Improving Hyperspectral Image Classification by Fusing Spectra and Absorption Features

Baofeng Guo, Honghai Shen, and Mingyu Yang

***Abstract*— Many features can be extracted to classify hyperspectral imagery. Classification relying on a single feature set may lose some useful information due to the intrinsic limitation of each feature extraction model. To improve classification accuracy, we propose an information fusion approach, in which both the global and the local aspects of hyperspectral data are taken into account and are combined by a decision-level fusion method. The global features are hyperspectral reflectance curves representing the holistic response to the incident light, and the local features are absorption characteristics reflected by materials' individual constituents. The decision-level fusion is carried out by analyzing the entropy of the classification output from the global feature set and modifying this output via the results of a multilabel classification using the local feature set. Simulations of classification performance on 16 classes of vegetation from the AVIRIS 92AV3C and Salinas data set show the effectiveness of the method, which increases the classification accuracy compared to a popular support vector machine-based method and a production-rule-based decision fusion method.**

***Index Terms*— Decision-level fusion, hyperspectral image classification, multilabel classification.**

I. Introduction

HYPERSPECTRAL sensors simultaneously measure hundreds of contiguous spectral bands with a fine spectral resolution, e.g., 0.01 μm [1]. This makes it possible to reduce overlaps between classes, and therefore enhances the potential to discriminate subtle spectral differences. Using data from hyperspectral sensors, the classification is carried out by analyzing the electromagnetic reflectance as a function of the wavelength or band, i.e., "the spectral signature." In recent years, the hyperspectral image classification has received significant attention in many applications [2].

To effectively classify hyperspectral reflectance curves, defining appropriate features is one of the major challenges. Ideal features should characterize the intrinsic distinctness among different types of materials, and should be robust to atmospheric and neighboring pixel interferences. The commonly used features include the complete spectra [3], the spectral bands from feature selection methods, the transformed features from project pursuit, and so on.

Recently, absorption features have been considered as an effective alternative for identifying materials by imaging spectroscopy [4]. Absorptions are the dips or valleys in the reflectance spectra due to the incident light absorbed by the constituent atoms or molecules. There are evidences showing that the absorptions are connected with the material's chemical constituents, surface roughness, and so on. Since the complete spectra is a reflectance curve in a wide range of wavelength, the hyperspectral features based on the *complete* spectra can be considered as a *global* view of the hyperspectral data. On the other hand, the absorption features are the dips or valleys appeared around certain *narrow* wavelengths. So they can be considered as *local* views of the hyperspectral data. Because the spectral reflectance and the absorption features are two complementary responses (i.e., reflectance versus absorption) from the material to the incident light and describe the data from different views (i.e., the global view versus the local view), it is potential to improve the hyperspectral classification accuracy by combining two sets of features through information fusion.

In this letter, a hyperspectral classification approach based on a novel decision-level fusion is proposed. First, the spectra of a pixel are extracted as features to describe the holistic characteristics of the interaction between the materials and the incident light. Second, the valleys in a spectral reflectance curve are detected. An absorption feature vector is formed by labeling the vector with positive ones in the elements corresponding to these particular wavelength or bands. Associated with this feature categorization, an information fusion scheme is introduced to combine the above two sets of features.

In remote sensing, many studies have been carried out to employ information fusion for better performance. Unlike data-level and feature-level fusion that merge multisource data or multiple feature-sets to improve performance, decision-level fusion adopts rules to combine the results of individual classifiers to make the final decision. In this sense, information fusion can be achieved not only by combining different *sources* of data, such as in the traditional sensor fusion, but also by different feature extraction or "experts' views," which can compensate the deficiency in a *single* view or knowledge acquisition. In [5], a support vector machine (SVM)-based decision fusion is proposed, where SVMs are trained individually on different data sources and then are fused by another SVM. In [6], a decision fusion approach is investigated to combine results from a supervised SVM classifier and an unsupervised K-means classifier. Then a weighted majority voting rule is applied to the decision stage. In [7], a method based on factor graphs is used for multiple sensors data fusion. A context-sensitive object recognition method is discussed in [8], where a decision-level fusion is used to combine

