A NEW SEMANTIC WAVELET-BASED SPECTRAL REPRESENTATION

Mario Parente and Marco F. Duarte

University of Massachusetts Amherst, Department of Electrical and Computer Engineering
{mparente, mduarte}@ecs.umass.edu

ABSTRACT

We consider the problem of designing features for the identification of reflectance spectra that capture the semantic information used by experts in ad-hoc labeling rules, such as the shape and position of absorption bands in spectra. We propose the use of statistical models on wavelet coefficients that allow us to quantify the presence of spectrum discontinuities at multiple scales. Such models are now feasible to train due to the availability of large-scale datasets of hyperspectral images. Using a non-homogeneous hidden Markov model, we can succinctly express the semantic information in the spectrum in terms of binary state labels representing its energy content in a multiscale time-frequency analysis. Experimental results show that the features succeed in outperforming existing approaches in encoding and quantifying the semantic features of the spectra that are relevant in classification tasks.

Index Terms— Classification, Spectral Libraries, Feature Selection, Wavelets, Hidden Markov Chains

1. INTRODUCTION

Recent improvements in the spatial and spectral resolution of hyperspectral imaging systems (HSIs) have enabled scientists and engineers to consider complex information extraction applications including spectral identification, hyperspectral image segmentation [1] and spectral unmixing [2]. Unfortunately, such improvements also strain the computational and visualization resources that are available to manage, understand, and decode the information that is present in the acquired hyperspectral data. Practitioners must therefore navigate through massive quantities of very high-dimensional data to identify relevant features that encode the information desired for the application of interest.

We are interested in facilitating the information extraction process via the use of mathematical models for hyperspectral signals, which we can easily adapt to be relevant to the particular problem of interest. We expect to extract features from the model that can capture scientifically meaningful cues and serve as spectral signatures of the materials under study. The aim is to encode the physical information used by scientists to discriminate between the spectra of different minerals into numerical features, which we refer to as *semantic* features.

Experienced researchers characterize reflectance spectra in terms of their shape and the positions/widths of their absorption bands. Such "diagnostic" information is encoded by complicated *ad-hoc* rules. One such approach is the tetracorder [3] which compares continuum-removed signatures from observed pixels to a library of laboratory samples by means of a modified least squares fit. Other approaches use parametric models to represent the absorption features (e.g. [4]) but still rely on rules to match observations to a library. Physically motivated rule-based methods suffer from the need to create new rules when previously unknown spectral species are added to the library.

Mathematical *signal models* have also been proposed to represent reflectance spectra. Some recent approaches based on cubic splines have been successful in spectral unmixing [5]. Models that leverage wavelet decompositions are of particular interest because they allow the representation of diagnostic absorption features at different scales. Current wavelet-based approaches are either *ad-hoc* in nature or limited to filtering techniques that manipulate the data but do not extract information [6, 7].

We propose a model for wavelet-domain representations that is able to encode the same diagnostic information as physically-based models without relying on ad hoc rules, interpreting unknown spectra based on training data obtained automatically from the library. Our approach applies Hidden Markov Models (HMM) [8] to wavelet coefficients from hyperspectral signals. HMMs succinctly capture correlations between wavelet coefficients in overlapping spectral ranges and at adjacent scales. As in [7], we use a continuous wavelet transform to obtain maximum flexibility on the set of scales and offsets considered. The resulting model provides a collection of N Non-Homogeneous Hidden Markov Chains (NHMCs), each corresponding to a particular spectral band. As an example, a set of *energy labels* generated by the model provides a map from each signal spectrum to a binary space that encodes the presence of fluctuations at different scales and wavelengths, effectively representing the semantic cues that allow for discrimination of spectra.

2. CONTINUOUS WAVELET TRANSFORM

The wavelet transform is a widely used tool in signal and image processing that provides a multiscale analysis of a signal's content, effectively encoding in a compact fashion the locations and scales at which the signal structure is present [9].

