IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 29, 2019, accepted September 2, 2019, date of publication September 11, 2019, date of current version September 24, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2940268*

# Hyperspectral Image Classification via Matching Absorption Features

**BAOFENG GUO**
School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China
e-mail: gbf@hdu.edu.cn

This work was supported by the National Natural Science Foundation of China under Grant 61375011.

**ABSTRACT** In this paper, we propose to extract spectral absorptions as the discriminative features to classify hyperspectral imagery. Different from previous researches that mainly take hyperspectral curves as high-dimensional inputs, we analyze hyperspectral data more from its physical and chemical origins. In the proposed approach, the discriminatory information, which is characterized by the observed materials' constituents, is extracted as a group of absorption features. First, the original hyperspectral spectra are transformed to a normalized spectra, in which a modified continuum removal algorithm is adopted to highlight all spectral valleys. Next, a standard peak detection method is applied to the continuum-removed spectra, and all effective absorptions are found as the candidate features. Then, to obtain the most informative absorptions to classification, a novel mutual-information based feature selection method is used to search for the key absorption spectra. Finally, we put forward a matching algorithm to classify the absorption features using the multi-label learning. To testify the proposed method, both laboratory and remotely sensed hyperspectral data are used to evaluate the classification performance. Experimental results show that the proposed method achieves competitive classification accuracy against the state-of-the-art methods, but with an advantage of more compact feature representation.

**INDEX TERMS** Hyperspectral imagery classification, absorption features, feature matching.

## I. INTRODUCTION

With the advance of remote sensing technology, hyperspectral imaging is becoming more and more important for earth observation. After success of NASA's Hyperion mission, the satellite-borne hyperspectral sensor has proved their capability to earth observation by providing critical ground and atmospheric information. Comparing to the conventional panchromatic imaging, hyperspectral remote sensing is based on the technology of imaging spectroscopy, dated back to more than one hundred years ago in the laboratory by physicists and chemists to identify materials according to their compositions. Actually, it can be considered as a sensor that combines two individual parts, namely imaging and spectroscopy, into a single system. Therefore hyperspectral sensors can simultaneously measure hundreds of contiguous spectral bands with a fine spectral resolution without complicated image registration. Comparing to the multispectral remote sensing, hyperspectral sensors can be seen as successors to the multispectral sensors. Specifically, the hyperspectral data sets, captured by one spectrometer, are generally composed of more than 100 spectral bands with relatively narrow bandwidths (5-10 nm). On the contrary, multispectral data sets, captured by several radiometers (for example the Landsat Thematic Mapper, LTM), are usually composed of about 5 to 10 bands of relatively large bandwidths (70-400 nm) [1], [2]. The increased number of bands and the narrowed bandwidth make it possible to reduce overlaps between classes, and give better capability to discriminate subtle spectral differences. In recent years, hyperspectral imaging has been found to be a powerful technique in numerous areas, such as mining, geology, agriculture, surveillance, ecology, oceanography, environmental monitoring, etc.. And the interest in applying hyperspectral imagery classification to earth observation is growing continuously [3], [4].

Hyperspectral imagery is typically collected as a data cube with spatial information in the usual $x - y$ coordinates, and spectral information represented in the another $z$ coordinate. Sampling each voxel from the $x - y$ plane, the classification is carried out by analyzing the electromagnetic reflectance as a function of the wavelength or band, i.e., $z$ axis. The higher spectral resolution in the $z$ axis

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

provides markedly potentials to object classification in earth observation, but a new range of challenges are also introduced by the hyperspectral imaging. First, the level of complexity for data processing is increased significantly comparing to the traditional panchromatic or the multispectral techniques. To work with more than 200 narrow spectral bands, it is necessary to reduce redundancy between spectral bands and the dimensionality reduction techniques are in great demand. Second, the need for data storage is rising dramatically for high-volume data. The facilities to accommodate the high-volume data cube and enough bandwidth for the space-earth telecommunication become decisive factors to make the mission of the hyperspectral-based earth observation success. Finally, the increased number of the spectral bands causes an expanded feature space, which makes the feature extraction or feature selection more difficult and more complicated. In this paper, we address the last challenge by investigating effective approaches for hyperspectral feature presentation and the corresponding classification algorithms.

To use hyperspectral data cubes effectively, the first priority is to extract appropriate features from the original high-dimensional data. Ideally, we expect that the extracted features contain enough discriminatory information and can characterize the intrinsic distinctness among different materials. This will help us to meet our main goal of identifying materials for earth observation. On the other hand, since atmospheric absorption, particles scattering, sensors noise and spatial interference from neighboring pixels are almost inevitable during measurement, we hope that the features should have a certain level of invariability and are robust to the aforementioned interference. In recent years, many studies have been carried out to address this problem. Typically used features are the complete spectra [5], the spectral bands from feature selection [6], various transformed features [7], spectral-spatial features [8], etc. Other newly developed techniques applied to hyperspectral feature extraction are kernel based methods [5], deep-learning based approaches [9], active learning based approach [10], etc.

Roughly speaking, the above hyperspectral feature extraction methods are developed by following three main technical routes, namely pattern analysis, signal processing, and machine learning. The methods that may be categorized as the pattern analysis route include the Principal Component Analysis (PCA) [7], [11], [12], Independent Component Analysis (LDA) [13], Linear Discriminant Analysis (LDA) [14], feature selection [6], etc. The methods based on the signal processing are the wavelet transform [15], Empirical Mode Decomposition (EMD) [16], Maximum Noise Fraction (MNF) [17], [18], Non-negative Matrix Factorization (NMF) [19], Intrinsic Mode Functions [20], etc. And the methods based on the machine learning are the Support Vector Machines (SVMs) [5], Convolutional Neural Network (CNN) [21]–[23], active learning [10], etc. The effectiveness of the these methods is mainly derived from their mathematical foundation and the applied optimization techniques. In this paper, apart from inheriting the solid basis from the classic pattern recognition theory, we are considering the hyperspectral structural feature extraction from a perspective of spectral absorption [24]–[27]. The idea is derived from the physical background associated with the hyperspectral imaging, and is different from the previous methods that are driven mainly by the pattern analysis theory.

The proposed method consists of the following steps. First, to avoid the effects from changing illumination the original high-dimensional hyperspectral curves are normalized into magnitude range of [0,1]. Then, a modified continuum removal algorithm is applied to the normalized spectra, leaving highlighted extrema (i.e., all maximal and minimal) at the spectra. Next, we use a peak detection algorithm to find all absorptions, which are served as the feature candidates. Then, we adopt a mutual-information based feature selection method to search for the most informative absorptions. The in-effect absorption feature vector is formed by labeling a binary vector with positive ones in the elements corresponding to the selected wavelength or bands. Finally, according to the nature of the binary absorption features, we propose a modified multi-label classification method.

The main contributions of the paper can be summarized as follows. First, we proposed an improved absorption valley detection approach, which consists of a modified continuum removal algorithm and an adaptive valley detection algorithm. Comparing to the traditional peak detection methods, the proposed approach can catch absorption valleys more accurately through the highlighted absorption peaks with more flexible magnitude comparing. Second, rather than relying on the uniqueness of the absorption valleys entirely, we investigated an alternative way for the absorption-based materials identification. The new scheme is carried out by comparing a group of selected absorption features collectively, where the absorptions are chosen by a mutual-information based algorithm. Finally, since in the new scheme an absorption feature could be associated with multiple material classes simultaneously, we proposed a binary feature-vector matching algorithm, which can handle each absorption feature individually.

