

TAILORING NON-HOMOGENEOUS MARKOV CHAIN WAVELET MODELS FOR HYPERSPECTRAL SIGNATURE CLASSIFICATION

Siwei Feng Yuki Itoh Mario Parente Marco F. Duarte

Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003

ABSTRACT

We consider the application of *non-homogeneous hidden Markov chain* (NHMC) models to the problem of hyperspectral signature classification. It has been previously shown that the NHMC model enables the detection of several semantic structural features of hyperspectral signatures. However, there are some aspects of the spectral data that are not fully captured by the proposed NHMC models such as the relatively smooth but fluctuating regions and the fluctuation orientations. In order to address these limitations, we propose an improved NHMC model based on Daubechies-1 wavelets in conjunction with an increased the model complexity. Experimental results show that the revised approach outperforms existing approaches relevant in classification tasks.

Index Terms— Classification, Hyperspectral Signal Processing, Wavelet, Hidden Markov Model

1. INTRODUCTION

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 library of spectra obtained in the laboratory from well characterized samples. Practitioners recognize several semantic 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 [1] have been proposed to encode the aforementioned semantic information. However, such techniques rely on ad-hoc rules to characterize instances of each material and require the creation of new rules for additional spectral species which were not previously analyzed. While automatic techniques for spectral matching (overviewed in [2]) aim to discriminate spectral shapes without hard-coded rules, they do not encode the aforementioned semantic information.

Parente and Duarte [3,4] have previously proposed a *non-homogeneous hidden Markov chain* (NHMC) model that implements semantic information extraction from hyperspectral signals by encoding the wavelet-domain representations of the spectra into binary features using a two-state zero-mean

Gaussian mixture model (GMM). The model enables the extraction of semantic information without relying on ad-hoc rules. The interpretation of unknown spectra is based on training data obtained automatically from the library. The NHMC model encodes the structural information in the wavelet decomposition of all library spectra into a set of discriminative features. Experimental results show that the model performs well in characterizing diagnostic features in mineral reflectance spectra like the positions/widths of their absorption bands.

However, the definitive success and widespread diffusion of this technique is determined by the degree to which we can characterize all relevant structural features of the spectra observed. Our focus in this paper is to review cases in which certain structural features are not properly characterized by the current NHMC model, which is seen to negatively affect the performance of spectral classification from NHMC-derived features. In the following sections, we show that binary features generated by the previously proposed model exhibit limited sensitivity to slow-varying fluctuations in reflectance spectra. We therefore propose the use of lower-order wavelet functions that introduce the desired sensitivity of the model to such fluctuations. Additionally, we increase the complexity of the signal model in order to improve the detectability of such subtler structural information of reflectance spectra.

2. BACKGROUND

The capability of wavelet coefficients to characterize signal discontinuities at different scales and positions makes wavelet transforms a popular tool in many signal processing applications. Parente and Duarte [3,4] exploited multi-resolution wavelet decompositions to detect the presence of diagnostic absorptions in reflectance spectra. We provide a summary of the approach below and point the reader to [3,4] for additional detail.

More specifically, we use the *undecimated wavelet transform* (UWT) of an N -sample signal $x \in \mathbb{R}^N$, which is composed of wavelet coefficients $w_{s,n}$. Each coefficient is labeled by a scale $s \in \{1, \dots, S\}$ and offset $n \in \{1, \dots, N\}$, where $S \leq N$, and is defined using inner products as $w_{s,n} = \langle x, \phi_{s,n} \rangle$, where $\phi_{s,n}$ denotes the mother wavelet function ϕ dilated to scale s and translated to offset n : $\phi_{s,j(n)} =$

This work was supported by the National Science Foundation under grant number IIS-1319585.

$\phi((n-j)/s)/\sqrt{s}$. Low values of s correspond to fine scales, while large values of s correspond to coarse scales.

Until now, several hyperspectral classification methods based on wavelet transforms have been proposed. However, most classification approaches do not exploit the statistics of the wavelet coefficients observed [5–7]. Crouse *et al.* [8] proposed the application of *hidden Markov trees* (HMTs) to the construction of statistical models of dyadic wavelet coefficients. The HMT uses two-state mixture of gaussian models for wavelet coefficients to provide a binary coefficient labeling scheme, grouping them into large and small-magnitude classes. Such models are used for hyperspectral classification in [9], which uses HMTs trained on individual classes to evaluate the likelihood of the spectrum observed under each class. The use of multiple HMT models can significantly increase the computational complexity of the training, as well as the potential for overfitting.

