A Novel Approach to Unsupervised Change Detection Based on a Semisupervised SVM and a Similarity Measure

Francesca Bovolo, *Member, IEEE*, Lorenzo Bruzzone, *Senior Member, IEEE*, and Mattia Marconcini, *Student Member, IEEE*

***Abstract*—This paper presents a novel approach to unsupervised change detection in multispectral remote-sensing images. The proposed approach aims at extracting the change information by jointly analyzing the spectral channels of multitemporal images in the original feature space without any training data. This is accomplished by using a selective Bayesian thresholding for deriving a pseudotraining set that is necessary for initializing an adequately defined binary semisupervised support vector machine (S^3VM) classifier. Starting from these initial seeds, the S^3VM performs change detection in the original multitemporal feature space by gradually considering unlabeled patterns in the definition of the decision boundary between changed and unchanged pixels according to a semisupervised learning algorithm. This algorithm models the full complexity of the change-detection problem, which is only partially represented from the seed pixels included in the pseudotraining set. The values of the classifier parameters are then defined according to a novel unsupervised model-selection technique based on a similarity measure between change-detection maps obtained with different settings. Experimental results obtained on different multispectral remote-sensing images confirm the effectiveness of the proposed approach.**

***Index Terms*—Bayesian thresholding, change vector analysis (CVA), multispectral images, multitemporal images, remote sensing, semisupervised support vector machine (S^3VM), unsupervised change detection.**

I. Introduction

THE EVER increasing number of operational satellites for Earth observation results in a growing interest of the remote-sensing community in the analysis of images acquired on the same geographical area at different times. Depending on the specific problem addressed, the analysis of multitemporal data can be carried out according to supervised classification techniques (e.g., for producing thematic maps or maps of land-cover transitions) or to unsupervised change-detection procedures (e.g., for generating change-detection maps associated with damages caused by natural disasters or with land-cover modifications due to anthropic phenomena) [1]–[3]. The methods adopted for data analysis also depend on the kind of images considered (e.g., multispectral or SAR) and on the related properties. In the aforementioned scenario, this paper focuses the attention on unsupervised change detection in multispectral images.

Several unsupervised change-detection methods for multispectral images acquired by passive sensors have been proposed in the remote-sensing literature [3]–[7]. Among them, a widely used technique is the change vector analysis (CVA) [3], [8]. The CVA technique computes a multidimensional difference image by subtracting the spectral feature vectors associated with each pair of corresponding pixels in two images acquired on the same scene at two different times. In the resulting image, each pixel is represented by a spectral change vector (SCV). Usually, the desired change-detection map is obtained by thresholding the magnitude of SCVs [3], [8], [9]. A major drawback of the CVA technique applied to the magnitude of SCVs is that the difference and the magnitude operators are not bijective (one-to-one) and result in a loss of information with respect to the original multitemporal and multispectral feature space. Nonetheless, if unsupervised change detection should be applied (i.e., no training data are available), the CVA technique allows one to establish a relatively simple criterion (based on thresholding) for identifying the changed patterns. If a pixel has magnitude value that is lower than a given threshold, then it can be associated with the class of unchanged pixels; else, it is labeled as changed.[1] The aforementioned criterion is more difficult to use for performing change detection in the original multidimensional feature space, where an unsupervised analysis would require the application of clustering algorithms to multispectral vectors. However, in real applications, the complexity of the data affects the change-detection map obtained with clustering, which, in many cases, results in change-detection accuracies that are lower than those provided by thresholding the magnitude of SCVs. In addition, clustering techniques require a further manual postprocessing for associating the identified clusters with changed or unchanged labels. This increases the level of required supervision and decreases the level of automation. In this context, other effective transformation methods (alternative to the magnitude of SCVs) have been proposed in the literature for projecting the information of original data into subspaces suitable to change

Manuscript received October 3, 2007; revised January 4, 2008. This work was supported by the Italian Ministry of Education, University and Research.

The authors are with the Department of Information Engineering and Computer Science, 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.2008.916643

[1]The direction information of the SCVs can be used for distinguishing different kinds of changes [3], [8].

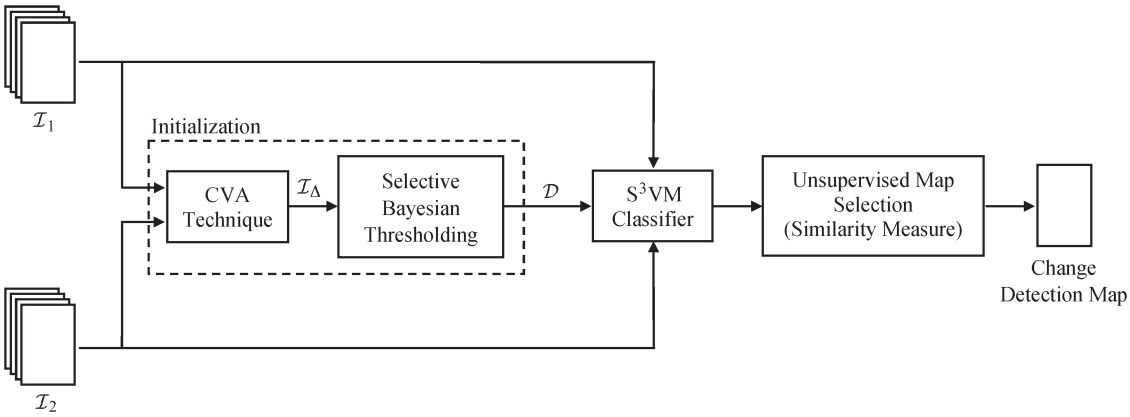


Fig. 1. Block scheme of the proposed approach.

detection [10], [11]. However, in all cases, we have to face a tradeoff between exploitation of all the information present in the original multitemporal feature space and possibility to define efficient unsupervised and automatic criteria for the generation of change-detection maps.

In this paper, we address the aforementioned problem by proposing a method for unsupervised change detection in multispectral images, which has the following properties: i) it performs change detection directly on the original spectral channels of multitemporal images, thus exploiting all the available information for increasing the accuracy of the process; ii) it is unsupervised (no training data are required); and iii) it exhibits an intrinsic robustness to the noise affecting multitemporal images (e.g., residual registration noise [12], [13]). The rationale of the proposed method is to exploit the intrinsic unsupervised properties of the CVA technique for deriving an initial set of seed pixels having a high probability to be correctly assigned to the classes of changed and unchanged pixels. These seeds are then jointly exploited with all the other unlabeled pixels of the images in a properly designed binary semisupervised support vector machine (S^{3}VM) classifier for solving the change-detection problem in the original multitemporal feature space. This classification technique models the complexity of the change-detection problem (which is only partially represented by the seed pixels), thanks to a semisupervised learning procedure that derives the decision boundary analyzing both the seed patterns and the structure of the unlabeled pixels. In order to obtain a completely unsupervised method, we also propose a novel unsupervised model-selection procedure for the selection of the S^{3}VM parameters, which is based on a similarity measure applied to the candidate change-detection maps. This procedure does not require any prior information.

The main novel contributions of this paper are as follows: i) the basic idea (which consists in the synergistic exploitation of both the simplicity associated with the thresholding of the magnitude of SCVs and the completeness of information contained in the original feature space) and the definition of the architecture of the change-detection approach; and ii) the procedure proposed for unsupervised model selection in the S^{3}VM.

This paper is organized into four sections. The next section presents the proposed change-detection approach, detailing the architecture of the method, as well as its single components. Section III describes the experimental setup and reports the results obtained by applying the proposed technique to two different remote-sensing data sets. Finally, Section IV draws the conclusion of this paper.

II. Proposed Change-Detection Approach

Let $\mathcal{I}_1$ and $\mathcal{I}_2$ be two coregistered multispectral images of size $P \times Q$ acquired on the same geographical area at different times t_1 and t_2, respectively. Let D be the number of spectral channels of each image, and let $\Omega = \{\omega_u, \omega_c\}$ be the set of classes of unchanged and changed pixels to be identified. The proposed technique consists of three main parts (see Fig. 1): i) an initialization that exploits a selective Bayesian thresholding of the magnitude of SCVs for generating, in an unsupervised way, a set of seed pixels (pseudotraining set); ii) a properly designed S^{3}VM classifier that exploits the pseudotraining set and the unlabeled pixels of images $\mathcal{I}_1$ and $\mathcal{I}_2$ for producing the change-detection map by analyzing the original multidimensional feature space; and iii) a novel similarity measure that is applied to the change-detection maps obtained with different values for the S^{3}VM parameters for deriving, in an unsupervised way, the final change-detection result. These three parts are described in detail in the following.

A. Bayesian Initialization

Let $\mathcal{I} = \{\mathbf{x}_p\}_{p=1}^{P \times Q}$ be the set of $P \times Q$ d-dimensional vectors (where $d = 2 \times D$) whose components are the channels associated with the generic pth pixel in $\mathcal{I}_1$ and $\mathcal{I}_2$. $\mathcal{I}$ represents the multitemporal data set under analysis achieved by stacking the coregistered multispectral images $\mathcal{I}_1$ and $\mathcal{I}_2$.