The *continuous wavelet transform* (CWT) of a spectral signal $x(f)$ supported over $[0, F]$ is composed of *wavelet coefficients* $w_{s,u}$ labeled by a scale $s \in \mathbb{R}^+$ and offset $u \in$

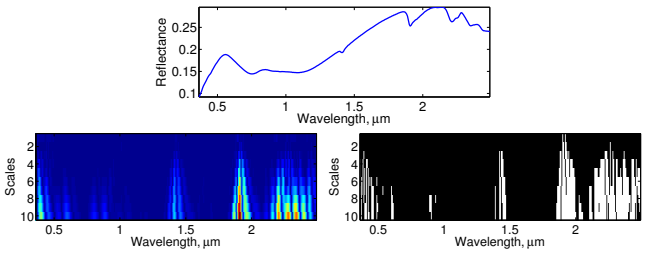

Fig. 1: *Top: Example spectral signature ($N = 325$ samples). Left: Corresponding CWT coefficient array ($s = 9$) using a Daubechies-4 wavelet. Small/large coefficient magnitudes are shown in blue/red. Right: State labels obtained for the example spectral signature using a NHMC model. White corresponds to large labels; black to small.*

$[0, F]$, and defined as $w_{s,u} = \langle x, \psi_{s,u} \rangle = \int x(f)\psi_{s,u}(f)df$, where $\psi_{s,u}$ denotes the mother wavelet ψ dilated to scale s and translated to offset u, i.e., $\psi_{s,u}(f) = \psi\left((f-u)/s\right)/\sqrt{s}$. Each coefficient $w_{s,u}$ at scale s can be described as a discontinuity detector over a portion of the signal of size proportional to s around the offset u.

The wavelet coefficients over a grid of scales $s = 1, \ldots, S$ and offsets $u = 0, F/N, 2F/N, \ldots, F - F/N$ can be organized into a 2-D array $W \in \mathbb{R}^{S \times N}$, where each row represents a different scale and each column represents a different band (cf. Fig. 1). For simplicity, we identify each offset $u = nF/N$ by its index $n = 0, \ldots, N-1$. A large wavelet coefficient (in magnitude) generally indicates the presence of a singularity inside its support; a small wavelet coefficient indicates a smooth region. The wavelet coefficients of a piecewise smooth signal exhibit a peaky non-Gaussian distribution.

A pair of coefficients $w_{s,n}$ and $w_{s+1,n}$ at the same offset and neighboring scales are referred to as a *parent* and *child* coefficient, respectively; the columns of coefficients in Fig. 1 correspond to sequences of parent/child wavelet coefficients. It is easy to see that parent and child wavelet functions are nested inside one another. Due to such nesting, edges and discontinuities in general manifest themselves in the wavelet domain as chains of large coefficients propagating across scales, a phenomenon known as *persistence*. Thus, wavelets both encode and exhibit structure from piecewise smooth signals, a property that is exploited repeatedly in signal processing.

In contrast to the commonplace dyadic wavelet transform, the CWT provides maximum flexibility on the choice of scales and offsets used in the multiscale analysis, as both parameters can take arbitrary real values. The tradeoff for such flexibility is the redundancy that arises in the wavelet representation: the size of the wavelet representation is S times larger than the length of the sampled signal.

3. NON-HOMOGENEOUS HIDDEN MARKOV CHAINS

Inspired by the use of Hidden Markov Trees (HMTs) for statistical modeling of dyadic wavelet coefficients [10], we rely on an NHMC to model the CWT coefficients. In contrast to HMTs, the choice of CWT yields a collection of non-homogeneous hidden Markov chains connecting each

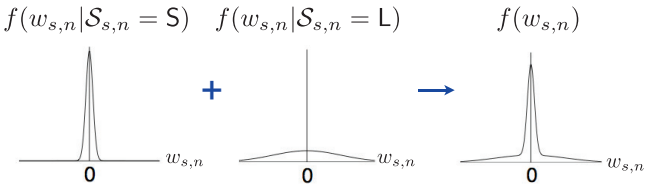

Fig. 2: *Modeling of the heavy-tailed wavelet coefficient distribution as a mixture of Gaussians corresponding to large and small states.*

wavelet coefficient with its parent and child (if they exist). As in HMTs, each wavelet coefficient $w_{s,n}$ is statistically modeled using a mixture of two Gaussians (cf. Fig. 2): the first component features a large variance $\sigma^2_{\mathsf{L},s,n}$ that models large nonzero coefficients and receives a small weight $p^{\mathsf{L}}_{s,n}$ (to encourage few such coefficients and preserve energy compaction), while the second component features a small variance $\sigma^2_{\mathsf{S},s,n}$ that models small and zero-valued coefficients and receives a large weight $p^{\mathsf{S}}_{s,n} = 1 - p^{\mathsf{L}}_{s,n}$. We distinguish these two components by associating to each wavelet coefficient $w_{s,n}$ an unobserved hidden state $\mathcal{S}_{s,n} \in \{\mathsf{S}, \mathsf{L}\}$, with probabilities $p(\mathcal{S}_{s,n} = \mathsf{S}) = p^{\mathsf{S}}_{s,n}$ and $p(\mathcal{S}_{s,n} = \mathsf{L}) = p^{\mathsf{L}}_{s,n}$. The value of $\mathcal{S}_{s,n}$ determines which of the two components of the mixture model is used to generate the probability distribution $f(w_{s,n})$ for $w_{s,n}$: $f(w_{s,n}|\mathcal{S}_{s,n} = \mathsf{S}) = \mathcal{N}(0, \sigma^2_{\mathsf{S},s,n})$ and $f(w_{s,n}|\mathcal{S}_{s,n} = \mathsf{L}) = \mathcal{N}(0, \sigma^2_{\mathsf{L},s,n})$, with $\sigma^2_{\mathsf{L},s,n} > \sigma^2_{\mathsf{S},s,n}$.