Manuscript received November 29, 2016; revised April 1, 2017 and May 1, 2017; accepted May 25, 2017. Date of publication June 28, 2017; date of current version July 20, 2017. This work was supported by the National Natural Science Foundation of China under Grant 61375011. *(Corresponding author: Baofeng Guo.)*

B. Guo is with the School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China (e-mail: gbf@hdu.edu.cn).

H. Shen and M. Yang are with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China (e-mail: shenhh@ciomp.ac.cn; ymy1983@163.com).

Digital Object Identifier 10.1109/LGRS.2017.2712805

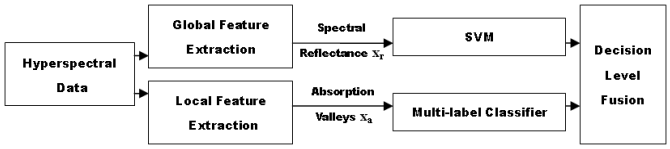

Fig. 1. Fusion diagram for two sets of hyperspectral features.

multiview remotely sensed data using the scene contextual information. In [9], a hyperspectral classification method is proposed based on a probabilistic weighted fusion rule for multiple spectral-spatial features. Other relevant researches on remote sensing can be found in [10]–[12]. For a comprehensive review of various decision fusion schemes, please refer to [13].

According to the nature of the above two sets of features, a popular SVM is chosen to classify the global reflectance features and a multilabel classifier [14] is used to classify the local absorption features. Furthermore, using an information entropy rule, a new scheme of decision-level fusion is proposed to better combine the results of the SVM and the multilabel classifier.

II. Decision-Level Fusion Based on Entropy

In the above arguments, we examined the ability of the absorption features to classification, and contend that they may lose efficacy due to the leak of information and the loss of exclusivity of absorption features. Therefore, for a more effective hyperspectral data classification, we consider an information fusion diagram, in which the global features of hyperspectral data (represented by the traditional spectral reflectance curve) and the local features (represented by the new-added absorption features) are combined. Because the two sets of features characterize the hyperspectral from the "reflectance" view and the "absorption" view, respectively, they are complementary to each other. Fusing them, therefore, will provide more knowledge about the desired hyperspectral signature for classification.

A. Fusion Framework

The proposed fusion diagram, shown in Fig. 1, consists of two sets of hyperspectral features. The upper branch of the features are the hyperspectral reflectance curves (i.e., the global features), which are given by

$$\mathbf{x}_r = \left(X_r^1, X_r^2, \ldots, X_r^L\right)$$

where X_r^i represents the ith band's reflectance value, $i = 1, 2, \ldots, L$, and L is the total number of bands in the hyperspectral data. SVMs have shown competitive performance with the best available algorithms in many classification areas including the hyperspectral data classification [3], and so were chosen as the classifiers for the global feature set $\mathbf{x}_r$.

The lower branch of features are new-added, which are based on the absorption valleys (i.e., the local features), and are represented by a binary vector

$$\mathbf{x}_a = \left(X_a^1, X_a^2, \ldots, X_a^L\right) \tag{1}$$

where the binary variable $X_a^i = 1$ if an absorption valley is detected in the ith band and $X_a^i = 0$ when no absorption occurs in this band. As discussed in Section I, if multiple classes are involved in the classification, the chance of finding unique absorption features for material identification becomes much slim. This is because the chance of the absorption features of one material coinciding with others increases significantly when more materials are involved. Fig. 2 shows the locations of the absorption features for all 16 classes of materials listed in AVIRIS 92AV3C. Before selecting absorption features, the spectra of the hyperspectral data are normalized to the range of $[0, 1]$. Then, a peak detection algorithm is carried out on the normalized spectra to find all absorption valleys. To avoid the impact from noise, only the absorption valleys that can meet the following two criteria will be selected as the absorption features. First, the depth of the absorption valley should be larger than 0.005 that is chosen by empirical observation. Second, the absorption features should appear on more than half of the training spectra, which can reduce the interference from noise. From Fig. 2, it is seen that the majority of the absorption features are shared by each other, and it is almost impossible to find a unique absorption feature for identification in this case. Therefore, a method based on a multilabel classification [14], depicted as follows, is proposed to better handle the absorption features.