The first step of the proposed unsupervised approach to change detection aims at identifying a pseudotraining set $\mathcal{D} = \{\mathcal{X}_l, \mathcal{Y}_l\}$ made up of pairs $(\mathbf{x}_n^l, y_n^l)$ to be used as seed patterns for initializing the S^{3}VM, where $\mathcal{X}_l = \{\mathbf{x}_n^l, \mathbf{x}_n^l \in \mathcal{I}\}$ and $\mathcal{Y}_l = \{y_n^l, y_n^l \in \Omega\}$ ($\mathcal{Y}_l$ is the set of pseudolabels corresponding to seed patterns in $\mathcal{X}_l$). From a theoretical point of view, this subset should contain patterns that are associated either to changed or unchanged areas on the ground with no uncertainty. As we detect these patterns in an automatic and

unsupervised way, we accept to relax the ideal assumption with the more realistic constraint that patterns included in the subset $\mathcal{X}_l$ are associated with a high probability to belong to changed or unchanged areas. Accordingly, we propose to identify $\mathcal{X}_l$ by applying the CVA technique to $\mathcal{I}_1$ and $\mathcal{I}_2$ and by selectively thresholding the magnitude $\mathcal{I}_\Delta$ of SCVs. Any unsupervised threshold-selection method presented in the change-detection literature can be used for identifying the threshold value T, which separates changed from unchanged patterns [9], [14]–[16]. Among them, we propose the use of unsupervised threshold-selection techniques based on the Bayesian decision theory, which showed to be effective in change detection on multispectral images [8], [9], [15]. The application of these techniques requires the explicit estimation of the statistical parameters of classes of changed and unchanged pixels (i.e., the class-prior and class-conditional probabilities). As we are dealing with an unsupervised change-detection problem, these statistical quantities are estimated from the observed statistical distribution of the magnitude of the SCVs $p(i^\Delta)$ (where i^Δ is the random variable associated with the magnitude of the SCVs in $\mathcal{I}_\Delta$) according to the expectation–maximization (EM) algorithm [9]. Depending on the kind of data considered, the estimation can be carried out according to the following models: i) a parametric model based on the Gaussian [9] or the generalized Gaussian [16] functions; or ii) a semiparametric model based on a mixture of Gaussian distributions [15]. The estimated class-statistical parameters are then used with the Bayes decision rule for minimum error for identifying the decision threshold T that separates changed from unchanged patterns. However, if we apply the Bayesian threshold to $\mathcal{I}_\Delta$, we obtain a change-detection map affected by errors resulting from the uncertainty that characterizes pixels with a magnitude value that is close to the threshold. This problem is partially due to the loss of information associated with the difference and magnitude operators, which do not allow to exploit all the information of the original feature space in the change-detection process. On the contrary, the threshold value T represents a reasonable reference point for identifying the subset $\mathcal{X}_l$. According to this observation, the desired set of pixels with a high probability to be correctly assigned to one of the two classes is obtained by defining an *uncertainty* region around T. This region conceptually identifies uncertain patterns and separates them from nearly certain pixels. Therefore, $\mathcal{X}_l$ is defined as

$$\mathcal{X}_l = \left\{ \mathbf{x}_n^l | i_n^\Delta \leq T - \delta_1 \quad \text{and} \quad i_n^\Delta \geq T + \delta_2 \right\}_{n=1}^{P \times Q} \tag{1}$$

where i_n^Δ is the magnitude of the nth SCV in $\mathcal{I}_\Delta$ and where δ_1 and δ_2 are positive constants that tune the left and right boundaries, respectively, of the margin of uncertain pixels around the threshold value T (see Fig. 2). δ_1 and δ_2 should be selected in order to obtain a high probability that patterns in $\mathcal{X}_l$ have a correct label. It is worth noting that, in general, the margin can be approximated as symmetric with respect to the threshold; thus, we can assume $\delta_1 = \delta_2 = \delta$. A reasonable strategy for selecting the value of δ is to relate it to the dynamic range of the difference image. Although more complex strategies could be defined, taking into account that the selection of δ does not affect significantly the final change-detection results, an empirical rule of thumb is to set its value equal to the 15% of the dynamic range of $\mathcal{I}_\Delta$.

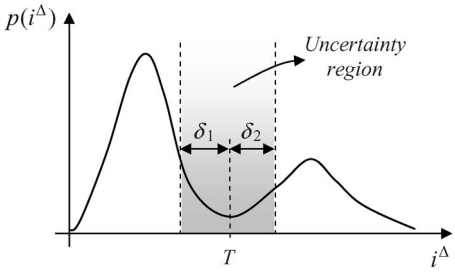


Fig. 2. Example of distribution of the magnitude of SCVs $p(i^\Delta)$ and of definition of the uncertainty region.

According to the properties of SCVs, pseudolabels of pixels in $\mathcal{X}_l$ are assigned as follows:

$$y_n^l = \begin{cases} \omega_u, & \text{if } i_n^\Delta \leq T - \delta_1 \\ \omega_c, & \text{if } i_n^\Delta \geq T + \delta_2 \, . \end{cases} \tag{2}$$

The set $\mathcal{X}_u = \{\mathbf{x}_m^u, \mathbf{x}_m^u \in \mathcal{I} \backslash \mathcal{X}_l\}$ of unlabeled patterns complementary to $\mathcal{X}_l$ includes all the pixels that fall inside the margin and that are thus associated with a high degree of uncertainty. These pixels should be analyzed in the original feature space.

B. Change Detection Based on S^{3}VM

The main idea of the second step of the proposed technique is to define a discriminant function in the original multitemporal feature space of the multitemporal images, which can accurately separate changed pixels from unchanged ones. In this way, it is possible to integrate the incomplete seed information extracted in the previous phase (on the basis of the behavior of the magnitude of SCVs) with the complete information associated with the uncertain unlabeled pixels in the original multitemporal images. In the presence of a reliable training set, the aforementioned problem could be easily solved according to any binary classification technique applied to the multitemporal feature space. However, in our case, we have a pseudotraining set which is made up of samples that satisfy a tradeoff between high probability to be correctly labeled and complete representation of the problem. In other terms, the margin imposed in the thresholding procedure used in the first step results in a pseudotraining set that does not model completely the statistics of changed and unchanged classes. In particular, we expect that the pseudotraining set does not contain enough information for modeling the distribution of patterns that are close to the decision boundary (which is the most critical portion of the multitemporal feature space for obtaining an effective discriminant function and, thus, an accurate change-detection map).

In order to address this point, in the proposed system, we exploit a binary S^{3}VM classifier [17], which is an improvement to the technique (properly adapted to the considered problem) presented in [18]. The main idea is to use the properties of the semisupervised learning to tune the separation hyperplane between the class of changed and unchanged pixels (which, like in standard supervised SVM, are mapped into a

high-dimensional kernel space through a nonlinear mapping) by jointly exploiting the pseudotraining samples in $\mathcal{D}$ and unlabeled samples in $\mathcal{X}_u$. This is accomplished by positioning the hyperplane in regions of the kernel space having a low density of samples (*cluster assumption*[2]). In other words, the available labeled samples play the role of an inertial term in the definition of the decision boundary, whereas the unlabeled pixels allow one to tune the discriminant function for better modeling the data distribution according to the cluster assumption.

Note that, depending on the size of the considered images, the cardinality of $\mathcal{X}_l$ and $\mathcal{X}_u$ can be high. To decrease the computational load, it is reasonable to randomly subsample both $\mathcal{X}_l$ and $\mathcal{X}_u$. The subsampling should be carried out in a uniform way in order to reach a tradeoff between a statistically significant representation of the classes and a reasonable computational load in the learning of the system. Let us assume that, after subsampling, the pseudotraining set $\mathcal{D} = \{\mathcal{X}_l, \mathcal{Y}_l\}$ is made up of N pairs $(\mathbf{x}_n^l, y_n^l)$ of seed patterns, i.e., $\mathcal{X}_l = \{\mathbf{x}_n^l\}_{n=1}^N$ and $\mathcal{Y}_l = \{y_n^l\}_{n=1}^N$, and $\mathcal{X}_u$ is made up of M unlabeled patterns $\mathcal{X}_u = \{\mathbf{x}_n^u\}_{m=1}^M$. The S^{3}VM algorithm performs an initial inductive learning (similar to that of standard supervised SVMs) [19], [20] on the subsampled pseudotraining set $\mathcal{D} = \{\mathcal{X}_l, \mathcal{Y}_l\}$ and then applies an iterative semisupervised strategy by gradually considering unlabeled samples $(\mathcal{X}_u)$. In this way, the initial position of the separation hyperplane in the kernel space is defined according to seed patterns, whereas its final position (at the convergence of the iterative learning algorithm) is tuned on the basis of both the seeds and the unlabeled patterns in the multitemporal images. From a formal viewpoint, the proposed S^{3}VM technique is based on the following two main phases: 1) *Initialization* and 2) *Semisupervised Learning*. The main modifications with respect to the algorithm presented in [18] concern the *Semisupervised Learning* phase and, in particular, the following: i) the criterion adopted for the iterative selection of the unlabeled patterns that are most suitable to be labeled correctly and ii) the weighting strategy employed for the regularization parameter associated with the semilabeled samples. These modifications, which result both in a slight decrease of computational load and in an increase of stability of the learning procedure, will be better explained in the following.