The remainder of this paper is organized as follows. In Section II-III, several key algorithms regarding the proposed method are presented, including a continuum removal algorithm and a mutual-information based feature selection algorithm. Then a binary vector matching scheme is introduced in Section IV, which is used to classify the absorption features. Experiments are carried out to assess the performance of the proposed method, which are discussed in Section V. Finally, we end this paper with conclusions and a proposal for future work in Section VI.

## II. ABSORPTION FEATURES EXTRACTION

In this research, we first study the absorptions of seven materials at a laboratory environment. The investigated materials include 'Aluminum', 'Yellow polyester film', 'White polyester film', 'Mono-crystalline silicon', 'Titanium', 'Silicon dioxide', and 'Carbon fiber'. We use an Analytical

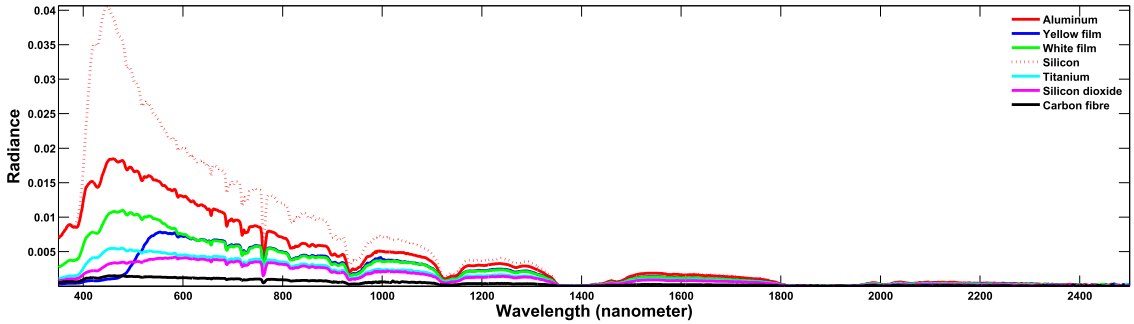



**FIGURE 1.** Samples of the ASD data set.

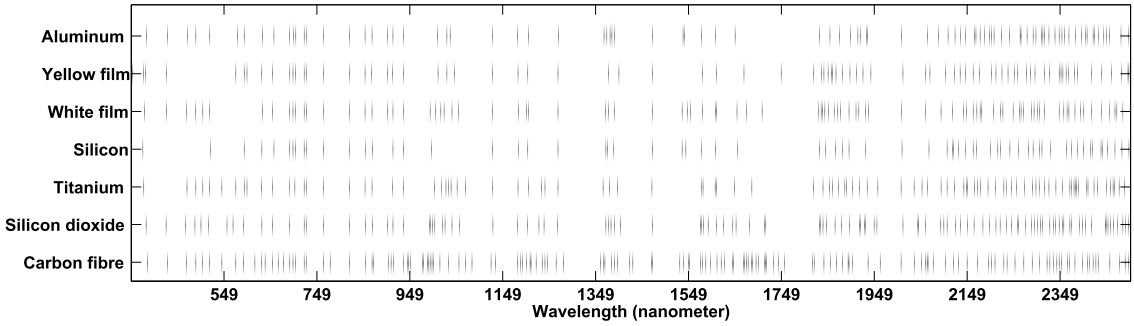



**FIGURE 2.** Detection results of absorption valleys.

Spectral Devices (ASD [28]) to recode the seven materials' spectra. Figure 1 shows the samples of the spectral curves from the seven studied materials, where the data are acquired by the author using the ASD portable spectrometer. Figure 2 shows the absorption valleys of the seven materials, where the detected absorption valleys are represented as black lines. These results are obtained by using a peak detection algorithm.

To use the detected absorptions for material identification, we can directly compare the absorptions of each of the materials against those of all other remaining materials. In this research, we use 'eigen-absorptions' to denote the absorptions that are different from all other materials that are involved in classification. We search for the 'eigen-absorptions' of each class based on their training samples. In details, the searching is carried out by a pair-wise comparison. For example, the first class's absorptions are compared against the second class's absorptions, and the different absorptions are remained and the identical absorptions are removed off. Such a searching is carried on by $N-1$ times ($N$ is the number of the classes), and during each round of the searching the $(N+1)$-th class's absorptions is applied for comparison. The resulting absorptions are therefore unique absorptions within the set of compared classes, and can be used for material identification. This strategy is straightforward and simple. But when the classification tasks become more complicated, for example a great amount of materials are involved for classification, applying this strategy to classification may no longer work. To avoid the aforementioned situation, we propose an improved hyperspectral absorption classification method. It is based on three foundations, namely the 'structural feature extraction', 'feature selection' and 'feature matching', which are discussed as follows.

To find the suitable absorption features, we are considering a structural feature extraction method consisting of two steps, including 1) Continuum removal and 2) Absorption detection. In the first step, we use continuum removal or envelop removal techniques to highlight all absorption valleys. In hyperspectral remote sensing, the continuum is an envelop that is just above the top of a spectral curve. By removing the continuum, we can normalize the reflectance or radiance spectral curve. This is helpful to compare different absorptions by providing a standard baseline. Considering the characteristics of the absorption features, we develop a continuum removal algorithm, which is presented at Algorithm 1. Figure 3 shows the result of Algorithm 1, where the solid line is the original spectral curve, the dashed line is its envelop and the dotted line is the continuum removed spectral curve. It is seen from Figure 3 that after continuum removal every absorptions are pulled up to the top baseline, and it becomes possible to compare absorption intensities for different valleys.

In the second step, we are using an adaptive absorption detection algorithm to avoid those less important absorptions. Directly using the original spectral curve, it is very difficult

**Algorithm 1** Continuum Removal

**Input:** Original spectral curve $\mathbf{x} = (X_1, X_2, \cdots, X_n)$, $n$ is the total number of bands

**Output:** Continuum removed curve $\mathbf{x}_c$

1: Calculating the first order of derivative as: $\mathbf{x}'$ =diff($\mathbf{x}$), diff($\mathbf{x}$) = $\{X_{m+1} - X_m\}$, $m = 1, 2, \cdots, n-1$
2: Calculating the second order of derivative as: $\mathbf{x}''$ =diff($\mathbf{x}'$), diff($\mathbf{x}'$) = $\{X'_{m+1} - X'_m\}$, $m = 1, 2, \cdots, n-2$
3: Finding the maxim values of $\mathbf{x}$, i.e., $\mathbf{x_m} = (X_m^1, X_m^1, \ldots, X_m^k)$, by solving $\mathbf{x}'' = 0$
4: Filling the start and the end values to $\mathbf{x_m}$, i.e., $\mathbf{x_m} = (X_1, X_m^1, X_m^1, \ldots, X_m^k, X_n)$
5: Using straight-line segments to connect each points of $\mathbf{x_m}$, i.e., the local spectra maximal, and getting the envelop $\mathbf{x}_e$
6: **if** $(\mathbf{x} - \mathbf{x}_e) < 0$ **then**
7: $\quad \mathbf{x}'_e \leftarrow \mathbf{x}$
8: **else**
9: $\quad \mathbf{x}'_e \leftarrow \mathbf{x}_e$
10: **end if**
11: Outputting the continuum removed spectral curve $\mathbf{x}_c = \mathbf{x}'_e / \mathbf{x}$

