

Semisupervised One-Class Support Vector Machines for Classification of Remote Sensing Data

Jordi Muñoz-Marí, Francesca Bovolo, *Member, IEEE*, Luis Gómez-Chova, *Member, IEEE*, Lorenzo Bruzzone, *Fellow, IEEE*, and Gustavo Camps-Valls, *Senior Member, IEEE*

Abstract—This paper presents two semisupervised one-class support vector machine (OC-SVM) classifiers for remote sensing applications. In *one-class* image classification, one tries to detect pixels belonging to one of the classes in the image and reject the others. When few labeled pixels of only one class are available, obtaining a reliable classifier is a difficult task. In the particular case of SVM-based classifiers, this task is even harder because the free parameters of the model need to be finely adjusted, but no clear criterion can be adopted. In order to improve the OC-SVM classifier accuracy and alleviate the problem of free-parameter selection, the information provided by unlabeled samples present in the scene can be used. In this paper, we present two state-of-the-art algorithms for semisupervised one-class classification for remote sensing classification problems. The first proposed algorithm is based on modifying the OC-SVM kernel by modeling the data marginal distribution with the graph Laplacian built with both labeled and unlabeled samples. The second one is based on a simple modification of the standard SVM cost function which penalizes more the errors made when classifying samples of the target class. The good performance of the proposed methods is illustrated in four challenging remote sensing image classification scenarios where the goal is to detect one of the classes present on the scene. In particular, we present results for multisource urban monitoring, hyperspectral crop detection, multispectral cloud screening, and change-detection problems. Experimental results show the suitability of the proposed techniques, particularly in cases with few or poorly representative labeled samples.

Index Terms—Change detection, one-class classification, one-class support vector machine (OC-SVM), semisupervised learning (SSL), support vector domain description (SVDD), target detection.

I. INTRODUCTION

IN REMOTE sensing image classification, it is quite common to deal with reduced sets of labeled samples when developing classifiers. Support-vector-machine (SVM)-based classifiers excel in using the labeled information, with (regularized) maximum margin classifiers also being equipped with an appropriate loss function [1], [2]. However, applicability of

the SVM is only possible when labeled samples of all the land-cover classes present in the scene are available. When such information is only available for one class of interest (or few), other techniques should be used. In particular, high interest has been devoted to the learning frameworks of the following: 1) *anomaly detection*, where one tries to identify pixels differing significantly from the background; 2) *target detection*, where the target spectral signature is assumed to be known and the goal is that of detecting pixels that match the target; and 3) *one-class* classification, where one tries to detect one class and reject the others.

This paper is focused on one-class classification. In the past, several kernel-based methods have been developed to this purpose. The use of kernel methods offers many advantages with regard to other approaches, as they alleviate the curse of dimensionality in hyperspectral images, increase the robustness of the method to noise, and allow flexible and smooth nonlinear mappings [1]. Kernel methods in general, and kernel-based classifiers in particular, rely on the proper definition of a kernel (or similarity) function between samples. In particular, the one-class SVM (OC-SVM) [3] aims at identifying samples of one particular class while rejecting all the others. In the remote sensing literature, the method was originally introduced for anomaly detection [4], [5], then exercised in incomplete and unreliable training data problems [6], and recently engineered for change detection [7]. Nevertheless, when very few or less representative training samples are available, the OC-SVM may result in unreliable classification results. Thus, for dealing with this kind of problems, OC-SVM can be reformulated in the framework of semisupervised learning (SSL) to exploit not only labeled but also unlabeled sample [8]–[10] information.

In this paper, we introduce two semisupervised OC-SVM methods. The first method, named semisupervised OC-SVM (S^2 OC-SVM), uses the available supervised information (*labeled*) and also the data with no *a priori* class information (*unlabeled*) to encode some knowledge about the geometry and data distribution. The exploration of the shape of the marginal distribution of data adds significant information in order to better position the decision boundary. It is worth noting that this kind of procedure can be applied to any kernel-based classification method and results in a modification of the measure of similarity in the kernel space according to the geometry of the unlabeled samples. This is done by including an additional regularization term on the geometry of both labeled and unlabeled samples by using the *graph Laplacian* [11], [12]. The second method, named biased-SVM (b-SVM), was originally proposed for text classification in [13], and it

Manuscript received July 27, 2009; revised December 18, 2009 and February 25, 2010. Date of publication May 6, 2010; date of current version July 21, 2010. This work was supported in part by Projects EODIX/AYA2008-05965-C04-03 and CONSOLIDER/CSD2007-00018.

J. Muñoz-Marí, L. Gómez-Chova, and G. Camps-Valls are with the Department Ingeniería Electrónica, Escola Tècnica Superior d'Enginyeria, Universitat de València, 46100 València, Spain (e-mail: jordi@uv.es; luis.gomez-chova@uv.es; gustavo.camps@uv.es).

F. Bovolo and L. Bruzzone are with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy (e-mail: francesca.bovolo@dit.unitn.it; 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.2010.2045764

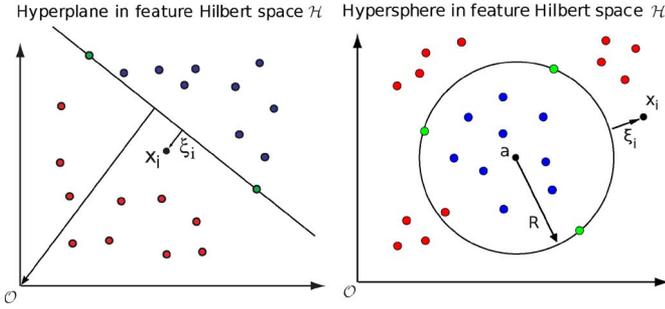


Fig. 1. (Left) OC-SVM: The hyperplane separates with maximum margin all target data from the origin by mapping all targets (blue) to the upper side of the hyperplane and outliers (red) to the lower side. Only the green points are needed to make predictions. \mathbf{x}_i is an outlier sample. ξ_i is one of the slack variables introduced to deal with outliers that may be present in the training set. (Right) SVDD: The hypersphere surrounds all target samples. \mathbf{a} and R are the center and the radius of the hypersphere, respectively.

is based on using the standard binary SVM classifier, where labeled samples represent the target class and unlabeled data are treated as being the outlier class. The SVM cost function is built such that errors committed on labeled target samples are much more penalized than errors occurred on nontarget class. This is because samples belonging to the outlier class show low (or lower) certainty as they are in fact unlabeled. The performance of the proposed S^2OC -SVM and b-SVM methods is illustrated in four different challenging problems: urban monitoring, crop detection, cloud masking, and change detection. Results are compared with the ones achieved by other one-class classifiers and, when possible, by a standard binary classifier.

The rest of this paper is organized as follows. Section II briefly reviews the formulation of the OC-SVM. Noting that the key for obtaining successful performance is the definition of the kernel structure and the adequate tuning of free parameters, Section III is devoted to introduce the proposed semisupervised version of the OC-SVM. Section IV introduces the b-SVM method. Section V analyzes the data sets used in this paper and presents experimental results. Section VI analyzes the obtained results, and finally, Section VII concludes this paper.

