Research Article

A Research on the Combination Strategies of Multiple Features for Hyperspectral Remote Sensing Image Classification

Yuntao Ma,¹ Ruren Li¹, ,¹ Guang Yang,² Lishuang Sun,¹ and Jingli Wang¹

¹School of Transportation Engineering, Shenyang Jianzhu University, Shenyang 110168, China

²School of Geography, South China Normal University, Guangzhou 510631, China

Correspondence should be addressed to Ruren Li; rurenli@163.com

Received 30 December 2017; Accepted 4 April 2018; Published 13 May 2018

Academic Editor: Yichun Xie

Copyright © 2018 Yuntao Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It has been common to employ multiple features in the identification of the images acquired by hyperspectral remote sensing sensors, since more features give more information and have complementary properties. Few studies have discussed the combination strategies of multiple feature groups. This study made a systematic research on this problem. We extracted different groups of features from the initial hyperspectral images and tried different combination scenarios. We integrated spectral features with different textural features and employed different dimensionality reduction algorithms. Experimental results on three widely used hyperspectral remote sensing images suggested that “dimensionality reduction before combination” performed better especially when textural features performed well. The study further compared different combination frameworks of multiple feature groups, including direct combination, manifold learning, and multiple kernel method. The experimental results demonstrated the effectiveness of direct combination with an autoweight calculation.

1. Introduction

The analysis of hyperspectral images has been more and more discussed in recent years. In the classification problem, it has been widely accepted that features from different views help to better recognize objects. It is common to extract multiple features before the classification procedure. Therefore, researchers have made great efforts in several aspects. First, for a long time, people have been striving to extract desirable and suitable features from hyperspectral remote sensing images for better representation [1–6]. It has been recognized that linear features are less effective than nonlinear features [7], while it reduces efficiency to obtain nonlinear features in many cases. Some efforts have been made to combine both features simultaneously [8]; second, more recently, many researchers have successfully designed different frameworks [9–12] to organize different types of features, like texture features, shape features, and spectral features, because the views from different feature spaces have particular statistical

properties [13]. Finally, great efforts have been made in the classification process. Classifiers based on kernels [14] have dealt well with the Hughes phenomenon [15]. It has led to a trend towards multiple kernel learning (MKL) [16–18], in which different groups of features have different kernel matrices and finally a composite kernel is yielded.

In general, the above approaches combine features extracted from peculiar bands or features (e.g., top principal component) with the initial hyperspectral bands and then put different groups of features into dimensionality reduction (DR) frameworks or classifiers [19, 20]. However, complementary properties of multiple features have not been widely considered or analyzed. Few approaches have taken global features which can be transformed or converted from the complete hyperspectral bands as a group of input features. More frequently, approaches like principal component analysis (PCA) [21], linear discriminant analysis (LDA) [22], isometric feature mapping (ISOMAP) [23], and Laplacian eigenmaps (LE) [24] have been only exploited as global

dimensional reduction techniques. The low-dimensional output features have reduced the contribution of each group of features [9, 10].

In this paper, we first made a systematic research on two schemes for hyperspectral image classification based on multiple features by utilizing different DR tools (linear and nonlinear) and different types of features. One scheme was combining spectral features with other features before the DR process; the other was reducing the dimensionality of the original hyperspectral features before combining with other extracted features. Based on the experimental results on three hyperspectral datasets, we suggested that alternative decisions should be made in different circumstances. Based on the research, we further compared different combination frameworks in hyperspectral remote sensing image classification. We selected complementary features including linear and nonlinear global features (in this paper, we take features converted from all the bands as “global features”) and two kinds of textural features (extracted from certain bands or layers). Three combination frameworks were tested on three frequently used hyperspectral datasets, which comprised two scenes collected by airborne visible/infrared imaging spectrometer (AVIRIS) over the Indian Pine region and Salinas valley, and one scene collected by the reflective optics spectrographic imaging system (ROSIS) over Pavia University.

The remainder of the paper is organized as follows. Section 2 provides the details about the source of three hyperspectral remote sensing datasets employed in the experiments and the process of the proposed method. Then, the experimental results are reported in detail in Section 3, including the comparison of two classification schemes based on multiple features and the test of the combination frameworks with the results both in accuracy and visual perspectives. Discussions based on the classification results are also included in this section. Finally, a general summary of the paper is represented in Section 4.

2. Methodology

2.1. Multiple Feature Extraction

2.1.1. Dimensionality Reduction of Spectral Features. In this paper, we consider features achieved from all the spectral bands as global features (GF). Generally, GF are divided into two categories. One is linear features, which commonly derive the information from the original spectral image bands by multiplying a transformation matrix. Among them, PCA is a conventional linear transition without class label information and has quite high time efficiency; the other is nonlinear features including manifold learning features and kernel features [25, 26]. More and more investigators have focused on absorbing or exploiting nonlinear features in the classification problems because linear features do not take into account the underlying nonlinear class boundaries [27]. However, linear and nonlinear global features are rarely combined in previous studies and the complementary properties of two types of features have not been widely discussed.

