A Support Vector Domain Description Approach to Supervised Classification of Remote Sensing Images

Jordi Muñoz-Marí, Lorenzo Bruzzone, *Senior Member, IEEE*, and Gustavo Camps-Valls, *Senior Member, IEEE*

***Abstract*—This paper addresses the problem of supervised classification of remote sensing images in the presence of incomplete (nonexhaustive) training sets. The problem is analyzed according to two different perspectives: 1) description and recognition of a specific land-cover class by using single-class classifiers and 2) solution of multiclass problems with single-class classification techniques. In this framework, we analyze different one-class classifiers and introduce in the remote sensing community the support vector domain description method (SVDD). The SVDD is a kernel-based method that exhibits intrinsic regularization ability and robustness versus low numbers of high-dimensional samples. The SVDD technique is compared with other standard single-class methods both in problems focused on the recognition of a single specific land-cover class and in multiclass problems. For the latter, we properly define an easily scalable multiclass architecture capable to deal with incomplete training data. Experimental results, obtained on different kinds of data (synthetic, hyperspectral, and multisensor images), point out the effectiveness of the SVDD technique and provide important indications for driving the choice of the classification technique and architecture in the presence of incomplete training data.**

***Index Terms*—Image classification, incomplete training data, kernel methods, one-class domain description, remote sensing, support vector domain description (SVDD).**

I. Introduction

CLASSIFICATION of remote sensing images is a complex task whose accuracy strongly depends on the available prior information. Given the general high complexity of the problem, usually supervised classifiers are preferred to unsupervised clustering algorithms, which are intrinsically less suitable to obtain accurate classification maps. However, supervised classifiers require the availability of a complete and representative training set for a proper learning of the classification algorithms. In particular, the training set should satisfy the following constraints.

1) It should be exhaustive (i.e., it should contain samples describing all the land-cover classes present in the investigated area).
2) It should have a sufficient number of independent samples for each land-cover class for a reliable estimation of the classifier parameters (this is related to the small-size training problem that may result in the Hughes phenomenon [1]).
3) It should be made up of samples that completely describe the intraclass variability due to the nonstationary behavior of the spectral signature in the spatial domain (this may result in unrepresentative training sets).

However, these constraints are difficult to satisfy in real applications. This is particularly critical in the light of the last generation of hyperspectral remote sensing sensors, which allows a dense sampling of the spectral signature of land-cover classes, and results in the following: 1) high-dimensional feature spaces; 2) high sensitivity to the intraclass variation of the spectral signature of a given land-cover class; and 3) high capability in discriminating similar land-cover classes (involving a higher number of classes to be recognized). These properties make more critical the definition of reliable training sets, i.e., those capable of a complete representation of land-cover classes and made up of a sufficient number of representative training samples. Similar considerations can be done when the integration of multisensor data is considered in the solution of the classification problem. In the recent literature, some papers addressed the learning problem in the presence of small-size or unrepresentative training sets [2]–[4]. Less attention has been devoted to the more critical (and in some applications very important) case of the incomplete knowledge of the classes present in the investigated scene. In this case, effective strategies should be considered for a proper design of a classification architecture capable to produce reliable classification maps by exploiting the available incomplete (nonexhaustive) training information. The aforementioned problem can be also analyzed from a different perspective. In some applications, in spite of the richness of the information present in the available remote sensing data, the goal is to recognize one specific land-cover class of interest and to discriminate it from the other classes present in the investigated area. This means that we are interested in using the available training information for characterizing and describing the behavior of

Manuscript received January 12, 2007; revised March 8, 2007. This work was supported by the Spanish Ministry for Education and Science and the Italian Ministry for University and Research under the "Integrated Actions Program" MEC/HI2005-0228.

J. Muñoz-Marí and G. Camps-Valls are with the Grup de Processament Digital de Senyals, Department Enginyeria Electrònica, Escola Tècnica Superior d'Enginyeria, Universitat de València, 46100 Burjassot, València, Spain (e-mail: jordi@uv.es; gcamps@uv.es).

L. Bruzzone is with the Department of Information and Communication Technology, University of Trento, 38050 Trento, Italy (e-mail: lorenzo.bruzzone@ing.unitn.it).

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.2007.897425

one class in order to make it possible to recognize this class independently on the other classes present in the considered scene. This formulation of the problem relaxes the constraint of having an exhaustive training set but requires the availability of representative training data for the analyzed class and, if possible, some training samples representative of other classes.[1] This problem can be solved according to single hypothesis testing methods, which are intrinsically devoted to recognize the samples of one specific class from a heterogeneous distribution.

On the basis of the previous discussion, it is important to define effective classification systems capable to properly model and recognize a single land-cover class and/or to address multiclass problems in the presence of a nonexhaustive representation of classes in the training set. Of course, in this last case, the system should be able to recognize the classes modeled in the training set, by properly rejecting samples belonging to other classes.

In this paper, we address this critical issue by considering classification systems made up of one-class classifiers, which are specialized in identifying a given class and rejecting the others present in the scene. The main contributions of this work consists of the following: 1) presenting an exhaustive analysis on the effectiveness of the different one-class classifiers proposed in the literature in the presence of either hyperspectral or multisensor data sets for one-class modeling and recognition; 2) assessing the effectiveness of architectures made up of one-class classifiers (one per each known class) in solving multiclass problems when only a subset of classes is known; and 3) introducing in the remote sensing community the use of the support vector domain description (SVDD) one-class classifier [5] in multiclass detection problems with incomplete training data.

Concerning the first contribution, an experimental analysis related to the effectiveness of different one-class classifiers is presented. In particular, several single hypothesis testing methods (including Gaussian, mixture of Gaussians, and k-nearest neighbor (k-NN) domain description classifiers [6]) are compared with the SVDD technique through both synthetic datasets (assuming different data distributions and complexity) and real remote sensing datasets with different properties (i.e., multispectral, hyperspectral, and radar data). The analysis of the different scenarios allows us to derive interesting conclusions on the effectiveness of the different classification techniques with different kinds of data. These conclusions provide indications that can drive a proper choice of the one-class classification technique to be used in different real applications.

With regard to the second contribution, in order to properly consider the nonexhaustive class representation in the phase of the classification system design, we address the multiclass problem by defining and analyzing the effectiveness of a classification architecture made up of one-class classifiers, one per each known class. Each one-class classifier has the ability to recognize samples from a given class of interest (targets) but also the capability to reject samples from other classes (outliers). Combining one-class classifiers, we can get a scalable multiclass classifier that assigns correctly the samples of known classes and, at the same time, rejects samples belonging to unknown classes. Thus, we can easily adapt our solution according to the available ground truth information, since adding up one more class-dedicated classifier does not force us to retrain the whole multiclass scheme.

