Multiclass and Binary SVM Classification: Implications for Training and Classification Users

A. Mathur and G. M. Foody, *Member, IEEE*

***Abstract*—Support vector machines (SVMs) have considerable potential for supervised classification analyses, but their binary nature has been a constraint on their use in remote sensing. This typically requires a multiclass analysis be broken down into a series of binary classifications, following either the one-against-one or one-against-all strategies. However, the binary SVM can be extended for a one-shot multiclass classification needing a single optimization operation. Here, an approach for one-shot multiclass classification of multispectral data was evaluated against approaches based on binary SVM for a set of five-class classifications. The one-shot multiclass classification was more accurate (92.00%) than the approaches based on a series of binary classifications (89.22% and 91.33%). Additionally, the one-shot multiclass SVM had other advantages relative to the binary SVM-based approaches, notably the need to be optimized only once for the parameters C and γ as opposed to five times for one-against-all and ten times for the one-against-one approach, respectively, and used fewer support vectors, 215 as compared to 243 and 246 for the binary based approaches. Similar trends were also apparent in results of analyses of a data set of larger dimensionality. It was also apparent that the conventional one-against-all strategy could not be guaranteed to yield a complete confusion matrix that can greatly limit the assessment and later use of a classification derived by that method.**

***Index Terms*—Accuracy, binary and multiclass classification, confusion matrix, image classification, support vector machine (SVM).**

I. Introduction

THE VALUE of a supervised classification is typically a function of its accuracy. One of the key objectives in classification is often to achieve a high accuracy with, if possible, a small number of training samples to make the classification process as useful and economical as possible. One attractive classifier for this application is a support vector machine (SVM).

SVM classifications may be more accurate than the widely used alternatives such as classification by maximum likelihood, decision tree, and neural network-based approaches [1], [2]. An SVM aims to fit an optimal separating hyperplane (OSH) between classes by focusing on the training samples that lie at the edge of the class distributions, the support vectors. The OSH is oriented such that it is placed at the maximum distance between the sets of support vectors. It is because of this orientation that SVM is expected to generalize more accurately on unseen cases relative to classifiers that aim to minimize the training error such as neural networks. Thus, with SVM classification only some of the training samples that lie at the edge of the class distributions in feature space (support vectors) are needed in the establishment of the decision surface unlike statistical classifiers such as the widely used maximum-likelihood classifiers in which all training cases are used to characterize the classes. Conventional maximum-likelihood classification may, therefore, require much larger training sample size than SVM to derive an accurate classification. Moreover, it may sometimes be possible to identify the most useful training sites for the provision of support vectors before the classification [3], [4]. This potential for accurate classification based on small training sets means that the adoption of SVM classification can provide the analyst with considerable savings in training data acquisition [5].

SVMs were designed for binary classification. The binary SVM approach can, however, be extended for multiclass scenarios that are commonly encountered in remote sensing. This is generally achieved by decomposing the multiclass problem into a series of binary analyses which can be addressed with a binary SVM by following either the one-against-one or one-against-all strategies [1], [6]. Alternatively, a multiclass classification based on a single optimization may also be undertaken. A major attraction of the multiclass SVM is that the classification to all the classes occurs in a single step. This approach is very different to that adopted in multiclass classification based on binary SVM. With the latter, a series of analyses, the number of which is a positive function of the number of classes, must be undertaken to derive the classification [7]. In addition, by reducing the classification to a single optimization problem, the multiclass SVM approach may also require fewer support vectors than a multiclass classification based on binary SVMs [8] although this potential has been rarely explored.

This letter aims to evaluate the multiclass and binary-based SVM approaches for the derivation of a multiclass land cover classification from remotely sensed data. Attention is focused mainly on the nature of the confusion matrices derived, especially with regard to classification accuracy (the key index of classification quality) and the number of support vectors required (an indicator of training data cost). The fundamentals of SVM classification are discussed in Section II before presentation of the data sets and methods used in Section III. Section IV provides a summary of the results and a discussion before conclusions are presented in Section V.

Manuscript received February 26, 2007; revised August 29, 2007. This work was supported in part by a British Council Commonwealth Scholarship.

A. Mathur is with the Punjab Remote Sensing Centre, Punjab Agricultural University Campus, Ludhiana, 141 004, India (e-mail: mathurajay12@rediffmail.com).

G. M. Foody is with the School of Geography, University of Nottingham, NG7 2RD Nottingham, U.K. (e-mail: giles.foody@nottingham.ac.uk).