B. Multilabel Classification for Absorption Features

Multilabel classification studies the problem where multiple target labels must be assigned to each instance or feature. This is exactly the problem that we encounter for classifying the absorption features, where an absorption valley may be associated with a set of materials simultaneously (see Fig. 2). Generally, the multilabel learning algorithms can be categorized as two groups, namely, the problem transformation methods and the algorithm adaptation methods [14]. Here we adapt a binary relevance method in our application. The basic idea of the binary relevance method is to decompose the multilabel classification into N independent binary classification problems, where N is the total number of the classes or labels. The traditional relevance method is not designed for inputs of binary vectors, so it cannot be used in our application straightforwardly. Hence, we modified the method as follows.

Given a training set as a group of binary vectors $\mathbf{x}_i = (X_i^1, X_i^2, \ldots, X_i^L)$, $X_i^j \in [0, 1]$ and their corresponding class labels $\mathbf{y}_i \in \{1, 2, \ldots, N\}, i = 1, 2, \ldots, M$, where $\mathbf{x}_i$ represents an absorption feature vector defined in (1), L is the total number of bands, and M is the number of training samples, the absorption feature vector $\mathbf{x}'_k$ for class $\mathbf{y}_k = n$, $n \in \{1, 2, \ldots, N\}$, is summarized by ensuring that the majority (e.g., 95%) of samples have the absorption valleys represented by $\mathbf{x}'_k$. Then we get a new training set as follows:

$$T = \left\{\left(\mathbf{x}'_k, \mathbf{y}_k\right)\right\}, \quad k = 1, 2, \ldots, N$$

where $\mathbf{x}'_k = [(X')_k^1, (X')_k^2, \ldots, (X')_k^L]$ representing the kth materials' absorption feature vector, $(X')_k^j \in \{0, 1\}$. An example of the training set T can be demonstrated by Fig. 2, where each row of the figure illustrates a 220-dimensional absorption feature vector $\mathbf{x}'_k$ corresponding to the vegetation labeled by the y-axis. In Fig. 2, the black dots represent the positive ones (i.e., presence of absorption) and

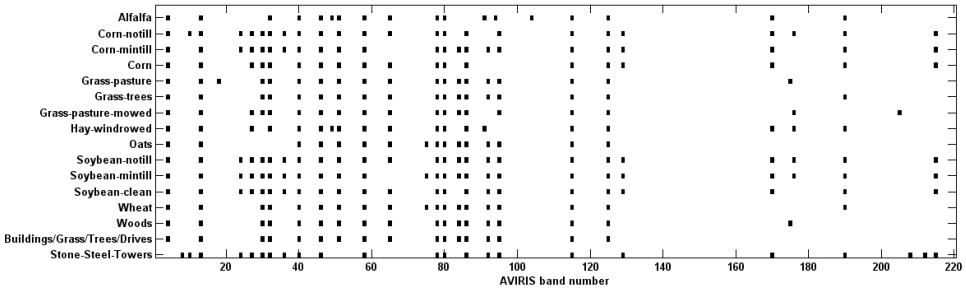

Fig. 2. Absorption features (labeled by black dots) for 16 classes of materials in AVIRIS 92AV3C data set.

the white areas are zeros (i.e., absence of absorption). Because in the binary vector $\mathbf{x}'_k$ the positive value of the component $(X')_k^j$ (i.e., the absorption valley is found in the jth band) could be the result of a multiclass set Y_j, $Y_j = \{\mathbf{y}_1, \mathbf{y}_2, \ldots\}$, it becomes a classical problem of multilabel classification, and can be solved by the following steps [14].