Concerning the third contribution, although one-class detection has been recently used in remote sensing images [7], we introduce here a general framework to use the SVDD to solve multiclass detection problems with incomplete training data. The SVDD is a recent kernel-based development which has demonstrated very effective (in terms of accuracy and robustness to high-dimensional problems) in different application domains, when no *a priori* knowledge on the data distribution is available [8], [9]. Unlike the standard support vector machine (SVM), the SVDD only considers samples belonging to the class of interest in order to learn the underlying data distribution. Hence, it becomes an efficient methodology for characterizing a given class and offers good rejection capabilities. The SVDD is, like SVM, a kernel method, thus inheriting all the related advantages [5], [8], [9].

The rest of this paper is organized as follows. Section II describes standard one-class methods evaluated in this paper, pointing out their properties, advantages, and disadvantages. Section III presents the SVDD one-class classifier. Section IV addresses the architecture adopted for solving multiclass problems. Section V reports the experiments considered in this paper, analyzes the properties of each problem, explains the model development, and shows the experiments results. Section VI discusses the obtained results from a theoretical viewpoint and also provides recommendations for further research.

II. One-Class Classification Methods

In this section, we briefly review the one-class formulations used in this paper for comparisons with the SVDD technique: Gaussian, mixture of Gaussians, and k-NN domain description.

A. Gaussian Domain Description Classifier

In the one-class Gaussian domain description method, the target class is modeled as a Gaussian distribution. The function is expressed as

$$f(\mathbf{x}) = (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \tag{1}$$

where μ is the class mean, and Σ represents the covariance matrix estimated using the training samples. According to this function, the classifier is defined as

$$h(\mathbf{x}) = \begin{cases} \text{target} & \text{if } f(\mathbf{x}) \leq \theta \\ \text{outlier} & \text{if } f(\mathbf{x}) > \theta. \end{cases} \tag{2}$$

Note that a threshold θ has to be adjusted heuristically, usually by evaluating the accuracy in an out-of-sample dataset.

[1] It is worth noting that, from a theoretical viewpoint, also in this case the availability of a complete training set would result in potentially more accurate classification results.

B. Mixture of Gaussian Domain Description Classifier

In order to improve the model of the Gaussian domain description method, here, the target class is modeled using a mixture of K Gaussians. The model is defined as

$$f(\mathbf{x}) = -\sum_{i=1}^{K} P_i \exp\left(-(\mathbf{x}-\mu_i)^T \Sigma_i^{-1}(\mathbf{x}-\mu_i)\right). \quad (3)$$

Parameters P_i, μ_i, and Σ_i represent the contribution, mean, and sample covariance for each Gaussian in the mixture, respectively. These parameters are optimized using the expectation maximization (EM) algorithm. The main drawback of this method is that when working with high-dimensional data and a large number of clusters, the number of free parameters to be tuned can be huge.

This method can also be formulated in order to use outlier objects in the training process. In this case, different mixtures of Gaussians are defined both for target and outlier samples (K_{t} and K_{o}, respectively). To ensure that the decision boundary will be closed around the target class, an extra outlier cluster with a wide covariance matrix should be introduced, giving the model

$$\begin{aligned} f(\mathbf{x}) = &-\sum_{i=1}^{\mathrm{Kt}} P_i \exp\left(-(\mathbf{x}-\mu_i)^T \Sigma_i^{-1}(\mathbf{x}-\mu_i)\right) \\ &+ P_* \exp\left(-(\mathbf{x}-\mu)^T \Sigma_*^{-1}(\mathbf{x}-\mu)\right) \\ &+ \sum_{j=1}^{\mathrm{Ko}} P_j \exp\left(-(\mathbf{x}-\mu_j)^T \Sigma_i^{-1}(\mathbf{x}-\mu_j)\right). \quad (4) \end{aligned}$$

Here, μ and Σ are the sample mean and covariance matrix of the complete dataset, respectively, and Σ_* represents a regularization covariance which is usually defined as 10Σ. The parameter P_* is optimized with the EM algorithm in such a way that $P_* + \sum_j P_j = 1$.

Note that the mixture of Gaussian domain description model constitutes a generalization of the Gaussian domain description, and that each data class distribution is assumed to be drawn from a Gaussian distribution, thus yielding a final mixture of Gaussian model.

The classification function for the mixture of Gaussian method is defined in (2), and again, a threshold θ is required and has to be adjusted in the classifier training process.

C. k-NN Domain Description Classifier

In this model, labels are assigned to test samples by computing the kth normalized distance to their NN. This method is the natural extension of the NN algorithm, which is defined as a given set of P points in an N-dimensional space $\mathbb{R}^N$, which constructs a data structure given any query point $\mathbf{q}$ that finds the point in P with the smallest distance to $\mathbf{q}$ [10]. Usually, the distance is induced by an L_s-norm, so the distance between two points $\mathbf{p}$ and $\mathbf{q}$ is defined as $f(\mathbf{p},\mathbf{q}) = (\sum_{i=1}^{N} |p_i - q_i|^s)^{1/s}$, $s \in [0, \infty)$. However, other notions of distance can be used as well.

In the k-NN algorithm, instead of searching for the nearest point, a set of k nearest points is retrieved using the defined distance, and then the point is assigned to the most common class in these k points.

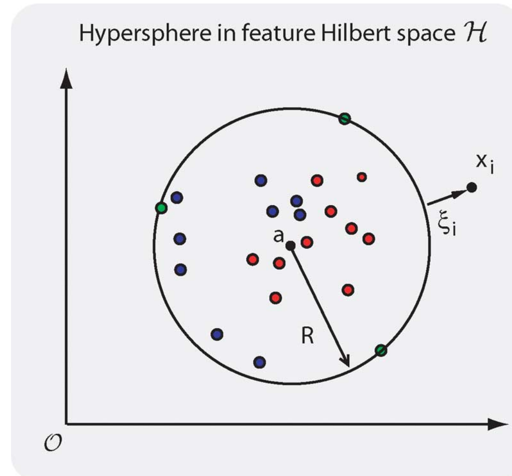


Fig. 1. Hypersphere containing the (colored) target data is described by the center $\mathbf{a}$ and radius R, in which the samples in the boundary are the support vectors (green), and samples outside the ball are assigned a positive constrained to deal with outliers.

The k-NN method works well when having a large training set to train the classifier and can work equally well on different probabilistic distributions, but its accuracy decreases severely if not enough training data are available or when dealing with high-dimensional feature spaces [11], [12].

III. SVDD Classifier

In remote sensing data processing, the use of the SVM binary classifier resulted in very high accuracies both in binary and multiclass hyperspectral image classification [2], [13]–[15]. However, the formulation of binary SVM cannot be used directly for one-class problems. In this context, Tax and Duin [5] originally proposed a support vector (SV) method for one-class classification.

A. Notation

Let us consider a dataset $\{\mathbf{x}_i \in \mathbb{R}^N, i = 1, \ldots, n\}$ belonging to a given class of interest. The goal here is to find a hypersphere (in a high-dimensional Hilbert feature space $\mathcal{H}$ where the samples have been mapped through a nonlinear transformation $\boldsymbol{\phi}$) of radius $R > 0$ and center $\mathbf{a}$ with a minimum volume containing most of these data objects [5] (see Fig. 1).

