A Novel Hierarchical Semisupervised SVM for Classification of Hyperspectral Images

Zhenfeng Shao, Lei Zhang, Xiran Zhou, and Lin Ding

***Abstract*—This letter presents a novel hierarchical semisupervised support vector machine (SVM) for classification of hyperspectral images. The method exploits the wealth of unlabeled samples by means of their cluster features. The method learns a suitable framework for classifying cluster features by a semisupervised SVM and thus makes use of advantages of clustering and classification. Experimental results demonstrate that the proposed classification method is effective for hyperspectral image classification when a few labeled samples are available. Another advantage of the proposed method is that the hierarchical structure can simultaneously take clustering and classification information into consideration.**

***Index Terms*—Hyperspectral image classification, semisupervised learning, support vector machine (SVM).**

I. Introduction

SINCE the detailed spectral information makes it possible to discriminate different classes of objects through spectral feature similarity measuring, wealthy spectral information of hyperspectral images holds great value in image classification and recognition. However, these applications of hyperspectral images are generally limited by accuracy of image classification [1]–[3]. In order to obtain reliable image classification results, labeled samples required 10–100 times more than feature dimensions in conventional methods [4]. Therefore, time and labor cost for collecting ground reference data as labeled samples are often high [5]. The quantity of available training data, which is known as Hughes phenomenon [6], and the quality of training data, including mixed pixels, result in ill-posed image classification problems, which are very critical in analysis of hyperspectral images [7].

One of the main difficulties in supervised methods for hyperspectral image classification is that their learning process heavily depends on the quality of the training data set. In order to deal with this problem, some semisupervised methods are proposed [8]–[10]. A semisupervised support vector machine (SVM) has been successfully employed in hyperspectral image classification [11]–[13]. The SVM is appealing in the remote sensing field due to its good generalization ability with limited training samples and globally optimal performance with high dimensions.

However, SVM classifiers can only use the labeled samples to provide predicted classes for new samples. In order to make data structure into consideration during the classification process, some clustering algorithms have been used gradually [14], [15]. Although these methods demonstrated good results in hyperspectral image classification, there are still some problems while adopting the semisupervised SVM or clustering methods separately. On the one hand, semisupervised SVMs do not consider relationship between different clusters and classes. On the other hand, semisupervised SVMs are not favored by data cluster features provided by clustering.

In this letter, we propose a hierarchical semisupervised classification method based on SVM. The proposed approach is referred to as a two-stage algorithm. First, cluster features of images are obtained using a new clustering method called kernel-spectral fuzzy C-means (KSFCM). Then, once the clustering is performed, the semisupervised SVM is used to classify cluster features. The classification process is implemented as a hierarchical system. Finally, QaR tree is introduced to optimize the proposed method, maximizing efficiency of this algorithm. This approach owns both advantages of clustering and classification, which makes classification results effective and robust.

The remainder of this letter is organized as follows. Section II presents the rationale and formulation of the proposed hierarchical semisupervised SVM. Experiments and analyses are demonstrated in Section III. Finally, Section IV draws the conclusion of the work.

II. Hierarchical Semisupervised SVM

Here, we describe the hierarchical semisupervised SVM by presenting first the rationale and then the mathematical formulation.

A. Rationale of the Proposed Method

The new method aimed at resolving the following problems.

1) Kernel fuzzy C-means (KFCM) is proposed by Zhang and Chen [16] to handle high dimensionality of the data. In comparison to other clustering algorithms, this algorithm provides more robust to noise and less sensitive to cluster shapes [17]. Different from the KFCM algorithm with equal partition trend for data sets, we introduce the

Manuscript received December 18, 2013; accepted January 16, 2014. Date of publication February 7, 2014; date of current version April 23, 2014. This work was supported by the National Basic Research Program of China under Grant 2010CB731801, by the National Natural Science Foundation of China under Grant 61172174, by the National Science and Technology Specific Projects under Grant 2012YQ1601850 and Grant 2013BAH42F03, and by the Program for New Century Excellent Talents in University under Grant NCET-12-0426.

