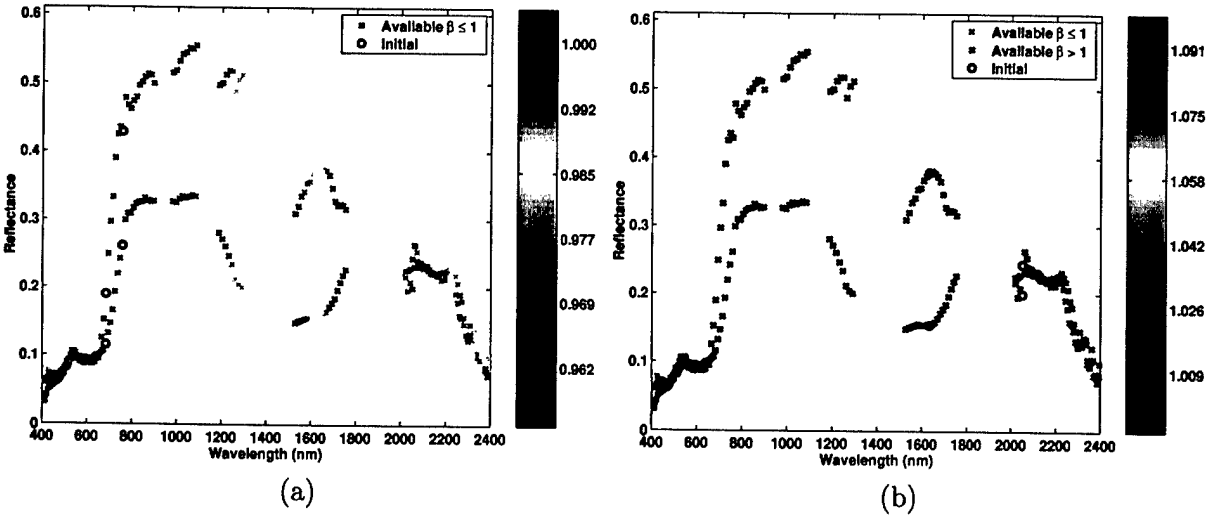


Figure 16. Plots of spectra with color shading from Figure 15 during first iteration that illustrate values of β as a function of wavelength for (a) minimum 2-angle (BAO-MIN) as starting bands; (b) maximum 2-angle (BAO-MAX) as starting bands.

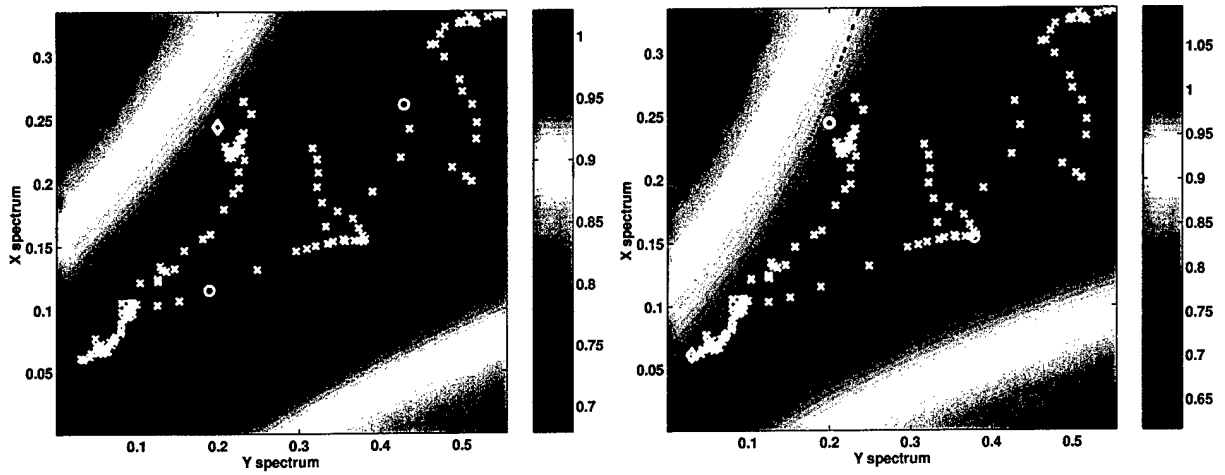


Figure 17. Scatterplots of band values for the two spectra in Figure 13 during the second iteration of BAO when using (a) minimum 2-angle (BAO-MIN) as starting bands; (b) maximum 2-angle (BAO-MAX) as starting bands. The color shading indicates the associated value of β, and the black line corresponds to values where $\beta = 1$.

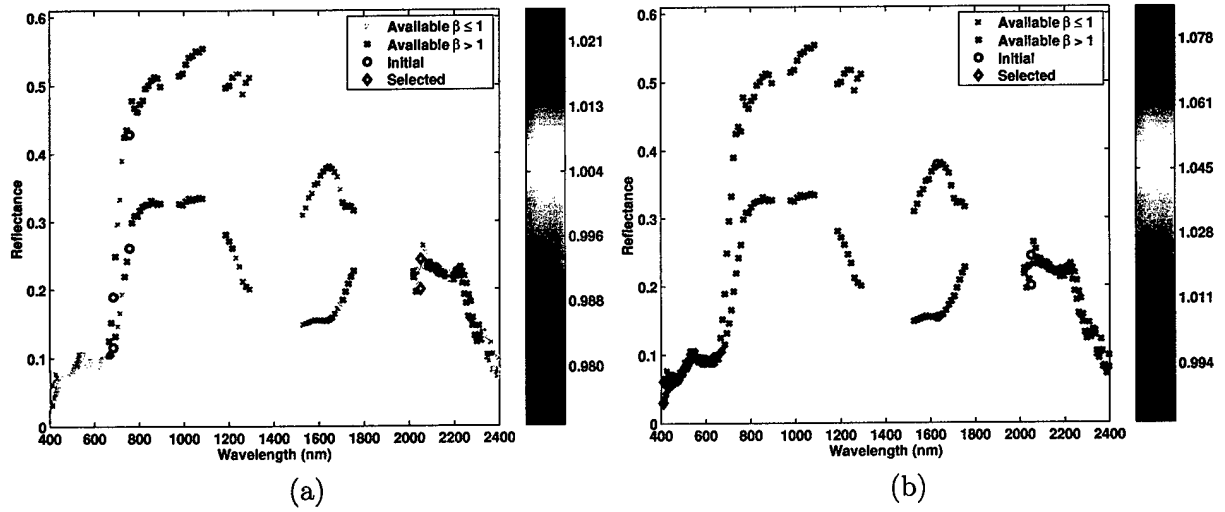


Figure 18. Plots of spectra with color shading from Figure 17 during second iteration that illustrate values of β as a function of wavelength for (a) minimum 2-angle (BAO-MIN) as starting bands; (b) maximum 2-angle (BAO-MAX) as starting bands.

	Bands in initial set (nm)	Initial angle (^{o})	Available bands with $\beta \leq 1$	Range of available β
BAO-MIN	684, 757	0.00012	143	[0.954, 1]
BAO-MAX	1632, 2051	29.98	15	[0.993, 1.091]

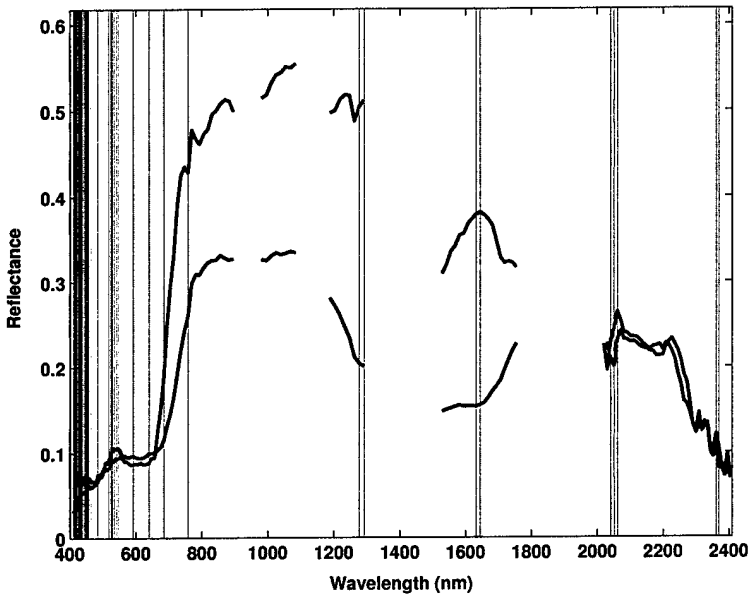
TABLE 8. Summary of results from Figure 15 for BAO-MIN and BAO-MAX.

	Selected bands (nm)	Existing angle (^{o})	Available bands with $\beta \leq 1$	Range of available β
BAO-MIN	684, 757, 2051	17.42	82	[0.977, 1.021]
BAO-MAX	1632, 2051, 411	30.70	11	[0.994, 1.095]

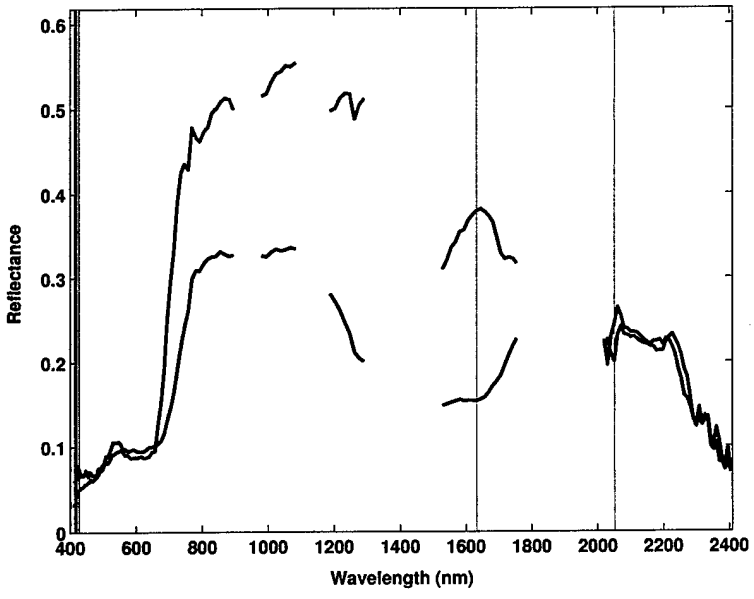
TABLE 9. Summary of results from Figure 17 for BAO-MIN and BAO-MAX.

	Initial bands (nm)	Initial angle (^{o})	Selected bands (nm)	Total number of bands	Final angle (^{o})
BAO-MIN	684, 757	0.00012	{2051, 1291, 2061, 425, 411, 2370, 414, 1632, 2041, 428, 421, 418, 447, 1644, 2362, 435, 432, 439, 443, 529, 450, 458, 535, 1276, 547, 541, 639, 524, 585, 592, 463, 484, 454, 457, 518}	37	27.34
BAO-MAX	1632, 2051	29.98	{412, 425, 414, 421, 418, 428}	8	32.10
Complete angle			All	145	16.71

TABLE 10. Final results for maximizing the angle between $\mathbf{x}$ *and* $\mathbf{y}$ *in Figure 13(a) using BAO-MIN and BAO-MAX.*



(a)



(b)

Figure 19. Band selection for (a) BAO-MIN and (b) BAO-MAX for the two spectra in Figure 13.

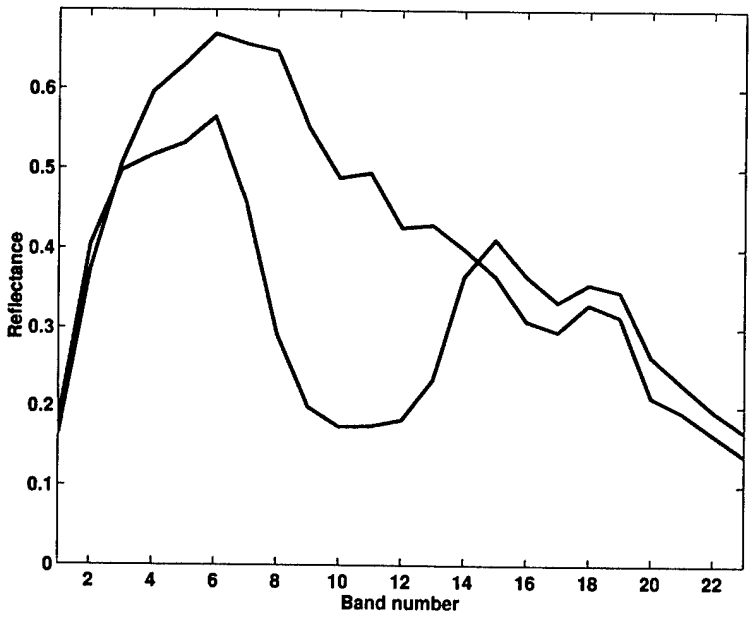


Figure 20. Two length-23 spectra derived by truncating and spectrally degrading two length-145 spectra.

3.6 COMPARISONS WITH EXACT ANSWERS

The results in Section 3.3, 3.4, and 3.5 indicate that while it is possible to select bands that yield a sub-angle greater than the complete angle, none of the approaches can guarantee that it always selects those bands that yield the biggest sub-angle. The attempts using single and multiple contiguous segments are intuitive, but severely limit the number of admissible solutions. The segments also become prohibitively expensive to compute when more segments are permitted. From the scatterplots in Figures 15 and 17, it is evident that the subset of bands selected using BAO depends on the rule used to select the bands at each iteration, as well as the starting bands. There are numerous opportunities to vary this algorithm, with each alteration possibly yielding a different result.

Thus, the approaches we have explored are inherently sub-optimal, i.e., they are not guaranteed to yield the set of bands possessing the largest sub-angle. Ultimately, the largest sub-angle between two spectra involves a multi-dimensional search with no known closed-form, analytical solution. In addition, the decision space over which an discrete optimization would search for the biggest sub-angle is not necessarily convex, meaning that any gradient-based, or greedy, approach may find a local point of optimality, instead of a global optimum.