Phase 1—Initialization: Let $\mathcal{D}^{(i)} = \{\mathcal{X}^{(i)}, \mathcal{Y}^{(i)}\}$ and $\mathcal{X}^{*(i)}$ denote the pseudotraining set and the unlabeled set (i.e., the set containing all the samples that are not associated with a label yet) at the generic iteration i, respectively. The first phase corresponds to the initial step of the entire process $(i = 0)$. We have that $\mathcal{D}^{(0)} = \{\mathcal{X}^{(0)}, \mathcal{Y}^{(0)}\} \equiv \{\mathcal{X}_l, \mathcal{Y}_l\}$ and $\mathcal{X}^{*(0)} \equiv \mathcal{X}_u$. The learning cost function of a standard supervised SVM is used to obtain an initial separation hyperplane based only on pseudotraining data $(\mathbf{x}_n^l, y_n^l)_{n=1}^N$. Let $\boldsymbol{\xi}^{l(0)} = \{\xi_1^{l(0)}, \ldots, \xi_N^{l(0)}\}$ be the vector of the slack variables associated with the patterns of $\mathcal{X}_l$. Accordingly, the optimization problem to be solved is the following:

$$\min_{\mathbf{w}^{(0)}, b^{(0)}, \boldsymbol{\xi}^{l(0)}} \left\{ \frac{1}{2} \left\| \mathbf{w}^{(0)} \right\|^2 + C \sum_{n=1}^{N} \xi_n^{l(0)} \right\}$$
$$\text{s.t.} \quad y_n^{l(0)} \cdot \left[\mathbf{w}^{(0)} \cdot \Phi\left(\mathbf{x}_n^l\right) + b^{(0)} \right] \geq 1 - \xi_n^{l(0)}$$
$$\xi_n^{l(0)} \geq 0 \quad \text{and} \quad n = 1, \ldots, N \tag{3}$$

where $\mathbf{w}^{(0)}$ is the vector that is normal to the separation hyperplane, $b^{(0)}$ is the bias of the separation hyperplane, C is the regularization parameter, and $\Phi(\mathbf{x})$ is a nonlinear mapping function [19]. According to standard SVM notation, the labels ω_u and ω_c are represented with values "+1" and "−1," respectively (i.e., $y_n^l \in \{\omega_c, \omega_u\} \equiv \{+1, -1\}$).

Phase 2—Semisupervised Learning: The second phase of the proposed S^{3}VM starts with iteration $i = 1$ and represents the core of the algorithm. At the generic iteration i, pseudolabels y^u are given to unlabeled pixels belonging to $\mathcal{X}^{*(i)} \subseteq \mathcal{X}_u$ according to the current separation hyperplane. These pixels are called pseudolabeled patterns. As support vectors (i.e., the samples of the current training set $\mathcal{D}^{(i)}$ belonging to the margin $\mathcal{M} = \{\mathbf{x} || \mathbf{w}^{(i)} \cdot \Phi(\mathbf{x}) + b^{(i)}| \leq 1, \mathbf{x} \in \mathbb{R}^d\}$) are the only patterns that affect the position of the discriminant hyperplane, unlabeled samples which fall into the margin and are closest to the margin bounds have the highest probability to be correctly classified. Accordingly, the ρ pseudolabeled samples lying into the margin (where, unlike in [18], $\rho \geq 1$ is defined *a priori* by the user) that are closest either to the lower or the upper margin bound are selected and denoted as semilabeled patterns [see Fig. 3(a) and (b)]. The set containing all the semilabeled samples defined at iteration i is called $\mathcal{H}^{(i)}$. Patterns of $\mathcal{H}^{(i)}$ and their corresponding semilabels are then merged with $\mathcal{X}^{(i)}$ and $\mathcal{Y}^{(i)}$, respectively.

Let $\mathcal{J}^{(i)}$ represent the set of all the pixels selected from $\mathcal{X}_u$, which have been always assigned the same label until iteration i. A dynamical adjustment is necessary for considering the modifications in the position of the separation hyperplane. Let $\mathcal{S}^{(i)}$, shown at the bottom of the page, represent the set of samples belonging to $\mathcal{J}^{(i-1)}$ whose labels obtained according to the separation hyperplane at iteration i are different than those at iteration $i-1$ (*label inconsistency*). Patterns belonging to $\mathcal{S}^{(i)}$ are reset to the unlabeled state and moved again into $\mathcal{X}^{*(i)}$. In this way, it is possible to reconsider these patterns at the following iterations. Therefore, we have

$$\mathcal{J}^{(i)} = \left(\mathcal{J}^{(i-1)} - \mathcal{S}^{(i)}\right) \cup \mathcal{H}^{(i)}$$
$$\mathcal{X}^{*(i)} = \left(\mathcal{X}^{*(i-1)} - \mathcal{H}^{(i-1)}\right) \cup \mathcal{S}^{(i-1)}. \tag{5}$$

[2]The cluster assumption states that each cluster of samples belongs to one class. Thus, the decision boundary is defined between clusters, i.e., in low-density regions of the feature space.

$$\mathcal{S}^{(i)} = \begin{cases} \varnothing, & i = 0, 1 \\ \left\{\mathbf{x}^u | (\mathbf{x}^u \in \mathcal{X}^{(i)}, x^u \in \mathcal{X}^{(i-1)}), y^{u(i)} \neq y^{u(i-1)}\right\}, & i > 1 \end{cases} \tag{4}$$

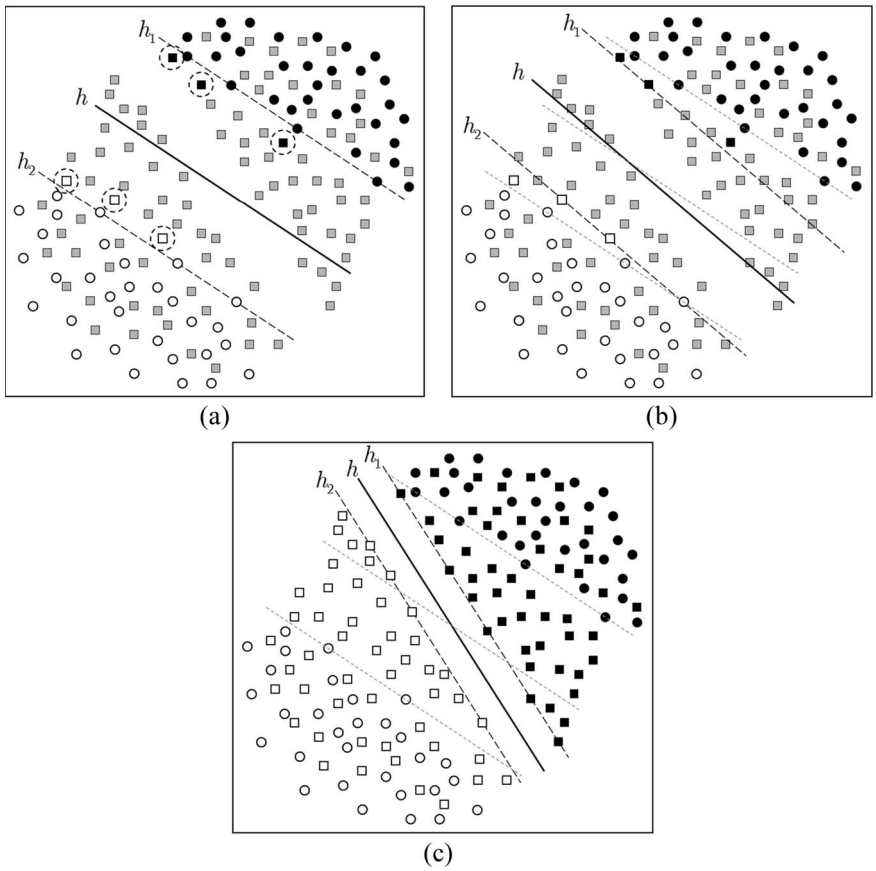


Fig. 3. Separation hyperplane (solid line) and margin bounds (dashed lines) resulting at different stages of the S^{3}VM algorithm for a simulated change-detection problem. Patterns in the pseudotraining set are shown as white and black circles. Corresponding semilabeled patterns are shown as white and black squares, respectively. Unlabeled patterns are represented as gray squares. Kernel space structure obtained: (a) at the first iteration (the dashed circles highlight the ρ semilabeled patterns selected from both sides of the margin; in the example, $\rho = 3$); (b) at the second iteration (the dashed gray lines represent both separation hyperplane and margin bounds at the beginning of the learning process); and (c) at the last iteration in an ideal situation (the dashed gray lines represent both the separation hyperplane and the margin bounds at the beginning of the learning process).