The authors are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: shaozhenfeng@whu.edu.cn; xysyleilei@126.com; zhouxiranjuven@163.com; dinglin@whu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2014.2302034

weighted information between cluster centers and their adjacent samples into the standard KFCM algorithm in this letter. Since the samples with similar spectral feature may belong to the same class, we utilize the spectral angle of samples to determine the weight values. If the spectral angle between two samples is very small, their weight will be large as they may belong to the same class. Therefore, the KSFCM proposed in this letter takes spectral feature information of remote sensing images into consideration in the cluster center computation process of KFCM.

2) From KSFCM, we can obtain cluster features of the images, which represent the inherent structure and spectral relationship of samples from clustering perspective. Actually, the results of KSFCM could be considered as cluster features since the clustering result reveals the cluster structure. Then, the semisupervised SVM is used to classify cluster features. Because final class information of the clustering method and the semisupervised SVM should be consistent, the difference between KSFCM and the semisupervised SVM should be minimized. Thus, we establish a semisupervised SVM framework with several judgment factors to make the objective function reach its maximum for obtaining high classification accuracy.

B. Formulation of the Proposed Method

Similar to KFCM, KSFCM clusters the data set by optimizing the following object function under the same constraints:

$$J(X;S,U,C)=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^{m}s_{ij}\left\|\phi(x_j)-\phi(c_i)\right\|^2$$
$$\text{s.t.}\begin{cases}\sum_{i=1}^{c}u_{ij}=1, & \forall j\in\{1,\ldots,n\}\\ \sum_{j=1}^{n}u_{ij}>0, & \forall i\in\{1,\ldots,c\}\end{cases}\tag{1}$$

where $X=\{x_1,\ldots,x_n\}$ is the original hyperspectral data set, $C=\{x_1,\ldots,x_c\}$ presents the cluster centers, n denotes the number of samples, c denotes the number of clusters, and S is the weighting matrix whose values are obtained from the spectral angle information between cluster centers and their adjacent samples, i.e.,

$$S_{ij}=\begin{cases}\dfrac{\sum_{i=1}^{b}\phi(x_i)\phi(x_j)}{\left[\sum_{i=1}^{b}\phi(x_i)^2\right]^{1/2}\left[\sum_{i=1}^{b}\phi(x_j)^2\right]^{1/2}}, & \phi(x_i)\neq\phi(x_j)\\ 1, & \phi(x_i)=\phi(x_j)\end{cases}\tag{2}$$

where b stands for the band number of the remote sensing image. In (1), U is the fuzzy partition matrix whose entries are membership degrees, that is, the degree of similarity between pixels and cluster centers. The parameter $m\in[1,\infty]$ is the weighting exponent determining the fuzziness of the clusters. $\|\cdot\|^2$ denotes the Euclidean norm, which is obtained by the following equation:

$$\begin{aligned}&\left\|\phi(x_j)-\phi(c_i)\right\|^2\\ &\quad=\langle\phi(x_j),\phi(x_j)\rangle+\langle\phi(c_i),\phi(c_i)\rangle-2\langle\phi(x_j),\phi(c_i)\rangle\\ &\quad=k(x_j,x_j)+k(c_i,c_i)-2k(x_j,c_i).\end{aligned}\tag{3}$$

By adapting the Gaussian kernel as a kernel function, the objective function of KSFCM can be reformulated as follows:

$$J(X;S,U,C)=2\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^{m}s_{ij}\left(1-k(x_j,c_i)\right).\tag{4}$$

Then, the cluster centers and fuzzy partition matrix are given by the following equations:

$$u_{ij}=\frac{\left[S_{ij}\left(1-k(x_j,c_i)\right)\right]^{-\frac{1}{m-1}}}{\sum_{l=1}^{C}\left[S_{ij}\left(1-k(x_j,c_l)\right)\right]^{-\frac{1}{m-1}}}$$
$$c_i=\frac{\sum_{j=1}^{n}u_{ij}^{m}S_{ij}k(x_j,c_i)x_j}{\sum_{j=1}^{n}u_{ij}^{m}S_{ij}k(x_j,c_i)}.\tag{5}$$