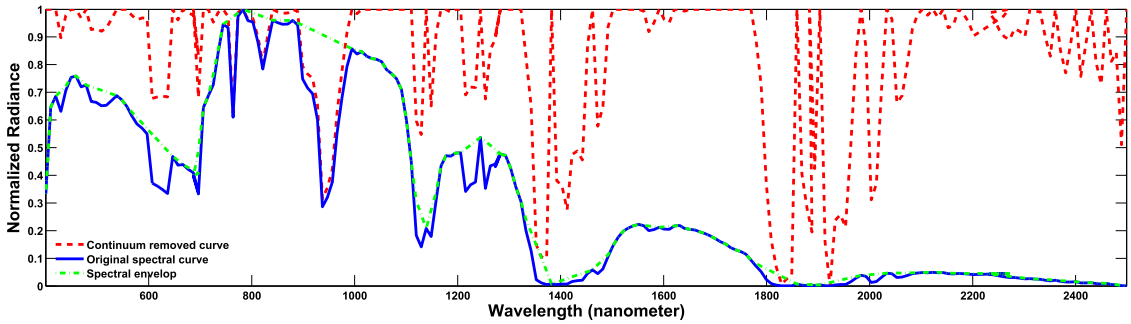



**FIGURE 3.** Results of continuum removal.

to measure the level of absorption intensity across different bands. It is seen that both in the ASD (see Figure 1) and the AVIRIS sensor, the levels of radiance drop significantly from the band of the visible light (VIS, 400-700 nm) to the band of the near-infrared light (NIR, 700-1000 nm). Therefore, without further processing majority of the absorptions located at the near infrared band will be omitted or be neglected. However, after removing the continuum we may compare individual absorption features from a *common* baseline (see Figure 3), and we are able to preclude the absorptions with the less absorption intensity. Specifically, by setting two thresholds we can prevent choosing the absorptions caused by sensors' noise or transmission interferences, whist preserving the absorptions that are important to materials identification. The details are depicted as the following equations.

$$A_c^i = \begin{cases} 1, & \text{if } (X_c^{i+1} - X_c^i) \geq \delta_l \text{ and } (X_c^{i-1} - X_c^i) \leq \delta_r, \\ & \qquad i = 1, 2, \cdots, L \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $\mathbf{x}_c = \left(X_c^1, X_c^2, \cdots, X_c^L\right)$ is a $L$-dimensional continuum removed spectral curve, and the binary scalar $A_c^i$ indicates whether an effective absorption feature is found at the $i$-th band. Since an absorption valley has two shoulders, two parameters $\delta_l$ and $\delta_r$ have been used as thresholds to control above what levels of intensity an absorption can be considered as an important feature. In our method, the parameters $\delta_l$ and $\delta_r$ are set up adaptively as in Equation (2). Comparing with the conventional methods, the proposed method has an advantage of adaptivity, achieved by using a common baseline for intensity comparison and two adaptive thresholds.

$$\begin{aligned} \delta_l &= (X_c^{i-1} - X_c^i)/2 \\ \delta_r &= (X_c^{i+1} - X_c^i)/2 \end{aligned} \tag{2}$$

## III. ABSORPTION FEATURES SELECTION

In the absorption features extraction, we were using a continuum removal algorithm and an adaptive absorption detection algorithm to get a series of absorptions that are qualified candidates in the sense of the level of the absorption intensity. But for the purpose of materials identification or objects classification, the high level of the absorption intensity, or the magnitudes of the features, is not sufficient to achieve a better performance. Hence to search for the most informative absorptions from all the possible candidates, we investigate a novel mutual-information based feature selection method [29].

The mutual information is a concept originated from the area of Information Theory, and is used to measure the mutual independence between two random variables. For the case of continual random variables, it is defined as follows:

$$I(A, B) = -\int_{a \in A} \int_{b \in B} p(a, b) \log \frac{p(a, b)}{p(a)\, p(b)}\, da\, db \tag{3}$$

where $p(a)$, $p(b)$, and $p(a, b)$ are the marginal and the joint probability density functions of the continual random variables $A$ and $B$, respectively. In the case of discrete random variables, the mutual information can also be defined as:

$$I(A, B) = -\sum_{a\in A}\sum_{b\in B} p(a, b)\log\frac{p(a, b)}{p(a)\,p(b)}. \tag{4}$$

Previous research has shown that the mutual information is related with the Bayesian classification error, which is usually considered as a golden standard in many classification applications, by the following inequality:

$$\frac{H(B) - I(A, B) - C_1}{C_2} \le P_e. \tag{5}$$

where $P_e$ is the Bayesian classification error, and $C_1 = H(P_e)$ and $C_2 = \log(l-1)$ are two constants. $H(B)$ is the entropy of the random variable $B$:

$$H(B) = -\sum_{b\in B} p(b)\log p(b). \tag{6}$$

From (5), it is seen that the classification error $P_e$ is lower bounded by $I(A, B)$ and $H(B)$. In a specific remote sensing application, given $B$ be a random variable standing for the ground truth, $H(B)$ is a fixed number due to the same prior distribution. Therefore, maximization of mutual information can optimize the lower bound of the classification error. In this sense, using mutual information for feature selection is actually based on the criterion of classification accuracy.

The above arguments pave the way of our proposal, namely using the mutual information as the objective function to search for the most discriminative absorptions for hyperspectral classification. To accomplish the proposed mutual-information based absorption selection, we have to solve one major problem in the first place, i.e., how to calculate a multidimensional mutual information. Formally, the mutual information between a multidimensional random vector, i.e, the multiple absorption candidates, and a one-dimensional random variable, i.e., the ground truth, can be written as:

$$I(\mathbf{x}, Y) = I\left((X_1, X_2, \ldots, X_M), Y\right). \tag{7}$$

where $\mathbf{x} = (X_1, X_2, \ldots, X_L)$ is a random vector representing multiple features, and $Y$ is a random variable denoting the ground truth.

In this paper, we extend a multidimensional mutual information decomposition scheme put forward by our previous research [29]. Specifically, we demonstrate that a multidimensional mutual information can be approximated by the following decomposition formula.

$$I(\mathbf{a}, B) \approx \sum_i I\left(A_i, B\right) - \sum_i \sum_{j>i} I\left(A_i, A_j\right) + \sum_i \sum_{j>i} I\left(A_i, A_j|B\right). \tag{8}$$

where $\mathbf{a} = (A_1, A_2, \ldots, A_L)$ is a multidimensional absorption feature vector with each of the components $A_i$ ($i = 1, 2, \ldots, L$) standing for one selected absorption feature.

Based on (8), we can select the informative absorption features by maximizing $I(\mathbf{a}, B)$ step by step. In details, i.e., by maximizing the following cost function:

$$A_1^0 = \underset{A_i, i=1,2,\ldots L}{\operatorname{argmax}}\ I\left(A_i, B\right), \tag{9}$$

where $A_1^0$ is the first selected absorption feature.