Inspired by HMTs, Parente and Duarte [3,4] proposed a *non-homogeneous hidden Markov chain* (NHMC) to model the UWT coefficients of reflectance spectra. In contrast to HMTs, the UWT provides a separate NHMC for each offset, since each wavelet coefficient has only one parent and one child (if they exist). A NHMC is trained on each of the N wavelengths sampled by the hyperspectral sensor separately. Training is performed via an *expectation maximization* (EM) algorithm that maximizes the likelihood of a library of training data given the model parameters. The set of NHMC parameters θ_n include the probabilities for the first hidden states, the state transition matrices, and the variances of the Gaussians in each GMM; these parameters are distinct for each wavelet coefficient.

Given the model, the observation is translated into a set of state labels $\{\mathcal{S}_{s,n}\}$ that encode the mixture used in each two-state GMM, obtained via a Viterbi algorithm [8, 10] that employs the Gaussian parameters and transition probabilities in θ_n . Such labels are used as features for classification via a basic nearest neighbor approach.

3. PROPOSED MODIFICATIONS

The NHMC model proposed in [3,4] is based on the Daubechies-4 wavelet. The characteristics of this wavelet make it very attractive in detecting higher order discontinuities at different scales. On the other hand, slow-varying fluctuations and slopes are generally overlooked. Consider the example in Figure 1, where the top row represents two spectra exhibiting gently sloping features, while the middle row shows the undecimated Daubechies-4 wavelet coefficient matrix of such signals. The Daubechies-4 wavelet transform of such signals is sensitive to higher order fluctuations (e.g., the rapid change of slope around $0.9\mu\text{m}$ and the small discontinuities after $2\mu\text{m}$ in the second spectrum); however, this wavelet does not express the differences between the spectral slopes in the ranges $0.5-0.8\mu\text{m}$ and $1-1.5\mu\text{m}$, as shown in the middle row of Figure 1. Note that the most significant differences

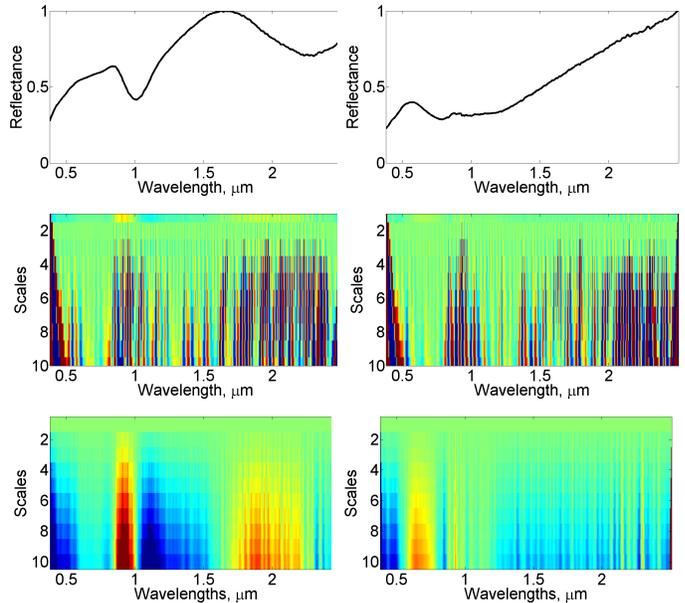


Fig. 1. Top row: Normalized reflectance spectra of two material samples from the Augite mineral family. Middle row: Corresponding UWT coefficient arrays ($S = 10$) using a Daubechies-4 wavelet. Bottom row: Corresponding UWT coefficient arrays ($S = 10$) using a Daubechies-1 wavelet. In both cases, small/large coefficient values are shown in blue/red.

in the spectra are not reflected in the Daubechies-4 wavelet coefficients. In other words, the Daubechies-4 wavelet cannot fully characterize the structural information of these two reflectance spectra. Nonetheless, we can find that in the two Daubechies-1 wavelet coefficient arrays, the differences between the spectra are captured. The Daubechies-1 wavelet coefficients encode the drastic fluctuation around $0.9\mu\text{m}$ as well as the small discontinuities after $2\mu\text{m}$ in the second spectrum. Additionally, the relatively smooth fluctuations like those between $1-1.5\mu\text{m}$ in the first spectrum and those between $0.5-0.8\mu\text{m}$ in the second spectrum are captured by the wavelet coefficient arrays.