Therefore, one has to minimize R^2 constrained to $\|\boldsymbol{\phi}(\mathbf{x}_i) - \mathbf{a}\|^2 \leq R^2$, $\forall i = 1, \ldots, n$. In addition, since the (training) distribution may contain outliers, one introduces a set of slack variables $\xi_i \geq 0$, as usual in the SVM framework, and the problem becomes

$$\min_{R,\mathbf{a},\xi} \left\{ R^2 + C \sum_i \xi_i \right\} \quad (5)$$

which is constrained to

$$\|\boldsymbol{\phi}(\mathbf{x}_i) - \mathbf{a}\|^2 \leq R^2 + \xi_i, \qquad \xi_i \geq 0 \qquad \forall i = 1, \ldots, n \quad (6)$$

where the parameter C controls the tradeoff between the volume of the hypersphere and the permitted errors (regularization parameter).

The primal function (5) is usually solved through its Lagrangian dual problem [8], which consists of solving

$$\max_{\alpha_i}\left\{\sum_i \alpha_i \left\langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_i)\right\rangle - \sum_{i,j} \alpha_i\alpha_j \left\langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j)\right\rangle\right\} \quad (7)$$

subject to $0 \le \alpha_i \le C$. This constitutes a quadratic programming problem that yields a set of α_i fulfilling the following.

1) If $\|\boldsymbol{\phi}(\mathbf{x}_i) - \mathbf{a}\|^2 < R^2$ then $\alpha_i = 0$.
2) If $\|\boldsymbol{\phi}(\mathbf{x}_i) - \mathbf{a}\|^2 = R^2$ then $0 < \alpha_i < C$.
3) If $\|\boldsymbol{\phi}(\mathbf{x}_i) - \mathbf{a}\|^2 > R^2$ then $\alpha_i = C$.

As a result, SVs are those satisfying $\alpha_i > 0$, while samples associated with $\alpha_i = C$ are considered outliers.

It is worth noting that all $\boldsymbol{\phi}$ mappings (which are in principle unknown) used in the SV method occur in the form of inner products. This allows us to define a kernel function K

$$K(\mathbf{x}_i, \mathbf{x}_j) = \left\langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j)\right\rangle \quad (8)$$

and the nonlinear SVDD can be constructed using only the kernel function, without needing to know the mapping $\boldsymbol{\phi}$ explicitly.

After solving the dual problem, the decision function implemented by the classifier for any test vector $\mathbf{x}$ is given by

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{n} \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b\right) \quad (9)$$

where b can be easily computed from the α_i that are neither 0 or C.

B. Equivalent Formulation

An equivalent formulation was proposed by Schölkopf in [16], in which one places a hyperplane that separates the data from the origin with maximum margin. In this case, the problem reduces to

$$\min_{\mathbf{w},\rho,\boldsymbol{\xi}}\left\{\frac{1}{2}\|\mathbf{w}\|^2 - \rho + \frac{1}{\nu n}\sum_i \xi_i\right\} \qquad \forall i = 1, \ldots, n \quad (10)$$

constrained to $\xi_i \ge 0$, and where $\nu \in (0, 1]$ is a regularization parameter (resembling C) controlling the tradeoff between accepting data into the class and having $\|\mathbf{w}\|$ small. Introducing again Lagrange multipliers, we reach the (equivalent) dual problem which, by using (8), becomes:

$$\min_{\boldsymbol{\alpha}}\left\{\frac{1}{2}\sum_{i,j} \alpha_i\alpha_j K(\mathbf{x}_i, \mathbf{x}_j)\right\} \quad (11)$$

which is subject to $0 \le \alpha_i \le 1/\nu n$ and $\sum_i \alpha_i = 1$, which once again is a quadratic programming problem.

C. Remarks

The SVDD, like binary SVMs, is a kernel-based method. In the framework of kernel methods, the key point is the use of a valid kernel matrix. The Mercer's theorem states the condition for a similarity matrix to be a valid kernel and constitutes the key requirement to obtain a unique global solution when developing kernel-based classifiers [9]. Some popular kernels are linear ($K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j\rangle$), polynomial ($K(\mathbf{x}_i, \mathbf{x}_j) = (\langle \mathbf{x}_i, \mathbf{x}_j\rangle + 1)^d$, $d \in \mathbb{Z}^+$), or Gaussian radial basis function (RBF) ($K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)$; $\sigma \in \mathbb{R}^+$, where σ is the Gaussian width). In all our tests, we used a Gaussian RBF kernel function because it has less numerical difficulties, and only the Gaussian width has to be tuned. In addition, the Gaussian RBF kernel is a universal kernel that includes other valid kernels as particular cases [17].

IV. One-Class Multiclass Architecture

Several multiclass architectures for one-class classifiers have been proposed in the literature with different degrees of sophistication [18]–[20]. Here, we propose a simple but very efficient architecture, particularly convenient when the information about labeled classes might potentially increase.

Essentially, one-class classifiers can provide two types of outputs: 1) a density estimation $\hat{p}(\mathbf{x}|\omega_k)$ or 2) a distance to a model $\delta(\mathbf{x}|\omega_k)$, being $\{\omega_k\}_{k=1}^{N_C}$ the class. Gaussian and mixture of Gaussians are classifiers of the first group, while k-NN and SVDD are classifiers of the second group.

Unlike [18], here we are not dealing with the more difficult problem of combining the outputs of different families of one-class classifiers. Instead, we only need to check the outputs of N_C one-class classifiers of the same type (each one trained to recognize a single class) and decide which class ω_k a given input vector $\mathbf{x}$ belongs to.

In the ideal case, each input test pattern should be recognized only from the classifier trained to identify that class. If the input test vector belongs to an unknown class, then it should be rejected by all classifiers. However, the problem arises when several classifiers accept an input vector simultaneously. In this work, in order to solve this situation, we assign the input vector to the most likely class. In the case of the Gaussian and mixture of Gaussians classifiers, we adopt the following criterion:

$$\arg\left\{\max_{k=1,\ldots,N_C} \hat{p}_k\right\}. \quad (12)$$

For the k-NN and SVDD classifiers, the smallest distance to the cluster and the smallest distance to the hypersphere, respectively, are used to classify the input samples. The decision is taken according to the following rule:

$$\arg\left\{\min_{k=1,\ldots,N_C} \delta_k\right\} \quad (13)$$

being δ_k as the distance from the test sample to the model, which is $\delta(\mathbf{x}|\omega_k)$ in the k-NN and SVDD one-class methods.

V. Experimental Results

In this section, we present an extensive evaluation of the SVDD technique and the other one-class methods for remote sensing data classification. Synthetic and real scenarios are considered using both hyperspectral images and fusion of multispectral and synthetic aperture radar (SAR) data.

The following problems are analyzed in this section.