II. OC-SVM

The OC-SVM was introduced by Schölkopf *et al.* as a support vector method for novelty detection [3]. Notationally, let us consider a data set $\{\mathbf{x}_i\}_{i=1}^l \in \mathbb{R}^N$ belonging to a given *class of interest*, named *target class*. The idea behind the OC-SVM is to describe the target class by a function that maps most part of it to a region where the function is nonzero. To this end, the origin is treated as the only available member of the nontarget class (i.e., as an *outlier*), and then, the problem is solved by finding a *hyperplane* with maximum margin separation from the origin. To deal with nonlinearly problems, the hyperplane is defined in a high-dimensional Hilbert feature space \mathcal{H} where the samples are mapped through a nonlinear transformation $\phi(\cdot)$ (Fig. 1, left). To separate the data set from the origin, one has to minimize

$$\min_{\omega, \rho, \xi} \left\{ \frac{1}{2} \|\omega\|^2 - \rho + \frac{1}{\nu l} \sum_i \xi_i \right\} \quad \forall i = 1, \dots, l \quad (1)$$

constrained to

$$\langle \omega, \phi(\mathbf{x}_i) \rangle \geq \rho - \xi_i. \quad (2)$$

Here, ω is a vector perpendicular to the hyperplane in \mathcal{H} , and ρ is the distance to the origin. Since the training data distribution may contain outliers, a set of *slack variables* $\xi_i \geq 0$ is introduced to deal with them, as usual in the SVM framework. The parameter $\nu \in (0, 1]$ controls the tradeoff between the number of examples of the training set mapped as positive by the decision function $f(x) = \text{sgn}(\langle \omega, \phi(\mathbf{x}_i) \rangle - \rho)$ and having a small value of $\|\omega\|$ to control model complexity.

After including restrictions (2) into (1) through the use of Lagrange multipliers α_i and deriving w.r.t. ω , the relation between primal and dual weights is expressed as a linear combination of the mapped samples with $\alpha_i \neq 0$, i.e., $\omega = \sum_i \alpha_i \phi(\mathbf{x}_i)$. Introducing a kernel function whose elements are defined as $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$, the direct computation of the nonlinear mapping $\phi(\cdot)$ can be avoided, and the dual problem becomes

$$\min_{\alpha} \left\{ \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \right\} \quad (3)$$

constrained to

$$0 \leq \alpha_i \leq \frac{1}{\nu l} \quad \text{and} \quad \sum_i \alpha_i = 1. \quad (4)$$

After solving the dual problem, a set of model weights (α_i) is obtained, and the decision function for any test vector \mathbf{x}_* is given by

$$f(\mathbf{x}_*) = \text{sgn} \left(\sum_i \alpha_i K(\mathbf{x}_i, \mathbf{x}_*) - \rho \right). \quad (5)$$

There exists in the literature an alternative and, under some conditions, equivalent formulation of the OC-SVM called support vector domain description (SVDD) which was originally introduced by Tax and Duin [14]. SVDD, instead of finding a hyperplane separating the samples of the target class from the origin with maximum separation, tries to find a hypersphere with minimum volume surrounding only the target class samples. In this way, all samples lying inside the hypersphere belong to the target class, whereas the samples lying outside are outliers (Fig. 1, right). The problem is defined as

$$\min_{R, \mathbf{a}} \left\{ R^2 + \frac{1}{\nu n} \sum_i \xi_i \right\} \quad \forall i = 1, \dots, l \quad (6)$$

subject to

$$\|\phi(\mathbf{x}_i) - \mathbf{a}\|^2 \leq R^2 + \xi_i \quad \forall i = 1, \dots, l \quad (7)$$

$$\xi_i \geq 0 \quad \forall i = 1, \dots, l. \quad (8)$$

It can be shown that, when working with isotropic kernels (for example, the radial basis function (RBF) Gaussian kernel) and normalized data, both formulations yield the same solutions [3].

III. PROPOSED S²OC-SVM

In this section, we pay attention to the appropriate definition of the kernel under semisupervised criteria. Essentially, we propose to deform the kernel using the graph Laplacian. This idea was originally presented in [11] for the inductive SVM and has been recently presented for kernel orthogonal subspace projection target detection [15]. Here, it is applied to the OC-SVM.

A. Learning a Suitable Kernel From Unlabeled Samples

The performance of any kernel method strongly depends on the adequate definition of the kernel structural form, which can be casted as a similarity measure among samples. Traditionally, the kernel form has been chosen to be either 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 RBF ($K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)$, $\sigma \in \mathbb{R}^+$) to fulfill Mercer’s conditions. However, despite the good performance offered, by imposing such “*ad hoc*” signal relations, the underlying data structure is obviated. To properly define a suitable kernel, unlabeled information and geometrical relationships between labeled and unlabeled samples may be useful to “deform” the core measure of distance, i.e., to estimate a likelihood kernel according to the unlabeled data structure that modifies the assumed prior kernel encoding signal relations.

B. Deforming the Kernel With the Graph Laplacian

For the formulation of the S²OC-SVM, let us consider the data set $\{\mathbf{x}_i, y_i\}_{i=1}^{l+u}$, where $\mathbf{x}_i \in \mathbb{R}^N$ and $y \in \{+1, -1\}$, made up of l labeled samples and u unlabeled samples, $l + u = n$. For the first l samples, the label is known, $y_i = 1$. For the next u samples, the label is unknown.

The classical kernel-based learning approach is based on solving the regularization problem given by

$$f = \arg \min_{h \in \mathcal{H}} \left\{ \frac{1}{l} \sum_{i=1}^l V(h, \mathbf{x}_i, y_i) + \gamma_A \|h\|_{\mathcal{H}}^2 \right\} \quad (9)$$

where $V(\cdot)$ is a loss function (hinge loss for the SVM), $\|h\|_{\mathcal{H}}$ is the norm of the classification function h in the reproducing kernel Hilbert space (RKHS), and γ_A controls the complexity of function h . The Representer Theorem [16] states that a solution can be found in the form $f(\mathbf{x}) = \sum_{i=1}^l \alpha_i K(\mathbf{x}, \mathbf{x}_i)$.

Graph-based SSL methods, based on the manifold assumption¹, define a nearest neighbor graph using all data (labeled and unlabeled), with edge weights $W_{ij} > 0$ for neighbors and 0 otherwise, and then solve the following regularization problem:

$$f = \arg \min_{h \in \mathcal{H}} \left\{ \frac{1}{l} \sum_{i=1}^l V(h, \mathbf{x}_i, y_i) + \gamma_A \|h\|_{\mathcal{H}}^2 + \frac{\gamma_I}{n^2} \sum_{i=1}^n \sum_{j=1}^n W_{ij} (h(\mathbf{x}_i) - h(\mathbf{x}_j))^2 \right\} \quad (10)$$

¹In the manifold assumption, the classification function is assumed to be smooth with respect to the underlying marginal data distribution, which is estimated using labeled and unlabeled data.

where γ_I controls the complexity of the function h according the intrinsic geometry of the data. Comparing (9) and (10), we can see that the difference with the graph-based SSL problem is the introduction of the last regularization term based on the nearest neighbor graph.

