

A Spatial–Contextual Support Vector Machine for Remotely Sensed Image Classification

Cheng-Hsuan Li, Bor-Chen Kuo, *Member, IEEE*, Chin-Teng Lin, *Fellow, IEEE*, and Chih-Sheng Huang

Abstract—Recent studies show that hyperspectral image classification techniques that use both spectral and spatial information are more suitable, effective, and robust than those that use only spectral information. Using a spatial–contextual term, this study modifies the decision function and constraints of a support vector machine (SVM) and proposes two kinds of spatial–contextual SVMs for hyperspectral image classification. One machine, which is based on the concept of Markov random fields (MRFs), uses the spatial information in the original space (SCSVM). The other machine uses the spatial information in the feature space (SCSVMF), i.e., the nearest neighbors in the feature space. The SCSVM is better able to classify pixels of different class labels with similar spectral values and deal with data that have no clear numerical interpretation. To evaluate the effectiveness of SCSVM, the experiments in this study compare the performances of other classifiers: an SVM, a context-sensitive semisupervised SVM, a maximum likelihood (ML) classifier, a Bayesian contextual classifier based on MRFs (ML_MRF), and k nearest neighbor classifier. Experimental results show that the proposed method achieves good classification performance on famous hyperspectral images (the Indian Pine site (IPS) and the Washington, DC mall data sets). The overall classification accuracy of the hyperspectral image of the IPS data set with 16 classes is 95.5%. The kappa accuracy is up to 94.9%, and the average accuracy of each class is up to 94.2%.

Index Terms—Classification, Markov random fields (MRFs), spatial–contextual information, support vector machines (SVMs).

I. INTRODUCTION

MOST classification studies in the remote sensing literature encounter great difficulties in two classification problems. One is the Hughes phenomenon [1]–[4], a kind of problem of the curse of dimensionality. When the dimension of the space grows and the size of the training set is fixed, the classification accuracy reaches a maximum for a given size and then decreases. The other problem is that spectral-domain-based classifiers often lead to imprecise estimation

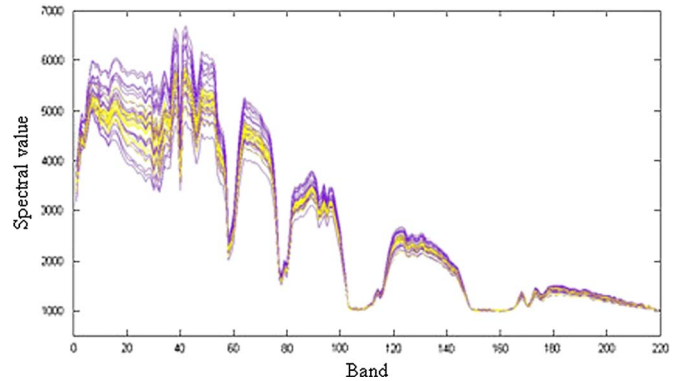


Fig. 1. These spectral values were obtained from the IPS data set. The purple represents the Soybeans-min till patterns, and the yellow represents the Corn-no till patterns. These two classes have very similar spectral properties.



Fig. 2. SVM classification result of the IPS image, containing specklelike errors.

of samples which come from different land-cover classes but have very similar spectral properties, which makes it difficult to distinguish unlabeled patterns [8], [9]. Fig. 1 shows the spectral values obtained from patterns of two categories in the Indian Pine site (IPS) data set: Soybeans-min till (purple color) and Corn-no till (yellow color) [23]. These two different classes have very similar spectral properties. Hence, employing these classes to train conventional classifiers (e.g., maximum likelihood (ML) classifier [5], k nearest neighbor (k -NN) classifier [5], and support vector machine (SVM) [6], [7]) would cause poor classification performance, producing a specklelike classification map [8], [9], [41]. Fig. 2 shows that the SVM classification map of the IPS includes a number of specklelike errors.

Considering both spectral and spatial–contextual information in the interpretation of a hyperspectral image is an effective

Manuscript received September 16, 2010; revised January 18, 2011 and May 11, 2011; accepted June 11, 2011. Date of publication August 30, 2011; date of current version February 24, 2012. This work was supported in part by the National Science Council under Grant NSC 100-2628-E-142-001-MY3 and 99-2221-E-142-002-.

C.-H. Li is with the Institute of Electrical Control Engineering, National Chiao Tung University, Hsinchu 30010, Taiwan, and also with the Graduate Institute of Educational Measurement and Statistics, National Taichung University of Education, Taichung 40306, Taiwan (e-mail: ChengHsuanLi@gmail.com).

B.-C. Kuo is with the Graduate Institute of Educational Measurement and Statistics, National Taichung University of Education, Taichung 40306, Taiwan (e-mail: kbc@mail.ntcu.edu.tw).

C.-T. Lin and C.-S. Huang are with the Brain Research Center and the Institute of Electrical Control Engineering, National Chiao Tung University, Hsinchu 30010, Taiwan (e-mail: ctlin@mail.ntcu.edu.tw; chih.sheng.huang821@gmail.com).

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/TGRS.2011.2162246

way to decrease specklelike errors. There are two main methods for combining spectral and spatial-contextual information. The graph-based technique [32]–[35] uses the typical method of building a regularization in which “similar” features belong to the same class. This method associates the vertices of a graph with the complete set of samples and then builds the regularization depending on the variables defined on the vertices [32]. The other approach is to use fixed-window-based methods, such as Markov random fields (MRFs) [8], [9], morphological filtering [36], or morphological leveling [37], [38]. This approach improves the classification performance compared to pixelwise methods when applied to hyperspectral images [12].

Jackson and Landgrebe [8] applied Gaussian function into Bayesian decision rule with MRFs, Bayesian contextual classifier based on MRF (ML_MRF), to mitigate the specklelike errors. Their method achieves improved performance in classification maps. Other research suggests applying similar concepts to develop an MRF-based k nearest neighbor classifier and Parzen classifier [9]. However, the MRF-based classifiers are still constrained by the statistical estimation (e.g., covariance matrix of ML base on Gaussian distribution) or the amount of learning data.

The SVM [6] is a pattern classification technique proposed by Boser *et al.*. Unlike traditional methods, which minimize empirical training errors, SVM attempts to minimize the upper bound of the generalization error by maximizing the margin between the separating hyperplane and the training data. Hence, SVM is a distribution-free algorithm that can overcome the problem of poor statistical estimation. SVM also achieves greater empirical accuracy and better generalization capabilities than other standard supervised classifiers [15], [16]. In particular, SVM has shown a good performance for high-dimensional data classification with a small size of training samples [17], [18] and is robust to the Hughes phenomenon [1], [2], [15], [17], [18].

Recent studies [1], [2], [8]–[12] show that SVMs with both spectral and spatial information achieve effective and stable hyperspectral image classification. A context-sensitive semisupervised SVM (CS⁴VM) [1] uses the context of neighborhood patterns as semipatterns to solve the problem of noisy training patterns. In this case, noisy training patterns are mislabeled patterns that bring distorted information to a classifier. CS⁴VM is a semilearning approach, in which the computational cost increases as the number of semisamples increases.

Tarabalka *et al.* [12] presented a spectral-spatial classification scheme based on partitional clustering techniques (SVM+EM). This approach segments an image into more homogeneous regions and combines the results of these regions using pixelwise SVM classification. A spatial postregularization (PR) of the classification map was performed to reduce the noise. This approach is particularly suitable for classifying images with large spatial structures, when spectral responses of different classes are dissimilar, and the classes contain a comparable number of pixels. If the spectral responses are not significantly different, this approach may result in misclassification [12].

This study proposes two kinds of spatial-contextual SVM classification algorithms. The proposed method modifies the

decision function and the constraints of SVM based on the spatial-contextual term of each training pattern. The corresponding separating hyperplane in the feature space is adjusted by the neighborhood system in the original space (SCSVM) or feature space (SCSVMF), making it more suitable for classification. This paper is organized as follows. Section II introduces the proposed algorithm and the corresponding classification system. Section III describes the experimental designs of the different spatial-contextual information-based approaches, along with the two hyperspectral images: the IPS and the Washington, DC mall. Section IV presents the experimental results, including classification accuracies and classification maps. Finally, Section V draws conclusions for the present study.

II. SPATIAL-CONTEXTUAL SVM

Let X be a hyperspectral d -dimensional image of size $I \times J$ pixels. Assume that a set of training data set $D = \{x_i | x_i \in X \subset R^d, i = 1, 2, \dots, n\}$ is available and $\{y_i \in \{+1, -1\}\}_{i=1}^n$ is the corresponding label set. The learning process of the proposed SCSVM (SCSVMF) classification system consists of three steps: 1) learning the standard SVM to classify the image; 2) learning SCSVM (SCSVMF) with both spectral and spatial-contextual information; and 3) repeating step 2) to update the unlabeled patterns until convergence.

A. Standard SVM

SVM tries to find a separating hyperplane in the feature space, a Hilbert space H , for a binary classification problem [6]. The soft-margin SVM algorithm is based on the following constrained minimization optimal problem:

$$\begin{aligned} \min_{w, \xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \\ \text{subject to} \quad & y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, n \end{aligned} \quad (1)$$

where w is a vector normal to the hyperplane, b is a constant such that $b/\|w\|$ represents the distance between the hyperplane and the origin, $\phi: R^d \rightarrow H$ is a nonlinear mapping function, ξ_i 's are slack variables to control the training errors, $\xi = [\xi_1, \dots, \xi_n]^T$, and $C \in R^+$ is a penalty parameter for tuning the generalization capability. Trying to solve this optimal problem directly with inequality constraints is generally difficult. However, the original optimal problem has an equivalent dual representation using the Lagrange optimization. The corresponding dual Lagrange function is defined as

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \phi(x_i)^T \phi(x_j) \\ \text{subject to} \quad & \sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, i = 1, 2, \dots, n \end{aligned} \quad (2)$$

where the artificial variables α_i 's are Lagrange multipliers and $\alpha = [\alpha_1, \dots, \alpha_n]^T$.

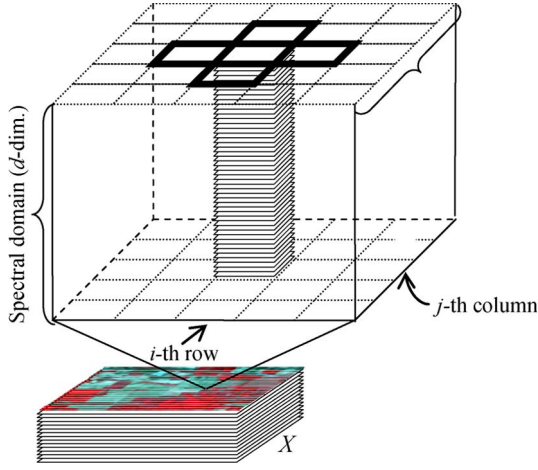


Fig. 3. Pixels enclosed by bold lines represent the first-order neighborhood system used in SCSVM.