1) *Synthetic problems*. Two synthetic problems are considered, and the behavior of the domain description classifiers presented in this paper under different probability density functions are studied.
2) *Real problems*. The proposed methods are tested in real remote sensing images under the following scenarios.
   a) *Hyperspectral images*.
      i) *Very little prior knowledge of the problem*. This scenario deals with the case in which one has training samples for few classes (N_C) and a relatively high number of test samples of unknown classes (m): ($m \gg N_C$).
      ii) *Reasonable prior knowledge about most of the classes*. In this case, one has training samples for almost all classes (N_C), but there are still some unknown classes (m): ($m > N_C$).

      We have identified two real scenarios considering hyperspectral images for crop cover and urban classification, respectively. Each problem considers different spectral and spatial resolutions; thus, a more complete comparison can be done.
   b) *Multispectral and SAR images*. We also consider the classification problem when multisource data are available, and thus, different distributions of the features are present. In particular, we treat the problem of using multispectral and SAR data jointly for urban monitoring. This leads to a complex classification problem.

A. Model Development

The four one-class domain description classifiers described in the previous sections have been tested in this work: Gaussian (GDD), mixture of Gaussians (MoGDD), k-nearest neighbor (KnnDD), and the SVDD (with the RBF kernel). All one-class domain description classifiers have a parameter to adjust, which is the fraction rejection, that controls the percentage of target samples the classifier can reject during training. Some other parameters need to be tuned depending on each classifier. For instance, the Gaussian classifier has a regularization parameter in the range [0–1] to estimate the covariance matrix, using all training samples or only diagonal elements. With regard to the mixture of Gaussian classifier, four parameters must be adjusted: the shape of the clusters to estimate the covariance matrix (e.g., full, diagonal, or spherical), a regularization parameter in the range [0–1] used in the estimation of the covariance matrices, the number of clusters used to model the target class, and the number of clusters to model the outlier class. The most important advantage of this classifier with respect to the Gaussian classifier, apart from the obvious improvement of modeling a class with several Gaussian distributions, is the possibility of using outliers information when training the classifier. This allows tracing a more precise boundary around the target class, improving classification results notably. However, having a total of five parameters to adjust constitutes a serious drawback for this method, making it difficult to obtain a good working model. With respect to the k-NN classifier, only the number of k neighbors used to compute the distance of each new sample to its class must be tuned. Finally, for the SVDD, the width of the RBF kernel (σ) has to be adjusted. In all the experiments, except the last one, σ is selected among the following set of discrete values: $\{\sqrt{10}, 5, 10, 25, 50, 100, 250, 500, 1000\}$. It is worth stressing here that the SVDD is one of the methods among those considered, together with the mixture of Gaussians that allows to use outliers information to better define the target class boundary. In addition, it has the important advantage that only two free parameters have to be adjusted, thus making relatively easier to define a good model.

TABLE I
Kappa Coefficient of Accuracy (κ) and Overall Accuracy (OA) Obtained for the Synthetic Problems

	GDD	**MoGDD**	**KnnDD**	**SVDD**
Distribution	κ (OA %)	κ (OA %)	κ (OA%)	κ (OA%)
2 Gaussians	0.52 (76.0)	0.67 (83.5)	0.56 (78.1)	0.63 (81.7)
Mix. Gauss/Log	0.65 (82.7)	0.77 (88.8)	0.48 (74.1)	0.78 (89.4)

In all the experiments, 30% of the samples available are used to train each classifier. In order to adjust free parameters, a cross-validation strategy with four folds is employed. Once the classifier is trained and adjusted, the final test is done using the remaining 70% of the samples. In all our experiments, we used the dd_tools [21] and LibSVM [22] software packages.

B. Synthetic Problems

Two different synthetic problems are tested here. In each problem, binary classification is addressed, thus having two sets of targets and outliers. Each class contains 500 samples generated from different distributions. The two problems analyzed are mixture of two Gaussian distributions and mixture of Gaussian and logarithmic distributions; both of them are strongly overlapped. Our goal here is testing the effectiveness of the different methods with standard and simple models but in very critical conditions on the overlapping of class distributions.

In order to measure the capability of each classifier to accept targets and reject outliers, the confusion matrix for each problem is obtained, and the kappa coefficient (κ) is estimated [23], which gives a good tradeoff between the capability of the classifier to accept its samples (targets) and reject the others (outliers).

Table I shows the results for the synthetic tests, and Fig. 2 shows graphical representations of the classification boundary defined by each method. In each plot, the classifier decision boundary is represented with a dotted curve.

Several conclusions can be obtained from these preliminary tests. First, MoGDD and SVDD obtain the highest accuracies

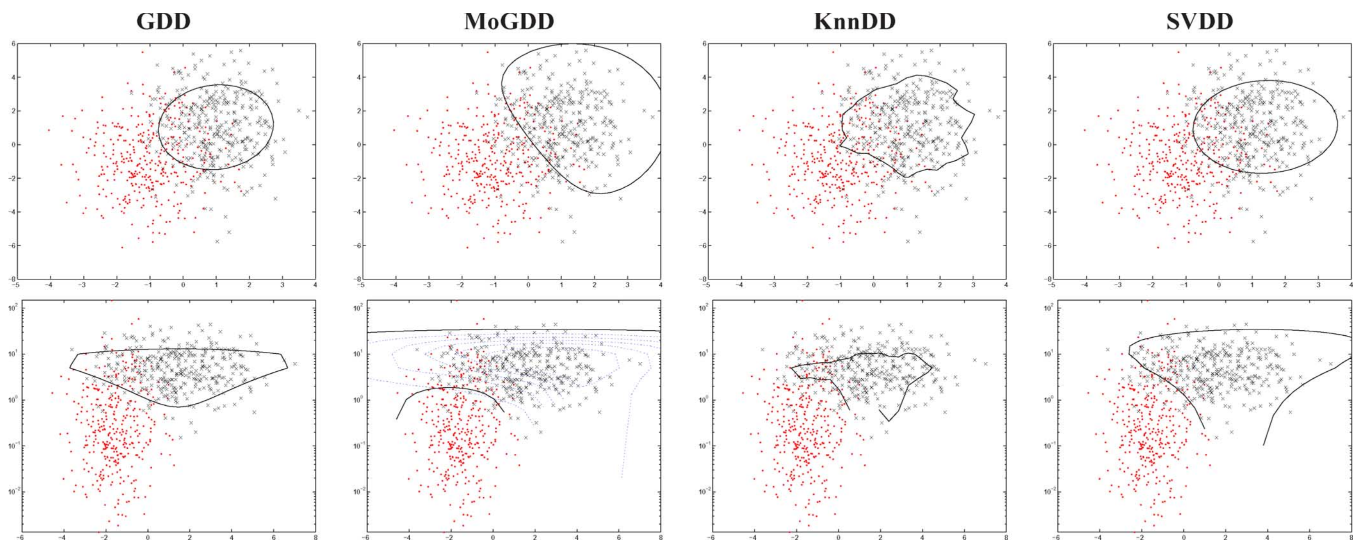


Fig. 2. Plot of the decision boundaries obtained with GDD, MoGDD, KnnDD, and SVDD for the two Gaussians problem (top) and the mixed Gaussian-logarithm problem (bottom). Note the semilogarithmic scale in this later case.