By the same way, the second absorption feature is chosen as the following maximization:

$$A_2^0 = \operatorname{argmax}_{A_i|A_i\neq A_1^0, i=1,2,\ldots L}\Big[I\left(A_i, B\right) - I\left(A_i, A_1^0\right) + I\left(A_i, A_1^0|B\right)\Big]. \tag{10}$$

After repeating the above maximization, we get a series of informative absorption features, such as follows:

$$A_n^0 = \operatorname{argmax}_{A_i|A_i\neq A_j^0,, i=1,2,\ldots L}\Big[I\left(A_i, B\right) - \sum\nolimits_j I\left(A_i, A_j^0\right) + \sum\nolimits_j I\left(A_i, A_j^0|B\right)\Big], \tag{11}$$

where $A_j^0, j = 1, 2, \cdots, n-1$ denote the absorption features that are already selected.

## IV. MATCHING OF ABSORPTION FEATURE

Through the above structural feature extraction and selection, the resulting absorption features are encoded as a binary vector or a bit array, in which each component of value one means selected absorption feature and the components of value zero stand for absent absorption. As mentioned before, we do not require that the absorption features are unique among all classes in the revised version of the proposed approach. Thus, it becomes possible that an absorption feature is associated with multiple material classes simultaneously. For example in the AVIRIS 92AV3C data set, it is found that the absorption feature at the fourth band is an effective absorption feature for all 16 classes. The absorption feature at the band 170 contributes to four different classes. In view of pattern recognition theory, this is a classification problem where multiple target labels are assigned to one instance or feature. Considering the fact, i.e., one absorption feature may correspond to more than one class, we propose the following binary feature-vector matching algorithm, which can manage each absorption feature individually.

First, we decompose the overall matching into $N$ independent binary classification problems, where $N$ is the total number of the classes or labels. Next, we adopt a binary relevance method depicted as follows.

Given a training set as a group of binary vectors $\mathbf{x}_i = \left(X_i^1, X_i^2, \cdots, X_i^L\right)$ and $X_i^j \in \{0, 1\}$ representing a binary absorption feature at the $j$-th band for the $i$-th instance or sample, we use $\mathbf{y}_i \in \{1, 2, \ldots, N\}, i = 1, 2, \ldots, M$ to denote the class label of the absorption feature vector $\mathbf{x}_i$ ($L$ is the total number of the bands and $M$ is the total number of the training samples). If we treat each band $X^j$ individually, we can formulate the following multi-label learning data set:

$$D_k = \{(x_i^j, \phi(Y_i^j, k)), |1 \le i \le M\}, \quad k = 1, 2, \cdots, N \tag{12}$$

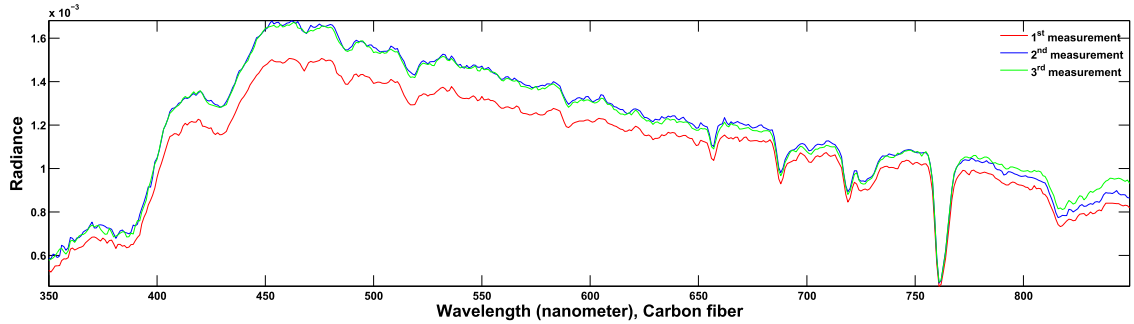



**FIGURE 4.** Three measurements of Carbon fiber's spectral radiance curves.

where $x_i^j \in R$ is the spectral radiance value at the $j$-th band of the $i$-th sample,

$$\phi(Y^j, k) = \begin{cases} +1, & \text{if } k \in Y^j \\ -1, & \text{otherwise,} \end{cases} \tag{13}$$

and

$$Y^j = \cup \mathbf{y}_i, \quad \text{if } X_i^j = 1,\ i = 1, 2, \cdots, M. \tag{14}$$

In (14), $Y^j$ is a subset of the class-label set $\{1, 2, \ldots, N\}$, associated with a scenario where the $j$-th band of the $i$-th sample is found to be positive to an effective absorption. In the new set of training samples $D_k$, for the relevant label $k \in Y^j$, the reflectance value $x_i^j$ is regarded as one positive example because an absorption feature is found at the $j$-th band, and vice verse. By customizing training sets as $D_k, (k = 1, 2, \cdots, N)$ and following the general procedure of the binary relevance method [30], $N$ binary classifiers $f_k$ can be trained.

Prediction of an unseen instance $\mathbf{x} = \left(X^1, X^2, \cdots, X^L\right)$ is carried out by two steps. First, we query each absorption feature, i.e., $X^j$, $j = 1, 2, \cdots, L$ and its radiance value $x^j$ individually, by the trained binary classifier $f_k$. Next we obtain the final prediction result of the feature $j$ by combining the relevant results from the output of $f_k$, such as follows:

$$R^j = \{k \mid f_k(x^j) > 0,\ 1 \le k \le N\}. \tag{15}$$

where $R^j \subseteq \{1, 2, \cdots, N\}$ is the predicted multi-label set based on the feature at the $j$-th band individually. Combining the predictions from all bands $j \in \{1, 2, \cdots, L\}$, the overall result is finally reached by a majority voting.

## V. EXPERIMENTS AND RESULTS

In order to evaluate the classification performance of the proposed method, we carry out experiments based on two hyperspectral data sets, corresponding to a hand-set field test (see Section V-A) and an air-borne remote sensing test (see Section V-B), respectively.

### A. RESULTS OF LABORATORY DATA

The first data set is acquired by using the ASD (Analytical Spectral Devices [28]) sensor on a controlled laboratory environment. Seven materials, including include 'Aluminum', 'Yellow polyester film', 'White polyester film', 'Mono-crystalline silicon', 'Titanium', 'Silicon dioxide', and 'Carbon fiber', are studied. In this data set, each sample is measured in the wavelengths range from the visible light band to the near-infrared light band, i.e., from 350 nm to 2,500 nm. The spectral resolutions is 1 nm and an overall 2,151 measured bands are given for each of the samples. For the studied materials, a white board is also used to give a reference of all reflectance, and three times of measurements are carried out to obtain three samples for each material.

Figure 4 shows three spectral curves for the material 'Carbon fiber', which are measured at three different times. It is seen from Figure 4 that the magnitudes of the 'Carbon fiber's spectral radiance are varying with different measurement times. This could be caused by the changing intensity of the illumination or the random shifting by the sensor's gain. But it is observed that the positions of the absorption valleys remain quite stable, with no regard to the changing illumination. In contrast to the volatile radiance magnitudes, this finding exhibits the superiority of using the absorptions as the features for material identification and object classification.

According to the proposed method introduced in IV, for each of the materials, one sample is used as the training data to obtain the necessary 'eigen-absorption'. The remaining two samples are used as the testing data to assess the classification accuracy. The proposed method is compared with two classical hyperspectral classification method, i.e., the nearest neighbor ($k$-NN) method [31], and the spectral angle mapping (SAM) method [32]. Because in this test there is only one training sample from each of the classes, some modern learning methods, such as the support vector machines (SVMs), are not suitable and are not adopted as the benchmark approaches.