One alternative to ascertain the largest sub-angle between two spectra would require exhaustively measuring every sub-angle that exists between two spectra. For hyperspectral signals having a dimension of approximately 150, Table 3 demonstrated that the number of sub-angles is well beyond reasonable calculation. For lower dimensions, however, such a complete search is possible.

Rank	Sub-angle (°)	Bands
1	33.68	$\{11, 20, 23\}$
2	33.51	$\{10, 20, 23\}$
3	33.42	$\{11, 20, 22\}$
4	33.31	$\{11, 20, 21\}$
5	33.25	$\{11, 20\}$
6	33.24	$\{10, 20, 22\}$
7	33.19	$\{11, 20, 22, 23\}$
8	33.12	$\{10, 20\}$
9	33.11	$\{10, 20, 21\}$
10	33.10	$\{10, 11, 20, 21, 22, 23\}$

TABLE 11. Ten highest sub-angles for the pair of spectra in Figure 20.

	Sub-angle (°)	Bands	Number of bands	Percentile rank (%)
Exhaustive (θ_{max})	33.68	$\{11, 20, 23\}$	3	100
BAO-MIN	24.36	$\{5, 6, 9, 11, 10, 16, 20, 12, 21, 23, 22, 8, 15, 17\}$	14	92.43
BAO-MAX	33.68	$\{11, 20, 23\}$	3	100
Single segment	27.95	$\{9, 10, 11, 12, 13, 14, 15, 16, 17\}$	9	99.37
Double segment	33.31	$\{11, 20, 21\}$	3	99.99
Triple segment	33.68	$\{11, 20, 23\}$	3	100
Complete angle	20.37	All	23	53.12

TABLE 12. Comparison of band selection techniques with exhaustive solution for length-23 spectra in Figure 20.

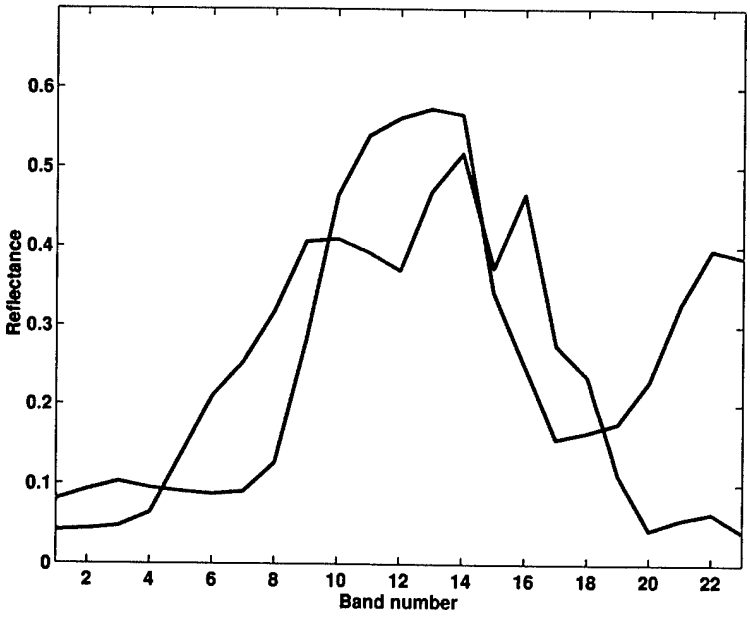


Figure 21. Two length-23 spectra derived by truncating and spectrally degrading two length-145 spectra.

Figure 20 shows two length-23 spectra that have been derived by truncating and then spectrally degrading two length-145 hyperspectral pixel spectra. The complete angle is 20.37^{o}. An exhaustive search calculated every sub-angle, and the maximum sub-angle, θ_{max}, of 33.68^{o} was induced by bands $\{11, 20, 23\}$. The ten highest sub-angles are tabulated in Table 11. While it is evident that there are several sub-angles yielding almost the same value as θ_{max}, nearly the same subset of bands re-appear in various combinations: $\{10, 11, 20, 21, 22, 23\}$. Table 12 demonstrates that when the band selection algorithms discussed in Section 3 are applied to the same pair of spectra, several of the approaches find subsets of bands that appear in Table 11. The rightmost column indicates the percent of the total set of sub-angles that a method equals or exceeds. BAO-MAX and the triple-segment approach both find θ_{max}. BAO-MIN finds a significantly lower angle than θ_{max}, while using 14 bands, but still exceeds 92.43% of all sub-angles. Interestingly enough, the complete angle, using all bands, is only greater than 53.12% of all sub-angles.

A similar set of effects is noticed in Table 13 and Table 14 for the two spectra plotted in Figure 21.

3.7 DISCUSSION

There are several conclusions that can be made about the band selection techniques and results that have been discussed. We consider them independently.

Rank	Sub-angle (°)	Bands
1	67.82	$\{7, 23\}$
2	66.61	$\{1, 7, 23\}$
3	66.42	$\{6, 23\}$
4	66.40	$\{2, 7, 23\}$
5	66.13	$\{3, 7, 23\}$
6	66.10	$\{7, 20, 23\}$
7	65.65	$\{4, 7, 23\}$
8	65.27	$\{1, 2, 7, 23\}$
9	65.14	$\{7, 21, 23\}$
10	65.11	$\{6, 7, 23\}$

TABLE 13. Ten highest sub-angles for the pair of spectra in Figure 21.

	Sub-angle (°)	Bands	Number of bands	Percentile rank (%)
Exhaustive (θ_{max})	67.82	$\{7, 23\}$	3	100
BAO-MIN	48.07	$\{4, 12, 23, 8, 22, 7, 21, 6, 20\}$	9	99.13
BAO-MAX	67.82	$\{7, 23\}$	2	100
Single segment	54.04	$\{17, 18, 19, 20, 21, 22, 23\}$	7	99.76
Double segment	67.82	$\{7, 23\}$	2	99.99
Triple segment	66.61	$\{1, 7, 23\}$	3	99.99
Complete angle	29.98	All	23	53.58

TABLE 14. Comparison of band selection techniques with exhaustive solution for length-23 spectra in Figure 20.

3.7.1 Number of Bands and Robustness

The exhaustive determination of sub-angles in Tables 12 and 14, as well as the techniques based on contiguous segments and band add-on, demonstrate that sub-angles greater than the complete angle can frequently be found having five or less bands. In Section 3.2, we suggested that the increased contrast can lead to more robust discrimination. However, are there drawbacks to using a very small number of bands?

For realistic conditions, the answer is yes, but the effects can be mitigated and controlled. Given a sub-angle larger than the complete angle, this indicates the presence of a feature that possesses greater contrast between two spectra (see Section 3.2). However, in order for this feature to be useful for discriminating between two classes, this feature must be robust to variability in the incoming signal. We can express this concept mathematically. Let there be two classes, having template spectra, $\mathbf{t}_1$ and $\mathbf{t}_2$, respectively. An unknown test pixel, $\mathbf{r}$, that arises from one of the two classes may also have additive interference, $\mathbf{w}$:

$$\mathbf{r} = \mathbf{t}_1 + \mathbf{w} \quad \text{OR} \quad \mathbf{r} = \mathbf{t}_2 + \mathbf{w}. \tag{27}$$

The interference may arise from numerous sources, including differences in observation angle, atmospheric effects, and illumination. If $\mathbf{r}$ is compared with $\mathbf{t}_1$ and $\mathbf{t}_2$ using SAM, and $\theta(\mathbf{t}_1, \mathbf{t}_2)$ is small, then if $\mathbf{w}$ is large enough, it will cause $\mathbf{r}$ to be misclassified. Increasing the angle using a subset of bands, $\mathbf{B}$, increases the angular contrast between the two classes, but, depending upon the structure of the noise, can also amplify the variability in the angular measurements, and consequently the angular comparisons with $\mathbf{t}_1$ and $\mathbf{t}_2$, also leading to misclassification. In short, a very low number of bands may not be robust to variability in the input signal, and hence, it may be preferable to accept a lower sub-angle (but still greater than the complete angle) in order to have a larger set of bands. This important topic is the subject of Section 4.

3.7.2 Starting Bands

We demonstrated two conditions for selecting a pair of starting bands to initialize the BAO optimization. Experiments using BAO-MIN and BAO-MAX have demonstrated a consistent behavior. BAO-MIN arrives at a lower sub-angle than BAO-MAX, and in doing so, utilizes more bands than BAO-MAX. This conclusion is confirmed in Section 3.6, where the maximum sub-angle was found by BAO-MAX and only had 2 or 3 bands. This result is not surprising. BAO-MAX starts with the largest 2-angle and adds bands that further increase that sub-angle. Repeated experiments have demonstrated that the largest sub-angle is often a combination of 2 to 5 bands, of which the initial bands for BAO-MAX are frequently a subset. Hence, BAO-MAX starts with a large angle and adds only a few, if any, bands, before it must stop. On the other hand, BAO-MIN achieves its final angle, albeit smaller than the maximum sub-angle, but using more bands.

3.7.3 Optimality

We explored several techniques for determining the largest sub-angle between two spectra. However, all of the approaches are sub-optimal, and none are guaranteed to yield the largest angle. The experiments in Section 3.6 demonstrate that the sub-optimal approaches appear to arrive at

the optimal solution, or very near to it. More importantly, the independent approaches focus on the same parts of the spectrum, leading to the conclusion that, despite being based on different formulations, the same bands of interest can be found by different means.

3.7.4 Non-Intuitive Results

The figures and tables summarizing the band selection demonstrate that the results are not always intuitive. Despite the obvious differences between spectra in Figures 7 and 8, the angular structure of data is difficult to infer from two-dimensional plots (see Section 2.2.1). The angular interpretation of data occurs in a high-dimensional space, and SAM imposes its own mathematical structure in that domain.

3.7.5 Phenomenology

In Section 3.1, we discussed how a priori physical knowledge of phenomenology has influenced the design specifications of sensors. The spectral intervals, as well as their bandwidths, are chosen to observe the important phenomenology, but little consideration is given to how the data will ultimately be processed. In contrast, we have demonstrated that band selection algorithms formulated around a given distance metric, such as SAM, selectively induce phenomenology into the mathematical analysis from bands that yield greater contrast, while omitting those that do not.

3.8 SECTION SUMMARY

In this section, we discussed ways in which bands have been selected to perform tasks in hyperspectral processing. We focused first on how a priori physical knowledge of the electromagnetic intervals where spectral features are present has driven the design of sensors as well as the algorithms that process the measured data. This approach has been prominent in the observation and analysis of environmental data where bands are carefully chosen to match the phenomenology under observation. In military scenarios, where the degree of accuracy required is much higher, physical knowledge about the objects of interest may be limited, and, hence, hyperspectral data is collected in many narrow bands over wide intervals. Distance metrics perform the fundamental task of comparing two spectra, and it was shown that by selecting the appropriate subset of bands, the most commonly used metric for hyperspectral processing, SAM (Spectral Angle Mapper), can be mathematically optimized to increase the angular separation between two spectra. This form of band selection "induces" the appropriate phenomenology to increase the contrast between two signals. Different approaches were demonstrated for selecting bands that increase SAM for two spectra. Contiguous segments of data were identified by exhaustive searches, but the approach severely limits the number of admissible solutions and is computationally demanding. A more efficient technique, BAO (Band Add-On), is a framework for analytically selecting bands that is based on a mathematical decomposition of SAM. The technique was demonstrated in detail for two sample spectra. Although it is impractical to exhaustively find the largest sub-angle for typical hyperspectral data, an exhaustive search of all possible sub-angles was performed for a pair of length-23 spectra. The results were compared to the bands selected by the different techniques discussed in this section, and showed that the methods, while mathematically sub-optimal, succeed in finding angles that are close to the maximum sub-angle. The robustness of band selection algorithms to target variability is an important issue that will be addressed in the next section.

4. DISCRIMINATING TARGETS HAVING VARIABILITY

In Section 3, we demonstrated how the angular separability of two target spectra can be increased by selecting bands through a mathematical optimization known as Band Add-On (BAO). In this section, we extend that formulation to select bands that improve the ability to distinguish two classes of targets, where each class is described by at least one sample spectrum. This scenario is important for material identification, which is the principle capability that hyperspectral sensing possesses over other sensor modalities (e.g., radar, sonar, etc.). Hence, material identification is a unique capability that hyperspectral sensors offer to spectral processing, as well as to the fusion of multi-sensor data.

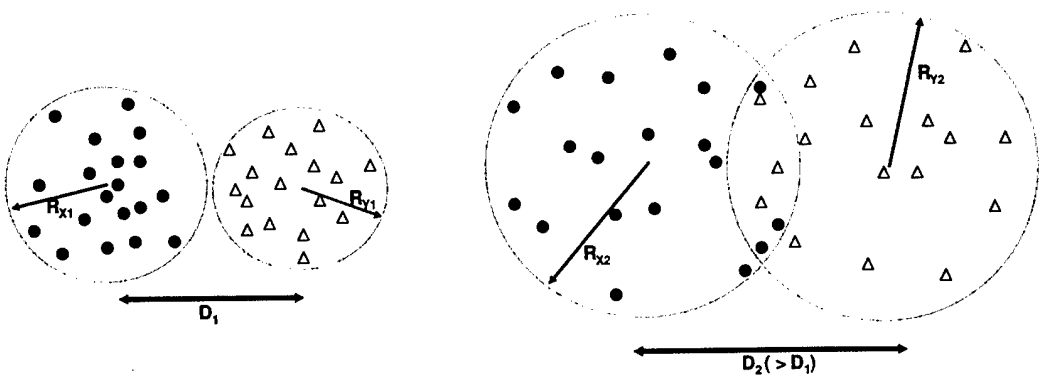
In Section 1.2, we noted that most hyperspectral processing involves a measurement of the similarity between two spectra, and this is where distance metrics perform an important role. The ability to distinguish two classes is not just a function of the distance between the means of two classes, also referred to as the inter-class separation. If the variability of the members, referred to as the intra-class variability, of one or both classes is great, then even if the inter-class separation is great, distinguishing one class from the other class may still be problematic. Figure 22 demonstrates this graphically. Essentially, successful discrimination of one class from another is a function of both the inter-class separation and the intra-class variability.