Representative manifold learning algorithms for DR comprise locally linear embedding (LLE) [28], ISOMAP,

Laplacian eigenmaps (LE), and local tangent space alignment (LTSA) [29]. This paper extracts ISOMAP and LE features from the hyperspectral datasets.

2.1.2. Textural Feature Extraction. In this paper, we take textural features derived from certain spectral bands or features as TF (textural feature). Many approaches have supplemented TF to spectral features because they complement features from different perspectives and give some detailed information. Two frequently used textural features are exploited in the scheme, including filter-based Gabor features and statistical GLCM features.

(1) 2D Gabor Textural Feature. We consider the procedure in [30]; a Gabor function is defined as

$$\Psi_{s,d}(x, y) = \Psi_{\bar{\mathbf{k}}}(\bar{\mathbf{x}}) = \frac{\|\bar{\mathbf{k}}\|}{\delta^2} \cdot \exp\left(-\frac{\|\bar{\mathbf{k}}\|^2 \cdot \|\bar{\mathbf{x}}\|^2}{2\delta^2}\right) \cdot \left[\exp(i\bar{\mathbf{k}}\bar{\mathbf{x}}) - \exp\left(-\frac{\delta^2}{2}\right)\right], \quad (1)$$

where $\bar{\mathbf{x}} = (x, y)$ is the image location in spatial domain and frequency vector $\bar{\mathbf{k}}$ determines the scales and directions of Gabor functions. It is defined as

$$\bar{\mathbf{k}} = \frac{\pi}{2f^s} \cdot \exp\left(i \cdot \left(\frac{\pi d}{8}\right)\right). \quad (2)$$

In our experiment, parameter f is fixed to 2. The scale parameter s ranges from 0 to 3, and the direction parameter d ranges from 0 to 7, which stands for 4 scales and 8 directions. s and d are both integers. Parameter δ is fixed to 2π which represents the number of oscillations under the Gaussian envelope. According to [10], the textural images derived from Gabor filters are the real part of convolving the image $I(x, y)$ with different s and d .

$$F_{s,d}(x, y) = G_{s,d}(x, y) * I(x, y). \quad (3)$$

(2) GLCM Textural Feature. The gray-level cooccurrence matrix (GLCM) [31] textural feature is a widely used statistical feature. Given a certain distance and a direction, a gray-level cooccurrence matrix is built by calculating the probability of the occurrence of two gray levels from a pixel. Various features can be obtained from GLCM, and we extract 8 features for the combination in the experiment, including mean value, variance, homogeneity, contrast, dissimilarity, entropy, second moment, and correlation (details of calculation process can be found in [31]). In the experiment, the grayscale quantization level is fixed to 64 and the preprocess window is 3×3 .

2.2. Combination Scenarios

2.2.1. Weight Estimation by Average Distance Measurement. Different types of features make different contributions to various classification problems. The paper automatically estimates the weights of different feature groups based on the

average distance measurement in Euclidean space. The procedure is as follows:

- (1) Concatenate and normalize different groups of features.
- (2) Calculate the mean values and variances of all classes for each feature from the sample set and compute the standardized distances between each two classes for each feature. The standardized distance is calculated as follows:

$$d = \frac{\mu_s - \mu_t}{\sigma_s + \sigma_t}, \quad (4)$$

where d is the standardized distance for each feature, μ_s and μ_t represent the mean values of class s and class t , and σ_s and σ_t represent the standard deviation of class s and class t . According to the standardized distances, if a couple of selected features give rise to the smallest distance within classes and the biggest distance between classes, the best classification result is likely to be yielded. Next, we extend the method to fit multiple classes and multiple features.

- (3) Calculate the sum of the standardized distances between each two classes for each feature and allocate the weight multipliers for different feature groups: the expression of a sample is defined as $X = [x_{i1}, x_{i2}, \dots, x_{iL(i)}]_{i=1}^n$, in which $L(i)$ is the number of the i th group of features and n is the number of feature groups ($n = 4$; PC, LE or ISOMAP, Gabor, and GLCM). We define w_i as the weight multiplier of the i th group. w_i can be calculated by

$$w_i = \frac{\sum_{k=1}^{L(i)} d_k^i / L(i)}{\sum_{i=1}^n \sum_{k=1}^{L(i)} d_k^i / \sum_{i=1}^n L(i)}, \quad (5)$$

in which

$$d_k^i = \frac{\sum_{p=1}^l \sum_{q=p}^l d_{pqk}^i}{l(l-1)/2}, \quad (6)$$

where d_{pqk}^i is the average standardized distance of the k th feature in the i th group between classes p and q (4) and l is the number of all the classes.

- (4) Renew the representation of the sample $X = [w_i x_{i1}, w_i x_{i2}, \dots, w_i x_{iL(i)}]_{i=1}^n$.

The advantages of the weight estimation method are as follows:

- (i) Each type of features has its own weights, which maintains the specific properties of different features.
- (ii) The weight multipliers are around 1, so the method does not influence much on the normalized values of the features.

