

## Research Article

# Active Learning Algorithms for the Classification of Hyperspectral Sea Ice Images

Yanling Han,<sup>1</sup> Jing Ren,<sup>1</sup> Zhonghua Hong,<sup>1</sup> Yun Zhang,<sup>1</sup>  
Long Zhang,<sup>2</sup> Wanting Meng,<sup>1</sup> and Qiming Gu<sup>1</sup>

<sup>1</sup>College of Information, Shanghai Ocean University, Shanghai 201306, China

<sup>2</sup>College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China

Correspondence should be addressed to Zhonghua Hong; zhong@shou.edu.cn

Received 29 August 2014; Accepted 9 December 2014

Academic Editor: L. W. Zhang

Copyright © 2015 Yanling Han 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.

Sea ice is one of the most critical marine disasters, especially in the polar and high latitude regions. Hyperspectral image is suitable for monitoring the sea ice, which contains continuous spectrum information and has better ability of target recognition. The principal bottleneck for the classification of hyperspectral image is a large number of labeled training samples required. However, the collection of labeled samples is time consuming and costly. In order to solve this problem, we apply the active learning (AL) algorithm to hyperspectral sea ice detection which can select the most informative samples. Moreover, we propose a novel investigated AL algorithm based on the evaluation of two criteria: uncertainty and diversity. The uncertainty criterion is based on the difference between the probabilities of the two classes having the highest estimated probabilities, while the diversity criterion is based on a kernel  $k$ -means clustering technology. In the experiments of Baffin Bay in northwest Greenland on April 12, 2014, our proposed AL algorithm achieves the highest classification accuracy of 89.327% compared with other AL algorithms and random sampling, while achieving the same classification accuracy, the proposed AL algorithm needs less labeling cost.

## 1. Introduction

As a member of the global marine and atmospheric system, sea ice with the high albedo has impacted the marine, power between the atmospheres, and heat and material exchange. Also sea ice plays a key role in the radiation balance, energy balance, and mass balance on the ocean surface [1]. Expect for the influence on marine hydrology, atmospheric circulation, and ecosystems, sea ice has a great threat to the shipping and the facilities of marine resource development and has become one of the most prominent oceanic disasters in the polar and high latitude regions.

For the prevention and mitigation of the ice disasters and the hazard assessment, we not only need to obtain real-time sea ice area and the outer line information, but also need more detailed data about types, thickness, and distribution of sea ice. However, for the traditional methods of sea ice detection, it is very difficult to get the continuous and large-area sea ice condition, while, remote sensing is

an effective mean that can get the large area of sea ice data rapidly. Currently, the main research areas of sea ice detection with the remote sensing are in the polar and high latitude regions. Furthermore, the countries carrying out the relevant research include America, Canada, Norway, Australia, and German. These researches mainly aimed at the moderate-resolution remote sensing detections, such as airborne remote sensing, moderate-resolution imaging spectrometer (MODIS) [2], and synthetic aperture radar (SAR). Shi et al. [3] extracted the information of sea ice by the surface temperature based on NOAA/AVHRR and got the relation between ice thickness and reflectivity by empirical formula; Meyer et al. [4] introduced an approach to map landfast ice extent based on L-band SAR data; Ozsoy-Cicek et al. [5] verified that active microwave can depart the ice edge and floating ice by field survey; Hong [6] proposed to use passive microwave for the inversion of small-scale roughness on ice surface and refractive index of sea ice. In contrast with the traditional remote sensing technology,

the hyperspectral remote images contain nearly continuous spectral information and the abundant spatial information, which has a higher capability of the target recognition and can greatly improve the accuracy of target detection. From the published literatures, the researches on sea ice detection with hyperspectral technology are rarely involved by far.

Many methods have been developed in the classification of hyperspectral images, which can be mainly concluded in two types [7, 8]: the unsupervised classification and supervised classification. Unsupervised classification does not need priori knowledge and can classify the primary hyperspectral images directly, which is simple and easy to implement, but the classification accuracy is low; the supervised classification needs some priori knowledge in advance and to get the classifier by training the labeled samples; finally we can use the trained classifier to categorize the unlabeled samples, which can get accurate classification accuracies. The classifier based on supervised classification can be obtained by the following ways: probabilistic model, empirical risk minimization (ERM) and structural risk minimization (SRM) [9, 10]. The most classic method based on probabilistic model is maximum likelihood classifier (MLC), which has higher computational complexity and requires a large number of training samples in order to get better classification results [11]; the commonly used methods based on ERM theory include the decision tree and neural network, which are easier to face the problems of “Hughes” and “overfitting” in the case of high dimension and small sample; the SRM principle considers the ERM and SRM simultaneously and increases the generalization capabilities for future samples, where the classic method based on the SRM principle is support vector machine (SVM). At present, SVM has got great progress in the way of theory research and algorithm realization and has obtained better classification results compared with traditional classifiers [12–16]. For example, Melgani and Bruzzone [17] compared the different SVM methods with the  $k$ -means and neural network based on radial basis function (RBF) in the original feature space and the feature subspace. Camps-Valls et al. [18] put forward to classify the crop by SVM and compare with other neural network methods, such as a multilayer perception neural network and RBF. Pal and Mather [19] compared the SVM with MLC and multilayer perception neural network and verified the results with Landsat-7.

