March 2015

Wavelet-Based Non-Homogeneous Hidden Markov Chain Model For Hyperspectral Signature Classification

Siwei Feng
University of Massachusetts Amherst

WAVELET-BASED NON-HOMOGENEOUS HIDDEN MARKOV CHAIN MODEL FOR HYPERSPECTRAL SIGNATURE CLASSIFICATION

A Thesis Presented

by

SIWEI FENG

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL AND COMPUTER ENGINEERING

February 2015

Electrical and Computer Engineering

WAVELET-BASED NON-HOMOGENEOUS HIDDEN MARKOV CHAIN MODEL FOR HYPERSPECTRAL SIGNATURE CLASSIFICATION

A Thesis Presented

by

SIWEI FENG

Approved as to style and content by:

Marco F. Duarte, Chair

Mario Parente, Member

Patrick A. Kelly, Member

Christopher V. Hollot, Department Chair
Electrical and Computer Engineering

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my advisor Professor Marco F. Duarte and Professor Mario Parente for the continuous support of my study and research during my master years. Their guidance helped me in all my research and writing of this thesis.

Besides my advisors, I would like to thank the another member of my thesis committee: Professor Patrick A. Kelly, for his encouragement, insightful comments, and hard questions.

My sincere thanks also goes to fellow labmate in UMass Remote Sensing Group: Yuki Itoh, Ping Fung and Carl Senecal for the stimulating discussions, and for all the fun we have had in the research time.

Last but not the least, I would like to thank my parents, for giving birth to me at the first place and supporting me spiritually throughout my life.

ABSTRACT

WAVELET-BASED NON-HOMOGENEOUS HIDDEN MARKOV CHAIN MODEL FOR HYPERSPECTRAL SIGNATURE CLASSIFICATION

FEBRUARY 2015

SIWEI FENG

B.Sc., SOOCHOW UNIVERSITY

M.S.E.C.E., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Marco F. Duarte

Hyperspectral signature classification is an approach for which performs classification of the constituent materials at pixel level in a hyperspectral image. The classification procedure can be operated directly on hyperspectral data or performed by using some features extracted from corresponding hyperspectral signatures containing information like signature energe or shape. In this thesis, we describe a technique that applies non-homogeneous hidden Markov chain (NHMC) models to hyperspectral signature classification. The basic idea is to use statistical models (NHMC models) to characterize wavelet coefficients which capture the spectrum structural information at multiple levels. Experimental results show that the approach based on NHMC models outperforms existing approaches relevant in classification tasks.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS ..... **iv**

ABSTRACT ..... **v**

LIST OF TABLES ..... **viii**

LIST OF FIGURES ..... **ix**

CHAPTER

1. INTRODUCTION **1**

2. BACKGROUND **4**

2.1 An Introduction of Hyperspectral Imaging 4
2.2 The Imaging Spectrometer 5
2.3 Spectral Reflectance 6
2.4 Mineral Spectra 6
2.5 Illumination Factors 7

2.5.1 Source Illumination 7
2.5.2 Illumination Geometry 7
2.5.3 Shadowing 7

2.6 Match Each Image Spectrum 8

3. RELATED WORK AND MATHEMATICAL BACKGROUNDS **9**

3.1 Spectral Matching Methods 10

3.1.1 Spectral Angle Measure 10
3.1.2 Euclidean Distance Measure 10
3.1.3 Spectral Correlation Measure 11
3.1.4 Spectral Information Divergence Measure 11

3.2 Wavelet Analysis 11
3.3 Advantages of Haar Wavelet 13
3.4 Statistical Modeling of Wavelet Coefficients 14
3.5 Wavelet-based Spectral Matching 16

4. NHMC-BASED SIGNATURE CLASSIFICATION 18

4.1 System Overview 18
4.2 Multi-State Hidden Markov Chain Model 19
4.3 Label Computation 21

5. PROPOSED MODIFICATIONS 23

5.1 Denoising 23
5.2 Comparison of Haar Wavelet with Daubechies-4 Wavelet 23
5.3 Spectral Fluctuation Orientation Characterization 24
5.4 Mixture of Gaussian 25

6. SECOND STAGE EXPERIMENT 32

6.1 Study Data and Performance Evaluation 32
6.2 Results for Comparison Approaches 36

7. THIRD STAGE EXPERIMENT 38

8. CONCLUSION 41

APPENDIX: PROOF OF LEMMA 43

BIBLIOGRAPHY 45

LIST OF TABLES

Table **Page**

6.1 Details on Best-performing Features from Cross Validation Experiments with Several Classifiers 35

6.2 Final Classification Rates on Models with Best CV Classification Accuracy by Using Nearest Neighbor Classifier and Support Vector Machine Classifier 36

6.3 Experimental Results of Comparison Approach which Performs Classification on the Whole Spectrum 37

6.4 Experimental Results of Approach Introduced in [18] 37

LIST OF FIGURES

Figure **Page**

3.1 Top: an example of normalized mineral reflectance (Garnet). Middle: corresponding UWT coefficient matrix (9-level wavelet decomposition) using a Haar wavelet. Bottom: corresponding UWT coefficient matrix 9-level wavelet decomposition) using a Daubechies-4 wavelet 14

4.1 System overview. Top: The NHMC Training Module collects a set of training spectra, computes UWT coefficients for each, and feeds then to a NHMC training unit that outputs Markov model parameters and state labels for each of the training spectra, to be used as classification features. Bottom: The Performance Testing Module considers a test spectrum, obtains its UWT coefficients, and extracts a state array from the NHMC obtained during training. A nearest-neighbors classifier searches for the most similar state array among the training data, and returns the class label for the corresponding spectrum. 19

5.1 CV classification rates of different NHMC modeling approaches. Daubechies-1 wavelets with denoising consistently provides the best performance, with 4 GMM states providing best overall success rate. 25

5.2 Examples of signed state labels as classification features. Top: Example normalized reflectance spectrum (Ilmenite). Second: Corresponding 9-label UWT coefficient matrix using a Haar wavelet. Third: Corresponding state label matrix from an NHMC model using a zero-mean two-state GMM. Blue represents smooth regions, while red represents fluctuations. Bottom: Corresponding features consisting of the state labels with added signs from the Haar wavelet coefficients. Green represents smooth regions, while red represents decreasing fluctuations and blue represents increasing fluctuations. 30

5.3 Comparison of classification features obtained from several statistical models for wavelet coefficients. Top: example normalized reflectance, same as in Fig. 5.2. Second: Corresponding state label matrix from a 2-state GMM NHMC model. Third: Corresponding state label matrix from a 6-state GMM NHMC model. Bottom: Corresponding state label matrix from a MOG NHMC with $k = 6$. 31

6.1 CV classification rates of different NHMC modeling approaches. Top left: NN classifiers with ℓ_1 distance; top right: NN classifier with Euclidean distance; bottom left: NN classifier with cosine similarity; bottom right: SVM classifier. 36

7.1 Classification rates of different NHMC modeling approaches as functions of diminant material percentage using NN classifier with ℓ_1 distance, Euclidean distance and cosine similarity (from top to bottom), respectively. Here "gmm" means Gaussian mixture model (GMM), "MOG" means mixture of Gaussian (MOG). The suffix "sign" means the case of adding Daubechies-1 wavelet coefficient signs to the corresponding state labels. "rivard" means the approach proposed in [18]. "wvlt" corresponds to the approach which performs classification on wavelet coefficients. 40

A.1 Directed graphic model for state labels 44

CHAPTER 1

INTRODUCTION

Hyperspectral remote sensors collect reflected image data simultaneously in hundreds of narrow, adjacent spectral bands which make it possible to derive a continuous spectrum curve for each image cell. Such hyperspectral reflectance curves can provide insight into the on-ground (or near ground) constituent materials in a single remotely sensed pixel.

The identification of ground materials from hyperspectral images often requires comparing the reflectance spectra of the image pixels, extracted endmembers, or ground cover exemplars to a training library of spectra obtained in the laboratory from well characterized samples. There is a rich literature on hyperspectral classification methods [1]. On one hand, many methods rely on nearest neighbor classification schemes based on one of many possible spectral similarity measures to match the observed test spectra with training library spectra. On the other hand, practitioners have designed feature extraction schemes that capture information relevant to be applied to the training and testing spectra, in conjunction with appropriate similarity metrics, in order to discriminate between different materials.

Classification methods based on spectral similarity measures can provide researchers with simple implementation and relatively small computational requirements; however, there is a tradeoff with the amount of storage required for the training spectra as well as with the uneven performance of nearest neighbor methods. For example, in some cases considering the original spectral allows for a large amount of redundant information to be considered while the role of relevant structural features is weakened.

Practitioners recognize several structural features in the spectral curves of each material as "diagnostic" or characteristic of its chemical makeup, such as the position and shape of absorption bands. Several approaches like the Tetracorder [3] have been proposed to encode such characteristics. However, such techniques require the construction of ad-hoc rules to characterize instances of each material while new rules must be created when spectral species which were not previously analyzed are added. Parente *et al.* [16] proposed an approach using parametric models to represent the absorption features. However, it still requires the construction of specific rules to match observations to a training library.

We focus on the facilitation of the information extraction process via the use of mathematical models for hyperspectral signals. Our goal is to encode the features that capture scientifically meaningful cues to discriminate between the spectral of different minerals into numerical features, which are referred to as semantic features. Furthermore, no new rules need to be constructed when mineral species which were not previously analyzed before are added.

Mathematical signal models have been used to represent reflectance spectra. More specifically, models leveraging wavelet decompositions are of particular interest because they enable the representation of structural features at different scale. The wavelet transform is a popular tool in many signal processing applications due to the capability of wavelet coefficients to characterize signal discontinuities at different scales and offsets. As mentioned above, the semantic information utilized by researchers is heavily related to the shape of reflectance spectra, which is succinctly represented in the magnitudes of its wavelet coefficients. A coefficient with large magnitude generally indicates a rapid change in its support while a small wavelet coefficient generally implies a smooth region. Existing wavelet approaches are limited to filtering techniques but do not extract features [18].

In this thesis, we apply hidden Markov models (HMMs) to the wavelet coefficients derived from the observed hyperspectral signals so that the correlations between wavelet coefficients in overlapping spectral ranges and at adjacent scales can be captured by the models. This idea is inspired by the hidden markov tree (HMT) model proposed in [4]. As for the wavelet transform, we use an undecimated wavelet transform (UWT) in order to obtain maximum flexibility on the set of scales and offsets (spectral bands or wavelengths[1]) considered.

Our model for a spectrum encompassing N spectral bands takes the form of a collection of N non-homogeneous hidden Markov chains (NHMCs), each corresponding to a particular spectral band. Such a model provides a map from each signal spectrum to a binary space that encodes the structural features at different scales and wavelengths, effectively representing the semantic features that allow for the discrimination of spectra.

This thesis is organized as follows. Section 2 introduces the mathematical background behind our hyperspectral signature classification system and reviews relevant existing approaches for the hyperspectral classification task. Section 4 provides an overview of the proposed hyperspectral signature classification system as well as the details about the choice of mother wavelet, model training and label computing. Section 6 describes our experimental validation setup as well as the corresponding results. Final conclusions are provided in Section 8.