The kernel trick uses a kernel function $\kappa : R^d \times R^d \rightarrow R$ to implicitly map the data from the original space R^d to H without knowing the feature mapping ϕ . The inner product of samples in the feature space can be computed directly from the original data items using a kernel function. This is because a kernel function κ satisfies the Mercer's theorem [16], i.e., there is a feature map ϕ into a Hilbert space H such that $\kappa(x, z) = \phi(x)^T \phi(z)$, where $x, z \in R^d$, if and only if κ is a symmetric function for which the matrices $K = [\kappa(x_i, x_j)]_{1 \leq i, j \leq n}$ formed by restriction to any finite subset $\{x_1, \dots, x_n\}$ of the space R^d are positive semidefinite. Hence, (2) can be written as in (3) by applying the kernel trick. Since, for a kernel function, the corresponding kernel matrix is positive semidefinite for all training sets, this in turn means that the optimization problem of (3) is always convex [16].

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j) \\ \text{subject to} \quad & \sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, n. \end{aligned} \quad (3)$$

Once the α_i 's are determined, the decision function for an unlabeled pattern x is defined as

$$f_{\text{SVM}}(x) = \sum_{i=1}^n y_i \alpha_i \kappa(x_i, x) + b$$

where b is chosen so that $y_j (\sum_{i=1}^n y_i \alpha_i \kappa(x_i, x_j) + b) = 1$ for any x_j with $0 < \alpha_j < C$, and a corresponding forecasting label is $\text{sgn}(f_{\text{SVM}}(x))$.

The following two sections propose two kinds of spatial-contextual SVMs with different neighborhood systems: the original space (SCSVM) and the feature space (SCSVMF).

B. Spatial-Contextual SVM in the Original Space

In SCSVM, spatial information exploits the semilabels for the pixels belonging to the neighborhood system in the original space (see Fig. 3) of each sample from the preceding discrim-

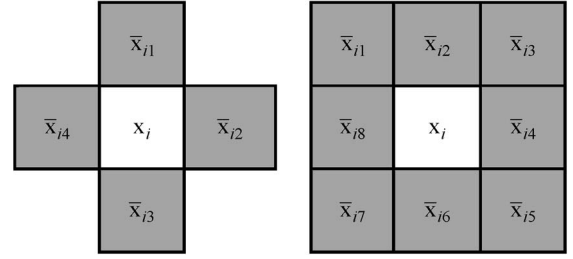


Fig. 4. Left and right images above represent the first-order and second-order neighborhood systems in the original space, respectively.

$y_{i1}^s = +1$	$y_{i2}^s = -1$	$y_{i3}^s = -1$	$m^+(x_i) = 6$ $m^-(x_i) = 2$
$y_{i8}^s = +1$	y_i	$y_{i4}^s = +1$	
$y_{i7}^s = +1$	$y_{i6}^s = +1$	$y_{i5}^s = +1$	

Fig. 5. Example of the spatial-contextual information with the second-order neighborhood system of pattern x_i in the original space.

inated process by applying standard SVM to overcome similar spectral properties. SCSVM can have good generalization, particularly for the pixels with similar spectral attributes but located in different regions. This approach decreases specklelike errors and significantly improves classification performance. Let $\partial x_i^O = \{\bar{x}_{ij} | j = 1, \dots, M\}$ represent a neighborhood system of the pixel x_i in the original space, where M is four or eight to represent that ∂x_i^O is a first-order or second-order neighborhood system, respectively (Fig. 4).

After performing the standard SVM, SCSVM can obtain the semilabels of ∂x_i , which is equal to ∂x_i^O , and denote them as $\{y_{ij}\}_{j=1}^M$, i.e., $y_{ij} = \text{sgn}(f_{\text{SVM}}(\bar{x}_{ij}))$, $i = 1, \dots, n$, $j = 1, \dots, M$. The constrained minimization problem associated to SCSVMs by taking into account the semilabels of the whole image is defined as follows:

$$\begin{aligned} \min_{w, \xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \\ \text{subject to} \quad & y_i (w^T \phi(x_i) + b + \gamma(m^+(x_i) - m^-(x_i))) \\ & \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, n \end{aligned} \quad (4)$$

where $\gamma \in [0, \infty)$ is a nonnegative parameter that controls the effect of spatial-contextual information. $m^+(x_i)$ and $m^-(x_i)$ are the numbers of pixels in the neighborhood system ∂x_i which belong to class +1 and class -1, respectively. Fig. 5 shows an example of the spatial-contextual information of the pattern x_i with the second-order neighborhood system ∂x_i^O employed in the spatial domain.

The SCSVM cost function does not need to be modified and still maintains the property of convex property. Since the objective function of the minimization problem of the SCSVM only contains training samples without samples with semilabels, the decision hyperplane is not influenced by samples with similar spectral, and the computational cost of each iteration in SCSVMs is similar to that of SVM.

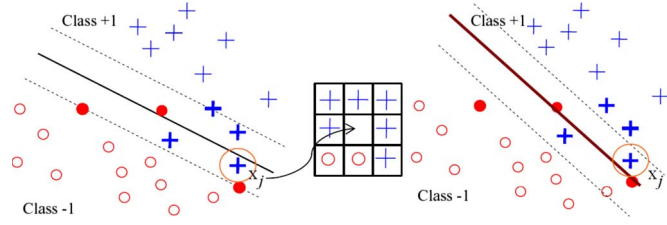


Fig. 6. Left panel shows the (solid black line) decision boundary obtained by SVM. The center panel shows the semilabels of the patterns in the second-order neighborhood system of x_j . The right panel shows the (solid red line) decision boundary obtained by applying SCSVM.

According to the Lagrange theorem, the corresponding dual problem is as follows:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n (1 - y_i \gamma (m^+(x_i) - m^-(x_i))) \alpha_i \\ & - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j) \\ \text{subject to} \quad & \sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, n. \end{aligned} \quad (5)$$

When $\alpha_i, i = 1, \dots, n$, are determined, the decision function for an unlabeled pattern x is defined as

$$f_{\text{SCSVM}}(x) = \sum_{i=1}^n y_i \alpha_i \kappa(x_i, x) + b + \gamma (m^+(x) - m^-(x)). \quad (6)$$

Moreover, any generic pattern belonging to the investigated image can be classified according

$$\text{sgn}(f_{\text{SCSVM}}(x)).$$

If some training patterns are located in the margin, these patterns may lead to similar spectral properties. Hence, these patterns may be noisy patterns in standard SVM learning. To overcome this problem, the spatial terms are added to the constraints and the decision function of SCSVM. If $m^+(x_j) - m^-(x_j) > 0$ and $m^+(x_j) - m^-(x_j) < 0$, then $f_{\text{SCSVM}}(x_j) > f_{\text{SVM}}(x_j)$ and $f_{\text{SCSVM}}(x_j) < f_{\text{SVM}}(x_j)$, respectively. This means that if the semilabels of most patterns in the neighborhood system ∂x_j are +1, then the signed distance from x_j to the decision hyperplane by applying SCSVMs will tend to be positive. If the semilabels of most patterns in the neighborhood system ∂x_j are -1, then the signed distance from x_j to the decision hyperplane by applying SCSVM will tend to be negative. The parameter γ controls the effect of the spatial-contextual information, i.e., the term $m^+(x_j) - m^-(x_j)$. If γ is set to zero, then SCSVM degenerates to the standard SVM. When γ increases, the impact of neighborhood points (spatial information) increases. If γ approaches ∞ , then the semilabel of x_j is determined by the sign of $m^+(x_j) - m^-(x_j)$, i.e., the spatial-contextual information.

Fig. 6 shows the effect of applying SCSVM. The left panel shows the decision boundary (solid black line) obtained by applying standard SVM. Note that the training sample x_j with

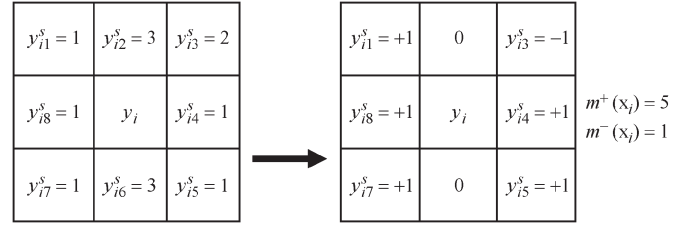


Fig. 7. Multiclass case of the spatial-contextual information defined by the OAO strategy (class 1 versus class 2) for pattern x_i in the neighborhood system ∂x_i^O . The labels of class 1 and class 2 are defined as +1 and -1, respectively.

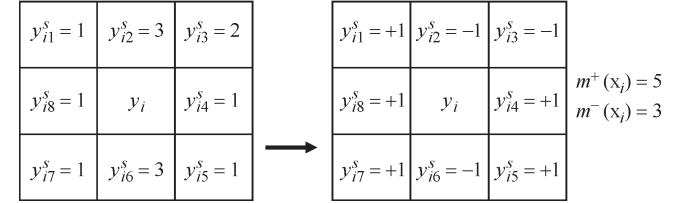


Fig. 8. Multiclass case of the spatial-contextual information defined by the OAA strategy (class 1 versus all others) for pattern x_i in the neighborhood system ∂x_i^O . The label of class 1 is defined as +1, and the labels of the remaining classes (class 2 and class 3) are defined as -1.

$y_j = +1$ is in the opposite area (class -1) but in the area between margins. After performing standard SVM, $m^+(x_j) - m^-(x_j) = 4 > 0$. The spatial-contextual information in the center panel of Fig. 6 shows that the training sample x_j should be in the area in which sample labels are one. Since $f_{\text{SCSVM}}(x_j) > f_{\text{SVM}}(x_j)$, thus x_j would be in the expected area (class +1), as shown in the right panel of Fig 6.

As mentioned earlier, SCSVM depends on the spectral information and the spatial-contextual information, which is based on the neighborhood system in the original system. Hence, the similar spectral properties can be solved by applying SCSVM. Note that SCSVM applies the spatial-contextual information to emphasize the effects of this pattern on the learning phase.

For the multiclass classification problem, the binary SCSVM to the solution of two multiclass strategies, one-against-one (OAO) strategy [20]–[22] and one-against-all (OAA) strategy [22], is described in the following. The OAO strategy separates each pair of classes. Thus, for a classification problem with L classes, $L(L-1)/2$ SCSVMs are trained to distinguish the samples of one class from the samples of another class. The classification result of an unlabeled pattern is based on the maximum voting, where each SCSVM votes for one class. When an SCSVM is trained by two classes of training data, it ignores the spatial-contextual information of other classes to avoid misjudgments in the training process.

Fig. 7 shows the OAO strategy for computing $m^+(x_i)$ and $m^-(x_i)$ in the neighborhood system ∂x_i^O of x_i . Suppose that there are three classes and SCSVM is trained by the training samples in class 1 and class 2. Thus, the semilabels equal to three will be omitted. Since $y_{i2} = 3$ and $y_{i6} = 3$, the spatial-contextual information is ignored, and hence, $m^+(x_i) = 5$ and $m^-(x_i) = 1$.