Cluster features are used to establish the connection of clustering and classification to realize advantageous complementarities. Cluster features r_i of sample x_i stand for the membership degree of each sample, i.e.,

$$r_i=\{u_{1i},u_{2i},\ldots,u_{ki}\}.\tag{6}$$

Since r_i is the membership degree vector of each sample according to membership matrix U; thus, $e^T r_i=1$, $i\in\{1,\ldots,n\}$, and e is the unit vector, $r_{ki}\geq 0$, $k\in\{1,\ldots,c\}$.

After obtaining cluster features of hyperspectral data by KSFCM, the semisupervised SVM is used to classify cluster features. The whole process of the semisupervised SVM algorithm is shown as follows.

Step 1: The input sample data set is $X=\{X_l,X_u\}$. X_l is the labeled sample set, and X_u is the unlabeled sample set. Cluster centers and limited labeled samples are processed as the labeled data set, and other cluster features as the unlabeled data set.
Step 2: Using the SVM train X_l to get classifiers C1 and C2. C1 uses the polynomial kernel; C2 uses the radial basis function.
Step 3: Classifiers C1 and C2 forecast classes of X_u, respectively, and get forecasting results P1 and P2.
Step 4: Compare P1 and P2 and make results with high creditability join into the X_l, update X_l.
Step 5: Back to step 2. If all samples are classified, exit loop.

In the aforementioned steps, the default polynomial order is set as 4, and the default penalty parameter is given as 2 in classifier C1. The radial basis function kernel with length scale $\sigma\in[10^{-2},\ldots,10^{2}]$ and the penalty parameter $C\in[1,\ldots,10^{3}]$ are used in classifier C2.

Then, in this letter, we propose judgment factors, including clustering evaluation, classification difference, classification consistence, and sample difference, to build a semisupervised SVM framework, which can make full use of limited labeled samples and many unlabeled data.

Clustering evaluation CuE helps ensure clustering accuracy. The higher clustering evaluation leads to better clustering, i.e.,

$$\text{CuE} = \frac{1}{\sum_i^n \sum_k^c r_{ki}^2 \sqrt{\|x_i - c_k\|}}. \tag{7}$$

The fuzzy membership degree is defined by both distance and spectral angle between samples and their cluster centers. Applying the spectral angle to clustering can accurately locate cluster centers and is superior to KFCM clustering.

In order to guarantee class consistency, classification difference CD is used to judge the difference between KSFCM and the semisupervised SVM, i.e.,

$$\text{CD} = \sum_i^n \text{KL}\left(p(c_i|x_i)\,\|p(c_i|r_i)\right). \tag{8}$$

KL stands for Kullback–Leibler divergence. $p(c_i|x_i)$ and $p(c_i|r_i)$ are cluster posterior probabilities of cluster membership, which are defined as follows:

$$p(c_i|x_i) = \frac{\text{dist}(x_i, c_i)}{\sum\limits_{i=1}^n \text{dist}(x_i, r_i)} \tag{9}$$

$$p(c_i|r_i) = \frac{\text{dist}(r_i, c_i)}{\sum\limits_{i=1}^n \text{dist}(x_i, r_i)}. \tag{10}$$

Through the kernel substitution, the distance is obtained as follows:

$$\begin{aligned}\text{dist}(x_i, c_i) &= \|\phi(x_i) - \phi(c_i)\|^2 \\ &= \langle\phi(x_i), \phi(x_i)\rangle + \langle\phi(c_i), \phi(c_i)\rangle - 2\langle\phi(x_i), \phi(c_i)\rangle \\ &= k(x_i, x_i) + k(c_i, c_i) - 2k(x_i, c_i).\end{aligned} \tag{11}$$