As it will be pointed out in the following, the S^{3}VM algorithm aims at gradually increasing the regularization parameter for the semilabeled patterns on the basis of a time-dependent criterion. Accordingly, $\mathcal{J}^{(i)}$ is partitioned into a finite number of subsets γ (where γ is a free parameter representing the maximum number of iterations for which the user allows the regularization parameter to increase). Each subset $\mathcal{J}_k^{(i)}$ includes all the samples that belong to $\mathcal{J}_{k-1}^{(i-1)}$ and are labeled in the same way after the tuning of the current separation hyperplane. The optimization problem to be solved becomes

$$\min_{\mathbf{w}^{(i)},b^{(i)},\boldsymbol{\xi}^{l(i)},\boldsymbol{\xi}^{u(i)}} \left\{ \frac{1}{2}\left\|\mathbf{w}^{(i)}\right\|^2 + C\sum_{n=1}^{N}\xi_n^{l(i)} + \sum_{m=1}^{\eta^{(i-1)}} C_m^*\xi_m^{u(i)} \right\}$$

$$\text{s.t.} \quad y_n^{l(i)} \cdot \left[\mathbf{w}^{(i)} \cdot \mathbf{\Phi}\left(\mathbf{x}_n^l\right) + b^{(i)}\right] \geq 1 - \xi_n^{l(i)}, \mathbf{x}_n^l \in \mathcal{X}_l$$

$$y_m^{u(i)} \cdot \left[\mathbf{w}^{(i)} \cdot \mathbf{\Phi}\left(\mathbf{x}_m^u\right) + b^{(i)}\right] \geq 1 - \xi_m^{u(i)}, \mathbf{x}_m^u \in \mathcal{J}^{(i-1)}$$

$$\xi_n^{l(i)}, \xi_m^{u(i)} \geq 0, \quad n = 1, \ldots, N, \quad m = 1, \ldots, \eta^{(i-1)} \tag{6}$$

where $\mathcal{X}^{(i)} = [\mathcal{X}^{(i-1)} \cup \mathcal{H}^{(i-1)}] - \mathcal{S}^{(i-1)}$ and $\eta^{(i)} = |\mathcal{J}^{(i)}|$. The semilabeled samples in the training set are associated with a regularization parameter $C_m^* = C_m^*(k) \in \mathbb{R}^+$. The purpose of C_m^* is to control the number of misclassified samples of the current training set $\mathcal{D}^{(i)}$ that originally belonged to the unlabeled set. On increasing their values, the penalty associated with errors increases. As the statistical distribution of semilabeled patterns can be different compared with that of the original training data, they should be considered gradually in the learning process in order to avoid instabilities. For this reason, C_m^* increases in a quadratic way, depending on the number of iterations that $\mathbf{x}_m^u$ had last inside $\mathcal{J}^{(i)}$. The following updating rule is adopted (see Fig. 4):

$$C_m^* = \frac{C^{*\max} - C^{*(0)}}{(\gamma-1)^2}(k-1)^2 + C^{*(0)} \Leftrightarrow \left(\mathbf{x}_m^u \in \mathcal{J}_k^{(i-1)}\right) \qquad \forall m = 1, \ldots, \eta^{(i-1)} \tag{7}$$

where $C^{*(0)}$ is the initial regularization value for semilabeled samples (this is a user-defined parameter that must be much smaller than the C value as it grows quadratically with the number of algorithm iterations; a reasonable choice has empirically proved to be $C^{*(0)} \simeq 0.01 \cdot C$), k is the index corresponding to the number of iterations for which the semilabeled sample $\mathbf{x}_m^u$ is labeled in the same way since it has been included into the pseudotraining set, and $C^{*\max} = \tau \cdot C$, with $0 < \tau \leq 1$ (a reasonable choice has proved to be $\tau = 0.5$). It is worth

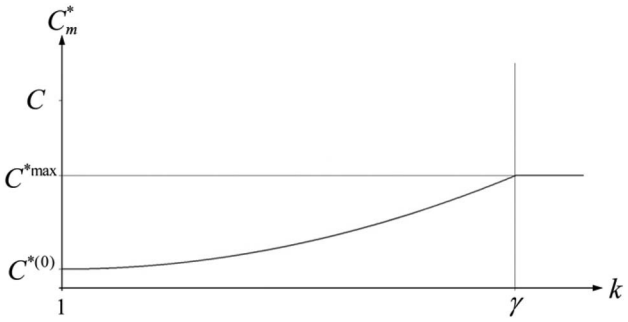


Fig. 4. Regularization parameter for the generic mth semilabeled sample $\mathbf{x}^u_m$ versus k (index corresponding to the number of iterations for which $\mathbf{x}^u_m$ is labeled in the same way since it has been included into the pseudotraining set).

noting that this rule is different from that presented in [18], where all the semilabeled samples were associated at each iteration with the same regularization parameter defined according to an absolute temporal criterion irrespectively on the iteration in which a given unlabeled sample was included in the learning procedure. The main advantage of this modification is that the more gradual and sample-dependent weighting strategy increases the robustness of the learning algorithm by avoiding possible instabilities associated with the algorithm in [18].

The semisupervised learning is iterated until convergence, which is obtained when none of the originally unlabeled samples lies into the margin band [see Fig. 3(c)]. In order to decrease the computational complexity, this criterion can be relaxed by accepting to have a small number of unlabeled samples in the margin. When the convergence is reached (i.e., $i = \text{end}$), a label $\hat{y}_p \in \{+1, -1\} \equiv \{\omega_c, \omega_u\}$ is assigned to all pixels $\mathbf{x}_p$ belonging to $\mathcal{I}$ according to

$$\hat{y}_p = \text{sgn}\left[\sum_{n=1}^{N} \alpha^l_n y^l_n K\left(\mathbf{x}^l_n, \mathbf{x}\right) + \sum_{m=1}^{\eta} \alpha^u_m y^u_m K\left(\mathbf{x}^u_m, \mathbf{x}\right) + b\right] \quad \mathbf{x}^l_n, \mathbf{x}^u_m \in \mathcal{X} \qquad (8)$$

where $y^u_m = y^{u(\text{end})}_m$, $\eta = \eta^{(\text{end})}$, and $\mathcal{X} = \mathcal{X}^{(\text{end})}$, whereas $\alpha^l_n = \alpha^{l(\text{end})}_n$ and $\alpha^u_m = \alpha^{u(\text{end})}_m$ are Lagrange multipliers that permit one to solve (6) in the dual formulation, and $K(\cdot,\cdot) = \Phi(\cdot) \cdot \Phi(\cdot)$ is a kernel function.[3] The reader is referred to [17]–[20] for more details on the kernel trick and kernel functions, as well as on their role and use in SVM.

Assigning labels according to (8) results in the definition of the desired change-detection map.

C. Novel Unsupervised Model-Selection Procedure Based on a Similarity Measure

In the design of the S^{3}VM architecture, it is necessary to select the values of the parameters of the considered kernel functions, as well as of ρ, C, C^*, and γ. This phase is called model selection of the classifier. Unfortunately, in the considered unsupervised change-detection problem, we do not have any labeled test data for carrying out the model-selection procedure. This issue deserves to be properly analyzed, as the model selection imposes some constraints on the analytical form of the decision boundary that is used for solving the change-detection problem. We expect that the choice of the model significantly affects the absolute accuracy of the final results. In the proposed architecture, we analyzed the following two different strategies: i) model selection based on seed pixels included in the pseudotraining set and ii) model selection based on a novel procedure that exploits both the pixels included in the pseudotraining set and a similarity measure between the change-detection results on unlabeled pixels. The first trivial strategy does not allow a fine tuning of the model (we expect that many different models can result in very similar accuracies on the change-detection problem defined from the seed pixels). This depends on the simplicity of the problem designed from pseudotraining patterns, which does not contain the most critical pixels that are close to the decision boundary.

The second strategy, which represents a novel methodological contribution of this paper, integrates the first one (requiring, as a basic constraint, a high accuracy on the seed pixels) with an analysis of the effects of different models on the change-detection map. The rationale of this strategy is to observe that: a) we expect that there exist different values of the parameters that result in proper change-detection maps; and b) values of the parameters that do not properly classify the unlabeled pixels in the margin result in different random spatial distributions of the wrong samples in the map (in other words, proper values of the parameters provide similar classifications of the samples in corresponding spatial positions). The basic idea is that correct solutions are more regular than wrong solutions.[4] In the proposed strategy, the most reliable solution (change-detection map) is selected according to the following criteria.

1) Among all the considered solutions, select those that result in an accuracy on the pseudotraining set $\mathcal{D}$ that is higher than a predefined value θ. We compute the accuracy in terms of kappa coefficient [22] to properly evaluate the tradeoff between missed and false alarms.[5] A reasonable choice is to set θ equal to the 90% of the maximum kappa accuracy obtained among all the considered solutions.
2) For all solutions that satisfy 1), evaluate the ratio between the number of detected changed and unchanged pixels. If this value differs significantly from the same quantity computed on the pseudotraining set (e.g., if the difference is higher than 30%), then discard the solution. This condition assures that unreliable solutions that strongly change the prior distributions of changed and unchanged pixels with respect to the pseudotraining set are rejected. An alternative formulation of this condition can be exploited if prior information on the expected balance between

[3]It is possible to prove that the learning problem addressed at each iteration of the algorithm results in a convex cost function and, thus, in the possibility to identify the optimal solution. Nonetheless, as in all transductive/semisupervised SVMs, at the end of the learning, we cannot assure to find a final solution that corresponds to the absolute minimum of the cost function for the considered semisupervised problem. Anyway, the convergence can be always obtained.