1) A binary training set D_k is formed by assigning the relevance of each absorption feature to $\mathbf{y}_k$

$$D_k = \left\{\left((X')_k^j, \phi(Y_j, \mathbf{y}_k)\right\}, \quad j = 1, 2, \ldots, L$$

where

$$\phi(Y_j, \mathbf{y}_k) = \left\{ \begin{array}{ll} +1, & \text{if } \mathbf{y}_k \in Y_j \\ -1. & \text{otherwise.} \end{array} \right\}$$

2) A binary classifier f_k is trained based on D_k, $(k = 1, 2, \ldots, L)$, where for relevant label $\mathbf{y}_k \in Y_j$, $(X')_k^j$ is regarded as one positive example, and vice versa;
3) Prediction for unseen instance $\mathbf{x}$ is carried out by querying $(X')^j$, $j = 1, 2, \ldots, L$ on each of the individual binary classifier, and the result is given by

$$\mathbf{y} = \underset{k}{\operatorname{argmax}}(\mathbf{y_k} | f(X^j) > 0, 1 \le k \le N, 1 \le j \le L).$$

C. Fusion Rule Based on Entropy

After extracting the reflectance feature vector $\mathbf{x}_r$ and the absorption feature vector $\mathbf{x}_a$, a natural way to combine them for classification is by information fusion. In this application, the two feature sets have different utilities to represent the hyperspectral data. To meet their specific classifier requirements, it is better to assign a corresponding classifier to each feature set individually. Meanwhile, as we discussed in Section II-B it is preferred to use the multilabel classification for the absorption features. So to achieve a better performance, a decision-level fusion is investigated.

The decision-level fusion is a high-level operation, where separate intermediate decisions are drawn from each individual feature set and then combined to reach a global decision. As we discussed above, we choose an SVM to classify the reflectance feature vector $\mathbf{x}_r$ and a multilabel algorithm to classify the reflectance feature vector $\mathbf{x}_a$. As indicated by our simulation, the classification accuracy of the reflectance feature vector $\mathbf{x}_r$ is higher than that of the absorption feature vector $\mathbf{x}_a$, therefore, a customized decision-level fusion rule is designed for this scenario as follows:

1) applying SVM to the reflectance feature vector $\mathbf{x}_r$, and getting an initial classification result $\mathbf{y}_r$;
2) applying the multilabel algorithm to the absorption feature vector $\mathbf{x}_a$, and getting a complementary classification result $\mathbf{y}_a$;
3) assessing the accuracy of the initial classification result $\mathbf{y}_r$;
4) if the accuracy of $\mathbf{y}_r$ is satisfactory, the final decision is given by the primary classification result from the reflectance features $\mathbf{x}_r$, i.e., $\mathbf{y} = \mathbf{y}_r$; otherwise, the complementary result from the absorption features $\mathbf{x}_a$ is selected as the final decision, i.e., $\mathbf{y} = \mathbf{y}_a$.

In 1), i.e., the assessment of the accuracy of the classification result for the reflectance feature vector $\mathbf{x}_r$, we first calibrate the SVMs' output to probability. Since standard SVMs do not provide posterior probability directly, a mapping method [15] is applied, where an additional sigmoid function is used to approximate the necessary posterior probability.

After getting the SVMs' probability outputs, we measure the accuracy of the SVMs' results by calculating the uncertainty of $p(\mathbf{y}|\mathbf{x_r})$. In this sense, the higher the uncertainty of the output is, the lower the accuracy of the classification is. In information theory, the uncertainty is often measured by entropy. So to assess the accuracy of the classification result $\mathbf{y}_r$, one convenient approach (based on the above basic concept in information theory) is to calculate the entropy of $p(\mathbf{y}|\mathbf{x_r})$ as follows:

$$H(Y|\mathbf{x_r}) = -\sum_{\mathbf{y} \in Y} p(\mathbf{y}|\mathbf{x_r}) \log p(\mathbf{y}|\mathbf{x_r}). \tag{2}$$

Based on (2), we propose a decision-level fusion rule, which can be depicted in Algorithm 1.