Digital Object Identifier 10.1109/LGRS.2008.915597

II. SVM CLASSIFICATION

An SVM seeks to find the OSH between classes by focusing on the training cases that lie at the edge of the class distributions, the support vectors, with the other training cases effectively discarded. The basis of the SVM approach to classification is, therefore, the notion that only the training samples that lie on the class boundaries are necessary for discrimination. Detailed discussion of SVM classification is available in the literature (e.g., [1], [9]–[11]) but here we draw on an earlier discussion [12] to highlight basic issues.

With a SVM, the training data set of r cases represented by $\{\boldsymbol{x}_i, y_i\}$, $i = 1, \ldots, r$, $y_i \in \{1, -1\}$ in a q dimensional space are used to form a classifier that can generalize accurately. The OSH runs between the classes in a way that maximizes the margin between them. A hyperplane is defined as $\boldsymbol{w} \cdot \boldsymbol{x} + \boldsymbol{b} = 0$, where x is a point lying on the hyperplane, $\boldsymbol{w}$ is normal to the hyperplane, $\boldsymbol{b}$ is the bias. For the linearly separable case, a separating hyperplane can be defined for the two classes as: $\boldsymbol{w} \cdot \boldsymbol{x_i} + \boldsymbol{b} \geq +1$ (for $y_i = +1$) and $\boldsymbol{w} \cdot \boldsymbol{x}_i + \boldsymbol{b} \leq -1$ (for $y_i = -1$). The two equations can be combined as

$$\boldsymbol{y}_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + \boldsymbol{b}) - 1 \geq 0. \tag{1}$$

The training data points on these two hyperplanes, that are parallel to the OSH and defined by $\boldsymbol{w} \cdot \boldsymbol{x}_i + \boldsymbol{b} = \pm 1$, are the support vectors. The margin between the planes is $2/|\boldsymbol{w}|$ and its maximization leads to the constrained optimization problem

$$\min \left\{ \frac{1}{2} \|\boldsymbol{w}\|^2 \right\} \tag{2}$$

under the inequality constraints of (1).

For nonlinearly separable classes, slack variables, $\{\xi_i\}_{i=1}^{r}$ may be introduced with the constraint becoming

$$\boldsymbol{y}(\boldsymbol{w} \cdot \boldsymbol{x}_i + \boldsymbol{b}) > 1 - \xi_i. \tag{3}$$

A penalty term, $C \sum_{i=1}^{r} \xi_i$, is generally added to penalize solutions for which ξ_i are very large. The optimization from (1) then becomes

$$\min \left[\frac{\|\boldsymbol{w}\|^2}{2} + C \sum_{i=1}^{r} \xi_i \right] \tag{4}$$

under the constraints of (3).

To allow for nonlinear decision surfaces, the data may be mapped into a high-dimensional space through some nonlinear mapping which has the effect of spreading the distribution of the data in a way that facilitates the fitting of a linear hyperplane. Specifically, the training data may be projected into a high-dimensional space $\boldsymbol{H}$ through a mapping function $\boldsymbol{\varphi}$, or $\boldsymbol{\varphi} : \boldsymbol{R}^q \rightarrow \boldsymbol{H}$. An input data point $\boldsymbol{x}$ can be represented as $\boldsymbol{\varphi}(\boldsymbol{x})$ in the high-dimensional space $\boldsymbol{H}$. The expensive computation of $(\boldsymbol{\varphi}(\boldsymbol{x}) \cdot \boldsymbol{\varphi}(\boldsymbol{x}_i))$ in a high-dimensional space is reduced by using a positive definite kernel such that

$$(\varphi(x) \cdot \varphi(x_i)) = k(x, x_i) \tag{5}$$

leading to decision functions of the form

$$\boldsymbol{f}(\boldsymbol{x}) = \mathrm{sgn} \left(\sum_{i=1}^{r} \alpha_i \boldsymbol{y}_i \boldsymbol{k}(\boldsymbol{x}, \boldsymbol{x}_i) + \boldsymbol{b} \right) \tag{6}$$

where α_i is a Lagrange multiplier. The kernel must meet Mercer's condition and one such kernel is the radial basis function

$$k(\boldsymbol{x}, \boldsymbol{x}_i) = \boldsymbol{e}^{-\gamma \|x - x_i\|^2} \tag{7}$$

where γ controls the width of the kernel function.

