

Received 20 October 2023, accepted 2 December 2023, date of publication 5 December 2023,
date of current version 13 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3339565

RESEARCH ARTICLE

Wavelet-Based Semantic Feature Representation for Hyperspectral Signature Discrimination

SIWEI FENG 

School of Computer Science and Technology, Soochow University, Suzhou 215000, China
e-mail: swfeng@suda.edu.cn

This work was supported by the National Natural Science Foundation of China under Grant NSFC 62106167.

ABSTRACT Hyperspectral signature discrimination is a quantitative analysis approach for hyperspectral imagery which performs detection and classification of the constituent materials at the pixel level in the scene. However, the substantial volume of data in hyperspectral images poses challenges for practitioners in terms of computation and storage, making the management of acquired hyperspectral data difficult. One alternative to directly operating on raw data is to perform classification using extracted features from hyperspectral signatures, such as energy or shape information. Therefore, there is a need to design features that capture scientifically meaningful cues of the materials under study from hyperspectral signatures. The wavelet transform is widely recognized for its ability to highlight signal discontinuities, making it suitable for representing spectral fluctuations with semantic value. By appropriately modeling wavelet coefficients, it becomes possible to distinguish between fluctuations that have discriminatory value and those that do not. In this paper, we describe a technique that applies non-homogeneous hidden Markov chain (NHMC) models to hyperspectral signature discrimination. The basic idea is to use statistical models (such as NHMC) to characterize wavelet coefficients which capture the spectrum semantics (i.e., structural information) at multiple levels. Experimental results show that the approach based on NHMC models can outperform existing approaches relevant in classification tasks.


INDEX TERMS Hyperspectral signature discrimination, non-homogeneous hidden Markov chain, semantic feature representation.

I. INTRODUCTION

Hyperspectral remote sensors capture reflected image data across hundreds of narrow, adjacent spectral bands, enabling the derivation of continuous spectrum curves for each pixel. These hyperspectral reflectance curves offer valuable insights into the constituent materials present on or near the ground within a single remotely sensed pixel.

The identification of ground materials in hyperspectral images often involves comparing the reflectance spectra of image pixels, extracted endmembers, or ground cover exemplars to a training library of spectra obtained from well-characterized samples in the laboratory. While there is a substantial body of literature on hyperspectral image classification [1], methods emphasizing spectral library matching and material identification have received comparatively less

attention [2], [3], [4]. Some approaches rely on nearest neighbor classification schemes utilizing various spectral similarity measures, while others employ feature extraction techniques to discriminate between different materials. Classification methods based on spectral similarity measures offer simplicity and lower computational requirements, but require significant storage for training spectra and may exhibit uneven performance. In contrast, practitioners recognize specific structural features in spectral curves, such as the position and shape of absorption bands, as characteristic of a material's chemical makeup. Techniques like Tetracorder [5] have been proposed to encode such features, but often require the construction of ad-hoc rules and lack flexibility for analyzing previously unexplored spectral species. Similarly, parametric models [6] have been used to represent absorption features, but still necessitate the formulation of specific rules for matching observations to the training library.

The associate editor coordinating the review of this manuscript and approving it for publication was Brian Ng .

In the past decade, two significant trends have emerged that enhance the performance of hyperspectral signature discrimination. The first one is the integration of deep learning [7], [8], [9], [10], [11], [12], [13], [14], which has gained significant attention in recent years [15], [16], [17], [18], [19], [20]. These approaches have shown promise in enhancing hyperspectral signature discrimination performance and eliminating the need for ad-hoc rules. However, it is important to acknowledge the limitations associated with these methods. These limitations include the requirement for abundant labeled training data, challenges posed by the curse of dimensionality arising from the high number of spectral bands, difficulties in capturing subtle spectral variations and addressing the high correlation between adjacent bands, the necessity to effectively incorporate spatial-spectral context, the lack of interpretability and explainability in deep learning models, and the high computational demands during training and deployment. The second one is the combination with spectral and spatial information [21], [22], [23], [24], [25]. The use of spectral-spatial information enhances classification accuracy by capturing contextual relationships and differentiating similar spectral signatures. But on the other side, integrating spatial information increases computational complexity, requiring additional resources and potentially leading to longer processing times. Spatial-spectral methods are sensitive to spatial variability, and their effectiveness depends on the alignment between spectral and spatial characteristics. In cases where these do not align or exhibit significant variability, the performance of these methods may be compromised. Furthermore, spatial-spectral methods may not be applicable to all hyperspectral classification tasks, as the inclusion of spatial information may not provide significant improvements when spectral information alone is sufficient. Factors such as image resolution, scene complexity, and class distribution can also affect the effectiveness of these methods. In this paper, our objective is to develop a computationally efficient and resource-friendly hyperspectral signature discrimination framework that prioritizes portability. Consequently, we exclude the utilization of deep learning-based approaches and spatial information.

Mathematical signal models have proven valuable in representing reflectance spectra, with wavelet decompositions being particularly appealing. Wavelet-based models offer the advantage of capturing structural features at various scales. The wavelet transform is widely employed in signal processing due to its ability to characterize signal discontinuities across scales and offsets through wavelet coefficients. In the context of hyperspectral signature analysis, the shape of reflectance spectra plays a crucial role in capturing semantic information. This information is succinctly encoded in the magnitudes of wavelet coefficients. Notably, a large magnitude coefficient indicates a rapid change in the signal, while a small coefficient suggests a smoother region. However, existing wavelet approaches primarily focus on filtering techniques and lack feature extraction capabilities [2], [3].

This paper introduces a novel information extraction process for hyperspectral signatures utilizing tailored mathematical models designed specifically for hyperspectral signals. Our primary goal is to capture the scientifically significant structural features inherent in the signatures and express them as numerical features, referred to as semantic features, without relying on ad-hoc rules specific to material types. Unlike traditional manual approaches used by expert practitioners to design diagnostic characteristics, our method automates the extraction of semantic information from hyperspectral signatures. Remarkably, our approach eliminates the need for creating new rules when previously unanalyzed mineral species are incorporated into the analysis. To achieve this, we employ hidden Markov models (HMMs) applied to wavelet coefficients derived from observed hyperspectral signals. This allows us to capture the correlations between wavelet coefficients at adjacent scales, facilitating the identification of significant and nonsignificant portions of hyperspectral signatures based on the training database. The inspiration for utilizing HMMs for this purpose stems from the hidden Markov tree (HMT) model proposed in [26]. For the wavelet transform, we utilize an undecimated wavelet transform (UWT) to maximize flexibility in the selection of scales and offsets (referred to interchangeably as spectral bands or wavelengths). Our model, designed for a spectrum encompassing N spectral bands, comprises a collection of N non-homogeneous hidden Markov chains (NHMCs), each corresponding to a specific spectral band. This model effectively maps each signal spectrum to a binary space, encoding the structural features at different scales and wavelengths and representing the semantic features necessary for spectrum discrimination. Importantly, to the best of our knowledge, the application of statistical wavelet models for the automatic selection of semantically meaningful features in hyperspectral signatures has not been previously proposed.

The key contributions of this paper are as follows:

- A novel lightweight hyperspectral signature discrimination framework for efficient material type classification of individual pixels in hyperspectral images is proposed. The proposed framework eliminates the reliance on expert knowledge and removes the need for ad-hoc rules, which is different from traditional manual approaches. Moreover, unlike the recently popular methods that heavily rely on deep learning or exploit spatial-spectral information, the proposed framework offers the advantage of minimal resource requirements in terms of computation and storage.
- A novel general framework to automatically select semantically meaningful features in hyperspectral signatures is proposed.
- Multiple experimental results to demonstrate the performance achieved by the proposed framework compared with relevant techniques from the state of the art and the corresponding analysis are provided.

This paper is organized as follows. In Section II, we present the mathematical background of our hyperspectral signature

classification system and provide a comprehensive review of existing approaches relevant to hyperspectral classification. Section III offers an in-depth overview of our proposed feature extraction method, including key details about the selection of the mother wavelet, training of the statistical model, and computation of labels. Additionally, we showcase examples that demonstrate how the proposed features capture semantic information within hyperspectral signatures. In Section IV and V, we outline our experimental test setup and present the corresponding results. Section VI contains our concluding remarks. Finally, we provide the proofs of our theoretical results in the appendix.

II. BACKGROUND AND RELATED WORK

In this section, we review the theoretical background for our proposed hyperspectral signature discrimination system, including wavelet analysis, hidden Markov chain models, and the Viterbi algorithm.

A. WAVELET ANALYSIS

Wavelet transforms encode signal structure at different scales effectively. Several hyperspectral classification methods utilize wavelet transforms, with most employing dyadic/decimated wavelet transforms (DWT) as a preprocessing step for concise representation. However, undecimated wavelet transforms (UWT) offer maximum flexibility in scales and offsets, allowing simple characterization of spectrum structure for each spectral band.