In Algorithm 1, the threshold η is used to measure whether the accuracy of the SVMs' output is satisfactory. The particular value of η can be determined by examining the training data, and the detailed discussion is given in Section III.

Comparing with previous research, the present approach is neither a "hard" decision fusion that usually uses a majority voting to get the final decision, nor a "soft" decision fusion that usually uses a Bayesian probability combination. It, however, can be considered as a *hybrid* decision fusion, in which an uncertainty measurement (e.g., the entropy) is drawn from the "soft" outputs (i.e., the probability outputs of the SVM), and is used to arbitrate the "hard" decisions before reaching the final decision. The strategy is particularly useful in our application

Algorithm 1 Decision-Level Fusion Rule Based on Entropy

Require: $\mathbf{x_r}$,$\mathbf{x_a}$, η
Ensure: $\mathbf{y}$
$\mathbf{y_r} \leftarrow$ classify $\mathbf{x_r}$ using SVMs
$\mathbf{y_a} \leftarrow$ classify $\mathbf{x_a}$ using the multi-label classifier
$p(\mathbf{y}|\mathbf{x_r}) \leftarrow$ transform SVMs' output using (1)
$H(Y|\mathbf{x_r})) \leftarrow$ calculate entropy based on $p(\mathbf{y}|\mathbf{x_r})$
if $H(Y|\mathbf{x_r})) < \eta$ **then**
  $\mathbf{y} \leftarrow \mathbf{y_r}$
else
  $\mathbf{y} \leftarrow \mathbf{y_a}$
end if

since only two decisions can be made from the two sets of features, in which the majority voting is not able to work.

III. Simulation Results

Simulations of classification performance have been carried out to assess the proposed band-selection method on the hyperspectral data set AVIRIS 92AV3C. The database is illustrative of the problem of hyperspectral image analysis to determine land use. It can be downloaded from ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/. The AVIRIS sensor collects nominally 224 bands of data, but 4 of these contain only zeros and so are discarded, leaving 220 bands in the 92AV3C data set. Each image is of size 145×145 pixels. The datacube was collected over a test site called Indian Pine in north-western Indiana, USA. The database is accompanied by a reference map, indicating partial ground truth. Each pixel is labeled as belonging to one of 16 classes of vegetation, including Alfalfa, Corn-notill, Corn-mintill, and so on. Pixels of all the 16 classes of vegetation are used in the simulation.

To implement the proposed fusion method, the parameter η is needed to be chosen at first. Two conditional probability distributions, i.e., $p_t = P(H(Y|\mathbf{y_r} = \mathbf{y})$ and $p_f = P(H(Y|\mathbf{y_r} \neq \mathbf{y})$, are estimated by normalizing the histograms of $H(Y|\mathbf{y_r} = \mathbf{y})$ and $H(Y|\mathbf{y_r} \neq \mathbf{y})$, where the random variable Y represents the output of the SVM, and conditions $\mathbf{y_r} = \mathbf{y}$ and $\mathbf{y_r} \neq \mathbf{y}$ represent correct decisions and wrong decisions, respectively. It is known that with the increase of the entropy of $H(Y)$, p_t decreases and p_f increases, respectively. In other words, with the increase of the entropy (i.e., the uncertainty of the output), the chance of incorrect classification will rise and the chance of correct classification will fall. Since in our algorithm the parameter η is designed to measure whether the accuracy of the SVMs' output is satisfactory, it can be chosen as a value where the probability of the misclassification is just above the probability of the correct classification, as follows:

$$\eta^* = \underset{\eta}{\operatorname{argmin}} \left(\int_{\eta}^{\infty} p_t dH(Y) \geq \int_{\eta}^{\infty} p_t dH(Y) \right). \quad (3)$$

For accuracy assessment, about 10% pixels from each class were randomly chosen as the training data set, with the remaining 90% forming the test set on which performance was assessed. The random sampling used here is to be consistent with previous researches and to allow an estimate of the standard deviation caused by sampling. However, it is noticed that this random sampling may tend to overestimate the classification accuracy due to the spatial autocorrelation between the test and training samples. To amend this, a more adequate assessment could be set up by a spatially disjointed sampling.