[4]It is worth noting that each solution is associated with a given set of values of the S^{3}VM parameters and, therefore, with a specific change-detection map.

[5]The kappa coefficient is computed by considering all the elements of the confusion matrix and usually assumes values in the [0, 1] range.

changed and unchanged areas is available. In this case, the aforementioned ratio can be related to the estimated ratio between changed and unchanged pixels rather than to the ratio obtained on the pseudotraining set. It is worth noting that the difference value could be in the range between 20% and 40% (depending on the flexibility that one would like to impose on this constraint); nonetheless, in general, we expect a minor impact of this choice on the final result of the model selection.

3) For a generic pair of solutions S_i and S_j $(i, j = 1, \ldots, N_S$ that satisfy both 1) and 2), compute a measure of similarity H_{ij} of the change-detection results on the $P \times Q$ pixels as follows:

$$H_{ij} = \frac{1}{P \times Q} \sum_{p=1}^{P \times Q} \hat{y}_p^i \cdot \hat{y}_p^j, \qquad H_{ij} \in [-1, +1] \tag{9}$$

where $\hat{y}_p^i$ and $\hat{y}_p^j$ are the estimated labels of the pth pixels in the change-detection maps associated with solutions S_i and S_j, respectively, computed according to (8). As $\hat{y}_p^i$ and $\hat{y}_p^j$ can assume values in $\{-1, +1\}$, their product is equal to 1 if $\hat{y}_p^i = \hat{y}_p^j$ and to -1 otherwise. Accordingly, the value of the similarity measure H_{ij} is equal to 1 if the two considered solutions S_i and S_j result in an identical change-detection map; otherwise, it is lower than 1. On the basis of this measure, we can define an absolute measure of similarity among a single solution S_i and all the others by computing the average value of H_{ij}, i.e.,

$$H_i = \frac{1}{N_s - 1} \sum_{j=1, j \neq i}^{N_s} H_{ij}, \qquad H_i \in [-1, +1]. \tag{10}$$

The solution (i.e., the change-detection map) that shows the highest average similarity with all the others is selected as the output of the proposed change-detection technique. This solution is the most regular among the considered ones (all associated with accuracy that is higher than θ on the pseudotraining set) and, thus [taking into account also conditions 1) and 2)], has the highest probability to be associated with a model that properly seizes the properties of the change-detection problem.

It is worth noting that the aforementioned criteria result in a reliable model selection under the following assumptions: i) the pseudotraining set contains samples that represent all kinds of changes present in the images (even if they only partially describe their distributions), and ii) all the three criteria are jointly considered. In general, we expect that the first assumption holds in real change-detection applications, as the EM-based thresholding algorithm used for defining the pseudotraining set proved to be reliable and accurate in many change-detection problems, also in the presence of different kinds of changes.

III. Experimental Results

In order to assess the effectiveness of the proposed approach, we considered two different data sets made up of multispectral images acquired by the Thematic Mapper (TM) scanner of the Landsat-5 satellite. The first data set refers to Elba Island (Italy), whereas the second one concerns an area in Mexico. In the following sections, we present the results obtained on these data sets.

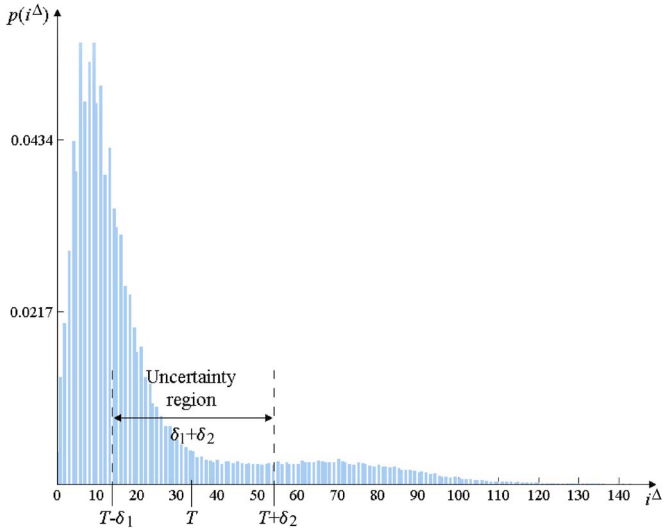


Fig. 5. Histogram of the magnitude of the difference image for the Mexico data set. The uncertainty region around the Bayesian decision threshold is highlighted.

A. Description of Experiments

In all experiments, we employed Gaussian kernel functions [18] for the S^3VM and a model-selection strategy based on the proposed similarity measure applied to a grid search in the following ranges:

1) number of pseudolabeled samples lying into the margin $\rho \in [5, 1000]$;
2) regularization parameter for samples in the pseudotraining $C \in [1, 700]$;
3) regularization parameter for semilabeled samples $C^{*(0)} \in [0.1, 1]$;
4) maximum number of iterations for which the user allows the C^* to increase $\gamma \in [5, 50]$;
5) Gaussian kernel width $2\sigma^2 \in [0.1, 1]$.

It is worth noting that we decided to apply a nearly exhaustive grid-search model selection in order to have a clear understanding of the potentialities of the method without any bias introduced by suboptimal search strategies. Nonetheless, like in standard supervised SVMs, any stochastic search algorithm could be used for speeding up the model selection (e.g., genetic algorithms with a fitness function defined in terms of kappa accuracy and similarity measure). In all the experiments, the threshold value T was derived according to the automatic algorithm proposed in [9], and the constant values $\delta_1 = \delta_2 = \delta$ (which identify the uncertainty region) were defined according to the empirical rule described in Section II-A. For the Elba Island data set, $\delta = 10$, whereas for the Mexico data set, $\delta = 20$. As an example, Fig. 5 shows the histogram of the magnitude of the difference image analyzed for the Mexico data set, where the threshold $T = 34$ and the uncertainty region are highlighted. It is worth noting that the estimation of δ_1

and δ_2 could be included in the model selection according to the grid-search strategy (or to more effective optimization algorithms). However, since the values of these parameters do not significantly affect the change-detection results (in our experiments, several trials were carried out with $\delta_1, \delta_2 \in [5, 30]$, obtaining very similar final results), we suggest to choose them on the basis of the statistical properties of the histogram of the difference image. With regard to the samples included in the pseudotraining set $\mathcal{D}$ and in the unlabeled sample set $\mathcal{X}_u$, in all our experiments, we randomly selected the 15% of patterns included in the two subsets after the initialization procedure. We also verified that increasing these numbers does not result in a significant change of the final results.

In order to assess the effectiveness of the proposed method, we compared the results obtained with those yielded by the standard CVA algorithm that was applied to the magnitude of SCVs according to a threshold derived exploiting class-statistical parameters that were computed with the EM algorithm. The choice of this comparison algorithm was carried out as it is very stable and is widely used in the literature [4], and, like the proposed method, is automatic and unsupervised. It is worth noting that, in this comparison, we did not consider context-sensitive methods [15], [21], as we are analyzing the effectiveness of the proposed technique at full resolution according to a pixel-based strategy. Nonetheless, it is possible to extend the proposed approach to a context-sensitive decision strategy. Moreover, in order to understand the effectiveness of the proposed unsupervised model-selection strategy for the S^3VM, we carried out experiments also performing the model selection on the basis of the real labels associated with pixels belonging to $\mathcal{X}_u$ (supervised model selection). Even if in practical applications no information is available about $\mathcal{X}_u$, with this kind of experiment, we aimed at determining an "upper bound" for the performances of the proposed technique and at comparing this upper bound with the results obtained with the proposed unsupervised similarity measure. A final analysis on the effectiveness of the presented approach was carried out by replacing, in the proposed architecture, the S^3VM with a standard SVM trained on the pseudotraining set and by applying both the presented (unsupervised) and the optimal (supervised) model selection. The results of this analysis allowed us to derive important information on the role of the semisupervised procedure in the proposed approach and to further assess the validity of the presented model-selection strategy.

The validation of the change-detection results in terms of false alarms, missed alarms, overall errors, and kappa coefficient of accuracy [22] was carried out according to the available reference maps concerning the location of changes both for the Elba Island and the Mexico data sets. It is worth noting that the reference maps were not used in the application of the proposed unsupervised approach, but they were considered only for validation purposes.

B. Results on the Elba Island Data Set

This data set is made up of a section $(414 \times 326$ pixels$)$ of three coregistered multispectral images acquired by the TM sensor on the western part of Elba Island, Italy. The three images were acquired in August 1992 $(\mathcal{I}_1)$, August 1994 $(\mathcal{I}_2)$, and September 1994 $(\mathcal{I}_3)$ [Fig. 6(a)–(c)]. Two wildfires occurred in 1993 and 1994 (between the acquisition dates of $\mathcal{I}_1$ and $\mathcal{I}_2$, and $\mathcal{I}_2$ and $\mathcal{I}_3$, respectively).