It is seen from Table 1 that in this relatively simple hyperspectral classification task, the traditional $k$-NN method and the SAM method achieved similar results to the proposed method at the six classes from the all seven classes. But for the class 'Silicon dioxide', both the $k$-NN method and the SAM method made a wrong identification, which lets their overall classification accuracies drop to 92.86%. Meanwhile, the overall classification accuracy of the proposed method

**TABLE 1. Comparison of classification accuracy, ASD data set, 1 training sample and 2 testing samples for each of the classes.**

| Class \ Accuracy (%) | Methods | | |
|---|---|---|---|
| | $k$-NN | SAM | The proposed |
| 1. Aluminum | **100** | **100** | **100** |
| 2. Polyester film (Yellow) | **100** | **100** | **100** |
| 3. Polyester film (White) | **100** | **100** | **100** |
| 4. Silicon | **100** | **100** | **100** |
| 5. Titanium | **100** | **100** | **100** |
| 6. Silicon dioxide | 50 | 50 | **100** |
| 7. Carbon fiber | **100** | **100** | **100** |
| Overall accuracy (%) | 92.86 | 92.86 | **100** |

still remains as 100%, making it goes beyond the two traditional competitors. Besides this controlled laboratory experiment, we also evaluate the proposed method by an air-borne remotely sensed data set, which is discussed in the next section.

### B. RESULTS OF REMOTELY SENSED DATA

For the purpose of hyperspectral earth observation, we also test the proposed method on AVIRIS (airborne visible/infrared imaging spectrometer) 92AV3C data set, which is a remotely sensed hyperspectral data set [33]. The AVIRIS hyperspectral sensor is developed by the Jet Propulsion Laboratory (JPL) for Earth remote sensing. It has been mounted on various aircraft platforms and can capture hyperspectral data at approximately 4-20 km height above sea level. The AVIRIS sensor provides 224 contiguous spectral channels at 10 nanometer intervals across the electromagnetic spectrum, covering a wavelength range from 400 nanometers to 2500 nanometers. The studied 92AV3C data set is acquired by the AVIRIS sensor in 1992 over the test site of Indian Pine in northwestern Indiana, USA. The data cube has a 145-pixel swath and 145-pixel scan length, giving a $145 \times 145 \times 224$ hyperspectral data cube. At the observed scene, each pixel is labeled as belonging to one of 16 classes of vegetation, including Alfalfa, Corn, Soybean, Wheat, etc. Both the data set and the reference map that indicates partial ground truth can be downloaded from Website [34].

In the test, pixels of all 16 classes vegetation are used for evaluation. Apart from the aforementioned two conventional hyperspectral classification methods, two machine-learning based approaches, including the classical support vector machine (SVM) [5] and the newly developed deep learning algorithm [9] (Transfer Learning method, TL), are also adopted for comparison. Both the SVM and TL methods are the machine-learning based approaches, so about 3% pixels from each class (or at least one sample if 3% samples is less than 1) were randomly chosen as the training data set. The training data are used to validate the models of the machine-learning based approaches, such as the optimization of the parameters of the SVM and TL classifiers. The training data are also used to help the mutual-information based feature selection. For classification accuracy assessment, the remaining 97% samples are formed as the testing set, on which the classification performance was assessed. Comparing to many previous researches that use the bigger size of training data (such as 20% 50%), we use only 3% training samples in this research. This is useful to demonstrate the advantage or capability regarding the proposed approach in the application scenarios where training samples are difficult to obtain.

During the training, the kernel function of the SVM is chosen as a Gaussian Radial Basis Function, i.e., $G(\mathbf{a}, \mathbf{b}) = \exp(-\frac{\|\mathbf{a}-\mathbf{b}\|^2}{\sigma})$. The penalty parameter $C$ is tested from $10^{-2}$ and $10^4$, and the kernel width $\gamma$ is tested between $10^{-2}$ and $10^3$ using a two folds validation procedure based on only the training data. The kernel width $\sigma = 15$ and penalty $C = 110$ were finally found as the optimal values for this training data set. The transfer learning (TL) [9], [10] is based on Convolutional Neural Networks [21]. The network used is the CNN-based VGGNet-VD16 (see website [35]) and this net is pre-trained by the visual data set ImageNet [36]. We use a transfer learning strategy that is adopted from Shao *et al.* [21].

Table 2 shows the performance of the proposed method against four competitive methods. It can be seen that the overall accuracy of the proposed method is 67.99%, which is very close to the state of the art methods, including the SVM's 68.03% and TL's 68.26%. It is significantly higher than those of the traditional methods, such as the $k$-NN and the SAM method. As for the two machine-learning based methods, the TL's result is similar to that of the SVM. This result is probably due to that fact that only a small amount of training samples (3%) are available for training in this special experiment. The advantages and potentials given by the deep learning approaches have not been exhibited sufficiently, because they are developed to accommodate and have to be supported by big data. Similar conclusions are also be drawn by comparing the Cohen's kappa coefficients of each of the methods, where the the proposed method has a kappa coefficients of 0.62, being very close or equal to TL's 0.63 and SVM's 0.62. From Table 2, it is also noted that the machine-learning based methods and the proposed method are biased to the individual classes that have more training samples. On the contrary, the traditional methods, such as the $k$-NN and the SAM method suffers less this bias. This result reveals one of the limitations of the learning-based methods. We also compare the proposed method with the method using the conventional peak detection method (CPD). It is shown that the performance of the CPD method is lower than the proposed method, with overall accuracy of 62.71% and Kappa coefficient of 0.59. This justifies the using the modified peak detection algorithm in the proposed method.

To further justify the above conclusions, we calculate the confusion matrices for each of the classification methods. To save space, we only list two of the confusion matrices in Table (3) and Table (4), corresponding to the SVM method and the proposed method, respectively. It is seen that the confusion matrix of the proposed method is very close to the SVM's confusion matrix, coinciding with the comparison results of the overall accuracies shown in Table 2.