Section 3.5 outlined the framework for an algorithm to select bands that increases the angle between two spectra, thereby increasing the inter-class separation. No mention was made of the concomitant impact of using a reduced set of bands on intra-class variability because there was no intra-class information. In this section, we account for the fact that there is variability in the reflectance spectrum measured for a distinct material, and we incorporate what is known about that variability into a band selection algorithm based upon BAO.

4.1 APPLICATIONS TO MATERIAL IDENTIFICATION

Material identification using hyperspectral data is the procedure by which an unknown pixel is classified as one of several materials whose reflectance spectra are known from reference measurements. Ideally, the reflectance spectrum of a material measured by a laboratory instrument should not vary, but, in reality, it does, due to numerous sources (e.g., sensor noise, atmospheric variability, target orientation). In practice, several reflectance measurements are usually collected using a spectroradiometer and then averaged to obtain a template spectrum for each class. In some HYMSMO (Hyperspectral MASINT Support to Military Operations) experiments with the HYDICE sensor, the number of reference measurements for targets of interest typically range from 3 to 10.

Considering the high dimensionality of hyperspectral data, the fact that only a handful of reference measurements may exist for a substance distinguishes the material identification problem from the approaches utilized for statistical pattern recognition. Traditional pattern recognition algorithms [12] require probability density functions (PDFs) to describe intra-class variability, but determining accurate PDFs empirically requires a large number of sample pixels for each class.



(a) (b)

Figure 22. Notional illustration of two target classes, where the inter-class distance between the class means in (b) is greater than that in (a), but the resulting increase in intra-class variability in (b) still makes perfect classification difficult.

Hence, the lack of an accurate description of how material spectra vary necessitates alternate methods.

In the absence of statistical formulations, a distance metric provides a meaningful way for comparing an unknown pixel spectrum with a library of template spectra, each corresponding to a specific material. A common distance metric for this application is SAM, which compares an unknown pixel spectrum, $\mathbf{r}$, to the template spectra, $\mathbf{t}_i, i = 1, \ldots, K$, for each of K templates and assigns $\mathbf{r}$ to the material having the smallest distance,

$$\text{Class}(\mathbf{r}) = \underset{1 \leq i \leq K}{\text{argmin}}\, \theta(\mathbf{r}, \mathbf{t}_i). \tag{28}$$

We will focus on SAM as a distance metric for material identification for the remainder of this report.

4.2 INCORPORATING VARIABILITY IN BAND SELECTION

In Section 3.5, we developed a method for iteratively selecting bands that increases the angular separation between two spectra. In Section 4.1, we saw, however, that a material may not be characterized by a single measured spectrum, but, in fact, may have several valid spectral signatures, but not enough for an accurate representation of statistical intra-class variability. Provided this, a band selection method that maximizes the angle between *sets* of spectra would increase the angular contrast between two materials, thus improving the ability to distinguish two classes of materials, even when interference and distortions are present.

Predicting variability in a reflectance spectrum for a unique material is difficult. The difference in spectroradiometer measurements indicate at least some variability that may arise from the instrument noise. Variability may also arise from other sources, including spatial non-homogeneity

of the material. More importantly, the cumulative information from the entire set of spectra is greater than the just the mean spectra, and the way in which the spectra vary, small or large as it may be, can still be exploited by band selection to increase the angular separability of two sets of spectra, while providing additional robustness to signature variability. The variability observed, and exploited, in the spectroradioimeter measurements, however, is significantly less than the worst signature variability that can be observed by an operational sensor. Yet, as stated before, incorporating the variability observed in the laboratory measurements provides at least some leverage to yield better classification results.

4.3 TWO PHILOSOPHIES: ADM AND MDM

The formulation in (28) provides the basis for classifying an unknown spectra with a material label, through a series of pairwise comparisons using SAM. Integral to optimizing this test are two quantities:

1. The template spectra, $\mathbf{t}_i$, that represent the spectra for each distinct material in a spectral library, and
2. The bands that are employed in the angular comparisons.

We consider two philosophies for selecting these quantities based on the BAO methodology explored in Section 3.5. Although a typical library of material spectra may contain hundreds or thousands of spectra, for our immediate purposes, we assume there are two material classes, class $\mathbf{X}$ and class $\mathbf{Y}$. Each pixel is a hyperspectral measurement having M bands, and each reference spectrum represents a different substance or material. Our goal, then, is to devise a way to classify pixels as belonging to either $\mathbf{X}$ or $\mathbf{Y}$. For each class, we assume there are a set of $M \times 1$ training pixels for each class, $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{N_X}\}, \mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_{N_Y}\}$. However, as mentioned earlier, there are not enough pixels to develop dependable, statistical representations of the intra-class variability. This is frequently the case for man-made targets of interest whose statistical variability under all possible observation conditions is hard to quantify.

4.3.1 Average Distance Method (ADM)

The first method uses BAO to select a subset of bands, $\mathbf{B}$, to maximize the average pairwise cosine between spectra in $\mathbf{X}$ and $\mathbf{Y}$. The Average Distance Method (ADM) is illustrated conceptually in Figure 23 and strives to minimize the average cosine of every pairwise angle between the entries in $\mathbf{X}$ and $\mathbf{Y}$. The quantity being optimized is

$$\underset{\mathbf{B}}{\operatorname{argmin}} \frac{1}{N_X N_Y} \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} \cos\theta(\mathbf{x}_i(\mathbf{B}), \mathbf{y}_j(\mathbf{B})). \tag{29}$$

The mean spectra of each class, μ_X and μ_Y, serve as templates during classification, but only using the selected bands. The implication here is that the bands in $\mathbf{B}$ will increase the *average* angle between the spectra in $\mathbf{X}$ and $\mathbf{Y}$, where the average is taken over all possible pairs of spectra in $\mathbf{X}$

and $\mathbf{Y}$. This approach differs from the simple maximization of the angle between μ_X and μ_Y, by including the influence of the member spectra in $\mathbf{X}$ and $\mathbf{Y}$ on band selection.

As in BAO, we choose two initial bands, $\mathbf{B}_1 = [\ b_1\ \ b_2\ ]$ to begin the iterations. Borrowing the same conclusions made in Section 3.5.2, we can extend the logic of BAO-MIN and BAO-MAX here and select the pair of bands that either minimize or maximize the cumulative sum of cosines for the spectra in $\mathbf{X}$ and $\mathbf{Y}$,

$$\text{BAO-MIN:} \quad \underset{\mathbf{B}_1}{\text{argmax}} \frac{1}{N_X N_Y} \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} \cos\theta(\mathbf{x}_i(\mathbf{B}_1), \mathbf{y}_j(\mathbf{B}_1)) \tag{30}$$

$$\text{BAO-MAX} \quad \underset{\mathbf{B}_1}{\text{argmin}} \frac{1}{N_X N_Y} \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} \cos\theta(\mathbf{x}_i(\mathbf{B}_1), \mathbf{y}_j(\mathbf{B}_1)). \tag{31}$$

It was observed, however, in Section 3.7 that the starting condition imposed by BAO-MAX often resulted in a low number of total bands (2-4). This is a consequence of picking an initial pair of bands that induces a large starting angle, leaving few, if any, valid candidates to further increase the angle. As Tables 11 and 13 demonstrate, BAO-MIN often arrives at similar, but lower, angles, but using significantly more bands. Despite the smaller angles, the luxury of more bands is important, since angular target variability is often mitigated. Hence, we use BAO-MIN as the starting condition for ADM. Likewise, Step 2 of ADM appends individual bands that induce the smallest average $\beta, \overline{\beta}$, such that $\overline{\beta} < 1$,

$$b_k = \underset{b_k}{\text{argmin}} \frac{1}{N_X N_Y} \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} \beta(\mathbf{x}_i(\mathbf{B}_1), \mathbf{y}_j(\mathbf{B}_1); x_i(b_k), y_j(b_k)), \tag{32}$$

for $k > 2$, where $\mathbf{x}_i(\mathbf{B}_1) = [x_i(b_1), \ldots, x_i(b_{k-1})]$, and $\mathbf{y}_i(\mathbf{B}_1) = [y_i(b_1), \ldots, y_i(b_{k-1})]$, and $b_k \notin \mathbf{B}_1$. If no band makes $\overline{\beta} < 1$, the procedure ends. The template spectra, $\mathbf{t_X}$ and $\mathbf{t_Y}$, for a subsequent test are $\mu_\mathbf{X}$ and $\mu_\mathbf{Y}$ and use only the selected bands in $\mathbf{B}$.

4.3.2 Minimum Distance Method (MDM)

In ADM, the spectra used as templates are the means of each class, and bands are chosen to maximize the average angular difference between $\mathbf{X}$ and $\mathbf{Y}$. However, the criteria in (31) does not guarantee that the elements in $\mathbf{X}$ and $\mathbf{Y}$ will be correctly classified by (28). In contrast to ADM, a different approach, called the Minimum Distance Method (MDM), chooses an initial pair of bands and template spectra such that the elements of $\mathbf{X}$ and $\mathbf{Y}$ are correctly classified by SAM, and then adds additional bands that increase the angular separation, while maintaining perfect classification.

To outline this technique, we define the *worst-case angle*, $\theta_w(\mathbf{X}, \mathbf{Y}, \mathbf{B}_1)$, for $\mathbf{X}$ and $\mathbf{Y}$ using a set of bands, $\mathbf{B}_1$, as

$$\theta_w(\mathbf{X}, \mathbf{Y}, \mathbf{B}_1) = \min_{\mathbf{x}_i \in \mathbf{X}, \mathbf{y}_j \in \mathbf{Y}} \theta(\mathbf{x}_i(\mathbf{B}_1), \mathbf{y}_j(\mathbf{B}_1)) \tag{33}$$

$$= \theta(\mathbf{x}_w, \mathbf{y}_w, \mathbf{B}_1), \tag{34}$$

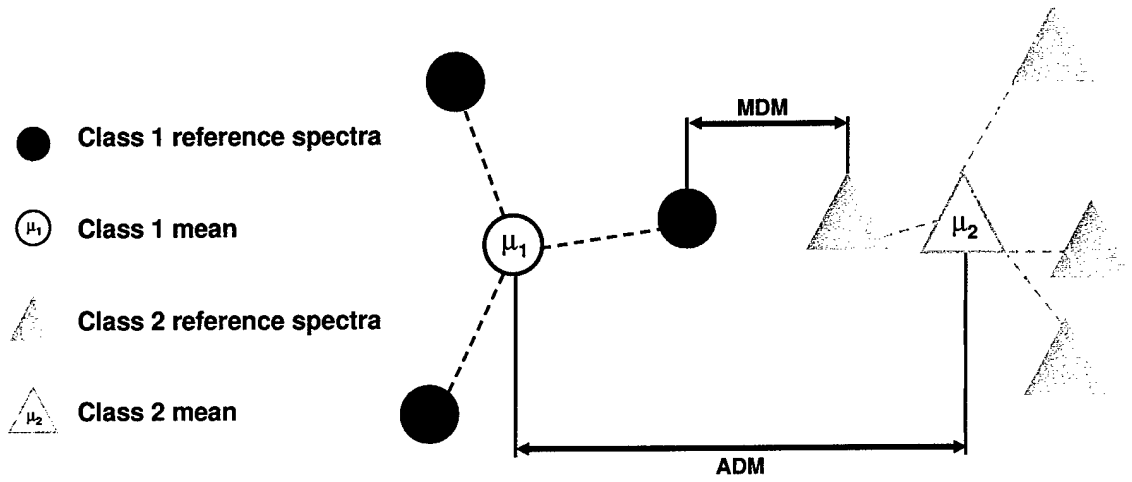


Figure 23. Conceptual difference between MDM and ADM. ADM selects bands that increase the separation between the means of each class. MDM selects bands to increase the separation between the closest, or worst-case, pixels from each class.

where $\mathbf{x}_w$ and $\mathbf{y}_w$ are the spectra in $\mathbf{X}$ and $\mathbf{Y}$ that create the smallest angle. We can search for a pair of initial bands, $\mathbf{B}_1$, having the largest $\theta(\mathbf{x}_w, \mathbf{y}_w, \mathbf{B}_1)$ that also must classify every pixel in $\mathbf{X}$ and $\mathbf{Y}$ correctly when the template spectra are $\mathbf{t}_x = \mathbf{x}_w$ and $\mathbf{t}_y = \mathbf{y}_w$,

$$\theta(\mathbf{t}_\mathbf{X}(\mathbf{B}_1), \mathbf{x}_i(\mathbf{B}_1)) < \theta(\mathbf{t}_\mathbf{Y}(\mathbf{B}_1), \mathbf{x}_i(\mathbf{B}_1)), \qquad \forall\, \mathbf{x}_i \in \mathbf{X} \tag{35}$$

$$\theta(\mathbf{t}_\mathbf{Y}(\mathbf{B}_1), \mathbf{y}_i(\mathbf{B}_1)) < \theta(\mathbf{t}_\mathbf{X}(\mathbf{B}_1), \mathbf{y}_i(\mathbf{B}_1)), \qquad \forall\, \mathbf{y}_i \in \mathbf{Y}. \tag{36}$$

Figure 23 illustrates this concept.