The OAA multiclass strategy trains L SCSVMs, one per class, using members of all other classes as negative examples if there are L classes. Fig. 8 shows the OAO strategy for computing $m^+(x_i)$ and $m^-(x_i)$ in the neighborhood system

of the pattern x_i . When the k th OAA SCSVM is trained, the class k is set as the positive class, and other classes are all set as negative class. Fig. 8 shows $m^+(x_i) = 5$ and $m^-(x_i) = 3$ for the example by considering the neighborhood system ∂x_i^O .

C. Spatial-Contextual SVM in the Feature Space

The SCSVMF approach uses the same concept as SCSVM except that the neighborhood system $\partial x_i = \partial x_i^F$ which contains M nearest neighbors in the feature space, i.e.,

$$\partial x_i^F = \{\bar{x}_{ij} | \|\phi(x_i) - \phi(x_{ij})\| \leq \|\phi(x_i) - \phi(z)\|, \\ \forall z \in X, z \neq x_i, z \neq x_{ij}, j = 1, \dots, M\}.$$

Similar spectral properties cannot be solved efficiently by applying SCSVMF because the nearest neighbors in the feature space are used in the neighborhood system ∂x_i^F , and hence, being neighborhoods of a given point in the feature space is caused by similar spectra. These neighborhood points may not have geographic relationship. In particular, the classification accuracy may decrease when the neighborhood points in the feature space are from different classes.

When SCSVMF is trained by the training samples in class k and class s , the semilabels of samples in ∂x_i^F , which are not equal to k and s in the OAO strategy, are ignored in the multiclass classification problem. The OAA multiclass strategy sets the semilabels of samples in ∂x_i^F , which belong to class k , as the positive class and other semilabels of samples in ∂x_i^F as negative classes when training the k th OAA SCSVM. Similarly, if γ is set to zero, then SCSVMF degenerates to the standard SVM.

D. Classification System by Applying SCSVM and SCSVMF

Based on the earlier descriptions and definitions, the following figure briefly describes the classification system by applying the proposed SCSVM and SCSVMF.

- Step 1) Obtaining the classification image with semilabels from the standard SVM.
- Step 2) Getting the spatial-contextual information for each training pattern with OAO or OAA multiclass architecture from the preceding classification result.
- Step 3) Training the proposed SCSVM (SCSVMF) with the spatial-contextual information from Step 2) and getting another classification image with the semilabels obtained from SCSVM (SCSVMF).
- Step 4) Repeating Steps 2) and 3) if an iteration is requested. The iteration may terminate when the difference of semilabels in this iteration step and the previous iteration step is smaller than a certain tolerance value.
- Step 5) Finally, the spatial PR of the classification map is performed for SCSVM [12]. The aim of the PR step is to reduce the noise in the classification map after the majority vote procedure.

Fig. 9 shows the framework of the SCSVM (SCSVMF) algorithm.

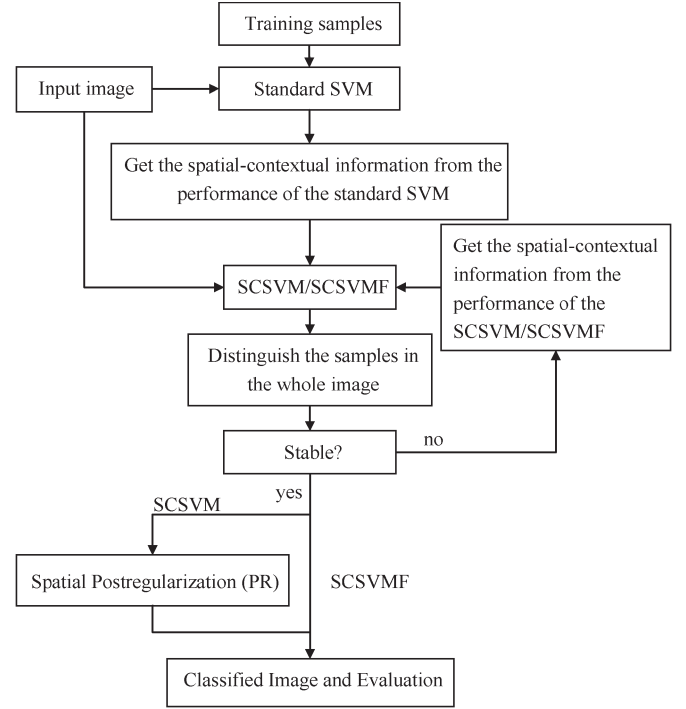


Fig. 9. Classification system by applying SCSVM and SCSVMF.

E. Relationship Between the SCSVM and the Standard SVM/LSVM With Generalized Bias Term and Regularized Kernel

The generalized bias term is based on a generalized regularizer, which leaves part of the hypothesis space unregularized, and a specific trick in constructing kernels [39]. The appropriate choice of the term will improve the applicability of the algorithm. The SVM with bias terms can be defined as

$$\min_{w, \xi} \quad \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \\ \text{subject to} \\ y_i (w^T \phi(x_i) + b_i) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, n \quad (7)$$

where b_i depends on the training sample x_i . The bias term is point adapted using the spatial-contextual information determined by the pixels in the neighborhood system of every training sample in this paper. That is, the SCSVM is equal to the SVM with the specific generalized bias term using the proposed regularization term

$$b_i = b + \gamma (m^+(x_i) - m^-(x_i)), \quad i = 1, 2, \dots, n.$$

If we consider the problem

$$\min_{\bar{w}, \xi} \quad \frac{1}{2} \bar{w}^T \bar{w} + C \sum_{i=1}^n \xi_i \\ \text{subject to} \\ y_i (\bar{w}^T \bar{\phi}(x_i) + \bar{b}_i) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, n \quad (8)$$

where $\bar{w}^T = [w^T, b]^T$, $\bar{\phi}(x_i)^T = [\phi(x_i)^T, 1]^T$, and $\bar{b}_i = \gamma(m^+(x_i) - m^-(x_i))$, then (8) is related to the Lagrangian

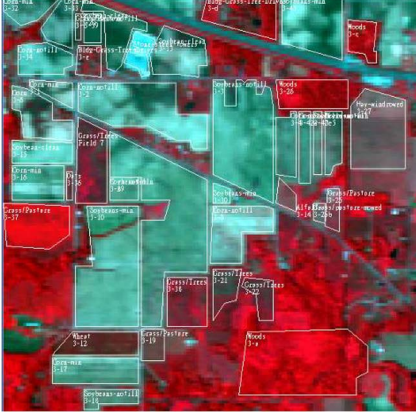
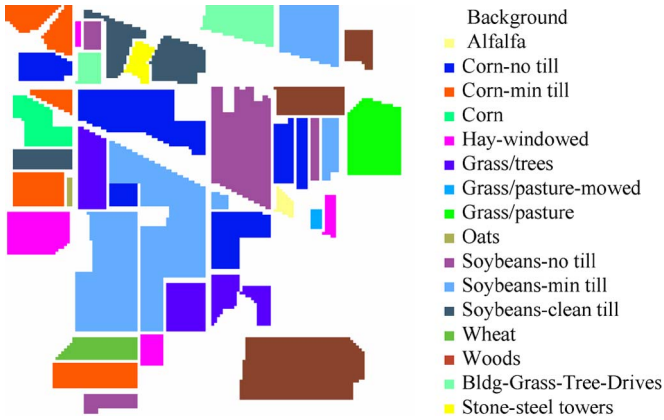
Fig. 10. Portion of the IPS image with a size of 145×145 pixels.

Fig. 11. Ground truth of the IPS data set.

SVM (LSVM) [40] with the proposed specific generalized bias term and regularized kernel. The corresponding dual problem is

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n (1 - y_i \bar{b}_i) \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \bar{\kappa}(x_i, x_j) \\ \text{subject to} \quad & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, n \end{aligned} \quad (9)$$

where $\bar{\kappa}(x_i, x_j) = \kappa(x_i, x_j) + 1$ is a regularized kernel.

III. EXPERIMENTAL DATA AND DESIGNS

This study uses two real data sets to evaluate the classification performance of the proposed SCSVM and SCSVMF, including the IPS, a mixed forest/agricultural site in Indiana [23], and the Washington, DC mall hyperspectral image [23] as an urban site.

The IPS data set was gathered by a sensor known as the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). This data set was obtained from an aircraft operated by the NASA/Jet Propulsion Laboratory and flown at an altitude of 65 000 ft. The images measured 145×145 pixels, with 220 spectral bands measuring approximately 20 m across the ground. Figs. 10 and 11 show the grayscale IR image and ground truth of IPS, respectively. There are 16 different land-cover classes available in the original ground-truth image. This study uses 16 categories: Alfalfa (class 1), Corn-no till (class 2), Corn-min till (class 3), Corn (class 4), Hay-windowed (class 5), Grass/trees (class 6),

TABLE I
SIXTEEN CATEGORIES AND CORRESPONDING
NUMBER OF PIXELS IN THE IPS IMAGE

No.	Category	#(pixels)	No.	Category	#(pixels)
1	Alfalfa	46	9	Oats	20
2	Corn-no till	1428	10	Soybeans-no till	972
3	Corn-min till	830	11	Soybeans-min till	2455
4	Corn	237	12	Soybeans-clean till	593
5	Hay-windowed	483	13	Wheat	205
6	Grass/trees	730	14	Woods	1265
7	Grass/pasture-mowed	28	15	Bldg-Grass-Tree-Drives	386
8	Grass/pasture	478	16	Stone-steel towers	93

No.	Category
1	Grass
2	Tree
3	Roof
4	Water
5	Road
6	Trail
7	Shadow



Fig. 12. False-color IR image of a portion of Washington, DC mall image measuring 205×307 pixels. There are seven categories: Grass, tree, roof, water, road, trail, and shadow.

Grass/pasture-mowed (class 7), Grass/pasture (class 8), Oats (class 9), Soybeans-no till (class 10), Soybeans-min till (class 11), Soybeans-clean till (class 12), Wheat (class 13), Woods (class 14), Bldg-Grass-Tree-Drives (class 15), and Stone-steel towers (class 16). Table I lists the number of pixels of each class.

This experiment randomly chose 10% of the samples for each class from the IPS reference data as training samples, which is the same method in [12]. The samples in the whole image were used as the testing set to evaluate the performance of the proposed algorithm.

The second data set, the Washington, DC mall, was obtained from the Hyperspectral Digital Imagery Collection Experiment (HYDICE) from an airborne hyperspectral data flight line over the Washington, DC urban area. Two hundred and ten bands were collected in the $0.4\text{--}2.4\text{-}\mu\text{m}$ region of the visible and infrared spectrum. Some water absorption channels were discarded, resulting in 191 channels. This data set is available in the student CD-ROM of [23]. This experiment used seven classes: grass (class 1), tree (class 2), roof (class 3), water (class 4), road (class 5), trail (class 6), and shadow (class 7). Fig. 12 shows the grayscale IR image of a portion of the image and the corresponding seven categories.