The persistence of large and small coefficients from parent to child is well-modeled by a Markov chain that links their coefficient states. This induces the NHMC graphical model on the coefficient array W, where the state $\mathcal{S}_{s,n}$ of a coefficient $w_{s,n}$ is affected only by the state $\mathcal{S}_{s-1,n}$ of its parent $w_{s-1,n}$. The NHMC is then completely determined by the set of state transition matrices for the different parent-child label pairs $(\mathcal{S}_{s,n}, \mathcal{S}_{s+1,n})$:

$$A_{s,n} = \begin{bmatrix} p^{\mathsf{S}\to\mathsf{S}}_{s,n} & p^{\mathsf{S}\to\mathsf{L}}_{s,n} \\ p^{\mathsf{L}\to\mathsf{S}}_{s,n} & p^{\mathsf{L}\to\mathsf{L}}_{s,n} \end{bmatrix}.$$

The persistence property implies that the values of $p^{\mathsf{L}\to\mathsf{L}}_{s,n}$ and $p^{\mathsf{S}\to\mathsf{S}}_{s,n}$ are significantly larger than their complements $p^{\mathsf{L}\to\mathsf{S}}_{s,n}$ and $p^{\mathsf{S}\to\mathsf{L}}_{s,n}$, respectively. Note that all the probabilities $p^{\mathsf{S}}_{s,n}$ and $p^{\mathsf{S}}_{s,n}$ can be computed from $\{A_{s,n}\}$, $p^{\mathsf{S}}_{1,n}$, and $p^{\mathsf{S}}_{1,n}$.

We separately train an NHMC on each of the N wavelengths or frequencies sampled by the hyperspectral acquisition device in order to capture the dynamics of observable spectral signatures for each wavelength individually. While the overlap between wavelet functions at a fixed scale and neighboring offsets introduces correlations between the corresponding wavelet coefficients, we consider each NHMC of parent-child wavelet coefficients independently for computational reasons. The set of NHMC parameters Θ_n include the probabilities for the first hidden states $p^{\mathsf{S}}_{1,n}$ and $p^{\mathsf{L}}_{1,n}$, the state transition matrices $\{A_{s,n}\}_{s=1}^{S}$, and Gaussian variances $\{\sigma^2_{\mathsf{L},s,n}, \sigma^2_{\mathsf{S},s,n}\}_{s=1}^{S}$ — each of these for $1 \le n \le N$.

NHMC training (e.g., picking the values of Θ_n) is performed via an expectation maximization (EM) algorithm that

maximizes the likelihood of a library of training data given the model parameters. The iterative algorithm can be briefly described as follows (with some initial set of parameters Θ_n^0):

1. **E step:** Calculate $p(\mathcal{S}_n|W_n, \Theta_n^i)$, the joint probability for the hidden state variables, using a forward-backward algorithm [8].
2. **M step:** Choose new model parameters by maximizing the expected likelihood of the coefficients and state labels: $\Theta_n^{i+1} = \arg\max_\Theta E_{\mathcal{S}_n}[\ln f(W, \mathcal{S}|\Theta)|W, \Theta]$. The values are obtained by computing weighted means of the likelihoods/variances over the training set [10].
3. If converged, stop; otherwise, increment i and repeat.

The model training dataset is, ideally, as large and diverse as possible, ultimately aiming to obtain a universal model for all observable samples. Given the model, the state labels $\{\mathcal{S}_{s,n}\}$ for a given observation are obtained using a Viterbi algorithm [8, 10] that employs the Gaussian parameters and transition probabilities in $\{\Theta_n\}$. The algorithm also returns the likelihood $f(W|\Theta)$ of a wavelet coefficient array W under the model Θ as a byproduct. We propose the use of the parameters Θ, state labels $\mathcal{S}$, and likelihood $f(W|\Theta)$ as representations of the original hyperspectral signal x. In the sequel, we discuss particular example problems that use the set of state labels $\mathcal{S}$ and present promising preliminary results.