**TABLE 2.** Classification results on AVIRIS 92AV3 dataset, 3% training set.

| Class \ Performance | Pixels in Testing | Accuracy (%) of Methods | | | | | |
|---|---|---|---|---|---|---|---|
| | | $k$-NN | SAM | SVM | TL | CPD | Proposed |
| 1. Alfalfa | 45 | 12.87 | 12.15 | 0 | 0 | 0 | 0 |
| 2. Corn/notill | 1385 | 52.55 | 56.66 | 74.16 | 74.35 | 65.08 | 69.98 |
| 3. Corn/min | 805 | 21.58 | 29.73 | 74.41 | 81.44 | 62.55 | 65.95 |
| 4. Corn | 230 | 12.23 | 10.11 | 46.25 | 55.95 | 31.94 | 41.46 |
| 5. Grass/Pasture | 469 | 42.5 | 33.33 | 91.47 | 93.07 | 79.90 | 89.11 |
| 6. Grass/trees | 708 | 43.75 | 65.89 | 78.58 | 78.98 | 76.13 | 76.66 |
| 7. Grass/Pasture/Mowed | 27 | 34.25 | 40.32 | 0 | 0 | 0 | 0 |
| 8. Hay/windrowed | 464 | 78.35 | 75.26 | 85.87 | 84.8 | 84.23 | 86.93 |
| 9. Oats | 19 | 2.431 | 3.008 | 0 | 0 | 0 | 0 |
| 10. Soybeans/notill | 943 | 28.25 | 31.19 | 60.83 | 61.52 | 55.70 | 61.18 |
| 11. Soybeans/min | 2381 | 50 | 50.47 | 55.94 | 55.32 | 46.37 | 55.8 |
| 12. Soybean/clean | 575 | 14.13 | 12.31 | 75.29 | 79.53 | 68.04 | 72.22 |
| 13. Wheat | 199 | 28.15 | 25.87 | 90 | 0 | 82.72 | 92.56 |
| 14. Woods | 1227 | 69.1 | 70.57 | 86.08 | 86.39 | 82.87 | 85.89 |
| 15. Building/Grass/Tree/Drives | 374 | 3.371 | 6.25 | 37.14 | 37.62 | 44.63 | 51.65 |
| 16. Stone/steel towers | 90 | 100 | 98.72 | 100 | 100 | 91.48 | 98.15 |
| Overall accuracy (%) | | 37.88 | 41.57 | 68.03 | 68.26 | 62.71 | 67.99 |
| Kappa coefficient | | 0.30 | 0.34 | 0.62 | 0.63 | 0.59 | 0.62 |

**TABLE 3.** Confusion matrix of the SVM method; Corn1: Corn-notill, Corn2: Corn-mintill, Corn3: Corn, Grass1: Grass-pasture, Grass2: Grass-trees, Grass3: Grass-pasture-mowed, Soybean1: Soybean-notill, Soybean2: Soybean-mintill, Soybean3: Soybean-clean, Hay: Hay-windrowed, Building: Buildings-Grass-Trees-Drives, Stone: Stone-Steel-Towers.

| | Alfalfa | Corn1 | Corn2 | Corn3 | Grass1 | Grass2 | Grass3 | Hay | Oats | Soybean1 | Soybean2 | Soybean3 | Wheat | Woods | Building | Stone |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Alfalfa | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 44 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Corn1 | 0 | 818 | 38 | 4 | 1 | 4 | 0 | 1 | 0 | 94 | 410 | 9 | 0 | 0 | 6 | 0 |
| Corn2 | 0 | 78 | 346 | 0 | 0 | 1 | 0 | 0 | 0 | 14 | 359 | 5 | 0 | 0 | 2 | 0 |
| Corn3 | 0 | 80 | 6 | 37 | 0 | 13 | 0 | 1 | 0 | 6 | 50 | 6 | 0 | 0 | 31 | 0 |
| Grass1 | 0 | 0 | 0 | 4 | 311 | 72 | 0 | 11 | 0 | 4 | 12 | 4 | 0 | 48 | 3 | 0 |
| Grass2 | 0 | 0 | 0 | 11 | 0 | 686 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 6 | 0 |
| Grass3 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 17 | 0 | 0 | 1 | 6 | 0 | 0 | 0 | 0 |
| Hay | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 462 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Oats | 0 | 0 | 0 | 0 | 0 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Soybean1 | 0 | 6 | 7 | 6 | 0 | 1 | 0 | 0 | 0 | 323 | 592 | 8 | 0 | 0 | 0 | 0 |
| Soybean2 | 0 | 79 | 60 | 11 | 1 | 9 | 0 | 2 | 0 | 27 | 2170 | 18 | 0 | 0 | 4 | 0 |
| Soybean3 | 0 | 39 | 7 | 3 | 0 | 1 | 0 | 0 | 0 | 37 | 282 | 195 | 0 | 0 | 11 | 0 |
| Wheat | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 9 | 0 | 187 | 0 |
| Woods | 0 | 0 | 0 | 0 | 13 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1193 | 14 | 0 |
| Building | 0 | 0 | 0 | 2 | 6 | 59 | 0 | 0 | 0 | 9 | 2 | 0 | 0 | 140 | 156 | 0 |
| Stone | 0 | 3 | 1 | 0 | 5 | 0 | 0 | 0 | 0 | 17 | 0 | 7 | 0 | 0 | 0 | 57 |

**TABLE 4.** Confusion matrix of the proposed method; for the detailed name of each class, see Table (3).

| | Alfalfa | Corn1 | Corn2 | Corn3 | Grass1 | Grass2 | Grass3 | Hay | Oats | Soybean1 | Soybean2 | Soybean3 | Wheat | Woods | Building | Stone |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Alfalfa | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 43 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Corn1 | 0 | 753 | 59 | 6 | 1 | 5 | 0 | 1 | 0 | 87 | 460 | 10 | 0 | 0 | 3 | 0 |
| Corn2 | 0 | 88 | 339 | 1 | 0 | 2 | 0 | 0 | 0 | 13 | 355 | 6 | 0 | 0 | 1 | 0 |
| Corn3 | 0 | 85 | 5 | 34 | 0 | 17 | 0 | 1 | 0 | 6 | 53 | 8 | 0 | 0 | 21 | 0 |
| Grass1 | 0 | 0 | 1 | 4 | 311 | 69 | 0 | 10 | 0 | 2 | 10 | 6 | 0 | 51 | 5 | 0 |
| Grass2 | 0 | 0 | 0 | 12 | 0 | 683 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 6 | 0 |
| Grass3 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 14 | 0 | 0 | 3 | 5 | 0 | 0 | 0 | 0 |
| Hay | 0 | 0 | 0 | 2 | 3 | 0 | 0 | 459 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Oats | 0 | 0 | 0 | 0 | 0 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Soybean1 | 0 | 7 | 18 | 4 | 1 | 1 | 0 | 0 | 0 | 331 | 573 | 8 | 0 | 0 | 0 | 0 |
| Soybean2 | 0 | 79 | 70 | 14 | 2 | 9 | 0 | 0 | 0 | 37 | 2136 | 30 | 0 | 0 | 4 | 0 |
| Soybean3 | 0 | 60 | 20 | 1 | 0 | 1 | 0 | 0 | 0 | 40 | 236 | 208 | 0 | 0 | 8 | 1 |
| Wheat | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 1 | 0 | 112 | 0 | 72 | 0 |
| Woods | 0 | 0 | 0 | 0 | 7 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 1199 | 12 | 0 |
| Building | 0 | 0 | 2 | 4 | 9 | 65 | 0 | 0 | 0 | 7 | 1 | 0 | 6 | 139 | 141 | 0 |
| Stone | 0 | 4 | 0 | 0 | 9 | 0 | 0 | 0 | 0 | 18 | 0 | 6 | 0 | 0 | 0 | 53 |

For visual observing, we also illustrate the classification maps of the $k$-NN, the SAM, the SVM, the TL method and the proposed method in Figure 6. Comparing with the ground truth map (Figure 5(a)), the distribution map of the training samples (Figure 5(b)) and the distribution map of the testing samples (Figure 5(c)), we may examine the