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

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

where λ is a scalar. 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 structure in the 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 .

B. ADVANTAGES OF HAAR WAVELET

The Haar wavelet captures rapid changes and gradual fluctuations in a signal, while being less sensitive to small discontinuities. The magnitudes of Haar wavelet coefficients indicate slope strength and direction. The Daubechies-4 wavelet responds to compact discontinuities but fails to capture gentle slopes and lower-order fluctuations effectively. It does not provide slope direction information. These limitations apply to other higher-order wavelets as well. In general, wavelet representations of spectral absorption

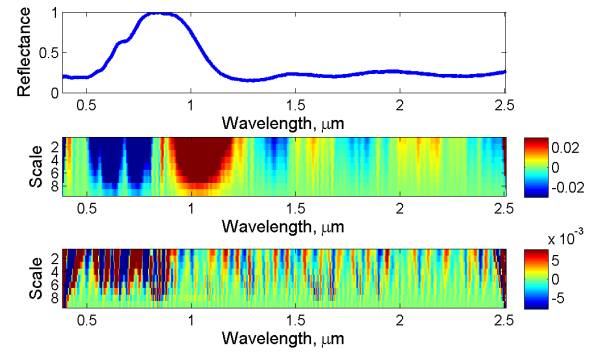


FIGURE 1. Top: an example of normalized mineral reflectance (Garnet). Middle: corresponding UWT coefficient matrix (9-level wavelet decomposition) using a Haar wavelet. Bottom: corresponding UWT coefficient matrix using a Daubechies-4 wavelet.

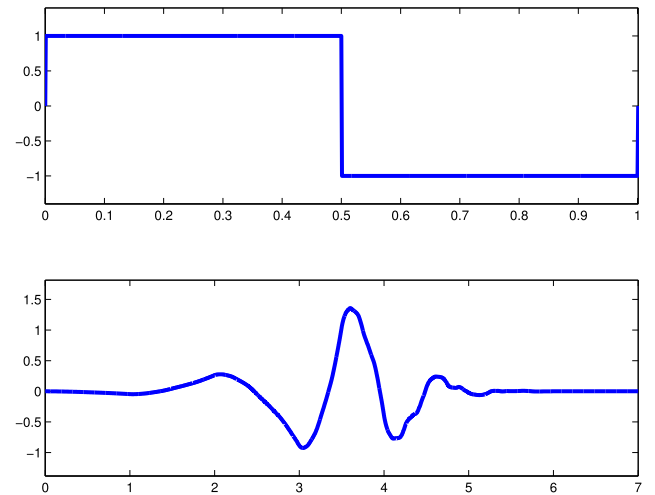


FIGURE 2. Illustration of wavelet base function. Top: Haar wavelet. Bottom: Daubechies-4 wavelet.

bands are less emphasized under Haar wavelet than under other higher order wavelets. However, this drawback can be alleviated using discretization, which will be described in the next subsection.

C. STATISTICAL MODELING OF WAVELET COEFFICIENTS

The pairwise statistics of DWT coefficients can be succinctly captured by a hidden Markov model [26]. 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 constructed based on the wavelet representation of spectra in a training library.

The statistical model employed here leverages the energy compaction property of the Discrete Wavelet Transform (DWT). While wavelet coefficient histograms generally exhibit a Gaussian-like shape with a peak around zero and tails on both sides, the actual peak is narrower than a Gaussian distribution, and the histograms tend to have heavy tails. These characteristics reflect the compression property of

wavelet coefficients in which a piecewise smooth signal typically consists of a small number of large coefficients and a large number of small coefficients. These properties cannot be adequately described by a Gaussian model. Hence, a zero-mean Gaussian mixture model (GMM) with two Gaussian components is used. The first component, labeled as “L”, has a high variance and represents the small number of “large” coefficients. The second component, labeled as “S”, has a low variance and represents the large number of “small” wavelet coefficients. The conditional probability of a wavelet coefficient w_s given the state S_s is expressed as $p(w_s|S_s = i) = \mathcal{N}(0, \sigma_{i,s}^2)$, where $i = S, L$. The hidden state $S_s \in S, L$ collects these labels and is not directly observed. The likelihoods $p_{S_s}(L)$ and $p_{S_s}(S)$ represent the probabilities of the two states, with the constraint that $p_{S_s}(L) + p_{S_s}(S) = 1$. Consequently, the distribution of the wavelet coefficient can be expressed as $p(w_s) = p_{S_s}(L)\mathcal{N}(0, \sigma_{L,s}^2) + p_{S_s}(S)\mathcal{N}(0, \sigma_{S,s}^2)$.

UWT coefficients demonstrate a persistence property [27], [28], where adjacent scales consistently exhibit high or low values with high probability. This property can be effectively captured by a non-homogeneous hidden Markov chain (NHMC) that connects the states of wavelet coefficients at the same offset. In this NHMC, the state of a coefficient w_s (denoted as S_s) is influenced solely by the state of its parent (S_{s-1} , if it exists) and the value of the coefficient w_s . The Markov chain in the NHMC is parameterized by the likelihoods for the first state $P_{S_1}(L)$ and $P_{S_1}(S)$, as well as the set of state transition matrices for the different parent-child label pairs (S_{s-1}, S_s) for $s > 1$:

$$A_s = \begin{pmatrix} p_{S \rightarrow S, s} & p_{L \rightarrow S, s} \\ p_{S \rightarrow L, s} & p_{L \rightarrow L, s} \end{pmatrix}, \quad (1)$$

where $p_{i \rightarrow j, s} := P(S_s = j | S_{s-1} = i)$ for $i, j \in \{L, S\}$. The parameters for the NHMC

$$\theta = \{p_{S_1}(S), p_{S_1}(L), \{A_s\}_{s=2}^L, \{\sigma_{S,s}, \sigma_{L,s}\}_{s=1}^L\}$$

(which include the probabilities for the first hidden states, the state transition matrices, and Gaussian variances for each of the states) that maximize the likelihood of a set of observations can be obtained via the expectation maximization (EM) algorithm [26], [29]. We define the $L \times N$ matrix S containing the collection of state values for all scales and spectral bands. We use a matrix w to denote the wavelet coefficients of spectra in a training library which can be referred to as the incomplete data. And the complete data is (w, s) , where s is the matrix of corresponding state labels of w . At the l th iteration, the E step computes an estimate of the state labels as $E_s[\ln f(w, s|\theta)|w, \theta^l]$ and then the M step returns the value of θ that maximizes the expectation as θ^{l+1} for the next iteration. 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 [30]:

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

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 [26]:

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

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

D. WAVELET-BASED SPECTRAL MATCHING

Several wavelet-based hyperspectral signature discrimination methods have been proposed [2], [3], [4], [31], [32]. In this paper, we focus on two benchmark approaches that exclusively use spectral information for individual pixel classification. We exclude methods with additional data or prior knowledge. Rivard et al. [2] employ wavelet decomposition, classifying wavelet coefficients into low-scale components of power (LCP) and high-scale components of power (HCP). LCP coefficients preserve structural features and are summed across scales for each spectral band, acting as a high-pass filter.

In [3], a wavelet-based approach applies Undecimated Wavelet Transform (UWT) on the entire database. Each wavelength's wavelet coefficients form a distinct feature vector. Linear Discriminant Analysis (LDA) reduces dimensionality. Outputs are grouped into C classes, training either a single Gaussian distribution or Gaussian Mixture Model (GMM) for each class. Classification labels or scores are obtained for each wavelength, and decision fusion yields a single label for the spectrum. Sufficient training samples are assumed for accurate model construction.

III. NHMC-BASED FEATURE REPRESENTATION DESIGN

In this section, we introduce a feature extraction scheme for hyperspectral signatures that exploits a Markov model for the signature's wavelet coefficients. A wavelet analysis is used in an UWT to capture information on the fluctuations of the spectra. The state labels extracted from the Markov model represent the semantic information relevant for hyperspectral signal processing.

Figure 3 presents an overview of the NHMC-based hyperspectral classification system. It comprises two modules: a training module and a classification module. In the training stage, the NHMC model is trained using a library of spectra containing samples from the target classes. The Viterbi Algorithm is employed to compute state estimates for each training spectrum. The obtained state arrays serve as classification features, which are combined with a chosen classification scheme such as nearest-neighbor (NN) or support vector machine (SVM). In the testing module, the NHMC model with the trained parameters is used to compute state estimates for a test spectrum. The classification scheme applied during training is then utilized to assign a label to the tested spectrum.