- (iii) The method has high time efficiency without iterative procedure.

2.2.2. Combination and Classification. As discussed before, we have got several global features and two kinds of textural features. For GF, we select the first 10 PC, 10 ISOMAP features, and 10 LE features, while for TF, 32 Gabor features (4 scales associated with 8 directions) and 8 GLCM features are extracted from the PCA top component. The two textural features have been recognized to have complementary properties [32]. For GF, 10 features not only avoid large amount of the calculation but also give sufficient information, while for TF, the 8 GLCM features are employed frequently as well as the Gabor features. Although different studies have selected different numbers of scales and directions for Gabor features, we use relatively fewer features in order to increase the efficiency of the calculation procedure. We do not analyze the properties of different parameters in detail in this paper. For the first 2 datasets, ISOMAP features are employed as the only features based on manifold learning while for the third dataset, we try to utilize LE features instead of ISOMAP. As described before, the features actually contain 4 different types, including 10 linear GF (PC), 10 nonlinear GF (ISOMAP or LE), 32 filter-based TF (2D Gabor), and 8 statistical TF (GLCM). Features are normalized before inputting the following approaches.

The paper addressed three scenarios to combine multiple features.

- (i) Scenario 1: direct combination

We directly combine different types of feature vectors, such as GF and TF; thus, a longer feature vector is formed.

- (ii) Scenario 2: dimensionality framework

The state-of-the-art dimensionality algorithm LE has been widely discussed and utilized in many studies [33]. We consider the method reported in [9].

- (iii) Scenario 3: multiple kernel method

The existing multiple kernel learning algorithms have to calculate the weight factors through an iterative process. To avoid the complex iterative calculation, we estimate the weight of different features before classification. We design the method by the distance measurement. Finally, the basis kernels are multiplied by the relevant weight factors. Then, it is converted to a simple kernel classifier. When the combinations are decided, the SVM [34] classifier is employed to test on these features with its parameters C and g confirmed through cross-validation with the training samples [35].

3. Experimental Section

3.1. Hyperspectral Image Data. Three commonly used datasets were tested in the experiments. All of them were acquired by hyperspectral sensors with different spatial resolutions. Researchers have been trying to improve the classification performance on these scenes for a long time. In this paper,

TABLE 1: Sample distribution of IP dataset.

Number	Class	Train sample	Number	Number	Class	Train sample	Number
1	Corn-notill	72	1434	7	Soybean-notill	48	968
2	Corn-mintill	42	834	8	Soybean-mintill	123	2468
3	Corn	12	234	9	Soybean	31	614
4	Grass-pasture	25	497	10	Wheat	11	212
5	Grass-trees	37	747	11	Woods	65	1294
6	Hay-windrowed	24	489	12	Building-road	19	380

FIGURE 1: IP dataset: (a) IP scene; (b) labeled pixels.

experiments are carried out on these scenes under similar experiment conditions.

3.1.1. Indian Pine Scene. The Indian pine (IP) dataset, derived from the airborne visible/infrared imaging spectrometer (AVIRIS), is one of the most commonly used hyperspectral image data for test. The resolution of the image is 30 m, and the size is 145×145 pixels. The sensor contains 220 bands in which 62 bands have to be discarded due to water absorption or noise and finally 158 valid bands are reserved in the area. The dataset mainly covers agricultural lands with 10171 labeled data points divided into 12 classes. In the experiment, 5% of the labeled data points are considered as training samples for each class. Table 1 lists the exhaustive class information and the number of samples. Figure 1 shows the image and the labeled condition.

3.1.2. Salinas Scene. Salinas (SL) dataset (Figure 2) was also acquired by AVIRIS. 204 valid bands are selected from the total 224 bands. The resolution of the scene is 3.7 m, and the size is 512×217 pixels in which 54129 pixels are labeled. 1% of the labeled pixels are considered as training samples for each class. The dataset is divided into 16 classes with the details listed in Table 2.

3.1.3. Pavia University Scene. The Pavia University (PU) dataset (Figure 3) was acquired by ROSIS, and the location is Pavia University, Italy. The resolution is 1.3 m which ranks

the highest among the three datasets. The image size is 610×340 pixels, including 207400 data points. 113 valid bands are selected from the total 115 bands with 2 noisy bands removed. Different from the former scenes, PU mainly covers artificial lands. 2% of 42776 labeled pixels are considered as training samples for each class. The land cover details are listed in Table 3.

3.2. Research on Different Dimensionality Reduction Scenarios. A systematic research is made on two DR scenarios for hyperspectral image classification based on multiple features. Different DR tools (linear and nonlinear) and different types of textural features are employed. One scenario is the conventional procedure characterized by combining hyperspectral bands with other features before reducing the dimensionality; the other is featured by reducing the dimensionality of the original hyperspectral bands before combining other extracted features. In addition, classification scenarios by only spectral features and only textural features help to compare and analyze the results. As a result, four scenarios are listed in Table 4. In scenario 3, we search the best output dimension d among 5-45 with the interval of 5 for each dataset. In scenario 4, the first 10 features are selected after the DR process.