[1]We use these three equivalent terms interchangeably in the sequel

CHAPTER 2

BACKGROUND

In this section, we first introduce some backgrounds about hyperspectral imaging. Then we discuss several existing spectral matching approaches. Next we review the theoretical background for our proposed hyperspectral signature classification system, including wavelet analysis, hidden Markov chain models, and the Viterbi algorithm.

2.1 An Introduction of Hyperspectral Imaging

In many of the past decade, hyperspectral imaging has been a field of active research and development, and hyperspectral images have been available only to scientifical practitioners. With the appearance of commercial airborne hyperspectral imaging systems in recent years, hyperspectral imaging has been gradually entering the mainstream of remote sensing. Hyperspectral images have found many and will find more applications in resource management, agriculture, mineral exploration, and environmental monitoring. But effective employment of hyperspectral images requires an understanding of the essence and limitations of the data and of various schemes for processing and interpreting it.

Multispectral remote sensors like the Landsat Thematic Mapper and SPOT XS generate images with a few relatively broad wavelength bands. Hyperspectral remote sensors, on the other hand, collect image data simultaneously in dozens or hundreds of narrow, adjacent spectral bands. These measurements make it possible to derive a continuous spectrum for each image pixell. After adjustments for sensor, atmospheric, and terrain effects are applied, these image spectra can be compared with field or

laboratory reflectance spectra in order to recognize and map surface materials such as particular types of vegetation or diagnostic minerals associated with ore deposits.

Hyperspectral images contain a wealth of data, but interpreting them requires an understanding of the exact properties of ground materials we are trying to measure, and how they relate to the measurements actually made by the hyperspectral sensor.

2.2 The Imaging Spectrometer

Hyperspectral images are generated by instruments called imaging spectrometers. The development of these sensors has converged the two related but distinct technologies: spectroscopy and the remote imaging of Earth and planetary surfaces.

Spectroscopy is a discipline about light that is emitted by or reflected from materials and its variation in energy with wavelength. As applied to the area of optical remote sensing, spectroscopy deals with the spectrum of sunlight that is diffusely reflected (or scattered) by materials at the Earths surface. Instruments called spectrometers (or spectroradiometers) are used to make measurements of the light reflected from a test material. An optical dispersing element such as a grating or prism in the spectrometer splits this light into many narrow, adjacent wavelength bands and the energy in each band is measured by a separate detector. By using hundreds or even thousands of detectors, spectrometers can make spectral measurements of bands as narrow as $10nm$ over a wide wavelength range, typically at least 400 to $2400nm$ (visible through middle infrared wavelength ranges).

Remote imagers are designed to focus and measure the light reflected from many adjacent areas on the surface of the earth. In many digital imagers, sequential measurements of small areas are made in a consistent geometric pattern as the sensor platform moves and subsequent processing is required to assemble them into an image. Until recently, imagers were restricted to one or a few relatively broad wavelength bands by limitations of detector designs and the requirements of data storage, trans-

mission, and processing. Recent advances in these areas have allowed the design of imagers that have spectral ranges and resolutions comparable to ground-based spectrometers.

2.3 Spectral Reflectance

In reflected-light spectroscopy the fundamental property that we want to obtain is spectral reflectance: the ratio of reflected energy to incident energy as a function of wavelength. Reflectance varies with wavelength for most materials because energy at certain wavelengths is scattered or absorbed to different degrees. These reflectance variations are obvious when we compare spectral reflectance curves for different materials. The overall shape of a spectral curve and the position and strength of absorption bands in many cases can be used to identify and discriminate different material types. For example, vegetation has higher reflectance in the near infrared range and lower reflectance of red light than soils.

2.4 Mineral Spectra

This is the main object in our research. In inorganic materials such as minerals, chemical composition and crystalline structure control the shape of the spectral curve and the presence and positions of specific absorption bands. Wavelength-specific absorption may be caused by the presence of particular chemical elements or ions, the ionic charge of certain elements, and the geometry of chemical bonds between elements, which is partially governed by the crystal structure.

In the spectrum of hematite (an iron-oxide mineral), the strong absorption in the visible light range is caused by ferric iron (Fe^{+3}). In calcite, the major component of limestone, the carbonate ion (CO_3^{-2}) is responsible for the series of absorption bands between 1800 and 2400nm. Kaolinite and montmorillonite are clay minerals that are common in soils. The strong absorption band near 1400nm in both spectra, along

with the weak $1900nm$ band in kaolinite, are due to hydroxide ions (OH^{-1}), while the stronger $1900nm$ band in montmorillonite is caused by bound water molecules in this hydrous clay. In contrast to these examples, orthoclase feldspar, a dominant mineral in granite, shows almost no significant absorption features in the visible to middle infrared spectral range.

2.5 Illumination Factors

2.5.1 Source Illumination

The incoming solar energy varies greatly with wavelength, peaking in the range of visible light. The spectrum of incoming solar energy at the time an image was acquired must be known, assumed, or derived indirectly from other measurements in order to convert image radiance values to reflectance.

2.5.2 Illumination Geometry

The amount of energy reflected by an area on the ground depends on the amount of solar energy illuminating the area, which in turn depends on the angle of incidence: the angle between the path of the incoming energy and a line perpendicular to the ground surface. Specifically, the energy received at each wavelength (Eg) varies as the cosine of the angle of incidence (θ): $Eg = Eox \cos\theta$, where Eo is the amount of incoming energy. The energy received by any ground area therefore varies as the suns height changes with time. If the surface is not flat, the energy received also varies instantaneously across a scene because of differences in slope angle and direction.

2.5.3 Shadowing

The amount of illumination received by an area can also be reduced by shadows. Shadows cast by topographic features or clouds can affect areas including many contiguous image pixels. Trees, crop rows, rock outcrops, or other small objects can also cast shadows that are confined to an individual image pixel. Both types of shadows

have the effect of lowering the measured brightness across all wavelengths for the affected pixels.

2.6 Match Each Image Spectrum

One approach to analyzing a hyperspectral image is to match each image spectrum individually to one of the reference reflectance spectra in a spectral library. It works best if the scene includes extensive areas of essentially pure materials that have corresponding reflectance spectra in the reference library. An observed spectrum will generally show varying degrees of match to a number of similar reference spectra.

Spectral matching is complicated by the fact that most hyperspectral scenes include many image pixels that represent spatial mixtures of different materials. The resulting composite image spectra may match a variety of pure reference spectra to varying degrees, perhaps including some spectra of materials that are not actually present. If the best-matching reference spectrum has a sufficient fit to the image spectrum, then this material is probably the dominant one in the mixture and the pixel is assigned to this material.

CHAPTER 3

RELATED WORK AND MATHEMATICAL BACKGROUNDS

The shape of a reflectance spectrum can usually be separated into two components: broad, smoothly changing regions that define the general shape of the spectrum and narrow, trough-like absorption features. This distinction leads to two different approaches to matching image spectra with reference spectra.

Many pure materials, such as minerals, can be recognized by their absorption features. One common matching strategy attempts to match only the absorption features in each candidate reference spectrum and ignores other parts. A unique set of wavelength regions is therefore examined for each reference candidate, determined by the locations of its absorption features. The local position and slope of the spectrum can affect the strength and shape of an absorption feature, so these parameters are usually determined relative to the continuum: the upper limit of the spectrums general shape. The continuum is computed for each wavelength subset and removed by dividing the reflectance at each spectral channel by its corresponding continuum value. Absorption features can then be matched using a set of derived values (including depth and the width at half-depth), or by using the complete shape of the feature. These types of procedures have been organized into an expert system by researchers at the U.S. Geological Survey Spectroscopy Lab.

Many other materials, such as rocks and soils, may lack distinctive absorption features. These spectra must be characterized by their overall shape. Matching procedures utilize full spectra or a uniform wavelength subset for all candidate materials.

One approach to matching seeks the spectrum with the minimum difference in reflectance (band per band) from the image spectrum (quantified by the square root of the sum of the squared errors). Another approach treats each spectrum as a vector in spectral space and finds the reference spectrum making the smallest angle with the observed image spectrum.

3.1 Spectral Matching Methods

A direct comparison of spectral similarity measures taken on the observed hyperspectral signals is the easiest and most direct way to do spectral matching. Generally speaking, spectral similarity measures can be combined with nearest neighbor classifiers. In this thesis we introduce four commonly used spectral similarity measures. In the sequel, we use $\boldsymbol{r}_i = (r_{i1}, r_{i2}, ..., r_{iN})^T$ and $\boldsymbol{r}_j = (r_{j1}, r_{j2}, ..., r_{jN})^T$ to denote the reflectance or radiance signatures of two hyperspectral image pixel vectors

3.1.1 Spectral Angle Measure

The spectral angle measure (SAM) [9] between two reflectance spectra is defined as

$$SAM(\boldsymbol{r}_i, \boldsymbol{r}_j) = \cos^{-1}\left(\frac{\langle \boldsymbol{r}_i, \boldsymbol{r}_j \rangle}{\sqrt{||\boldsymbol{r}_i||_2^2 ||\boldsymbol{r}_j||_2^2}}\right).$$

A smaller spectral angle indicates larger similarity between two reflectance spectra.

3.1.2 Euclidean Distance Measure

The Euclidean distance measure (ED) [19] between two reflectance spectra is defined as

$$ED(\boldsymbol{r}_i, \boldsymbol{r}_j) = ||\boldsymbol{r}_i - \boldsymbol{r}_j||_2.$$

As with SAM, smaller ED implies larger similarity between two vectors. The ED measure takes the intensity of two reflectance spectra into account, while the former is invariant with intensity.

3.1.3 Spectral Correlation Measure

The spectral correlation measure (SCM) [20] between two reflectance spectra is defined as

$$SCM(\boldsymbol{r}_i, \boldsymbol{r}_j) = \frac{\sum_{k=1}^{N}(r_{ik} - \bar{r}_i)(r_{jk} - \bar{r}_j)}{\sqrt{\sum_{k=1}^{N}(r_{ik} - \bar{r}_i)^2 \sum_{k=1}^{N}(r_{jk} - \bar{r}_j)^2}}.$$

where $\bar{r}_i$ is the mean of the values of all the elements in a reflectance spectrum vector $\boldsymbol{r}_i$. The SCM can take both positive or negative values; larger positive values are indicative of similarity between spectra.

3.1.4 Spectral Information Divergence Measure

The spectral information divergence measure (SID) [2] between two reflectance spectra is defined as

$$SID(\boldsymbol{r}_i, \boldsymbol{r}_j) = D(\boldsymbol{r}_i||\boldsymbol{r}_j) + D(\boldsymbol{r}_j||\boldsymbol{r}_i),$$

where $D(\boldsymbol{r}_i||\boldsymbol{r}_j)$ is regarded as the relative entropy (or Kullback-Leibler divergence) of $\boldsymbol{r}_j$ with respect to $\boldsymbol{r}_i$, which is defined as

$$D(\boldsymbol{r}_i||\boldsymbol{r}_j) = -\sum_{k=1}^{N} p_{ik}(\log p_{jk} - \log p_{ik}).$$

Here $p_{ik} = r_{ik} / \sum_{k=1}^{N} r_{ik}$ corresponds to a normalized version of the spectrum r_i at the k^{th} spectral band, which is interpreted in the relative entropy formulation as a probability distribution.