The Daubechies-1 wavelet is the simplest possible compact wavelet with the properties of square-like shape and discontinuity [11]. These two properties enable the Daubechies-1 wavelet to detect both slow-varying fluctuations and sudden changes in a signal [12]. The bottom row of Figure 1 shows that, in contrast from Daubechies-4, the Daubechies-1 wavelet transform captures the gently sloping fluctuations in the spectra. It is easy to show that the magnitude of the Daubechies-1 wavelet coefficients is proportional to the slope of the spectra. Furthermore, the sign of these coefficients capture whether the reflectance increases or decreases as a function of wavelength.

As a second consideration, note that the papers [3,4] use a two-state zero-mean GMM to model the distribution of the observed wavelet coefficients because such a mode, when used with higher-order wavelets, distinguishes between sharp

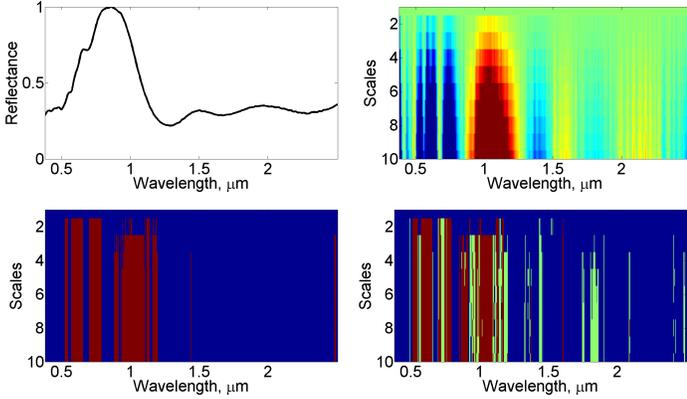


Fig. 2. Top left: example of normalized mineral reflectance spectrum (Almandine). Top right: corresponding UWT coefficient array ($S = 10$) using a Daubechies-1 wavelet. Negative/positive coefficient values are shown in blue/red; darker shades represent larger magnitudes. Bottom left: corresponding label array from a NHMC model using a zero-mean two-state GMM. Red represents fluctuations, while blue represents smooth regions. Bottom right: corresponding label array from a NHMC model using a zero-mean three-state GMM. Red/green represents drastic/smooth fluctuation, while blue represents smooth regions.

absorption bands and flat regions in a spectrum by assigning them large and small state labels, respectively. A straightforward improvement of the model consists of increasing the number of state labels (i.e. three or more) to represent intermediate level fluctuations, which were found to have influence on classification accuracy. This modification is particularly important for Daubechies-1, which is sensitive to a larger range of fluctuation frequencies than Daubechies-4 but with a lower discriminative power.

As a third consideration, in order to eliminate the differences between spectra caused by illumination conditions [13], 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 [14], a technique that applies a threshold on the maximum value of the signal coefficients and that is commonly applied to wavelet representations.

Figure 2 shows an example of the Daubechies-1 wavelet coefficients and labels corresponding to the proposed modifications for an example spectrum of the mineral almandine. The figure shows that the presence of additional states allows for a finer characterization of the slopes of the spectra, including the subtle features present after $1.5\mu\text{m}$.

4. EXPERIMENTAL RESULTS

We expand on the original classification experiments in [3, 4] by increasing the spectral complexity of the classification.

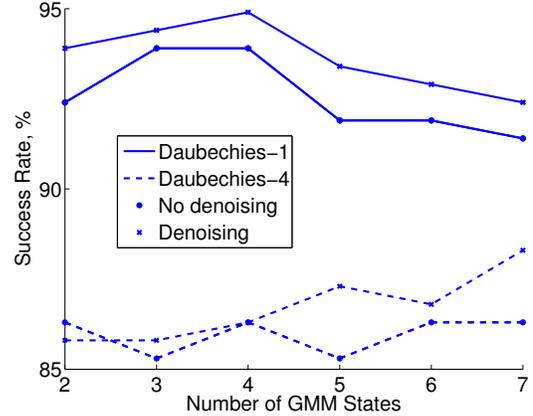


Fig. 3. CV classification rates of different NHMC modeling approaches. Daubechies-1 wavelets with denoising consistently provides the best performance, with 4 GMM states providing best overall success rate.