Instead of solving (10) directly, we will modify (or *deform*) the original kernel \mathbf{K} in order to adapt it to the underlying distribution geometry, defining a new kernel $\tilde{\mathbf{K}}$. The new problem to be solved becomes

$$\tilde{f} = \arg \min_{h \in \tilde{\mathcal{H}}} \left\{ \frac{1}{l} \sum_{i=1}^l V(h, \mathbf{x}_i, y_i) + \gamma_A \|h\|_{\tilde{\mathcal{H}}}^2 \right\}. \quad (11)$$

We can now see that (9) and (11) are the same, but solved with different kernels, and thus in different RKHS. The solution of (11) is equally obtained as a kernel expansion using only labeled samples

$$\tilde{f}(\mathbf{x}) = \sum_{i=1}^l \alpha_i \tilde{K}(\mathbf{x}, \mathbf{x}_i). \quad (12)$$

In order to deform \mathbf{K} and adapt it to the geometry of the underlying distribution, we first define $\tilde{\mathcal{H}}$ as RKHS. Let \mathcal{V} be a linear space with positive semidefinite inner product, and let $S : \mathcal{H} \rightarrow \mathcal{V}$ be a bounded linear operator. Defining $\tilde{\mathcal{H}}$ to be the space with the same functions as \mathcal{H} , but modified inner product $\langle f, g \rangle_{\tilde{\mathcal{H}}} = \langle f, g \rangle_{\mathcal{H}} + \langle Sf, Sg \rangle_{\mathcal{V}}$, makes $\tilde{\mathcal{H}}$ a valid RKHS (interested readers can see the proof in [11]).

In the specific considered problem, it is needed that S and \mathcal{V} depend on the data. Therefore, let \mathcal{V} be \mathbb{R}^N , and define S as the evaluation map $S(f) \equiv \mathbf{f} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_n))$. Thus, using a symmetric semidefinite matrix \mathbf{M} , the seminorm in \mathbb{R}^N can be written as $\|Sf\|_{\mathcal{V}}^2 = \mathbf{f}^T \mathbf{M} \mathbf{f}$. With such a norm, the regularization problem in (11) becomes

$$\tilde{f} = \arg \min_{h \in \tilde{\mathcal{H}}} \left\{ \frac{1}{l} \sum_{i=1}^l V(h, \mathbf{x}_i, y_i) + \gamma_A (\|h\|_{\tilde{\mathcal{H}}}^2 + \mathbf{h}^T \mathbf{M} \mathbf{h}) \right\} \quad (13)$$

where \mathbf{h} includes both labeled and unlabeled data and the matrix \mathbf{M} encodes smoothness w.r.t. the graph or manifold. Comparison between (10) and (13) reveals that (13) solves, in fact, an SSL problem, but with the advantage that it can be indirectly set out using (11) and solved using (12).

The geometry of the data is included through \mathbf{M} , defined proportional to the graph Laplacian \mathbf{L} , i.e., $\mathbf{M} = \gamma \mathbf{L}$, where $\gamma \in [0, \infty)$ is a free parameter that controls the “deformation” of the kernel. To incorporate the geometry information of the manifold, we build a graph G using labeled and unlabeled pixels. The graph Laplacian of G is a matrix defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where \mathbf{W} is the adjacency matrix, whose elements w_{ij} are a measure of the similarity between pixels \mathbf{x}_i and \mathbf{x}_j calculated using a distance function, and \mathbf{D} is a diagonal matrix whose elements are defined as $d_{ii} = \sum_{j=1}^n w_{ij}$, and they represent the degree of each one of the n vertices of the graph G .

The graph Laplacian \mathbf{L} measures the variation of the function \mathbf{f} along the graph built upon all labeled and unlabeled samples. It is worth noting that, by fixing $\gamma = 0$, the original

TABLE I
SUMMARY OF THE MAIN CHARACTERISTICS OF THE FOUR DATA SETS USED IN THE EXPERIMENTS.
BACKGROUND PIXELS ARE UNKNOWN OR INVALID AND CANNOT BE USED IN THE EXPERIMENTS

Dataset →	Naples	KSC	MERIS L1b	Mexico
Image size	400 × 400	512 × 614	2399 × 2241	360 × 512
Background pixels	18094	309157	1955629	0
Features used	10	176	6	2
Classes	2	13	2	2
Total target pixels	34651	431	313780	29506
Total outlier pixels	107255	4780	3106750	154814
Training targets (l)	60	50	200	453
Training outliers (o)	0	50	200	0
Training unlabeled (u)	1000	1000	1000	969

(undeformed) kernel is obtained. Therefore, in the limit case, the deformed kernel leads to the same results as the unmodified one.

The S²OC-SVM is obtained by replacing the nondeformed kernel in (3) with the modified one whose elements can be computed as

$$\tilde{K}(\mathbf{x}_i, \mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j) - \mathbf{K}_{\mathbf{x}_i}^\top (\mathbf{I} + \mathbf{M}\mathbf{K})^{-1} \mathbf{M}\mathbf{K}_{\mathbf{x}_j} \quad (14)$$

where \mathbf{K} is the complete kernel matrix formed with the l labeled plus the u unlabeled samples, $\mathbf{K}_{\mathbf{x}_i} = [K(\mathbf{x}_1, \mathbf{x}_i), \dots, K(\mathbf{x}_n, \mathbf{x}_i)]^\top$, and \mathbf{I} is the identity matrix. The deformed kernel fulfills Mercer’s conditions, being a valid kernel [11]. Moreover, any kernel-based method can easily incorporate information provided by unlabeled samples with this deformation mechanism.

Solving the new problem is computationally equivalent to the original supervised one. However, two more free parameters need to be adjusted: the number of neighbors in the graph Laplacian k and the amount of deformation γ . Moreover, the new kernel (14) has to be computed, which implies a matrix inversion of size $(l + u) \times (l + u)$. Note that this inversion scales exponentially with the number of samples. Therefore, one pays the cost of including more unlabeled samples to better model the data marginal distribution. To stem this problem, the number of unlabeled samples should be kept as small as possible. However, when dealing with nonrepresentative training samples, a significant amount of unlabeled samples are needed to improve classification accuracy. In order to effectively handle the matrix inversion problem, it is possible to use techniques that avoid the direct calculation of the inverse matrix, such as the Nyström method [17] or the incomplete Cholesky factorization [18].

IV. b-SVM

The second proposed one-class classifier is the b-SVM. This classifier was originally proposed in [13] for text classification using only target and unlabeled data. The roots of the b-SVM are in [19], where it is stated that if the sample size is large enough, minimizing the number of unlabeled samples classified as *targets* while correctly classifying the labeled target samples will give an accurate classifier.

Let us consider again the data set $\{\mathbf{x}_i\}_{i=1}^{l+u} \in \mathbb{R}^N$ made up of l labeled samples and u unlabeled samples ($n = l + u$). The l labeled samples all belong to the same class: the target

class. Concerning unlabeled samples, they are treated by the algorithm as being all outliers, although their class is unknown. This characteristic makes b-SVM a sort of semisupervised classifier. Under these assumptions, the b-SVM formulation is defined as

$$\min_{\omega, \xi} \left\{ \frac{1}{2} \|\omega\|^2 + C_t \sum_{i=1}^l \xi_i + C_o \sum_{i=l+1}^{l+u} \xi_i \right\} \quad \forall i = 1, \dots, n \quad (15)$$

subject to

$$\langle \omega, \phi(\mathbf{x}_i) \rangle \geq 1 - \xi_i \quad (16)$$

where $\xi_i \geq 0$ are slack variables and C_t and C_o are the costs assigned to errors on target and outlier (unlabeled) classes, respectively. The two cost values should be adjusted to fulfill the goal of classifying the target class correctly while, at the same time, trying to minimize the number of unlabeled samples classified as target class. To achieve this goal, C_t should have a large value, as initially we trust our labeled training set, and C_o should have a small value, because it is unknown whether the unlabeled samples are actually targets or outliers.