3.2 Wavelet Analysis

The wavelet transform of a signal provides a multiscale analysis of a signal's content which effectively encodes the locations and scales at which the signal structure is present in a compact fashion [10]. To date, several hyperspectral classification

methods based on wavelet transform have been proposed. Most of these classification approaches (e.g. [22, 13, 21]) employ a dyadic/decimated wavelet transform (DWT) as the preprocessing step. Compared with UWT, the DWT provides a smaller representation because it minimizes the amount of redundancy in the coefficients. However, the tradeoff for such redundancy is that UWT provides maximum flexibility on the choice of scales and offsets used in the multiscale analysis, which is desired because it allows for a simple characterization of the spectrum structure at each individual spectral band.

A one-dimensional real-valued UWT of an N-sample signal $x \in \mathbb{R}^N$ is composed of wavelet coefficients w_s, each labeled by a scale $l \in 1, ..., L$ and offset $n \in 1, ..., N$, where $S \leqslant N$. The coefficients are defined using inner products as $w_{l,n} = \langle x, \phi_{l,n} \rangle$, where $\phi_{l,n} \in \mathbb{R}^N$ denotes a sampled version of the mother wavelet function ϕ dilated to scale l and translated to offset n:

$$\phi_{l,n}(\lambda) = \frac{1}{\sqrt{l}} \phi \left(\frac{\lambda - n}{l} \right).$$

To improve the interpretability of the notation, we will change our notation for scales in the sequel from $l = 1, 2, \ldots, L$ to $s = L, L-1, \ldots, 1$ (i.e., we reverse the ordering of the scales). With this change, small values of s correspond to coarse scales while large values of s correspond to fine scales. All the coefficients can be organized into a two-dimensional matrix W of size $L \times N$, where rows represent scales and columns represent samples. In this case, each coefficient $w_{s,n}$, where $s < L$, has a child coefficient $w_{s+1,n}$ at scale $s+1$. Similarly, each coefficient $w_{s,n}$ at scale $s > 1$ has one parent $w_{s-1,n}$ at scale $s-1$. Such a chain structure of wavelet coefficients enables the representation of fluctuations in a spectral signature by chains of large coefficients appearing within the columns of the wavelet coefficient matrix W.

3.3 Advantages of Haar Wavelet

The Haar wavelet is the simplest possible compact wavelet which has the properties of square-like shape and discontinuity. These properties makes the Haar wavelet sensitive to a larger range of fluctuations than other mother wavelets and provides it with a lower discriminative power. Thus, the Haar wavelet enables the detection of both slow-varying fluctuations and sudden changes in a signal [10], while it is not particularly sensitive to small continuities (i.e., noise) on a signal, in effect averaging them out over the wavelet support.

Consider the example in Fig. 3.1, where the figure at the top represents an example hyperspectral signature, while the figures in the middle and at the bottom show the undecimated wavelet coefficient matrix of the spectrum under the Haar and Daubechies-4 wavelets, respectively. The middle figure of Fig. 3.1 shows the capability of Haar wavelet transform to capture both rapid changes and gently sloping fluctuations in the sample reflectance spectrum. The Daubechies-4 wavelet transform of a signal is sensitive to compact and drastic discontinuities (i.e., higher order fluctuations). From the bottom figure we can find a high density of large-magnituded wavelet coefficients in the spectral band range from $500nm$ to $900nm$, which corresponds to the rapid increasing slopes of the sample spectrum. Apart from that, the small discontinuities around $800nm$ are particularly clearly recorded and those from $1200nm$ to the end are also detected. However, the Daubechies-4 wavelet tends to show little response to relatively flat regions (i.e., lower order fluctuations). For example, the strong decreasing slope in the range $900nm - 1200nm$ is not reflected in the Daubechies-4 wavelet coefficients, and those gentling sloping fluctuations starting at $1200nm$ do not have an appearance as significant as those between $500nm - 900nm$. The coefficients in those regions mostly focus on small-scale, high-order fluctuations, and do not successfully capture the semantic information in the spectral signature. Thus, it appears that the Daubechies-4 wavelet does not provide a good match to

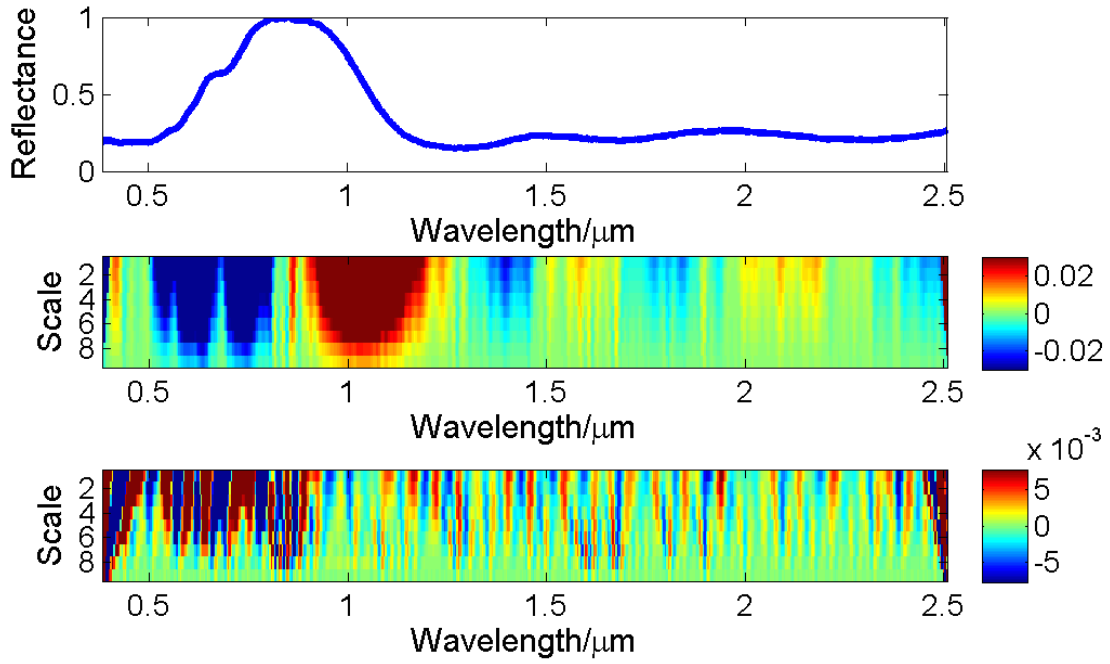

Figure 3.1. Top: an example of normalized mineral reflectance (Garnet). Middle: corresponding UWT coefficient matrix (9-level wavelet decomposition) using a Haar wavelet. Bottom: corresponding UWT coefficient matrix 9-level wavelet decomposition) using a Daubechies-4 wavelet

semantic information extraction for this example reflectance spectrum. Intuitively, these problems will be shared with other higher-order wavelets, which provide good analytical matches to functions with fast, high-order fluctuations.

3.4 Statistical Modeling of Wavelet Coefficients

Crouse *et al.* [4] proposed the use of hidden Markov models (HMM) to capture the statistics of DWT coefficients. In that paper, the dyadic nature of DWT coefficients gives rise to a hidden Markov tree (HMT) model that characterizes the clustering and persistence properties of wavelet coefficients.

The statistical model is motivated by the compression property of the DWT, which states that the wavelet transform of a piecewise smooth signal generally features a small number of large coefficients and a large number of small coefficients. This

property motivates the use of a zero-mean Gaussian mixture model (GMM) with two Gaussian components to capture the compression property, where one Gaussian component with a high-variance characterizes the small number of "large" coefficients (labeled with a state L), while a second Gaussian component with a low-variance characterizes the large number of "small" wavelet coefficients (labeled with a state S). The state $S_s \in \{\mathrm{S}, \mathrm{L}\}$ of a wavelet coefficient[1] is said to be hidden because its value is not explicitly observed. The likelihoods of the two Gaussian components $p_{S_s}(\mathrm{L}) = p(S_s = \mathrm{L})$ and $p_{S_s}(\mathrm{S}) = p(S_s = \mathrm{S})$ should meet the condition that $p_{S_s}(\mathrm{L}) + p_{S_s}(\mathrm{S}) = 1$. The conditional probability of a particular wavelet coefficient w_s given the value of the state S_s can be written as $p(w_s|S_s = i) = \mathcal{N}(0, \sigma_{i,s}^2)$, where $i = \{\mathrm{S}, \mathrm{L}\}$, and the distribution of the same wavelet coefficient can be written as $p(w_s) = p_{S_s}(\mathrm{L})\mathcal{N}(0, \sigma_{\mathrm{L},s}^2) + p_{S_s}(\mathrm{S})\mathcal{N}(0, \sigma_{\mathrm{S},s}^2)$.

In cases where a UWT is used, the persistence property of wavelet coefficients [11, 12] (which implies the high probability of a chain of wavelet coefficients to be consistently small or large across adjacent scales) can be accurately modeled by a non-homogeneous Markov chain (NHMC) that links the states of wavelet coefficients in the same offset. This means the state S_s of a coefficient w_s is only affected by the state S_{s-1} of its parent (if it exists) and by the value of its coefficient w_s. The Markov chain is completely determined by the likelihoods for the first state and the set of state transition matrices for the different parent-child label pairs (S_{s-1}, S_s) for $s > 1$:

$$A_s = \begin{pmatrix} p_{\mathrm{S}\to\mathrm{S},s} & p_{\mathrm{L}\to\mathrm{S},s} \\ p_{\mathrm{S}\to\mathrm{L},s} & p_{\mathrm{L}\to\mathrm{L},s} \end{pmatrix}, \tag{3.1}$$

where $p_{i\to j,s} := P(S_s = j|S_{s-1} = i)$ for $i, j \in \{\mathrm{L}, \mathrm{S}\}$. The training process of an HMM is based on the expectation maximization (EM) algorithm which generates a

[1]Since the same model is used for each chain of coefficients $\{S_{1,n}, \ldots, S_{L,n}\}$, $n = 1, \ldots, N$, we remove the index n from the subscript for simplicity in this sequel whenever possible.

set of HMM parameters $\boldsymbol{\theta} = \{p_{S_1}(\mathrm{S}), p_{S_1}(\mathrm{L}), \{A_s\}_{s=2}^{L}, \{\sigma_{\mathrm{S},s}, \sigma_{\mathrm{L},s}\}_{s=1}^{L}\}$ including the probabilities for the first hidden states, the state transition matrices, and Gaussian variances for each of the states. We define the $L \times N$ matrix S the collection of state values for all scales and spectral bands. The iterative parts of the algorithm can be briefly described as follows:

1. **E step**: Perform maximum likelihood estimation of the state labels using a forward-backward algorithm [17]:

$$S^l = \arg\max_S p(S|W, \boldsymbol{\theta}^l);$$

   this joint conditional probability mass function (PMF) will be used in the M step.

2. **M step**: Update model parameters to maximize the expected value of the joint likelihood of the wavelet coefficients and state estimates [4]:

$$\boldsymbol{\theta}^{l+1} = \arg\min_{\boldsymbol{\theta}} E_S[\ln f(W, S|\boldsymbol{\theta}^l)|W, \boldsymbol{\theta}^l].$$

3. Set $l = l + 1$. If converged, then stop; otherwise, repeat.

3.5 Wavelet-based Spectral Matching

Many hyperspectral signature classification approaches have been proposed in the literature, with a subset of them involving wavelet analysis. In this thesis, we review three approaches that are particularly close in scope to our proposed method, which will be used for comparison in our numerical experiments.