This study uses three distinct subsets, $N_u = 20 < N < d$ (case 1), $N_u = 40 < d < N$ (case 2), and $d < N_u = 300 < N$ (case 3), to investigate the influence of training sample size to the dimensionality in the Washington, DC hyperspectral image data set. In case 1, $N_u = 20 < N = 180 < d = 191$ is an ill-posed classification situation, which means that data dimensionality exceeds the number of independent training samples in every class. In case 2, $N_u = 40 < d = 191 < N = 360$

TABLE II
OVERALL ACCURACIES IN PERCENTAGES BY APPLYING SCSVM (OAA, M), SCSVM (OAO, M), SCSVMF (OAA, M), AND SCSVMF (OAO, M)
WITH DIFFERENT PARAMETERS M , THE SIZE OF THE NEIGHBORHOOD SYSTEM, AND γ IN THE IPS DATA SET

γ	0	0.05	0.1	0.3	0.5	1	10	100	500	1000	10000
SCSVM (OAA, 4)	86.5	89.0	90.2	92.4	92.5	92.5	91.7	89.1	91.9	91.6	88.8
SCSVM (OAA, 8)	86.5	91.5	93.3	94.7	95.2	95.4	94.8	94.7	94.6	94.3	93.7
SCSVM (OAO, 4)	84.4	86.9	89.6	90.1	87.3	84.0	84.0	84.0	84.3	84.2	84.2
SCSVM (OAO, 8)	84.4	90.1	92.9	88.3	86.5	88.0	87.0	87.3	87.1	87.0	87.0
SCSVMF (OAA, 4)	86.5	88.4	88.3	86.6	85.2	83.5	79.4	78.0	77.8	77.8	77.8
SCSVMF (OAA, 8)	86.5	88.3	87.6	84.6	83.6	82.3	78.5	77.8	77.8	77.8	77.8
SCSVMF (OAO, 4)	84.4	85.1	85.1	84.0	84.0	84.0	84.0	84.0	84.0	84.0	84.0
SCSVMF (OAO, 8)	84.4	85.3	84.4	84.0	84.0	84.0	84.0	84.0	84.0	84.0	84.0

is a poorly posed classification situation, which means that data dimensionality is greater than or comparable to the number of (independent) per-class representative training samples but smaller than the total number of representative samples. In case 3, there are enough independent training samples. MultiSpec [23] was used to select training and testing samples (100 testing samples per class) randomly in all experiments [29]–[31].

This study compares the classification performance of the proposed SCSVM and SCSVMF with OAO and OAA multiclass strategies and other reference classification algorithms, namely, ML classifier, ML_MRF classifier [8], k -NN classifier, standard SVM with OAO and OAA multiclass strategies, CS⁴VM (which is based on the OAA multiclass strategy) [1], and SVM+EM [12]. This experiment also compares the classification performance of SVM+EM and SCSVM with the PR step by a 3×3 mask and without the PR step. The SVM-based classifiers, including SVM, CS⁴VM, and SCSVM, employ the RBF kernel (i.e., the Gaussian radial basis function kernel). Both the IPS and the Washington, DC mall hyperspectral data sets were normalized to the range [0, 1]. A grid search with k -fold cross validation was used to find the proper $2\sigma^2$ within a range $[10^{-2}, 10]$ for the RBF kernel (suggested in [1]) and parameter C within a given set $\{0.1, 1, 10, 20, 60, 100, 160, 200, 1000\}$. Moreover, for CS⁴VM, the value κ_1/κ_2 was set to two and $C/\kappa_1 \in \{2, 4, 6, 8, 10, 12, 14\}$ following [1]. Since the semisamples were used to train CS⁴VM, only a first-order neighborhood system ∂x_i^O was considered for the context patterns to avoid spending too much time on training CS⁴VM. For SCSVM, only the decision function and constraints contain the spatial-contextual information of the neighborhood system. Using SCSVM, the training time increases a little for each round. The size of the neighborhood system M is set to four and eight in SCSVM for comparison. The term γ was set to 0.05, 0.1, 0.3, 0.5, 1, 10, 100, 500, 1000, and 10000 to determine the influence on spectral and spatial information.

The Gaussian function was adopted as the likelihood function of Bayesian decision rule for the ML classifier and ML_MRF classifier [5]. Several trials were carried out for the k -NN classifier, varying the value of k from 1 to 20 to identify the value that maximizes the accuracy. For simplicity, the model selection for the k -NN classifier was carried out based on the accuracy of the testing data set.

This study employs the following measures of classification accuracy to investigate classifier performance: 1) overall classification accuracy (the percentage of the correctly classified samples for all classes); 2) overall kappa coefficient (the percentage of the kappa coefficient for all classes); and 3) average

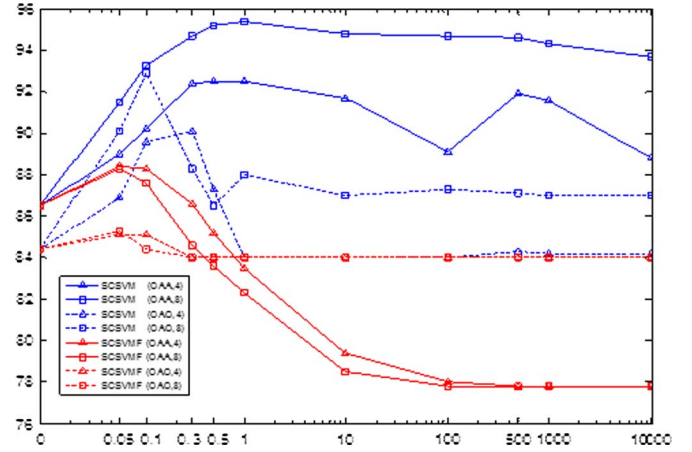


Fig. 13. Overall accuracies in percentages of the experimental classifiers, SCSVM and SCSVMF, for the IPS data set.

accuracy (the percentage of the average of correctly classified samples for each class). Since the number of testing data is the same for every class (i.e., $N_u = 100$) in the Washington, DC hyperspectral image data set, therefore the overall classification accuracy and the average accuracy are identical. In the IPS data set, the overall classification accuracy and the average accuracy are not identical due to the unequaled testing sample sizes of classes.

IV. EXPERIMENTAL RESULTS

This study compares the multiclass classification performance of the ML classifier, ML_MRF classifier, k -NN classifier, SVM, CS⁴VM, SVM+EM, and SCSVM. The following sections present the experimental results for the IPS data set and the Washington, DC mall data set.

A. IPS

According to the experimental design for IPS, 10% of the samples for each class were chosen as the training set. ML-based classifiers (ML and ML_MRF) require estimating the covariance matrices of the classes to classify all samples in the IPS hyperspectral image. These classifiers encounter the problem of covariance matrices and poor estimations because the number of training samples of the class is less than the dimensionality. Hence, the ML and ML_MRF performance in the IPS experiment is not compared and is denoted as N/A.

To investigate the effects of the neighborhood systems and parameters, M and γ , Table II and Fig. 13 show the overall accuracies by applying the SCSVM

TABLE III
OVERALL ACCURACIES, KAPPA COEFFICIENTS, AND AVERAGE ACCURACIES IN
PERCENTAGES OF THE EXPERIMENTAL CLASSIFIERS FOR THE IPS DATA SET

Classifier		Overall Accuracy (%)	Kappa Coefficient (%)	Average Accuracy (%)
ML		N/A	N/A	N/A
ML_MRF		N/A	N/A	N/A
k -NN		75.5	72.1	74.6
SVM (OAO)		84.4	82.3	85.5
SVM (OAA)		86.5	84.6	83.8
CS ⁴ VM		88.0	86.3	85.0
SVM+EM	<i>before PR</i>	91.3	90.0	81.6
	<i>after PR</i>	92.8	91.8	82.5
SCSVM (OAA, $M=8$, $\gamma=1$)	<i>before PR</i>	95.4	94.7	94.2
	<i>after PR</i>	95.5	94.9	94.1

TABLE IV
CLASS-SPECIFIC ACCURACIES IN PERCENTAGES FOR THE IPS DATA SET

No.	Class Sample size	k -NN	SVM(OAO)	SVM(OAA)	CS ⁴ VM	SVM+EM		SCSVM (OAA, $M=8$, $\gamma=1$)	
						<i>before PR</i>	<i>after PR</i>	<i>before PR</i>	<i>after PR</i>
1	46	78.3	91.3	95.7	95.7	93.5	93.5	93.5	100.0
2	1428	64.8	78.8	86.3	88.9	86.6	89.0	82.4	86.4
3	830	62.8	82.4	79.8	81.4	89.2	90.1	93.0	94.1
4	237	54.9	95.4	77.2	79.7	97.9	100.0	97.0	100.0
5	483	89.2	90.9	91.1	91.1	93.6	94.6	95.4	95.9
6	730	94.9	93.8	94.5	93.8	97.1	98.5	95.9	97.0
7	28	85.7	96.4	85.7	85.7	0.0	0.0	100.0	100.0
8	478	96.0	84.7	97.3	97.5	97.9	98.3	86.2	88.3
9	20	40.0	60.0	45.0	45.0	5.0	0.0	95.0	100.0
10	972	74.4	89.8	85.8	85.7	87.5	90.2	97.1	98.4
11	2455	76.7	79.6	86.6	88.7	92.7	94.2	94.4	96.4
12	593	50.8	77.7	75.4	82.6	92.6	92.9	90.2	91.2
13	205	98.0	99.5	99.5	99.5	99.0	99.0	99.5	100.0
14	1265	89.7	91.8	92.6	92.6	93.3	93.8	95.9	97.3
15	386	48.7	69.4	66.8	67.6	83.7	88.3	97.2	99.2
16	93	89.2	87.1	81.7	83.9	95.7	96.8	98.9	100.0

and SCSVMF with the grids of $M \in \{4, 8\}$ and $\gamma \in \{0, 0.05, 0.1, 0.3, 0.5, 1, 10, 100, 500, 1000, 10000\}$.

SCSVM is a generalized version of SVM, and SCSVM becomes the original SVM when $\gamma = 0$. These results show that SCSVM (OAA) can obtain higher overall accuracy than SCSVM (OAO) in the IPS data set regardless of M and γ . Fig. 13 shows that using the neighborhood system ∂x_i^O generally obtains better performance than using ∂x_i^F . Moreover, SCSVM with $M = 8$, a second-order neighborhood system, outperforms $M = 4$, a first-order neighborhood system, because the IPS is a larger spatial structure image in the original space. In hyperspectral image classification, many samples from different land-cover classes but having very similar spectral properties [8], [9] affect the performance of SCSVMF. Specifically, the SCSVMF performance increases only a little and is even worse than the SVM performance when γ exceeds a threshold. The highest overall SCSVM accuracy of 95.4% occurred at $M = 8$ and $\gamma = 1$ with the OAA multiclass strategy and the neighborhood system ∂x_i^O .

Tables III and IV present the validation measures of all samples in the IPS and class-specific accuracies from the best performance of k -NN classifier ($k = 1$), SVM (OAO and OAA multiclass strategy), CS⁴VM, SVM+EM with and without the PR step, and SCSVM, which has the highest accuracy in Table II, with and without the PR step. The best overall accuracy,

kappa coefficient, and average accuracy are highlighted in gray. “Before PR” means that the performance of the classifier does not include the PR step. “After PR” means that the performance of the classifier includes the PR step.