We repeat ten times of independent experiments for each case. In each trial, the samples are randomly selected from all the labeled pixels and the selection strategy is stratified by

TABLE 2: Sample distribution of SL dataset.

Number	Class	Train sample	Number	Number	Class	Train sample	Number
1	Broccoli 1	20	2009	9	Grape 2	62	6203
2	Broccoli 2	37	3726	10	Corn	33	3278
3	Fallow 1	20	1976	11	Lettuce 1	11	1068
4	Fallow 2	14	1394	12	Lettuce 2	19	1927
5	Fallow 3	27	2678	13	Lettuce 3	9	916
6	Stubble	40	3959	14	Lettuce 4	11	1070
7	Celery	36	3579	15	Grape 3	73	7268
8	Grape 1	113	11271	16	Grape 4	18	1807

FIGURE 2: SL dataset: (a) SL scene; (b) labeled pixels.

FIGURE 3: PU dataset: (a) PU image; (b) labeled pixels.

classes. We calculated the average overall accuracy, kappa index, and the best d for DR scheme in scenario 3. The parameters are optimized by the training samples.

Three datasets (IP, SL, and PU) are tested, and the performance is reported in Table 5. For each method, d relates to the best performance in scenario 3. We can get from Table 5 that the classification accuracies in scenario 4 are generally higher than those in scenario 3, especially when textural features perform well by themselves. Even if textural features do not perform well, scenario 3 is not always superior to scenario 4. We may explore the reason referring to scenario 1 and scenario 2. When textural features outperform spectral features and have less numbers, scenario 4 obviously outperforms scenario 3. Features with a larger number may be dominated during a global DR transformation after the feature combination, regardless of whether the DR tool is linear or nonlinear. As a result, a higher accuracy yield is in scenario 4, owing to the good performance and sufficient feature numbers of textural features, like Gabor features. On the contrary, the initial hyperspectral features with both larger numbers and worse performance influence and reduce the accuracy (scenario 3). However, when spectral features outperform textural features, accuracies in scenario 3 are close or superior to those in scenario 4 according to the “PC-GLCM” and “LE-GLCM” methods in Table 5. In addition,

Table 5 shows that, regardless of whether the DR algorithm is linear or nonlinear, it is not easy to find the empirical d during the procedure of DR.

With the development of textural feature extraction technique, the performance of textural features researchers exploited often outperforms the initial hyperspectral features with less numbers. As a result, it is proper to reduce the dimensionality of hyperspectral features before combining with textural features or other features. Generally, we cannot exactly predict the performance of different groups of features, so we just reduce the dimensionality of hyperspectral features to a certain extent and select moderate numbers of low representation.

3.3. Combination Frameworks. In the experiment, we design 5 feature selection scenarios for comparison: hyperspectral bands only, GF only, TF only, integrating GF and TF, and integrating GF, TF, and hyperspectral bands. The results of 5 groups of features associated with 3 combining strategies are listed. In addition, for IP scenes, details of the DR framework (with the dimensions no more than 60) will be presented; for SL scenes, the classification accuracies with different groups of features will be shown in an intuitive way; for PU scenes, we will investigate the complementary properties of different feature groups.

TABLE 3: Sample distribution of PU dataset.

Number	Class	Train sample	Number	Number	Class	Train sample	Number
1	Asphalt road	133	6631	6	Bare ground	101	5029
2	Grass	373	18649	7	Asphalt roof	27	1330
3	Gravel	42	2099	8	Tile	74	3682
4	Trees	61	3064	9	Shadow	19	947
5	Metal roof	27	1345				

TABLE 4: Details of experimental scenarios.

Scenario 1	Spectral features
Scenario 2	Textural features
Scenario 3	Combine both features and reduce the dimensionality as a whole
Scenario 4	Combine textural features with low-dimensional spectral features

TABLE 5: Experimental results of different datasets associated with different scenarios. “Sc.” represents different scenarios in Table 4.

Method	IP				SL				PU			
	Sc.1	Sc.2	Sc.3	Sc.4	Sc.1	Sc.2	Sc.3	Sc.4	Sc.1	Sc.2	Sc.3	Sc.4
PC	81.31	93.60	94.31	96.08	89.52	95.9	98.28	98.44	82.87	85.7	97.44	98.67
Gabor	0.784	0.927	0.935	0.955	0.833	0.954	0.981	0.983	0.765	0.807	0.966	0.983
			$d = 30$				$d = 35$				$d = 40$	
PC	81.31	48.19	78.52	77.25	89.52	57.16	89.69	89.72	82.87	65.83	91.17	87.36
GLCM	0.784	0.365	0.752	0.737	0.833	0.508	0.885	0.885	0.765	0.503	0.882	0.83
			$d = 20$				$d = 30$				$d = 25$	
LE	81.31	93.60	80.11	95.66	89.52	95.9	95.51	98.78	82.87	85.7	96.17	97.26
Gabor	0.784	0.927	0.772	0.95	0.833	0.954	0.95	0.986	0.765	0.807	0.949	0.964
			$d = 40$				$d = 40$				$d = 35$	
LE	81.31	48.19	61.92	68.02	89.52	57.16	83.66	85.34	82.87	65.83	85.51	81.98
GLCM	0.784	0.561	0.561	0.628	0.833	0.508	0.817	0.837	0.765	0.503	0.805	0.754
			$d = 35$				$d = 35$				$d = 35$