V. EXPERIMENTAL RESULTS

This section presents the experimental results obtained with the proposed S²OC-SVM and b-SVM and compares them against two classical one-class methods and, when outlier samples are available in the data sets, also against the standard binary SVM.

A. Data Description

Experiments in multisource (optical and radar) urban monitoring, hyperspectral crop detection, multispectral cloud screening, and multispectral change detection illustrate the capabilities of the proposed methods in different remote sensing applications. In the following, a description of the four data sets is given, and Table I summarizes their main characteristics.

1) *Multisource Urban Monitoring*: For the first set of experiments, we used images from the Urban Expansion Monitoring project [20], [21]. The considered test site was Naples (Italy), where images from ERS2 SAR and Landsat Thematic Mapper (TM) sensors were acquired in 1999 (Fig. 2). The available features were seven TM bands, two SAR backscattering



Fig. 2. RGB image composition of the Naples image using Landsat bands 3, 2, and 1. The red square surrounds the 400×400 area used in our experiments.

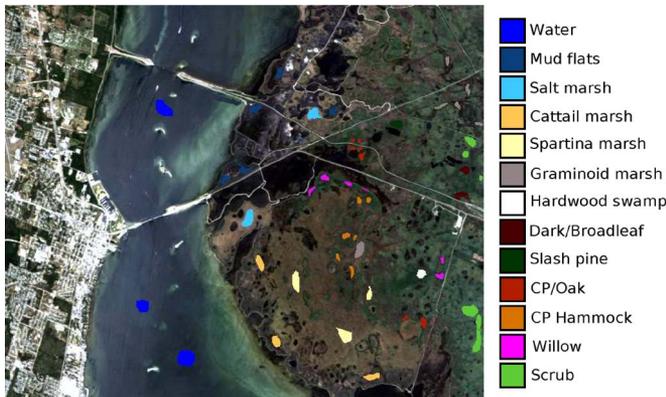


Fig. 3. RGB composition of the data acquired over the KSC by the NASA AVIRIS instrument. Thirteen classes were identified and labeled; they are superimposed in the image.

intensities (0–35 days), and the SAR interferometric coherence. Since these features come from different sensors, the first step was to perform a specific processing and conditioning of optical and SAR data and to coregister all images [20], [21]. After preprocessing, all features were stacked at a pixel level.

2) *Hyperspectral Crop Detection*: The second data set comes from the National Aeronautics and Space Administration (NASA) AVIRIS instrument that acquired data over the Kennedy Space Center (KSC), Florida, U.S., in 1996. The AVIRIS sensor acquires 224 bands of 10-nm width with center wavelengths from 400 to 2500 nm. The data were acquired from an altitude of 20 km and have a spatial resolution of 18 m. After removing low SNR and water absorption bands, a total of 176 bands remained for further analysis. The data set has 13 classes representing the various land cover types present in the KSC area. The reader is referred to <http://www.csr.utexas.edu/> for further detail on this data set. Fig. 3 shows an RGB composition with the labeled classes highlighted. For our experiments, class “Graminoid marsh” was selected as the target class, whereas the others are considered as outlier class. The interest in this class is motivated by its intrinsic complexity; it is underrepresented, and it can be confused with similar subclasses in the scene, as documented in [15].

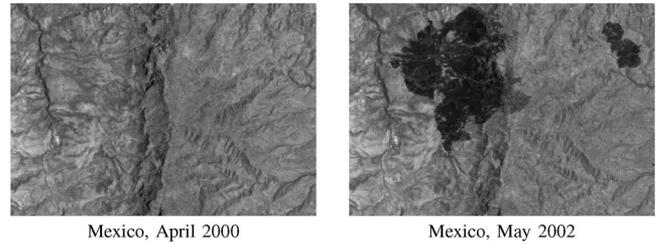


Fig. 4. Channel 4 of the Landsat-5 TM images acquired on the area of Mexico in (left) April 2000 and (right) May 2002. The darker area in the May 2002 image corresponds to the burnt area.

3) *Multispectral Cloud Screening*: The third data set is constituted by an ENVISAT MERIS Level 1b (L1b) image taken over France. The tested image was acquired on March 19, 2005. For our experiments, we used six physically inspired features extracted from the 15 MERIS spectral bands: cloud brightness and whiteness in the visible (VIS) and near-infrared (NIR) spectral ranges, an atmospheric oxygen feature, and a water vapor absorption feature [22].

4) *Change Detection*: The fourth data set is made up of a section (512×512 pixels) of two multispectral images acquired by the TM sensor of the Landsat-5 satellite in an area of Mexico in April 2000 and May 2002 (Fig. 4). Between the two aforementioned dates, a fire destroyed a large portion of the vegetation in the considered region. A detailed visual analysis of both the available multitemporal images and the multispectral difference image [23] allowed us to produce a reference map containing 25 599 changed and 236 545 unchanged pixels for test purposes. As the class of changed pixels is the most relevant from the application point of view and it usually shows a small prior probability, it was selected as the target class. Here, the multispectral difference image obtained subtracting the two dates according to the change vector analysis technique [23] was used as input to the considered classifiers.

B. Classifier Development and Experimental Setup

In this paper, the results obtained with S^2OC -SVM and b-SVM are compared with the ones achieved by two reference one-class classifiers: 1) the one-class Gaussian domain descriptor (Gaussian DD) [24], which is a standard and simple one-class classification algorithm, and 2) the OC-SVM, which is the base of the proposed S^2OC -SVM. Moreover, as for two of the data sets (KSC and MERIS L1b), some outlier samples are available; also, the standard binary SVM is trained for comparison.

In all the experiments, we trained the kernel-based methods using the RBF Gaussian kernel. For the S^2OC -SVM, the graph Laplacian L consisted of $l + u$ nodes connected using k nearest neighbors, and the edge weights W_{ij} were computed using the Euclidean distance among samples. The graph Laplacian deforms the kernel through (14), with $M = \gamma L$. Thus, the free parameter γ tunes the amount of kernel deformation. For the b-SVM, C_t and C_o were adjusted always such that $C_t > C_o$. Table II summarizes the tested algorithms along with their free parameters and their meaning, search ranges, and steps.