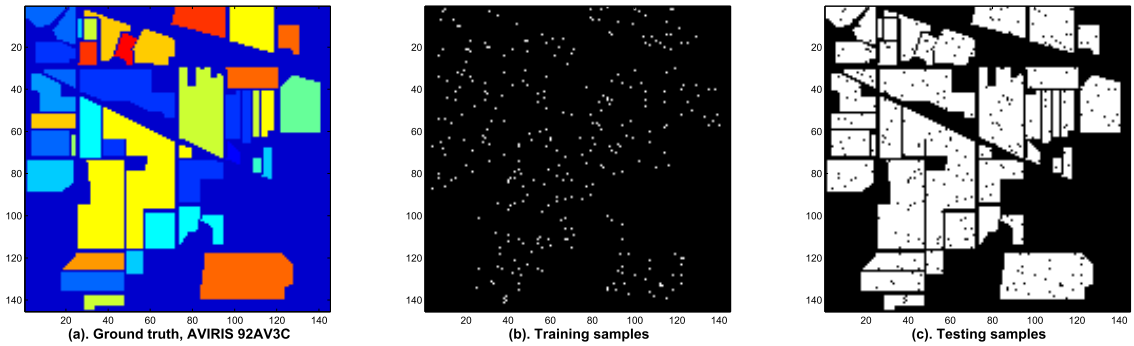



**FIGURE 5.** Ground truth (a), distributions of training data (b) and testing data (c), AVIRIS 92AV3C.

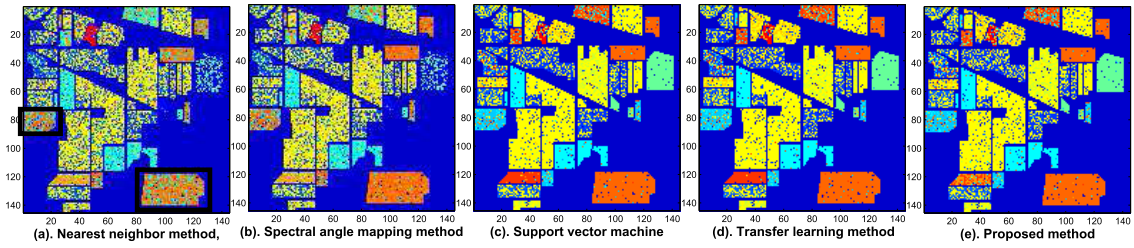



**FIGURE 6.** Comparison of classification maps.

**TABLE 5.** Comparison of dimensionality for different methods' feature-vectors, ASD data set.

| Class \ Dimension | Methods | | |
|---|---|---|---|
| | $k$-NN | SAM | The proposed |
| 1. Aluminum | 2151 | 2151 | **89** |
| 2. Polyester film (Yellow) | 2151 | 2151 | **79** |
| 3. Polyester film (White) | 2151 | 2151 | **90** |
| 4. Silicon | 2151 | 2151 | **64** |
| 5. Titanium | 2151 | 2151 | **93** |
| 6. Silicon dioxide | 2151 | 2151 | **108** |
| 7. Carbon fiber | 2151 | 2151 | **145** |

**TABLE 6.** Comparison of dimensionality for different methods' feature-vectors, AVIRIS data set.

| Class \ Dimension | Methods | | |
|---|---|---|---|
| | SVM | TL | The proposed |
| 1. Alfalfa | 220 | 220 | **21** |
| 2. Corn/notill | 220 | 220 | **21** |
| 3. Corn/min | 220 | 220 | **15** |
| 4. Corn | 220 | 220 | **19** |
| 5. Grass/Pasture | 220 | 220 | **28** |
| 6. Grass/trees | 220 | 220 | **25** |
| 7. Grass/Pasture/Mowed | 220 | 220 | **23** |
| 8. Hay/windrowed | 220 | 220 | **19** |
| 9. Oats | 220 | 220 | **19** |
| 10. Soybeans/notill | 220 | 220 | **19** |
| 11. Soybeans/min | 220 | 220 | **20** |
| 12. Soybean/clean | 220 | 220 | **18** |
| 13. Wheat | 220 | 220 | **25** |
| 14. Woods | 220 | 220 | **21** |
| 15. Building/Grass/Tree/Drives | 220 | 220 | **25** |
| 16. Stone/steel towers | 220 | 220 | **23** |

detailed differences of the classification results for the $k$-NN (Figure 6(a)), the SAM (Figure 6(b)), the SVM (Figure 6(c)), the TL method (Figure 6(d)) and the proposed method (Figure 6(e)). For example, in the two rectangles surrounded by black lines in Figure 6(a), we see that the proposed method indeed corrects some classification errors made by the $k$-NN and the SAM methods, which echoes the previous quantitative analysis.

### C. COMPARISON OF FEATURE VECTORS

It is shown that the proposed method achieves better classification accuracy than the two classical hyperspectral classification methods (i.e., the $k$-NN and the SAM) in the laboratory acquired ASD data set. And in the remotely sensed AVIRIS data set, the classification accuracy of the proposed method is just below the two state of the art methods (i.e., SVM and TL). However, the worth mentioning value of the proposed method is its compact feature representation. Table 5 and Table 6 compare the dimensionality of the feature-vectors for different methods based on the ASD data set and the AVIRIS data set respectively.

Due to different materials or objects have different numbers of absorption valleys, so the numbers of the 'eigen-absorption' features calculated for each of the classes in the ASD data set and AVIRIS data set are different (see the fourth columns of Table 5 and Table 6). On the other hand, the conventional hyperspectral classification methods and the state of the art machine learning based methods use the complete spectra as their features, so the numbers of their

features are 2151 and 220 for the ASD data set and AVIRIS data set respectively (see the second and the third columns of Table 5 and Table 6).

It is shown that the proposed method has a much sparser feature representation in contrast to the traditional methods that take the whole set of reflectance values as their feature vectors. In the laboratory ASD data set, the average dimension of the proposed method's feature-vector is about 95, much smaller than the 2151 dimensional full spectral vectors used in the classical $k$-NN and SAM methods. In the remotely sensed AVIRIS data set, the average dimension of the proposed method's feature vector is about 21, also smaller than the 220 dimensional full spectral vectors used in the state of the art SVM and TL methods.

Though the proposed method does not outperform the competitive machine learning based SVM and TL methods in the remotely sensed data set, it is observed that the difference in classification accuracy between them is quite narrow (i.e., less than 0.3%). Considering its advantage of the compact and sparse feature representation, the proposed method could still be useful in the applications where the storage space or transmission bandwidth are limited and the labeled training samples are expensive to acquire.

In the controlled laboratory experiment because the data are acquired with lower noise (both from the sensor and from the environmental lighting), the statistical variability of the spectral curves is relatively smaller. So we use one sample from each of the materials as the gallery signature, and the other two as the probe curves. Experimental results show that the classification accuracies of the k-NN method and the SAM method are inferior to the proposed method. This could be explained as the $k$-NN method and the SAM method take the whole set of the bands in classification and the redundant bands therein might deteriorate their performance (e.g., the 'Hughes Phenomenon'). In the remotely sensed AVIRIS data set, we use more training samples. However in remote sensing applications, the training samples have to be obtained by expensive field survey or heavy-labored manual labeling. Comparing to other general image classification applications, it tends to be more difficult to obtain a large number of samples for classifiers' training. So in our experiment, we randomly choose a relatively smaller set of data as the training samples. It is useful to testify the performance of the proposed method in a scenario where training samples are expensive to obtain.