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 3. 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.

Figure 4 shows example classification features for spectral signatures using the Daubechies-4 and Daubechies-1 wavelets in the case of GMM with 3 states. The top row

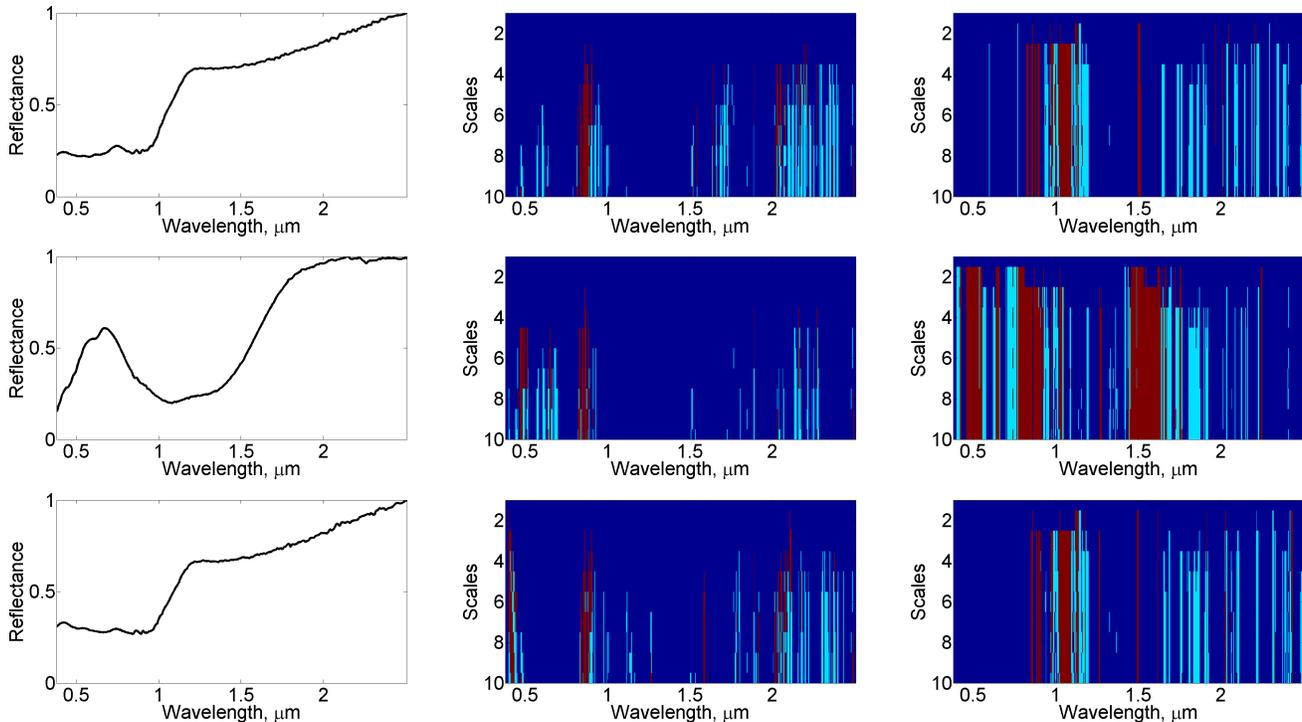


Fig. 4. Comparison of label-based classification using NHMC with three states and Daubechies-1 and Daubechies-4 wavelets. Top row: Spectrum being tested. Middle row: Nearest neighbor of top row spectrum in the whole training set in Hamming distance. Bottom row: Nearest neighbor of top row spectrum in the same class in Hamming distance. Left column: Normalized spectra. Middle column: Corresponding state label arrays to the normalized spectra using Daubechies-4 wavelet. Right column: Corresponding state label arrays to the normalized spectra using Daubechies-1 wavelet.

shows an example spectrum, the middle row shows the closest spectrum in state label space with Daubechies-4 wavelets, and the bottom row shows the closest spectrum in state label space with Daubechies-1 wavelets. From the state label arrays, we can see that as expected the Daubechies-4 wavelet is sensitive to compact and drastic discontinuities, while showing little response to slow-varying fluctuations. For example, the Daubechies-4 wavelet coefficients are quite sensitive to the small but drastic fluctuation near $1\mu\text{m}$ in all the 3 spectra. However, note that all three spectra receive similar labels in the $0.5 - 2\mu\text{m}$ range; these are important structural features in such kind of material (hematite), and their differences are not discriminated in the corresponding Daubechies-4 state label arrays. Thus, the Daubechies-4 wavelet does not fit this classification task. In contrast, the labels obtained for the Daubechies-1 wavelet coefficients using three states, shown in the third column, are able to discriminate between the structures of the first and third spectra and that of the second spectra by endowing them with different label arrays. Therefore, we deduce that the state labels of Daubechies-1 wavelet successfully characterizes the shape of these spectra.