TABLE II
CLASSES CONSIDERED IN THE AVIRIS DATA SET WITH THEIR LABELS AND NUMBER OF SAMPLES

Name	Class #	# of samples
Corn-notill	2	1434
Corn-min	3	834
Corn	4	234
Grass/Pasture	5	497
Grass/Trees	6	747
Grass/pasture-moved	7	26
Hay-windrowed	8	489
Soybeans-clean	12	614
Wheat	13	212
Woods	14	1294
Bldg-Grass-Tree-Drives	15	380

in our problems. When mixing two Gaussians, none of the classifiers shows a good behavior, as in our synthetic data the distributions are strongly overlapped. This is a common situation in the remote sensing field when trying to classify classes very similar to each other. Mixture of overlapped Gaussian and logarithmic features is a difficult problem. In this synthetic example, we can see that the SVDD performs slightly better, since it does not assume any *a priori* data distribution.

C. Medium-Resolution Hyperspectral Image Classification

These experiments were carried out using the hyperspectral images acquired by the Airborne Visible/Infrared Imaging Spectroradiometer (AVIRIS) sensor in June 1992 on the Indian Pines area, Indiana [24]. The hyperspectral image contains 220 bands. In our experiments, 20 bands were discarded because they were affected by atmospheric problems.

The hyperspectral image tested has a total of 16 classes. In order to simplify the problem and obtain interpretable results, we have selected ten classes in our experiments. Table II shows labels and number of samples of the selected classes.

The methodology used to obtain the one-class classifiers is described in Section V-A. The most problematic classifier in the training phase is the mixture of Gaussians due to its relatively high number of free parameters to tune. More specifically, regarding the selection of the number of Gaussians to model the target and the outlier classes in our tests with the AVIRIS image, we noticed that adding more Gaussians to model these classes did not improve results in an appreciable way, but considering outlier information to trace the boundary around the target class resulted in a significant improvement. Note, however, that including outliers information is different from including outliers labels, as in the usual supervised approach. Taking this into account, we used only a Gaussian to model the target and another one to model the rest of outliers classes.

We addressed two possible scenarios which are common in the remote sensing field. In the first scenario, we simulated having information about the samples of one class only for training the classifier. It could be possible having samples of some other classes too, but here, we were only interested in knowing whether a given sample belongs to a target class or not. Thus, in order to test each one-class classifier, we used samples of all the classes in the image. In the second scenario, we simulated having labels for almost all classes in the image, but still some classes remained unlabeled. In this case, we built a multiclass classifier based on one-class algorithms and tested it using samples of all the classes present in the image.

1) First Scenario—Samples of One Class Are Available: Table III shows the results for the methods tested in this situation in which we feed each classifier with samples of its class (targets) and samples of other classes (outliers) to evaluate both acceptance and rejection properties. Next, the confusion matrix was built, and from it, we computed the kappa coefficient (κ) and the overall accuracy (OA) [%]. Only seven out of the nine available classes were used to develop the classifiers. These are the same classes we used in the second scenario to build the one-class-based multiclassifier.

In general terms, the best results are achieved by the SVDD, even when classifying difficult classes like corn-min (3) and soy beans-clean (12), for which other methods almost fail. However, in some cases, due to the underlying Gaussian distribution of some of the classes of the image, GDD or MoGDD

TABLE III
KAPPA COEFFICIENT (κ) AND OVERALL ACCURACY (OA [%]) FOR THE AVIRIS IMAGE WITH GAUSSIAN CLASSIFIER, MIXTURE OF GAUSSIANS CLASSIFIER, k-NN CLASSIFIER, AND SVDD. THE CORRESPONDING FREE PARAMETER (PAR) AND FRACTION REJECTION (F.R.) RATIOS ARE INDICATED; FOR GDD AND MoGDD PAR IS A REGULARIZATION TERM AND THE TYPE OF COVARIANCE MATRIX (F: FULL, D: DIAGONAL); FOR KnnDD PAR IS THE NUMBER OF NEIGHBOURS; AND FOR SVDD σ IS THE RBF KERNEL WIDTH

	GDD			**MoGDD**			**KnnDD**			**SVDD**		
Class	par	f.r.	κ (OA %)	par	f.r.	κ (OA %)	par	f.r.	κ (OA %)	σ	f.r.	κ (OA %)
2	0.1	0.1	0.69 (89.6)	0.08 f	0.1	0.73 (90.9)	5	0.3	0.65 (88.8)	5	0.2	**0.83** (94.3)
3	0.3	0.2	0.51 (88.4)	0.17 f	0.2	0.57 (90.3)	5	0.4	0.58 (90.6)	5	0.3	**0.71** (93.7)
6	0.2	0.01	0.92 (98.4)	0.14 f	0.0001	**0.93** (98.6)	7	0.2	0.80 (96.4)	10	0.01	0.89 (97.9)
8	0.9	0.0001	**0.98** (99.7)	0.00 d	0.01	0.94 (99.2)	5	0.01	0.97 (99.5)	10	0.001	0.96 (99.4)
12	0.1	0.1	0.54 (91.4)	0.08 f	0.1	0.60 (93.3)	9	0.4	0.36 (86.7)	5	0.001	**0.77** (96.3)
13	0.9	0.0001	**0.91** (99.5)	0.15 d	0.1	0.85 (99.1)	6	0.1	0.89 (99.3)	25	0.1	0.85 (99.1)
14	0.1	0.01	0.87 (95.5)	0.10 f	0.01	**0.89** (96.5)	8	0.1	0.82 (94.4)	10	0.1	0.86 (95.8)

TABLE IV
KAPPA COEFFICIENT OF ACCURACY (κ) AND OVERALL ACCURACY (OA) FOR THE AVIRIS IMAGE WITH THE MULTICLASS SCHEME BUILT WITH SEVEN ONE-CLASS CLASSIFIERS

GDD	**MoGDD**	**KnnDD**	**SVDD**
κ (OA %)	κ (OA %)	κ (OA %)	κ (OA %)
0.75 (79.2)	0.71 (76.3)	0.65 (70.7)	**0.79** (82.4)

performed slightly better than SVDD. We can also see that KnnDD showed low accuracies in almost all cases.

2) Second Scenario—Having Labeled Samples of Almost All Classes: In this scenario, we used the seven previously trained one-class classifiers and the multiclass architecture described in Section IV to build a multiclass classifier for seven classes. Then, we tested it with all the classes in the image. It is worth noting that the main property of this multiclass classifier is that it is not only able to classify known classes (by accurately describing the domain or class distribution), but it also rejects samples from unknown classes instead of misclassifying them. Also, the combination of one-class classifiers worked better than each one-class classifier working alone.

Table IV shows the results of the seven-class multiclass classifiers tested. We can see that the best accuracies were obtained by the combination of the SVDD one-class classifiers.