Due to length constraints, the classification maps with the highest accuracies of each type of classifier are shown for comparison. Fig. 14(a)–(h) shows the classification maps of the IPS hyperspectral image by k -NN ($k = 1$), SVM (OAO), SVM (OAA), CS⁴VM, SVM+EM, and SCSVM (OAA) with $M = 8$ and $\gamma = 0.1$, respectively. To conveniently compare the performances, we put the ground truth of IPS image in Fig. 14(i).

The classification results from Table III, Table IV, and Fig. 14 present the following findings.

- 1) In terms of accuracy, SCSVM (OAA) with the PR step obtained the highest overall accuracy and kappa coefficient (95.5% and 94.9%, respectively) (Table III). However, SCSVM (OAA) without the PR step can obtain the highest average accuracy, up to 94.2%.
- 2) Only 10% of each sample was selected randomly from the reference data set to serve as the training set for each class. Therefore, some classes were represented by only few training samples (i.e., there are only three and two samples for class 7 (Grass/pasture-mowed) and class 9

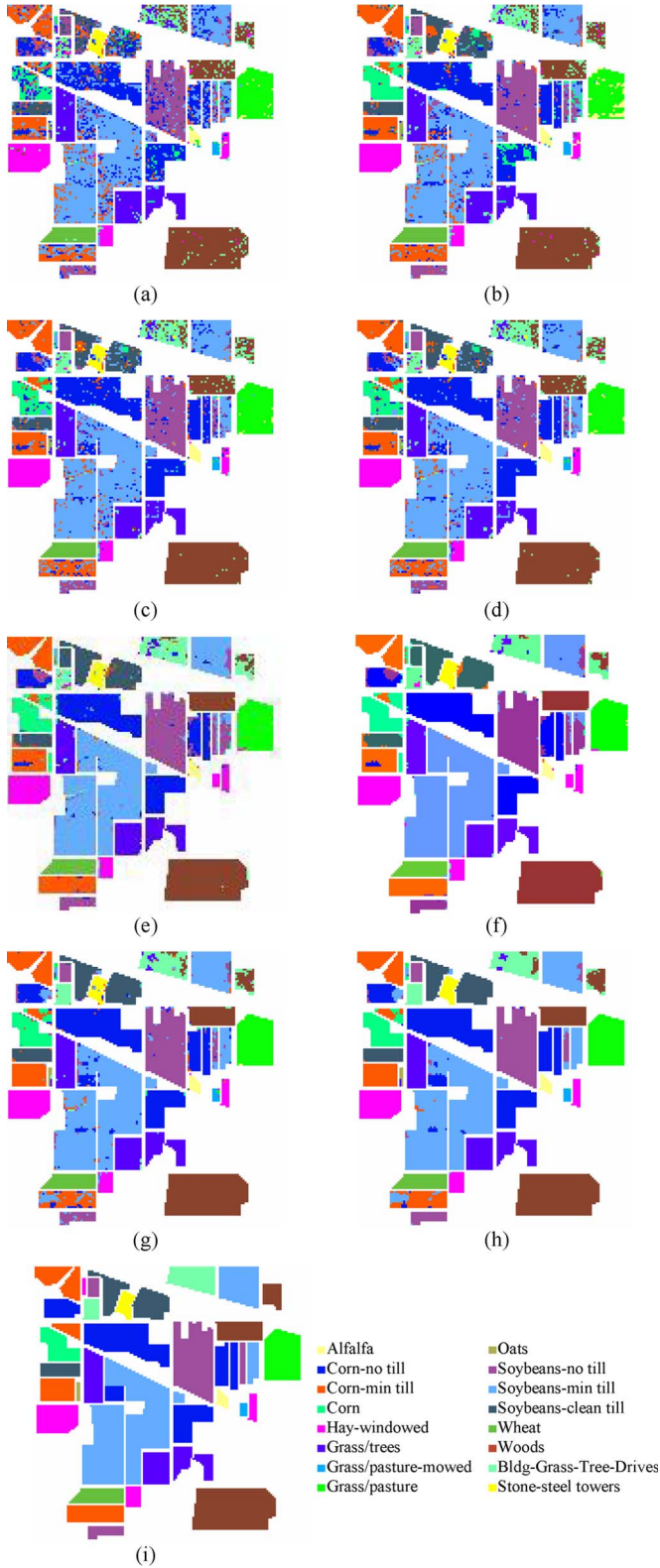


Fig. 14. Classification maps of the IPS data set by the highest performance of each type of classifier. (a) k -NN. (b) SVM (OAO). (c) SVM (OAA). (d) CS^4VM . (e) SVM+EM. (f) SVM+EM (PR). (g) SCSVM (OAA, $M = 8$, $\gamma = 1$). (h) SCSVM (OAA, $M = 8$, $\gamma = 1$, PR). (i) Ground truth.

(Oats), respectively). This may provide an unfair representation of this class in the training process. Table IV shows that the classification performance for class 7

with SCSVM (OAA) was better than that with k -NN, SVM (OAO), SVM (OAA), CS^4VM , and SVM+EM. Moreover, applying SCSVM (OAA) achieved better performance than other classifiers in class 9. Therefore, this situation was improved by SCSVM (OAA), even without the PR step, and the classification accuracies of class 7 and class 9 with the PR step were 100%.

- 3) The classifiers with the PR step efficiently reduced some noise in the classification map and slightly increased classification accuracy.
- 4) Table III and the classification maps in Fig. 14 show that classifiers based on spatial information (SVM+EM, CS^4VM , and SCSVM) were much better than the classifiers based on only spectral information [k -NN, SVM (OAO), and SVM (OAA)]. Furthermore, the spatial-contextual-based classifiers reduced the number of specklelike errors, particularly in the areas of Soybeans-min till, Soybeans-no till, and Corn-no till, which were the most difficult parts to classify accurately. SCSVM (OAA) achieves a great improvement in the classification map. The SCSVM (OAA) classification map [Fig. 14(h)] was approximate to the ground truth [Fig. 14(i)] of the IPS.
- 5) SVM+EM also achieved sound performance on the classification maps and obtained a 92.8% overall classification accuracy. However, this scheme relies on the partitional clustering results. Hence, if the partitional clustering technique is not good for partitioning these areas, which have similar spectral properties from different classes or come from the small-sample-size classes [i.e., class 7 (Grass/pasture-mowed) and class 9 (Oats)], then these areas will be misclassified and classified by the clustering technique into the same class [see Fig. 14(f)]. If the partitional clustering technique works very well, but the standard SVM classifier cannot sensitively distinguish the pixels, e.g., classes which have similar spectral properties from different classes, then these areas will be sacrificed. Table IV shows that SVM+EM achieved either 0%, e.g., class 7 and class 9, or low classification accuracies for the small classes.
- 6) Since CS^4VM is based on the OAA multiclass strategy [1], the class-specific accuracies for applying CS^4VM are higher than or similar to those for applying SVM (OAA).

B. Washington, DC Mall

The experiments in this study used three cases to investigate the effects of sample size on the dimensionality in the Washington, DC mall data set. Note that the sample covariance matrices of ML-based classifiers, ML and ML_MRF, of case 1 and case 2 will be singular. Hence, in case 1 and case 2, ML-based classifiers would not work, and the performances of ML-based classifiers are shown as N/A.

Similar to the IPS data set, Tables V–VII and Figs. 15–17, respectively, show the overall accuracies of SCSVM (OAA, OAO) and SCSVMF (OAA, OAO) with the grids of $M \in \{4, 8\}$ and $\gamma \in \{0, 0.05, 0.1, 0.3, 0.5, 1, 10, 100, 500, 1000, 10000\}$ to investigate the influence of the parameters M and

TABLE V
OVERALL ACCURACIES IN PERCENTAGES BY APPLYING SCSVM (OAA, M), SCSVM (OAO, M), SCSVMF (OAA, M), AND SCSVMF (OAO, M) WITH DIFFERENT PARAMETERS M , THE SIZE OF THE NEIGHBORHOOD SYSTEM, AND γ IN WASHINGTON, DC MALL DATA SET (CASE 1)

γ	0	0.05	0.1	0.3	0.5	1	10	100	500	1000	10000
SCSVM (OAA, 4)	86.9	91.9	91.3	89.7	90.0	90.0	91.1	91.0	91.0	91.0	91.0
SCSVM (OAA, 8)	86.9	90.4	90.6	90.6	90.6	90.6	90.6	90.6	90.6	90.6	90.6
SCSVM (OAO, 4)	86.9	91.0	91.7	90.1	90.4	90.9	89.6	89.6	89.6	89.6	89.6
SCSVM (OAO, 8)	86.9	90.7	91.3	90.7	91.0	90.7	90.4	90.6	90.6	90.6	90.6
SCSVMF (OAA, 4)	86.9	88.3	88.3	88.1	88.0	88.1	87.6	87.7	87.7	87.7	87.7
SCSVMF (OAA, 8)	86.9	87.7	87.9	88.0	87.9	87.9	87.6	87.6	87.6	87.6	87.6
SCSVMF (OAO, 4)	86.9	86.6	86.6	86.6	86.4	85.9	86.4	86.4	86.4	86.4	86.4
SCSVMF (OAO, 8)	86.9	85.7	85.7	85.7	85.7	86.0	86.0	85.7	85.7	85.7	85.7

TABLE VI
OVERALL ACCURACIES IN PERCENTAGES BY APPLYING SCSVM (OAA, M), SCSVM (OAO, M), SCSVMF (OAA, M), AND SCSVMF (OAO, M) WITH DIFFERENT PARAMETERS M , THE SIZE OF THE NEIGHBORHOOD SYSTEM, AND γ IN WASHINGTON, DC MALL DATA SET (CASE 2)

γ	0	0.05	0.1	0.3	0.5	1	10	100	500	1000	10000
SCSVM (OAA, 4)	89.0	93.4	93.3	93.4	93.7	93.7	93.4	93.3	93.1	93.1	93.1
SCSVM (OAA, 8)	89.0	91.9	92.7	93.1	92.9	93.0	92.9	92.4	92.4	92.4	92.4
SCSVM (OAO, 4)	88.6	92.4	93.6	94.0	93.7	92.4	91.0	90.4	90.4	90.4	90.7
SCSVM (OAO, 8)	88.6	93.0	94.1	92.4	92.6	92.4	92.0	91.9	91.9	91.9	91.9
SCSVMF (OAA, 4)	89.0	88.9	88.9	89.1	89.3	89.3	89.1	89.3	89.3	89.3	89.3
SCSVMF (OAA, 8)	89.0	88.9	88.7	88.7	88.9	89.0	88.9	88.9	88.9	88.9	88.9
SCSVMF (OAO, 4)	88.6	88.6	88.6	88.9	88.9	88.3	88.0	88.1	88.0	88.1	88.1
SCSVMF (OAO, 8)	88.6	88.1	88.3	88.3	88.1	87.9	88.0	88.0	88.0	88.0	88.0

TABLE VII
OVERALL ACCURACIES IN PERCENTAGES BY APPLYING SCSVM (OAA, M), SCSVM (OAO, M), SCSVMF (OAA, M), AND SCSVMF (OAO, M) WITH DIFFERENT PARAMETERS M , THE SIZE OF THE NEIGHBORHOOD SYSTEM, AND γ IN WASHINGTON, DC MALL DATA SET (CASE 3)