SVMs were designed for binary classification but can be extended for multiple class classification scenarios that are common in remote sensing. Two main approaches have been suggested for multiclass classification by SVM. In each, the underlying basis has been to reduce the multiclass problem to a set of binary problems, enabling a basic SVM approach to be used. The two approaches are, however, very different in detail. In the "one-against-all" approach, a set of binary classifiers, each trained to separate one class from the rest, is undertaken and each pixel is allocated to the class for which the largest decision value was determined. Specifically, with this approach after solving (4), for a case x_i there are n decision functions, where n is the number of classes

$$(\boldsymbol{w}_i)^{\mathrm{T}} \boldsymbol{\varphi}(\boldsymbol{x}) + b_j, \qquad \text{where } j = 1, \ldots, n. \tag{8}$$

For each of the n analyses, this allocation indicates that the case is either a member of the class under consideration or not. With the latter outcome, the approach does not indicate a specific class, only that the case is not a member of the class under consideration and so may be considered as a member of a class "other." After all n analyses the case is allocated to the class for which the above decision function has the largest value. This final allocation may sometimes be to "other."

The second method to reduce a multiclass problem to a set of binary ones to enable the use of the basic SVM model for multiclass classification is the "one-against-one" approach. In this, a series of classifiers are applied to each pair of classes, with the most commonly computed class label kept for each pixel. This method requires the training of $n(n-1)/2$ classifiers or machines using data from every pair of classes. A strategy to handle instances in which an equal number of votes are derived for more than one class for a pixel is also required. Finally, the training data vector $\boldsymbol{x}_i$ is predicted to belong to the class with maximum number of votes.

Multiclass classifications of remotely sensed data by SVM have to date been based mostly on the above approaches. While both strategies to reducing the multiclass problem to a set of binary classifications enable the SVM to be employed in a more appropriate approach that is less computationally demanding, is to consider all classes at one time, yielding a multiclass SVM. One means to achieve this, which is similar in basis to the "one-against-all" approach, is by solving a single optimization problem [8]. With this, n two class rules where the mth function $\boldsymbol{w}_m^{\mathrm{T}} \boldsymbol{\varphi}(\boldsymbol{x}) + b$ separates the training data vectors of class m from that of others are constructed. Hence, there are n decision functions or hyperplanes, but all are obtained by solving one problem

$$\min_{w,b,\xi} \frac{1}{2}\sum_{m=1}^{n} \boldsymbol{w}_m^{\mathrm{T}}\boldsymbol{w}_m + C\sum_{i=1}^{l}\sum_{m\neq y_i} \xi_i^m \boldsymbol{w}_{y_i}^{\mathrm{T}}\boldsymbol{\varphi}(\boldsymbol{x}_i) + b_{y_i} \geq \boldsymbol{w}_m^{\mathrm{T}}\boldsymbol{\varphi}(\boldsymbol{x}_i) + b_m + 2 - \xi_i^m \quad (9)$$

with $\xi_i^m \geq 0,\ i = 1,\ldots,l, m \in \{1,\ldots,n\}/y_i$. The resulting decision function is

$$\arg\max_{m=1,\ldots,n}\left(\boldsymbol{w}_m^{\mathrm{T}}\boldsymbol{\varphi}(\boldsymbol{x}_i) + b_m\right). \quad (10)$$

In reducing the classification to a single optimization problem, the approach may require fewer support vectors than a multiclass classification based on the combined use of many binary SVMs. Here, this one-shot multiclass approach is evaluated relative to the one-against-one and one-against-all binary based approaches to multiclass classification.

III. Data and Methods

Two data sets were used. Most attention is focused on analysis of a low-dimensional (3) multispectral data set with a further set of analyses undertaken on a widely available data set of larger dimensionality (12) for comparison and to increase the generalizability of this letter.

The low-dimensional data set was acquired by the LISS-III sensor carried on the Indian Remote Sensing Satellite (IRS-1D). These data had spatial resolution of ~24 m and were acquired on September 22 for a study area that comprised the south-western part of Punjab state, India. These data were acquired in red (0.62–0.68 μm), near-infrared (0.77–0.86 μm) and shortwave-infrared (1.55–1.75 μm) wavebands and used to classify the main land cover classes that occurred in the study area: cotton, basmati rice, a local variety of rice, built-up land and sand. Ground data on class membership were collected in the field immediately prior to image acquisition during the period September 15–21, 2003.