$K(x, y)$ is taken as the radial basis function kernel due to its robustness, i.e.,

$$K(x, y) = \exp\left(\frac{-\|x - y\|^2}{\sigma^2}\right). \tag{12}$$

Thus, the distance can be written as

$$\text{dist}(x_i, c_i) = 2 - 2\,\exp\left(\frac{-\|x_i - c_i\|^2}{\sigma^2}\right) \tag{13}$$

where σ is the kernel parameter and affects the classification result, which is defined as

$$\sigma^2 = \sum_{i=1}^n \|x_i - \bar{x}\|^2 \quad \bar{x} = \frac{1}{n}\sum_{i=1}^n x_i. \tag{14}$$

Classification consistence CaC is used to judge classification accuracy according to the ground truth data set, i.e.,

$$\text{CaC} = \sum_i^n p(y_i|x_i) + \sum_i^c p(y_i|r_i). \tag{15}$$

$Y = \{y_1, \ldots, y_n\}$ is the ground truth data set. $p(y_i|x_i)$ and $p(y_i|r_i)$ are posterior probabilities of class membership, which are defined as follows:

$$p(y_i|x_i) = \frac{\text{dist}(x_i, y_i)}{\sum\limits_{i=1}^n \text{dist}(x_i, r_i)} = \frac{\|\phi(x_i) - \phi(y_i)\|^2}{\sum\limits_{i=1}^n \|\phi(x_i) - \phi(r_i)\|^2} \tag{16}$$

$$p(y_i|r_i) = \frac{\text{dist}(y_i, r_i)}{\sum\limits_{i=1}^n \text{dist}(x_i, r_i)} = \frac{\|\phi(y_i) - \phi(r_i)\|^2}{\sum\limits_{i=1}^n \|\phi(x_i) - \phi(r_i)\|^2}. \tag{17}$$

Sample difference SD can judge similarities of samples in the same class, which is calculated by the Euclidean distance, i.e.,

$$\text{SD} = \sum_{ij}^n \sqrt{\|p(c_j|x_i) - p(c_i|x_j)\|^2 + \|p(c_j|r_i) - p(c_i|r_j)\|^2}. \tag{18}$$

The semisupervised SVM framework makes the clustering information guide classification process depending on cluster features that were obtained by the clustering algorithm. In this letter, we build a semisupervised SVM framework by CuE, CD, CaC, and SD, which can make the best of clustering information. The objective function is as follows:

$$\max S = \lambda_1\text{CuE} - \lambda_2\text{CD} + \lambda_3\text{CaC} - \lambda_4\text{SD}. \tag{19}$$

λ_1, λ_2, λ_3, and λ_4 stand for weights of judgment factors, ranging from 0 to 1, and $\sum_{i=1}^4 \lambda_i = 1$. The goal function is to get the maximum value under the constraints of judgment factors. CuE builds constraints of KSFCM to ensure that cluster features have the largest representation of internal data structures. CD compares differences between KSFCM and the semisupervised SVM to limit classification errors. CaC confirms classification consistency between KSFCM and semisupervised SVM through ground truth data set. SD is the evaluation factor of this method, which is used to judge the effectiveness of classification results.

If all cluster features obtained from KSFCM are embedded into the semisupervised SVM framework simultaneously, it will create data redundancy. QaR tree (Quad and R* tree) [18], as the postprocessing for image classification, is introduced to optimize the semisupervised SVM framework to improve the efficiency of this algorithm.

We use QaR tree to obtain different clustering regions in the KSFCM algorithm. Then different classification regions will be obtained by the semisupervised SVM, respectively. DR_{KSFCM} stands for image regions of KSFCM by QaR tree. DR_{SVM} stands for image regions of the semisupervised SVM by QaR tree. Comparing DR_{KSFCM} and DR_{SVM} to obtain different classification regions DR' between KSFCM and semisupervised SVM classification results. Finally, DR' is imbedded into the semisupervised SVM framework to obtain the optimization classification.