First, Rivard et al. [18] propose a method based on the wavelet decomposition of the spectral data. The obtained wavelet coefficients are separated into two categories: low-scale components of power (LCP) capturing mineral spectral features

(corresponding to the first fine scales), and high-scale components of power (HCP) containing the overall continuum (corresponding to later coarser scales). The coefficients for the LCP spectrum, which capture detailed structural features, are summed across scales at each spectral band. This process can conceptually be described as a filtering approach, since the division into LCP and HCP effectively acts as a high-pass filter that preserves only the fine-scale detailed portion of the spectrum. The proposed method then implements classification using a nearest neighbors approach with a SAM metric.

The next two wavelet-based classification approaches use the DWT to perform wavelet analysis. West et al. [21] propose a method based on a DWT decomposition of each spectrum. The set of wavelet detail coefficients for each separate scale is considered as a separate feature vector. Linear discriminate analysis (LDA) is performed on each one of these vectors for dimensionality reduction purposes; each output vector is then fed into a maximum likelihood classifier, and a majority vote is taken among the classification results corresponding to vectors from different scales. While this method provides a simple pipeline for processing the wavelet coefficients on a scale by scale basis, it ignores the fact that the coefficients across scales exhibit correlations that are not considered by the classifiers and majority vote fusion steps.

As an alternative, Zhang et al. [22] aim to capture the statistics of the DWT coefficients. Their approach trains a HMT (that is, an HMM model for DWT coefficients) on the coefficients of the spectra from each of the classes. A maximum likelihood classifier then evaluates the likelihood of the wavelet coefficients being observed under the HMT trained for each of the classes, and selects the class that provides the maximum value of the probability. It is implicitly expected by this method that the number of training samples for each one of the classes is sufficiently large so that the class-specific HMTs can be accurately trained.

CHAPTER 4

NHMC-BASED SIGNATURE CLASSIFICATION

In this section we introduce a hyperspectral signature classification system that exploits a Markov model for the wavelet coefficients for purposes of feature extraction. The Haar wavelet is used in a UWT to capture information on the fluctuations of the spectra. The state labels extracted from the Markov model are collected as classification features for each of the spectra. The features capture the semantic information that is used by practitioners to distinguish spectra from different minerals.

4.1 System Overview

We provide an overview of the hyperspectral classification system in Fig. 4.1. The system consists of two modules: an NHMC model training module and a performance testing module. While the figure assumes a binary-state Gaussian mixture model (GMM) in the NHMC, as described in Section 3.4, we explore the performance of the system for k-ary states extensions, $k = 2, 3, \ldots$. The training stage uses a training library of spectra containing samples from the classes of interest and runs them through the UWT. The wavelet representations are then used to train a single NHMC model, which is then used to compute state estimates for each of the training spectra using a Viterbi Algorithm. The state arrays will then be used as classification features coupled with a nearest-neighbor approach. The testing module considers a spectrum under test and computes the state estimates under the trained NHMC model using the parameters obtained during training. The module then searches for

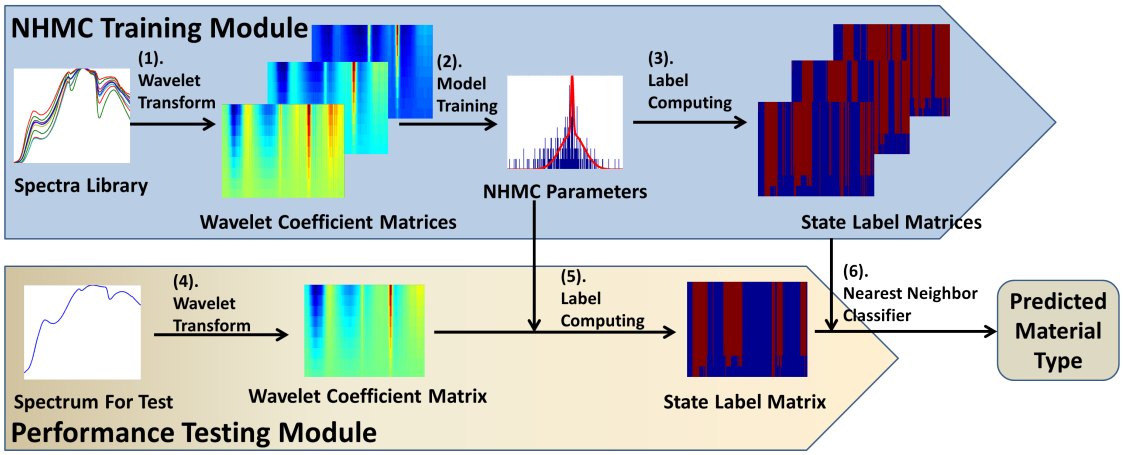

Figure 4.1. System overview. Top: The NHMC Training Module collects a set of training spectra, computes UWT coefficients for each, and feeds then to a NHMC training unit that outputs Markov model parameters and state labels for each of the training spectra, to be used as classification features. Bottom: The Performance Testing Module considers a test spectrum, obtains its UWT coefficients, and extracts a state array from the NHMC obtained during training. A nearest-neighbors classifier searches for the most similar state array among the training data, and returns the class label for the corresponding spectrum.

the most similar state array among the training dataset, returning the class label of the selected training spectrum.

4.2 Multi-State Hidden Markov Chain Model

In our system, we choose to use the NHMC model described in Section 3.4 applied to the UWT via the Haar wavelet. We select the Haar wavelet due to its special shape, which allows for the magnitude of the wavelet coefficients to be proportional to the slope of the spectra across the wavelet's support. Furthermore, the signs of these coefficients are indicative of the slope orientation (increasing or decreasing for positive and negative, respectively).

In contrast to the prior work of [4], we design our NHMC to feature k-state GMMs for the wavelet coefficients. We increase the number of states from 2 to $k > 2$ because a two-state zero-mean GMM provides an overly coarse distinction between sharper absorption bands (fluctuations) and flatter regions in a hyperspectral signature, which

are usually assigned large and small state labels, respectively. In our cases of interest, spectrum classification requires a labeling granularity for the signature fluctuations that is finer than that achieved by binary labels.

We associate each wavelet coefficient w_s with an unobserved hidden state $S_s \in \{0, 1, ..., k-1\}$, where the states have prior probabilities $p_{i,s} := p(S_s = i)$ for $i = 0, 1, ..., k-1$. Here the state $i = 0$ represents smooth regions of the spectral signature, in a fashion similar to the small (S) state for binary GMMs, while $i = 1, \ldots, k-1$ represent a more finely grained set of states for spectral signature fluctuations, similarly to the large (L) state in binary GMMs. All the weights should meet the condition $\sum_{i=0}^{k-1} p_{i,s} = 1$. Each state is characterized by a zero-mean Gaussian distribution for the wavelet coefficient with variance $\sigma_{i,s}$. The value of S_s determines which of the k components of the mixture model is used to generate the probability distribution for the wavelet coefficient w_s: $p(w_s|S_s = i) = \mathcal{N}(0, \sigma_{i,s}^2)$. We can then infer that $p(w_s) = \sum_{i=0}^{k-1} p_{i,s} p(w_s|S_s = i)$. In analogy with the binary GMM case, we can also define a $k \times k$ transition probability matrix

$$A_s = \begin{pmatrix} p_{0\to 0,s} & p_{1\to 0,s} & \cdots & p_{k-1\to 0,s} \\ p_{0\to 1,s} & p_{1\to 1,s} & \cdots & p_{k-1\to 1,s} \\ \vdots & \vdots & \ddots & \vdots \\ p_{0\to k-1,s} & p_{1\to k-1,s} & \cdots & p_{k-1\to k-1,s} \end{pmatrix},$$

where $p_{i\to j,s} = p(S_s = j|S_{s-1} = i)$. Note that the probabilities in the diagonal of A_s are expected to be larger than those in the off-diagonal elements due to the persistence property of wavelet transforms. Note also that all state probabilities $p_{i,s}$ for $s > 1$ can be derived from the matrices $\{A_s\}_{s=2}^{L}$ and $\{p_{i,1}\}_{i=0}^{k-1}$.

The training of the k-GMM NHMC is also performed via an EM algorithm. Because of the overlap between wavelet functions at a fixed scale and neighboring offsets, adjacent coefficients may have correlations in relative magnitudes [14]. However, for

computational reasons, in this thesis we only consider the parent-child relationship of the wavelet coefficients in the same offset. Namely, we train an NHMC separately on each of the N wavelengths sampled by the hyperspectral acquisition device. The set of NHMC parameters $\boldsymbol{\theta}_n$ of a certain spectral band n include the probabilities for the first hidden states $\{p_{i,1,n}\}_{i=0}^{k-1}$, the state transition matrices $\{A_{s,n}\}_{s=2}^{L}$, and the Gaussian variances $\{\sigma_{0,s,n}^2, \sigma_{1,s,n}^2, \ldots, \sigma_{k-1,s,n}^2\}_{s=1}^{L}$.

4.3 Label Computation

Given the model parameters $\boldsymbol{\theta}$, the state label values $\{S_s\}_{s=1}^{L}$ for a given observation are obtained using a Viterbi algorithm [17, 4]. For a particular wavelet coefficient w_s, a k-dimensional conditional probability vector is defined with elements being the conditional PMF of the wavelet coefficient

$$p(w_s|S_s = i) = \frac{1}{\sqrt{2\pi\sigma_s^2}} \exp\left(-\frac{{w_s}^2}{2\sigma_s^2}\right)$$

under each possible state value $i = 0, \ldots, k-1$. A variable $\delta_{i,s}$ is defined as the "best score" that ends in a particular state i at scale s from its previous state, while the variable $\psi_{i,s}$ is the most likely state at a particular scale $s-1$ to have children s with state i. The definitions of the two variables are

$$\psi_{i,1} = 0, \tag{4.1}$$

$$\delta_{i,1} = p_{i,1} \cdot p(w_1|S_1 = i), \tag{4.2}$$

$$\psi_{i,s} = \arg \max_{j=0,\ldots,k-1} (\delta_{j,s-1} p_{j \to i,s}), \tag{4.3}$$

$$\delta_{i,s} = \delta_{\psi_{i,s},s-1} p_{\psi_{i,s} \to i,s} \cdot p(w_s|S_s = i), \tag{4.4}$$

for $i = 1, \ldots, k-1$ and $s = 2, \ldots, L$. The algorithm also returns the likelihood $p(W|\boldsymbol{\theta})$ of a wavelet coefficient matrix W under the model $\boldsymbol{\theta}$ as a byproduct. We

propose the use of the state label array S as classification features for the original hyperspectral signal x. It is easy to identify the presence of such features simply by inspecting the labels obtained from the NHMC.

CHAPTER 5

PROPOSED MODIFICATIONS

5.1 Denoising

As mentioned earlier, in order to eliminate the differences between spectra caused by illumination conditions [5], we perform normalization on the whole database by dividing each reflectance spectrum by its maximum value. As a result of the normalization of the spectra in the database before applying the wavelet transform, the impact caused by small discontinuities (i.e., noise) might be enlarged, especially for some relatively flat spectra like galena. Such effect can be reduced by performing denoising on the normalized spectra. We use soft thresholding denoising [6], a technique that applies a threshold on the maximum value of the signal coefficients and that is commonly applied to wavelet representations.