γ	0	0.05	0.1	0.3	0.5	1	10	100	500	1000	10000
SCSVM (OAA, 4)	93.7	95.3	95.6	95.9	96.1	97.4	96.9	95.7	95.3	95.3	95.3
SCSVM (OAA, 8)	93.7	94.4	94.6	95.1	95.3	95.3	95.4	95.5	94.4	94.4	94.4
SCSVM (OAO, 4)	94.3	94.9	97.3	98.6	98.1	97.3	96.1	95.4	96.1	96.1	96.1
SCSVM (OAO, 8)	94.3	95.9	97.3	96.9	96.0	96.0	95.0	95.4	95.1	95.1	95.1
SCSVMF (OAA, 4)	93.7	94.3	94.4	94.3	94.1	94.3	94.9	94.1	94.1	94.1	94.1
SCSVMF (OAA, 8)	93.7	94.1	94.4	94.4	94.4	94.6	94.4	94.4	94.4	94.4	94.4
SCSVMF (OAO, 4)	94.3	93.7	94.0	93.9	93.6	93.4	93.4	93.1	93.1	93.1	93.1
SCSVMF (OAO, 8)	94.3	93.9	94.0	93.9	93.7	93.4	93.6	93.4	93.4	93.4	93.4

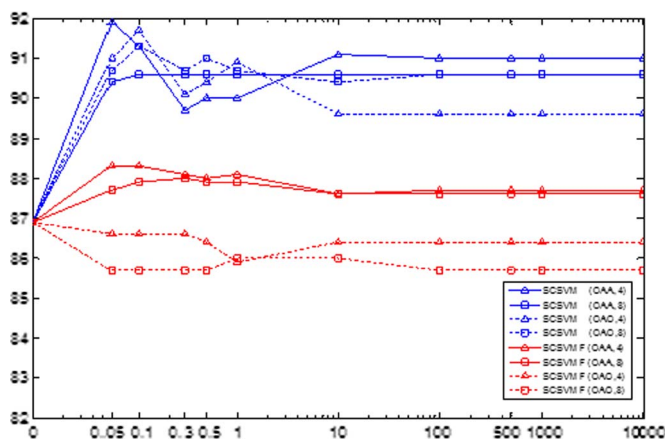


Fig. 15. Overall accuracies in percentages of the experimental classifiers, SCSVM and SCSVMF, for the Washington, DC mall data set in case 1.

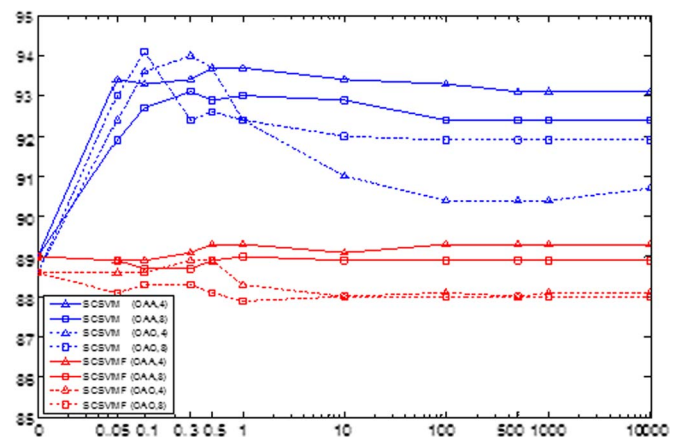


Fig. 16. Overall accuracies in percentages of the experimental classifiers, SCSVM and SCSVMF, for the Washington, DC mall data set in case 2.

γ in three cases. In case 1, the highest accuracy of 91.9% occurred at $M = 4$ and $\gamma = 0.05$ with the neighborhood system in the original space (SCSVM) and using the OAA multiclass strategy. However, SCSVM (OAO, 4) has a similar accuracy, i.e., 91.7%. In case 2, the highest accuracy is 94.1%

at $M = 8$ and $\gamma = 0.1$ with the neighborhood system in the original space (SCSVM) and OAO multiclass strategy. However, SCSVM (OAO, 4) has a similar accuracy, at 94.0%. In case 3, the highest accuracy of 98.6% occurred at $M = 4$ and $\gamma = 0.3$ with the neighborhood system in the original

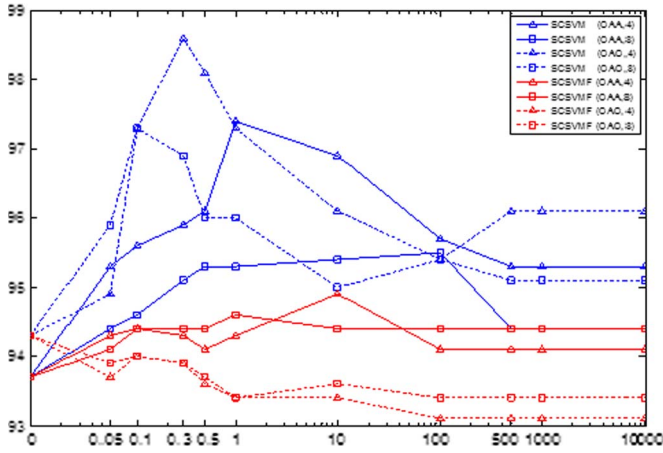


Fig. 17. Overall accuracies in percentages of the experimental classifiers, SCSVM and SCSVMF, for the Washington, DC mall data set in case 3.

space (SCSVM) with OAO multiclass strategy. The SCSVM (OAO) performance was generally better than or similar to that of SCSVM (OAA) among the three cases. Furthermore, SCSVM with $M = 4$ has better performance than or similar performance to SCSVM with $M = 8$. This is because the Washington, DC mall image is an urban site which is not a large spatial structure. Figs. 15–17 show that the overall accuracy increases as the training sample size increases. Similar to the IPS data set, a lot of samples, which come from different land-cover classes, but have very similar spectral properties [8], [9], affect the performance of SCSVMF. That is, SCSVMF only increases a little classification accuracy or is even worse than the performance by applying SVM when γ is greater than a threshold. The performance by applying SCSVMF (OAO) is much worse than that by applying SVM (OAO).

Tables VIII, X, and XII show the overall accuracies, kappa coefficients, and average accuracies by applying the k -NN classifier with $k = 1$, which has the best classification performance of testing set, and SVM (OAO and OAA), CS^4VM , SVM+EM with and without the PR step, and SCSVM with and without the PR step for all three cases of the Washington, DC mall data set. Tables IX, XI, and XIII display the class-specific accuracies of these classifiers. Since the number of testing samples is the same for every class in the Washington, DC hyperspectral image data set (i.e., equal to 100), the overall classification accuracy and the average accuracy are identical in Tables VIII, X, and XII, and the first decimal points are all zero in Tables IX, XI, and XIII.

Similar to the IPS data set, Tables VIII, X, and XII and Tables IX, XI, and XIII, respectively, compare the validation measures by applying SCSVM with different parameters, M and γ , which have the highest accuracies in Tables V–VII. Due to length constraints, Fig. 18 only compares the classification maps with the highest accuracies of each type of classifier in case 3.

The classification results in Tables VIII–XIII and Fig. 18 reveal the following findings.

- 1) SCSVM obtained the highest classification accuracies of the testing set in terms of overall accuracy, kappa

coefficient, and average classification for all cases. In case 1, SCSVM (OAA) with $M = 4$, $\gamma = 0.05$, and the PR step has the best classification accuracy. The overall accuracy, kappa coefficient, and average classification were 92.0%, 90.6%, and 92.0%, respectively. In case 2, SCSVM (OAA) with $M = 8$ and $\gamma = 0.1$ achieved the best classification accuracy. The overall accuracy, kappa coefficient, and average classification were 94.1%, 93.2%, and 94.1%, respectively. In case 3, SCSVM (OAO) with $M = 4$ and $\gamma = 0.3$ and without the PR step achieved the best classification accuracy. The overall accuracy, kappa coefficient, and average classification were 98.6%, 98.3%, and 98.6%, respectively. As the training sample size increased, the accuracies represent ascending tendencies in all classifiers.

- 2) Because the Washington, DC image is an urban site image, some areas in this image are small spatial structures, and some areas are large spatial structures. The PR step did not really work effectively for small spatial structure areas, e.g., class 6 (trail) in Fig. 18. Tables IX, XI, and XIII show that the accuracy decreased upon applying the PR step in class 6 (trail). However, some noise pixels in large structure areas, e.g., class 4 (water), were improved after the PR step (except for case 1, which encountered the small-sample-size problem).
- 3) The image contains too many kinds of roofs. Hence, when the training sample size is small, e.g., in class 3 (roof), some kinds of roofs may not be chosen as training samples. For this reason, the classification accuracies and maps of this class are poor in case 1 and case 2 (Tables IX and XI) regardless of the classifier. In the general case (case 3), these classes have been identified more correctly (Table XIII).
- 4) Fig. 19 shows classified images of SCSVM (OAO) and SCSVM (OAA) with $M = 4$ and $\gamma = 0, 0.1$, and 0.3 , revealing the effect of parameter γ on the classified image. As the gamma increased, the classified image exhibited more homogeneous groups of pixels.
- 5) Most of the spatial-information-based classifiers (ML_MRF, CS^4VM , and SCSVM) achieved better classification performance than the spectral-information-only-based classifiers (ML, k -NN, and SVM) except SVM+EM in cases 1 and 2, because SVM+EM is particularly suitable for classification of images with large spatial structures [12]. The drawback of SVM+EM is that, when including spatial information from the segmentation map or from the closest neighborhoods in a classifier, small spatial structures may be assimilated with larger neighboring structures if the spectral responses are not significantly different [12]. Hence, SVM+EM is not really suitable for the small areas of the Washington, DC image. However, the classification accuracy and maps of class 4 (water), which is a large structure, can be improved, and the class-specific accuracy of the class (water) is 100% (Tables IX, XI, and XIII) for all three cases.