Once $\mathbf{B}_1$ is found, additional bands are added using BAO to increase the angular distance between $\mathbf{x}_w$ and $\mathbf{y}_w$ (under the condition that $\theta_w(\mathbf{x}_w, \mathbf{y}_w)$ increases by the definition in (33)) and each pixel in $\mathbf{X}$ and $\mathbf{Y}$ continues to be correctly classified. It may occur that after adding a band to the current $\mathbf{x}_w$ and $\mathbf{y}_w$, that a different member of $\mathbf{X}$ and $\mathbf{Y}$ satisfies $\theta_w(\mathbf{x}_w, \mathbf{y}_w)$. In this case, the template pixels for $\mathbf{X}$ and $\mathbf{Y}$ are reset to those entries giving rise to $\theta_w(\mathbf{x}_w, \mathbf{y}_w)$. The goal here is to protect the pixels having the greatest chance of misclassification (i.e., $\mathbf{x}_w$, $\mathbf{y}_w$) by letting the template spectrum equal $\mathbf{x}_w$ and $\mathbf{y}_w$. Bands are added incrementally, as before, under the condition that it preserves perfect classification of the pixels in $\mathbf{X}$ and $\mathbf{Y}$ and also increases $\theta_w(\mathbf{x}_w, \mathbf{y}_w)$. The iterations end when no unused band remains that increases θ_w and still preserves perfect classification.

4.4 TWO-CLASS EXPERIMENTS WITH SIMILAR TARGETS

We choose to apply the band selection techniques, ADM and MDM, to improve the discriminability of two classes of targets that are very similar in their spectral signatures. While two classes

	Bands (nm)	Number of bands	$\theta_w(\mathbf{X}, \mathbf{Y})$ (°)	Correct Classification	
				Material X	Material Y
ADM	{994, 1009, 2291, 2281, 2308, 2300, 2317, 2272}	8	5.434	17/17	0/20
MDM	{769, 2281, 2300, 2020, 2317, 2291, 704, 694, 513, 508, 518, 2030, 2386, 2308, 503, 524, 529, 498, 535, 684}	20	7.171	17/17	20/20
All bands	{400 – 2405}	145	3.180	17/17	0/20

TABLE 15. Results of band selection and binary classification test to discriminate material X and material Y using data from Forest Radiance I, Run 05.

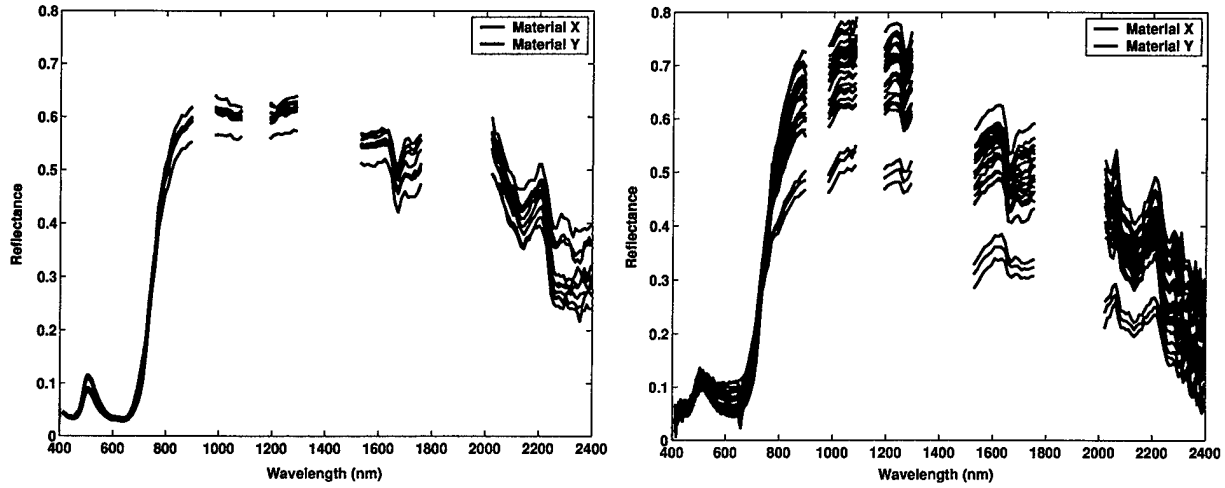
may have discernible laboratory reflectances, the process by which operational sensors collect data is not reliably clean enough to always maintain the differences. In addition to sensor noise, there are numerous sources of additional interference, including atmospheric compensation and target variability. Moreover, tactical scenarios may introduce occlusion as well as off-nadir viewing geometries. In short, the separability of materials in controlled laboratory setting cannot be expected in real airborne or spaceborne data collection environments.

In particular, military strategies often employ CC&D techniques that render targets indistinguishable to the natural environment in the visible spectrum. However, spectral differences in the near-infrared and shortwave infrared may exist that a hyperspectral sensor can exploit. The differences may be the inevitable dissimilarity between a man-made target and the environment, or they may be intentional, in order to discern one man-made target from another. Invariably, the spectral signatures of targets in CC&D environments may be very similar, and consequently, the ability to distinguish two materials with a high degree of accuracy, as well as certainty, is pivotal.

4.4.1 Two Similar Targets in HYDICE Data

In this section, we demonstrate the application of the band selection techniques in Section 4.3, ADM and MDM, in order to increase the discriminability of two target classes that are very similar spectrally. Our goal is to select bands that maximize the separability of two similar materials and to improve classification results with real data.

We will, again, refer to the two targets as material X and material Y. As part of the effort to provide ground-truth for the experiment, the reflectance spectra for X and Y were measured with a spectroradiometer. Figure 24(a) plots the six measured reflectance spectra (after resampling to the wavelengths of the sensor) for material X and the three measured reflectance spectra for material Y. In comparison, Figure 24(b) illustrates twelve reflectance spectra (after atmospheric compensation) from one target panel of each type as measured by the sensor when it was flown at an altitude of 5000 feet yielding approximately 1m × 1m pixels. These pixels were specifically identified to be full pixels of their respective materials by careful pixel-by-pixel analysis and ground-truth diagrams. It is worth noting the presence of a multiplicative scaling in the spectra in Figure 24(b).



(a) (b)

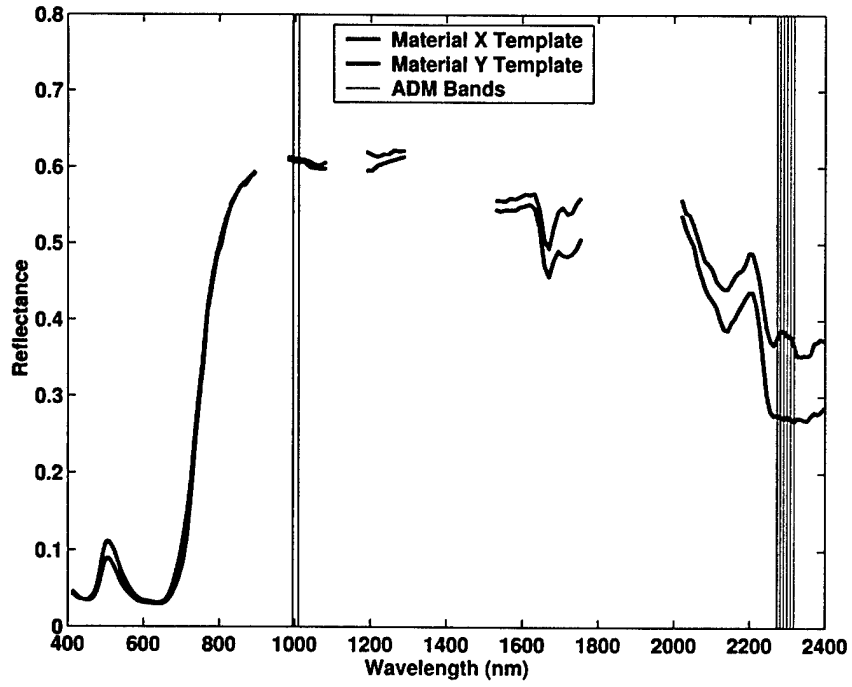
Figure 24. Data from Forest Radiance I, Run 05: (a) Reflectance spectra from a spectroradiometer for material X (blue) and material Y (red); (b) Atmospherically compensated data from HYDICE sensor collected at 5000 feet.

Upon examination of Figure 24(a), the difference between the two classes of spectra is most apparent in the spectral interval from 2200nm to 2400nm. There is also a discernible difference in the amplitude of the reflectance peak at 500nm. The reference spectra in each target class maintains the same distinctions in these intervals, but the separation between classes in other intervals is not as apparent.

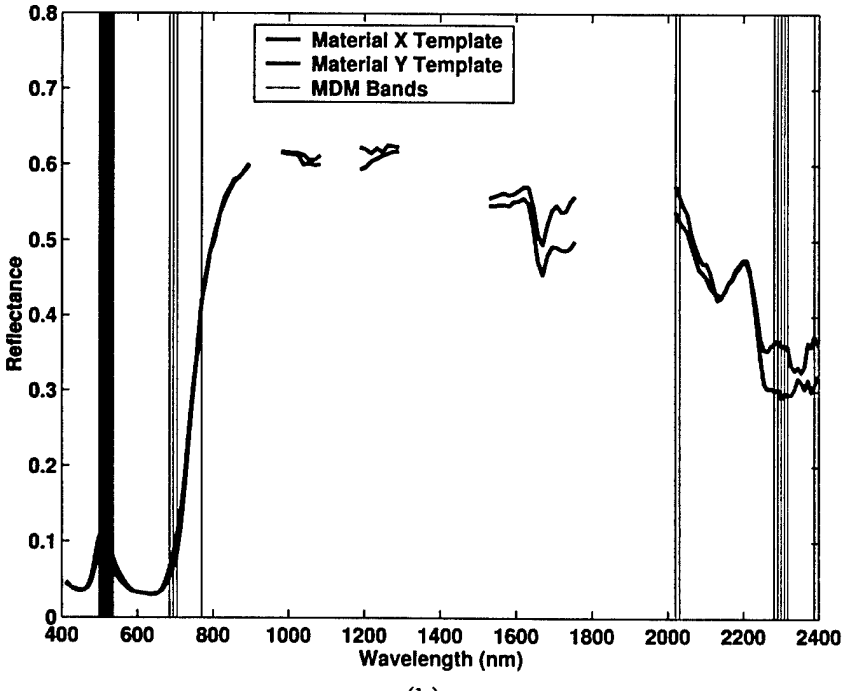
We can now compare the ability of SAM to correctly classify material X and material Y test pixels from the Forest Radiance I Run 05 data collection using 1) all bands and reference spectra means, 2) the ADM templates (class means) and bands, and 3) the MDM templates (worst-case reference pixels) and bands. From the collection of reference spectra, class means were derived by simple averaging to serve as class template spectra. Including the 12 full pixels for each class in Figure 24(b), 17 full pixels for material X and 20 full pixels for material Y were used. Using all 145 bands, the binary test in (28) classified all material X pixels as belonging to material X, but it misclassified every material Y pixel as material X.

ADM and MDM were executed on the reference spectra to select bands and class templates to increase angular separation using the spectra in Figure 24(a). The plots of the template spectra and selected bands for ADM appear in Figure 25(a) and the template spectra and bands for MDM are in Figure 25(b). The ADM band selection chose 8 bands, while MDM selected 20 bands. The bands and templates using ADM performed exactly the same as using all bands, classifying all material X pixels correctly and misclassifying every material Y pixel. MDM, however, correctly classified every pixel from both classes. Table 15 summarizes the results of the band selection and classification.

The same kind of results can be generated for similar targets employed in Desert Radiance II,

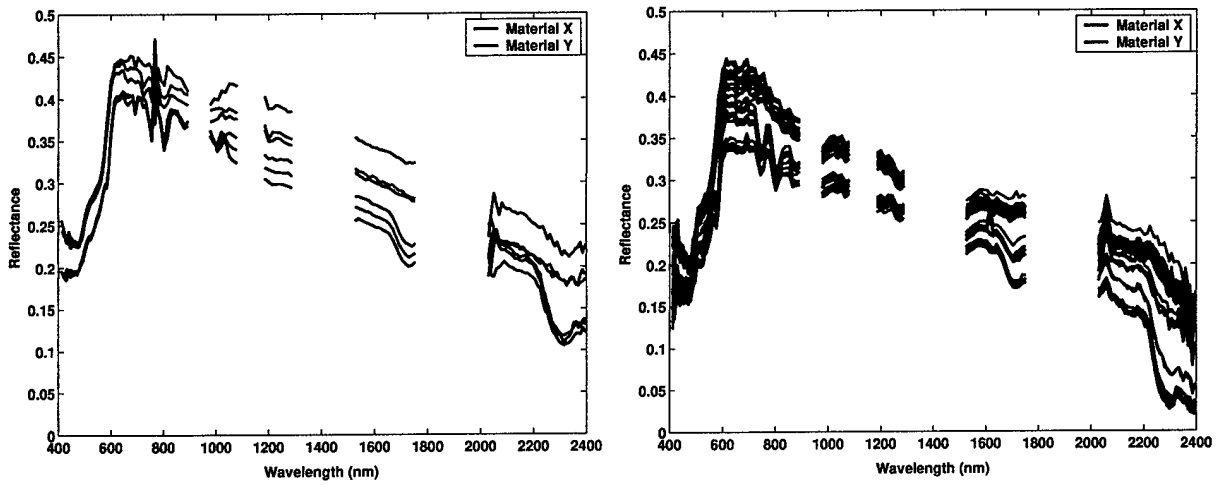


(a)



(b)

Figure 25. Data from Forest Radiance I, Run 05: (a) Template spectra for material X (blue) and material Y (red) with bands selected by ADM; (b) Template spectra for material X (blue) and material Y (red) with bands selected by MDM.



(a) (b)

Figure 26. Data from Desert Radiance II, Run 03: (a) Reflectance spectra from a spectroradiometer for material X (blue) and material Y (red); (b) Atmospherically compensated data from HYDICE sensor collected at 5000 feet.