5.2 Comparison of Haar Wavelet with Daubechies-4 Wavelet

We expand on the original classification experiments in [7, 15] by increasing the spectral complexity of the classification. We sample reflectance spectra from the USGS remote sensing database at AVIRIS wavelengths, totaling 244 samples with 26 spectral classes, including reflectance spectra of minerals, vegetation, and other materials. Each category contains at least 5 samples.

We compare NHMC models with a numbers of GMM mixtures/states varying from one to seven. We first randomly separate the database into a training set (including 197 samples with each category containing no less than 4 samples) and a test set (including 47 samples with each category containing no less than 1 sample).

In order to evaluate the performance of different NHMC models, we implement a 4-fold *cross validation* (CV) on the training set. We use three folds at a time to generate the parameters of the NHMC model. Then we use the Viterbi algorithm to obtain the corresponding state labels for both the training and test set and use nearest neighbor classification (via Hamming distance on the state labels) on the remaining fold to evaluate which of the models obtained by varying the number of states and by considering or not the wavelet signs achieves the highest (average) CV classification accuracy. Finally, we select the best performing model (in terms of the CV performance) and train it on the entire training set. We then use nearest neighbor classification (via Hamming distance on the state labels) on the test set to evaluate the models' generalization performance. For denoising, we performed a line search for the threshold value that provided best performance in our spectral matching task, finding it to be $\tau = 0.05$.

Our experimental results are shown in Figure 5.1. Overall, the classification accuracy for NHMC models using the Daubechies-1 wavelet are higher than that of Daubechies-4. The model achieving the highest classification rate (94.9%) uses the Daubechies-1 wavelet with denoising.

5.3 Spectral Fluctuation Orientation Characterization

As mentioned above, because of the shape of the Haar wavelet function, the signs of Haar wavelet coefficients of a reflectance spectrum capture whether the slopes increases or decreases as a function of wavelength. This characteristic of Haar wavelet coefficients can be utilized to design state labels that capture the slope orientations of the corresponding reflectance spectra. Thus, we make a simple modification by adding the sign of a Haar wavelet coefficient to its counterpart in the corresponding state label matrix. Figure 5.2 shows the effect of adding signs to state label matrices. The top two figures represent the reflectance spectrum of a sample material and

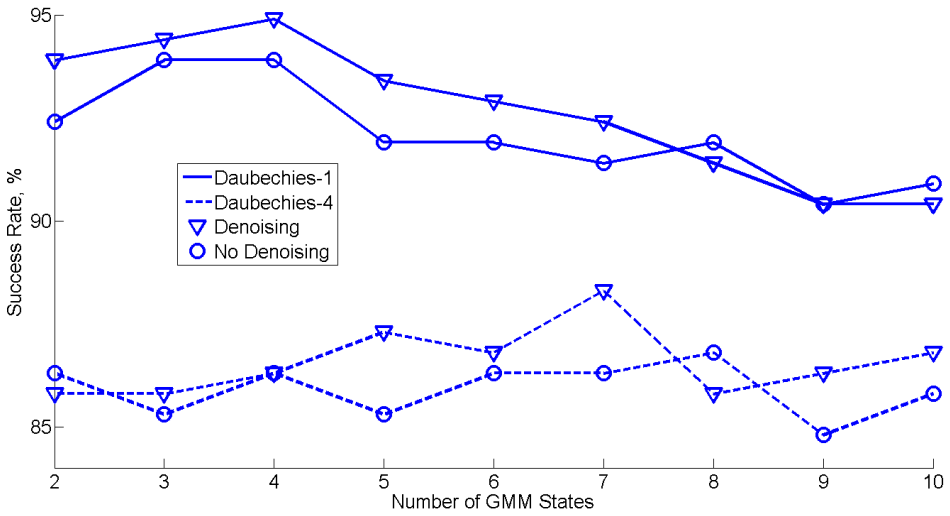

Figure 5.1. CV classification rates of different NHMC modeling approaches. Daubechies-1 wavelets with denoising consistently provides the best performance, with 4 GMM states providing best overall success rate.

the corresponding Haar wavelet coefficient matrix, while the bottom two show the corresponding state label matrices with and without being added wavelet coefficient signs, respectively. The figure shows that the fluctuations in the region $600nm - 800nm$ are predominantly not detected by state labels. Furthermore, one can see many narrow chains of "large" state labels starting at $1700nm$. Increasing the number of GMM state enables a finer-scale quantization of spectral signature fluctuations, which is somewhat analogous to increasing the quantization resolution for our wavelet analysis. This is quite important when the Haar wavelet is used due to its sensitivity to a large range of fluctuation orders, which implies a relatively low discriminative power when compared with higher-order wavelet transforms.

5.4 Mixture of Gaussian

Unfortunately, a large number of GMM states might also have negative influence on classification results. The GMM state of a particular wavelet coefficient $w_{s,n}$ is determined by the coefficient's magnitude with respect to those for the rest of the

NHMC training spectra, the state label of its parent $S_{s-1,n}$, and the transition probability matrix $A_{s,n}$. In practice, this dependence translates to the establishment of different ranges for the values of the coefficients in each GMM state across scales and offsets (s, n). This variance often makes it difficult to assess the semantic information in the label array of a spectral signature due to the variability in the value-to-state mappings among adjacent scales and bands. Furthermore, as the granularity of the GMM labeling keeps increasing, one runs the risk of characterizing the spectral variability of a given mineral using a multitude of GMM state labels. Thus, we may desire a model that both characterizes the presence of spectral fluctuations (high discriminability for low wavelet coefficient magnitudes) while being invariant to the presence of spectral variability in a given element (low discriminability for large wavelet coefficient magnitudes).

We propose a solution that combines the advantages of a binary-state GMM and a k-state GMM. Our modified wavelet coefficient statistical model consists of a binary-state NHMC with a "small" state (0) modeled by a standard nonzero-mean Gaussian distribution and a "large" state (1) modeled by a mixture of k-1 Gaussian distributions. Note that we use numbers here instead of letters for the state labels to distinguish between the 2-state GMM NHMC and the 2-state MOG NHMC. We denote this modified model mixture of Gaussians (MOG) NHMC in the sequel. As desired, this modified model maintains the discriminability between smooth regions and absorption bands in spectral signatures, while providing classification features (binary labels, in this case) that are invariant to the presence of spectral variability among the signatures for a given class in the training and testing sets.

Next, we describe several connections between the k-state GMM NHMC and the MOG NHMC that allow for a simple transition in the development of our classifier. Consider the transition from the GMM NHMC model that yields state labels $S_s \in \{0, \ldots, k-1\}$ to a MOG NHMC that yields state labels $Z_s \in \{0, 1\}$ with probabilities

$q_{i,s} = P(Z_s = i)$, $i = 1, 2$. One can show that the change of models lead to the following mapping for labels:

$$Z(S) = \begin{cases} 0 & \text{if } S = 0, \\ 1 & \text{if } S \neq 0. \end{cases} \tag{5.1}$$

Similarly to (3.1), we can define a transition probability matrix

$$B_s = \begin{pmatrix} q_{0\to 0,s} & q_{1\to 0,s} \\ q_{0\to 1,s} & q_{1\to 1,s} \end{pmatrix}$$

for the MOG NHMC, where $q_{i\to j,s} := P(Z_s = j | Z_{s-1} = i)$ for $i, j \in \{0, 1\}$ and $s = 1, \ldots, L$. We have the following intuitive result.

Lemma 1 *Denote the vector of state probabilities for a wavelet coefficient w_s under the k-state GMM NHMC as $\boldsymbol{P}_s = (p_{0,s}, p_{1,s}, ..., p_{k-1,s})^T$. The corresponding vector of probabilities for the MOG NHMC states $\boldsymbol{Q}_s$ can be written as follows:*

$$\boldsymbol{Q}_s = (q_{0,s}, q_{1,s})^T = \left(p_{0,s}, \sum_{i=1}^{k-1} p_{i,s} \right)^T = (p_{0,s}, 1 - p_{0,s})^T .$$

We also have the following lemma, proven in Appendix A.

Lemma 2 *The elements of the MOG NHMC transition matrix $\boldsymbol{B}_s$ can be written in terms of the elements of the GMM NHMC transition matrix $\boldsymbol{A}_s$ as follows:*

$$q_{0\to 0,s} = p_{0\to 0,s}, \tag{5.2}$$

$$q_{1\to 0,s} = \frac{\sum_{i=1}^{k-1} p_{i\to 0,s} p_{i,s-1}}{\sum_{i=1}^{k-1} p_{i,s-1}}, \tag{5.3}$$

$$q_{0\to 1,s} = \sum_{j=1}^{k-1} p_{0\to j,s}, \tag{5.4}$$

$$q_{1\to 1,s} = \frac{\sum_{i=1}^{k-1} p_{i,s-1} \sum_{j=1}^{k-1} p_{i\to j,s}}{\sum_{i=1}^{k-1} p_{i,s-1}}. \tag{5.5}$$

Here i and j represent state labels ranging from 1 to $k-1$.

Below is an example of the transform of a state probability vector and transition probability matrix, respectively, where the original number of state is 4:

$$(0.422, 0.3696, 0.1042, 0.1042)^T \to (0.422, 0.578)^T,$$

$$\begin{pmatrix} 1 & 0.0001 & 0 & 0 \\ 0 & 0.9999 & 0 & 0 \\ 0 & 0 & 0.5 & 0.4999 \\ 0 & 0 & 0.5 & 0.5001 \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Correspondingly, we also make small modifications to the label computation scheme from Section 4.3. For the MOG NHMC, equations (4.1–4.4) become

$$\begin{aligned} \psi_{i,1} &= 0, \\ \delta_{i,1} &= q_{i,1} \cdot p(w_1|Z_1 = i), \\ \psi_{i,s} &= \arg\max_{j=0,1}(\delta_{j,s-1} q_{j\to i,s}), \\ \delta_{i,s} &= \delta_{\psi_{i,s},s-1} q_{\psi_{i,s}\to i,s} \cdot p(w_s|Z_s = i), \end{aligned}$$

respectively, for $i = 0, 1$ and $s = 2, \ldots, L$. The required conditional probabilities involving Z_s can be written as given in the following lemma.

Lemma 3 *The state-conditional probabilities for the MOG NHMC can be given in terms of the state-conditional probabilities for the GMM NHMC as follows:*

$$\begin{aligned} p(w_s|Z_s = 0) &= p(w_s|S_s = 0), \\ p(w_s|Z_s = 1) &= \frac{\sum_{i=1}^{k-1} p_{i,s} p(w_s|S_s = i)}{\sum_{i=1}^{k-1} p_{i,s}}, \end{aligned}$$

where i denotes a state label ranging from 1 to $k-1$.

We provide an example comparison between labels obtained from the GMM NHMC and the MOG NHMC in Fig. 5.3.