The classification of hyperspectral remote sensing image often applies the supervised classification techniques, which require a large amount of labeled samples. The quality and the number of the available training samples are important for the accurate classification images. Because of the limitations of environment and conditions, the measured data used in the sea ice detection are very rare. The interpretation of hyperspectral images need to be analyzed by traditional remote sensing images with the higher spatial resolution at the same time and scene. But the available training samples are usually not enough for adequate learning of the classifier. How to mark as few samples as possible artificially and obtain better classification performance becomes the key issues of sea ice detection. In order to solve the classification problems, active learning (AL) approaches are proposed based on SVM

which have got remarkable success in the real-world learning. The SVM classifier is fit for AL because its classification rule can be characterized by a small number of support vectors that can be easy to update over the successive iterations [20]. At each iteration, the classifiers do not passively accept the training samples provided by the user but actively select the most valuable samples for the current classification model. The obtained labeled samples by the user are incorporated in the training set and the classifiers is retrained and updated. In this way, we can greatly reduce the labeling time and improve the classification accuracy. In recent years, researchers have conducted a large number of studies on the active learning and proposed many AL methods. Tong et al. [21, 22] proposed the margin sampling (MS), which selects the samples closest to the current separating hyperplane as the most uncertainty and informative samples. Another popular strategy is given by committee-based active learners. A set of unlabeled samples are trained by different classifiers. The approach selects the examples where the disagreement is maximal between the classifiers. Afterward, the method based on entropy is proposed. Examples with the highest value of entropy are selected to query the user. In [23], Joshi et al. put forward the best versus second-best (BvSB) approach, which is based on the difference between the probabilities of the two classes having the highest estimated probabilities as a measure of uncertainty. In addition, there are many other AL algorithms, such as the Fisher information matrix method. In this paper, we propose a novel investigated AL technology on the classification of hyperspectral sea ice images, which can select the most informative samples and get good classification results.

## 2. AL Algorithms

*2.1. AL Process Model.* AL was put forward by professor Angluin from Yale University in the paper of “Queries and concept learning” [24]. Currently, the AL algorithms are widely used in the text classification and image retrieval. However, AL can be applied to the classification of remote sensing images by considering the specific features of this area. In the remote sensing problems, the land-cover types of the area are selected by the three methods, such as photo interpretation, ground survey and mixed strategies. But, these strategies are implemented with high costs and much time. So we expect that the AL process can be conducted with few labeled training samples without reducing the convergence capability. The classification framework based on AL is described as shown as the Figure 1.

The active learning process is conducted according to an iterative process that can be described by the form  $(C, Q, E, T, U)$  [25], in which,  $C$  is a supervised classifier trained;  $T$  is the labeled training set;  $Q$  is the query function, in order to select the most informative unlabeled samples from the unlabeled sample pool  $U$ ;  $E$  is an human expert who can label the selected samples with the above mentioned three strategies.

The iterative process of AL can be described as follows. First, the classifier  $C$  is trained on the initial training set  $T$  which made up of few labeled samples. After

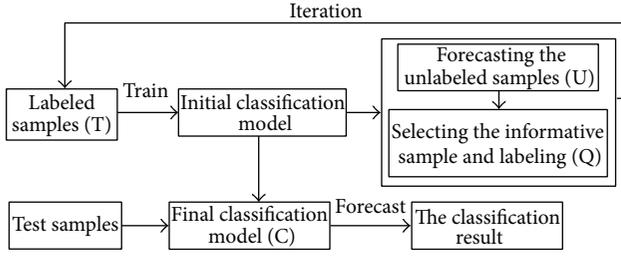


FIGURE 1: The classification framework based on AL.

the initialization, the set of samples are selected by  $Q$  query function from the pool  $U$  and query the expert  $E$ . Then, these labeled samples are included into  $T$  and the classifier  $C$  is retained by the updated training set. The retrained process continues until a stopping criterion (e.g., the labeling costs or the generalization accuracy reached some standard) is satisfied. The algorithm is described as follows.

*Algorithm 1* (active learning).

*Inputs.* The inputs are labeled sample set  $T$ , unlabeled sample set  $U$ , classifier  $C$  (Initial), and query function  $Q$ .

*Output.* the output is updated classifier  $C$ .

- (1) Train the initial classifier  $C$  on labeled sample set  $T$ .
- (2) Classify the unlabeled samples  $U$  by classifier  $C$ .
- (3) Repeat.
- (4) Select the unlabeled samples with the query function  $Q$  from the unlabeled sample pool  $U$ .
- (5) Label the unlabeled samples by human expert  $E$ .
- (6) Add the new labeled samples to the training set  $T$ .
- (7) Retrain the classifier  $C$ .
- (8) Until the stopping criteria is satisfied.

From the above mentioned descriptions, the selection of classifier and sampling strategy are two important components of AL.

**2.2. The Classification Model.** Because SVM shows the outstanding performance in solving small sample, nonlinear and high-dimensional pattern recognitions, we choose the SVM classifier in this paper. SVM is only directly applicable for two-class tasks. Aiming at solving the multiclass classification problems of hyperspectral sea ice images, the implementation of SVM is approached by multiclass strategy.

Supposing that the training sample set  $T$  is made up of  $N$  independent samples, which can be described  $(x_i, y_i)_{i=1}^N$ , where  $x_i$  denotes the training samples and  $y_i \in \{+1, -1\}$  denotes the associated labels. The basic thought of SVM is to map the data through a proper nonlinear transformation into a higher dimensional feature space, in order to find an optimal hyperplane which maximizes the margin between the two classes.

The classification problem can be transformed into a typical convex programming problem [5] on the basis of Kuhn-Tucker theorem. Accordingly, the convex programming problem can be converted into the following the dual linear programming problem by Lagrange multipliers  $\alpha_i$  associated with the original training patterns  $x_i$ :

$$\text{Max } J(a) = \text{Max} \left\{ \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right\} \quad (1)$$

$$\text{s.t. } \sum_{i=1}^N \alpha_i y_i = 0 \quad \alpha_i \geq 0; \quad i = 1, \dots, N.$$

The dual linear problem has global optimization. The  $\alpha_i$  values corresponding with the nonsupport vectors are zero, so the optimal classification decision function used for binary problem is obtained by solving above problems:

$$f(x) = \text{sgn} \left( \sum_{SV} \alpha_i y_i K(x_i \cdot x) + b \right), \quad (2)$$

where  $SV$  is the set of support vectors,  $\alpha_i$  and  $b$  are the parameters used to define the optimal hyperplane, and  $K(\cdot, \cdot)$  is the kernel function (we adopt the radial basis kernel function with better classification performance).