The available ground truth concerning the location of the two wildfires is shown in Fig. 6(d). The burned areas of the 1993 (black) and 1994 (gray) fires include 2842 and 2412 pixels, respectively. We carried out two kinds of experiments aimed at identifying the two burned areas. In the former, we considered the pair of images $(\mathcal{I}_2, \mathcal{I}_3)$; in the latter, we analyzed the pair of images $(\mathcal{I}_1, \mathcal{I}_2)$. For both pairs of images, only bands 4 and 7 were considered in the experimental analysis, as they proved to be the most effective for the detection of the burned areas [9], [15].[6] This selection is not required for the proposed technique (which can properly manage all the spectral channels), but it is important for the CVA technique with the EM-based thresholding, where the use of noisy channels may decrease the change-detection accuracy (adding ambiguity in the magnitude values and increasing the overlapping between the distributions of changed and unchanged classes).

In the first experiment, given the availability of a pair of images acquired few days before and after the 1994 fire, we obtained high change-detection accuracies [Table I(a)]. The proposed method sharply outperformed the standard CVA technique, increasing the value of the kappa coefficient of 0.187 (from 0.654 to 0.841) and decreasing false alarms from 1681 to 380 and missed alarms from 390 to 375. In addition, the presented unsupervised model-selection strategy based on the similarity measure applied to change-detection maps resulted in the choice of a solution that is rather close to the upper bound derived according to the supervised model selection (0.841 versus 0.890). Another interesting result is that, if in the proposed approach we replace the S^3VM with a conventional supervised SVM trained on the pseudotraining set (with both unsupervised and supervised model-selection strategies), we obtain kappa accuracies that are significantly smaller than those yielded with the S^3VM (i.e., 0.802 versus 0.890 in the upper bound case and 0.800 versus 0.841 with the unsupervised model selection). This confirms both the important role played from the S^3VM in the considered system and the effectiveness of the proposed unsupervised model-selection strategy also on the supervised SVM classifier.

The superiority of the proposed method with respect to the standard CVA can be assessed also from a qualitative viewpoint by comparing the change-detection maps shown in Fig. 7. As one can see, the map obtained with the proposed technique shows both a significantly lower amount of false alarms with respect to the map yielded with the CVA and a better identification of the burned area. In greater detail, it is possible to observe a significant robustness of the proposed method to the noise present between images, particularly to false alarms along the coastal areas and the ridges of the mountains (which are mainly associated with residual registration noise). This depends on the capability of the proposed method to discriminate the

[6]Note that, here, different from [9] and [15], no noise reduction filters were applied to the multitemporal images, as we expect that the proposed approach has a high capability to reduce the effects of noise (this will be explained later).

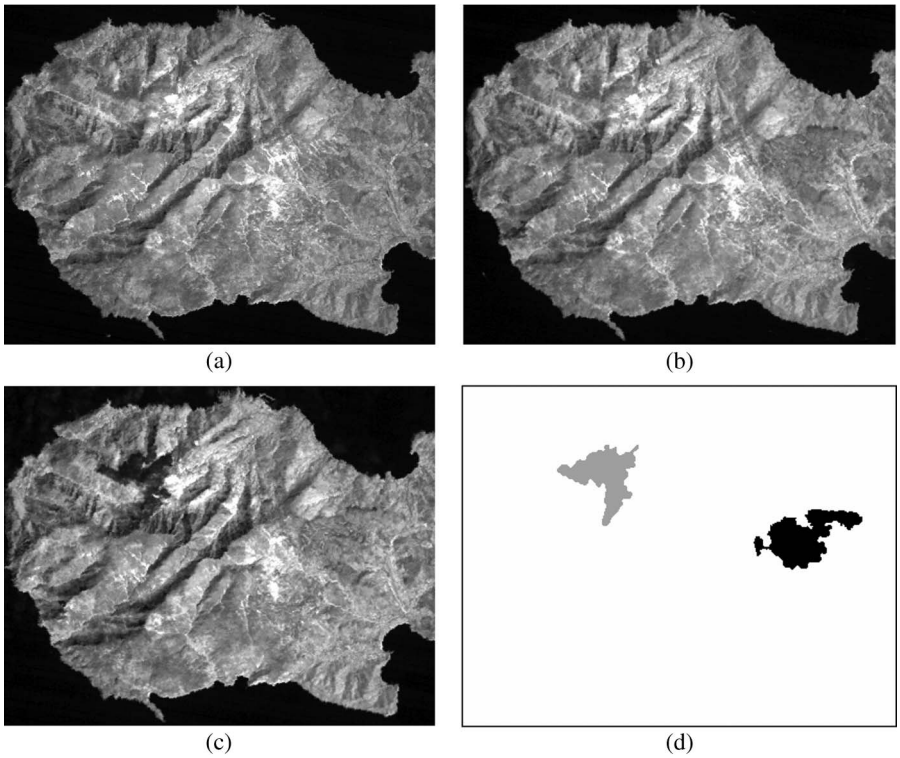


Fig. 6. Channel 4 of the Landsat-5 TM images acquired on the Elba Island (Italy) in (a) August 1992, (b) August 1994, and (c) September 1994. (d) Available reference map of the areas burned in 1993 (black) and 1994 (gray).

TABLE I
FALSE ALARMS, MISSED ALARMS, OVERALL ERRORS, AND KAPPA COEFFICIENTS OF ACCURACY FOR THE CHANGE-DETECTION MAPS OBTAINED FOR IDENTIFYING THE AREAS BURNED IN (a) 1994 AND (b) 1993 (ELBA ISLAND DATA SET)

Technique	Missed Alarms	False Alarms	Overall Error	Kappa Accuracy
CVA with EM-based thresholding [9]	390	1681	2071	0.654
Proposed with SVM (upper bound)	396	574	970	0.802
Proposed with SVM (unsupervised model selection)	385	604	989	0.800
Proposed with S^3VM (upper bound)	342	162	504	0.890
Proposed with S^3VM (unsupervised model selection)	375	380	755	0.841

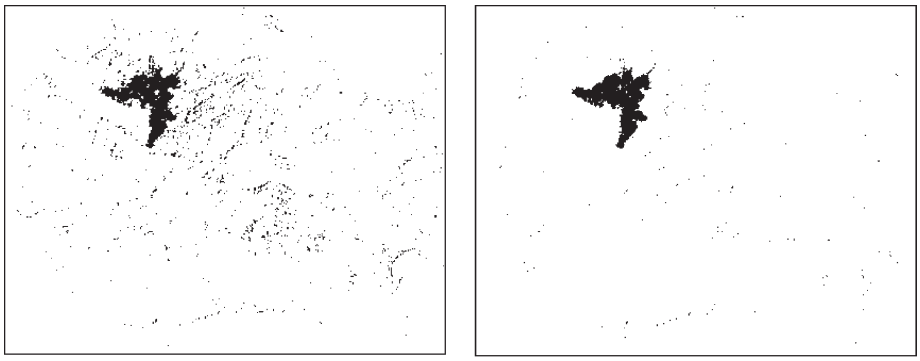
(a)

Technique	Missed Alarms	False Alarms	Overall Error	Kappa Accuracy
CVA with EM-based thresholding [9]	1270	3978	5248	0.357
Proposed with SVM (upper bound)	1189	1043	2232	0.5885
Proposed with SVM (unsupervised model selection)	1215	1081	2296	0.5776
Proposed with S^3VM (upper bound)	992	981	1973	0.645
Proposed with S^3VM (unsupervised model selection)	1103	1064	2167	0.608

(b)

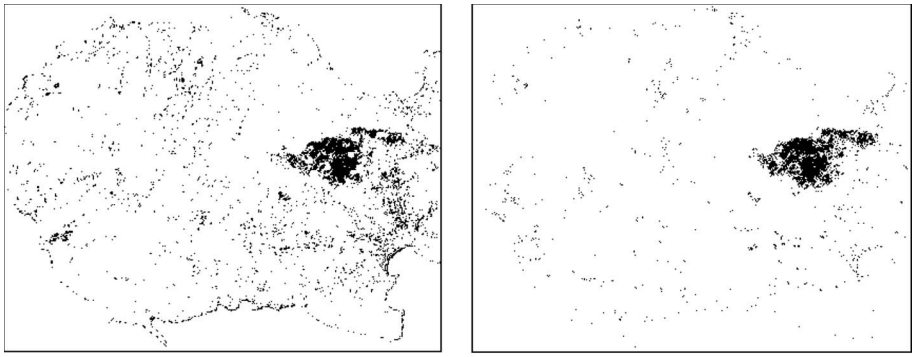
spectral–temporal signatures of mixed pixels associated with noise and misregistration (which often results in the uncertainty region of the distribution of the magnitude of SCVs) from the signature of true changes. It is worth noting that the different results obtained by the CVA with EM-based thresholding in these experiments and in those presented in [9] and [15] are due to the fact that, in this paper, we did not apply any filtering to the original images. This choice is motivated from the fact that the proposed method can intrinsically reduce the noise effects, thus allowing a more accurate geometrical representation of the scene. The residual noise in the map obtained with the proposed approach can be easily removed in the postprocessing, whereas the noise in the map generated by the CVA is more difficult to reduce without any filtering in the phase of image preprocessing.