	Bands (nm)	Number of bands	$\theta_w(\mathbf{X}, \mathbf{Y})$ (^{o})	Correct Classification	
				Material X	Material Y
ADM	{682, 702, 2316, 2289, 2307, 2298, 766, 2333, 2325, 2342, 2280, 2377, 2271, 2351, 2403, 2261}	16	10.168	15/17	17/17
MDM	{2028, 2316, 2069, 2307, 2289, 2059, 766, 2342, 2298, 2333, 2271, 2325, 2280}	13	11.526	16/17	17/17
All bands	{400 – 2405}	145	5.041	15/17	17/17

TABLE 16. Results of band selection and binary classification test to discriminate material X and material Y using data from Desert Radiance II, Run 03.

where the panels are tan in color, instead of green in Forest Radiance I. The reference spectra and sample spectra are in Figure 26(a) and Figure 26(b), respectively. The band selection for ADM and MDM are in Figure 27(a) and Figure 27(b). While using all bands succeeded in correctly classifying all pixels except two, MDM misclassified only one pixel, while only utilizing a fraction of the bands. ADM also misclassified two pixels. The results are summarized in Table 16.

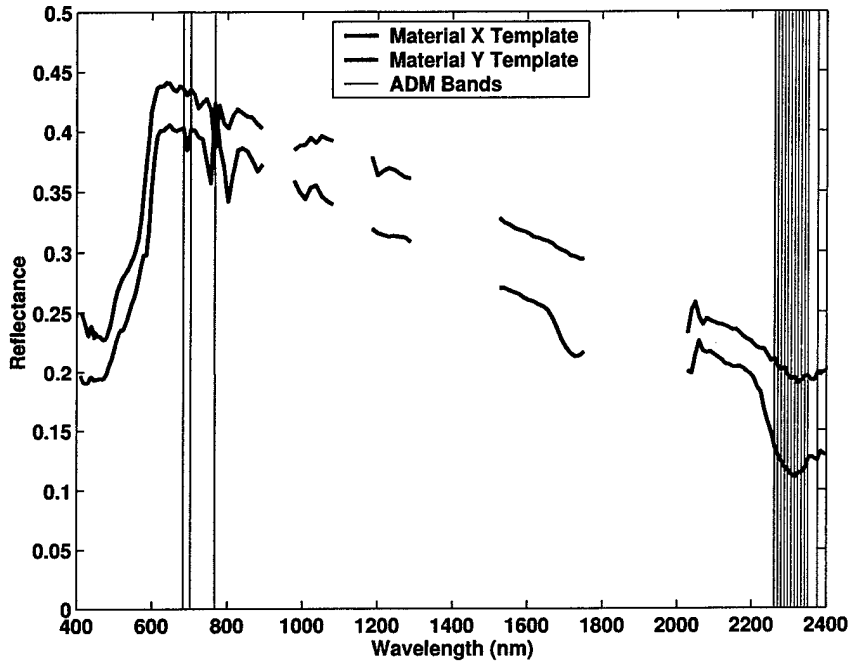
Despite employing two complementary philosophies, the band selection in Figure 25 and Figure 27 demonstrate that both approaches share common subsets of bands, while choosing others to meet their optimization criteria. The order in which the bands were selected by BAO was preserved in Table 15 and Table 16, and demonstrates that while ADM and MDM may start with different pairs of bands, they often converge on common bands, exploiting and inducing the phenomenology in those bands that increases the angular separation and yields superior classification performance.

4.5 DISCUSSION

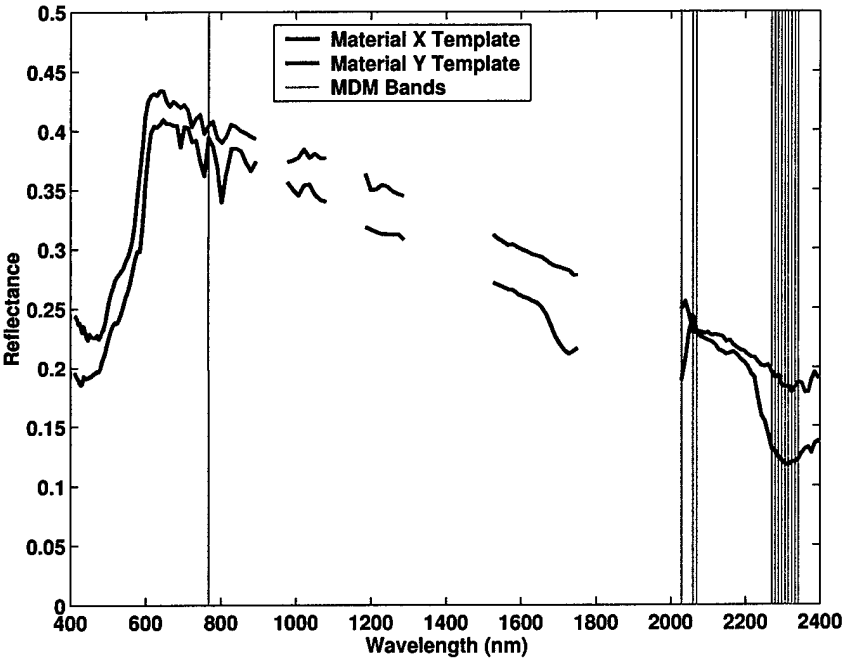
The results in Section 4.4.1 demonstrate how metric-driven band selection can help distinguish two spectrally similar target classes in a realistic, and noisy, sensing environment. Moreover, the larger point proven is that some applications may perform better using only a subset of the spectral information collected by the sensor. This is certainly apparent from the band selection illustrations in Figures 25 and 27 and Tables 15 and 16. Superior classification performance was achieved using only a fraction of the collected bands.

Of the two approaches, ADM more closely resembles the traditional approach of performing angular classification. The mean of the reference measurements provides the template spectra for a class, and ADM simply augments this by using a subset of bands on the template. It does not endeavor to bound the worst-case performance the way MDM does by making the template spectra for a class the one that is most "at-risk."

The small collection of laboratory reference measurements for a target class do not at all provide the best description of the target variability. It is not hard to find targets whose spectra vary more dramatically. To counteract other sources of variability requires a model. One possibility is to model the variability as arising from mismatch between the actual atmospheric conditions and the parameters used to perform atmospheric compensation. MODTRAN [5] is capable of performing these simulations, and in conjunction with an atmospheric compensation program, different reflectance estimates can be recovered, providing the inputs for ADM or MDM.



(a)



(b)

Figure 27. Data from Desert Radiance II, Run 03: (a) Template spectra for material X (blue) and material Y (red) with bands selected by ADM; (b) Template spectra for material X (blue) and material Y (red) with bands selected by MDM.

4.6 SECTION SUMMARY

In this section, we extended the BAO algorithm investigated in Section 3.5 to select bands that increase the angular separation of two classes whose spectral signature varies. The strong parallelism between this capability and the task of performing highly reliable and robust material identification using spectral libraries was stated as a practical motivation. Two techniques both based on BAO were explored to select bands and template signatures that may be used in an angle-based classifier. One approach, the Average Distance Method (ADM), simply selects bands that maximize the average angle created by the reference spectra in both classes. The second technique, the Minimum Distance Method (MDM), selects bands that improve the angular separation between the spectra in each class that are most likely to be misclassified. The applicability of the techniques to difficult CC&D problems was discussed, and the desire to maximize the angular separability of spectrally similar classes using band selection was motivated. ADM and MDM were applied to the task of accurately discriminating two spectrally similar materials using laboratory measurements and HYDICE sensor data from two data collections. Figure 25 and Figure 27 illustrate the template spectra and the selected bands based on the laboratory measurements. Table 15 and Table 16 illustrate the improvement in classification results over employing all bands when the bands and templates are used to classify actual pixels from both classes collected from Forest Radiance I and Desert Radiance II. Superior classification performance was achieved while using only a fraction of the available bands. Improvements in performance and robustness can be achieved through better models of class variability.

5. MATERIAL IDENTIFICATION AND SPECTRAL LIBRARIES

In this section, we extend the results for discriminating between two classes to the more general task of classifying an unknown pixel spectrum into one of many classes. The fundamental unit of this architecture will be the ability to discriminate two classes of spectra which was developed in Section 4.

In developing this architecture based on angle-based measures, it will become clear that there are many opportunities to streamline and expedite the processing. Some past efforts focused on arranging the entries in a spectral library in clusters that expedite efficient comparisons, while choosing a composite of different measures of similarity [2]. Other approaches focus on modelling target variability with simulations of different atmospheric conditions [17]. Spectral reflectance libraries may contain thousands of spectra, and when timely and accurate answers are required in real, operational scenarios, the ability to leverage gains in performance and efficiency from the fundamental properties of the mathematical operators is critical.

5.1 ARCHITECTURES FOR ANGLE-BASED MATERIAL ID

The most common technique for matching reflectance spectral with template spectra in a library utilizes SAM to provide sequential pairwise comparisons between the unknown spectrum, $\mathbf{r}$, and each of K library templates, $\mathbf{t}_i, i = 1, \ldots, K$, and chooses the material having the smallest distance,

$$\text{Class}(\mathbf{r}) \quad = \quad \underset{1 \leq i \leq K}{\text{argmin}} \, \theta(\mathbf{r}, \mathbf{t}_i). \tag{37}$$

Hence, for every unknown pixel spectrum, there must be K angle calculations. The linear architecture that describes this procedure appears in Figure 28 for the case of four classes, $K = 4$. Each test utilizes the exact same set of bands, and each SAM comparison is performed independently, and oblivious to, the other comparisons. A comparator collects every angle measurement and assigns the unknown pixel to the class having the smallest angle, or based on additional criteria, leaves the pixel unassigned.

This linear structure, however, is a specific case of a more general, hierarchical architecture that appears in Figure 29 for $K = 4$. Figure 29(b) describes the basic kernel of the architecture. At each stage the unknown pixel is compared to only two classes at a time using a set of bands and two templates that optimize the current binary test. A comparator rejects from further consideration the class having the larger SAM angle, and another binary test is formulated with the retained class and a new class, using bands and templates that optimize the new test. The procedure continues until only one class remains. Hence, $K - 1$ stages are required, and each stage consists of two angle calculations.

Unlike the linear architecture where the sequence of angular comparisons is irrelevant, the generalized architecture is hierarchical and sequences subsequent comparisons based on the outcome of the current comparison. The key difference between the architectures in Figure 28 and Figure

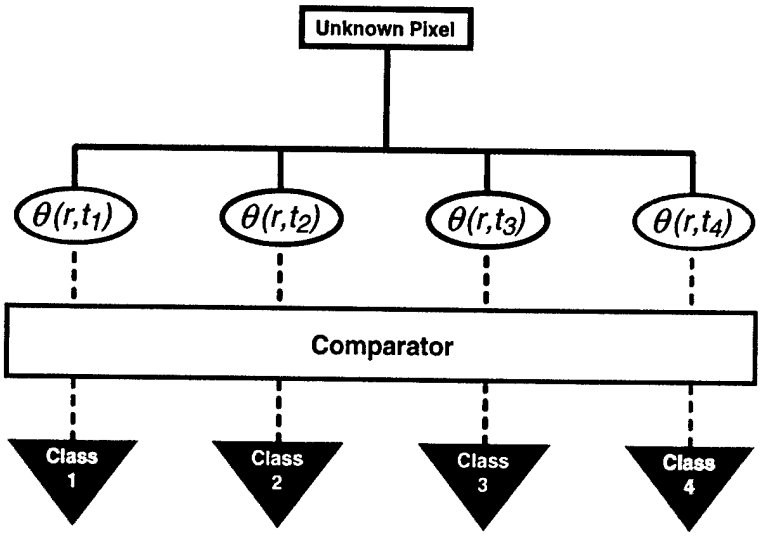


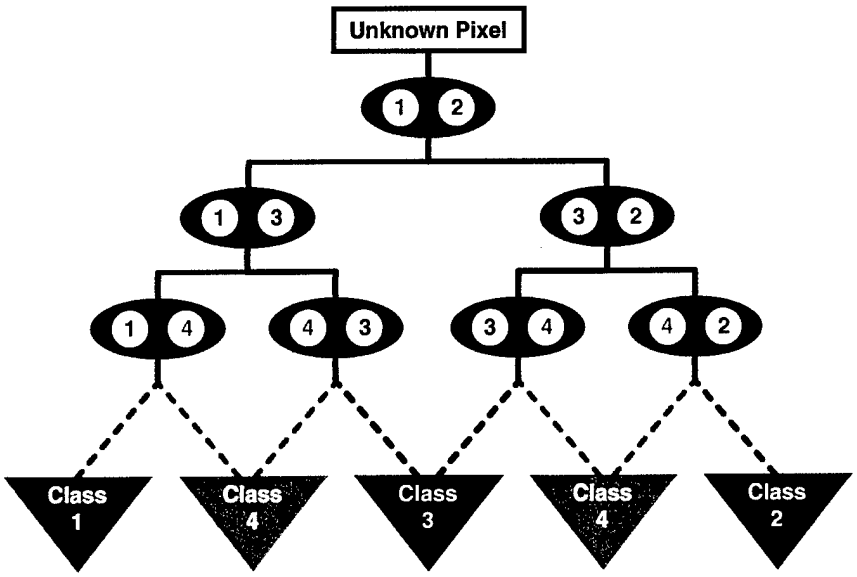
Figure 28. Linear architecture for material identification with spectral libraries.

29(a) is that the hierarchical architecture optimizes each binary test with the appropriate bands and templates to reveal the most contrast between the two classes. Moreover, the sequencing of subsequent tests can be formulated to efficiently arrive at the correct class with a minimum of computation.