TABLE VIII
OVERALL ACCURACIES, KAPPA COEFFICIENTS, AND AVERAGE ACCURACIES IN PERCENTAGES
OF THE EXPERIMENTAL CLASSIFIERS FOR THE WASHINGTON, DC MALL DATA SET (CASE 1)

Classifier		Overall Accuracy	Kappa Coefficient	Average Accuracy
ML		N/A	N/A	N/A
ML_MRF		N/A	N/A	N/A
k -NN		85.6	83.2	85.6
SVM (OAO)		86.9	84.7	86.9
SVM (OAA)		86.9	84.7	86.9
CS ⁴ VM		87.7	85.7	87.7
SVM+EM	<i>before PR</i>	82.0	79.0	82.0
	<i>after PR</i>	80.3	77.0	80.3
SCSVM (OAA, $M=4$, $\gamma=0.05$)	<i>before PR</i>	91.9	90.5	91.9
	<i>after PR</i>	92.0	90.6	92.0

TABLE IX
CLASS-SPECIFIC ACCURACIES IN PERCENTAGES FOR THE WASHINGTON, DC MALL DATA SET IN CASE 1

No.	Class Sample size	k -NN	SVM(OAO)	SVM(OAA)	CS ⁴ VM	SVM+EM		SCSVM (OAA, $M=4$, $\gamma=0.05$)	
						<i>before PR</i>	<i>after PR</i>	<i>before PR</i>	<i>after PR</i>
1	100	79.0	78.0	81.0	86.0	80.0	83.0	92.0	92.0
2	100	82.0	94.0	93.0	99.0	93.0	92.0	100.0	100.0
3	100	58.0	66.0	60.0	57.0	62.0	60.0	93.0	91.0
4	100	98.0	93.0	94.0	96.0	100.0	100.0	98.0	97.0
5	100	94.0	95.0	95.0	98.0	54.0	51.0	83.0	86.0
6	100	90.0	90.0	91.0	85.0	91.0	83.0	85.0	84.0
7	100	98.0	92.0	94.0	93.0	94.0	93.0	92.0	94.0

TABLE X
OVERALL ACCURACIES, KAPPA COEFFICIENTS, AND AVERAGE ACCURACIES IN PERCENTAGES
OF THE EXPERIMENTAL CLASSIFIERS FOR THE WASHINGTON, DC MALL DATA SET (CASE 2)

Classifier		Overall Accuracy	Kappa Coefficient	Average Accuracy
ML		N/A	N/A	N/A
ML_MRF		N/A	N/A	N/A
k -NN		87.1	85.0	85.6
SVM (OAO)		88.6	86.7	88.6
SVM (OAA)		89.0	87.2	89.0
CS ⁴ VM		89.4	87.7	89.4
SVM+EM	<i>before PR</i>	86.7	84.5	86.7
	<i>after PR</i>	84.6	82.0	84.6
SCSVM (OAO, $M=8$, $\gamma=0.1$)	<i>before PR</i>	94.1	93.2	94.1
	<i>after PR</i>	94.1	93.2	94.1

TABLE XI
CLASS-SPECIFIC ACCURACIES IN PERCENTAGES FOR THE WASHINGTON, DC MALL DATA SET IN CASE 2

No.	Class Sample size	k -NN	SVM(OAO)	SVM(OAA)	CS ⁴ VM	SVM+EM		SCSVM (OAO, $M=8$, $\gamma=0.1$)	
						<i>before PR</i>	<i>after PR</i>	<i>before PR</i>	<i>after PR</i>
1	100	86.0	95.0	95.0	97.0	98.0	99.0	98.0	98.0
2	100	84.0	95.0	95.0	100.0	96.0	95.0	100.0	100.0
3	100	66.0	60.0	59.0	54.0	58.0	56.0	82.0	82.0
4	100	98.0	98.0	98.0	98.0	100.0	100.0	100.0	100.0
5	100	96.0	90.0	91.0	98.0	70.0	66.0	98.0	98.0
6	100	87.0	91.0	92.0	85.0	91.0	83.0	84.0	84.0
7	100	93.0	91.0	93.0	94.0	94.0	93.0	97.0	97.0

6) The CS⁴VM classifier is based on the OAA multiclass strategy. This classifier achieved slightly better classification accuracy than SVM (OAA) (see Tables VIII, X, and XII), but SCSVM still achieved better performance than CS⁴VM. However, CS⁴VM distinguished some areas better than SCSVM.

V. CONCLUSION

This study has proposed two kinds of spatial-contextual SVMs. An SCSVM that is based on the neighborhood system in the original space can overcome similar spectral properties. SCSVM modifies the decision function and the constraints of SVM based on spatial-contextual information. To reduce the

TABLE XII
OVERALL ACCURACIES, KAPPA COEFFICIENTS, AND AVERAGE ACCURACIES IN PERCENTAGES
OF THE EXPERIMENTAL CLASSIFIERS FOR THE WASHINGTON, DC MALL DATA SET (CASE 3)

Classifier		Overall Accuracy	Kappa Coefficient	Average Accuracy
ML		94.1	93.2	94.1
ML_MRF		96.7	96.2	96.7
k -NN		94.4	93.5	94.4
SVM (OAO)		94.3	93.3	94.3
SVM (OAA)		93.7	92.7	93.7
CS ⁴ VM		94.1	93.2	94.1
SVM+EM	<i>before PR</i>	94.6	93.7	94.6
	<i>after PR</i>	92.9	91.7	92.9
SCSVM	<i>before PR</i>	98.6	98.3	98.6
(OAO, $M=4$, $\gamma=0.3$)	<i>after PR</i>	98.4	98.1	98.4

TABLE XIII
CLASS-SPECIFIC ACCURACIES IN PERCENTAGES FOR THE WASHINGTON, DC MALL DATA SET IN CASE 3

No.	Class Sample size	ML	ML_MRF	k -NN	SVM (OAO)	SVM (OAA)	CS ⁴ VM	SVM+EM		SCSVM (OAO, $M=4$, $\gamma=0.3$)	
								<i>before PR</i>	<i>after PR</i>	<i>before PR</i>	<i>after PR</i>
1	100	99.0	100.0	95.0	97.0	96.0	99.0	98.0	99.0	99.0	99.0
2	100	99.0	99.0	96.0	99.0	99.0	99.0	93.0	92.0	100.0	100.0
3	100	90.0	94.0	80.0	78.0	75.0	76.0	85.0	85.0	99.0	100.0
4	100	98.0	99.0	100.0	98.0	98.0	98.0	100.0	100.0	99.0	100.0
5	100	93.0	99.0	99.0	100.0	100.0	100.0	98.0	97.0	99.0	100.0
6	100	85.0	90.0	93.0	90.0	90.0	89.0	91.0	83.0	97.0	94.0
7	100	95.0	96.0	98.0	98.0	98.0	98.0	97.0	94.0	97.0	96.0

remaining noise in the classification map, a PR step, a fixed-window-based postfiltering, was employed. The experiments in this study compared and analyzed the effects of different types of classifiers on the classification accuracy and classification map by applying the proposed SCSVM, ML classifier, ML_MRF classifier, k -NN classifier, a standard supervised SVM, a CS⁴VM, and SVM+EM.

The experimental results obtained from two different hyperspectral image data sets, the IPS (a mixed forest/agricultural site in Indiana) and the Washington, DC mall hyperspectral image (an urban site in Washington, DC), confirm that the proposed SCSVM improves the classification accuracies and kappa coefficients.

The earlier discussion leads to the following conclusions.

- 1) SCSVM (OAA) is better than or similar to SCSVM (OAO) in the IPS data set. The classification map of IPS data set obtained from SCSVM (OAA) with the PR step [Fig. 14(h)] is very close to the ground truth, and the SCSVM classification accuracy and kappa coefficient are 95.5% and 94.9%, respectively. However, in the Washington, DC mall data set, SCSVM (OAO) is better than or similar to SCSVM (OAA), and SCSVMF (OAA) is better than or similar to SCSVMF (OAO).
- 2) This study shows that selecting a suitable spatial parameter γ improves the SCSVM performance, and the best choice of γ becomes larger as the training sample size increases. That is, γ has a strong influence on the performance, particularly for the SCSVM (OAA).
- 3) The computational cost of the learning phase in the proposed SCSVM is slightly higher than that of the standard

SVM in each round. From a theoretical point of view, a standard supervised SVM is a special case of SCSVM, if the parameter γ is equal to zero. However, CS⁴VM requires a huge semisample set from the neighborhoods of each training sample in the objective function. Hence, the computational cost of the CS⁴VM learning phase is slightly higher than that of the SCSVM learning phase because SCSVM only uses the same training sample in the objective function in each round. For example, in the IPS data set experiment, carried out on a PC mounting an Intel Core 2 Duo CPU at 2.4 GHz and a 4-GB DDR2 RAM, the training phase of a supervised SVM (OAA) took about 7.566 s, and the training phase of SCSVM (OAA) took about 7.909 s, but the training phase of CS⁴VM required about 185.56 s.

- 4) The SVM+EM method is particularly suitable for classifying images with large spatial structures, e.g., the IPS image, when the spectral responses of different classes are dissimilar and the classes contain a comparable number of pixels. However, most real data do not always satisfy this condition (e.g., the Washington, DC mall image). Hence, SVM+EM is not generally suitable for all situations. In the SCSVM classifier, for different data sets, the spatial neighborhood system can be modified according to the spatial structure.

ACKNOWLEDGMENT

The authors would like to thank two anonymous reviewers for their suggestions and Dr. Landgrebe for providing the Indian Pines site and The Washington DC mall data sets.

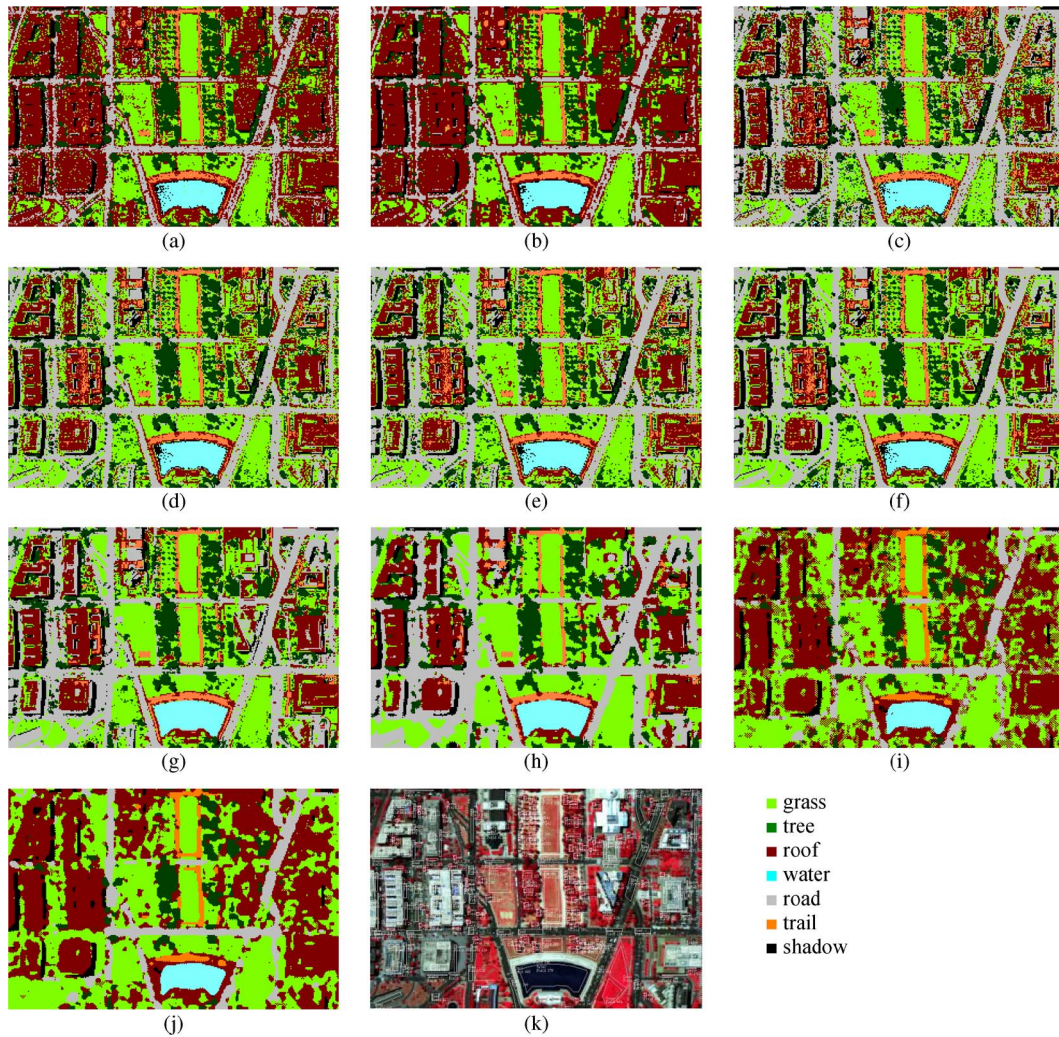


Fig. 18. Classification maps of a portion of the Washington, DC data set (case 3) by the highest performance of each type of classifier. (a) ML classifier. (b) ML_MRF. (c) k -NN. (d) SVM (OAO). (e) SVM (OAA). (f) CS^4VM . (g) SVM+EM. (h) SVM+EM (PR). (i) SCSVM (OAO, $M = 4$, $\gamma = 0.3$). (j) SCSVM (OAO, $M = 4$, $\gamma = 0.3$, PR). (k) False-color IR image of the portion of Washington, DC mall image.