FIGURE 4: Classification maps of the IP dataset with directly combined features; (a)–(e), respectively, represent different groups of input features: spectral features, GF, TF, GF combined with TF, and all features along with the spectral features. d is the feature dimension for input.

Also, complementary properties of linear and nonlinear features will be discussed. Finally, classification results with weight estimation and without weight estimation will be compared in accuracy.

10 independent experiments are repeated for each case. In each trial, the samples are randomly selected from all the labeled pixels and the selection strategy is stratified by classes. Overall accuracy (OA) is calculated by ten trails. We also get

FIGURE 5: Classification maps of the IP dataset with features under a manifold-based framework; (a)–(e), respectively, represent different groups of input features: spectral features, GF, TF, GF combined with TF, and all features along with the spectral features. d is the dimension associated with the best performance.

FIGURE 6: Classification maps of the IP dataset by the multiple kernel method; (a)–(e), respectively, represent different groups of input features: spectral features, GF, TF, GF combined with TF, and all features along with the spectral features. d is the feature dimension for input.

the average kappa index, execution times, and the best d for the DR scheme in scenario 3. The parameters are also optimized by the training samples.

3.3.1. IP Scenes. Figures 4–6 present classification results based on different groups of features associated with 3 combining strategies. The accuracies can be seen in Table 6. In each figure, GF + TF has the best performance both in accuracy and visual perspectives. Figures 4(a), 5(a), and 6(a) yield the most misclassifications, and Figures 4(b), 5(b), and 6(b) show better results than do Figures 4(a), 5(a), and 6(a) with less features. Figures 4(c), 5(c), and 6(c) have much better results than Figures 4(b), 5(b), and 6(b) have, owing to better textural features. With the help of the integration of GF and TF, Figures 4(d), 5(d), and 6(d) have the fewest classification errors. It has to be mentioned that there is no need to add initial hyperspectral features in the proposed method, because the accuracy reduces when the spectral features are added referring to the results in Figures 4(e), 5(e), and 6(e). As a result, 60 input features yield the best result. It can also be discovered from Table 6 that nonlinear LE dimensionality reduction tools spend more time when dealing with the combined features. According to the comparison of three organization schemes, it can be concluded that the overall calculation amount of the nonlinear DR procedure is greater than that of the classification procedure with relatively higher inputted dimensions. Also, we find that Figure 4 outperforms Figures 5 and 6. So the dimensionality framework or multiple kernel method has not given rise to a better classification

TABLE 6: Classification results by different features for IP dataset, including OA, kappa index, and the execution time.

IP		SPECT	GF	TF	GF + TF	TF + GF + SPECT
Direct combine	OA (%)	81.30	80.49	94.59	96.39	96.02
	Kappa	0.784	0.775	0.938	0.959	0.954
	Time (s)	3.3	0.99	1.46	1.77	4.18
LE framework	OA (%)	62.31	78.42	90.98	93.76	79.36
	Kappa	0.568	0.751	0.897	0.929	0.764
	Best d	50	20	35	55	40
Multiple kernel	Time (s)	4.23	2.44	2.73	3.38	4.06
	OA (%)	81.30	85.89	93.72	94.36	92.88
	Kappa	0.784	0.838	0.928	0.935	0.918
	Time (s)	3.3	4.48	4.62	6.91	7.48

result than the direct way of combination. In addition, it is not easy to find a desirable dimension in the dimensionality reduction process (Figure 7).

3.3.2. SL Scenes. The SL dataset has more feature numbers and a higher resolution. Extremely high accuracy is yielded

FIGURE 7: Relationship of d and accuracy in the IP dataset ($d \leq 60$).

FIGURE 8: Classification maps of the SL dataset with directly combined features; (a)–(e), respectively, represent different groups of input features: spectral features, GF, TF, GF combined with TF, and all features along with the spectral features. d is the feature dimension for input.

FIGURE 9: Classification maps of SL dataset with features under manifold based framework; (a)–(e), respectively, represent different groups of input features: spectral features, GF, TF, GF combined with TF, and all features along with the spectral features. d is the dimension associated with the best performance.

with the proposed scenario, and similar regulations in IP dataset can be found. The results are listed in Figures 8–10. Table 7 lists the accuracies by different scenarios. Among the 16 classes, it is challenging to distinguish class 8 and class 15, while the GF+TF strategy shows the perfect accuracy rate.