Training and testing data were acquired following a stratified, by class, random sampling design. This approach ensured that even relatively rare classes in the study area, such as sand and basmati rice, were adequately sampled. A total of 180 pixels of each class were acquired and divided by random sampling into equally sized training and testing sets. The training set, therefore, comprised 90 pixels of each class and a total of 450 pixels. This size of training set is compatible with the widely used heuristic that suggests use of ~30 times the number of discriminatory variables be used for a supervised image classification.

A Gaussian kernel was used for both the multiclass and binary SVM analyses. With the one-shot multiclass classification, the parameters C and γ were set at 0.25 and 0.005, respectively, following a five-fold cross-validation analysis. Note that for this one-shot multiclass SVM, these parameters need be optimized only once while with the one-against-one and one-against-all binary strategies a series of analyses is needed and so parameter selections are required. With the one-against-all approach, the parameters C and γ need to be optimized a number of times equal to the number of classes. Here, the one-against-all binary approach required five SVMs and the C and γ parameters for each were optimized individually using five-fold cross-validation. With the one-against-one strategy, ten optimizations were required. Again a five-fold cross-validation approach was used. With each binary SVM approach, a number of analyses (equal to the number of optimizations) are then also required to derive the final classification.

TABLE I
Confusion Matrix for the One-Shot Multiclass SVM. Overall Accuracy = 92.00%

Actual class →	Predicted class ↓					
	B	S	C	RB	RL	Total
Built-up (B)	89	1	0	0	0	90
Sand (S)	15	75	0	0	0	90
Cotton (C)	0	0	88	0	2	90
Rice Basmati (RB)	0	0	0	82	8	90
Rice Local (RL)	0	0	3	7	80	90
Total	104	76	91	89	92	450

Additional analyses were undertaken using a 12-waveband data set. This data set was the moderate-dimension example data set available with the Multispec system and described in associated documentation [13]. Here, attention focused on the classification of five crops: corn, oats, red clover, soya bean, and wheat. Using data in all 12 wavebands, a training set comprising 360 (30 times the number of wavebands) randomly selected pixels was generated for each class. An independent testing set comprising of 360 randomly selected pixels of each class was also derived. As with the analyses of the other data set, a series of SVM classifications were undertaken with five-fold cross validation used in the determination of parameter settings.

To facilitate comparison of the different approaches to classification, a confusion matrix was derived from each analysis. These matrices summarize the class allocations made in each analysis from which accuracy may be assessed and may also provide valuable information for later users of the classification. Here, classification accuracy was expressed in terms of the percentage of cases correctly allocated. To determine if the classifications derived from the three classification approaches differed significantly, a McNemar test was used in recognition of the use of a common testing set [14]. With this, the evaluation of the significance of the difference between two classifications is based on a computed z score, with a statistically significant difference at the 95% level of confidence if $z > |1.96|$.

IV. Results and Discussion

Focusing on the low-dimensional data set, all three approaches classified the testing data with a high accuracy, >89% (Tables I–III). However, it was apparent that the one-shot multiclass SVM classification yielded the most accurate classification, with an accuracy of 92.00% (Table I). The difference in the accuracy of the multiclass and one-against-all binary-based classification was also statistically significant, $z = 3.12$. Although the classification by the multiclass SVM was more

TABLE II
CONFUSION MATRIX FOR THE ONE-AGAINST-ALL BINARY SVM STRATEGY. OVERALL ACCURACY = 89.11%

Actual class →	Predicted class ↓						
	B	S	C	RB	RL	Others	Total
Built-up (B)	84	3	0	0	0	3	90
Sand (S)	9	69	0	0	0	12	90
Cotton (C)	0	0	88	0	2	0	90
Rice Basmati (RB)	0	0	0	79	8	3	90
Rice Local (RL)	0	0	2	7	81	0	90
Total	93	72	90	86	91	18	450

TABLE III
CONFUSION MATRIX FOR THE ONE-AGAINST-ONE BINARY SVM STRATEGY. OVERALL ACCURACY = 91.33%

Actual class →	Predicted class ↓					
	B	S	C	RB	RL	Total
Built-up (B)	85	5	0	0	0	90
Sand (S)	12	78	0	0	0	90
Cotton (C)	0	0	88	0	2	90
Rice Basmati (RB)	0	0	0	81	9	90
Rice Local (RL)	0	0	4	7	79	90
Total	97	83	92	88	90	450

TABLE IV
CONFUSION MATRIX FOR THE ONE-AGAINST-ALL STRATEGY USING THE PARAMETER VALUES SELECTED FOR THE ONE-SHOT MULTICLASS SVM. OVERALL ACCURACY = 88.22%

Actual class →	Predicted class ↓						
	B	S	C	RB	RL	Others	Total
Built-up (B)	84	3	0	0	0	3	90
Sand (S)	17	68	0	0	0	5	90
Cotton (C)	0	0	88	0	2	0	90
Rice Basmati (RB)	0	0	0	81	6	3	90
Rice Local (RL)	0	0	3	7	76	4	90
Total	101	71	91	88	84	15	450

accurate than that from the one-against-one binary SVM, the difference in accuracy was not significant, $z = 0.77$.