The multiclass classification problems depend on the binary classification, which can be transformed into multiple binary classification problems to solve. The construction of multiclass SVM classifier can be approached in two methods [26, 27]. Firstly, by constructing a series of binary classifiers, the decision is taken by combining the partial decisions of the single members of the ensemble [26, 27]. There are two common techniques, namely, the one-against-all (OAA) strategy and one-against-one (OAO) strategy. Secondly, the method is represented by SVM formulated directly as a multiclass optimization problem. The second method has poor stability and can affect the classification accuracy for the multiclass optimization. The OAO method for SVM is computationally efficient and shows good classification performance [28]. So we use OAO approach for multiclass classification in the paper. If there are  $k$  classes of data, then we need to construct  $k(k-1)/2$  binary classifiers in total. In this case, the binary classifier can separate class  $i$  from class  $j$  by means of a discriminant function  $f_{ij}(x)$ :

$$f_{ij}(x) = \text{sgn} [w^{ij} \cdot x + b^{ij}] \quad (3)$$

$$(i = 1, \dots, k; j = 1, \dots, k; i < j),$$

where  $w^{ij}$  is the normal vector of the hyperplane discriminating the class  $i$  and the class  $j$ . The final decision in the OAO strategy is taken on the basis of the "winner-takes-all" rule, which corresponds to the following maximization:

$$M = \text{argmax} \left( \sum_{j=1, j \neq i}^k \text{sgn} (f_{ij}(x)) \right). \quad (4)$$

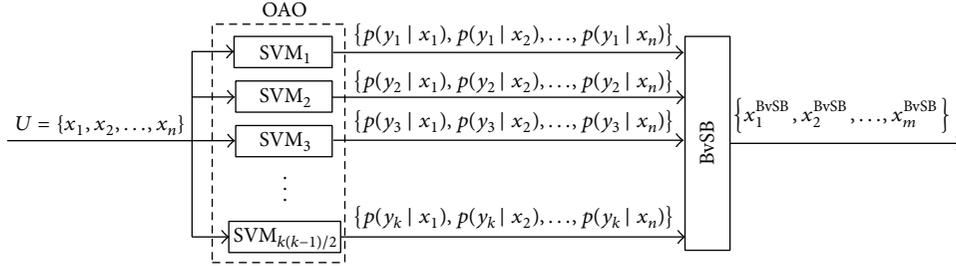


FIGURE 2: Architecture of the BvSB method.

Sometimes, conflict situations may occur between two different classes characterized by the same score. Such ambiguities can be solved by selecting the class with the smaller index value of the class.

**2.3. The Sampling Strategy.** The sampling strategy is crucial to distinguish the pros and cons of different AL algorithms. The selection of unlabeled samples with different strategies depends on the information content of the samples (i.e., the influence of the unlabeled samples on the generalization capabilities of classifier). Currently, the sampling strategy is widely used on the basis of uncertainty. In which, MS is one of the popular and effective measure for active SVM learning, but this method is only applicable to binary SVM classification problem. Although the entropy method is suitable for multiclass classification, it has a drawback. A major shortcoming of the entropy method is that its value is heavily affected by probability values of unimportant classes, which make the classifier confused [23]. In order to solve the aforementioned problems, the BvSB method was proposed which can achieve better performance in the multiclass classification.

It is very important to observe that the above mentioned strategies consider only the uncertainty. In the way, the selected samples may have the same label after querying the user. It means that there will be a lot of redundant samples selected, which do not provide additional information and are unfavorable for the algorithm convergence. In order to address the shortcomings, we adopted the enhanced clustering-based diversity criterion (ECBD) based on the diversity distribution of the samples. In the following sections, we will introduce the BvSB method and the EBCD method, respectively.

### 3. AL Algorithm Based on Uncertainty and Diversity

**3.1. The BvSB Method Based on the Uncertainty.** Let us assume that the unlabeled sample set is  $U = \{x_1, \dots, x_n\}$  and the associated labels are denoted by  $Y = \{y_1, \dots, y_k\}$ , where  $p(y_i | x_j)$  denotes the membership probability [29]. Under the BvSB criterion, we only consider the difference as a measure of uncertainty between the probabilities of the two classes having the highest estimated probability. The optimal label and the suboptimal label are denoted by  $y_{\text{Best}}$  and  $y_{\text{Second-best}}$  respectively, but the other classes are ignored.

The probabilities of the optimal label and the suboptimal label for sample  $x_i$  are represented by  $p(y_{\text{Best}} | x_i)$  and  $p(y_{\text{Second-best}} | x_i)$ ; then the criterion can be described as

$$c_{\text{uncertainty}}(x_i) = \arg \min_{x_i \in U} (p(y_{\text{Best}} | x_i) - p(y_{\text{Second-best}} | x_i)). \quad (5)$$

The BvSB method is taken as a more greedy approach as a measure of uncertainty. We use the OAO strategy for multiclass classification. We assume that  $C_{i,j}$  ( $i, j \in Y$ ) is the classifier used to discriminate the sample  $x$  between the class  $i$  and class  $j$ . If the true class label of an unlabeled sample  $x$  is  $l$ , once its label is marked and added to the training set, which will modify the boundary of the classifiers that separate class  $l$  from the other classes. We denote these classifiers by  $C_l = \{C_{l,i} (i, l \in Y, i \neq l)\}$ . Because the true label of the sample  $x$  is unknown, we use the optimal label  $y_{\text{best}}$  as the evaluation of the true label. Thus, the classification set in contention is called  $C_{y_{\text{best}}} = \{C_{y_{\text{best}}, y_i} (i, y_{\text{best}} \in Y, y_i \neq y_{\text{best}})\}$ . For the classifier set  $C_{y_{\text{best}}}$ , the uncertainty degree of sample  $x$  can be denoted by the difference in the estimated class probability value  $p_{y_{\text{best}}} - p_i$ , which can be taken as an indicator describing the information content of sample  $x$ . By minimizing the value of  $p_{y_{\text{best}}} - p_{y_i}$ , that is maximizing the classification uncertainty, the BvSB criterion is obtained:

$$\begin{aligned} c_{\text{uncertainty}}(x) &= \arg \min_{x_i \in U} \left( \min_{y \in Y, y_i \neq y_{\text{best}}} (p(y_{\text{best}} | x) - p(y_i | x)) \right) \quad (6) \\ &= \arg \min_{x_i \in U} (p(y_{\text{best}} | x) - p(y_{\text{second-best}} | x)). \end{aligned}$$

According to formula (6),  $m$  samples with lower  $c_{\text{uncertainty}}$  are selected as uncertainty samples. From the view of changing the classification boundaries, the BvSB criterion can be considered an efficient approximation for selecting the informative samples. Figure 2 shows the architecture based on the BvSB method.