However, no matter what kinds of combination strategy are employed, the results have not varied a lot for one group of features and often remain in a certain range. The accuracies change more obviously and regularly with the variation of the input features (Figure 11). So we can suggest that the

selection of features is more important than combining approaches in hyperspectral image classification.

However, no matter what kinds of combination strategy are exploited, the results do not vary a lot within a fix group of features. The accuracies just remain in a certain scope. The accuracies change more obviously and regularly with the variation of input features (Figure 11). So we can suggest that the selection of features is more important than the combination approaches in hyperspectral image classification. In addition, we can conclude that the proposed strategy applies to different combination schemes.

TABLE 7: Classification results by different features for SL dataset, including OA, kappa index, and execution time.

SL		SPEC	GF	TF	GF + TF	TF + GF + SPEC
Direct combine	OA (%)	89.68	89.89	97.47	98.95	98.5
	Kappa	0.885	0.887	0.972	0.988	0.983
	Time (s)	10.9	3.88	5.98	6.63	9.7
LE framework	OA (%)	83.28	89.86	95.57	98.18	95.69
	Kappa	0.813	0.887	0.951	0.980	0.952
	Best d	35	20	35	35	45
	Time (s)	11.21	5.42	7.82	7.29	9.743
Multiple kernel	OA (%)	89.68	90.01	96.83	97.74	97.34
	Kappa	0.885	0.899	0.964	0.974	0.970
	Time (s)	10.9	3.88	5.98	6.63	15.7

FIGURE 10: Classification maps of the SL dataset with directly combined features; (a)–(e), respectively, represent different groups of input features: spectral features, GF, TF, GF combined with TF, and all features along with the spectral features. d is the feature dimension for input.

TABLE 8: Classification results by different features for PU dataset, including OA, kappa index, and execution time.

PU		SPEC	GF	TF	GF + TF	TF + GF + SPEC
Direct combine	OA (%)	83.61	85	92.78	98.6	98.48
	Kappa	0.775	0.797	0.904	0.982	0.980
	Time (s)	8.91	4.36	2.38	6.6	8.5
LE framework	OA (%)	80.89	83.66	91.78	98.44	97.69
	Kappa	0.741	0.778	0.89	0.979	0.969
	Best d	25	15	40	55	45
	Time (s)	9.23	5.48	6.62	7.91	8.48
Multiple kernel	OA (%)	83.61	84.91	96.11	98.08	97.86
	Kappa	0.775	0.796	0.948	0.975	0.972
	Time (s)	8.91	4.76	5.96	7.03	12.05

3.3.3. *PU Scenes*. Among the three datasets, PU has the highest resolution and the most pixels. Table 8 lists the accuracies by different scenarios. In this high-resolution dataset, the complementary properties of GF and TF have a more apparent representation. For example, in Figures 12, 13, and 14(a) and 14(b), with GF, the misclassification occurs frequently

between class 2 (grass) and class 6 (bare soil), while TF discriminates the class pair well; class 1 (asphalt road) and class 4 (tree) are challenging classes for TF because of the close location, but GF or spectral features perform well between the 2 classes. Figure 15 presents a direct perspective of the complementary properties of different feature groups.

FIGURE 11: Accuracy change range with features and strategies on the SL dataset.

FIGURE 12: Classification maps of the PU dataset with features under a manifold-based framework; (a)–(e), respectively, represent different groups of input features: spectral features, GF, TF, GF combined with TF, and all features along with the spectral features. d is the feature dimension for input.

FIGURE 13: Classification maps of the PU dataset by the multiple kernel method; (a)–(e), respectively, represent different groups of input features: spectral features, GF, TF, GF combined with TF, and all features along with the spectral features. d is the dimension associated with the best performance.

FIGURE 14: Classification maps of the PU dataset with directly combined features; (a)–(e), respectively, represent different groups of input features: spectral features, GF, TF, GF combined with TF, and all features along with the spectral features. d is the feature dimension for input.

TABLE 9: The effect of autoweighting method in the GF + TF strategy.

Method	IP		SL		PU	
	Directly combine	LE framework	Directly combine	LE framework	Directly combine	LE framework
Autoweight	96.39	93.76	98.95	98.18	98.6	98.44
Without weight	95.87	93.01	98.47	97.61	97.23	96.9

As has been discussed before, different groups of features have their specific properties. Different feature extraction algorithms yield different feature numbers, so it seems necessary to add weight factors to different groups of features in different circumstances. We can find the improvement of the autoweighting method in the proposed GF + TF strategy in Table 9.

3.4. Discussions. It is necessary to discuss the experimental results on the hyperspectral datasets.

The summary of the experiments with different DR schemes is as follows: textural features combined with a low-dimensional spectral features prove to be a more appropriate strategy according to experiments in Section 3.2, especially when textural features perform well and have less numbers. When textural features do not work well, it is not always the case. However, great efforts have been made to extract favorable textural features for hyperspectral image expression, so for most circumstances, the textural features we select are empirically superior. As a result, scenario 4 in Table 4 is recommended.