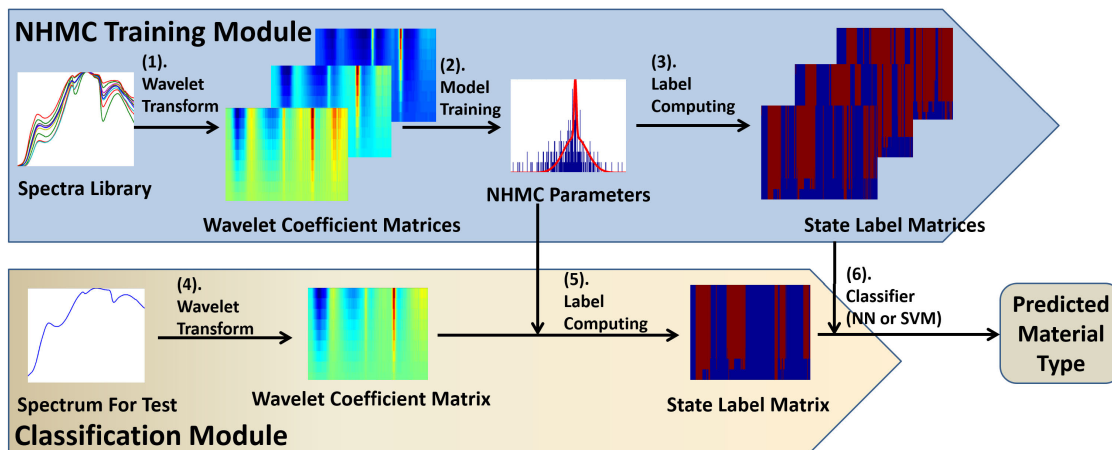


FIGURE 3. NHMC classification overview. Top: The classifier is trained from a set of training spectra; we compute UWT coefficients for each spectrum, and feed the wavelet coefficient vectors to an NHMC training module that outputs the model parameters and state label arrays for each of the training spectra, to be used as semantic features. Bottom: The classifier considers a test spectrum, obtains its UWT coefficients, and extracts a state array from the aforementioned NHMC model to use as semantic features. A nearest-neighbors search returns the most similar state array among the training data, as well as the class label for the corresponding spectrum.

A. SEMANTIC FEATURES FROM NHMC LABELS

After training the model, state values S can be estimated from wavelet coefficients w using the Viterbi algorithm [26], [30]. States are organized into an array S to form a semantic feature vector for the spectrum. “Large” states represent significant fluctuations, while “small” states indicate the absence of fluctuations or lack of statistical significance. Figure 4 illustrates an example of a state label array using a GMM with 2 Gaussian states, highlighting meaningful fluctuations in an ilmenite sample’s spectrum.

However, Figure 4’s third row lacks information about the direction of fluctuations, only indicating their magnitude. This limitation reduces the discriminative power of the designed features for hyperspectral pixels. Fortunately, when using the Haar wavelet in the UWT, the wavelet coefficient vector contains information about fluctuation direction [33]. To address this, we enhance the state array S by incorporating the sign of the corresponding wavelet coefficient for each state label. The bottom row of Figure 4 demonstrates that the augmented feature effectively distinguishes between rising and falling fluctuations, providing a clearer characterization of the structural information in the hyperspectral curve.

To enhance the discriminative power, a finer-scale quantization of spectral signature fluctuations is needed. Particularly, the state labels fail to detect fluctuations in the $0.6 - 0.8 \mu\text{m}$ region, and narrow chains of “large” labels appear at $1.7 \mu\text{m}$. Increasing the number of Gaussian components in the GMM allows for a finer quantization. This adjustment is crucial when using the Haar wavelet, known for its sensitivity to various fluctuation orders and relatively lower discriminative power. By increasing the components beyond $k \geq 2$, the NHMC model generates richer spectral features. The semantic feature array S contains entries from the set $0, 1, \dots, k - 1$ (or $-k + 1, \dots, -1, 0, 1, \dots, k - 1$

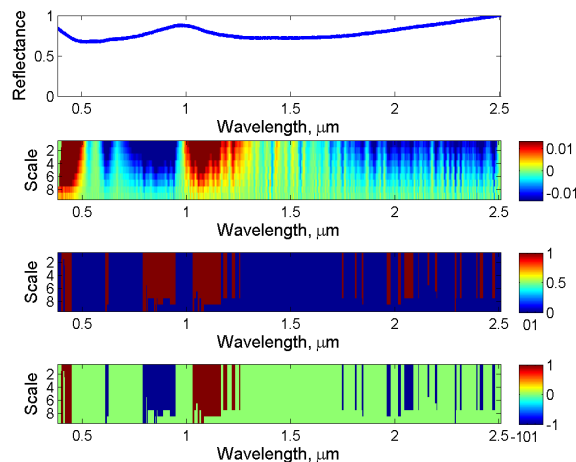


FIGURE 4. Examples of signed state labels as classification features. Top: Example normalized reflectance spectrum (ilmenite). Second: Corresponding 9-scale UWT coefficient matrix using a Haar wavelet. Third: Corresponding state label matrix from an NHMC model using a zero-mean two-state GMM. Blue represents 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.

with augmented wavelet coefficient sign). Class 0, with the smallest variance, effectively identifies non-fluctuating or statistically insignificant spectra, surpassing features obtained for $k = 2$. The subsequent section discusses increasing Gaussian components in a GMM.

B. MULTI-STATE HIDDEN MARKOV CHAIN MODEL

In our system, we combine the NHMC model and the Haar wavelet applied through the UWT (Section II-C). The Haar wavelet’s distinct shape allows the coefficients to capture

slope information in the spectra and their signs indicate the orientation of the slope (positive for increasing, negative for decreasing). Unlike Crouse et al. [26], our NHMC approach uses GMMs with $k > 2$ to model each wavelet coefficient. Simulations have shown that NHMC with $k = 2$ produces coarse state labels, often overlooking weak absorption bands. Larger values of k improve classification performance, as demonstrated in subsequent sections.

We associate each wavelet coefficient w_s with an unobserved hidden state $S_s \in \{0, 1, \dots, k - 1\}$, where the states have prior probabilities $p_{i,s} := p(S_s = i)$ for $i = 0, 1, \dots, k - 1$. 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 $i = 1, \dots, k - 1$ 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 $\sigma_{i,s}^2$. 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

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

where $p_{i \rightarrow j,s} = p(S_s = j | S_{s-1} = i)$. 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 $p_{i,s} > 1$ can be derived from the matrices $\{A_s\}_{s=2}^L$ and $\{p_{i,1}\}_{i=0}^{k-1}$.

The k -GMM NHMC is trained using an EM algorithm. Due to the overlap between wavelet functions at a fixed scale and neighboring offsets, adjacent coefficients may exhibit correlations in relative magnitudes [34]. However, for computational efficiency, we solely consider the parent-child relationship of wavelet coefficients within the same offset. Specifically, we train an NHMC independently for each of the N wavelengths captured by the hyperspectral acquisition device. The set of NHMC parameters θ_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 $\{\sigma_{0,s,n}^2, \sigma_{1,s,n}^2, \dots, \sigma_{k-1,s,n}^2\}_{s=1}^L$. In the sequel, we remove from the parameters θ the dependence on the wavelength index n whenever possible.

Given the model parameters θ , the state label values $\{S_s\}_{s=1}^L$ for a given observation are obtained using a Viterbi algorithm. 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 $i = 0, \dots, k - 1$. 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

$$\psi_{i,1} = 0, \tag{2}$$

$$\delta_{i,1} = p_{i,1} \cdot p(w_1|S_1 = i), \tag{3}$$

$$\psi_{i,s} = \arg \max_{j=0, \dots, k-1} (\delta_{j,s-1} p_{j \rightarrow i,s}), \tag{4}$$

$$\delta_{i,s} = \delta_{\psi_{i,s,s-1} p_{\psi_{i,s,s-1} \rightarrow i,s}} \cdot p(w_s|S_s = i), \tag{5}$$

for $i = 1, \dots, k - 1$ and $s = 2, \dots, L$. The algorithm also returns the likelihood $p(W|\theta)$ of a wavelet coefficient matrix W under the model θ 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.

C. ADDITIONAL MODIFICATIONS TO NHMC

A high number of GMM states can negatively impact semantic modeling due to spectral variability caused by diverse reflectance values across bands. In our experiments, we observe differences in the maps linking fluctuation rates to k state labels for neighboring bands’ wavelet coefficients. Assigning k -ary labels to fluctuations across bands is non-uniform, complicating the evaluation of spectral signatures using semantic features.

The determination of the GMM state for a wavelet coefficient $w_{s,n}$ depends on its magnitude relative to other coefficients, the parent’s state label $S_{s-1,n}$, and the transition probability matrix $A_{s,n}$. This leads to varying mappings between coefficient values and GMM states, making interpretation of the label array challenging. The variability can affect the interpretability of features derived from GMM labels and impact classification methods using NN classifiers. To address this, we propose the mixture-of-Gaussians (MOG) NHMC, combining benefits of binary-state and k -state GMMs. MOG-NHMC merges k -state GMM labels representing fluctuations into a single “large” state label, similar to label 1. Label 0 remains unchanged for “small” coefficients. Statistical models, likelihoods, and inferred labels for MOG-NHMC can be computed using existing methods for k -GMM NHMC, with the “large” state corresponding to a mixture of $k - 1$ Gaussians. Refer to [29] for more details.