**3.2. The ECBD Method Based on the Diversity.** Considering the distribution of uncertain samples at the diversity step, clustering is an effective solution to select the most diverse samples. In the previous section, the similar samples may be selected as the informative samples by the BvSB method. So we consider combining the sampling strategy with unsupervised clustering. In this case, the representative samples

are selected to label from different clusters. That is to say, the  $h < m$  samples are selected by clustering, where  $m$  samples are obtained in the uncertainty step. The standard  $k$ -means clustering algorithm is applicable to original feature space, while the SVM classification hyperplane works in the kernel space. Therefore, the selected samples in the original space may not be fit for the kernel space. To overcome this shortcoming, we adopt the enhanced clustering-based diversity method (ECBD) by clustering in kernel space, which is improved based on the standard  $k$ -means clustering. The ECBD is described as follows [20].

Assuming that  $m$  samples  $(x_1, x_2, \dots, x_m)$  are selected at the uncertainty step, the idea of kernel  $k$ -means is to divide  $m$  samples into  $h$  clusters  $(C_1, C_2, \dots, C_h)$  in the kernel space, then the most uncertainty samples of each cluster is taken as the representative sample. The center of each cluster is denoted by  $(\rho_1, \rho_2, \dots, \rho_h)$ . We suppose that sample  $x_i$  mapped into the kernel space is indicated by  $\varphi_i = \theta(x_i)$ . The Euclidean distance between the sample  $\varphi_i$  and the sample  $\varphi_j$  is written as

$$\begin{aligned} D^2(\varphi_i, \varphi_j) &= \|\theta(x_i) - \theta(x_j)\|^2 \\ &= \theta^2(x_i) - 2\theta(x_i) \cdot \theta(x_j) + \theta^2(x_j) \\ &= K(x_i, x_i) - 2K(x_i, x_j) + K(x_j, x_j). \end{aligned} \quad (7)$$

Let  $\nabla_k$  be the cluster center in the kernel space that

$$\nabla_k = \frac{1}{|C_k|} \sum_{i=1}^m \delta(\varphi_i, C_k) \varphi_i, \quad (8)$$

where  $|C_k|$  denotes the total number of samples in the cluster  $C_k$  and is computed as  $|C_k| = \sum_{i=1}^m \delta(x_i, C_k)$ .  $\delta(x_i, C_k)$  ( $1 \leq k \leq h$ ) shows the indicator function:

$$\delta(x_i, C_k) = \begin{cases} 1 & D(x_i, m_k) < D(x_i, m_j) \quad \forall j \neq k \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

The distance between  $\varphi_i$  and  $\nabla_k$  can be expressed as

$$\begin{aligned} D^2(\varphi_i, \nabla_k) &= \left\| \varphi_i - \frac{1}{|C_k|} \sum_{j=1}^m \delta(\varphi_j, C_k) \varphi_j \right\|^2 \\ &= K(x_i, x_i) + f(x_i, C_k) + g(C_k), \end{aligned} \quad (10)$$

where

$$f(x_i, C_k) = -\frac{1}{|C_k|} \sum_{j=1}^m \delta(\varphi_j, C_k) K(x_i, x_j), \quad (11)$$

$$g(C_k) = \frac{1}{|C_k|^2} \sum_{j=1}^m \sum_{s=1}^m \delta(\varphi_j, C_k) \delta(\varphi_s, C_k) K(x_j, x_s).$$

By applying (10) to the standard  $k$ -means clustering, we obtain the kernel-based  $k$ -means algorithm described as follows:

- (1) Assign the initial value of  $\delta(x_i, C_k)$  ( $i = 1, 2, \dots, m$ ,  $k = 1, 2, \dots, h$ ), and  $h$  initial clusters  $C_1, C_2, \dots, C_h$  are obtained.
- (2) For each cluster  $C_k$ , compute  $|C_k|$ ,  $f(x_i, C_k)$  and  $g(C_k)$ .
- (3) For each training sample  $x_i$  and cluster  $C_k$ , assign  $x_i$  to the closest cluster:

$$\begin{aligned} \delta(x_i, C_k) &= \begin{cases} 1 & f(x_i, C_k) + g(C_k) < f(x_i, C_j) + g(C_j) \quad \forall j \neq k \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (12)$$

- (4) Repeat step (2) and step (3) until converge.
- (5) For each cluster  $C_k$ , select the sample that is closest to the center in the kernel space as the pseudocenter of  $C_k$ :

$$m_k = \arg \min_{x_i \text{ that } \delta(x_i, C_k)=1} (D^2(\theta(x_i), \nabla_k)). \quad (13)$$

After  $C_1, C_2, \dots, C_h$  are obtained, the most informative sample is selected as the representative sample of each cluster. This sample is defined as follows:

$$\begin{aligned} x_k^{\text{BvSB+ECBD}} &= \arg \min_{x_i \text{ that } \delta(x_i, C_k)=1} \{c_{\text{uncertainty}}(x_i^{\text{BvSB}})\} \\ & \quad k = 1, 2, \dots, h, \end{aligned} \quad (14)$$

where  $x_k^{\text{BvSB+ECBD}}$  represents the  $k$ th sample chosen by the sampling strategy (i.e., BvSB + ECBD), and it is the most uncertain sample of the  $k$ th cluster (i.e., the sample that has minimum  $c_{\text{uncertainty}}(x)$  in the  $k$ th cluster). Totally,  $h$  samples are selected using (14), one for each cluster.

**3.3. AL Algorithm Based on BvSB + ECBD.** Based on the considerations of the uncertainty of the current classifier and the diversity of the sample distribution, we design the multiclass classification algorithm based on the BvSB + ECBD method. The BvSB criterion aims at selecting the most informative samples; ECBD criterion is used to select the diversity samples by clustering in the kernel space. In this case, the most representative samples are selected to query the user for sample labels. Then the obtained samples and the corresponding labels are together incorporated into the training samples and the classifier is retrained. The algorithm can be summarized as follows.