4. MATERIAL IDENTIFICATION

Material identification is one of the most common applications in hyperspectral data processing. The goal is to identify the dominant material species in an observed spectrum by comparing it with a library of reference signatures for common materials. Challenges in material identification arise due to the effect of the atmosphere, illumination geometry, instrumental artifacts and sensor noise on remotely sensed spectra.

Physically-based approaches for spectral representation are routinely used in material identification [3, 4, 11]. Automated approaches for material identification often require the computation of similarity measures between the unknown spectrum and each of the library spectra to determine the best match. Popular examples include cross-correlation spectral matching [12], spectral information divergence [13], and spectral angle mapping [14]. Such measures lack the sophistication needed to quantify subtle spectral differences.

The wavelet representations exploited by our model provide a succinct characterization of diagnostic mineral features. Intuitively, the presence of a fluctuation in a spectral signature is represented by chains of large state labels $\mathcal{S}_{s,n}$ appearing for the corresponding sampled wavelengths (cf. Fig. 1). Thus, it is easy to identify the presence of such semantic cues simply by inspecting the state labels obtained from the NHMC. The discriminating nature of the state labels allows for efficient and simple classification algorithms such as nearest-neighbor search, in contrast with existing ad-hoc approaches that aim to capture these cues using custom similarity measures.

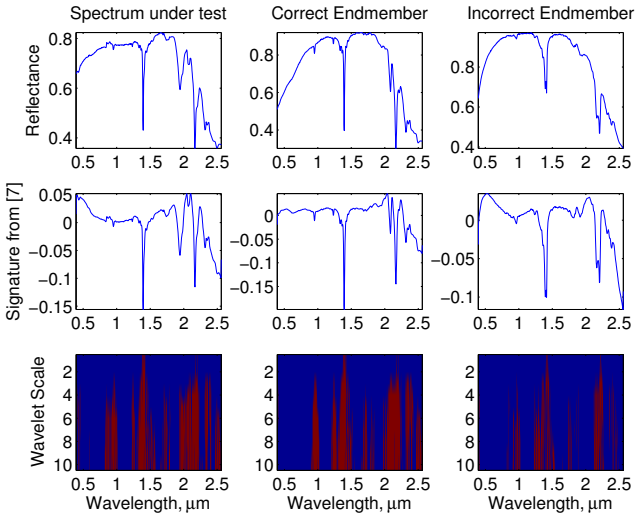

Fig. 3: *Comparison of classification results for USGS spectral library. Left column: spectrum being identified (pyrophy3); middle column: correct endmember (pyrophy2); right column: incorrect endmember returned by the approach of [7] (kaolini1). Top row: reflectances; middle row: filtered signatures from [7]; bottom row: labels from NHMC for wavelet coefficients.*

We showcase the potential of our NHMC-derived representation by studying two material identification scenarios. The first problem considers the classification of 57 clay mineral samples in 12 different classes from the USGS spectral library (cf. [7, Table 2]). The approach of [7] computes a 10-level continuous wavelet decomposition of the mineral spectra and produces a signature by filtering the wavelet decomposition to ignore the 4 coarsest scales. This signature is used as a feature for a minimum angle classifier; the reported accuracy is 89%, and we reached an accuracy of 84% in our implementation. In contrast, we use the state labels $\mathcal{S}$ assigned to each material sample as a feature vector and employ a simple nearest-neighbor search classifier. Since the state labels are binary, we use the Hamming distance as a metric of similarity. The results show a classification rate of 95%, significantly outperforming the approach of [7]. As can be seen in Fig. 3 (sample *pyrophy3*), the approach of [7] incorrectly selects an endmember that provides a better match to the remnants of the continuum after filtering. In contrast, our state labels selects the correct endmember: it encodes the presence of discontinuities and determines their relevance through the library's NHMC statistical model.

The second problem considers the segmentation of a hyperspectral image acquired by the AVIRIS spectrometer over Cuprite, NV in 1995 [15–17]. The hyperspectral image represents a section of Cuprite of size 350×400 pixels. Spectral bands covering the short wave infrared (SWIR) spectral range ($2.0 - 2.5$ μm) were selected due to the presence in that range of diagnostic absorptions for the minerals in the scene and in order to avoid residual atmospheric contributions to the spectra. A USGS library is used to train the NHMCs, and each spectrum in the image is mapped to one of the library spec-