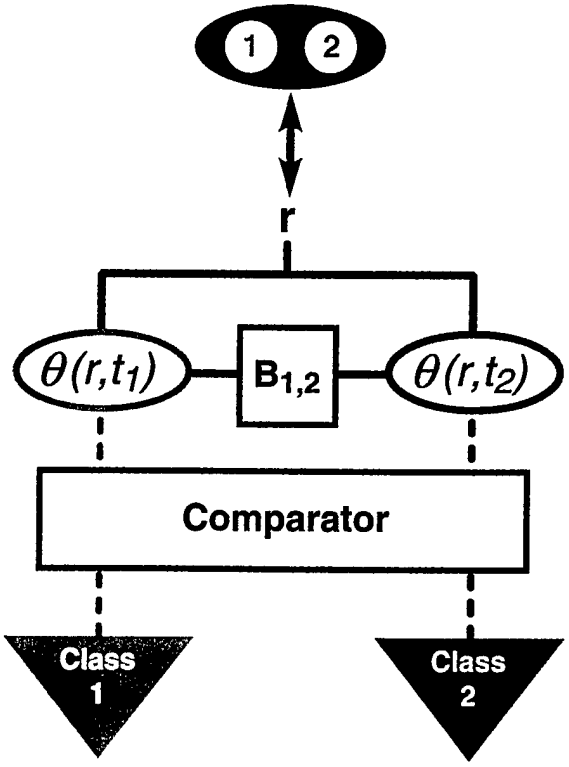
5.2 MULTI-CLASS MATERIAL ID WITH HYDICE DATA

We consider a multi-class experiment using data collected by the HYDICE sensor. Each of the classes has corresponding reference measurements taken by a spectroradiometer, from which both ADM and MDM have selected bands. We use ten classes ($K = 10$), and consequently, $\binom{10}{2} = \frac{10!}{8!2!} = 45$ sets of bands and templates are required. Using ground-truth, the locations of full pixels from each of the target classes have been verified in the atmospherically compensated sensor data. These pixels will serve as inputs to the classifier that utilizes the hierarchical architecture in Figure 29(a) and the 45 sets of bands and template spectra for ADM and MDM.

We chose ten target types that were all similar in their visible appearance, due to the desire to camouflage their appearance in the natural environment. The mean spectra for each class based on the spectroradiometer measurements is illustrated in Figure 30. Table 17 documents for Forest Radiance 1, Run 05, the angular classification results using all spectral bands, the MDM method, and the ADM method. For each class, the number of test pixels used is indicated, and then the percentage of correct classifications (P_{CC}) for that class is indicated for each method of band selection. Also shown are the average number of bands utilized. When using all bands, an unknown pixel requires K SAM angle comparisons in order to be classified. However, for both MDM and ADM, $K-1$ tests are performed that each require two SAM angle computations. For the results in Table 17, each SAM angle computation for MDM and ADM utilizes, on the average, only 15 and



(a)



(b)

Figure 29. (a) Hierarchical architecture for material identification with spectral libraries; (b) Kernel for binary SAM test using distinct bands and templates.

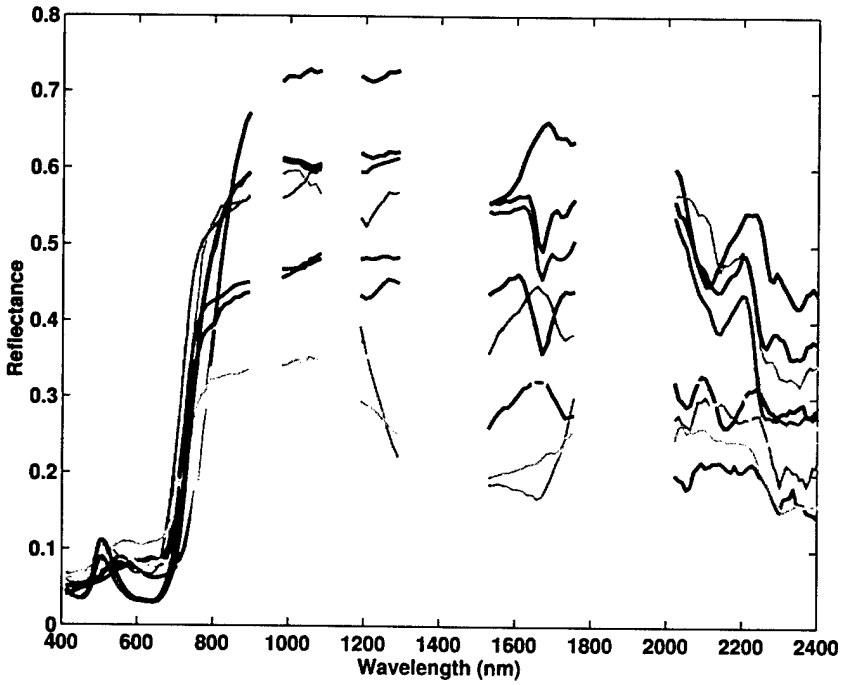


Figure 30. Mean reference spectra for ten target classes from Forest Radiance I, Run 05.

24 bands respectively. At the bottom of the table, the number of classes for which each of the three methods achieves the best/worst performance is provided. For most classes, band selection yields better classification performance than using all bands. Table 18 conveys the similar conclusions for Forest Radiance 1, Run 16, which like Run 05, was collected at an altitude of 5000 feet. Table 19 and Table 20 document similar results for Forest Radiance 1, Run 07 and Run 22, respectively, which were both collected at an altitude of 10000 feet.

5.3 DISCUSSION

The process of selecting bands for MDM or ADM must be done for all possible target pairs to be used in the hierarchical classification system in Figure 29. MDM needs a starting pair of bands, which requires a search over all inter-class reference pixel pairs and all band pairs. The subsequent procedure for adding bands also requires further searches and assessments for candidate bands. Likewise, ADM needs a search over all possible band pairs.

Although Tables 17 and 18 demonstrate that MDM and ADM provides superior classification, using all bands gave better results in some cases. Both MDM and ADM are greedy searches that add bands until no bands exist that provide additional angular separation. The occasional consequence of this strategy is that the selection of bands can terminate prematurely, selecting an extremely low number of bands (< 5). In such a case, this set of bands may not provide sufficient robustness to target variability, and an alternative may be to re-select bands on a less greedy pathway.

Target class	Number of pixels	P_{CC}		
		All bands	MDM	ADM
1	26	38%	31	31
2	18	0	6	0
3	18	56	39	50
4	21	43	10	33
5	18	22	44	100
6	16	63	100	56
7	266	46	79	94
8	258	58	98	98
9	16	50	81	81
10	16	0	81	6
Win/Lose		3/6	5/3	4/2
Avg. no. of bands used		145 (All)	15	24

TABLE 17. Probability of Correct Classification (P_{CC}) using all bands, MDM, ADM. Data was from HYDICE Forest Radiance I, Run 05, collected at 5000 feet. Win/Lose corresponds to the number of classes for which a technique achieves the comparatively best or worst P_{CC} for a class.

Target class	Number of pixels	P_{CC}		
		All bands	MDM	ADM
1	16	100%	88	100
2	18	0	0	0
3	16	94	100	94
4	18	100	100	100
5	25	24	0	8
6	22	82	100	100
7	293	70	100	99
8	166	93	99	93
9	15	6	73	100
10	12	0	100	0
Win/Lose		2/6	5/2	3/2
Avg. no. of bands used		145 (All)	13	23

TABLE 18. Probability of Correct Classification (P_{CC}) using all bands, MDM, ADM. Data was from HYDICE Forest Radiance I, Run 16, collected at 5000 feet. Win/Lose corresponds to the number of classes for which a technique achieves the comparatively best or worst P_{CC} for a class.

Target class	Number of pixels	P_{CC}		
		All bands	MDM	ADM
1	5	20%	20	20
2	5	0	0	0
3	5	60	60	60
4	6	0	0	0
5	5	40	0	0
6	6	33	100	100
7	67	3	97	91
8	114	39	93	76
9	5	0	40	40
10	7	0	43	0
Win/Lose		1/5	5/1	2/2
Avg. no. of bands used		145 (All)	13	25

TABLE 19. Probability of Correct Classification (P_{CC}) using all bands, MDM, ADM. Data was from HYDICE Forest Radiance I, Run 07, collected at 10000 feet. Win/Lose corresponds to the number of classes for which a technique achieves the comparatively best or worst P_{CC} for a class.

Target class	Number of pixels	P_{CC}		
		All bands	MDM	ADM
1	5	0%	20	20
2	8	0	0	0
3	5	20	20	20
4	5	0	0	0
5	8	25	0	88
6	8	88	100	63
7	78	12	100	94
8	79	47	92	63
9	3	67	67	100
10	6	0	17	0
Win/Lose		0/5	4/2	3/2
Avg. no. of bands used		145 (All)	15	24

TABLE 20. Probability of Correct Classification (P_{CC}) using all bands, MDM, ADM. Data was from HYDICE Forest Radiance I, Run 22, collected at 10000 feet. Win/Lose corresponds to the number of classes for which a technique achieves the comparatively best or worst P_{CC} for a class.

In other cases where the number of selected bands was acceptable, the selection of bands did not capture discriminating features that were robust enough to enable superior classification. Scrutinizing these results may provide additional insight on improved band selection algorithms. However, in several cases, it is clear that ADM and MDM failed in the same places that using all bands also failed. In such cases, the incorrectly classified pixel may contain artifacts that no method would be able to mitigate, without prior knowledge of such a distortion.

While experiments indicate that they yield better classification performance, ADM and MDM are neither uniquely superior, nor optimal. The techniques for selecting bands that have been outlined for ADM and MDM are amenable to numerous changes, and these have been discussed in detail in previous sections. For instance, the rule for selecting additional bands based on having the lowest value of β can be changed to select bands based on another criterion. The choice of initial bands is also another parameter that can be adjusted. Our efforts are intended to explore a few sample pathways for selecting hyperspectral bands, based on strong mathematical reasoning and repeatable empirical evidence, that convincingly demonstrates that better performance is achievable through a prudent selection of bands.

However, selecting bands that are appropriate to compare two classes leads to the concept of partitioning spectral libraries by their corresponding angular relationships. Considering the architecture in Figure 29, if an unknown pixel, $\mathbf{r}$, is closer to Class A than Class B, can that information rule out consideration of other classes from comparison? The likely answer is yes, with a considerable savings in overall computation, and this savings is achieved by, once again, exploiting the properties of the metric that performs the comparisons of spectra.

5.4 SECTION SUMMARY

In this section, we extended the capability of discriminating between two target classes, each defined by a set of reference spectra, to a hierarchical, multi-class architecture that is suitable for material identification with spectral libraries. The basic kernel for the architecture is a binary SAM angle test that compares an unknown pixel using bands and template spectra unique to the pair of target classes (see Section 4.3.1 and Section 4.3.2). The class with the larger angle is excluded from further consideration and a new binary test is created with the retained class and one of the remaining classes that employs a distinct set of bands and class templates. This hierarchical structure was implemented and compared to the traditional linear architecture in a ten-class material identification test using laboratory reference measurements and measured sensor data collected with the HYDICE sensor. As the experiments with spectrally similar targets showed in Section 4.4, Tables 17 and 18 demonstrate that band selection using ADM and MDM can provide superior material identification performance while using only a fraction of the available bands.

6. FURTHER APPLICATIONS

In Section 4 and Section 5 we employed band selection to increase the angular separation between two classes and created a hierarchical architecture for material identification with spectral libraries. This approach can be implemented as an independent mechanism for classifying unknown pixels that arise from hyperspectral imagery. For instance, material identification can be performed upon a single pixel in a scene. Or, endmember spectra that have been extrapolated from a scene can be compared to template spectra in a library to ascertain the associated material type. Similarly, in a multi-INT environment, material identification can provide complementary physical analysis to sensors, such as SAR, that are well-suited to detecting man-made and metallic objects. In either case, any individual spectrum may be submitted to a material identification architecture, such as the one in Figure 29. The approaches developed in Section 4 and Section 5 can also provide payoffs for other hyperspectral applications based on the ability to maximize the amount of contrast between two classes and the ability to reduce computation.

6.1 FALSE-ALARM MITIGATION FOR DETECTION

Target detection is one of the most important applications of hyperspectral data. In MTI (Moving Target Indicator) radar detection, a moving target is declared when the value in a range-Doppler cell exceeds a threshold. The magnitude of the return from the moving vehicle will exceed that of the surrounding natural background (at the same range-Doppler location) because it will produce more backscatter than its natural surrounding. Thus, detecting moving targets with radar relies on the difference in backscattering coefficients [25].

Detection of hyperspectral targets depends on more than the differences in magnitude. It also depends on the difference in shape between the desired target spectrum and the background. Section 2.3.3 discussed how statistical detectors for hyperspectral processing are frequently based on a measurement of spectral angle. Invariably, any kind of detection involves maximizing the probability of detection (P_D) and minimizing the probability of false alarms (P_{FA}). Perfect detection is only achieved when a threshold may be set that delineates all target test statistics from background test statistics. Figure 31 illustrates target and background test statistic distributions that result from a statistical detector and the relationship that a threshold has with P_D and P_{FA}.

While it was noted in Section 2.3.3 that many statistical detectors are essentially angle-based comparisons, the band selection techniques discussed in Section 4 are not directly applicable to improve detector performance. Statistical detectors extract better detection performance by exploiting the statistical covariance between different band values. Consequently, the correlations between one band and the remaining bands contribute to the separability between the target and background distributions. The important requirement, however, is that the estimate of the background covariance should be a reliable one, and this only occurs when a sufficient number of samples exist to gauge the intra-class variability. The relationship between the number of background training samples, the number of bands, the amount of target variability has been explored analytically [30]. It is important to note that the band selection methods in Section 4 become applicable when there are not enough samples to create a covariance and apply traditional statistical pattern recognition methods.