*Algorithm 2* (Proposed BvSB + ECBD).

*Inputs*

$m$  is the number of samples selected based on the BvSB method

$h$  is the number of samples selected to add to training set at each iteration

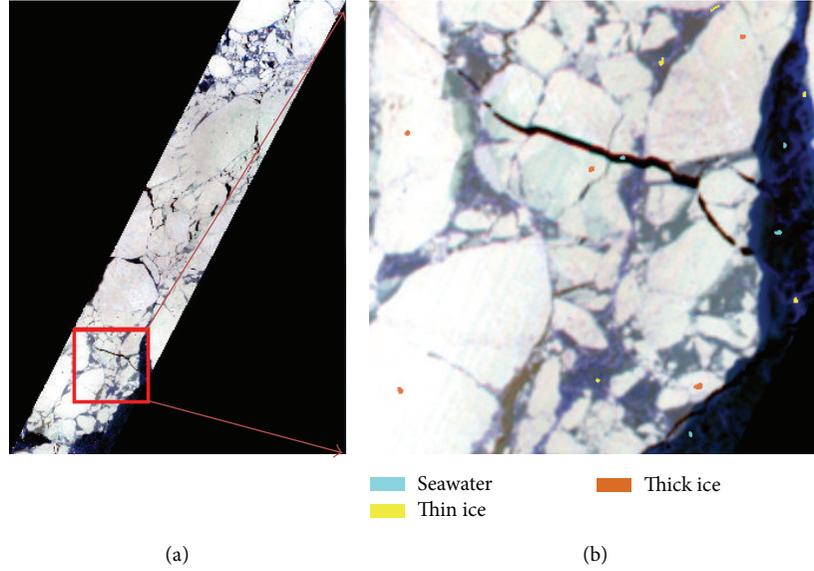


FIGURE 3: (a) Hyperspectral image marked with labeled samples (a false color image composed of R: 84, G: 49, and B: 25). (b) Partial hyperspectral image taken from (a).

$K_{\text{ini}}$  is the number of samples selected as initial training samples

#### Output

$X$  is set of unlabeled samples added in the training set

(1) *The Selection of Training Samples and the Classifier.* The training set and the unlabeled pool are marked as  $T$  and  $U$ , respectively. In the initial classification,  $K_{\text{ini}}$  samples from the pool  $U$  are selected and labeled by the user, and the sample set is denoted by  $S_{\text{ini}}$ . The training sample set  $T$  and the unlabeled pool  $U$  are updated:  $T = S_{\text{ini}}$ ,  $U \rightarrow U \setminus S_{\text{ini}}$ . The SVM classifier is trained by the training set  $T$ , then we compute the probabilities of all unlabeled samples in the pool  $U$  samples belonging to each class, which is denoted by  $p(y_i | x)$ ,  $y_i \in Y$ , and  $x \in U$ .

#### Repeat

(2) *AL Based on the BvSB Method.* According to aforeobtained probabilities  $p(y_i | x)$  of each unlabeled samples,  $m$  samples are chosen from  $U$  by formula (5), which is marked as  $X_{\text{BvSB}} = \{x_1^{\text{BvSB}}, x_2^{\text{BvSB}}, \dots, x_m^{\text{BvSB}}\}$ .

(3) *The Clustering Based on ECBD in the Kernel Space.* The  $m$  samples are clustered in the kernel space, and detailed description is shown in Section 3.2. The  $h$  samples are selected according to formula (14), which is marked as  $X_{\text{BvSB} + \text{ECBD}} = \{x_1^{\text{BvSB} + \text{ECBD}}, x_2^{\text{BvSB} + \text{ECBD}}, \dots, x_h^{\text{BvSB} + \text{ECBD}}\}$ . These selected samples are marked as the representative samples by the user and denote them by  $X = X_{\text{BvSB} + \text{ECBD}}$ .

(4) *Updating the Training Sample Set and Retraining the Classifier.* Renew the training sample set and unlabeled

sample set with new selected sample  $T = T \cup X$ ,  $U \rightarrow U \setminus X$ , then the SVM classifier are retrained with the novel training sample set.

Until algorithm converges or satisfies the number of iterations.

## 4. Experiment Analysis

4.1. *Data set Description.* Hyperion sensor is mounted on the Earth observation satellite which was launched by NASA in November 2000. Hyperspectral image has a total of 242 bands and spatial resolution of 30 m. There are 220 unique spectral channels collected with a complete spectrum covering from 357–2576 nm. Because it is being in the experimental stage, the coverage of hyperspectral image is small, only 7.7 km  $\times$  44 km [30]. Because hyperspectral images have the high resolution and continuous spectrums, it has been widely used in vegetation studies, geological surveys, fine agriculture, marine remote sensing and so on.

The data set is a hyperspectral image acquired on a marine area of Baffin Bay in northwest Greenland on April 12, 2014. The data are L1Gst level through geometric correction, projection registration and topographic correction. This image consists of 2395  $\times$  1769 pixels (which include background pixels). The number of bands is initially reduced to 176 by removing the bands with low signal-to-noise and water absorption. The available labeled samples (1678 samples) are collected by Landsat-8 image interpretation, which are illustrated in Figure 3(a). Figure 3(b) is a subset of the entire image in Figure 3(a). As can be seen from the image, there are three different classes available, namely seawater, thin ice and thick ice. All labeled samples are randomly divided to derive a pool  $U$  and the validation set  $V$ . Here we use Landsat-8 data with a spatial resolution of 15 m as the test set at the same time and same scene. The final classification performance is

TABLE 1: Number of samples of each class in  $U, V$  for the data set.

Class	$U$	$V$
Seawater	343	200
Thin ice	339	200
Thick ice	396	200
Total	<b>1078</b>	<b>600</b>

evaluated by calculating the Kappa coefficient and the overall classification accuracy. The classes and the related number of samples used in the experiments are shown in Table 1.

**4.2. Design of Experiments.** In our experiments, without losing generality, we adopt an SVM classifier with radial basis function (RBF) kernel. The values for the regularization parameter  $C$  and the spread  $\gamma$  of the RBF kernel parameters are acquired by the cross validation grid search method on the basis of the validation set. Finally, the best value of the parameter  $C$  is 32; the optimal kernel width parameter  $\gamma$  is found equal to 16.