TABLE II
FREE PARAMETERS, THEIR MEANING, SEARCH RANGES, AND STEPS FOR THE TESTED METHODS

Method	Free parameter	Range	Meaning
Gaussian DD	ν	$\{0.01, \dots, 0.1\}$, in 0.01 steps	Expected percentage of outliers in the training set
	R	$\{0.1, \dots, 0.9\}$, in 0.1 steps	Regularization parameter for the estimated covariance matrix
OC-SVM	σ	$\{10^{-2}, \dots, 10^2\}$, in powers of $\sqrt{10}$ steps	Width of the RBF Gaussian kernel
	ν	$\{0.01, \dots, 0.1\}$, in 0.01 steps	Expected percentage of outliers in the training set
S ² OC-SVM	σ	$\{10^{-2}, \dots, 10^2\}$, in power of $\sqrt{10}$ steps	Width of the RBF Gaussian kernel
	ν	$\{0.01, \dots, 0.1\}$, in 0.01 steps	Expected percentage of outliers in the training set
	γ	$\{10^{-2}, \dots, 10^3\}$, in power of 10 steps	Amount of introduced kernel deformation
	k	$\{2, \dots, 5\}$	Number of neighbors used to computed the graph Laplacian
SVM	σ	$\{10^{-2}, \dots, 10^2\}$, in power of $\sqrt{10}$ steps	Width of the RBF Gaussian kernel
	C	$\{10^{-1}, \dots, 10^3\}$, in power of $\sqrt{10}$ steps	Regularization parameter to deal with errors
bSVM	σ	$\{10^{-2}, \dots, 10^2\}$, in power of $\sqrt{10}$ steps	Width of the RBF Gaussian kernel
	C_o	$\{10^{-3}, \dots, 10^1\}$, in power of $\sqrt{10}$ steps	Regularization parameter accounting for outlier class errors
	C_t	$C_o \times \{10, 100, 1000\}$	Regularization parameter accounting for target class errors

Supervised one-class classifiers were trained according to the following three-step procedure.

- 1) *Free-parameter search.* This search is done using a four-fold cross-validation strategy on the training set². In this step, only labeled samples for the target class are used to tune the free parameters of the classifier (when available, also labeled samples for outlier class can be used).
- 2) *Classifier learning.* Using the optimal set of free parameters found in the previous step, the final classifier is trained using all the available labeled target samples in the training set.
- 3) *Classifier testing.* The trained classifier is used to label the whole image (except the training samples), the confusion matrix is computed, and the *kappa statistic* (κ) is estimated [25].

When semisupervised one-class classifiers are considered, steps 1) and 2) also involve unlabeled samples.

In general, free parameters are hard to adjust when only target labeled samples are available in the training set. The problem is that, in such situations, only the true positive rate measure (sensitivity) can be computed, while the other error counterpart (specificity) cannot. In these cases, the free-parameter selection was carried out by evaluating $\arg \max_{\theta} \{OA[\%]/\#SV\}$, where θ is the whole set of free parameters, OA is the overall accuracy (the percentage of pixels correctly assigned), and $\#SV$ is the number of support vectors³. This bound of performance enforces high overall accuracy while limiting model complexity keeping a low number of support vectors [5].

On the other hand, in the data sets where some labeled outlier samples were available for validation, the kappa statistic was used to measure the error of the classifier in the training set and to select the best free parameters.

²In v -fold cross-validation, the training set is split in v disjoint groups: $v - 1$ sets are used for training and the remaining one for validation. The procedure is repeated v times. The best combination of free parameters is chosen by minimizing an average error measurement computed with the predictions on the v different validation sets.

³Except for the Gaussian DD algorithm, for which no way to measure the model complexity is available, and thus, just the overall accuracy was used.

C. Results

In this section, quantitative and qualitative results are shown for the four considered data sets. In particular, attention is devoted to the performance of the S²OC-SVM and b-SVM methods in different remote sensing ill-posed and high-uncertainty conditions.

All the experiments were carried out using ten random realizations of labeled and unlabeled training sets. The kappa statistic is estimated from the confusion matrix for each realization, and as a final result, the average kappa statistic on the ten realizations is provided together with its standard deviation. Table I summarizes the number of labeled l , unlabeled u , and outlier o samples used in each experiment.

1) *Multisource Urban Monitoring:* The Naples data set constitutes a complex example of urban monitoring using multisensor data (optical and SAR data). The experiments were carried out using only labeled samples of the target (*urban*) class. No labeled samples of the outlier class were used in the training process, which makes it a very difficult task.

Here, the attention is payed to the complex scenario of working with only $l = 60$ labeled samples for the “urban” class using $u = 1000$ unlabeled samples. The first four rows of Table III show the results for this data set. The average kappa statistic on the test set for the OC-SVM was 0.8, while the S²OC-SVM improved results to 0.84, and the b-SVM obtained 0.85. For this data set, the Gaussian DD has a low performance, obtaining an average κ of 0.66. These results show that the proposed methods can improve the results obtained with standard one-class methods. In general, semisupervised methods obtain better results in all ten realizations, even in those (for instance, fifth realization) where standard one-class methods fail. Fig. 5(a) shows the best classification maps, obtained in the sixth realization.

2) *Hyperspectral Crop Detection:* The KSC hyperspectral data set constitutes a challenging problem of target detection due to the characteristics of the “*Graminoid marsh*” class selected as target. In this experiment, 50 labeled samples were used, and a small number of outlier labeled samples ($o = 50$) were involved in the free-parameter selection. For the semisupervised classifiers, $u = 1000$ unlabeled samples were used.

The one-class classifiers Gaussian DD and OC-SVM obtained an average κ of 0.77 and 0.58, respectively. The proposed

TABLE III
RESULTS FOR ALL DATA SETS AND CLASSIFIERS. THE KAPPA STATISTIC IS PRESENTED FOR THE TEN REALIZATIONS.
MEAN AND STANDARD DEVIATIONS ARE PRESENTED IN THE LAST COLUMN

Dataset	Method	r1	r2	r3	r4	r5	r6	r7	r8	r9	r10	κ mean (std)
Naples	Gaussian DD	0.65	0.68	0.63	0.60	0.55	0.69	0.67	0.79	0.79	0.60	0.66 (0.08)
	b-SVM	0.83	0.85	0.86	0.85	0.84	0.89	0.87	0.80	0.85	0.83	0.85 (0.03)
	OC-SVM	0.85	0.82	0.85	0.83	0.58	0.79	0.84	0.84	0.85	0.80	0.80 (0.08)
	S ² OC-SVM	0.87	0.82	0.85	0.83	0.84	0.84	0.84	0.84	0.85	0.84	0.84 (0.01)
KSC	Gaussian DD	0.86	0.77	0.80	0.80	0.75	0.89	0.79	0.79	0.61	0.66	0.77 (0.08)
	SVM	0.50	0.46	0.65	0.52	0.66	0.40	0.49	0.50	0.69	0.56	0.54 (0.09)
	b-SVM	0.63	0.71	0.62	0.76	0.66	0.64	0.72	0.67	0.72	0.73	0.69 (0.05)
	OC-SVM	0.67	0.32	0.62	0.58	0.62	0.59	0.58	0.60	0.59	0.65	0.58 (0.10)
MERIS L1b	S ² OC-SVM	0.68	0.71	0.65	0.58	0.62	0.64	0.64	0.60	0.65	0.66	0.64 (0.04)
	Gaussian DD	0.74	0.85	0.77	0.77	0.79	0.59	0.71	0.79	0.76	0.82	0.76 (0.07)
	SVM	0.87	0.83	0.85	0.83	0.85	0.90	0.91	0.84	0.84	0.94	0.87 (0.04)
	b-SVM	0.88	0.85	0.77	0.87	0.89	0.84	0.89	0.84	0.85	0.88	0.86 (0.04)
Mexico	OC-SVM	0.58	0.49	0.51	0.50	0.86	0.50	0.86	0.86	0.85	0.86	0.69 (0.18)
	S ² OC-SVM	0.76	0.81	0.71	0.72	0.87	0.72	0.87	0.86	0.86	0.88	0.81 (0.07)
	Gaussian DD	0.79	0.86	0.86	0.89	0.88	0.87	0.88	0.89	0.88	0.90	0.87 (0.03)
	b-SVM	0.87	0.88	0.78	0.83	0.88	0.89	0.83	0.78	0.85	0.71	0.83 (0.06)
Mexico	OC-SVM	0.89	0.84	0.87	0.88	0.88	0.88	0.62	0.88	0.87	0.88	0.85 (0.08)
	S ² OC-SVM	0.89	0.88	0.89	0.88	0.88	0.88	0.63	0.88	0.87	0.89	0.86 (0.08)

semisupervised S²OC-SVM improved the result obtained by the OC-SVM, obtaining 0.64 for the averaged κ , but did not improve the Gaussian DD classifier that worked well for this particular data set. The b-SVM outperformed all classifiers, except the Gaussian DD, obtaining an average κ of 0.69.