TABLE I
ACCURACY OF DIFFERENT CLASSIFICATION METHODS WITH 16 CLASSES

Class	Training samples	KSFCM		SVM		SVM+ cluster features		our method	
		PA	UA	PA	UA	PA	UA	PA	UA
Alfalfa	31	100	87.5	100	100	100	100	100	100
Corn-notill	952	32.3	25.6	48.3	66.7	80	61.5	85.7	85.7
Corn-mintill	553	64.6	69.5	57.1	66.7	50	75	75	75
Corn	158	81.4	39.0	100	80	100	100	100	100
Grass-pasture	322	59.8	84.7	100	78.6	75	100	75	100
Grass-trees	487	86.8	84.7	94.4	94.4	100	85.7	100	100
Grass-pasture-mowed	19	100	83.3	100	100	100	100	100	100
Hay-windrowed	319	98.0	100	100	100	100	100	100	100
Oats	13	92.9	50	100	100	90	100	100	100
Soybean-notill	648	54.8	55.6	72.2	65	60	60	60	75
Soybean-mintill	1637	63.0	77.9	70.9	66.7	100	81.8	100	77.8
Soybean-clean	395	48.7	58.3	77.8	63.6	66.7	85.7	100	100
Wheat	137	98	96.1	100	100	100	100	100	100
Woods	843	96	71.3	100	100	100	100	100	100
Buildings-Grass-Trees-Drives	257	52	86.7	83.3	83.3	100	100	100	100
Stone-Steel-Towers	62	100	100	100	100	100	100	100	100
overall accuracy		68.05%		78.92%		88.46%		92.96%	
kappa coefficient		0.6479		0.7681		0.8759		0.9245	

PA stands for production accuracy, UA stands for user accuracy and OA stands for overall classification accuracy.

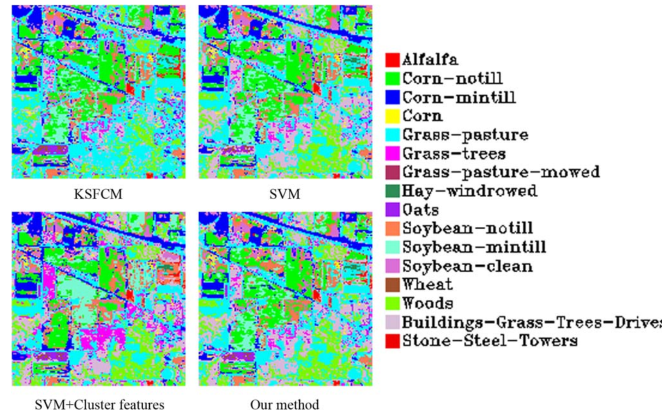

Fig. 1. Classification maps with 16 classes.

III. EXPERIMENT

The experimental data source came from the AVIRIS hyperspectral image of Indiana. It consists of 145 × 145 pixels by 220 bands of radiance data with about two-thirds agriculture and one-third forest or other natural perennial vegetation. We have reduced the number of bands to 200 by removing bands covering the region of water absorption. The full scene with 16 classes has been studied by Gualtieri and Chettri [19] that allows us a comparison to the SVM classifier.

Table I compares the results of the KSFCM, SVM classifier, and our method with 16 classes, and our method outperforms other classifiers. To investigate the performances of our method when limited labeled samples are available, the numbers of training sample sets are also shown in Table I. The remaining samples are used as independent test sets. Fig. 1 shows the corresponding classification maps.

The production accuracy and user accuracy per method are shown in Table I, from which we can make the following analyses.