The experiments are designed in order to compare the classification accuracy with different AL algorithms that are, respectively, BvSB and the investigated and proposed techniques (BvSB +  $k$ -means and BvSB + ECBD). In the experiments with AL algorithms, three samples of each class are randomly chosen from the pool  $U$  as initial training samples, and the rest are considered as unlabeled samples. All experimental results are referred to the average classification accuracies obtained in ten trials because of ten initial randomly selected training samples. At each round of active learning in the following, firstly, in the uncertainty step, we select  $m$  samples on the basis of uncertainty (the difference between the highest estimated probability values of the two classes, i.e., BvSB) to query the user for labels. In the diversity step, the most diverse  $h < m$  samples are chosen based on either standard  $k$ -means or ECBD to query the user for labels. Then the selected samples and the corresponding labels are together incorporated in the training set. Finally, the classifier is retrained. The related number of samples chosen by the different methods is shown in Table 2.

**4.3. Experimental Results.** This section reports experimental results with the random sampling and AL algorithms, that are respectively BvSB, BvSB +  $k$ -means, and BvSB + ECBD. Results are presented as learning rate curves, which show the relation between the average overall classification accuracies and the active learning rounds used to train the SVM classifier. By analyzing Figure 4, we can observe that three AL algorithms are generally better than random sampling. The results show that our proposed BvSB + ECBD technique shows the highest accuracies in most of the iterations. Furthermore, given the same size of training samples, as indicated by the same point on the  $x$ -axis, BvSB + ECBD shows significantly improved classification accuracy. From another perspective, in order to achieve the same value of classification accuracy (same point on the  $y$ -axis), our proposed active learning algorithm needs far fewer training

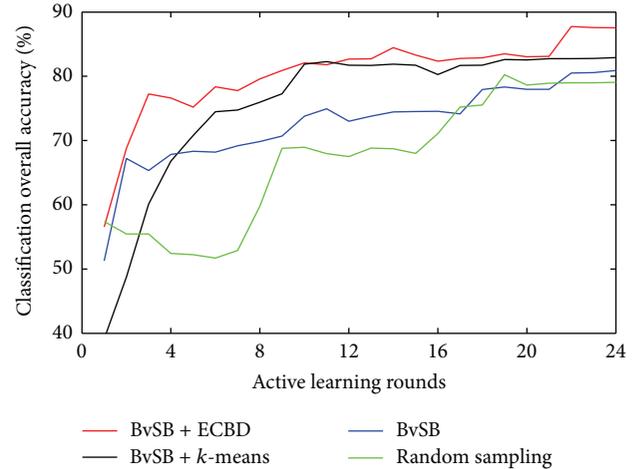


FIGURE 4: Classification of overall accuracy obtained with different AL algorithms and sampling.

TABLE 2: Number of samples chosen by the different methods.

Method	$m$	$h$
BvSB	3	—
BvSB + $K$ -means	12	3
BvSB + ECBD	12	3

TABLE 3: Percentage reduction in the number of training samples.

BvSB + ECBD selection rounds	Random selection rounds	%Reduction in training samples
1	8	63.64
2	9	58.33
5	17	60
8	19	50

samples than random selection from Table 3. The result indicates that the proposed method selects the most useful samples at each iteration, so that user input can be effectively utilized on the most relevant samples [23].

From Figure 4, one can know that the BvSB + ECBD method provides more informative samples compared to BvSB and BvSB +  $k$ -means methods and achieves higher accuracies with the same active learning rounds. Figures 5 and 6 show the distribution of the chosen training samples and the pool (considering bands 49 and 84 of the hyper-spectral image) after six iterations of the AL process with the BvSB method and BvSB + ECBD method, respectively. Note that, since the BvSB method considers only the uncertainty of samples, it may result in the selection of similar samples which can only provide redundant information. We can also find that performing the clustering in the kernel space can improve the classification accuracy compared with the standard  $k$ -means clustering. Indeed, because of the kernel mapping, the set of most diverse sample in the original space may not be the most diverse in the kernel space [20].

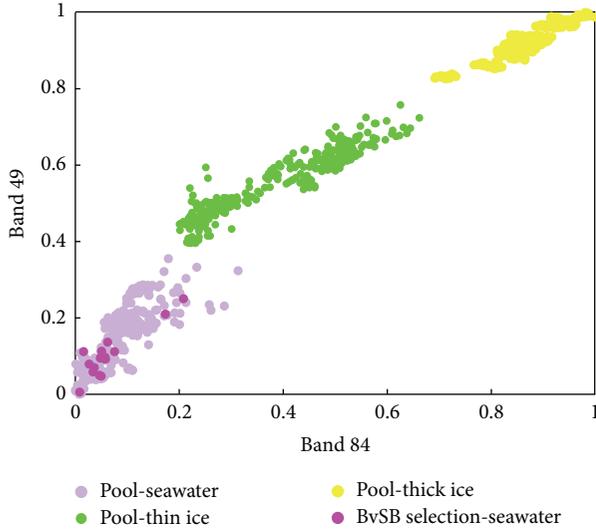


FIGURE 5: Distribution of the chosen training samples by the BvSB method and pool samples considering bands 49 and 84 of the hyperspectral image.

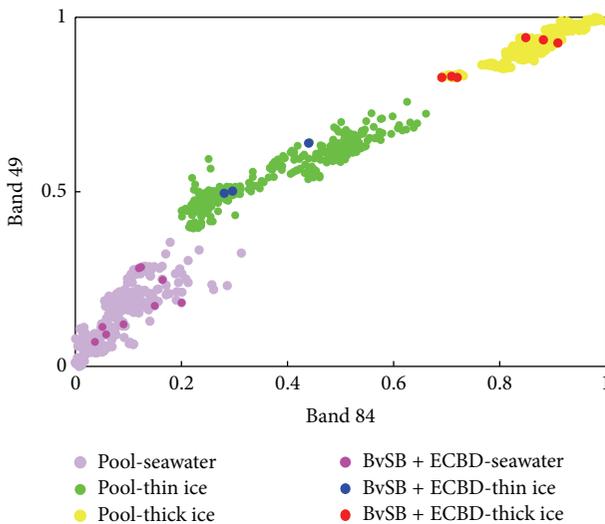


FIGURE 6: Distribution of the chosen training samples by the BvSB + ECBD method and pool samples considering bands 49 and 84 of the hyperspectral image.