Our modified wavelet coefficient model consists of a binary-state NHMC. The “small” state (0) is modeled by a standard zero-mean Gaussian distribution, while the “large” state (1) is modeled by a mixture of $k-1$ Gaussians. We use numbers for state labels to differentiate between the

2-state GMM NHMC and the 2-state MOG NHMC. This modified model, called MOG NHMC, maintains discriminability between smooth regions and absorption bands in spectral signatures while decreasing the likelihood of the mentioned variability.

In order to obtain a MOG NHMC model, the first step is to train a k -state GMM NHMC model that yields state labels $S_s \in \{0, \dots, k-1\}$. After that, all the states are quantized into two states so that we can get a MOG NHMC that yields state labels $Z_s \in \{0, 1\}$ with probabilities $q_{i,s} = P(Z_s = i)$, $i = 0, 1$. One can show that the change of models lead to the following mapping for labels:

$$Z(S) = \begin{cases} 0 & \text{if } S = 0, \\ 1 & \text{if } S \neq 0. \end{cases} \quad (6)$$

Similar to (1), we can define a transition probability matrix

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

for the MOG NHMC, where $q_{i \rightarrow j,s} := P(Z_s = j | Z_{s-1} = i)$ for $i, j \in \{0, 1\}$ and $s = 1, \dots, L$. We have the following pair of intuitive results, whose proves are presented in Appendices A and B.

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

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

The wavelet coefficient matrix shows compression, with most coefficients close to zero and a few larger ones [26]. The smallest state label represents the abundant group of small-valued coefficients. Summing probabilities of other labels avoids population gaps between zero and nonzero labels, preserving classification accuracy. The presence of nonzero labels is crucial to capture subtle differences between label matrices.

We use the same idea that separating the zero-valued state and nonzero-valued states when dealing with transition probability matrix. Let us use a new code name Z_s to denote the state label of a certain wavelet coefficient w_s in the new transitions matrices. Here we introduce some new notations:

$$\begin{aligned} q_{a \rightarrow bs} &= p(Z_s = b | Z_{s-1} = a), \\ p_{a \rightarrow bs} &= p(S_s = b | S_{s-1} = a), \end{aligned}$$

where a and b are appropriate integers for corresponding cases. The proof of Lemma 1 is in Appendix A.

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

Lemma 2: The parameters of the MOG NHMC can be written in terms of the parameters of the GMM NHMC.

$$q_{0 \rightarrow 0,s} = p_{0 \rightarrow 0,s}, \quad (7)$$

$$q_{1 \rightarrow 0,s} = \frac{\sum_{i=1}^{k-1} p_{i \rightarrow 0,s} p_{i,s-1}}{\sum_{i=1}^{k-1} p_{i,s-1}}, \quad (8)$$

$$q_{0 \rightarrow 1,s} = \sum_{j=1}^{k-1} p_{0 \rightarrow j,s}, \quad (9)$$

$$q_{1 \rightarrow 1,s} = \frac{\sum_{i=1}^{k-1} p_{i,s-1} \sum_{j=1}^{k-1} p_{i \rightarrow j,s}}{\sum_{i=1}^{k-1} p_{i,s-1}}. \quad (10)$$

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:

$$\begin{aligned} (0.422, 0.3696, 0.1042, 0.1042)^T &\rightarrow (0.422, 0.578)^T, \\ \begin{pmatrix} 1 & 0.0001 & 0 & 0 \\ 0 & 0.9999 & 0 & 0 \\ 0 & 0 & 0.5 & 0.4999 \\ 0 & 0 & 0.5 & 0.5001 \end{pmatrix} &\rightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

Correspondingly, we also make small modifications to the label computation scheme from Section III-A. For the MOG NHMC, equations (2-5) become

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

respectively, for $i = 0, 1$ and $s = 2, \dots, L$. 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 $k-1$.

Figure 5 provides an example comparison between labels obtained from the k -state GMM NHMC and the MOG-NHMC; the figure highlights the variability obtained when k labels are used in the feature, reducing its semantic significance, while MOG-NHMC retains semantic significance.

D. ILLUSTRATION OF EXTRACTED SEMANTIC INFORMATION

The proposed MOG-NHMC semantic features capture four behaviors of hyperspectral signatures. First, state labels reflect the direction of reflectance fluctuation. Second, label continuity across wavelengths captures fluctuation width.

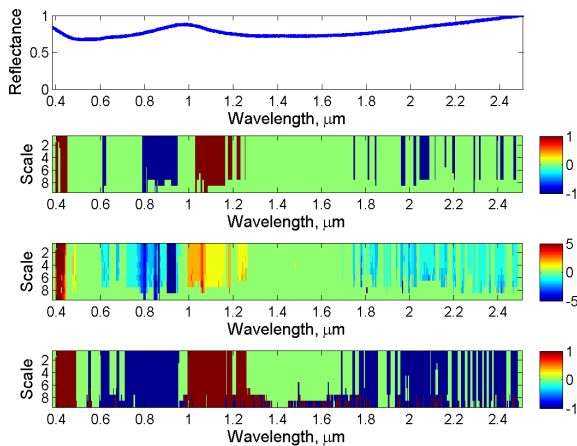


FIGURE 5. Comparison of label arrays obtained from several statistical models for wavelet coefficients. Top: example normalized reflectance, same as in Fig. 4. Second: Corresponding state label matrix from a 2-state GMM NHMC model. Third: Corresponding state label matrix from a 6-state GMM NHMC model. Bottom: Corresponding state label matrix from a MOG NHMC model with $k = 6$.

Third, consistent state values across scales represent fluctuation slope. Fourth, absorption band locations are indicated by transitions in the feature vector. Figure 6 illustrates these behaviors in reflectance spectra. To summarize, we transform the state array S into a vector S_v by selecting the most prevalent state for each band. Reflectance spectra are color-coded based on S_v . Absorption minima locations are labeled where S_v fluctuates from 1 to -1. NHMC extracts semantic features directly from data without relying on expert-defined criteria.

IV. CLASSIFICATION EXPERIMENTS AND RESULT ANALYSIS ON SYNTHETIC DATA

In this section, we present multiple experimental results that assess the performance of the proposed features in hyperspectral signature discrimination. We also study the effect of NHMC parameter selections on the classification performance from the corresponding extracted features. Note that all experimental results are presented in numerical values and curves instead of images due to the exclusion of certain hyperspectral signatures from the analysis.

A. STUDY DATA AND PERFORMANCE EVALUATION

In our simulations, we use the RELAB spectral database with 26 mineral reflectance spectrum classes. We focus on the $0.35 \mu\text{m}$ to $2.6 \mu\text{m}$ spectral region, covering the visible and near-infrared range. The spectra are discretized at 5nm intervals, resulting in 450 wavelengths. To address varying sample sizes, we generate additional samples for each class using the Hapke mixing model [35]. The dataset comprises 1690 normalized reflectance spectra, ensuring consistent illumination conditions for classification.

We evaluate NHMC models with k ranging from 2 to 10, with and without MOG conversion and wavelet coefficient sign augmentation. The data is randomly split into 80%/20% for training and testing, respectively. Classification accuracy is computed using five-fold cross-validation. However, the

high separation of classes leads to consistently high accuracies, making it challenging to distinguish model performance. To address this, we tackle a more challenging problem by identifying dominant elements in material mixtures within a hyperspectral image. The image pixels correspond to testing samples, randomly assigned in a 2D manner. We simulate the image formation process by introducing mixtures and applying spatial blurring using a 3×3 Gaussian kernel. The task difficulty is controlled by adjusting the blurring kernel's variance, quantified by the dominant material percentage (DMP) ranging from 70% to 100% with a step size of 5%.

B. FEATURE COMPARISON

Classification performance is evaluated using NN and SVM accuracies. For NN, three distance metrics are used: ℓ_1 distance, Euclidean (ℓ_2) distance, and cosine similarity. SVM employs RBF kernel with a grid search for optimal parameter values (cost and Gaussian variance). NHMC models (if applicable) and classifiers are trained on the training set and tested on the test set.

Figure 7 displays the classification rates of various NHMC models, employing both NN and SVM classifiers, across different dominant material percentages. The figure also includes the classification accuracies of related approaches outlined in Section II-D. Each classification feature is labeled as follows: “Rivard” represents the method proposed in [2];¹ “Wavelet Coefficient” indicates the scheme using wavelet coefficients as classification features;² “Spectral Similarity” denotes the classification approach based on spectral similarity matching (inputting the spectra themselves into each NN classifier); “GMM” represents an NHMC employing Gaussian mixture models; “MOG” signifies an NHMC utilizing mixtures of Gaussians; and “GMM+Sign” and “MOG+Sign” represent the two aforementioned approaches with the addition of Haar wavelet coefficient signs to state labels. Our NHMC experiments involve models with different numbers of mixed Gaussian components. Figure 7 highlights the highest performance achieved among all tested values, while Tables 1-4 provide the optimal values for each DMP.