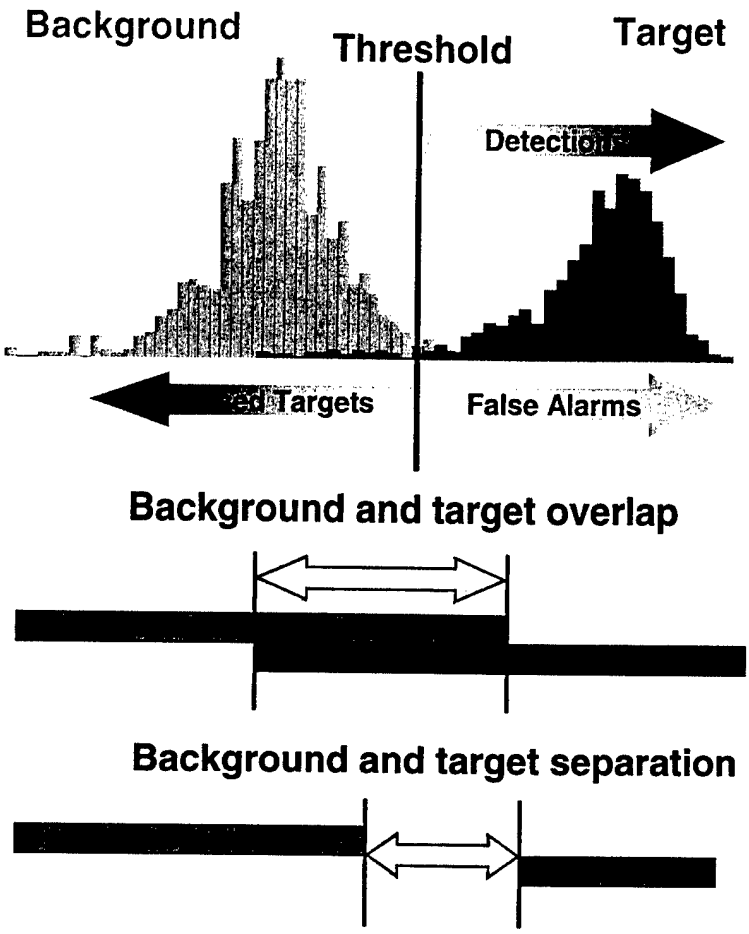
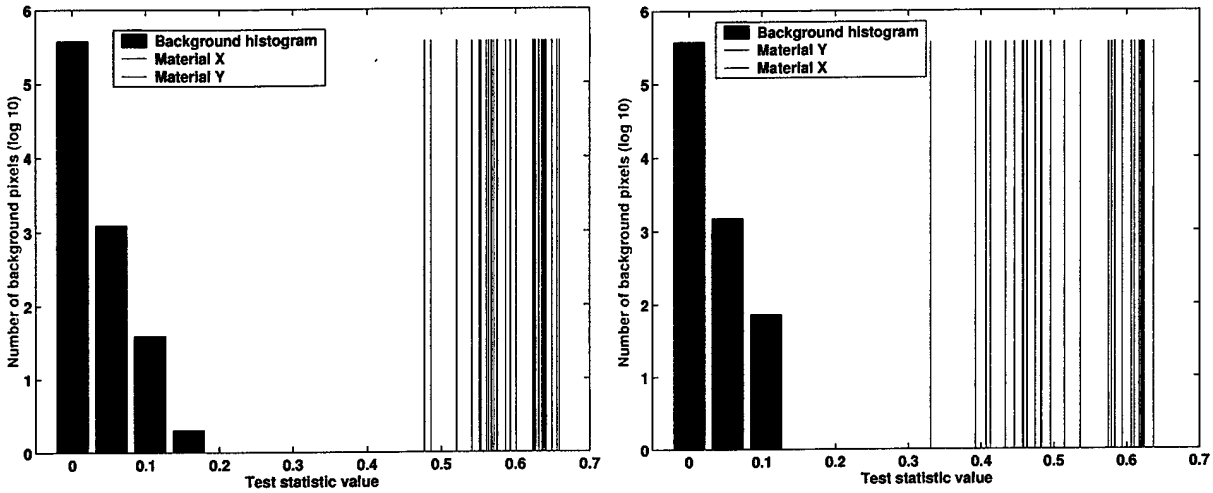


Figure 31. Notional distributions of background and target detection statistics.

Statistical descriptions of background variability, however, are rarely sufficient to completely distinguish targets and background. Nor is knowledge of the optimal location for the threshold always available. Hence, some level of post-detection processing can analyze pixels to minimize P_{FA} and maximize P_D. As we discussed in Section 5, in many CC&D environments, spectral signatures of different targets can be very similar. Statistical target detectors, despite their statistical optimality, can have difficulty distinguishing between two similar target types, and, consequently, lead to a higher P_{FA}. Post-processing detections using the technique in Section 5 is one method of refining the results of statistical detectors.

We can motivate this argument by revisiting the binary classification experiments in Section 4.4 to correctly discern material X and Y. Provided the mean reference spectra for either class in Figure 24(a), any of the statistical detectors discussed in Section 2.3.3 can be employed to adaptively detect the presence of the target amid background. The ACE detector in (16) was



(a) (b)

Figure 32. ACE detection histograms and target test statistics using Forest Radiance I, Run 05 data: (a) The desired target is material X (blue); (b) The desired target is material Y (red). The detector capably distinguishes the desired target from background, but is unable to distinguish similar targets.

run on the Forest Radiance I, Run 05 scene using the mean reference spectra for material X, and then again with the mean reference spectra for material Y. Multiple instances of numerous targets appear in the scene, and as part of an effort at MIT Lincoln Laboratory to provide canonical data sets for testing and evaluation to the hyperspectral community, target and background pixels have been scrupulously corroborated with ground-truth measurements. The covariance was estimated using only the background pixels and the desired target in the scene.

The results of the ACE detector when seeking material X appear in Figure 32(a). The green histogram represents the distribution of test statistics from the background pixels in the scene. The blue lines indicate the test statistics induced by the material X pixels. In red are the statistics for the material Y pixels, and they appear mixed with the material X test statistics. Clearly, the ACE detector is capable of distinguishing the desired target pixels from the background pixels. However, the ACE detector is unable to reject the material Y pixels. A similar result occurs when the ACE detector is employed to detect material Y. In both cases, further post-processing on the detector output is required to discern targets similar to the desired target, and hence, to reduce false alarms. The kind of material identification architecture discussed in Section 5 can be employed to further refine the results of statistical detection and provide more precise identification of pixels.

6.2 DIMENSION REDUCTION

The most significant challenge in hyperspectral processing is to develop automated techniques for exploiting hyperspectral data that achieve optimal performance while processing a minimum amount of measured data. Optimal performance is always desirable from an operational viewpoint,

but minimizing the amount of data required has numerous practical advantages. First, real-time processing is more likely when the amount of data to be processed is minimized, and less latency translates into faster application of results to tactical scenarios. Second, the hardware and software requirements for on-board or off-line processing are loosened, which results in savings in cost and complexity. Finally, the required downlink bandwidth from sensor platforms can be minimized if the measured data is appropriately pre-processed to retain only information that is key for subsequent applications to succeed.

Numerous methods have been undertaken to compress hyperspectral data into efficient lower-dimensional representations, and they borrow much of their intuition from the large research literature devoted to the compression of video imagery. Most of these techniques are centered on optimal statistical representations for scenes using principal components analysis [36, 45], entropy models, and Markov structures [1]. Implicit in most of these approaches is the understanding that compression may lead to some degradation in the performance of applications that subsequently exploit the uncompressed data. This follows from the parallel logic that uncompressed video imagery, at best, will only be an approximation to the original video sequence, and the quality of the reproduction is measured by the human visual system. In hyperspectral processing, however, the usefulness of a dimension reduction approach is not measured by visual inspection, but by more tangible, mathematical measures of application algorithm performance: P_D, P_{FA}, P_{CC}.

Effective dimension reduction, therefore, must take into account the measure(s) of performance that subsequent processing will utilize. Only then will dimension reduction techniques retain information that applications require to succeed. Our efforts to perform band selection have been motivated by the desire to increase the performance of algorithms that use SAM in their processing. In doing so, Tables 15, 16, 17, and 18 demonstrate that an increase in performance can also be accompanied by a dramatic degree of dimension reduction. Compression algorithms whose principle objective is dimension reduction, and not algorithm performance, may provide some form of statistical optimality, but without any consideration of subsequent processing, algorithms that process the reconstructed data will almost invariably underperform. This is especially true for military scenarios which possess a wholly different set of standards and requirements than commercial video processing. So, while dimension reduction was not the intended goal of performing angle-based band selection, it is a concomitant by-product that is nevertheless useful.

7. FUTURE WORK

This project report has documented the importance of a thorough understanding of how distance metrics compare two hyperspectral signals. As an example of the benefits, we have explored how bands can be selected from hyperspectral signals to yield better application performance while using only a fraction of the data collected by the sensor. This capability has important benefits for the design of efficient hyperspectral sensing platforms.

There are numerous opportunities to extend the work in this report, and they are discussed in detail in the following sections.

7.1 BOUNDS ON TARGET VARIABILITY

In Section 4, bands were selected to increase the angular separation between two classes of pixels. Doing this permitted the variability observed in the reference measurements taken by the spectroradiometer to be included. However, these samples do not represent the total amount of variability that can exist for that target class. As noted before, variability can be introduced by the conditions in which the target is observed (e.g., observation angle), the atmospheric conditions, sensor artifacts, and the atmospheric compensation algorithm. Combined, the recovered reflectance spectra from the scene may deviate significantly from the actual material spectra.

One way of incorporating the variability of a spectrum is to provide upper and lower bounds for the reflectance values at each wavelength for a spectrum that demarcates the acceptable range of variability that still defines a material. As an example, Figure 7.1 illustrates a reflectance spectrum with two additional spectra that indicate the upper and lower bound on acceptable reflectance values for that class. If two classes are defined with upper and lower bounds, a high-dimensional volume can be defined describing the range of values for each class, and bands can be selected to reduce each class to a lower-dimensional space while increasing the angular separation between each class.

7.2 PHYSICAL MODELS FOR TARGET VARIABILITY

Another possible way of modelling target variability involves simulating the process by which a material reflectance spectrum is first measured by a sensor as a radiance measurement and is then converted to reflectance by atmospheric compensation. The initial step of moving to radiance can be accomplished by MODTRAN, and the procedure can be repeated with a variety of atmospheric conditions and viewing geometries. This approach has been used to perform hyperspectral processing through forward modelling of reflectance spectra into radiance values [17]. A set of spectra describing the variability can then be obtained by considering all possible pairs of parameters that define the MODTRAN reflectance-to-radiance procedure and the corresponding atmospheric compensation. The recovered reflectance estimates of the original spectrum will then demonstrate the variability that exists when the two procedures are mismatched.

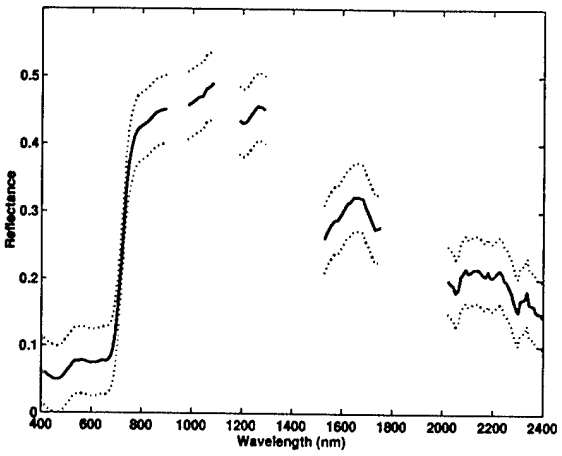


Figure 33. Mean reflectance spectrum for a class with upper and lower bounds.

7.3 FAST ARCHITECTURES FOR SPECTRAL LIBRARIES

In a typical scenario, a spectral library may contain hundreds, even thousands, of spectral signatures. Therefore, the ability to rapidly assign a class label to an unknown pixel becomes a challenge for real-time operation. Band selections for pairs of classes can be calculated off-line and then recalled as needed in the material identification architecture given in Figure 29. We have already demonstrated that band selection invariably leads to dimension reduction, and in most cases, the reduction in bands is significant, thus providing a significant decrease in computation.

In addition, another source of computational savings can be exploited from a spectral library. In a comparison between two classes, the locus of points residing exactly halfway between the two templates provides a partition in high-dimensional space between the two classes. An unknown pixel will necessarily fall on one side of the partition, ruling out the other class from consideration. By virtue of the triangle inequality possessed by valid distance metrics, it is possible to rule out other classes that also fall on the other side of the partition. Thus, for each pair of classes, a list of classes that reside on each side of their partition can also be stored, in addition to the bands that optimize their comparison.

7.4 ALTERNATIVE COST FUNCTIONS

In Section 2.4, the possibility of other distance metrics and cost functions was considered. While few candidates have appeared, the possibility of optimizing other cost functions for hyperspectral processing through band selection still exists. Our efforts have been focused on distance metrics because they provide the foundation for many hyperspectral algorithms. The key factor in evaluating whether a performance measure can be optimized is determining its mathematical properties in the way that the properties of SAM and EMD were enumerated in Table 1.

7.5 TUNABLE SENSING

The ability to identify which bands provide increased discrimination between two classes directly impacts what subset of data collected by a sensor is to be processed. It also has the potential to impact what data is collected by the sensor, if the sensor can be tuned to collect only certain spectral intervals. Electronically tunable filters (ETFs) have utilized sophisticated technologies based on liquid crystal technology, acousto-optic filters, and Fourier transform spectrometers (FTS) [14]. They require, as inputs, the boundaries of the spectral intervals in which they are to collect measurements. A band selection analysis can be adapted to the practical characteristics of a tunable sensor to enable measurements in only spectral intervals that yield the desired discriminability.

7.6 ANGULAR INFORMATION THEORY

Since the angle between two hyperspectral signals can be decomposed into a virtually infinite number of sub-angles, an interesting question is where the information requisite for applications to succeed resides in terms of the associated sub-angles? Just as information theory [37] provided bounds on noise and efficiency for reliable digital communications, there may exist bounds on the spatial and spectral resolution and sensor performance necessary to meet a prescribed performance bound.