D. High-Resolution Hyperspectral Image Classification

In this section, one-class classifiers on a high-resolution hyperspectral urban image of the area of Pavia, Italy, are tested. The image is a nine-class data set, acquired from DAIS 7915 airbone imaging spectrometer of Deutsches Zentrum für Luft- und Raumfahrt (DLR) [25]. This is a challenging urban classification problem dominated by directional features and relatively high spatial resolution (5-m pixels). The image has a size of 400 $\times$ 400 pixels (2000 $\times$ 2000 m), 40 bands and 9 labeled classes. Class labels and their number of samples are shown in Table V.

We followed the same procedure used in the section before to train and test the one-class classifiers. In this example, we trained classifiers for the following classes: $\{1, 3, 4, 6, 7, 9\}$, which correspond to water, brick roofs, asphalt, meadows, bitumen, and shadows, respectively. To test the classifiers, samples of all classes in the image were used.

TABLE V
DAIS 7915 PAVIA IMAGE CLASSES, LABELS, AND NUMBER OF SAMPLES

Name	**Class #**	**# of samples**
Water	1	4281
Trees	2	2424
Brick roofs	3	2237
Asphalt	4	1699
Bare soil	5	1475
Meadows	6	1245
Bitumen	7	685
Parking lot	8	287
Shadows	9	241

1) First Scenario—Samples of One Class Are Available: Table VI shows the accuracy for each one-class classifier. Note that only the six classes used in the second scenario to build the multiclass classifier are considered but here each classifier was used to recognize only its own class. Best results are boldfaced. The main conclusion is that, again, the best overall accuracies are those obtained by using SVDD one-class classifiers. Only in class shadows (9), MoGDD obtained better results. However, this can be due to the fact that this is the class with the lowest number of labeled samples.

2) Second Scenario—Having Labeled Samples of Almost All Classes: In a similar way as we did in the previous section, in this scenario, we built a six-class multiclass classifier made-up of six one-class classifiers. Table VII shows the obtained accuracies. In agreement with the results in Table VI, the combination of SVDD one-class classifiers clearly provides the highest accuracies, yielding a gain in the overall accuracy between 2% and 4%.

E. Multispectral and SAR Image Fusion

The images used in this section were collected in the Urban Expansion Monitoring (UrbEx) ESA-ESRIN DUP project [26]. Results from UrbEx project were used both to perform the analysis of the selected test site and for validation purposes (for further details, visit http://dup.esrin.esa.int/ionia/projects/summaryp30.asp). The considered test site is the city of Rome, Italy, where images from ERS2 SAR and Landsat Thematic Mapper (TM) sensors were acquired in 1999. In the classification, we considered the seven Landsat TM optical

TABLE VI

KAPPA COEFFICIENT (κ) AND OVERALL ACCURACY (OA [%]) FOR THE DAIS 7915 IMAGE WITH GAUSSIAN CLASSIFIER, MIXTURE OF GAUSSIANS CLASSIFIER, k-NN CLASSIFIER, AND SVDD. THE CORRESPONDING FREE PARAMETER (PAR) AND FRACTION REJECTION (F.R.) RATIOS ARE INDICATED; FOR GDD AND MoGDD PAR IS A REGULARIZATION TERM AND THE TYPE OF COVARIANCE MATRIX (F: FULL, D: DIAGONAL); FOR KnnDD PAR IS THE NUMBER OF NEIGHBOURS; AND FOR SVDD σ IS THE RBF KERNEL WIDTH

	GDD			MogGDD			KnnDD			SVDD		
Class	par	f.r.	κ (OA %)	par	f.r.	κ (OA %)	par	f.r.	κ (OA %)	σ	f.r.	κ (OA %)
1	0.9	0.0001	0.99 (99.9)	0.00 d	0.0001	**1.00** (100)	5	0.0001	0.99 (99.9)	$\sqrt{10}$	0.0001	0.99 (99.9)
3	0.4	0.1	0.94 (98.5)	0.13 f	0.1	0.93 (98.3)	5	0.1	0.93 (98.4)	100	0.0001	**0.99** (99.8)
4	0.3	0.01	0.94 (98.7)	0.19 f	0.01	**0.96** (99.1)	6	0.1	0.93 (98.7)	10	0.001	**0.96** (99.3)
6	0.6	0.1	0.89 (98.4)	0.04 f	0.1	0.91 (98.7)	6	0.1	0.89 (98.4)	10	0.0001	**0.96** (99.4)
7	0.1	0.01	0.96 (99.6)	0.03 f	0.01	0.96 (99.6)	7	0.1	0.91 (99.2)	25	0.0001	**0.98** (99.8)
9	0.1	0.1	0.92 (99.7)	0.02 f	0.0001	**0.94** (99.8)	6	0.3	0.81 (99.4)	10	0.0001	0.88 (99.6)

TABLE VII

KAPPA COEFFICIENT OF ACCURACY (κ) AND OVERALL ACCURACY (OA) FOR DAIS 7915 IMAGE WITH THE MULTICLASS SCHEME BUILT WITH SIX ONE-CLASS CLASSIFIERS

GDD	MoGDD	KnnDD	SVDD
κ (OA %)	κ (OA %)	κ (OA %)	κ (OA %)
0.94 (95.9)	0.95 (96.1)	0.92 (94.4)	**0.97** (98.2)

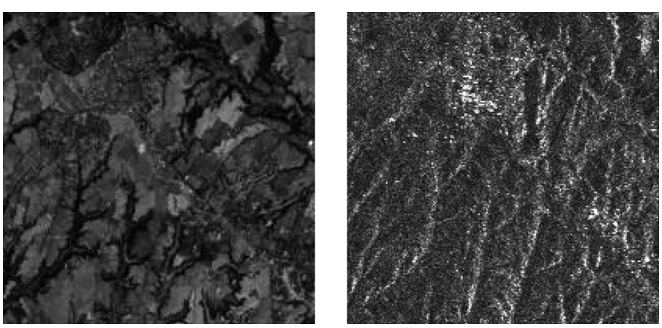


Fig. 3. Images of the test areas of Rome acquired in 1999 both in (left) band L3 from Landsat and (right) SAR first log-intensity band.

spectral bands, the two SAR backscattering intensities (0–35 days), the coherence between the two SAR signals, and an additional feature computed by applying a multistage spatial filter to the coherence images. Once that features were extracted from optical and SAR images, we analyzed their potential use for multisource image classification (see more details in [27]).

1) Model Development and Free Parameter Selection: In this experiment, we selected a subset image containing 200 $\times$ 200 pixels (see Fig. 3). We randomly selected 25% of the changed pixels for training and used five-fold cross-validation for free parameter tuning. Then, we tested the built classifier on the whole subset image. The RBF kernel widths were tuned in the range $\sigma = \{10^{-3}, \ldots, 10^{3}\}$, and the rejection fraction parameter for the SVDD method was tuned in the range $\{10^{-3}, \ldots, 10^{0}\}$.

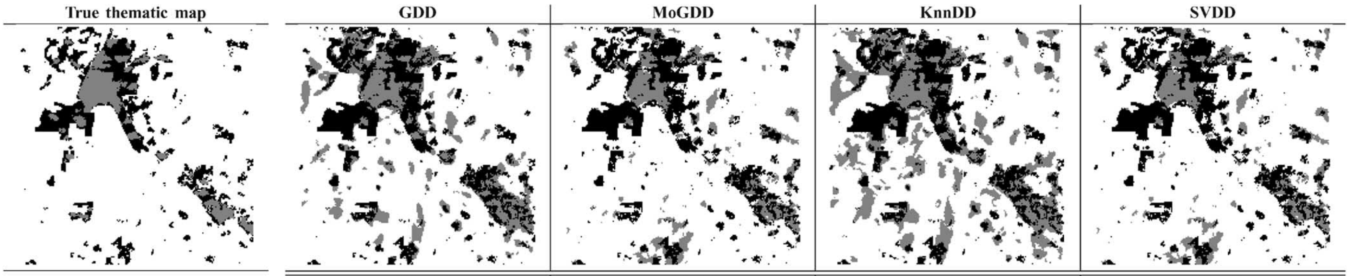
2) Model Comparison: Fig. 4 shows the true classification map for the Rome dataset 1999 and the test results obtained by the classifiers (average of 10 realizations) along with the classification maps for all of them. Several conclusions can be obtained. First, SVDD offers very good results since rather than building a separating hyperplane "urban"/"nonurban," the method tries to model the "urban" class accurately. Second, the standard Gaussian classifier fails due to the fact that the data do not fit the assumed distribution. The mixture of Gaussian classifier improves the results obtained by the Gaussian classifier, but its accuracies are slightly worse than those provided by the SVDD. Furthermore, the KnnDD classifier produces the worst results, suggesting that data lie in much more complex underlying subspace than that detected locally through neighbors. Visually, results offered by the SVDD method are much better than the rest of classifiers, yielding more homogeneous areas and lower number of false detections.

VI. DISCUSSION AND CONCLUSION

This paper has addressed the problem of supervised classification of remote sensing images in the presence of incomplete (nonexhaustive) training sets. The analysis has been carried out according to two different objectives: 1) description and recognition of a specific land-cover class by using single-class classifiers and 2) solution of multiclass problems with single-class classifiers integrated in a classification architecture capable to properly consider the lack of training information for some land-cover classes. In this framework we introduced the SVDD technique in multiclass detection problems with incomplete datasets and analyzed its effectiveness compared to other one-class standard techniques.

The experimental part of this paper has been conducted on different data sets. First, we have tested the different techniques by using synthetic data, to show their behavior when dealing with typical (but controlled) probability distribution models adopted in remote sensing problems which include mixtures of two Gaussians, and of Gaussian and logarithmic distributions. Second, we have analyzed these methods by considering two hyperspectral and one multisensor data sets. These last experiments have been carried out in order to: 1) test the ability of different techniques in modeling and recognizing a single class and 2) analyze the effectiveness of the combination of different one-class classifiers for solving multiclass problems.

An important conclusion of this paper, which is derived from all the experimental results obtained, is that one-class domain description classifiers represent a very promising methodology in the presence of incomplete training knowledge, allowing a flexible adaptation of the classification architecture to the available information. In our experiments, owing to their rejection properties, these classifiers provided good classification accuracies in problems associated with both the recognition of



	GDD	MoGDD	KnnDD	SVDD
	κ (OA %)	κ (OA %)	κ (OA %)	κ (OA %)
	0.58 (66.2)	0.61 (69.7)	0.48 (59.1)	**0.63** (77.3)

Fig. 4. True thematic map for the Rome dataset (1999). (Middle) Obtained classifications for the Rome dataset (1999). In classification maps, white pixels represent the class "nonurban," black pixels are "unknown class," and gray pixels are "urban." (Bottom) Kappa coefficient and overall accuracy (average results over ten realizations) with Gaussian domain description (GDD), mixture of Gaussians (MoGDD), k-NN domain description (KnnDD), and SVDD.

a single-class and the recognition of a subset of the unknown complete set of classes characterizing a given study area.

Concerning the accuracy obtained from the single classification techniques, the SVDD classifier provided very good results, with significant improvement of the classification accuracy with respect to other well-established one-class techniques, particularly on the multisensor data set. This is in agreement with what was expected, given the distribution-free property of this method, which is particularly suitable to address problems in which data acquired by different sensors do not allow a rigid parametric modeling of the class distributions. SVDD proved effective also in addressing hyperspectral image classification, yielding the highest classification accuracy in most of the cases. This proves the robustness of this kernel-based method to deal with hyperdimensional feature spaces. The accuracy of the SVDD classifier is obtained at the expense of a relatively high computational cost since a quadratic programming problem containing as many restrictions as the available training samples must be solved. However, this does not constitute a critical issue in our case for two reasons: first, only the training phase is computationally demanding as the classification function is based on a sparse model (few SVs are typically selected to characterize the domain); and second, only two free parameters (kernel function parameter, if any, and the fraction rejection ratio) must be tuned. Another good property inferred from the proposed multiclass strategy is that it is not mandatory to retrain the scheme if an additional class comes to the scene, as occurs in the majority of inductive global domain description classification schemes like SVMs.

With regard to the behavior of the other analyzed methods, as expected, Gaussian and mixture of Gaussian domain description classifiers obtained good results when dealing with data that can be reasonably modeled with Gaussian distributions. In these situations (which include also hyperspectral data sets), the use of Gaussian-like classifiers results in classification accuracies similar (or in few cases slightly better) than those provided by the SVDD technique. In general, the mixture of Gaussian domain description method can improve the results obtained by the more simple Gaussian domain description classifier. However, we should take into account that an important drawback of the mixture of Gaussian method is the relatively high number of free parameters to adjust compared with other methods (i.e., the number of Gaussians to mix both for target and outlier classes, and the shape of the data used to estimate the covariance matrices). In addition, the computational time associated with the training of a mixture of Gaussian classifier can be twofold that of a simple Gaussian classifier. Regarding the k-NN domain description classifier, in spite of its distribution-free properties, poor accuracies are obtained, particularly in high-dimensional spaces or when data lie in a complex subspace. Also, it is worth stressing here that k-NN classifiers do not perform any learning task but, rather, a memorization process, and thus, the training phase is very fast (only comparison of a sample to its neighbors is performed), but the time required from the operational test phase dramatically increases.

Finally, considering all the obtained results, we can conclude that, in general, the SVDD classifier (and the multiclass scheme proposed here) provides better results than the rest of the methods, in presence of incomplete training data. These results are particularly noticeable in complex environments such as high-dimensional data and multisource data.

ACKNOWLEDGMENT

The authors would like to thank Prof. P. Gamba (University of Pavia, Italy) for providing the reference data for the Pavia dataset, Prof. D. Landgrebe (Purdue University, USA) for providing the AVIRIS data, Dr. C.-J. Lin (Taiwan University) for providing the libSVM library, and Dr. D. M. J. Tax (Delft University of Technology) for kindly offering his dd_tools package.