We highlight some features of the obtained results:

- In most cases, the use of signs in the NHMC features improves performance with respect to their original counterparts.
- In the NN classifiers, GMM performs better than MOG for lower DMPs, which are more challenging settings, while MOG with additional signs outperforms GMM for DMPs closer to 100%. Nonetheless, in most cases MOG without wavelet coefficient signs provides the worst performance.
- While the performance of NHMC methods with SVM classifiers is higher than that obtained with NN classifiers, they are outperformed by the wavelet coefficient

¹Please note that “Rivard” is only depicted in the bottom left figure of Fig. 7 since it specifically refers to a NN classifier with cosine distance [2].

²In our implementation, we utilize the Haar wavelet with 9 scales.

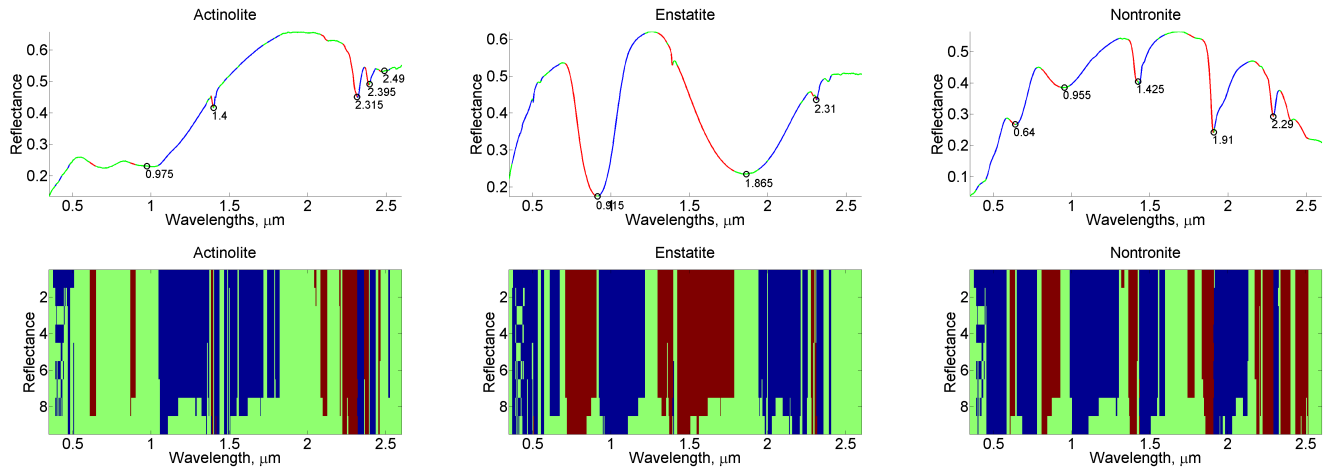


FIGURE 6. Semantic information extracted in some sample spectral curves based on MOG with 2 states. Top row: Sample spectral curves with extracted semantic information. Bottom row: corresponding state label array.

TABLE 1. Number of gaussian components mixed achieving highest classification accuracy with nearest neighbor classifier in conjunction with ℓ_1 distance.

Model \ DMP	DMP						
	70%	75%	80%	85%	90%	95%	100%
GMM	10	6	10	2	7-10	8-10	3-10
GMM+sign	10	6	4,5,10	2	7-9	2-10	2-10
MOG	4,7,8	5	7	3,5,7-9	5-7,9	6-10	3-10
MOG+sign	4	10	3	10	9	3-10	3-10

TABLE 2. Number of gaussian components mixed achieving highest classification accuracy with nearest neighbor classifier in conjunction with Euclidean distance.

Model \ DMP	DMP						
	70%	75%	80%	85%	90%	95%	100%
GMM	5,10	5	10	2	9	6-8,10	3-10
GMM+sign	6	3	4,9,10	5	6,9	2-10	2-10
MOG	7,8	5	7	3,5,7-9	5-7,9	6-10	3-10
MOG+sign	4,9	4,10	3	7-10	9	3-10	3-10

TABLE 3. number of gaussian components mixed achieving highest classification accuracy with nearest neighbor classifier in conjunction with cosine similarity.

Model \ DMP	DMP						
	70%	75%	80%	85%	90%	95%	100%
GMM	6	6	5,9,10	4	5,9	6,7,9,10	3-10
GMM+sign	9	6,9	3,5	5	5,7	2-10	2-10
MOG	4	4	7	8	5-7,9	6-10	3-10
MOG+sign	3,5	5	3	6-8	9	3-10	3-10

approach. We conjecture that this is due to the discrete nature of NHMC labels, which are not as easily leveraged in the SVM’s search for a separation boundary from support vectors.

We also attempted to implement the approach proposed in [3]. However, because of the lack of sufficient data for individual classes, we obtained several ill-conditioned covariance matrices when constructing multivariate GMMs.

Thus, we do not include the comparison with this approach in this paper.

C. NHMC PARAMETERS

We examine the impact of the number of states in the NHMC model on classifier performance. Specifically, we focus on a DMP of 85% and assess the classification accuracy of the proposed NHMC features with NN and SVM classifiers.

TABLE 4. Number of gaussian components mixed achieving highest classification accuracy with support vector machine classifier in conjunction with radial basis function.

Model \ DMP	70%	75%	80%	85%	90%	95%	100%
	GMM	8	10	4,7-10	3	4-10	3-10
GMM+sign	2	10	2	3,5,8,9	4-10	2-10	2-10
MOG	6	7	6,7,9,10	9	4	5-10	3-10
MOG+sign	3	3,5-7	4,5	3-10	3-10	3-10	3-10

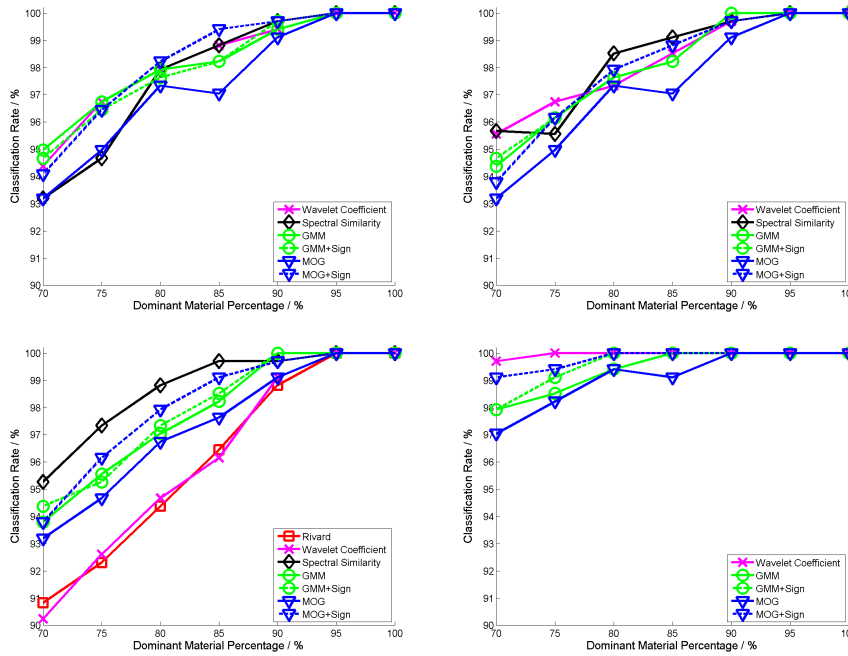


FIGURE 7. Classification rates of different NHMC modeling approaches and other relative classification approaches under different dominant material percentages. Top left: NN classifiers with ℓ_1 distance; top right: NN classifier with Euclidean distance; bottom left: NN classifier with cosine similarity; bottom right: SVM classifier. For NHMC models, the highest classification rate among the models tested is listed for each DMP value.

The number of states varies from 2 to 10 for GMM and from 3 to 10 for MOG. Figure 8 illustrates the trend of classification accuracy with an increasing number of mixed Gaussian components, considering different classifiers and similarity metrics.

The four figures reveal that MOG with the inclusion of wavelet coefficient signs demonstrates relatively consistent performance compared to other NHMC-based models. Furthermore, in terms of classification accuracy, the two MOG model configurations present two performance extremes: incorporating wavelet coefficient signs yields the highest classification performance, while MOG without signs produces the lowest. As previously mentioned, MOG combines the simplicity of a binary-state GMM with the ability to characterize spectral fluctuations in a multistate GMM. Without considering the wavelet coefficient signs, spectra with similar locations of fluctuations but varying magnitudes and orientations are likely to be assigned similar MOG label vectors. This is due to the binary-state GMM’s tendency

to assign the same labels to different fluctuations with different levels and orientations. However, the inclusion of Haar wavelet coefficient signs enhances the state labels by capturing more accurate spectral fluctuation orientation information.