1) In the KSFCM method, the production accuracy and user accuracy are generally low. In addition, there are several significant gaps among different class accuracy. For example, some production accuracy of Corn-mintill and the user accuracy of Corn-mintill, Grass-pasture, Soybean-mintill, and Buildings–Grass–Trees–Drives are higher than the SVM classifier. That is to say, KSFCM has high classification accuracy for some special classes.
2) Through exploiting the SVM method, most classes have higher production accuracy and user production accuracy than KSFCM. Although the overall accuracy of SVM is higher than KSFCM, SVM has low classification accuracy for some special classes involving Corn-mintill, Grass-pasture, and Soybean-mintill. Thus, the table shows that KSFCM and SVM have different advantages in different classes. Moreover, the aforementioned result also implies that KSFCM and SVM can compensate for each other's weaknesses.
3) In the SVM method based on cluster features directly, the overall accuracy is higher than KSFCM and SVM. Additionally, the gaps between different types of class accuracy are lower than KSFCM and SVM classifiers. Thus, using cluster features in the classifier is meaningful in some degree. However, there are some classes such as Corn-mintill, Grass-pasture, Oats, Soybean-notill, and Soybean-clean that have low accuracy than the SVM method based on original data. From Table I, it is easy to find that if we use SVM to classify cluster features directly, although most classes will obtain advanced classification accuracy, the classification accuracy of some classes may be influenced by cluster features and lead to low classification accuracy. Thus, we need to find certain constraints to make full use of advantages of both KSFCM and SVM classifiers.
4) In our method, both production accuracy and user accuracy are higher than other methods. Moreover, each class has high accuracy. The semisupervised SVM framework can own both advantages of clustering and classification, which makes clustering and classification results effective and robust.

From Table I and Fig. 1, the semisupervised SVM framework with several judgment factors proposed in this letter raises the classification accuracy effectively and can make full use of the respective advantages of clustering and classification, which can avoid the difficulties in gathering training samples of high quality.

From the 16 different land-cover classes available in the original ground truth, we use the 8 largest land-cover classes to

TABLE II
ACCURACY OF DIFFERENT CLASSIFICATION METHODS WITH 8 CLASSES

	KSFCM		SVM		SVM+ cluster features		our method	
	PA	UA	PA	UA	PA	UA	PA	UA
Corn-notill	64.9	82.8	95.6	87.8	89.6	93.3	98	96.8
Corn-mintill	64.3	90	100	86.1	100	100	100	100
Grass-trees	100	91.7	97.8	68.6	97.1	97.8	99.3	100
Soybean-notill	91.7	52.4	43.7	83	61	84.4	99.1	96.8
Soybean-mintill	78.3	81.8	76.7	71	70.9	42.1	91.9	97.5
Soybean-clean	83.3	62.5	99.6	98.6	100	100	100	100
Woods	100	100	98.3	90.7	100	99.4	100	100
Buildings-Grass-Trees-Drives	80	100	79.7	92.6	90.7	87	96.6	97.4
OA	78.74%		86.72%		89.68%		98.52%	
Kappa	74.87%		84.67%		0.881		98.28%	

PA stands for production accuracy, UA stands for user accuracy and OA stands for overall classification accuracy.

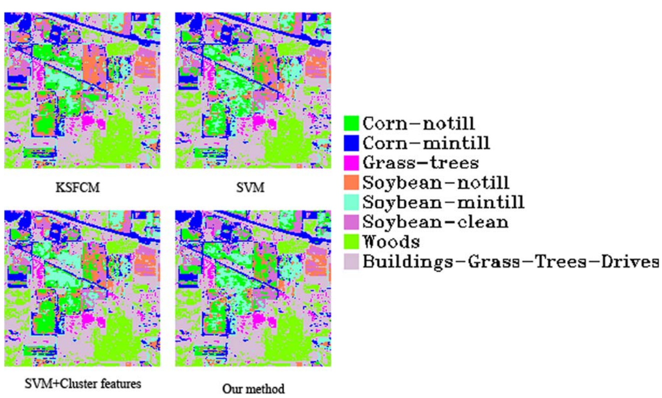

Fig. 2. Classification maps with 8 classes.

verify the validity of the proposed algorithm. Table II compares the results of the KSFCM, SVM classifier, and our method, respectively, with 8 classes, and our method outperforms than other classifiers, as well. From Table II and Fig. 2, we can make the following analyses.

1) If we use SVM to classify cluster features directly with the 8-class data, although the classification accuracy can be increased, the improvement is little. Thus, we need to establish certain constraints to combine cluster features and the SVM classifier in order to fully use the advantages of KSFCM and SVM classifiers.
2) The experiment result indicates that our method takes effect to compensate the disadvantage of KSFCM and SVM and enhance the classification accuracy at a large extent, when few classes are available.