In Figure 7, we demonstrate the original hyperspectral image and the classification results, in which, Figure 7(a) is an original hyperspectral image composed of band 25, 49 and 84. Figure 7(b) is the classification image of the Landsat-8 data obtained by the standard SVM classifier, which used as a reference image for evaluating the classification performance. Further, Figure 7(c) is the classification image of the hyperspectral image with the BvSB + ECBD method.

In order to assess the classification performance of our proposed method, in Table 4, we also report the confusion matrix, the accuracy per class and Kappa coefficient at the last iteration of the BvSB + ECBD method. It is important to observe that the accuracies of seawater and thin ice are low.

TABLE 4: Confusion matrix for the classification with the proposed BvSB + ECBD method.

Class	Seawater	Thin ice	Thick ice	Total
Seawater	45476	9967	2450	57893
Thin ice	21741	139924	39873	201538
Thick ice	3729	59508	963416	1026653
Total	<b>70946</b>	<b>209399</b>	<b>1005739</b>	<b>1286084</b>
%Accuracy per class	64.099	66.822	95.792	—
Kappa = 0.693			Overall accuracy = 89.327%	

Since hyperspectral image was acquired in April, 2014, thin ice began to melt with the increasing temperature. So there are lots of pixels of seawater and thin ice that were confused and wrongly classified.

Finally, we carry out an analysis of the sensibility of our proposed BvSB + ECBD method with different number of initial training samples. In Figure 8,  $n$  denotes the total number of initial training samples of three classes, where, the selected number of samples of each class is the same. One can see that, selecting different  $n$  values results in similar classification accuracies. The experiment result indicates that the different initial training samples do not provide a large benefit in the BvSB + ECBD method, that is to say, the classification accuracy is not sensitive to the selection of initial training samples [23]. Furthermore, we can also observe that, when using high  $n$  values, convergence is easily achieved than when using small  $n$  values. That is because the greater  $n$  values, the more the number of training samples, when given the same number of rounds.

## 5. Conclusions

In this paper, AL algorithms in the classification of hyperspectral image have been addressed, which can reduce the number of labeled samples added to training set and improve the classification accuracy with respect to traditional passive techniques. Query function based on BvSB in the uncertainty step, and standard  $k$ -means clustering and ECBD in the diversity step have been generalized to multiclass problems. Moreover, our proposed novel BvSB + ECBD method is compared with BvSB, BvSB +  $k$ -means and random sampling in the classification accuracy. By analyzing the experiment results, we can summarize as follows: (1) The proposed BvSB + ECBD method gets the best performance in terms of classification accuracy and can reduce a large amount of labeled samples compared with random sampling; (2) the BvSB +  $k$ -means method provides slightly lower classification accuracies than the BvSB + ECBD technique. At the diversity step, because of kernel mapping, the most diverse samples in the original space by the standard  $k$ -means clustering may not be the most diverse in the kernel space, which means that the most informative samples cannot be selected for the current classifier; (3) the BvSB method leads to poorer classification accuracies with respect to other AL algorithms. Therefore, we can conclude that obtained uncertainty samples based on

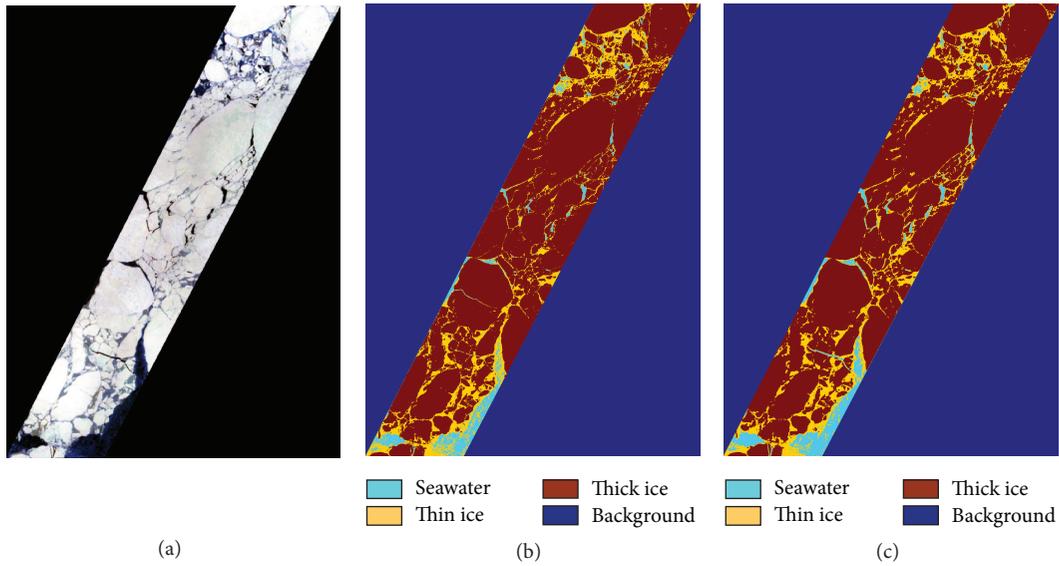


FIGURE 7: (a) Hyperspectral image (a false color image composed of R: 84, G: 49, and B: 25). (b) Result of the classification of the Landsat-8 data. (c) Result of the classification of the hyperspectral image.

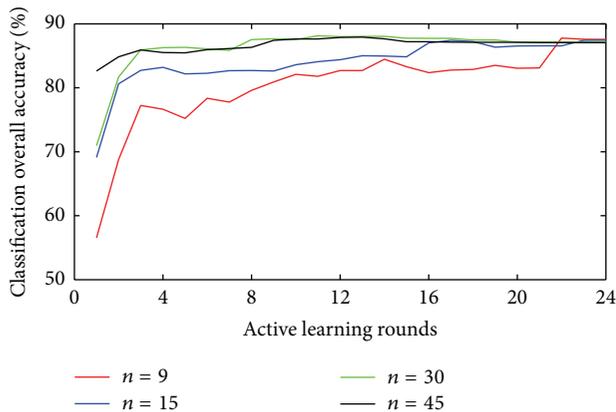


FIGURE 8: Classification overall accuracy obtained by the BvSB + ECBD method with different  $n$  values.

the BvSB technique may be the similar samples and cannot provide available information; (4) our proposed BvSB + ECBD method is not sensitive to the selection of initial training samples.