TABLE I
Classification Performance (SVM-Based Method Using Reflectance Features Versus the Proposed Fusion Scheme Using Reflectance and Absorption Features, AVIRIS 92AV3C Data Set, 10% Training Samples)

Class	Pixels in testing	Accuracy (%) SVM method	Accuracy (%) Fusion method
1.Alfalfa	41	19.51	17.07
2.Corn-notill	1285	41.94	64.90
3.Corn-min	747	42.30	72.42
4.Corn	213	21.12	30.51
5.Grass/Pasture	435	47.59	82.07
6.Grass/Trees	657	87.82	93.15
7.Grass/pasture-mowed	25	0.00	0.00
8.Hay-windrowed	430	98.37	99.53
9.Oats	18	0.00	0.00
10.Soybeans-notill	875	66.74	87.20
11.Soybeans-min	2209	79.86	71.80
12.Soybean-clean	534	25.28	69.66
13.Wheat	184	98.37	98.91
14.Woods	1138	96.13	97.10
15.Bldg-Grass-Tree-Drives	347	30.55	48.99
16.Stone-steel towers	84	91.67	92.86
Overall accuracy (%)		65.67	76.99
Kappa coefficient		0.60	0.74

As discussed previously, SVMs are chosen as the classifier for the reflectance feature vector. The kernel function used is a heterogeneous polynomial. The penalty parameter C is tested between 10^{-3} and 10^5, and the polynomial order is tested from 1–10 by a twofold validation procedure using only training data. The polynomial order 4 and $C = 1500$ were finally found as the best values for this training set, and applied to the following testing stage.

Table I compares the performance of the fusion method and the SVM-based method [3]. It can be seen that the overall accuracy of the fusion methods is 76.99%, which is significantly higher than that of the SVM-based method using the reflectance features. In all 16 classes, there are 12 classes whose individual classification accuracies are improved by the fusion method. The Cohen's kappa coefficient of the fusion method is also higher than that of the SVM-based method (0.74 versus 0.60).

To avoid bias on random samplings of the training data set, the testing was repeated 10 times to allow an estimate of the error in this sampling process. The 10 times classification results based on the reflectance features and three fusion approaches are shown in Table II. The first method is the SVM-based method that uses only the reflectance features. The second method is a fusion method based on the production rule [13]. Both the third and the fourth methods are based on the proposed fusion scheme except that the third one uses the spectral angle mapping (SAM) algorithm to classify the reflectance features but the fourth one uses the SVM. It is seen

TABLE II

PERFORMANCE COMPARISON OF THE PROPOSED FUSION SCHEME WITH OTHER METHODS, AVIRIS 92AV3C DATA SET, 16 CLASSES, 10% TRAINING SAMPLES

	Overall Accuracy (%±STD)	Cohen's Kappa Coefficient
SVM method [3]	64.98±1.55	0.62
Fusion by product rule [13]	76.24±1.44	0.72
Fusion using SAM	76.20±1.01	0.72
Proposed fusion using SVM	77.20±1.12	0.74

TABLE III

PERFORMANCE COMPARISON OF THE PROPOSED FUSION SCHEME WITH OTHER METHODS, AVIRIS 92AV3C DATA SET, 7 CLASSES, 5% TRAINING SAMPLES

	Overall Accuracy (%)	Standard deviation
SVM method [3]	73.47	1.39
Fusion by product rule [13]	82.52	0.77
Fusion ABS+SVM [4]	83.84	0.74
Proposed fusion using SVM	84.24	1.87

TABLE IV

PERFORMANCE COMPARISON OF THE PROPOSED FUSION SCHEME WITH OTHER METHODS, SALINAS DATA SET, 16 CLASSES, 5% TRAINING SAMPLES

	Overall Accuracy (%±STD)	Averaged Cohen's Kappa Coefficient
SVM method [3]	88.41±0.60	0.87
Fusion by product rule [13]	91.03±0.22	0.90
Proposed fusion using SVM	92.09±0.47	0.91

that the overall classification accuracies of the fusion methods are higher than that of the SVM-based method using only the reflectance features. It is reasonable due to the new-added information from the absorption features. The differences among the three fusion methods show that the proposed fusion scheme is slightly better than other two alternatives, in which the choice of using SVM as the classifier to the reflectance features is found to be better than that of using the SAM classifier.