In our final test, we compare the generalization performance of the best NHMC model (4 states, Daubechies-1 wavelet, with denoising), which achieved a classification accuracy of 91.5%, with other spectrum matching approaches. The first competitor approach is the one introduced in [5]. The

method calculates a 10-level undecimated wavelet decomposition of the database spectra then generates a signature by combining the wavelet coefficients of the 6 finest scales at each wavelength. This amounts to removing the continuum of a spectrum while only considering the most essential structural features of the spectral curve. We evaluate the method by finding the nearest neighbor in the training set for each test spectrum but in this case we use cosine distance as similarity measurement. The classification accuracy is 85.1%. We also evaluate other spectral matching approaches that operate on the spectra themselves. Four similarity measures (reviewed recently in [2]): *spectral angle measure* (SAM), *euclidean distance measure* (ED), *spectral correlation measure* (SCM) and *the spectral information divergence* (SID) are evaluated with the same approach as above. The classification rates are 80.9% for SAM, ED and SCM and 83.0% for SID.

5. CONCLUSIONS AND FUTURE WORK

We improved the previously proposed non-homogeneous hidden Markov chain model by leveraging the Daubechies-1 wavelet, increasing number of Gaussian mixture model states, and denoising the spectra used in training and testing. The model outperformed some classic approaches in the case of material classification tasks. Further work will expand the size of database in order to provide enough number of spectra thus decreasing the randomness in experimental results.

6. REFERENCES

- [1] R. Clark, G. A. Swayze, K. Livo, S. Sutley, J. Dalton, R. McDougal, and C. Gent, "Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems," *J. Geophys. Res.*, vol. 108, no. E12, Dec. 2003.
- [2] F. van der Meer, "The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery," *Int. J. Applied Earth Observation and Geoinformation*, pp. 3–17, Jan. 2006.
- [3] M. Parente and M. F. Duarte, "A new semantic wavelet-based spectral representation," in *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Gainesville, FL, June 2013.
- [4] M. F. Duarte and M. Parente, "Non-homogeneous Hidden Markov Chain models for wavelet-based hyperspectral image processing," in *Allerton Conf. Communication, Control, and Computing*, IL, Oct. 2013.
- [5] B. Rivard, J. Feng, A. Gallie, and A. Sanchez-Azofeifa, "Continuous wavelets for the improved use of spectral libraries and hyperspectral data," *Remote Sensing Environ.*, vol. 112, pp. 2850–2862, 2008.
- [6] H. Okamoto, T. Murata, T. Kataoka, and S. Hata, "Plant classification for weed detection using hyperspectral imaging with wavelet analysis," *Weed Biology and Management*, pp. 31–37, March 2007.
- [7] T. West, S. Prasad, and L. M. Bruce, "Multiclassifiers and decision fusion in the wavelet domain for exploitation of hyperspectral data," in *IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Barcelona, Spain, July 2007, pp. 4850–4853.
- [8] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using Hidden Markov Models," *IEEE Trans. Signal Processing*, vol. 46, no. 4, pp. 886–902, Apr. 1998.
- [9] X. Zhang, N. H. Younan, and C. G. O'Hara, "Wavelet domain statistical hyperspectral soil texture classification," *IEEE Trans. Geoscience and Remote Sensing*, pp. 615–618, March 2005.
- [10] L. R. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–285, Feb. 1989.
- [11] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, 1992.
- [12] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1999.
- [13] P. Dennison, K. Halligan, and D. Roberts, "Application of the discrete wavelet transform to the monitoring of tool failure in end milling using the spindle motor current," *Remote Sensing Environ.*, vol. 93, pp. 359–367, Nov. 2004.
- [14] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Info. Theory*, vol. 41, pp. 613–627, May 1995.