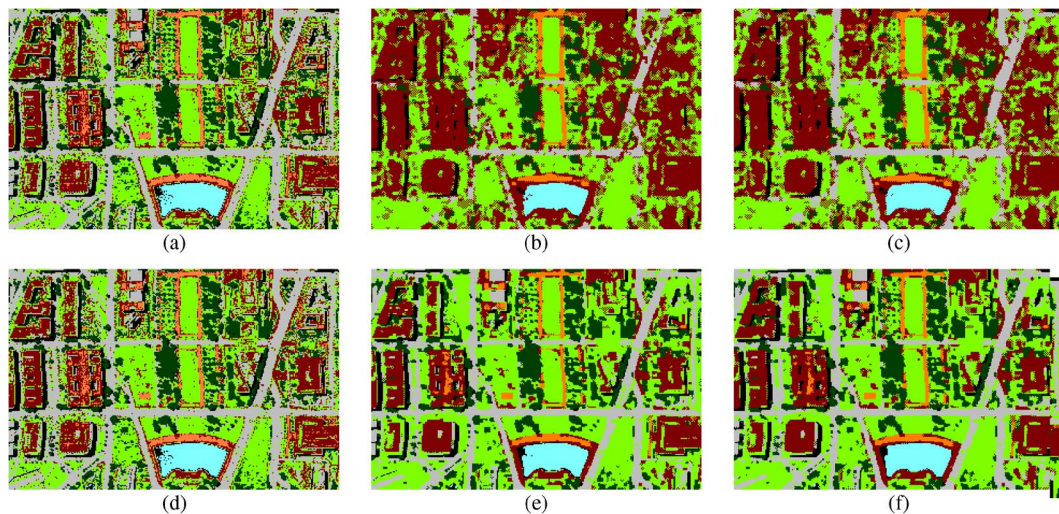


Fig. 19. Classification maps of a portion of the Washington, DC data set (case 3) by applying SCSVM (OAO) and SCVM (OAA) with $M = 4$ and different parameters $\gamma = 0, 0.1$, and 0.3 . (a) SCSVM (OAO), $\gamma = 0$. (b) SCSVM (OAO), $\gamma = 0.1$. (c) SCSVM (OAO), $\gamma = 0.3$. (d) SCSVM (OAA), $\gamma = 0$. (e) SCSVM (OAA), $\gamma = 0.1$. (f) SCSVM (OAA), $\gamma = 0.3$.

REFERENCES

- [1] L. Bruzzone and C. Persello, "A novel context-sensitive semisupervised SVM classifier robust to mislabeled training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2142–2154, Jul. 2009.
- [2] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
- [3] 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.
- [4] B.-C. Kuo and K.-Y. Chang, "Feature extractions for small sample size classification problem," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 756–764, Mar. 2007.
- [5] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. San Diego, CA: Academic, 1990.
- [6] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. Workshop Comput. Learn. Theory*, 1992, pp. 144–152.
- [7] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. New York: Springer-Verlag, 2001.
- [8] Q. Jackson and D. A. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, Nov. 2002.
- [9] B.-C. Kuo, C.-H. Chuang, C.-S. Huang, and C.-C. Hung, "A nonparametric contextual classification based on Markov random fields," in *Proc. 1st WHISPERS—Evolution in Remote Sensing*, pp. 1–4.
- [10] 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.
- [11] J. Li and R. M. Narayanan, "Integrated spectral and spatial information mining in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 673–685, Mar. 2004.
- [12] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitioning clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.
- [13] G. Ball and D. Hall, "ISODATA, a novel method of data analysis and classification," Stanford Univ., Stanford, CA, Tech. Rep. AD-699616, 1965.
- [14] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B*, vol. 39, no. 1, pp. 1–38, 1977.
- [15] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [16] S. T. John and C. Nello, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [17] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.
- [18] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Evaluation of kernels for multiclass classification of hyperspectral remote sensing data," in *Proc. ICASSP*, May 2006, pp. II-813–II-816.
- [19] G. Camps-Valls, L. Gómez-Chova, J. Calpe-Maravilla, E. Soria, J. D. Martín-Guerrero, L. Alonso-Chorda, and J. Moreno, "Robust support vector method for hyperspectral data classification and knowledge discovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1530–1542, Jul. 2004.
- [20] S. Knerr, L. Personnaz, and G. Dreyfus, "Single-layer learning revisited: A stepwise procedure for building and training a neural network," in *Neurocomputing: Algorithms, Architectures and Applications*, J. Fogelman, Ed. New York: Springer-Verlag, 1990.
- [21] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.
- [22] L. Bottou, C. Cortes, J. Denker, H. Drucker, I. Guyon, L. Jackel, Y. LeCun, U. Muller, E. Sackinger, P. Simard, and V. Vapnik, "Comparison of classifier methods: A case study in handwriting digit recognition," in *Proc. Int. Conf. Pattern Recognit.*, 1994, pp. 77–87.
- [23] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ: Wiley, 2003.
- [24] B.-C. Kuo and D. A. Landgrebe, "A covariance estimator for small sample size classification problems and its application to feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 4, pp. 814–819, Apr. 2002.
- [25] R. Kinderman and J. L. Snell, "Markov random fields and their applications," *Amer. Math. Soc.*, vol. 1, pp. 1–142, 1980.
- [26] B.-C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 5, pp. 1096–1105, May 2000.
- [27] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
- [28] S. B. Serpico and G. Moser, "Extraction of spectral channels from hyperspectral images for classification purposes," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 484–495, Feb. 2007.
- [29] B.-C. Kuo, C.-H. Li, and J.-M. Yang, "Kernel nonparametric weighted feature extraction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1139–1155, Apr. 2009.
- [30] J.-M. Yang, P.-T. Yu, and B.-C. Kuo, "A nonparametric feature extraction and its application to nearest neighbor classification for hyperspectral image data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1279–1293, Mar. 2010.
- [31] J.-M. Yang, B.-C. Kuo, P.-T. Yu, and C.-H. Chuang, "A dynamic subspace method for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2840–2853, Jul. 2010.
- [32] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.
- [33] K. Ersahin, I. G. Cumming, and R. K. Ward, "Segmentation and classification of polarimetric SAR data using spectral graph partitioning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 164–174, Jan. 2010.
- [34] G. Camps-Valls, T. V. B. Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.
- [35] G. Camps-Valls, N. Shervashidze, and K. M. Borgwardt, "Spatio-spectral remote sensing image classification with graph kernels," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 741–745, Oct. 2010.
- [36] M. Fauvel, "Spectral and spatial methods for the classification of urban remote sensing data," Ph.D. dissertation, Grenoble Inst. Technol., Grenoble, France, 2007.
- [37] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.
- [38] M. Fauvel, J. Chanussot, J. A. Benediktsson, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.
- [39] W. Li, K.-S. Leung, and K.-H. Lee, "Generalizing the bias term of support vector machines," in *Proc. Int. Joint Conf. Artif. Intell.*, Jan. 2007, pp. 919–924.
- [40] O. Mangasarian and D. Musicant, "Lagrangian support vector machines," *J. Mach. Learn. Res.*, vol. 1, pp. 161–177, Sep. 2001.
- [41] V. Madhok and D. A. Landgrebe, "A process model for remote sensing data analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 3, pp. 680–686, Mar. 2002.
- [42] Y. P. Zhao and J. G. Sun, "A fast method to approximately train hard support vector regression," *Neural Netw.*, vol. 23, no. 10, pp. 1276–1285, Dec. 2010.



Cheng-Hsuan Li received the B.S. and M.S. degrees from the National Chung Hsing University, Taichung, Taiwan, in 2001 and 2003, respectively. He is currently working toward the Ph.D. degree in the Institute of Electrical Control Engineering, National Chiao Tung University, Hsinchu, Taiwan.

He is currently an Executive Assistant with the Graduate Institute of Educational Measurement and Statistics, National Taichung University, Taichung. His research interests include pattern recognition and machine learning.



Bor-Chen Kuo (S'01–M'02) received the B.S. degree in mathematics education and the M.S. degree in statistics and assessment from the National Taichung Teachers College, Taichung, Taiwan, in 1993 and 1996, respectively, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, in 2001.

From 2002 to 2005, he was an Associate Professor with the National Taichung University of Education, Taichung, where he is currently a Professor and the Director of the Graduate Institute of Educational Measurement and Statistics. His research interests include pattern recognition, feature extraction and classification of high-dimensional data, computerized adaptive testing, and nonparametric functional estimation.

Dr. Kuo is a member of the Chinese Association of Psychological Testing AdCom and the Editor of the *Journal of Educational Measurement and Statistics*, Taiwan. He was the recipient of an Outstanding and Excellence Research Award from the R.O.C. Education and Research Society in 2009.



Chih-Sheng Huang received the B.S. degree from Fu Jen Catholic University, Taipei, Taiwan, in 2007 and the M.S. degree from the National Taichung University, Taichung, Taiwan, in 2010. He is currently working toward the Ph.D. degree in the Institute of Electrical Control Engineering, National Chiao Tung University, Hsinchu, Taiwan.

His current research interests include statistical forecast, pattern recognition, machine learning, and bio-signal processing.



Chin-Teng Lin (F'05) received the B.S. degree from the National Chiao Tung University (NCTU), Hsinchu, Taiwan, in 1986 and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, in 1992.

He is currently the Chair Professor of electrical and computer engineering with NCTU. His research interests include biologically inspired information systems. He has authored or coauthored more than 120 journal papers in the areas of neural networks, fuzzy systems, multimedia hardware/software, and cognitive neuroengineering, including approximately 74 IEEE journal papers.