The summary of the experiments with the combination scenarios is as follows: for all the datasets, the GF + TF method performs best both in accuracy and visual perspectives; “global features” and “textural features” present clear complementary properties in the classification results. The GF + TF combination with only 60 input features works well without the incorporation of the original hyperspectral bands. The weight estimation method proves to be effective when dealing with multiple features. In addition, the number of the input features of the GF + TF strategy is independent of the band number of the images or the types of sensors.

The summary of the experiments with different feature combination frameworks is as follows: the results yielded by different combination strategies do not vary a lot if the input

features are confirmed; compared with the “framework” strategy, the “direct combination” strategy with the only 60 GF + TF input features not only performs better but also avoids great calculating amount and large number of dimensions; multiple kernel method and nonlinear DR framework sometimes lead to desirable results, but the best d is hard to determine, which can be reflected both in our experiments and in other studies.

We can conclude by the experimental results and the discussions above that what kinds of features to be combined influences the classification results to a larger extent than the strategies of integrating multiple features. In addition, multiple features with complementary properties lead to good classification results. In this paper, we exploit 4 different features for combination, including linear and nonlinear “global features” and filter-based and statistics-based “textural features,” which ensure the complementary and diverse properties.

Pattern recognition has been applied in many fields. Hyperspectral remote sensing image classification is a peculiar application and has its own characteristics. Compared with image recognition (like in [9, 13]) and medicine field (such as gene and protein classification [7]), remote sensing image classification problems have a relatively lower dimension or at least a smaller d/N (d represents dimension and N represents labeled samples). According to an overall statistics in [15], when $d \gg N$ (like some cases in body, face, object image recognition, or gene classification), the accuracy appears low (Figure 16). In this case, a DR algorithm will help to improve both the results and the classification efficiency. However, in the case of $N > d$, the accuracy appears pretty well and a DR framework does not lead to evidently better results. In addition, it is not easy to find the best d in different cases according to our experimental results and Figure 7. In fact, researchers have not found an effective way of

FIGURE 15: Complementary properties of PC and LE features for all classes in the PU dataset.

FIGURE 16: Hughes curves labeled by common application of remote sensing image classification.

confirming the most desirable d . The latter case (a smaller d/N) may be the common application of hyperspectral remote sensing image classification in the temporary period of development (Figure 16). With the development of the hyperspectral sensors, the bands may increase; thus, the area of the red frame in Figure 16 may grow. However, the proposed GF + TF strategy may still be practical for use because it is independent of the bands of the sensors.

4. Conclusions

Hyperspectral sensors provide more details in spectra; however, problems are yielded along with the advantages. One is the high-dimension problem, and the other is the possibility of extracting multiple features from the hyperspectral images. In this paper, a systematic research is made to find an appropriate strategy to deal with classification problems

of multiple features. Then, we further exploit the complementary features to improve hyperspectral image classification performance. Experiments on 3 hyperspectral datasets suggest that the scheme GF + TF is effective. The main contribution can be concluded as follows: first, based on the experiments, we suggest DR algorithms work better as feature acquirement methods than just reducing the dimensions. The paper further selects features from a different perspective in multiview problem. In the previous work, there have been feature combination ideas characterized by “spectral and nonspectral” or more recent schemes like “linear and nonlinear.” However, we present “global and nonglobal” strategy and take linear and nonlinear global features (like PC, ISO-MAP, and LE) as a portion of features for the first time. We have also got the conclusion that features with different types are more likely to have complementary properties. Second, concluded from the experimental results, feature selection

proves to be more important than how to organize multiple features in hyperspectral image classification problems. Third, a systematic research has been made on the combination frameworks of multiple features, including direct combination, manifold learning, and multiple kernel method. We have found that complex methods like manifold framework and multiple kernel do not lead to the increase in accuracy; instead, the direct combine strategy with an autoweight calculation performs the best. Finally, we have compared hyperspectral image classification problems with other applications of pattern recognition and clearly analyzed the characteristics of the former.

As future work, we will continue to find more complementary features for integration in hyperspectral remote sensing image classification based on the experiments. For example, shape features have been widely developed recently and have not been considered in the study. In addition, we will further discuss the internal redundancy of each group of features.

Conflicts of Interest

The authors declare no conflict of interests.

Authors' Contributions

Yuntao Ma is the main author who proposed the basic idea, completed the experiments, and carefully revised this manuscript. Ruren Li and Guang Yang provided the useful suggestions on designing the approaches involved in our proposed strategy. Lishuang Sun and Jingli Wang helped to modify the manuscript.

Acknowledgments

The study was funded by the National Natural Science Foundation of China "Study on environmental impact mechanism of surface deformation monitoring in open pit mine and inversion methods under low coherence environment based on GB-SAR" (no. 51774204) and "TGRA Forest Cover and Dynamic Change Detection Based on Time Series Remote Sensing Images" (no. 2014QC018).