V. CLASSIFICATION EXPERIMENTS AND RESULT ANALYSIS ON NATURAL IMAGES

In this section, we present the classification results obtained from a natural hyperspectral image. The experimental setup for the natural image is identical to that of the synthetic data, except for the manual introduction of mixtures into the hyperspectral data. Additionally, we compare the classification performance achieved using Haar wavelet (V-B) and Daubechies-4 wavelet (V-D). Both experiments evaluate the classification accuracy using NN and SVM classifiers. For the NN classifier, we focus on the results obtained using Euclidean (ℓ_2) distance and cosine similarity measure, as the performance with ℓ_1 distance is similar and space limitations

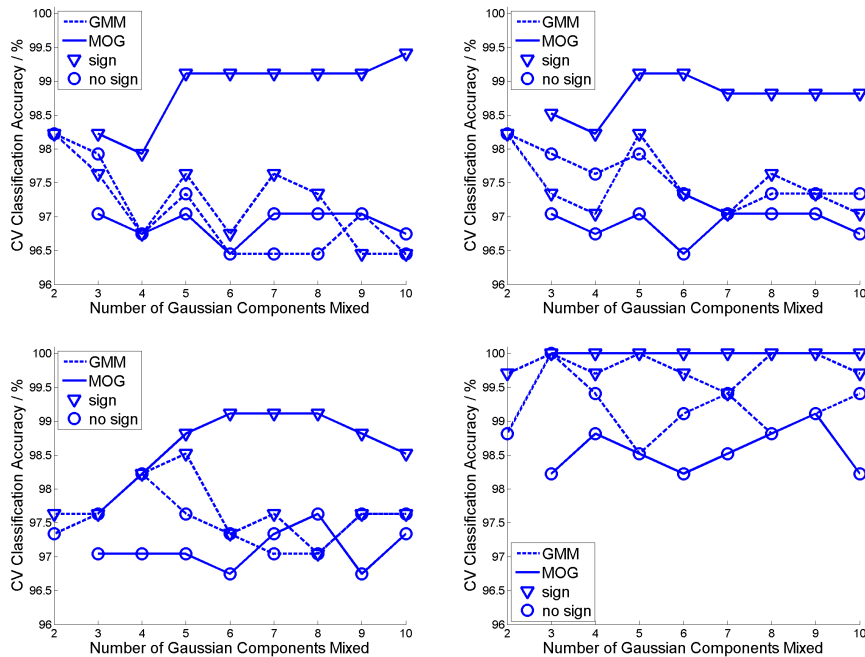


FIGURE 8. Classification rates of using different number of mixed Gaussian states with DMP of 85%. **Top left:** NN classifiers with ℓ_1 distance; **top right:** NN classifier with Euclidean distance; **bottom left:** NN classifier with cosine similarity; **bottom right:** SVM classifier.

prevent us from presenting all the results. For the SVM classifier, we employ RBF as the kernel and perform a grid search to determine the optimal parameters. Note that all experimental results are presented in numerical values and curves instead of images due to the exclusion of certain hyperspectral signatures from the analysis.

A. STUDY DATA

In this work, we perform experiment on three natural hyperspectral images. Below is a brief description of the three images we employ

- 1) Samson. This image includes 952×952 pixels, with 156 spectral bands covering the wavelengths from 401 nm to 889 nm in each pixel. In order to reduce the computational load, a region of 95×95 pixels is used in this experiment. All the pixels belong to three different categories, which are soil, tree, and water.
- 2) Jasper Ridge. There are 512×614 pixels in this image. Each pixel contains 224 spectral bands ranging from 380 nm to 2500 nm. A subimage of 100×100 pixels is used in this experiment. And after removing the channels 1–3, 108–112, 154–166 and 220–224 due to dense water vapor and atmospheric effects, there are 198 bands remained. There are four endmembers in this image: road, soil, water, and tree.
- 3) Urban. There are 307×307 pixels, with each pixel containing 210 wavelengths ranging from 400 nm to 2500 nm. After removing the bands 1–4, 76, 87, 101–111, 136–153 and 198–210 due to dense water vapor and atmospheric effects, we remain 162 bands.

There are six endmembers in the image, which are asphalt, grass, tree, roof, metal, and dirt.

B. FEATURE COMPARISON WITH HAAR WAVELET

Table 5 presents the classification results obtained from using the four NHMC-based models and the related approaches discussed in Section II-D. The conclusions drawn from these results differ somewhat from those obtained from the synthetic data in Section IV-B. Firstly, adding wavelet signs generally improves the classification rates for MOG, but it only enhances the performance of GMM in the Urban image. Additionally, MOG outperforms GMM only in the Samson image. In the other two images, GMM achieves better results than MOG, regardless of whether wavelet coefficient signs are included. Furthermore, the overall findings indicate that utilizing raw data yields the best classification performance, with the wavelet coefficient approach slightly trailing behind. Interestingly, in the Samson image, Rivard's method performs better when using both cosine measure and Euclidean distance, while the opposite is observed in the other two images.

C. NHMC PARAMETER ANALYSIS FOR EXPERIMENTAL RESULTS BASED ON HAAR WAVELET

We also examine the impact of the number of Gaussian components in the NHMC model on the classification performance of testing samples. Similar to Section IV-C, we evaluate the classification performance of all proposed NHMC features using NN and SVM classifiers, varying the number of states between 2 and 10 for GMM and 3 and 10

TABLE 5. Performance loss evaluation with haar wavelet.

Samson	GMM	MOG	GMM+sign	MOG+sign	Spectral Similarity	Rivard	Wavelet Coefficient
Cos	95.8061	95.9003	96.2632	96.3906	98.8172	96.6648	97.4377
Euc	96.2687	96.2687	96.3878	96.4515	98.1108	96.8947	97.5762
Jasper Ridge							
Cos	93.9850	93.4700	94.0475	93.9550	97.8000	92.8900	94.3800
Euc	94.0275	93.3450	94.0275	93.9825	96.7200	93.7450	94.8250
Urban							
Cos	87.1000	87.1000	87.6208	87.5500	93.7457	86.6750	89.1333
Euc	87.8708	87.8708	88.3500	88.3500	91.0542	87.6542	90.3042

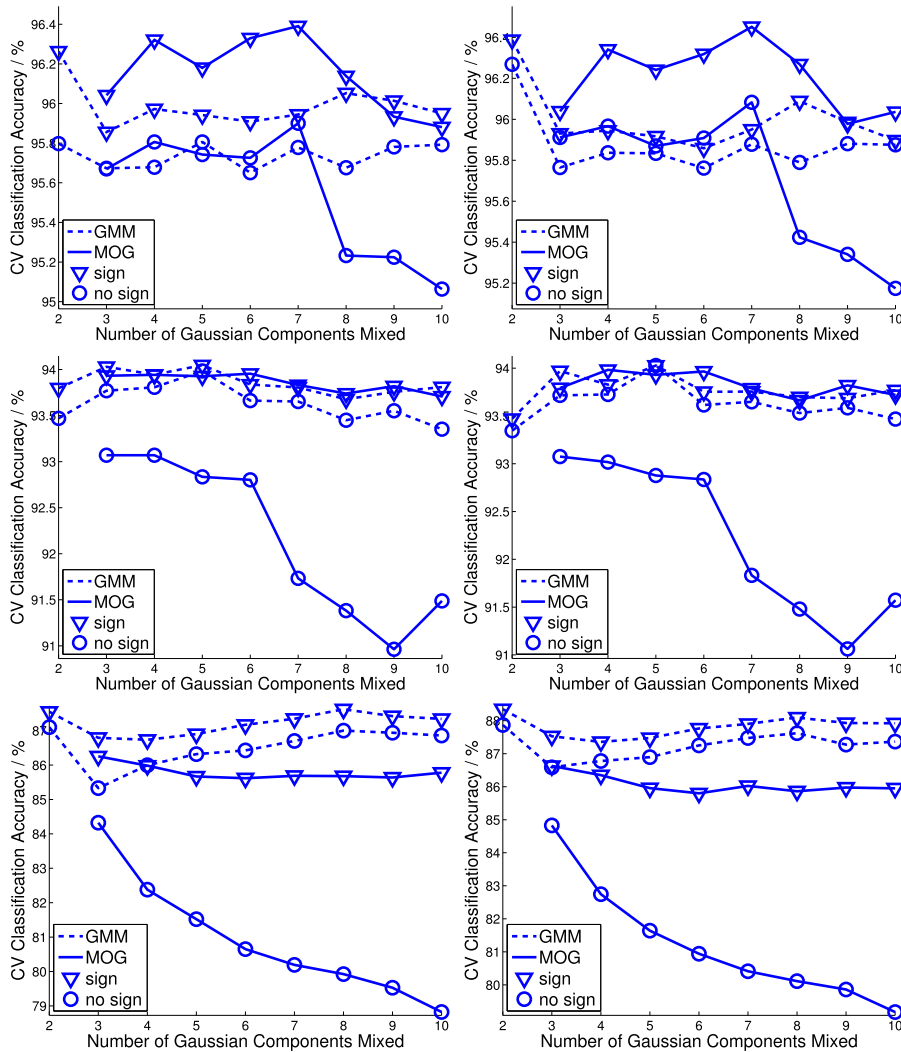


FIGURE 9. Classification rates for all three images of using different number of mixed Gaussian states with Haar wavelet. Top row: Samson; Second row: Jasper Ridge; Bottom row: Urban. Left row: NN classifier with cosine similarity; Right row: NN classifier with Euclidean distance.