To compare with more fusion schemes, another assessment is carried out based on the AVIRIS 92AV3C data set. To keep the same experimental setting with [4], seven classes of vegetations, including Corn-notill, Corn-mintill, Grass-trees, Soybean-notill, Soybean-mintill, Soybean-clean, and Woods, are chosen for classification, and about 5% of samples are used for training. The experimental results are shown in Table III. It is seen that the performance of the proposed fusion scheme is competitive to the state-of-the-art fusion approaches [4], [13], and outperforms the popular SVM-based method [3].

To further testify the proposed fusion scheme, another hyperspectral data set, Salinas scene, is assessed. This scene was collected by the 224-band AVIRIS sensor over Salinas Valley, California, USA. The area covered comprises 512 lines by 217 samples and includes vegetables, bare soils, and vineyard fields. In the experiment, about 5% of samples are used for training and the remaining 95% of samples are used for testing. Table IV shows the overall classification accuracies based on 10 times random samplings of the training data. It can be seen from Table IV that the classification accuracy of the proposed fusion is slightly higher than the product fusion scheme [13] and the SVM-based method [3]. Compared to the 92AV3C scene, the classification of the Salinas data is relatively easier, which makes the improvement by decision fusion limited.

IV. CONCLUSION

A decision-level fusion method for hyperspectral image classification has been described based on the reflectance features and the absorption features. Given the complementary relation between the reflectance features and the absorption features, combining them is analogous to the idea of sensors fusion that many fusion literatures employ. It can better characterize the spectral signature from both a global reflectance-view and a local absorption-view. Experiments on the AVIRIS 92AV3C and Salinas data sets show that the proposed method outperformed the popular SVM-based method and was competitive with the state-of-the-art fusion methods. The future work will concentrate on new methods for better classifying the absorption features.

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] 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.

[3] 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. 2005.

[4] 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.

[5] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3858–3866, Dec. 2007.

[6] H. Yang, Q. Du, and B. Ma, "Decision fusion on supervised and unsupervised classifiers for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 875–879, Oct. 2010.

[7] A. Makarau, G. Palubinskas, and P. Reinartz, "Alphabet-based multisensory data fusion and classification using factor graphs," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 969–990, Apr. 2013.

[8] F. T. Mahmoudi, F. Samadzadegan, and P. Reinartz, "Object recognition based on the context aware decision-level fusion in multiviews imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 1, pp. 12–22, Jan. 2015.

[9] Z. Chunsen, Z. Yiwei, and F. Chenyi, "Spectral–spatial classification of hyperspectral images using probabilistic weighted strategy for multifeature fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 10, pp. 1562–1566, Oct. 2016.

[10] B. Luo, M. M. Khan, T. Bienvenu, J. Chanussot, and L. Zhang, "Decision-based fusion for pansharpening of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 19–23, Jan. 2013.

[11] D. G. Stavrakoudis, E. Dragozi, I. Z. Gitas, and C. G. Karydas, "Decision fusion based on hyperspectral and multispectral satellite imagery for accurate forest species mapping," *Remote Sens.*, vol. 6, no. 8, pp. 6897–6928, 2014.

[12] J. Wu, Z. Jiang, J. Luo, and H. Zhang, "Composite kernels conditional random fields for remote-sensing image classification," *Electron. Lett.*, vol. 50, no. 22, pp. 1589–1591, 2014.

[13] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.*, vol. 6, no. 3, pp. 21–45, Sep. 2006.

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

[15] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Advances in Large Margin Classifiers*. Cambridge, MA, USA: MIT Press, 1999, pp. 61–74.