References

- [1] H. Song and Y. Wang, "A spectral-spatial classification of hyperspectral images based on the algebraic multigrid method and hierarchical segmentation algorithm," *Remote Sensing*, vol. 8, no. 4, p. 296, 2016.
- [2] M. Diani, N. Acito, M. Greco, and G. Corsini, "A new band selection strategy for target detection in hyperspectral images," *Lecture Notes in Computer Science*, vol. 5179, pp. 424–431, 2008.
- [3] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "A spatial-spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recognition*, vol. 45, no. 1, pp. 381–392, 2012.
- [4] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 466–479, 2005.
- [5] S. Z. Wang, Y. Li, and W. Cheng, "Distributed classification of localization attacks in sensor networks using exchange-based feature extraction and classifier," *Journal of Sensors*, vol. 2016, Article ID 8672305, 18 pages, 2016.
- [6] P. H. Hsu, "Feature extraction of hyperspectral images using wavelet and matching pursuit," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, no. 2, pp. 78–92, 2007.
- [7] G. Lee, C. Rodriguez, and A. Madabhushi, "Investigating the efficacy of nonlinear dimensionality reduction schemes in classifying gene and protein expression studies," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 5, no. 3, pp. 368–384, 2008.
- [8] G. Yang and S. Fang, "Improving remote sensing image classification by exploiting adaptive features and hierarchical hybrid decision trees," *Remote Sensing Letters*, vol. 8, no. 2, pp. 156–164, 2016.
- [9] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1299–1313, 2009.
- [10] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 3, pp. 879–893, 2012.
- [11] J. Li, X. Huang, P. Gamba et al., "Multiple feature learning for hyperspectral image classification," *IEEE Transactions on Geoscience & Remote Sensing*, vol. 53, pp. 1592–1606, 2015.
- [12] A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet, "Simple MKL," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.
- [13] T. Xia, D. Tao, T. Mei, and Y. Zhang, "Multiview spectral embedding," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 6, pp. 1438–1446, 2010.
- [14] H. Liu, J. Yang, H. Yang, and F. Yi, "Soft sensor of vehicle state estimation based on the kernel principal component and improved neural network," *Journal of Sensors*, vol. 2016, Article ID 9568785, 8 pages, 2016.
- [15] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 55–63, 1968.
- [16] S. Sonnenburg, G. Rätsch, C. Schäfer, and B. Schölkopf, "Large scale multiple kernel learning," *Journal of Machine Learning Research*, vol. 7, pp. 1531–1565, 2006.
- [17] M. Gönen and E. Alpaydm, "Multiple kernel learning algorithms," *Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, 2011.
- [18] P. Du, A. Samat, P. Gamba, and X. Xie, "Polarimetric sar image classification by boosted multiple-kernel extreme learning machines with polarimetric and spatial features," *International Journal of Remote Sensing*, vol. 35, no. 23, pp. 7978–7990, 2014.
- [19] Q. Zhang, Y. Tian, Y. Yang, and C. Pan, "Automatic spatial-spectral feature selection for hyperspectral image via discriminative sparse multimodal learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 261–279, 2015.
- [20] C. F. Tsai and Y. C. Hsiao, "Combining multiple feature selection methods for stock prediction: union, intersection, and multi-intersection approaches," *Decision Support Systems*, vol. 50, no. 1, pp. 258–269, 2010.

- [21] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933.
- [22] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
- [23] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [24] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2006.
- [25] C. Chen, W. Li, H. Su, and K. Liu, "Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine," *Remote Sensing*, vol. 6, no. 6, pp. 5795–5814, 2014.
- [26] S. C. H. Hoi, R. Jin, P. Zhao, and T. Yang, "Online multiple kernel classification," *Machine Learning*, vol. 90, no. 2, pp. 289–316, 2013.
- [27] A. Mohan, G. Sapiro, and E. Bosch, "Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, pp. 206–210, 2007.
- [28] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [29] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimension reduction via local tangent space alignment," *SIAM Journal on Scientific Computing*, vol. 26, pp. 313–338, 2005.
- [30] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1700–1715, 2007.
- [31] M. Hanta-Kasari, J. Parkkinen, and T. Jaaskelainen, "Generalized co-occurrence matrix for multispectral texture analysis," in *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria, August*, pp. 25–29, IEEE Computer Society Press, Los Alamitos, USA, 1996.
- [32] Y. O. Ouma, J. Tetuko, and R. Tateishi, "Analysis of co-occurrence and discrete wavelet transform textures for differentiation of forest and non-forest vegetation in very-high-resolution optical-sensor imagery," *International Journal of Remote Sensing*, vol. 29, no. 12, pp. 3417–3456, 2008.
- [33] Y. Zhou, *Nonlinear Feature Extraction of Hyperspectral Remote Sensing Data Based on a Fast Manifold Learning Strategy*, PhD, Wuhan University, China, 2013.
- [34] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [35] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society Series B-Statistical Methodology*, vol. 36, pp. 111–147, 1974.