for MOG. Figure 9 illustrates the trend of classification accuracy with the number of Gaussian components in each model. Consistent with the results discussed in Section IV-C, MOG without the inclusion of wavelet coefficient signs yields the poorest classification results across all cases in the Jasper Ridge and Urban images. In Samson, MOG without signs performs slightly better than GMM without signs in certain instances. Furthermore, except for MOG without signs, both

GMM with and without signs consistently produce similar results as the number of Gaussian components increases in the Jasper Ridge and Urban images. In those two images, the performance of all models, except MOG without signs, is quite comparable. However, in the Samson image, none of the four models' results appear to be stable. Overall, the inclusion of wavelet coefficient signs contributes to improved classification accuracy.

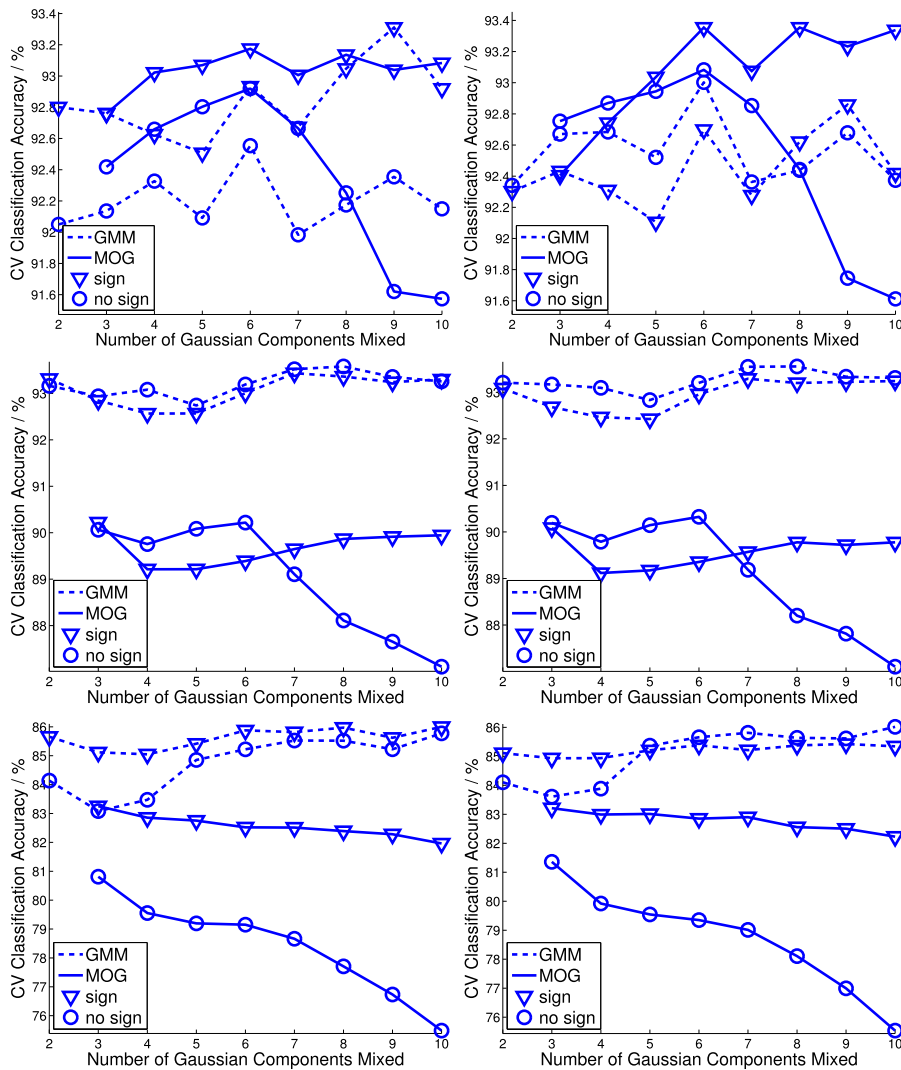


FIGURE 10. Classification rates for all three images of using different number of mixed Gaussian states with Daubechies-4 wavelet. Top row: Samson; Second row: Jasper Ridge; Bottom row: Urban. Left row: NN classifier with cosine similarity; Right row: NN classifier with Euclidean distance.

TABLE 6. Performance loss evaluation with Daubechies-4 wavelet.

	Samson	GMM	MOG	GMM+sign	MOG+sign	Spectral Similarity	Rivard	Wavelet Coefficient
Cos		92.5540	92.9197	93.3102	93.1745	98.8172	92.2354	95.2687
Euc		93.0027	93.0831	92.8587	93.3456	98.1108	91.8920	95.1745
Jasper Ridge								
Cos		93.5800	93.1625	93.3250	93.3125	97.8000	89.5825	93.6400
Euc		93.5550	93.2075	93.2900	93.0825	96.7200	89.9550	93.8650
Urban								
Cos		85.7750	84.1375	86.0000	85.6583	93.7457	83.6208	88.2708
Euc		86.0167	84.1042	85.4250	85.1125	91.0542	84.2375	88.9000

D. FEATURE COMPARISON WITH DAUBECHIES-4 WAVELET

Table 6 presents the Daubechies-4 version of Table 5. The results in Table 6 indicate that the inclusion of wavelet coefficient signs has a negative impact on classification accuracy. This can be attributed to the fact that the signs of Daubechies-4 wavelet coefficients do not convey the structural information of hyperspectral signals as effectively

as Haar wavelet coefficients, which indicate the direction of slope (rising or falling). When comparing GMM and MOG, the numbers in the table generally favor GMM, except for a few cases in Samson. Additionally, except for Samson, MOG without additional wavelet coefficient signs performs the poorest. The relative performance of our proposed models and the competing methods remains consistent with that shown in Table 5: raw data achieves the best performance,

the wavelet coefficient approach performs slightly worse, and our proposed models perform better than the wavelet coefficient approach but not as well as Rivard’s method. Importantly, the use of Daubechies-4 wavelet coefficients leads to a decrease in classification accuracy for all methods in the table, except when using raw data. Referring to the middle and bottom rows of Fig. 3 and Fig. 1, we can conclude that the semantic characterization of structural information in hyperspectral signatures using Haar wavelet coefficients contributes to improved classification performance.

E. NHMC PARAMETER ANALYSIS FOR EXPERIMENTAL RESULTS BASED ON DAUBECHIES-4 WAVELET

In this section, we present the Daubechies-4 wavelet version of NHMC parameter analysis, using the same setup as in Section V-C with Haar wavelet. The classification accuracy trends with varying numbers of mixed Gaussian components are displayed in Fig. 10. The results differ considerably from those obtained using Haar wavelet. First, the addition of wavelet coefficient signs does not consistently impact classification accuracy. In Samson, wavelet coefficient signs improve performance with cosine similarity but harm it with Euclidean distance. In Jasper Ridge, the signs generally decrease classification accuracy. However, in Urban, the signs have a relatively stable positive effect on classification rates. As mentioned in Section V-D, Daubechies-4 wavelet coefficients do not reflect the diagnostic information of hyperspectral signature structures. Therefore, the unstable influence of added wavelet coefficients is expected. Regarding the comparison between GMM and MOG, GMM consistently outperforms MOG in Jasper Ridge and Urban. However, in Samson, MOG performs slightly better than GMM.