IV. CONCLUSION

In order to solve difficulties in gathering training samples, we proposed a hierarchical semisupervised classification method based on SVM. This method can make full use of little labeled samples and many unlabeled samples that can reflect the characteristic of the sample distribution. The semisupervised SVM framework can own both advantages of clustering and classification, which makes clustering and classification results achieved simultaneously effective and robust. In addition to that, the experiment shows that the generalization performances of the semisupervised SVM are improved with our method and avoid support vector increases linearly with the number of training samples.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe for providing the Indian Pine data sets.

REFERENCES

[1] S. Serranti, D. Cesare, F. Marini, and G. Bonifazi, “Classification of oat and groat kernels using NIR hyperspectral imaging,” *Talanta*, vol. 103, pp. 276–284, Jan. 2013.
[2] A. Villa, J. Chanussot, and J. A. Benediktsson, “Spectral unmixing for the classification of hyperspectral images at a finer spatial resolution,” *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 521–533, Jun. 2011.
[3] L. Shijin, W. Hao, and W. Dingsheng, “An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine,” *Knowl.-Based Syst*, vol. 24, no. 1, pp. 40–48, Feb. 2011.
[4] P. H. Hsu, “Feature extraction of hyperspectral images using wavelet and matching pursuit,” *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 2, pp. 78–92, Jun. 2007.
[5] W. Kim and M. Crawford, “Adaptive classification for hyperspectral image data using manifold regularization kernel machines,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4110–4121, Nov. 2010.
[6] G. F. Hughes, “On the mean accuracy of statistical pattern recognizers,” *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.
[7] M. Chi and L. Bruzzone, “Semisupervised classification of hyperspectral images by SVMs optimized in the primal,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1870–1880, Jun. 2007.
[8] L. Bruzzone and M. M. Chi, “A novel transductive SVM for the semisupervised classification of remote-sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3363–3373, Nov. 2006.
[9] V. G. Camps, T. V. Bandos, and D. Y. Zhou, “Semisupervised graph-based hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.
[10] F. Ratle, G. Camps-Valls, and J. Weston, “Semisupervised neural networks for efficient hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2271–2282, May 2010.
[11] M. Marconcini, G. Camps-Valls, and L. Bruzzone, “A composite semisupervised SVM for classification of hyperspectral images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 234–238, Apr. 2009.
[12] H. Su, H. Yang, and Q. Du, “Semisupervised band clustering for dimensionality reduction of hyperspectral imagery,” *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 1135–1139, Nov. 2011.
[13] L. Gómez-Chova, G. Camps-Valls, and J. Munoz-Mari, “Semisupervised image classification with Laplacian support vector machines,” *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 336–340, Jul. 2008.
[14] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, “Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.
[15] C. L. Stork and M. R. Keenan, “Advantages of clustering in the phase classification of hyperspectral materials images,” *Microsc. Microanal.*, vol. 16, no. 6, pp. 810–820, Dec. 2010.
[16] D. Q. Zhang and S. C. Chen, “Kernel-based fuzzy and possibilistic *c*-means clustering,” in *Proc. Int. Conf. Artif. Neural Netw.*, 2003, pp. 122–125.
[17] L. Wang, K. Inoue, and K. Urahama, “Robust kernel fuzzy clustering,” in *Proc. Fuzzy Syst. Knowl. Discov.*, 2005, pp. 454–461.
[18] S. Zhenfeng, Z. Xiran, and L. Jun, “QaR tree method for optimal image particle decomposing,” *Geomatics Inf. Sci. Wuhan Univ.*, vol. 38, no. 2, pp. 204–207, 2013.
[19] J. A. Gualtieri and S. Chettri, “Support vector machines for classification of hyperspectral data,” in *Proc. IEEE IGARSS*, 2000, vol. 2, pp. 813–815.