One problem encountered with the one-against-all approach was an inability to label all cases appropriately. For 18 cases ($\sim$4% of the test set), none of the five component binary SVMs indicated membership to any one of the classes. These cases were labeled "other" and cannot be located within the confusion matrix. For presentational purposes, these cases are shown in an additional column added to the matrix (Table II).

A key attraction of the one-shot multiclass SVM is the requirement to optimize the parameter values only once. The one-against-all and one-against-one strategies required five and ten optimizations, respectively. It would, of course, be possible to use a single set of values but these would yield a suboptimal classification. For example, using the single set of parameter values defined for the one-shot multiclass SVM in all five analyses of the one-against-all approach yielded a classification with an accuracy of 88.22% (Table IV). Because the parameter values determined by five-fold cross-validation for each component of the one-against-all approach (Table V) sometimes differed greatly from the set used in the one-shot multiclass SVM, this result highlights the importance of selecting appropriate parameter values and potential value of an approach requiring only a single optimization analysis.

TABLE V
PARAMETER VALUES DETERMINED FROM FIVE-FOLD CROSS-VALIDATION FOR THE ONE-AGAINST-ALL STRATEGY

Combination	Parameters	
	C	γ
Built-up *v* rest	.5	.01
Sand *v* rest	2	.02
Cotton *v* rest	.5	.00125
Rice Basmati *v* rest	.2	.0025
Rice Local *v* rest	2	.0025

TABLE VI
NUMBER OF SUPPORT VECTORS DEFINED FROM THE TRAINING SET OF 450 CASES

	Number of support vectors		
Class	**One-shot multi-class**	**One-against-all**	**One-against-one**
Built-up (B)	51	54	56
Sand (S)	52	64	68
Cotton (C)	29	40	36
Rice Basmati (RB)	39	37	38
Rice Local (RL)	46	48	48
Total	**215**	**243**	**246**

As expected, many of the training cases used in the classifications were unnecessary as they were not support vectors (Table VI). The smallest number of support vectors used was 215 for the classification with the one-shot multiclass SVM classifier. The one-against-one and one-against-all strategies for multiclass classification required a larger number of support vectors, 246 and 243, respectively. Thus, not only was the classification derived from the one-shot multiclass SVM the most accurate but it could be derived with the least training data. Given that it may sometimes be possible to identify sites likely to furnish support vectors in advance of the analysis [3], [4] the results indicate that the one-shot multiclass SVM may be able to derive the most accurate classification with the least expensive training set.

Similar trends were observed with the analyses of the 12-waveband data set. Here, the highest accuracy, 98.1%, was achieved by the one-shot approach. The one-against-one strategy also yielded a classification with an accuracy of 98.1% but the one-against-all approach yielded a significantly different classification ($z = 2.66$) with a lower accuracy of 97.6% (Table VII). More importantly, the one-shot approach required fewest support vectors (Table VIII) and the one-against-all approach did not always produce a class allocation, leaving some cases ($\sim$1%) to be labeled as "other" [Table VII(b)].

V. CONCLUSION

The results indicate the potential of one-shot multiclass SVM for the classification of remotely sensed data. The accuracy of the one-shot multiclass SVM classification was not only the highest derived but also required the smallest number of support vectors, providing a potential for accurate classification from a small training set. The binary-based approaches to classification also suffered a disadvantage of requiring a series of parameter optimizations and classifications. The number of optimization and classification analyses required was a function of the number of classes and greatest for the one-against-one strategy. It was also apparent that the widely used one-against-all strategy could not be guaranteed to yield a complete confusion matrix.