The second investigated change-detection problem, due to the fact that $\mathcal{I}_2$ was acquired approximately one year after the 1993 fire, was much more complex, as confirmed by the poor results obtained by the standard CVA algorithm [Table I(b)]. This mainly depends on the fact that the vegetation had a fast growth after the fire, reducing the possibility to identify the burned area. However, the proposed technique was able to improve the performances by sharply increasing the kappa coefficient of 0.257 and by decreasing both false (i.e., -2914) and missed (i.e., -167) alarms, as confirmed by comparing Fig. 8(a) and (b). Also in this case, the unsupervised similarity measure for model selection chose a solution associated with change-detection results rather close to those related to the upper bound (i.e., 0.608 versus 0.645). Furthermore, the kappa accuracy yielded with the proposed S^3VM confirmed significantly higher than that obtained with the supervised SVM in the case of both the supervised model selection (upper bound) and the presented unsupervised model-selection strategy.



(a) (b)

Fig. 7. Change-detection maps obtained for the 1994 wildfire by (a) the standard CVA with EM-based thresholding [9] and (b) the proposed technique (Elba Island data set).



(a) (b)

Fig. 8. Change-detection maps obtained for the 1993 wildfire by (a) the standard CVA with EM-based thresholding [9] and (b) the proposed technique (Elba Island data set).

As for the previous experiment, a qualitative comparison between the change-detection maps in Fig. 8 confirms the effectiveness of the proposed technique, which resulted in a change-detection map where the impact of false alarms is significantly smaller than in the one yielded by the standard CVA algorithm. This also confirms the robustness of the proposed method to the residual registration noise.

All these results prove the effectiveness of the proposed approach on the Elba Island data set.

C. Results on the Mexico Data Set

This data set is made up of a section $(512 \times 360$ pixels$)$ of two coregistered multispectral images acquired by the TM sensor on an area in Mexico. The two images were acquired in April 2000 $(\mathcal{I}_1)$ and May 2002 $(\mathcal{I}_2)$ [Fig. 9(a) and (b)]. Two wildfires occurred between the two acquisition dates. A reference map concerning the location of the wildfire was available [Fig. 9(c)]. This map includes 29 506 changed pixels. Also in this case, the images were coregistered. A preliminary analysis pointed out that spectral channels 4 and 5 are the most relevant ones for discriminating the burned area on this data set. Accordingly, we used these channels in our trials. Also, in this case, we obtained high change-detection accuracy with the proposed technique (see Table II). In greater detail, the proposed approach sharply increased the kappa accuracy provided by the standard CVA technique from 0.844 to 0.913. The improvement is associated to a very significant decrease of false alarms (from 3840 to 1298) and to a slight reduction of missed alarms (from 3879 to 3827). Furthermore, the kappa accuracy yielded with the proposed unsupervised model-selection strategy (i.e., 0.913) is very close to the upper bound of accuracy obtained according to a supervised model selection (i.e., 0.915). Finally, also on this data set, the proposed S^3VM outperformed the standard supervised SVM trained on the pseudotraining set in the case of both the supervised model selection (upper bound) and the presented unsupervised model selection.

The change-detection maps shown in Fig. 10 confirm the quantitative results. The map obtained with the proposed approach exhibits a reduction of both false and missed alarms with respect to the map derived with the CVA technique, resulting in a better identification of the burned areas. This further proves the effectiveness of the proposed method.

IV. Discussion and Conclusion

A novel approach to unsupervised change detection in multispectral remote-sensing images has been presented, which is based on a properly defined S^3VM technique (initialized with a selective Bayesian thresholding) and an unsupervised model-selection strategy based on a similarity measure. On the one hand, with respect to the standard CVA algorithm, the proposed architecture properly exploits the information present in the original images by jointly analyzing the bitemporal spectral features (and not the results obtained by applying to the images the

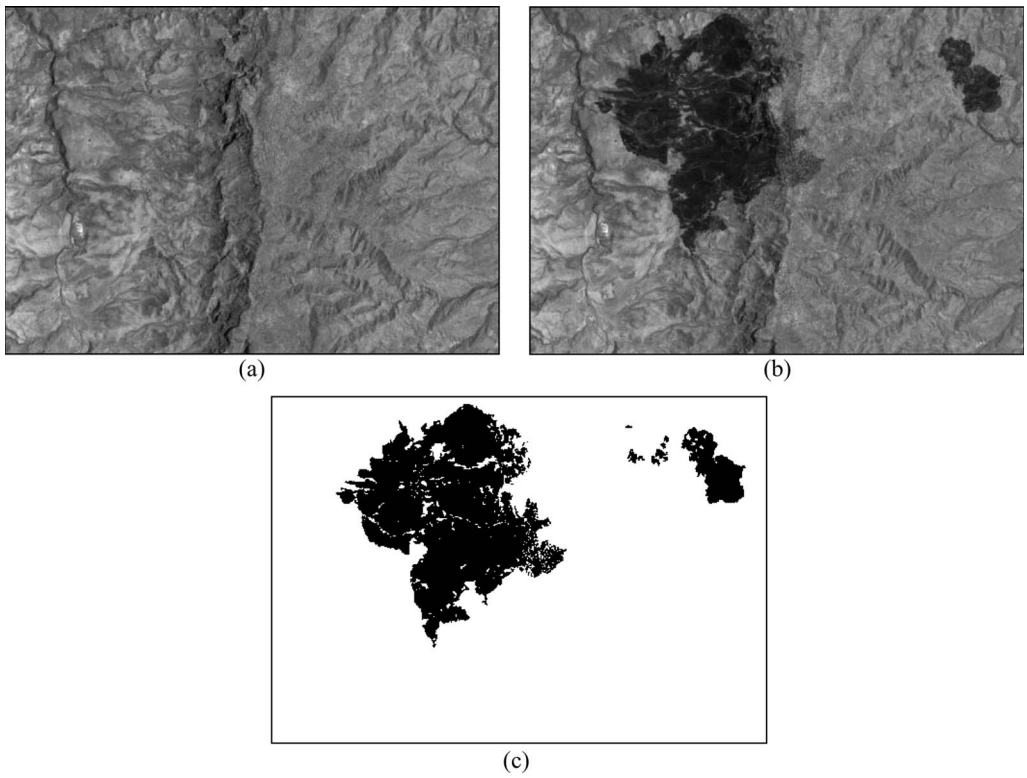


Fig. 9. Channel 4 of the Landsat-5 TM images acquired on an area in Mexico in (a) April 2000 and (b) May 2002. (c) Available reference map of the burned areas.

TABLE II
FALSE ALARMS, MISSED ALARMS, OVERALL ERRORS, AND KAPPA COEFFICIENTS FOR THE CHANGE-DETECTION MAPS OBTAINED FOR IDENTIFYING THE BURNED AREAS (MEXICO DATA SET)

Technique	Missed Alarms	False Alarms	Overall Error	Kappa Accuracy
CVA with EM-based thresholding [9]	3879	3840	7719	0.844
Proposed with SVM (upper bound)	3794	1234	5028	0.8949
Proposed with SVM (unsupervised model selection)	3827	1308	5135	0.8927
Proposed with S^3VM (upper bound)	2854	1283	4137	0.915
Proposed with S^3VM (unsupervised model selection)	2941	1298	4239	0.913

difference and magnitude operators, which are not bijective). On the other hand, thanks to the initialization seeds extracted by the Bayesian thresholding of the SCV magnitude and the joint use of unlabeled patterns, it results in an increased capability to adequately model the decision boundary between changed and unchanged classes in a completely unsupervised way. The proposed method exploits a novel validation procedure for the unsupervised definition of the parameters of the S^3VM technique. This validation procedure is based on the concept of similarity of the solutions and takes advantage from the rationale that correct solutions provide change-detection maps that are more correlated among them than wrong solutions.

Experimental results, obtained on two different remote-sensing data sets, confirm the effectiveness of the proposed technique that, thanks to the direct use of the spectral features of the considered images, sharply increased the change-detection accuracy with respect to the CVA algorithm applied with the EM-based thresholding. In addition, the proposed unsupervised model-selection procedure resulted in the choice of models associated with change-detection accuracies similar to the best possible ones (i.e., those identified according to a supervised model selection) on all data sets. It is worth noting that, although in this paper we carried out experiments on Landsat TM images in scenarios related to burned-area detection, the method is general and can be applied to any kind of multispectral images also in other application domains.

In all the experiments, the proposed approach proved effective also in reducing the effects of the registration noise present in the multitemporal images, without the application of any spatial filtering (and thus without degrading the spatial accuracy of the change-detection map). This mainly depends on the possibility to distinguish the spectral signatures of the different kinds of changes in the temporal and spectral feature space, thus separating true investigated changes from noise deriving from residual misalignment between multitemporal data (this noise usually results in mixed pixel that has a spectral signature in the two images different from that of true changes). It is worth noting that the mixed temporal–spectral signatures of pixels affected from registration noise result in SCVs that have a high probability to belong to the uncertainty region in $p(i^{\Delta})$.