8. SUMMARY

In this report, we have derived practical benefits for hyperspectral processing from a thorough mathematical and physical understanding of distance metrics. Most importantly, we have employed a technique for band selection based on mathematical principles to improve the performance of applications that use the Spectral Angle Mapper (SAM) to compare two spectra. Starting with the task of selecting bands to increase the angle between two spectra, we proceeded to extend the approach to select bands that increase the angular separation between two classes of pixels. The strong parallelism between this capability and the problem of material identification using spectral libraries was highlighted. Many examples with spectrally similar targets used for CC&D demonstrated the ability of band selection to provide better classification performance while using only a fraction of the available bands, thereby yielding significant benefits for dimension reduction.

Perhaps the most important by-product of this report is the confirmation that significant performance gains (e.g., P_{CC}, P_D, P_{FA}, computational speed, throughput) can be achieved by a thorough mathematical understanding of the algorithms and operators that are employed to process hyperspectral data. In the case of SAM, we exploited a single property, its non-monotonicity, to achieve gains in classification performance, dimension reduction, and robustness, and we also recognized the applicability of the band selection algorithm to the material identification problem. Similar gains may be possible in other areas, but they will also require a significant exploration of the mathematical behavior, as well as the realistic physical limitations, that underscore the problem.

The gains made in reducing the dimension of the data were by-products of an optimization that focused on maximizing angular separation, and thereby, the overall performance. Optimizations that place a priority only on dimension reduction (or compression) will almost surely lead to a degradation in algorithm performance. So, while optimizations of end-performance are invariably more complex than straightforward, statistical compression techniques (e.g., PCA, JPEG), they are more likely to deliver better performance for the application for which they are optimized.

There are numerous extensions to the research presented in this report. The few methods for band selection techniques presented are only examples from a general framework based on a mathematical decomposition of SAM. There must also exist many other approaches. Section 7 outlines some of the extensions involving most sophisticated physical modelling of target variability, as well as numerous opportunities to streamline the architecture of spectral libraries. Likewise, band selection can provide the inputs to tunable sensors, thus minimizing not only the data to be processed, but the data to actually be collected.

ACRONYMS

ACE	Adaptive Cosine/Coherence Estimator
ADM	Average Distance Method
AMF	Adaptive Matched Filter
ATREM	Atmosphere Removal Program
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
BAO	Band Add-On
CC&D	Camouflage, Concealment, and Deception
DUSD	Deputy Under Secretary of Defense
ETF	Electronically Tunable Filter
EMD	Euclidean Minimum Distance
ESM	Exemplar Selector Module
FTS	Fourier Transform Spectrometers
GLRT	Generalized Likelihood Ratio Test
HSI	Hyperspectral Imaging
HTAP	Hyperspectral Technology Assessment Program
HYDICE	Hyperspectral Digital Imagery Collection Experiment
HYMSMO	Hyperspectral MASINT Support to Military Operations
LMM	Linear Mixing Model
LSE	Least Squares Error
MAP	Maximum a Posteriori
MDM	Minimum Distance Method
ML	Maximum Likelihood
MSE	Mean Squared Error
MTI	Moving Target Indicator
ORASIS	Optical Real-time Adaptive Spectral Identification System
PCA	Principal Components Analysis
P_{CC}	Probability of Correct Classification
P_D	Probability of Detection
PDF	Probability Density Function
P_{FA}	Probability of False Alarm
RF	Radio Frequency
SAM	Spectral Angle Mapper
SeaWIFS	Sea-viewing Wide Field-of-view Sensor
SNR	Signal to Noise Ratio

REFERENCES

[1] G. P. Abousleman, M. W. Marcellin, and B. R. Hunt. Hyperspectral image compressiong using entropy-constrained predictive trellis coded quantization. *IEEE Transactions on Image Processing*, 6(4):566–573, April 1997.

[2] E. M. Bassett and L. S. Kalman. Material taxonomy for object identification in HYDICE imagery. In *Hyperspectral Remote Sensing and Applications*, volume 2821 of *Proceedings of the SPIE*, pages 268–79, August 1996.

[3] E. M. Bassett and S. S. Shen. Information theory-based band selection for multispectral systems. volume 3118 of *Proceedings of the SPIE*, pages 28–35, August 1997.

[4] M. Bassevile. Distance measures for signal processing and pattern recognition. *Signal Processing*, 18(4):349–368, December 1989.

[5] A. Berk, L.S. Bernstein, G.P. Anderson, P.K. Acharya, D.C. Robertson, J.H. Chetwynd, and S.M. Adler-Golden. MODTRAN cloud and multiple scattering upgrades with application to aviris. *Remote Sensing of the Environment*, 65(3):367–375, 1998.

[6] J. C. Bezdek, R. Ehrlich, and W. Full. FCM: The fuzzy c-means clustering algorithm. *Computers and Geosciences*, 10(2-3):191–203, 1984.

[7] J. Bowles, J. Antoniades, M. Baumback, J. Grossman, D. Haas, P. Palmadesso, and J. Stracka. Real time analysis of hyperspectral data sets using NRL's ORASIS algorithm. In *Proceedings of the SPIE*, pages 38–45, May 1997.

[8] J. Bowles, P. Palmadesso, J. Antoniades, M. Baumback, and L. J. Rickard. Use of filter vectors in hyperspectral data analysis. In *Proceedings of the SPIE*, pages 148–157, May 1995.

[9] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–718, March 1992.

[10] F. Csillag, L. Pasztor, and L. L. Biehl. Spectral band selection for the characterization of salinity status of soils. *Remote Sensing of the Environment*, 43(3):231–242, March 1993.

[11] G.M. Foody and D.P. Cox. Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership model and fuzzy membership functions. *International Journal of Remote Sensing*, 15:619–631, 1994.

[12] K. Fukunaga. *Introduction to Statistical Pattern Recognition.* Academic Press, 1990.

[13] B-C. Gao, K. B. Heidebrecht, and A. F. H. Goetz. Derivation of scaled surface reflectances from AVIRIS data. *Remote Sensing of the Environment*, 44(2-3):165–178, May 1993.

[14] N. Gat. Imaging spectroscopy using tunable filters: A review. In *Wavelet Applications VII*, volume 4056 of *Proceedings of the SPIE*, pages 50–64, April 2000.

[15] G. H. Golub and C. F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, 1996.

[16] J. D. Gorman, N.S. Subotic, and B.J. Thelen. Robust material identification in hyperspectral data via multiresolution wavelet techniques. volume 3662 of *Proceedings of ICASSP*, pages 2805–2808, May 1995.

[17] G. Healey and D. Slater. Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6):2706–2717, November 1999.

[18] SeaWiFS home page. "http://seawifs.gsfc.nasa.gov/SEAWIFS.html".

[19] S. D. Hunt and M. Velez-Reyes. Band selection for lossless image compression. In *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, volume 4381 of *Proceedings of the SPIE*, pages 76–83, April 2001.

[20] T. Kailath. The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15(1):52–60, February 1967.

[21] E. J. Kelly. An adaptive detection algorithm. *IEEE Transactions on Aerospace and Electronic Systems*, 22(5):115–127, March 1986.

[22] M. G. Kendall. *A Course in the Geometry of n Dimensions.* Griffin's Statistical Monographs & Courses. Hafner Publishing Company, Charles Griffin & Company Limited, 42 Drury Lane, London, WC 2, 1961.

[23] N. Keshava and P. Boettcher. On the relationships between physical phenomena, distance metrics, and best bands algorithms in hyperspectral processing. In *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VII*, volume 4381 of *Proceedings of the SPIE*, pages 55–67, April 2001.

[24] N. Keshava, J. Kerekes, D. Manolakis, and G. Shaw. An algorithm taxonomy for hyperspectral unmixing. In *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, volume 4049 of *Proceedings of the SPIE*, pages 42–63, April 2000.

[25] N. Keshava, S. M. Kogon, and D. Manolakis. The relationship between detection algorithms for hyperspectral and radar applications. In *Ninth Annual Adaptive Sensor Array Processing Workshop*, March 2001.

[26] S. Kraut and L. Scharf. The CFAR adaptive subspace detector is a scale-invariant GLRT. *IEEE Transactions on Signal Processing*, 47:2538–2541, 1999.

[27] S. Kraut, L. Scharf, and L.T. McWhorter. Adaptive subspace detectors. *IEEE Transactions on Signal Processing*, 49(1):1–16, 2001.

[28] C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems.* Series in Automatic Computation. Prentice-Hall, 1974.

[29] D. Manolakis, G. Shaw, and N. Keshava. Comparative analysis of hyperspectral adaptive matched filter detectors. In *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, volume 4049 of *Proceedings of the SPIE*, pages 2–17, April 2000.

[30] D. Manolakis, C. Siracusa, and G. Shaw. Hyperspectral subpixel target detection using the linear mixing model. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1392–1409, July 2001.

[31] J. N. McDonald and N. A. Weiss. *A Course in Real Analyis.* Academic Press, 1999.

[32] J. C. Price. Band selection procedure for multispectral scanners. *Applied Optics*, 33(15):3281–3288, May 1994.

[33] J. C. Price. Spectral band selection for visible-near infrared remote sensing: Spectral-spatial resolution tradeoffs. *IEEE Transactions on Geoscience and Remote Sensing*, 35(5):1277–1285, September 1997.

[34] F.C. Robey, D.R. Fuhrmann, E.J. Kelly, and R. Nitzberg. A CFAR adaptive matched filter detector. *IEEE Transactions on Aerospace and Electronic Systems*, 28(1):208–218, January 1992.

[35] S. Ross. *A First Course in Probability.* MacMillan Publishing Company, 866 Third Avenue, New York, NY 10022, third edition, 1988.

[36] J. A. Saghri, A. G. Tescher, and J. T. Reagan. Practical transform coding of multispectral imagery. *IEEE Signal Processing Magazine*, 12(1):32–43, January 1995.

[37] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication.* University of Illinois Press, 1963.

[38] C. Sheffield. Selecting band combinations from multispectral data. *Photogrammetric Engineering and Remote Sensing*, 51(6):681–687, June 1985.

[39] V. K. Shettigara, D. O'Mara, T. Bubner, and S. G. Kempinger. Hyperspectral band selection using entropy and target to clutter ratio measures. In *Proc. 10th Australasian Remote Sensing and Photogrammetry Conference, Adelaide, Australia*, volume I, pages 1008–18, August 2000.

[40] D. L. Snyder, J. A. O'Sullivan, D. R. Fuhrmann, and W. H. Smith. Estimation of overlapping spectral signatures from hyperspectral data. In *Proceedings of the SPIE*, volume 3718, pages 470–479, April 1999.

[41] D. M. Y. Sommerville. *An Introduction to the Geometry of N Dimensions.* Dover Books on Physics, Engineering. Dover Publications, Inc., 920 Broadway, New York 10, NY, 1958.

[42] A. Tarantola and B. Valette. Generalized nonlinear inverse problems solved using the least squares criterion. *Review of Geophysics and Space Physics*, 20:219–232, 1982.

[43] A. Tarantola and B. Valette. Inverse problems: Quest for information. *Review of Geophysics and Space Physics*, 50:159–170, 1982.

[44] H. L. Van Trees. *Detection, Estimation, and Modulation Theory: Part I.* John Wiley and Sons, 1968.

[45] V. D. Vaughn and T. S. Wilkinson. System considerations for multispectral image compression designs. *IEEE Signal Processing Magazine*, 12(1):19–31, January 1995.

[46] D. J. Wiersma and D. A. Landgrebe. Analytical design of multispectral sensors. *IEEE Transactions on Geoscience and Remote Sensing*, GE18(2):180–189, April 1980.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	18 December 2002	Project Report

4. TITLE AND SUBTITLE

Distance Metrics and Band Selection in Hyperspectral Processing with Applications to Material Identification and Spectral Libraries

5. FUNDING NUMBERS

C — F19628-00-C-0002

6. AUTHOR(S)

N. Keshava

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Lincoln Laboratory, MIT
244 Wood Street
Lexington, MA 02420-9108

8. PERFORMING ORGANIZATION REPORT NUMBER

PR-HTAP-12

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

DUSD (S&T)
Attn: CAPT F. Garcia
Rosslyn Plaza North, Suite 9030
1777 N. Kent St.
Rosslyn, VA 22209

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ESC-TR-2001-080

11. SUPPLEMENTARY NOTES

None

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

In this report, we investigate and exploit the properties of distance metrics in hyperspectral processing to achieve superior algorithm performance as well as dimension reduction. Distance metrics are mathematical operators that provide a scalar measure of similarity for two hyperspectral (vector) signals, and they are at the nucleus of many application algorithms. However, the similarity between two signals, or conversely the contrast, can be measured by various means, and different distance metrics offer distinct notions of similarity. After enumerating the mathematical properties of two common distance metrics, the Spectral Angle Mapper (SAM) and Euclidean Minimum Distance (EMD), it is demonstrated that the angle measured between two spectra by SAM can change as a function of the bands that are selected and omitted. Several band selection algorithms are derived and compared that maximize the angle between two spectra. A technique called Band Add-On (BAO) is introduced that is based on an exact decomposition of SAM to maximize the angle, and its performance is compared to exhaustive searches through all possible band combinations. BAO is extended to increase the angle between two classes of spectra, and the parallelism with the material identification problem is noted. The Average Distance Method (ADM) and the Minimum Distance Method (MDM) are used to select bands that discriminate two spectrally similar target classes, demonstrating that, in comparison to using all bands collected by a hyperspectral sensor, superior discrimination can be achieved using only a fraction of the available bands. Similar results are obtained for a ten-class material identification test. Several potential benefits and areas of future research are enumerated.

14. SUBJECT TERMS

hyperspectral	material	metric
band	identification	spectral
selection	angle	library

15. NUMBER OF PAGES
98

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Same as Report	Same as Report	Same as Report