TABLE VII
CONFUSION MATRICES FROM (A) ONE-SHOT SVM, ACCURACY = 98.1%, (B) ONE-AGAINST-ALL, ACCURACY = 97.6%, AND (C) ONE-AGAINST-ONE, ACCURACY = 98.1%

Actual class →	Predicted class ↓						
	C	O	RC	S	W	Others	Total
Corn (C)	354	2	0	4	0	n/a	360
Oats (O)	2	344	14	0	0	n/a	360
Red Clover (RC)	1	7	352	0	0	n/a	360
Soyabean (S)	2	1	0	357	0	n/a	360
Wheat (W)	0	0	0	0	360	n/a	360
Total	359	354	366	361	360	n/a	1800

(a)

Actual class →	Predicted class ↓						
	C	O	RC	S	W	Others	Total
Corn (C)	346	1	0	2	0	11	360
Oats (O)	0	344	11	0	0	5	360
Red Clover (RC)	0	8	350	0	0	2	360
Soyabean (S)	2	0	0	357	0	1	360
Wheat (W)	0	0	0	0	360	0	360
Total	348	353	361	359	360	19	1800

(b)

Actual class →	Predicted class ↓						
	C	O	RC	S	W	Others	Total
Corn (C)	352	4	0	4	0	n/a	360
Oats (O)	0	343	17	0	0	n/a	360
Red Clover (RC)	1	5	354	0	0	n/a	360
Soyabean (S)	2	0	0	358	0	n/a	360
Wheat (W)	0	0	0	0	360	n/a	360
Total	355	352	371	362	360	n/a	1800

(c)

TABLE VIII
NUMBER OF SUPPORT VECTORS DEFINED FROM THE TRAINING SET OF 1800 CASES

	Number of support vectors		
Class	**One-shot multi-class**	**One-against-all**	**One-against-one**
Corn (C)	85	165	106
Oats (O)	109	148	128
Red Clover (RC)	126	181	144
Soyabean (S)	76	110	78
Wheat (W)	100	104	47
Total	**496**	**709**	**503**

This indicates a problem with the approach which may require attention as it can limit classification evaluation and later uses of the classification. Given the difficulty of anticipating later user needs and the ability to derive higher classification accuracy with a single optimization and classification analysis it is suggested that the one-shot multiclass approach to classification rather than a binary-based strategy be used for multiclass classifications from remotely sensed data.

ACKNOWLEDGMENT

The authors would like to thank Dr. P. K. Sharma, Director Punjab Remote Sensing Centre (PRSC), P. Litoria, Dr. D. Loshali, and S. Singh for assistance as well as the British Council's award to A. Mathur while on leave from PRSC. The SVM were constructed with BSVM (version 2.01) developed by C.-W. Hsu and C.-J. Lin, National Taiwan University, Taipei, Taiwan, R.O.C. This letter was initiated while the authors were at Southampton University. Finally, the authors would also like to thank the three referees for their helpful comments.

REFERENCES

[1] C. Huang, L. S. Davis, and J. R. G. Townshend, "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens.*, vol. 23, no. 4, pp. 725–749, Feb. 2002.
[2] Y. Bazi and F. Melgani, "Toward an optimal SVM classification system for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3374–3385, Nov. 2006.
[3] G. M. Foody and A. Mathur, "Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification," *Remote Sens. Environ.*, vol. 93, no. 1/2, pp. 107–117, Oct. 2004.
[4] G. M. Foody, A. Mathur, C. Sanchez-Hernandez, and D. S. Boyd, "Training set size requirements for the classification of a specific class," *Remote Sens. Environ.*, vol. 104, no. 1, pp. 1–14, Sep. 2006.
[5] A. Mathur and G. M. Foody, "Crop classification by a SVM with intelligently selected training data for an operational application," *Int. J. Remote Sens.,* doi 10.1080/01431160701395203.
[6] G. H. Halldorsson, J. A. Benediktsson, and J. R. Sveinsson, "Support vector machines in multisource classification," in *Proc. IGARSS*, Toulouse, France, Jul. 21–25, 2003, pp. 2054–2056.
[7] J. A. Gualtieri and R. F. Cromp, "Support vector machines for hyperspectral remote sensing classification," *Proc. SPIE*, vol. 3584, pp. 221–232, 1998.
[8] 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.
[9] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
[10] 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.
[11] K.-R. Muller, S. Mika, G. Ratch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–201, Mar. 2001.
[12] G. M. Foody and A. Mathur, "A relative evaluation of multiclass image classification by support vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1335–1343, Jun. 2004.
[13] *Documentation for Multispec*, Dec. 2005. [Online]. Available: http://cobweb.ecn.purdue.edu/~biehl/MultiSpec/documentation.html
[14] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, 2004.