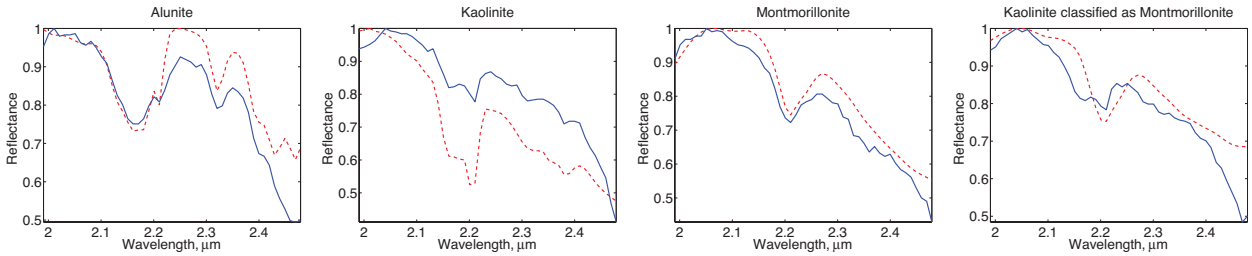

Fig. 4: *Mineral identification results for AVIRIS image data with USGS library. Continuous line: image sample; dotted line: library sample.*

tra using Hamming distances on the obtained state labels $\mathcal{S}$. Figure 4 shows example matches in the library for several image spectra. The first three examples show correct matchings for alunite, kaolinite, and montmorillonite samples. The last figure shows a kaolinite sample that is incorrectly matched to montmorillonite in the library. Note that the kaolinite sample in the second figure is more reminiscent of the library sample of the same material, while the ridges of the last kaolinite sample are less prominent, causing a mislabeling. A preliminary study on the overall accuracy improvement of the proposed method over traditional automated approached used a random sampling of spectral matches, under visually assessment by researchers. The proposed method exhibited an accuracy 15% higher than the classical SAM method.

5. CONCLUSIONS AND FUTURE WORK

We presented a novel mathematical representation of hyperspectral signals that encode scientifically meaningful information by means of a statistical model of the signal's wavelet representation. We leveraged the multiscale analysis properties of the wavelet transform to succinctly express informative semantic cues on the spectrum's structure. The method was validated on material identification tasks where it outperformed some classic classification approaches. Further work will expand the set of applications of the proposed model to include image segmentation, where an augmented model can capture spatial correlations in the field of view and provide further discrimination between classes. Finally, we will characterize the effect of the physics of acquisition and other non-idealities in the applicability of the proposed models.

REFERENCES

[1] A. Plaza, J. Benediktsson, J. Boardman, and 10 others, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. 1, pp. S110–122, 2009.

[2] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Select. Top. Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[3] R Clark, Gregg 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.

[4] M. Parente, H. D. Makarewicz, and J. L. Bishop, "Decomposition of mineral absorption bands using nonlinear least squares curve fitting: Application to martian meteorites and CRISM data," *Planet. Space Sci.*, vol. 59, no. 5–6, pp. 423 – 442, 2011.

[5] M. Parente, J. F. Mustard, S. L. Murchie, and F. P. Seelos, "Robust unmixing of hyperspectral images: Application to Mars," in *IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Vancouver, Canada, July 2011, pp. 1291–1294.

[6] L. M. Bruce, C. H. Koger, and J. Li, "Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction," *IEEE Trans. Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2331–2338, Oct. 2002.

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

[8] 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.

[9] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, USA, 1999.

[10] 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.

[11] S. M. Pelkey, J. F. Mustard, S. L. Murchie, and 11 others, "CRISM multispectral summary products: Parameterizing mineral diversity on Mars from reflectance," *Journal of Geophysical Research*, vol. 112, no. E08, pp. July, 2007.

[12] F. van der Meer and W. Bakker, "CCSM: Cross correlogram spectral matching," *Int. J. Remote Sens*, vol. 18, no. 5, pp. 1197–1201, 1997.

[13] C.-I Chang, "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1927–1932, 2000.

[14] R. Yuhas, A. Goetz, and J. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *JPL AVIRIS Workshop*, Pasadena, CA, June 1992, pp. 147–149.

[15] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for Earth remote sensing," *Science*, vol. 228, pp. 1147–1153, 1985.

[16] G. Swayze, R. Clark, F. Kruse, S. Sutley, and A. Gallagher, "Ground-truthing AVIRIS mineral mapping at Cuprite, NV," in *JPL AVIRIS Workshop*, Pasadena, CA, June 1992, pp. 47–49.

[17] F. A. Kruse, J. W. Boardman, and J. F. Huntington, "Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping," *IEEE Trans. Geosci. and Remote Sens.*, vol. 41, no. 6, pp. 1388–1400, 2003.