As outliers were available for this data set, a binary SVM was also trained and tested, which resulted in an average κ of 0.58. This result confirms that, in multiclass problems where the user is interested in only one class, one-class classifiers can obtain better results than traditional supervised (binary) classifiers. Fig. 5(b) shows the classification maps obtained in realization 7 with the algorithms.

3) *Multispectral Cloud Screening*: In this experiment, we aim to detect cloudy areas in a MERIS L1b multispectral image. Cloud screening is specially well suited to semisupervised approaches since cloud features change to a great extent depending on the cloud type, thickness, transparency, height, and background, making extremely difficult the definition of a representative training set. Cloud screening is carried out before atmospheric corrections (i.e., with the data being affected by the atmospheric conditions), thus increasing the solution complexity. Moreover, in this particular case, the presence of snow, which is easily confused with clouds, further increases the problem complexity.

Here, 200 labeled target samples were used to build a reliable OC-SVM, 200 samples of cloud-free areas (outliers) were involved to properly adjust free parameters, and 1000 unlabeled samples were used for SSL. Fig. 5(c) shows the RGB composite image of the area. Note that only the upper left and the bottom right parts of the image are covered by clouds, while the remaining white areas are snow covered. As can be deduced from the low κ values in Table III, standard one-class classifiers are unable to distinguish clouds from snow, whereas semisupervised classifiers discriminate much better and successfully reject most of the snow areas (higher κ coefficient). In this experiment, the proposed S²OC-SVM semisupervised method significantly improves the OC-SVM classifier in all realizations. This is also clear when observing maps in Fig. 5(c).

4) *Change Detection*: In the last experiment, the goal is to detect the location of changed areas, assuming that no ground truth information is available, as what usually happens in change-detection applications. An initial set of labeled samples having a high probability to be correctly assigned to the classes of changed pixels is detected by exploiting the intrinsic unsupervised properties of the change vector analysis technique following [26]. Nearly certain changed patterns show a magnitude greater than $T + \delta$, where T is the decision threshold that separates changed from unchanged patterns and is selected according to the Bayes decision rule for minimum error [27] and δ is a positive constant that allows one to exclude uncertain pixels (the ones with magnitude too close to T) from the training set. Patterns with magnitude smaller than $T + \delta$ are considered as unlabeled.

For this particular problem, $l = 453$ and $u = 969$. Different from [28], here, outliers (unchanged pixels) were not used in the training process. As one can see from Table III, the S²OC-SVM performs better than the OC-SVM in most of the cases. The improvement introduced by the S²OC-SVM is small due to the simple structure of the considered problem and the low dimensionality of the input space. These results suggest that the proposed method is better suited for high-dimensional problems. However, S²OC-SVM significantly reduced the impact of false detected targets on the change-detection map. The b-SVM obtains an average κ value that is slightly lower for this data set, but on the other hand, we can see that its results are more uniform and consistent across all runs, suggesting less dependence on the training set than the OC-SVM or the S²OC-SVM. Fig. 5(d) shows the reference and classification maps obtained for realization 3.

VI. DISCUSSION

In the experiments, the proposed S²OC-SVM and b-SVM obtained more uniform (i.e., less noisy) classification cover maps and significantly reduced the number of false positives. Comparing both methods, in general, the b-SVM worked