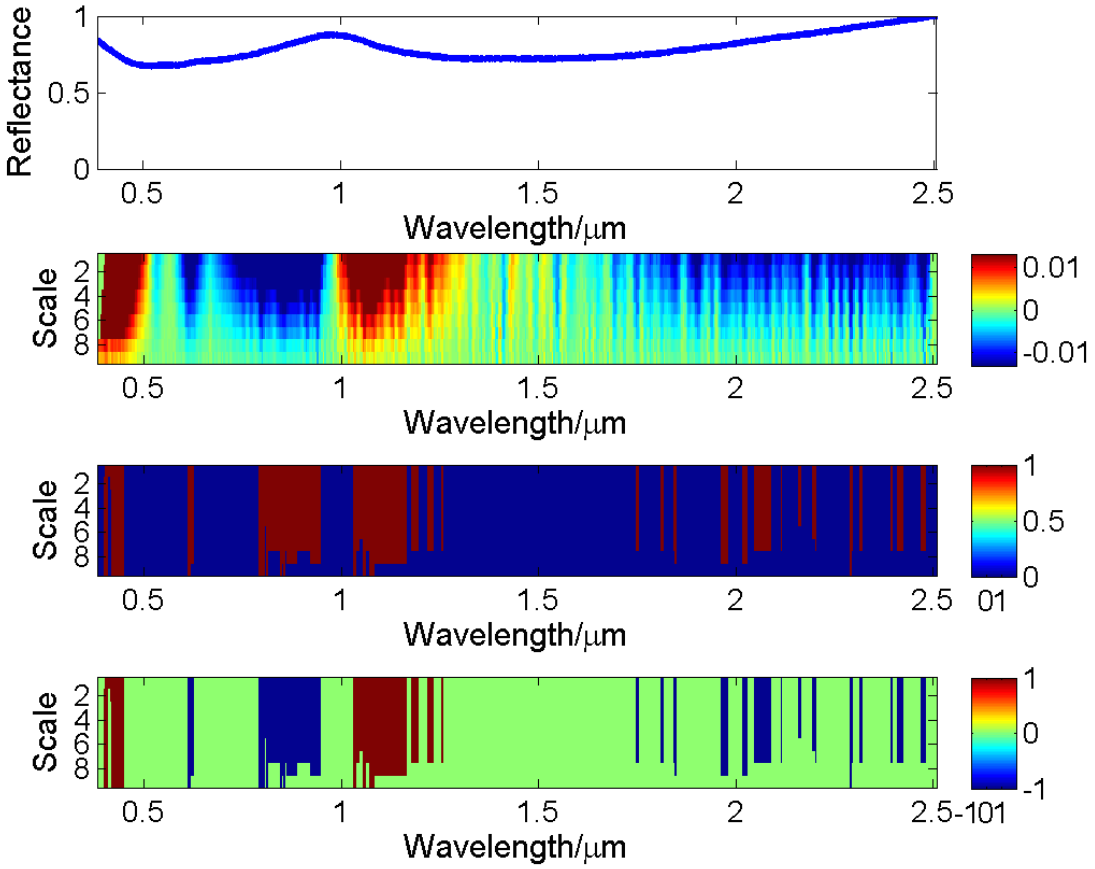

Figure 5.2. Examples of signed state labels as classification features. Top: Example normalized reflectance spectrum (Ilmenite). Second: Corresponding 9-label UWT coefficient matrix using a Haar wavelet. Third: Corresponding state label matrix from an NHMC model using a zero-mean two-state GMM. Blue represents smooth regions, while red represents fluctuations. Bottom: Corresponding features consisting of the state labels with added signs from the Haar wavelet coefficients. Green represents smooth regions, while red represents decreasing fluctuations and blue represents increasing fluctuations.

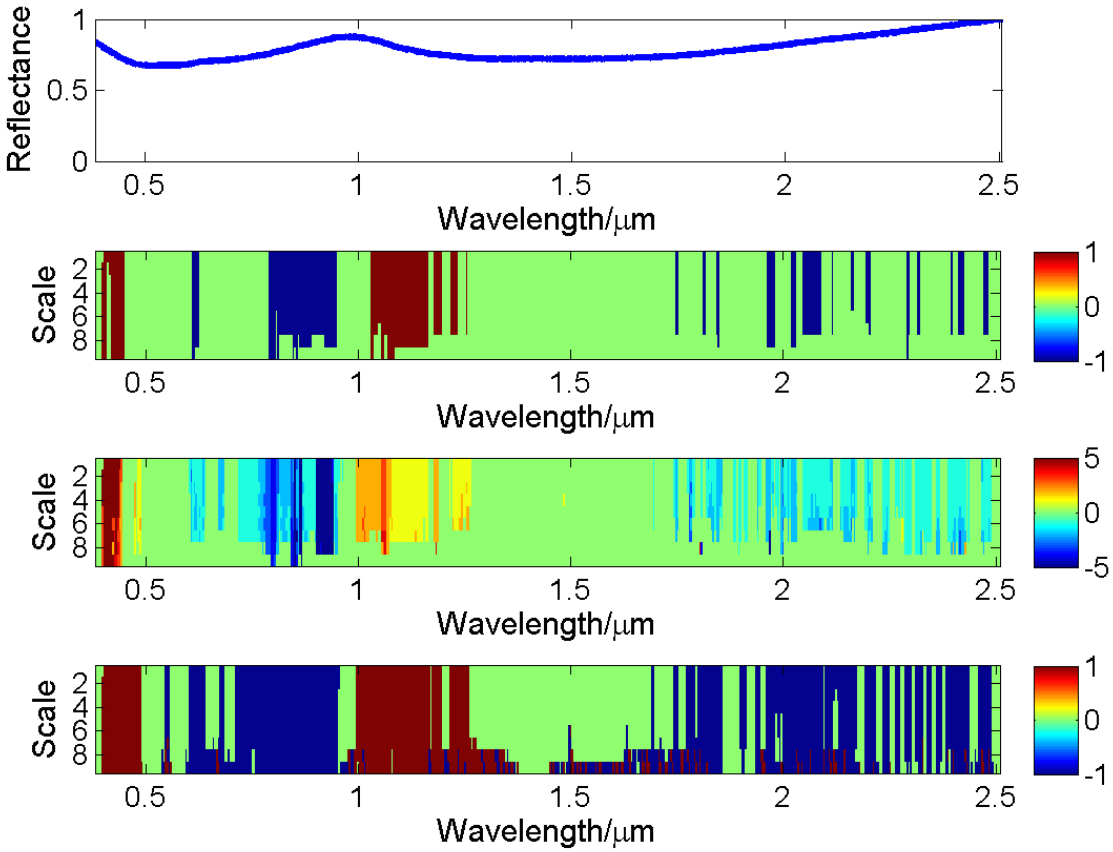

Figure 5.3. Comparison of classification features obtained from several statistical models for wavelet coefficients. Top: example normalized reflectance, same as in Fig. 5.2. Second: Corresponding state label matrix from a 2-state GMM NHMC model. Third: Corresponding state label matrix from a 6-state GMM NHMC model. Bottom: Corresponding state label matrix from a MOG NHMC with $k = 6$.

CHAPTER 6

SECOND STAGE EXPERIMENT

6.1 Study Data and Performance Evaluation

The dataset used in this thesis is a part of the RELAB spectral database with 26 mineral reflectance spectrum classes. Since the spectra in the original database have different wavelength ranges, we only use the spectral region from $350nm$ to $2600nm$ (if applicable) which contains almost all of the visible and near-infrared region of the electromagnetic spectrum. We also resample the spectra to $5nm$ to eliminate the differences in spectral resolution in different spectra. However, the number of samples in each mineral category is different. In order to ensure the same prior probability of each category in the training process. Otherwise different mineral types will have different contributions to the model obtained and it may have influence on the final classification accuracy, we use the Hapke mixing model [8] to generate enough mixtures of existing spectra in a given class until the number of samples in each class is the same. The final dataset contains 1690 reflectance spectra with each category including 65 reflectance spectra. Additionally, in order to eliminate the influences caused by illumination conditions [5], we perform normalization on the whole database by dividing each reflectance spectrum by its maximum value.

We compare different NHMC models (GMM/MOG, different number of mixed Gaussian components, with/without assigning Daubechies-1 wavelet coefficient signs to state labels). We first randomly separate the dataset into a training library (including 1352 samples with each category containing 52 reflectrance spectra) and a validation set (including 338 samples with each category containing 13 reflectance

spectra). In order to evaluate the performance of different NHMC-based features, we implement a 5-fold *cross validation* (CV) on the training library. We use four folds at a time to generate the parameters of the NHMC model. Then we use the Viterbi algorithm to obtain the corresponding state labels for both the folds for training and the one for testing and use several classifiers (*nearest neighbor* (NN) classifier, *support vector machine* (SVM) classifier) on the testing fold to evaluate which of the model achieves the highest CV classification accuracy. Finally, we select the best performing model (in terms of the CV performance) and train it on the entire training library. We then perform classification on the validation set to evaluate the models generalization performance.

Unfortunately, the resulting dataset features a significant separation between the different classes, and so the performance of all proposed methods is equal and very high. We tried to introduce some mixing into the database with the attempt to enlarge the differences between reflectance spectra in the same class. In this thesis, our solution is to imitate the image blurring process using Gaussian smoothing operator. We first randomly set the order of reflectance spectra in the database to get a new database in the form of a 2-D matrix with the first dimension being wavelengths and the second being spectra. Then we generate a hyperspectral volumn by reshaping the obtained 2-D reflectance spectra matrix. The first dimension is still wavelengths while the other two are spectra. This amounts to mimicking a hyperspectral image. In this volumn, each pixel corresponds to a spectrum from a library. And each intersecting surface contains the reflectance value of spectra at a certain wavelength. We perform a spatial blurring on each wavelength using the same Gaussian smoothing operator. Finally, we build a new library from the blurred pixels. The Gaussing smoothing operater we employ has the size of 3-by-3 and Gaussian variance of 0.6. Note that here we use the Gaussian smoothing operator with a constant variance. We adjust the mixing level, which refers to the degree to which a pure reflectance spectrum in

the database will be mixed with other spectra, by randomly interpolating different numbers of vectors with constant values (0 or 1) into the hyperspectral volumn. By doing that, every spectrum has a certain probability to be mixed with constant vectors which changes all its reflectance values to the same degree and thus has no influence on the overall shape of it.

For this study, classification performance is evaluated by using NN and SVM classification accuracies. For NN classifier, three distance metrics are employed: ℓ_1 distance, Euclidean distance, and cosine similarity. For SVM classifier, we use *radial basis function* (RBF) as the kernel and perform a grid search for the corresponding parameter values that provided best performance in each NHMC model. Figure 6.1 and Figure **??** exhibit the CV classification rates for different NHMC models using NN classifier and SVM classifier, respectively. It is easy to find that the model configuration of MOG in conjunction with Daubechies-1 wavelet coefficient signs being added to corresponding state labels could consistently provide us with a better performance than all three other model configurations. In terms of classification performance stability, it seems that the two model configurations using MOG are the two extremes: by adding wavelet coefficient signs we could get the most stable classification performance while without the signs we could get the opposite thing. As I mentioned earlier, MOG combines the simplicity of a binary-state GMM and the spectral fluctuation characterization capability of a multi-state GMM. However, if we do not consider the wavelet coefficient, the fact that state labels only reflect spectral fluctuation magnitudes could cause several similar state label matrices which actually correspond to reflectance spectra with different shape. The reason is that the use of a binary-state GMM form could make several fluctuations of different level have the same state labels and the possibility of two different spectra have similar state label matrices could increase. However, if we add the Daubechies-1 wavelet coefficient signs to the corresponding state labels, the state labels could reflect more

Table 6.1. Details on Best-performing Features from Cross Validation Experiments with Several Classifiers

Classifier	NN			SVM
Similarity Metric	ℓ_1 Distance	Euclidean Distance	Cosine Similarity	
CV Accuracy	0.983	0.987	0.983	0.990
GMM or MOG	MOG	MOG	MOG	MOG
Gaussian State #	3	4	4 and 5	5
Sign Added	Yes	Yes	Yes	Yes

information (the spectral fluctuation orientation information) in the corresponding reflectance spectrum. In this case, the appearance of similar state label matrices with actually different spectra could decrease. Furthermore, in this case the form of a binary-state GMM could help maintain the stability of classification performance because no matter how many Gaussian components are mixed, the state labels always have only three possible values (0,1,-1). Thus the state label matrix of a certain reflectance spectrum will be similar by using different numbers of mixed Gaussian components. Table 6.1 lists the models achieving highest CV classification rates for each classifiers.