Another important property of the proposed approach is that, due to the use of the distribution-free S^3VM classifier, it can be extended to the analysis of bitemporal images acquired in a multisensor framework (i.e., using images acquired by different sensors at the two considered dates). This can be easily obtained according to adequate modifications of the Bayesian

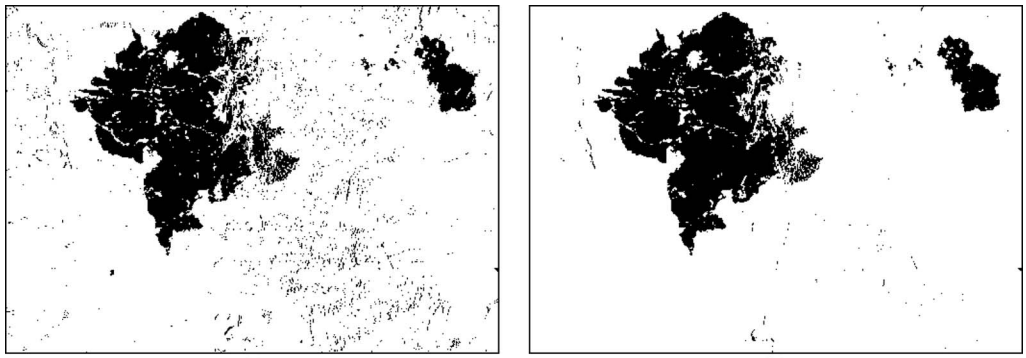


Fig. 10. Change-detection maps obtained for the Mexico data set by (a) the standard CVA with EM-based thresholding [9] and (b) the proposed technique.

initialization technique defined on the basis of the peculiarities of the considered multisensor images. In particular, we expect that different initialization strategies can be used for different sensors (e.g., CVA with EM-based Gaussian thresholding for multispectral images, and image ratioing with EM-based generalized Gaussian thresholding for SAR data [16]).

A critical issue (which is common to the standard CVA technique) to take into account in the use of the proposed method is that, when large-size images are considered (and thus, the prior probability of the class of changes may become very small), the selective Bayesian initialization and the learning of the S^3VM should be carried out by exploiting the split-based approach presented in [23]. In greater detail, the split-based approach automatically divides the analyzed scene into a set of nonoverlapping subimages (frames) of user-defined size. Then, frames are sorted out according to their probability to contain a significant amount of changed pixels (estimated according to the value of the variance of the magnitude of the SCVs on each frame). Afterward, a subset of frames having a high probability to contain changes is selected and analyzed. The initialization procedure and the semisupervised learning are applied to samples randomly extracted from these frames. This strategy allows one a proper modeling of the statistic of the changed pixels both in the initialization phase and in the learning of the S^3VM. In addition, if we fix the number of semilabeled samples to extract from the selected frames, we can obtain similar learning complexity in the case of large- and small-size images.

As a final remark, it is worth noting that the computational time required from the proposed technique is significantly higher than that taken from the CVA with EM-based thresholding algorithm (about 1 h against 2 min for generating a map on the considered data sets with a personal computer equipped with 512-MB RAM and Intel Core 2 Duo 1.8-GHz processor). This mainly depends on the model-selection strategy that is necessary for the definition of a proper S^3VM architecture. However, this time (which does not change significantly by increasing the size of the images, as it is mainly associated with the learning phase of the system) can be reduced by replacing the grid search with a suboptimal stochastic strategy for the estimation of the S^3VM parameters.

As future developments of the proposed work, we are considering the following: i) to extend the change-detection technique to the case of the analysis of very high resolution (VHR) images (this requires some methodological modifications for taking into account the multiscale properties of these data); ii) to extensively analyze the property of robustness to registration noise of the proposed approach on VHR data (in which residual registration noise after coregistration is usually sharply higher than in medium-resolution images); and iii) to extend the use of the unsupervised model-selection procedure for S^3VM to the more complex case of semisupervised image classification in the presence of multiclass problems.

REFERENCES

[1] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 858–867, Jul. 1997.

[2] L. Bruzzone and S. B. Serpico, "Detection of changes in remotely sensed images by the selective use of multi-spectral information," *Int. J. Remote Sens.*, vol. 18, no. 18, pp. 3883–3888, Dec. 1997.

[3] A. Singh, "Digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989.

[4] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 294–307, Mar. 2005.

[5] P. R. Coppin, I. Jonckheere, and K. Nachaerts, "Digital change detection in ecosystem monitoring: A review," *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1565–1596, May 2004.

[6] D. Lu, P. Mausel, E. Brondízio, and E. Moran, "Change detection techniques," *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2407, Jun. 2004.

[7] M. J. Carlotto, "Detection and analysis of change in remotely sensed imagery with application to wide area surveillance," *IEEE Trans. Image Process.*, vol. 6, no. 1, pp. 189–202, Jan. 1997.

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

[9] L. Bruzzone and D. Fernández 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.

[10] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF processing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sens. Environ.*, vol. 64, no. 1, pp. 1–19, 1998.

[11] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 463–478, Feb. 2007.

[12] J. R. G. Townshend, C. O. Justice, and C. Gurney, "The impact of misregistration on change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 5, pp. 1054–1060, Sep. 1992.

[13] J. Flusser and T. Suk, "A moment-based approach to registration of images with affine geometric distortion," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 2, pp. 382–387, Mar. 1994.

[14] P. L. Rosin, "Thresholding for change detection," *Comput. Vis. Image Underst.*, vol. 86, no. 2, pp. 79–95, May 2002.
[15] L. Bruzzone and D. Fernández Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 452–466, Apr. 2002.
[16] Y. Bazi, L. Bruzzone, and F. Melgani, "Image thresholding based on the EM algorithm and the generalized Gaussian distribution," *Pattern Recognit.*, vol. 40, no. 2, pp. 619–634, Feb. 2007.
[17] O. Chapelle, B. Schölkopf, and A. Zien, Eds., *Semi-Supervised Learning.* Cambridge, MA: MIT Press, 2006. AA.VV.
[18] L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive SVM for the semisupervised classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3363–3373, Nov. 2006.
[19] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[20] V. N. Vapnik, *Statistical Learning Theory*. Hoboken, NJ: Wiley, 1998.
[21] S. Ghosh, L. Bruzzone, S. Patra, F. Bovolo, and A. Ghosh, "A context-sensitive technique for unsupervised change detection based on Hopfield-type neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 778–789, Mar. 2007.
[22] R. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sens. Environ.*, vol. 37, no. 1, pp. 35–46, 1991.
[23] F. Bovolo and L. Bruzzone, "A split-based approach to unsupervised change detection in large-size multitemporal images: Application to tsunami damage assessment," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1658–1670, Jun. 2007.

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 with 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 interests are related to 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 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the *International Journal of Remote Sensing*, *Photogrammetric Engineering and Remote Sensing Journal*, *Remote Sensing of Environment Journal*, and *Pattern Recognition Journal*.

Dr. Bovolo ranked first place in the Student Prize Paper Competition of the 2006 IEEE International Geoscience and Remote Sensing Symposium (Denver, August 2006). Since 2006, she has been serving on the Scientific Committee of the SPIE International Conference on "Signal and Image Processing for Remote Sensing." In 2007, she served on the Scientific Committee of the IEEE Fourth International Workshop on the Analysis of Multi-Temporal Remote Sensing Images (MultiTemp 2007), Leuven, Belgium, July 2007, and of the IEEE International Geoscience and Remote Sensing Symposium 2007 (IGARSS'07) (Barcelona, Spain, July 2007).

Lorenzo Bruzzone (S'95–M'98–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.

From 1998 to 2000, he was a Postdoctoral Researcher with the University of Genoa. In 2000, he was with the University of Trento, Trento, Italy, where he his currently a Full Professor of telecommunications. He teaches remote sensing, pattern recognition, 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 area of remote-sensing image processing and recognition (analysis of multitemporal data, 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 an Evaluator of project proposals for many different governments (including the European Commission) and scientific organizations. He is the author (or coauthor) of 60 scientific publications in referred international journals, more than 120 papers in conference proceedings, and 7 book chapters. He is a Referee for many international journals and has served on the Scientific Committees of several international conferences. He is a member of the Managing Committee of the Italian Inter-University Consortium for Telecommunications and a member of the Scientific Committee of the India–Italy Center for Advanced Research.

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 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Best Reviewers in 1999 and was a Guest Editor of a Special Issue of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING on the subject of the analysis of multitemporal remote-sensing images (November 2003). He was the General Chair and Cochair of the First and Second IEEE International Workshop on the Analysis of Multitemporal Remote-Sensing Images (MultiTemp) and is currently a member of the Permanent Steering Committee of this series of workshops. Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. From 2004 to 2006, he served as an Associated Editor for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS and is currently an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He is also a member of the International Association for Pattern Recognition and of the Italian Association for Remote Sensing.

Mattia Marconcini (S'06) 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 2002, 2004, and 2008, respectively.

He is currently with the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento. His current research activities are in the area of machine learning, pattern recognition, and remote sensing. In particular, his interests are related to transfer learning and domain-adaptation classification and to image-segmentation problems. He conducts research on these topics within the frameworks of several national and international projects.