As a future development, we consider to extend our proposed AL technique by integrating the semisupervised method in the classification of hyperspectral image. During the iterative process, we can also make full use of abundant spectral information to select the more respective samples and more accurately identify the types of ice.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

This research is supported by National Natural Science Foundation of China (Grant nos. 41376178 and 41401489), the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, Shanghai Science and Technology Committee (Project 11510501300), and Shanghai Foundation for University Youth Scholars (ZZHY13033), and special funding for the development of science and technology of Shanghai Ocean University (A2-0209-14-200070).

### References

- [1] C. Deser, J. E. Walsh, and M. S. Timlin, “Arctic sea ice variability in the context of recent atmospheric circulation trends,” *Journal of Climate*, vol. 13, no. 3, pp. 617–633, 2000.
- [2] C. Drüe and G. Heinemann, “Accuracy assessment of sea-ice concentrations from MODIS using in-situ measurements,” *Remote Sensing of Environment*, vol. 95, no. 2, pp. 139–149, 2005.
- [3] J. Shi, J. Guo, T. Zheng, Q. Wang, and J. Zhu, “Monitoring sea ice based on NOAA/AVHRR data,” in *Earth Observing Systems XVII*, vol. 8510 of *Proceedings of SPIE*, August 2012.
- [4] F. J. Meyer, A. R. Mahoney, H. Eicken, C. L. Denny, H. C. Druckenmiller, and S. Hendricks, “Mapping arctic landfast ice extent using L-band synthetic aperture radar interferometry,” *Remote Sensing of Environment*, vol. 115, no. 12, pp. 3029–3043, 2011.
- [5] B. Ozsoy-Cicek, S. F. Ackley, A. Worby, H. Xie, and J. Lieser, “Antarctic sea-ice extents and concentrations: comparison of satellite and ship measurements from International Polar Year cruises,” *Annals of Glaciology*, vol. 52, no. 57, pp. 318–326, 2011.
- [6] S. W. Hong, “Detection of small-scale roughness and refractive index of sea ice in passive satellite microwave remote sensing,” *Remote Sensing of Environment*, vol. 114, no. 5, pp. 1136–1140, 2010.

- [7] F. A. Mianji and Y. Zhang, "Robust hyperspectral classification using relevance vector machine," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2100–2112, 2011.
- [8] B. C. Kuo and D. A. Landgrebe, "A robust classification procedure based on mixture classifiers and nonparametric weighted feature extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 11, pp. 2486–2494, 2002.
- [9] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, Cambridge, UK, 2000.
- [10] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, NY, USA, 1998.
- [11] X. Jia and J. A. Richards, "Efficient maximum likelihood classification for imaging spectrometer data sets," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 2, pp. 274–281, 1994.
- [12] J. Chen, C. Wang, and R. S. Wang, "Using stacked generalization to combine SVMs in magnitude and shape feature spaces for classification of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2193–2205, 2009.
- [13] M. Marconcini, G. Camps-Valls, and L. Bruzzone, "A composite semisupervised SVM for classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 2, pp. 234–238, 2009.
- [14] J. Chen, C. Wang, and R. Wang, "Fusion of SVMs in wavelet domain for hyperspectral data classification," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO '09)*, pp. 1372–1375, December 2009.
- [15] R. S. Hosseini and S. Homayouni, "A SVMs-based hyperspectral data classification algorithm in a similarity space," in *Proceedings of the 1st Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS '09)*, pp. 1–4, Grenoble, France, August 2009.
- [16] B. Demir and S. Erturk, "Accurate SVM classification using border training patterns," in *Proceedings of the 1st Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS '09)*, pp. 1–4, Grenoble, France, August 2009.
- [17] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.
- [18] G. Camps-Valls, L. Gómez-Chova, J. Calpe-Maravilla et al., "Robust support vector method for hyperspectral data classification and knowledge discovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 7, pp. 1530–1542, 2004.
- [19] M. Pal and P. M. Mather, "Support vector machines for classification in remote sensing," *International Journal of Remote Sensing*, vol. 26, no. 5, pp. 1007–1011, 2005.
- [20] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 3, pp. 1014–1031, 2011.
- [21] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *Proceedings of the 9th ACM International Conference on Multimedia*, pp. 107–118, October 2001.
- [22] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *The Journal of Machine Learning Research*, vol. 2, pp. 45–66, 2002.
- [23] A. J. Joshi, F. Porikli, and N. Papanikolopoulos, "Multi-class active learning for image classification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 2372–2379, June 2009.
- [24] D. Angluin, "Queries and concept learning," *Machine Learning*, vol. 2, no. 4, pp. 319–342, 1988.
- [25] M. Li and I. K. Sethi, "Confidence-based active learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1251–1261, 2006.
- [26] J. A. Gualtieri and R. F. Crompt, "Support vector machines for hyperspectral remote sensing classification," in *The 27th AIPR Workshop Advances in Computer-Assisted Recognition*, vol. 3584 of *Proceedings of SPIE*, pp. 221–232, Washington, DC, USA, January 1999.
- [27] D. J. Sebald and J. A. Bucklew, "Support vector machines and the multiple hypothesis test problem," *IEEE Transactions on Signal Processing*, vol. 49, no. 11, pp. 2865–2872, 2001.
- [28] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.
- [29] R. Zhang and A. I. Rudnicky, "A large scale clustering scheme for kernel k-means," in *Proceedings of the 16th International Conference on IEEE Pattern Recognition*, vol. 4, pp. 289–292, IEEE, 2002.
- [30] P. Barry, C. Segal, and S. Carman, *EO-1/Hyperspectral Science Data User's Guide*, 2001.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