Finally, we select those best performing models and train it on the entire training library. We then use NN (via the aforementioned three similarity metrics) and SVM classification on the validation set to evaluate the models' generalization performance. The final test results are listed in Table 6.2 and the classification rates are corresponding to the NHMC models of the same order as those in Table 6.1. Note that in Table 6.1 two models achieve the same highest classification accuracy value in the case of cosine similarity. However, in the final test, the model with 4 Gaussian components performs better than the one with 5 Gaussian components. Thus, Table 6.2 only contains the classification rate of the former model under cosine similarity.

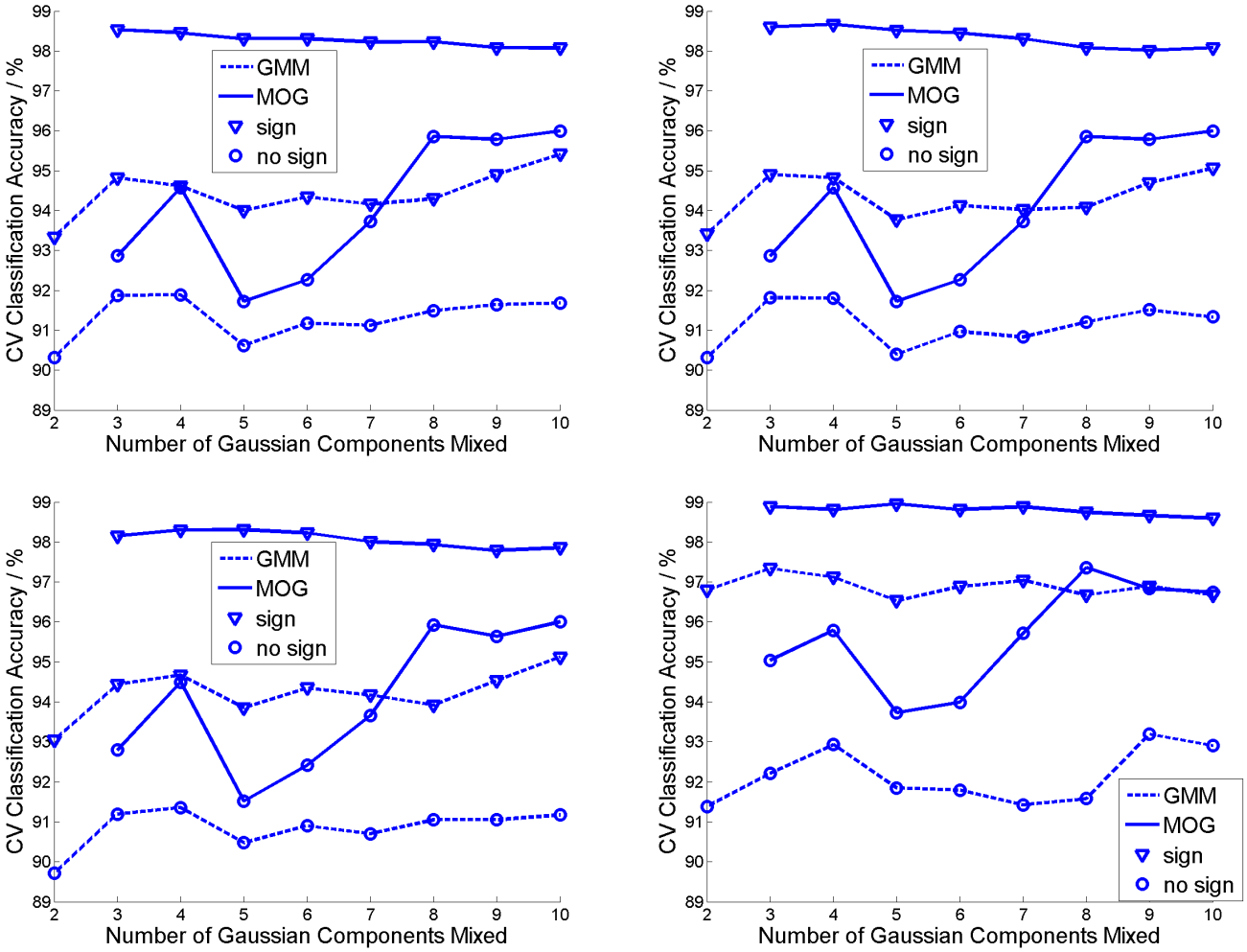

Figure 6.1. CV classification rates of different NHMC modeling approaches. Top left: NN classifiers with ℓ_1 distance; top right: NN classifier with Euclidean distance; bottom left: NN classifier with cosine similarity; bottom right: SVM classifier.

Table 6.2. Final Classification Rates on Models with Best CV Classification Accuracy by Using Nearest Neighbor Classifier and Support Vector Machine Classifier

Classifier	NN			SVM
Similarity Metric	ℓ_1 Distance	Euclidean Distance	Cosine Similarity	
Accuracy	0.967	0.967	0.959	0.967

6.2 Results for Comparison Approaches

In our final test, we compare the generalization performance of the best NHMC models, which are listed in Table 6.1, with the four hyperspectral classification ap-

Table 6.3. Experimental Results of Comparison Approach which Performs Classification on the Whole Spectrum

Classifier	NN				SVM
Similarity Metric	SAM	ED	SCM	SID	
Accuracy	0.959	0.950	0.953	0.959	0.964

Table 6.4. Experimental Results of Approach Introduced in [18]

Classifier	NN			SVM
Similarity Metric	ℓ_1 Distance	Euclidean Distance	Cosine Similarity	
Accuracy	0.964	0.956	0.941	0.967

proaches mentioned in Section II. For all the comparison experiments we use the same dataset as in the experiments above and separate it into training library and validation set with the same scheme.

First, we compare our method with the four spectral similarity measures: spectral angle measure (SAM), Euclidean distance measure (ED), spectral correlation measure (SCM) and the spectral information divergence (SID). We evaluate the method by finding the nearest neighbor in the training library for each testing spectrum. To have an all-around comparison with our approach, we also perform SVM classification on spectra themselves in the database. The classification rates of the four spectral similarity measures are listed in Table 6.3.

We also tried the approach introduced in [18]. In that paper the author only employs SAM, which is the same as cosine distance, as the similarity metric. Here we expand on their experiment by using the aforementioned three similarities performed on our NHMC models when using NN classifier: ℓ_1 distance, euclidean distance and cosine similarity. Still, in order to conduct an overall comparison with our NHMC model, we perform SVM classification on the same signature proposed in [18]. The results are shown in Table 6.4.

CHAPTER 7

THIRD STAGE EXPERIMENT

In last chapter, we perform our experiment with a Gaussian smoothing operator with a fixed variance. And we use constant vectors to control the mixing level. However, by doing that we can not get an quantitative relationship between the classification of a certain model configuration and the mixing level. In this stage of experiment, we control the mixing level by adjusting the Gaussian variance of Gaussian smoothing operator. By doing that, the dominant material percentage (the percentage of an original pure spectrum in the mixed spectrum) will be equivalent to the weight of the element in the central place of the Gaussian smoothing operator. The example below shows the mask of a Gaussian smoothing operator with the Gaussian variance of 0.7. If we using this mask to perform image blurring, then the dominant material percentage in any mixed spectrum should be 33.77%.

$$\begin{pmatrix} 0.0439 & 0.1027 & 0.0439 \\ 0.1217 & 0.3377 & 0.1217 \\ 0.0439 & 0.1217 & 0.0439 \end{pmatrix}$$

In this experiment, we will control the pure spectrum percentage from 50% to 100% with the step of 5%. And for the convenience of comparison, here we use a competitor which performs classification on the wavelet coefficients rather than the one performs classification on the whole spectra. Figure 7.1 exhibits the classification rates for different NHMC models using NN classifier. Here we directly perform finally classification on the test set without any model selection process. So all the classification

rates in the figures below are final classification rates. Under all the three similarity measures, in any certain dominant material percentage value our models seem to have similar results. However, cosine similarity seems to have influences on the two competitors which decreases the classification rates. Under another two spectral similarity measures, we can find that higher mixing level could have larger negative influence on our models than the two competitors. In the case of using ℓ_1 distance, our models start to catch up with the two competitors from the dominant material percentage of 70%. And under Euclidean distance it is 80%. And when the dominant material percentage is higher than 90%, all the six approaches in the figures could achieve excellent classification performance. So the results in this dominant material percentage period is not informative to us. These results show that our model, especially the case of using MOG with sign being added to corresponding state labels, could outperform the two competitors with relatively pure spectra. However, we could not ignore the fact that mixing could have larger influence on our models. The reason behind this could lie in the classification features we use. In this study, we quantize the spectral fluctuation magnitudes into several different integers depending on the number of Gaussian components we use. In some cases, such a quantization could enlarge the difference between spectra. Because of the randomness of mixing, the differences between spectra after mixing are larger than those before mixing. Thus, the disadvantages of our models will be amplified. Now let's look at the two competitors. Wavelet coefficients could be regarded as variables which could almost continuously reflect the change in spectral fluctuations. And the approach proposed in [18] removes the continuum in a spectrum which could even decrease the differences between spectra. So the impact from mixing on these two approaches will not be so large as on our models.

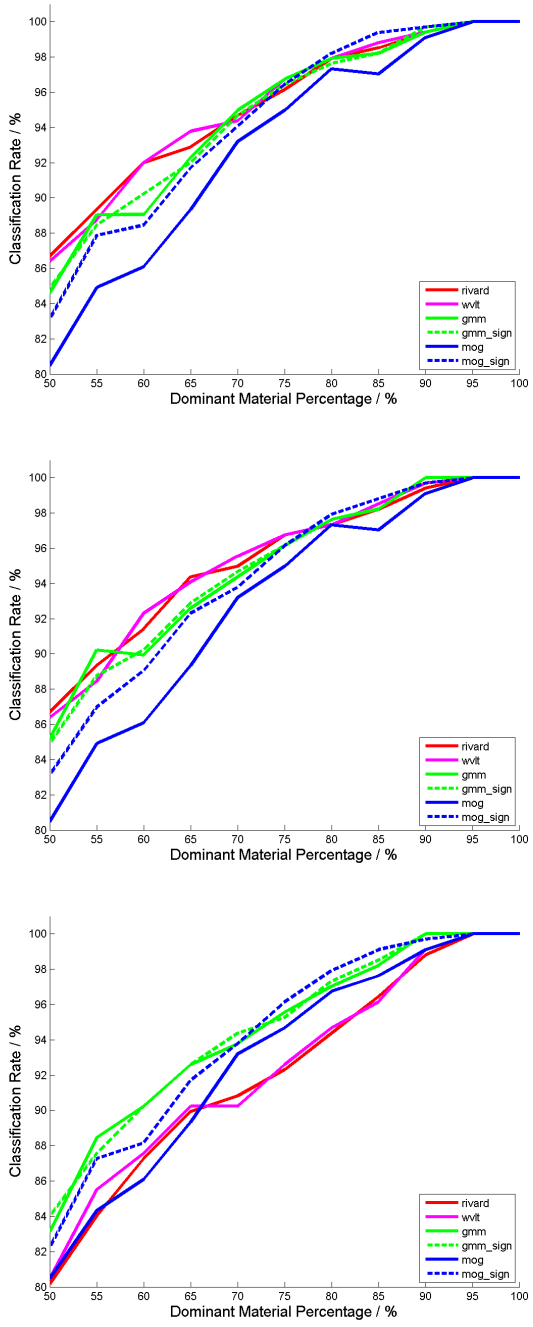

Figure 7.1. Classification rates of different NHMC modeling approaches as functions of diminant material percentage using NN classifier with ℓ_1 distance, Euclidean distance and cosine similarity (from top to bottom), respectively. Here "gmm" means Gaussian mixture model (GMM), "MOG" means mixture of Gaussian (MOG). The suffix "sign" means the case of adding Daubechies-1 wavelet coefficient signs to the corresponding state labels. "rivard" means the approach proposed in [18]. "wvlt" corresponds to the approach which performs classification on wavelet coefficients.

CHAPTER 8

CONCLUSION

We developed a wavelet-based non-homogeneous Markov chain model for hyperspectral signature classification. We used an undecimated wavelet transform to get multiresolution information on the shape of hyperspectral signatures. The Haar wavelet function was chosen for its ability of detecting both slow-varying fluctuations and rapid changes in spectral curves, as well as characterizing slope orientations via wavelet coefficient signs. An NHMC model was trained at each wavelength via EM algorithm and then a set of NHMC parameters were generated. The Viterbi algorithm was employed to transform wavelet coefficient into discrete state label matrices which captured the semantic information of corresponding hyperspectral signatures. Finally, the hyperspectral classification system was tested by using different classifiers and spectral similarity measures. Two modifications were also proposed. The modifacation of the addition of wavelet coefficient signs to the corresponding state labels enabled the state label matrices to contain slope orientation information, while the combination of several GMM states into two reduced the complexity of state label matrices and increased classification accuracies on the whole. In general, spectral matching using cosine similarity as distance metric performed the best.

We also compared our method with two other classification approaches. One performed spectral matching directly on spectra. One performed spectral matching on the sum of wavelet coefficients of certain scales. Furthermore, the small size of training data also limited the performance of the two DWT-based approaches. Thus we recommend further work that expands the size of database and investigate

more mother wavelets for both UWT and DWT-based hyperspectral classification researches. Additionally, NHMC based on nonzero-mean GMM might be advantageous in the future because in some cases the histogram of wavelet coefficients cannot be accurately modeled by zero-mean Gaussian mixture models.

APPENDIX

PROOF OF LEMMA

The relationship between the original state labels S_{s-1}, S_s and the combined state labels Z_{s-1}, Z_s can be characterized by a directed graphical model which is shown in Fig. A.1. By considering all possible transitions from Z_{s-1} to Z_s through the state transitions S_{s-1} to S_s and the map above, and denoting

$$p_{a\to b}^{s,S\to Z} = p(Z_s = b|S_s = a),$$
$$p_{a\to b}^{s,Z\to S} = p(S_s = b|Z_s = a),$$

we can write

$$q_{b\to a,s} = \sum_{x=0}^{k-1}\sum_{y=0}^{k-1} p_{x\to a}^{s,S\to Z} p_{y\to x,s} p_{b\to y}^{s-1,Z\to S}. \tag{A.1}$$

From the $Z(S)$ map in equation (5.1), we can infer that

$$p_{0\to 0}^{s,S\to Z} = 1,$$
$$p_{0\to 1}^{s,S\to Z} = 0,$$
$$p_{i\to 0}^{s,S\to Z} = 0,$$
$$p_{i\to 1}^{s,S\to Z} = 1,$$
$$p_{0\to 0}^{s-1,Z\to S} = 1,$$
$$p_{0\to i}^{s-1,Z\to S} = 0,$$
$$p_{1\to 0}^{s-1,Z\to S} = 0,$$
$$p_{1\to i}^{s-1,Z\to S} = \frac{p(S_{s-1} = i)}{\sum_{j=1}^{k-1} p(S_{s-1} = j)},$$

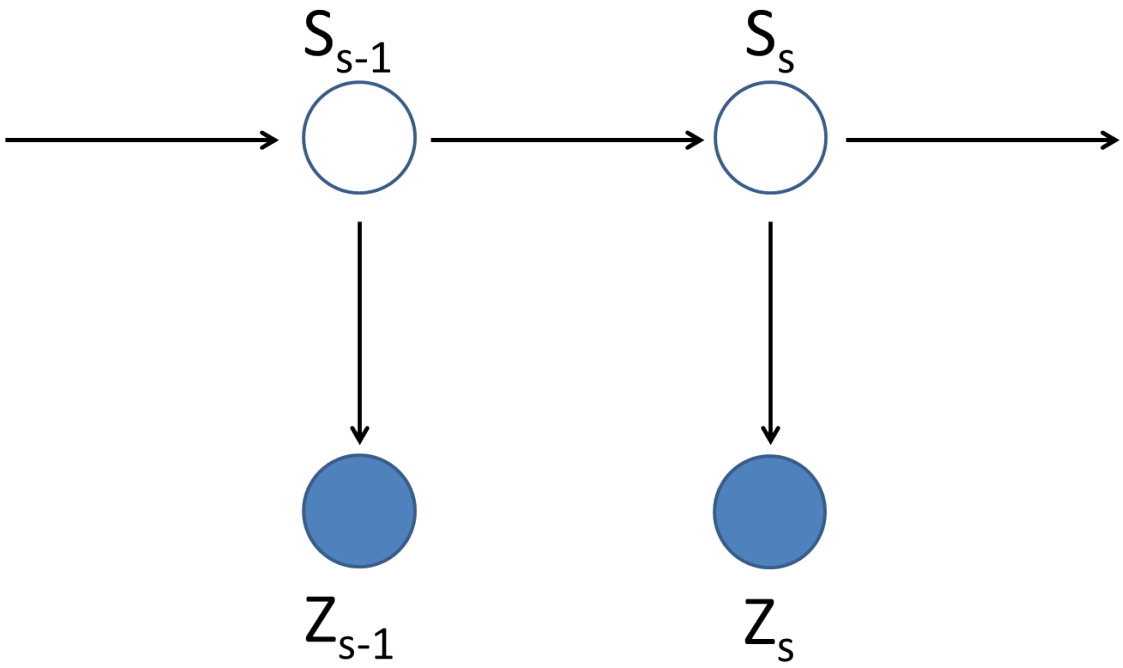

Figure A.1. Directed graphic model for state labels

where $i = 1, ..., k-1$. After combining the 8 equalities above with $(A.1)$ for $a, b \in \{0, 1\}$, we can get the four elements in new matrices expressed in $(5.2 - 5.5)$, proving the lemma.

BIBLIOGRAPHY

[1] Camps-Valls, G., Tuia, D., Bruzzone, L., and Atli Benediktsson, J. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE Signal Proc. Mag. 31*, 1 (Jan. 2014), 45–54.

[2] Chang, C. I. An information-theoretic approach to spectral variability, similarity, and discrmination for hyperspectral image analysis. *IEEE Trans. Info. Theory 46*, 5 (May 2000), 1927–1932.

[3] Clark, R., Swayze, G. A., Livo, K., Sutley, S., Dalton, J., Mc-Dougal, R., and Gent, C. Imaging spectroscopy: Search and planetary remote sensing with the USGS Tetracorder and expert systems. *J. Geophysical Research 108*, E12 (Dec. 2003).

[4] Crouse, M. S., Nowak, R. D., and Baraniuk, R. G. Wavelet-based statistical signal processing using hidden Markov models. *IEEE Trans. Signal Proc. 46*, 4 (Apr. 1998), 886–902.

[5] Dennison, P., Halligan, K., and Roberts, D. Application of the discrete wavelet transform to the monitoring of tool failure in end milling using the spindle motor current. *Remote Sensing of Environment 93* (Nov. 2004), 359–367.

[6] Donoho, D. L. De-noising by soft-thresholding. *IEEE Trans. Info. Theory 41* (May 1995), 613–627.

[7] Duarte, M. F., and Parente, M. Non-homogeneous hidden Markov chain models for wavelet-based hyperspectral image. In *Allerton Conf. Communication, Control, and Computing* (Monticello, IL, Oct. 2013), pp. 154–159.

[8] Hapke, B. *Theory of reflectance and emittance spectroscopy.* Cambridge University Press, 2012.

[9] Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., and Goetz, F. H. The Spectral Image Processing System (SIPS) – Interactive visulization and analysis of imaging spectrometer data. *Remote Sensing of Environment 44*, 2–3 (May 1993), 145–163.

[10] Mallat, S. *A wavelet tour of signal processing.* Academic Press, San Diego, CA, 1999.

[11] Mallat, S., and Hwang, W. Singularity detection and processing with wavelets. *IEEE Trans. Info. Theory 38*, 2 (Mar. 1992), 617–643.

[12] Mallat, S., and Zhong, S. Characterization of signals from multiscale edges. *IEEE Trans. Pattern Analysis and Machine Intelligence 14*, 7 (Jul. 1992), 710–732.

[13] Masood, K. Hyperspectral imaging with wavelet transform for classification of colon tissue biopsy samples. In *Applications of Digital Image Proc. XXXI* (San Diego, CA, Aug. 2008), vol. 7073 of *Proc. SPIE*.

[14] Orchard, M. T., and Ramchandran, K. An investigation of wavelet-based image coding using an engropy-constrained quantization framework. In *Data Compression Conf. (DCC)* (Snowbird, UT, Mar. 1994), pp. 341–350.

[15] Parente, M., and Duarte, M. F. A new semantic wavelet-based spectral representation. In *IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)* (Gainesville, FL, June 2013).

[16] Parente, M., Makarewicz, H. D., and Bishop, J. L. Decomposition of mineral absorption bands using nonlinear least squares curve fittening: application to martian meteorites and CRISM data. *Planetary and Space Science 59*, 5–6 (2011), 423–442.

[17] Rabiner, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE 77*, 2 (Feb. 1989), 257–285.

[18] Rivard, B., Feng, J., Gallie, A., and Sanchez-Azofeifa, A. Continuous wavelets for the improved use of spectral libraries and hyperspectral data. *Remote Sensing of Environment 112* (2008), 2850–2862.

[19] Sweet, J. N. The spectral similarity scale and its application to the classification of hyperspectral remote sensing data. *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data* (1993), 92–99.

[20] van der Meer, F., and Bakker, W. CCSM: Cross correlogram spectral matching. *Int. J. Remote Sensing 18*, 3 (1997), 1197–1201.

[21] West, T., Prasad, S., and Bruce, L. M. Multiclassifiers and decision fusion in the wavelet domain for exploitation of hyperspectral data. *IEEE Int. Geoscience and Remote Sensing Symposium (IGARSS)* (2007), 4850–4853.

[22] Zhang, X., Younan, N. H., and Hara, C. G. O'. Wavelet domain statistical hyperspectral soil texture classification. *IEEE Trans. Geoscience and Remote Sensing 43*, 3 (Mar. 2005), 615–618.