From the above experiments, we show that the proposed method can outperform many conventional hyperspectral classification approaches. In contrast to the state of the art methods, the proposed method needs a size of training samples much smaller than the deep learning based methods or support vector machines. In some applications of materials identification, like the aforementioned ASD test, it is difficult or costly to obtain enough amount of training data due to the expensiveness of the materials. This will hamper the using of the machine learning based approaches, and make the proposed method useful. So in our experiments, to demonstrate the advantage of the proposed method for this scenario of applications we use one sample (in the above ASD data set) and a small percentage of the total data (3% in the AVIRIS data set) as the training data. With more samples are adopted for training, the machine learning based methods will surpass the proposed method but at the cost of highly expensive data acquisition.

## VI. CONCLUSION AND DISCUSSION

In this paper, we discuss a hyperspectral classification method based on a novel feature representation, namely the spectral absorption features. This idea is originated from the research of spectrometry, but it has not been fully exploited in the area of remotely sensed hyperspectral classification. Given the capability of spectral absorptions to characterize a material's constituents, the spectral absorptions are found to be powerful to hyperspectral classification. Aiming at an effective hyperspectral classification, first we propose to search for the unique spectral absorptions, i.e., the 'eigen-absorptions', for materials identification. But further research shows that this straightforward approach may fail due to the empty set of 'eigen-absorptions'. It is usually happened when a large number of objects are involved in the classification and the increased amount of pair-wise comparisons dramatically reduce the possibility of finding unique spectral absorptions. Therefore, we propose an improved approach, which does not require the distinctiveness or uniqueness for the absorption features. In this method we designed three algorithms, including the continuum removal algorithm, the absorption selection algorithm and the feature matching algorithm. The techniques from the envelop detection, the mutual-information based feature selection and the multi-label learning are adopted in our algorithms. To assess the classification performance, experiments have been carried out based on a laboratory ASD hyperspectral data set and a remotely sensed AVIRIS 92AV3C data set. The experimental results show that the proposed method outperformed the traditional nearest neighbor classifier and the spectral angle mapping method. Its classification accuracy is also competitive to the support vector machine method (SVM) and the deep-learning based transfer learning (TL) method. Comparing to the state-of-the-art methods, such as the SVM and the TL, the proposed method has significant advantages of compact feature representation and shorter training time. Future research directions regarding this approach include more accurate extraction of effective absorptions and better classifying strategies.

## REFERENCES

[1] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, 1985.
[2] A. F. H. Goetz, "Three decades of hyperspectral remote sensing of the Earth: A personal view," *Remote Sens. Environ.*, vol. 113, pp. S5–S16, Sep. 2009.
[3] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[4] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Campsvalls, J. Chanussot, M. Fauvel, P. Gamba, and A. J. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, Sep. 2009.

[5] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2004.

[6] B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson, "Band selection for hyperspectral image classification using mutual information," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 522–526, Oct. 2006.

[7] L. O. Jimenez and D. A. Landgrebe, "Hyperspectral data analysis and supervised feature reduction via projection pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2653–2667, Nov. 1999.

[8] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[9] J. Lin, C. He, Z. J. Wang, and S. Li, "Structure preserving transfer learning for unsupervised hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1656–1660, Oct. 2017.

[10] J. Lin, L. Zhao, S. Li, R. Ward, and Z. J. Wang, "Active-learning-incorporated deep transfer learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4048–4062, Nov. 2018.

[11] G. Licciardi, P. R. Marpu, J. Chanussot, and J. A. Benediktsson, "Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 447–451, May 2012.

[12] Q. Du and J. E. Fowler, "Low-complexity principal component analysis for hyperspectral image compression," *Int. J. High Perform. Comput. Appl.*, vol. 22, no. 4, pp. 438–448, 2008.

[13] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.

[14] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.

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

[16] Z. Ziyong, "Super-resolution reconstruction of hyperspectral images using empirical mode decomposition and compressed sensing," *J. Appl. Remote Sens.*, vol. 10, no. 4, 2016, Art. no. 042011.

[17] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. GRS-26, no. 1, pp. 65–74, Jan. 1988.

[18] C.-I. Chang and Q. Du, "Interference and noise-adjusted principal components analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2387–2396, Sep. 1999.

[19] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.

[20] B. Demir, S. Erturk, and M. K. Gullu, "Hyperspectral image classification using denoising of intrinsic mode functions," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 220–224, Mar. 2011.

[21] M. Shao, D. Kit, and Y. Fu, "Generalized transfer subspace learning through low-rank constraint," *Int. J. Comput. Vis.*, vol. 109, nos. 1–2, pp. 74–93, 2014.

[22] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.

[23] G. A. Licciardi and F. D. Frate, "Pixel unmixing in hyperspectral data by means of neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4163–4172, Nov. 2011.

[24] 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. 1–44, Dec. 2003.

[25] Z. Fu and A. Robles-Kelly, "Discriminant absorption-feature learning for material classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1536–1556, May 2011.

[26] B. Guo, H. Shen, and M. Yang, "Improving hyperspectral image classification by fusing spectra and absorption features," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1363–1367, Aug. 2017.

[27] B. Guo, "Entropy-mediated decision fusion for remotely sensed image classification," *Remote Sens.*, vol. 11, no. 3, p. 352, 2019.

[28] *ASD Visible, NIR (and SWIR) Spectrometers*. Accessed: Apr. 15, 2019. [Online]. Available: https://www.malvernpanalytical.com/en/products/product-range/asd-range/

[29] B. Guo, R. I. Damper, S. R. Gunn, and J. D. B. Nelson, "A fast separability-based feature-selection method for high-dimensional remotely sensed image classification," *Pattern Recognit.*, vol. 41, no. 5, pp. 1653–1662, 2008.

[30] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *Int. J. Data Warehousing Mining*, vol. 3, no. 3, pp. 1–13, 2007.

[31] T. Hastie and R. Tibshirani, "Discriminant adaptive nearest neighbor classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 6, pp. 607–616, Jun. 1996.

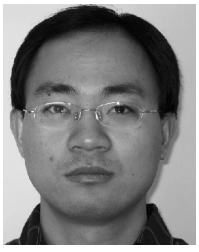
[32] Y. Sohn and N. S. Rebello, "Supervised and unsupervised spectral angle classifiers," *Photogramm. Eng. Remote Sens.*, vol. 68, no. 12, pp. 1271–1282, 2002.

[33] AVIRIS. *Airborne Visible/Infrared Imaging Spectrometer*. Accessed: Mar. 15, 2018. [Online]. Available: http://aviris.jpl.nasa.gov/

[34] *Aviris 92av3c Data*. Accessed: Jan. 15, 2019. [Online]. Available: https://ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/

[35] *Caffe Website*. Accessed: Apr. 15, 2019. [Online]. Available: https://github.com/BVLC/caffe/

[36] *Imagenet Website*. Accessed: Apr. 15, 2019. [Online]. Available: https://www.image-net.org



**BAOFENG GUO** received the B.Eng. degree in electronic engineering and the M.Eng. degree in signal processing from Xidian University, in 1995 and 1998, and the Ph.D. degree in signal processing from the Chinese Academy of Sciences, Beijing, in 2001. From 2002 to 2004, he was a Research Assistant with the University of Bristol. From 2004 to 2009, he was a Research Fellow with the University of Southampton, U.K. Since 2009, he has been with the School of Automation, Hangzhou Dianzi University, China. His current research interests include pattern recognition, image processing, and machine learning.