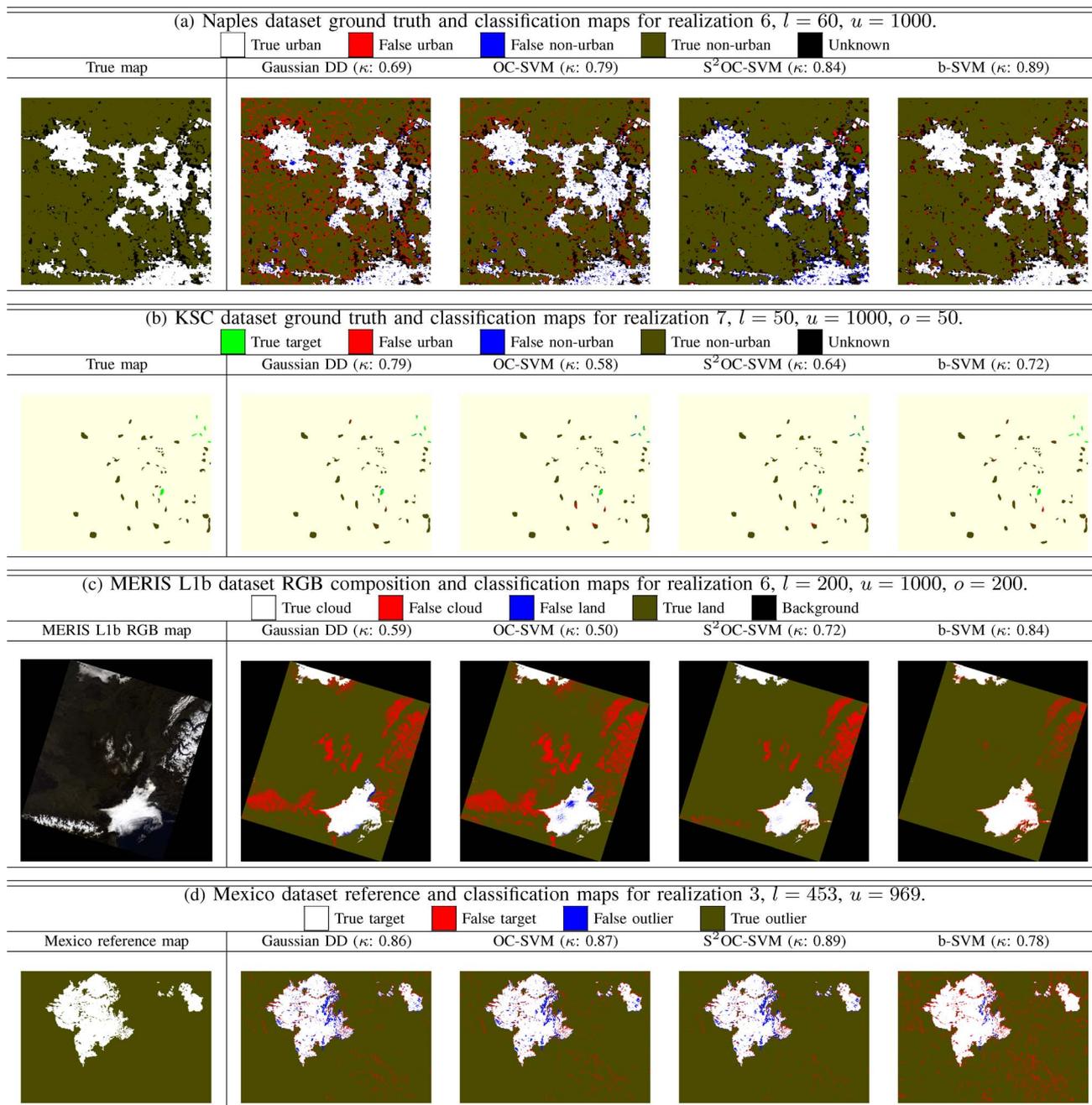


Fig. 5. True and classification maps for all data sets obtained with Gaussian DD, OC-SVM, S²OC-SVM, and b-SVM for some specific realizations. (a) Naples data set ground truth and classification maps for realization 6 ($l = 60$ and $u = 1000$). (b) KSC data set ground truth and classification maps for realization 7 ($l = 50$, $u = 1000$, and $o = 50$). (c) MERIS L1b data set RGB composition and classification maps for realization 6 ($l = 200$, $u = 1000$, and $o = 200$). (d) Mexico data set reference and classification maps for realization 3 ($l = 453$ and $u = 969$).

slightly better, reducing not only the number of false positives but also the number of false negatives.

From the results in Table III, one of the advantages of the presented S²OC-SVM method is that it will always perform, at least, as well as the OC-SVM: This is the case, for example, in KSC realization 8, where the S²OC-SVM optimal kernel was the original OC-SVM kernel (the best value for γ , controlling the amount of kernel deformation, was zero).

The Gaussian DD obtained the best average κ values for two data sets, KSC and Mexico. In both data sets, these good results reveal the Gaussian underlying distribution of the target

class. Therefore, the Gaussian DD, despite being simple, is a clear candidate in situations where a Gaussian distribution is suspected. It also has the advantage of being fast, because only the covariance matrix has to be estimated. On the other hand, when the target class does not follow a Gaussian distribution, or it is underrepresented by the training set, this classifier obtains poor results. This is the case for the Naples and MERIS L1b data sets.

For comparison purposes, in the KSC and MERIS L1b data sets, where outlier samples were available, the standard binary SVM classifier was also trained. In the case of the

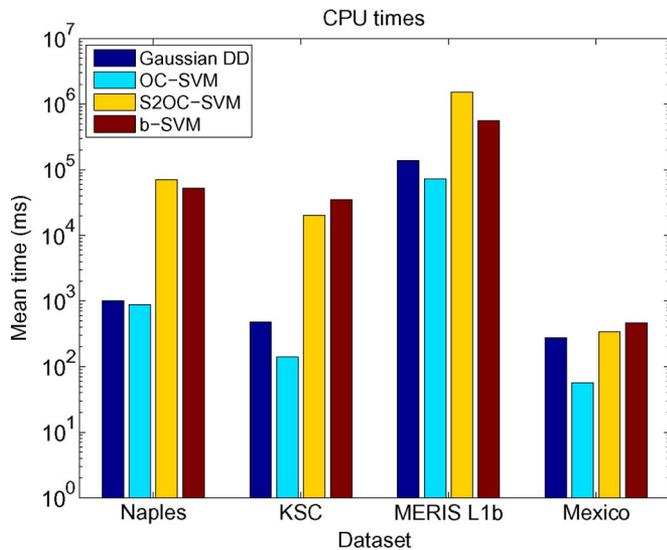


Fig. 6. Mean CPU times of the algorithms measured using Matlab `cputime` function.

KSC, the SVM obtained worst results than the one-class or semisupervised methods. This result was expected, taking into account that the SVM tries, in this case, to model all classes being nontarget (12 in total) as one single class. On the other hand, one-class methods focus in obtaining a good model for the selected class of interest. For the MERIS L1b data set, the SVM obtained the best average κ value. In this case, given that this was more clearly a binary classification problem (cloudy versus cloud free), the SVM has the advantage, over one-class and semisupervised methods, of having, in the training phase, information about both classes.

As a final analysis, the different algorithms were compared in terms of computational cost (here, the `cputime` Matlab function was used to evaluate it). All the algorithms run under the Matlab environment. For the Gaussian DD, we used the `dd_tools` toolbox [29]. The rest of the algorithms were implemented using the LibSVM [30]. As can be seen from Fig. 6, standard one-class methods took a second or less to train the models and make predictions, whereas S²OC-SVM and bSVM methods needed up to 100 times more to carry out the same process. This was an expected result, given that the semisupervised methods have to deal with many more pixels than the standard one-class methods. However, depending on the complexity of the considered problem, the gain in accuracy obtained by semisupervised methods justifies the higher computational cost.

VII. CONCLUSION

Two methods for one-class classification and target detection for remote sensing problems that take advantage of unlabeled sample information for improving classification accuracy have been presented and evaluated. The S²OC-SVM method works by taking as core learner the OC-SVM and uses the graph Laplacian to incorporate the unlabeled information into the kernel machine by deforming the supervised kernel matrix. The b-SVM method treats all unlabeled samples as outliers and uses different weights for each class (target and outlier) in

the training process. Four real one-class classification scenarios have been evaluated: urban monitoring, target detection, cloud masking, and change detection. The obtained results reveal the potential of the proposed methods. It has been shown that the presented methodology can improve results by exploiting the information contained in the unlabeled data, particularly when the information in the labeled samples is poor or lacks representativeness. Further work is tied to analyze the improvement of semisupervised classifiers as a function of the unlabeled samples and input dimensionality.

REFERENCES

- [1] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [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] B. Schölkopf, R. C. Williamson, A. Smola, and J. Shawe-Taylor, "Support vector method for novelty detection," in *Proc. Adv. Neural Inf. Process. Syst.*, Denver, CO, 1999, vol. 12, pp. 582–588.
- [4] G. Mercier and F. Girard-Ardhuin, "Partially supervised oil-slick detection by SAR imagery using kernel expansion," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, pt. 1, no. 10, pp. 2839–2846, Oct. 2006.
- [5] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2282–2291, Aug. 2006.
- [6] J. Muñoz-Marí, L. Bruzzone, and G. Camps-Valls, "A support vector domain description approach to supervised classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 8, pp. 2683–2692, Aug. 2007.
- [7] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, M. Martínez-Ramón, and J. L. Rojo-Álvarez, "Kernel-based framework for multi-temporal and multi-source remote sensing data classification and change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1822–1835, Jun. 2008.
- [8] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning*, 1st ed. Cambridge, MA: MIT Press, 2006.
- [9] C. Elkan and K. Noto, "Learning classifiers from only positive and unlabeled data," in *Proc. Int. Conf. Knowl. Discovery Data Mining*, 2008, pp. 213–220.
- [10] C. Scott and G. Blanchard, "Novelty detection: Unlabeled data definitely help," in *Proc. Int. Conf. Artif. Intell. Statist., JMLR*, 2009, vol. 5, pp. 464–471.
- [11] V. Sindhwani, P. Niyogi, and M. Belkin, "Beyond the point cloud: From transductive to semi-supervised learning," in *Proc. 22nd ICML*, 2005, pp. 1–6.
- [12] L. Gómez-Chova, G. Camps-Valls, J. Muñoz-Marí, and J. Calpe-Maravilla, "Semisupervised image classification with Laplacian support vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 336–340, Jul. 2008.
- [13] B. Liu, Y. Dai, X. Lia, W. S. Lee, and P. S. Yu, "Building text classifiers using positive and unlabeled examples," in *Proc. Int. Conf. Data Mining*, 2003, pp. 179–186.
- [14] D. Tax and R. P. Duin, "Support vector domain description," *Pattern Recognit. Lett.*, vol. 20, no. 11–13, pp. 1191–1199, Nov. 1999.
- [15] L. Capobianco, A. Garzelli, and G. Camps-Valls, "Target detection with semisupervised kernel orthogonal subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3822–3833, Nov. 2009.
- [16] B. Schölkopf and A. Smola, *Learning With Kernels. Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, MA: MIT Press, 2002.
- [17] C. K. I. Williams and M. Seeger, "Using the Nyström method to speed up kernel machines," in *Proc. NIPS*, T. K. Leen, T. G. Dietterich, and V. Tresp, Eds., Vancouver, BC, Canada, Dec. 2001, vol. 13, pp. 682–688.
- [18] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
- [19] B. Liu, W. S. Lee, P. S. Yu, and X. Li, "Partially supervised classification of text documents," in *Proc. ICML*, 2002, pp. 387–394.
- [20] L. Gómez-Chova, D. Fernández-Prieto, J. Calpe, E. Soria, J. Vila-Francés, and G. Camps-Valls, "Urban monitoring using multitemporal SAR and multispectral data," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 234–243, Mar. 2006.