VI. CONCLUSION

In this paper, we proposed a feature extraction scheme for hyperspectral signatures that preserves semantic information used in signature discrimination. Our approach utilizes statistical models for wavelet transform coefficients, capturing fluctuations and discontinuities in spectra. The statistical model enables informative segmentation of spectra. We evaluated our features using a simple hyperspectral classification system and compared them with three existing approaches: spectral matching, direct classification on wavelet coefficients, and feature computation based on wavelet coefficient sums. Our proposed features outperformed or matched the baselines, particularly for high DMP in synthetic data experiments. However, in real data experiments, our method generally performed better than the third approach but was outperformed by spectral matching and direct classification on wavelet coefficients. Comparing Haar wavelet and Daubechies-4 wavelet, Haar wavelet showed clear advantages, likely due to its ability to capture semantic structural information in hyperspectral signatures. In experiments with reduced DMP, the performance of all methods decreased, but MOG and GMM methods exhibited

a stronger reduction compared to other methods. This could be attributed to the difficulty of modeling signal classes with increased variability using binary features. However, the best combinations of NHMC features and NN classifiers still outperformed the baseline methods when DMP was sufficiently large. The reason could be the partial sacrifice of performance resulted from the light weight of NHMC. We acknowledge that validation on large-area images are valuable in hyperspectral image studies. Therefore, future work will focus on expanding the database size and exploring modifications to the feature extraction scheme and statistical models. One potential direction is exploring NHMC models based on nonzero-mean GMM, as zero-mean Gaussian mixture models may not accurately model the histogram of wavelet coefficients in some cases.

**APPENDIX
PROOF OF LEMMA 1**

By denoting $p_{a \rightarrow b}^{s,S \rightarrow Z} = p(Z_s = b | S_s = a)$ and using law of total probability, we can get $p(Z_s = b) = \sum_a p_{a \rightarrow b}^{s,S \rightarrow Z} p(S_s = a)$. From the $Z(S)$ map in eq. (6), we can infer that $p_{0 \rightarrow 0}^{s,S \rightarrow Z} = 1$ and $p_{i \rightarrow 1}^{s,S \rightarrow Z} = 1$. Therefore, it is easy to derive the conclusion in Lemma 1.

**APPENDIX
PROOF OF LEMMA 2**

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 shown in Fig. 11. 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

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

we appeal to the law of total probability to write

$$q_{b \rightarrow a, s} = \sum_{x=0}^{k-1} \sum_{y=0}^{k-1} p_{x \rightarrow a}^{s,S \rightarrow Z} p_{y \rightarrow x, s} p_{b \rightarrow y}^{s-1, Z \rightarrow S}. \quad (11)$$

From the $Z(S)$ map in equation (6), we can also infer that $p_{0 \rightarrow 1}^{s,S \rightarrow Z} = 0, p_{i \rightarrow 0}^{s,S \rightarrow Z} = 0, p_{0 \rightarrow 0}^{s-1, Z \rightarrow S} = 1, p_{0 \rightarrow i}^{s-1, Z \rightarrow S} = 0, p_{1 \rightarrow 0}^{s-1, Z \rightarrow S} = 0$, and

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

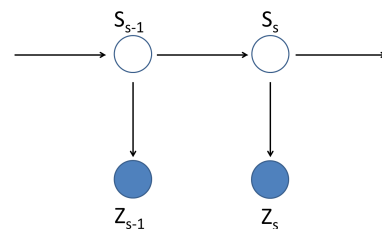
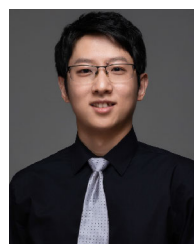


FIGURE 11. Directed graphic model for MOG and GMM state labels.

where $i = 1, \dots, k - 1$. After combining the equalities above with (11) for $a, b \in \{0, 1\}$, we can get the four elements in new matrices expressed in (7 – 10), proving the lemma.

REFERENCES

- [1] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.
- [2] B. Rivard, J. Feng, A. Gallie, and A. Sanchez-Azofeifa, "Continuous wavelets for the improved use of spectral libraries and hyperspectral data," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2850–2862, Jun. 2008.
- [3] S. Prasad, W. Li, J. E. Fowler, and L. M. Bruce, "Information fusion in the redundant-wavelet-transform domain for noise-robust hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 9, pp. 3474–3486, Sep. 2012.
- [4] X. Zhang, N. H. Younan, and C. G. O'Hara, "Wavelet domain statistical hyperspectral soil texture classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 615–618, Mar. 2005.
- [5] R. N. Clark, G. A. Swayze, K. E. Livo, R. F. Kokaly, S. J. Sutley, J. B. Dalton, R. R. McDougal, and C. A. Gent, "Imaging spectroscopy: Earth and planetary remote sensing with the USGS tetracorder and expert systems," *J. Geophys. Res., Planets*, vol. 108, no. E12, pp. 1–10, Dec. 2003.
- [6] M. Parente and M. F. Duarte, "A new semantic wavelet-based spectral representation," in *Proc. 5th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2013, pp. 1–4.
- [7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–15.
- [8] J. Gao, L. Jiao, X. Liu, L. Li, P. Chen, F. Liu, and S. Yang, "Multiscale dynamic curvelet scattering network," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, 2022, doi: [10.1109/TNNLS.2022.3223212](https://doi.org/10.1109/TNNLS.2022.3223212).
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [10] X. Liu, L. Jiao, L. Li, L. Cheng, F. Liu, S. Yang, and B. Hou, "Deep multiview union learning network for multisource image classification," *IEEE Trans. Cybern.*, vol. 52, no. 6, pp. 4534–4546, Jun. 2022.
- [11] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*. Fort Lauderdale, FL, USA: ML Research Press, 2017, pp. 1273–1282.
- [12] X. Liu, L. Li, F. Liu, B. Hou, S. Yang, and L. Jiao, "GAFnet: Group attention fusion network for PAN and MS image high-resolution classification," *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 10556–10569, Oct. 2022.
- [13] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2935–2947, Dec. 2018.
- [14] M. Liu, L. Jiao, X. Liu, L. Li, F. Liu, S. Yang, and X. Zhang, "Bio-inspired multi-scale contourlet attention networks," *IEEE Trans. Multimedia*, early access, 2023, doi: [10.1109/TMM.2023.3304448](https://doi.org/10.1109/TMM.2023.3304448).
- [15] N. Audebert, B. Le Saux, and S. Lefevre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 159–173, Jun. 2019.
- [16] N. Wambugu, Y. Chen, Z. Xiao, K. Tan, M. Wei, X. Liu, and J. Li, "Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks: A review," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, Dec. 2021, Art. no. 102603.
- [17] X. Li, B. Liu, K. Zhang, H. Chen, W. Cao, W. Liu, and D. Tao, "Multi-view learning for hyperspectral image classification: An overview," *Neurocomputing*, vol. 500, pp. 499–517, Aug. 2022.
- [18] Z. Zhang, Y. Ding, X. Zhao, L. Siye, N. Yang, Y. Cai, and Y. Zhan, "Multireceptive field: An adaptive path aggregation graph neural framework for hyperspectral image classification," *Expert Syst. Appl.*, vol. 217, May 2023, Art. no. 119508.
- [19] L. Jiao, X. Zhang, X. Liu, F. Liu, S. Yang, W. Ma, L. Li, P. Chen, Z. Feng, Y. Guo, X. Tang, B. Hou, X. Zhang, J. Bai, D. Quan, and J. Zhang, "Transformer meets remote sensing video detection and tracking: A comprehensive survey," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1–45, 2023.
- [20] L. Jiao et al., "Brain-inspired remote sensing interpretation: A comprehensive survey," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2992–3033, 2023.
- [21] D. He, Y. Zhong, and L. Zhang, "Spatiotemporal subpixel geographical evolution mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2198–2220, Apr. 2019.
- [22] D. He, Y. Zhong, and L. Zhang, "Spectral-spatial-temporal MAP-based sub-pixel mapping for land-cover change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1696–1717, Mar. 2020.
- [23] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, "Deep pyramidal residual networks for spectral-spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 740–754, Feb. 2019.
- [24] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2018.
- [25] X. Kang, S. Li, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.
- [26] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.
- [27] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 617–643, Mar. 1992.
- [28] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 7, pp. 710–732, Jul. 1992.
- [29] S. Feng, M. F. Duarte, and M. Parente, "Universality of wavelet-based non-homogeneous hidden Markov chain model features for hyperspectral signatures," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 19–27.
- [30] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Oct. 1989.
- [31] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2276–2291, Apr. 2013.
- [32] L. Shen and S. Jia, "Three-dimensional Gabor wavelets for pixel-based hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 5039–5046, Dec. 2011.
- [33] S. Mallat, *A Wavelet Tour of Signal Processing*. New York, NY, USA: Academic, 1999.
- [34] M. T. Orchard and K. Ramchandran, "An investigation of wavelet-based image coding using an entropy-constrained quantization framework," in *Proc. IEEE Data Compress. Conf.*, Oct. 1994, pp. 341–350.
- [35] B. Hapke, *Theory of Reflectance and Emittance Spectroscopy*. Cambridge, U.K.: Cambridge Univ. Press, 2012.



SIWEI FENG received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Massachusetts Amherst. He is currently an Assistant Professor with the School of Computer Science and Technology, Soochow University. He has published over 20 research papers in leading international journals and conferences. His research interests include machine learning and signal processing. He is a member of CCF.