REFERENCES

[1] 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.
[2] 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.
[3] M. Chi and L. Bruzzone, "A semilabeled-sample-driven bagging technique for ill-posed classification problems," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 1, pp. 69–73, Jan. 2005.
[4] L. Bruzzone, M. Chi, and M. Marconcini, "Transductive SVMs for semi-supervised classification in ill-posed problems," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3363–3373, Nov. 2006.
[5] D. Tax and R. P. Duin, "Support vector domain description," *Pattern Recognit. Lett.*, vol. 20, no. 11–13, pp. 1191–1199, Nov. 1999.

[6] D. Tax, "One-class classification, Concept-learning in the absence of counter-examples," Ph.D. dissertation, Delft Univ. Technol., Delft, The Netherlands, 2001.

[7] G. Mercier and F. Girard-Ardhuin, "Partially supervised oil-slick detection by SAR imagery using kernel expansion," *IEEE Trans. Geosci. Remote Sens.*, vol. 4, no. 10, pp. 2839–2846, Oct. 2006.

[8] B. Schölkopf and A. Smola, *Learning With Kernels-Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, MA: MIT Press, 2002.

[9] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[10] G. Shakhnarovich, T. Darrell, and P. Indyk, *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice*. Cambridge, MA: MIT Press, Mar. 2006.

[11] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is 'nearest neighbor' meaningful?" in *Lecture Notes in Computer Science*, vol. 1540. Berlin, Germany: Springer-Verlag, 1999, pp.217–235.

[12] P. Vincent and Y. Bengio, "K-local hyperplane and convex distance nearest neighbor algorithms," in *Advances in Neural Information Processing Systems 14*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds. Cambridge, MA: MIT Press, 2002.

[13] J. A. Gualtieri and R. F. Cromp, "Support vector machines for hyperspectral remote sensing classification," in *Proc. SPIE, 27th AIPR Workshop*, Feb. 1998, pp. 221–232.

[14] G. Camps-Valls, L. Gómez-Chova, J. Calpe, E. Soria, J. D. Martín, L. Alonso, 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.

[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. 1178–1790, Aug. 2004.

[16] B. Schölkopf, R. C. Williamson, A. Smola, and J. Shawe-Taylor, "Support vector method for novelty detection," in *Advances in Neural Information Processing Systems*, vol. 12. Cambridge, MA: MIT Press, 1999, pp. 582–588.

[17] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," *Neural Comput.*, vol. 15, no. 7, pp. 1667–1689, Jul. 2003.

[18] R. Duin and D. Tax, "Experiments with classifier combining rules," in *Proc. Int. Workshop MCS*, Cagliari, Italy, Jun. 2000, vol. 1875, pp. 16–29.

[19] D. Tax and R. Duin, "Combining one-class classifiers," in *Proc. Int. Workshop MCS*, Cambridge, U.K., Jul. 2001, vol. 2096, pp. 299–308.

[20] D. Lee and J. Lee, "Domain described support vector classifier for multi-classification problems," *Pattern Recognit.*, vol. 40, no. 1, pp. 41–51, Jan. 2007.

[21] D. Tax, *DDtools, the Data Description Toolbox for Matlab*, Jun. 2006. version 1.5.3.

[22] C.-C. Chang and C.-J. Lin, *LIBSVM: A library for support vector machines*, 2001. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin

[23] K. G. Russell and G. Congalton, *Assessing the Accuracy of Remotely Sensed Data*. Boca Raton, FL: CRC Press, 1998.

[24] D. Landgrebe, *AVIRIS NW Indiana's Indian Pines 1992 data set*, 1992. [Online]. Available: http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/documentation.html

[25] German Aerospace Center (DLR), *DLR Airbone Image Spectroscopy, DAIS 7915*, 2001. [Online]. Available: http://www.op.dlr.de/DAIS/

[26] P. Castracane, F. Iavaronc, S. Mica, E. Sottile, C. Vignola, O. Arino, M. Cataldo, D. Fernandez-Prieto, G. Guidotti, A. Masullo, and I. Pratesi, "Monitoring urban sprawl and its trends with EO data. UrbEx, a prototype national service from a WWF-ESA joint effort," in *Proc. 2nd GRSS/ISPRS Joint Workshop Remote Sens. and Data Fusion Over Urban Areas*, 2003, pp. 245–248.

[27] L. Gómez-Chova, D. Fernández-Prieto, J. Calpe, E. Soria, J. Vila, and G. Camps-Valls, "Urban monitoring using multitemporal SAR and multispectral data," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 234–243, Mar. 2006.

Jordi Muñoz-Marí was born in València, Spain, in 1970. He received the B.Sc. degree in physics, the B.Sc. degree in electronics engineering, and the Ph.D. degree in electronics engineering from the Universitat de València, València, in 1993, 1996, and 2003 respectively.

He is currently an Associate Professor in the Department of Electronics Engineering, Universitat de València, where teaches analysis of circuits and linear systems, introduction to programmable logical devices, digital electronic systems, and electronic systems with microprocessor. His research interests include kernel methods for signal and image processing.

Lorenzo Bruzzone (S'95–M'99–SM'03) received the laurea (M.S.) degree in electronic engineering (*summa cum laude*) and the Ph.D. degree in telecommunications from the University of Genoa, Genoa, Italy, in 1993 and 1998, respectively.

He is currently Full Professor of telecommunications and Head of the Remote Sensing Laboratory at the Department of Information and Communication Technologies, University of Trento, Trento, Italy. He teaches remote sensing, pattern recognition, and electrical communications. His current research interests are in the area of remote sensing image processing and recognition (analysis of multitemporal data, feature selection, classification, regression, data fusion, and neural networks). He conducts and supervises research on these topics within the frameworks of several national and international projects. He is the author (or coauthor) of more than 150 scientific publications, including journals, book chapters, and conference proceedings. He is a referee for many international journals and has served on the scientific committees of several international conferences.

Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium, Seattle, July 1998. He was a recipient of the recognition of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Best Reviewers in 1999. He was the General Chair and Cochair of the First and Second, respectively, IEEE International Workshop on the Analysis of Multitemporal Remote-Sensing Images. Since 2003, he has been the Chair of the International Society for Optical Engineers Conference on Image and Signal Processing for Remote Sensing. Since 2005, he has been an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING.

Gustavo Camps-Valls (M'04–SM'07) was born in València, Spain, in 1972. He received the B.Sc. degree in physics, the B.Sc. degree in electronics engineering, and the Ph.D. degree in physics from the Universitat de València, València, in 1996, 1998, and 2002, respectively.

He is currently an Associate Professor in the Department of Electronics Engineering, Universitat de València, where teaches electronics, advanced time series processing, and signal processing. His research interests include neural networks and kernel methods for signal and image processing. He is the author (or coauthor) of 40 journal papers, more than 60 international conference papers, several book chapters, and editor of the book *Kernel Methods in Bioengineering, Signal and Image Processing* (IGI, 2007). He is a referee of many international journals and currently serves on the program committees of the International Society for Optical Engineers Europe, the International Geoscience and Remote Sensing Symposium (IGARSS), and the International Conference on Image Processing.