[21] L. Gómez-Chova, D. Fernández-Prieto, J. Calpe, E. Soria, J. Vila-Francés, and G. Camps-Valls, "Urban monitoring using multitemporal SAR and multispectral data," in *Proc. 3rd Pattern Recogn. Remote Sens. Workshop*, Kingston Upon Thames, U.K., Aug. 27, 2004.

[22] L. Gómez-Chova, G. Camps-Valls, J. Calpe, L. Guanter, and J. Moreno, "Cloud screening algorithm for ENVISAT/MERIS multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, pt. 2, no. 12, pp. 4105–4118, Dec. 2007.

[23] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in polar domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 218–236, Jan. 2007.

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

[25] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, FL: Lewis Publishers, 1999.

[26] F. Bovolo, L. Bruzzone, and M. Marconcini, "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2070–2082, Jul. 2008.

[27] L. Bruzzone and D. Fernandez Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.

[28] F. Bovolo, G. Camps-Valls, and L. Bruzzone, "A support vector domain method for change detection in multitemporal images," *Pattern Recognit. Lett.*, 2010, to be published.

[29] D. Tax, Ddtools, the data description toolbox for Matlab, Dec. 2009. [Online]. Available: http://ict.ewi.tudelft.nl/~davidt/dd_tools.html

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



Luis Gómez-Chova (S'08–M'09) received the B.Sc. (first-class honors), M.Sc., and Ph.D. degrees in electronic engineering from the Universitat de València, València, Spain, in 2002 and 2008, respectively.

Since 2000, he has been with the Department of Electronic Engineering, Universitat de València, where he was first enjoying a research scholarship from the Spanish Ministry of Education and is currently an Associate Professor. He is also a Researcher with the Imaging Processing Laboratory. His work is mainly related with pattern recognition and machine

learning applied to remote sensing multispectral images and cloud screening. He conducts and supervises research on these topics within the frameworks of several national and international projects. He is the author (or coauthor) of 25 international journal papers, more than 80 international conference papers, and several international book chapters.

Dr. Gómez-Chova was awarded by the Spanish Ministry of Education with the National Award for Electronic Engineering.



Lorenzo Bruzzone (S'95–M'98–SM'03–F'10) received the Laurea (M.S.) degree in electronic engineering (*summa cum laude*) and the Ph.D. degree in telecommunications from the University of Genoa, Genova, Italy, in 1993 and 1998, respectively.

He is currently a full Professor of telecommunications with the University of Trento, Trento, Italy, where he teaches remote sensing, pattern recognition, radar, and electrical communications. He is the Head of the Remote Sensing Laboratory, Department of Information Engineering and Computer Science,

University of Trento. His current research interests are in the areas of remote sensing, signal processing, and pattern recognition (analysis of multitemporal images, feature extraction and selection, classification, regression and estimation, data fusion, and machine learning). He conducts and supervises research on these topics within the frameworks of several national and international projects. He is the author (or coauthor) of 90 scientific publications in refereed international journals (61 in IEEE journals), more than 140 papers in conference proceedings, and 13 book chapters. He is an Editor/Coeditor of ten books/conference proceedings. He is a referee for many international journals.

Dr. Bruzzone is a member of the Managing Committee of the Italian Inter-University Consortium on Telecommunications and a member of the Scientific Committee of the India-Italy Center for Advanced Research. Since 2009, he has been a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society. He has served on the scientific committees of several international conferences.



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 electronic engineering, and the Ph.D. degree in physics from the Universitat de València, València, Spain, in 1996, 1998, and 2002, respectively.

He is currently a full Professor with the Department of Electronic Engineering, Universitat de València, where he teaches electronics, advanced time series processing, and machine learning for remote sensing. His research interests are tied to the develop-

ment of machine learning algorithms for signal and image processing with special focus on remote sensing data analysis. He conducts and supervises research within the frameworks of several national and international projects, and he is an Evaluator of project proposals and scientific organizations. He is the author (or coauthor) of 70 international journal papers, more than 80 international conference papers, and several international book chapters. He is the Editor of the books "Kernel methods in bioengineering, signal and image processing" (IGI, 2007) and "Kernel methods for remote sensing data analysis" (Wiley & Sons, 2009). He is a referee for many international journals and conferences.

Dr. Camps-Valls has been a member of the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society since 2007. Since 2009, he has been a member of the Machine Learning for Signal Processing (MLSP) Technical Committee of the IEEE Signal Processing Society. He currently serves on the Program Committees of the International Society for Optical Engineers (SPIE) Europe, MLSP, and International Conference on Image Processing.



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

He is currently an Associate Professor with the Department of Electronic Engineering, Universitat de València, where he teaches analysis of circuits and linear systems, introduction to programmable logical devices, digital electronic systems, and micro-

processor electronic systems. His research interests include the development of machine learning algorithms for signal and image processing.



Francesca Bovolo (S'05–M'07) received the Laurea (B.S.) degree in telecommunication engineering, the Laurea Specialistica (M.S.) degree in telecommunication engineering (*summa cum laude*), and the Ph.D. degree in communication and information technologies from the University of Trento, Trento, Italy, in 2001, 2003, and 2006, respectively.

She is currently a Post-Doc Member of the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento. Her main research activity is in the area of remote-sensing image processing. In particular, her research interests are related to multitemporal remote-sensing image analysis and change detection in multispectral and SAR images and very high resolution images. She conducts research on these topics within the frameworks of several national and international projects. She is a referee for the *International Journal of Remote Sensing*, *Pattern Recognition*, *Pattern Recognition Letters*, *Remote Sensing of Environment*, *Photogrammetric Engineering and Remote Sensing*, *Photogrammetry and Remote Sensing*, and *Sensors*.

Dr. Bovolo is also a referee for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, and IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS.