Title: Updating Land-Cover Maps by Classification of Image Time Series: A Novel Change-Detection-Driven Transfer Learning Approach

This paper appears in: IEEE Transactions on Geoscience and Remote Sensing

Date of Publication: 2012

Author(s): Begum Demir, Francesca Bovolo, Lorenzo Bruzzone

Volume:51 , Issue: 1

Page(s): 300-312

DOI: 10.1109/TGRS.2012.2195727

Updating Land-Cover Maps by Classification of Image Time Series: A Novel Change-Detection-Driven Transfer Learning Approach

Begüm DEMİR, *Member IEEE*, Francesca BOVOLO, *Member IEEE*, *and* Lorenzo BRUZZONE, *Fellow*, *IEEE*

Dept. of Information Engineering and Computer Science, University of Trento,
Via Sommarive, 14, I-38123 Trento, Italy
e-mail: demir@disi.unitn.it, francesca.bovolo@disi.unitn.it, lorenzo.bruzzone@ing.unitn.it.

***Abstract*—** This paper proposes a novel change-detection-driven transfer learning approach to update land-cover maps by classifying remote sensing images acquired on the same area at different times (i.e., image time series). The propose approach requires that a reliable training set is available only for one of the images (i.e., the source domain) in the time series, whereas it is not for another image to be classified (i.e., the target domain). Unlike other literature transfer learning methods, no additional assumptions on either the similarity between class distributions or the presence of the same set of land-covers classes in the two domains are required. The proposed method aims at defining a reliable training set for the target domain, taking advantage of the already available knowledge on the source domain. This is done by applying an unsupervised change detection method to target and source domains, and transferring class labels of detected unchanged training samples from the source to the target domain to initialize the target domain training set. The training set is then optimized by a properly defined novel Active Learning (AL) procedure. At the early iterations of AL, priority in labeling is given to samples detected as being changed, whereas in the remaining ones the most informative samples are selected from changed and unchanged unlabeled samples. Finally, the target image is classified. Experimental results

show that transferring the class-labels from the source domain to the target domain provides a reliable initial training set and that the priority rule for AL results in a fast convergence to the desired accuracy with respect to standard AL.

Index Terms – Transfer learning, active learning, automatic classification, remote sensing, time series

I. INTRODUCTION

Updating of land-cover maps by classification of remote sensing images is an important issue due to the availability of increased numbers of images regularly acquired by satellite-borne sensors on the same areas at different times (i.e., time series of remotely sensed images, temporally shifted images). Because of the new policies related to free availability of data (e.g., Landsat archive, future ESA Sentinel missions) this issue is becoming more and more strategic as time series are accessible to each potential users in a systematic way. Land-cover maps can be updated by direct supervised classification of each image in the time series. However, such an approach requires reliable ground reference data for all the available temporal images in order to properly train the classifier. In operational scenarios, gathering a sufficient number of labeled training samples for each single image to be classified is not realistic due to the high cost and the related time consuming process of this task. Moreover, although the images in the time series refer to the same area, ground reference samples available on one of the images may not follow the same distribution in other acquired images due to several reasons, such as differences in the atmospheric conditions at the image acquisition dates, different acquisition system state, different levels of soil moisture, changes occurred on the ground, etc. In these situations, exploiting the classifier trained on the image for which training data are available may result in poor classification performance,

and therefore recollection of labeled samples is necessary. To reduce the need and effort to recollect labeled samples, it is desirable to reuse the already available information on images acquired on the same area of interest (source domain) to classify new images acquired on the same area (target domain). To deal with this problem, transfer learning (TL) techniques, and more in detail domain adaptation (DA) methods in transfer learning, have been recently introduced in the remote sensing literature [1]-[8]. DA (also known as partially supervised/unsupervised learning) methods define strategies that use the information available on the source domain to classify the target domain for which no prior information is available, assuming that the two domains may have different (but related) distributions [9]-[11]. In [1]-[5], DA problems are addressed by semi-supervised learning (SSL) that exploits a classifier trained on the source domain for the target domain after tuning its parameters by using unlabeled data of the target domain. These methods are defined under two assumptions: i) the set of land-cover classes that characterizes the target domain should be the same as those included in the source domain, and ii) the land-cover class statistical distributions should be sufficiently correlated (but not necessarily identical) between the domains. However, in some real remote sensing classification problems these assumptions could not be satisfied due to i) the possible appearance and/or disappearance of the land-cover classes during time, and ii) the possible high differences on the class statistical distributions in the image time series.

To overcome the limitations of the former assumption, a DA method with SSL is presented in [6], which automatically identifies the differences between the set of classes in the target and source domains by exploiting unlabeled samples of the target domain together with the labeled samples of the source domain, and considers these differences in the map updating process. In [7]-[8] DA problems are addressed with Active Learning (AL), which iteratively selects the most informative unlabeled samples of the target domain to be included in the training set after manually labeling by a supervisor. Thus, differently from [6], a small number of labeled training

samples for the target domain is exploited in addition to the labeled samples of the source domain. By these methods, it is possible to include the information on new appeared classes in the training set via manually labeling process. However, according to our knowledge, in the remote sensing literature, there are not DA methods that can work efficiently also when the second above-mentioned assumption does not hold, i.e., significant differences between statistical distribution of the source and target domains are present.

In order to deal with the above-mentioned problems, we propose a novel change-detection-driven transfer learning (CDTL) approach to the classification of multitemporal images that overcomes the limitations about the possible differences on both i) the land-cover class statistical distributions and ii) the set of land-cover classes present in the source and target domains. The main idea of the proposed approach is to fuse TL with AL for exploiting at the best the available training data for the source domain together with few specific new labeled data of the target domain selected for optimizing the classification accuracy. In greater detail, unlike other DA approaches, the proposed one takes advantage of the properties of time series for defining TL in terms of label propagation of source training patterns rather than in terms of adaptation of the classification parameters of the source domain to the target domain (as it is usually done in the TL literature). This is accomplished according to three steps. The first step is the TL step and is devoted to define an initial training set for the target domain without collecting labels on its samples. To this end, unsupervised change detection is applied to target and source domains, and class labels of detected unchanged training samples are propagated from the source to the target domain. This novel approach, unlike the DA methods proposed in remote sensing literature, allows one to estimate the classification parameters of the target domain directly from the target domain samples. Thus there is no need to adapt the classification parameters of the source domain to the target domain. Accordingly the proposed system can handle possible significant differences between statistical distributions of the source and target domains. In the second step, the initial

training set is enriched by a novel AL procedure, which gives priority to the labeling of the samples detected as changed at early iterations (Priority AL), and selects samples among all unlabeled samples (i.e., changed and unchanged ones) at the remaining iterations (Standard AL). This novel procedure results in a fast increase of the classification accuracy versus the number of new labeled samples of the target domain. Moreover, due to the manual labeling process associated with AL, unlike other literature methods the proposed approach does not require that the same set of land-cover classes describes the two domains. At convergence of the AL process, in the third step the target image is classified by a supervised classifier. The experiments conducted on two different multitemporal and multispectral data sets show the effectiveness of the proposed technique.

The paper is organized into six sections. In Section II, related works in AL and TL are surveyed. Section III defines the considered problem and describes the proposed CDTL approach. Section IV illustrates the considered data sets and the design of experiments. Section V shows the experimental results. Finally, Section VI draws the conclusion of this work.

II. RELATED WORK

In this section, we review some AL and TL techniques presented in the literature for classification of remote sensing images.

A. Active Learning

AL methods defined for the supervised classification of single images assume that a small initial training set is available for the image to be classified and aim at expanding it in the most effective way to define an optimized training set. This is done by an iterative procedure. At each iteration, the most informative samples among a pool of unlabeled samples are selected by a query function and included in the current training set after manually labeling by a supervisor. The most informative unlabeled samples are the samples that have the lowest probability to be correctly classified by the current classification model, and thus have maximum uncertainty among all

unlabeled samples [12]. The supervisor is an expert who is able to reliably assign the correct label to selected samples (note that this is the standard assumption behind every AL approach used in remote sensing). In some cases the labeling process can be done by photointerpretation, whereas in other cases it may require ground data collection. When the AL process is completed, the training set consists of a minimum number of most informative samples for the related classifier. The main advantages of AL are: i) the reduced labeling cost (as a result of avoiding redundant sampling), and ii) the reduced computational complexity for training the classifier (as a result of the selection of an optimal small size training set). In the remote sensing literature several AL techniques have been presented to optimize the training set for the single-date image classification. In [12], the unlabeled sample that is closest to the classification boundary (i.e., classification margin) of each binary Support Vector Machine (SVM) is considered as the most informative and therefore included in the current training set at each iteration of the AL process. An AL technique that selects the unlabeled sample that maximizes the information gain is presented in [7]. To estimate the information gain, the Kullback–Leibler (KL) divergence is calculated between the posterior probability distribution of the current training set and the training set obtained by including each unlabeled sample into the training set. In [13], different AL techniques proposed in the machine learning literature are investigated for the multi-class SVM classification problems, and also a novel AL method is proposed. The latter firstly selects the most informative unlabeled samples by the Multiclass-Level Uncertainty strategy. Then it analyzes their distribution by using the *k*-means clustering in the kernel space. Finally, the most informative (i.e., most uncertain) sample of each cluster is added to the training set at each iteration of AL. In [14], a cluster assumption based AL method is presented for addressing critical problems where significantly biased initial training sets are available. Label acquisition costs sensitive AL techniques are proposed in [15], [16]. In [15], the cost is measured with respect to the distance traveled during the labeling process, whereas in [16] it is measured either in units of time (which depend on the vehicles average speed and the cost

of labeling each sample) or in terms of distance traveled during the labeling process.

B. Transfer Learning

The problem of updating land-cover maps by classifying image time series when ground reference samples are available only for one time image is addressed by DA methods in the framework of TL. In the last years, TL has obtained an increasing interest in the remote sensing community due to the increased availability of time series of remotely sensed images. TL methods address the problem of identifying which knowledge can be transferred and how to transfer it across the domains [9]. A DA method (named as a partially unsupervised classification method) is presented in [1]. This method is able to update the parameters of an already trained parametric maximum-likelihood (ML) classifier trained on the source domain on the basis of the distribution of the target domain for which training data are not available. In order to better exploit the temporal correlation between images, the method has been generalized in the context of the Bayesian rule for cascade classification in [2]. A further improvement of this approach is proposed in [3] by presenting a multiple cascade classifier system that is made up of ML and radial basis function neural-network classifiers. Another DA method based on the SVM classifier is presented in [4] where a novel circular validation strategy for the accuracy assessment of the classification results is also described. In this work, firstly labeled samples of the source domain are exploited to initialize the discriminant function for the target domain. Then the unlabeled patterns of the target domain that have a high probability to be correctly classified are iteratively included in the training set, whereas the labeled samples of the source domain are gradually removed. A DA method for the binary hierarchical classifier (BHC) is presented in [5]. This method aims to update the parameters of a BHC classifier trained on the source domain on the basis of the distribution of the target domain. The presented algorithm can be used when either no labeled training samples are available or a small number of training samples exist for the target domain. The DA methods presented in [2], [3] are further improved in [6] by addressing the problems related to the differences on the set

of land-cover classes between the domains. This is done by the joint use of a change-detection method and of the Jeffreys-Matusita (*JM*) statistical distance measure. AL methods for DA are presented in [7] and [8]. In [7], the unlabeled samples from target domain that have the maximum information gain measured by the KL divergence are included in the training set of the target domain after manual labeling, whereas the initial classification parameters are obtained by the distributions estimated on the labeled samples of the source domain. In [8], the initial statistical parameters of an ML classifier are calculated by exploiting the labeled samples from the source domain. Then, the most informative samples are selected from the target domain by AL for manual labeling and inclusion in the training set, whereas the source domain samples that do not fit with the distributions of the classes in the target domain are removed. Because of the manual labeling process of AL, these methods are able to detect new appeared classes but not the disappeared ones. All the above mentioned methods may achieve low classification accuracy for the target domain when the class statistical distributions in the target domain differ significantly from those of the source domain. Moreover, except [6], these methods result in a poor classification performance when the two domains do not share exactly the same set of land-cover classes since classes might be appeared or disappeared between available multitemporal acquisitions. Therefore, it is necessary to develop TL methods that are not considerably affected from both the distribution differences in land-cover classes and the possible appearance and disappearance of the land-cover classes between the domains.

III. PROPOSED CHANGE-DETECTION-DRIVEN TRANSFER LEARNING APPROACH

Let $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_P]$ be a time series made up of P co-registered remote-sensing images acquired on the same area at different times, where $\mathbf{X}_p = \{x_{p,1}, x_{p,2}, ..., x_{p,B}\}$ is the p-th multispectral remote sensing image in the series acquired at time t_p and made up of C spectral channels and B pixels. Without losing in generality, let us consider two images extracted from the

time series and let us define them for simplicity as $\mathbf{X}_1 = \{x_{1,1}, x_{1,2}, ..., x_{1,B}\}$ and $\mathbf{X}_2 = \{x_{2,1}, x_{2,2}, ..., x_{2,B}\}$. Let $\left(x_{1,j}, x_{2,j}\right)$ be the *j*-th pair of temporally correlated pixels made up of a pixel $x_{1,j}$ acquired at time t_1 and a spatially corresponding pixel $x_{2,j}$ acquired at time t_2. Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_R\}$ be the set of land-cover classes at time t_1, and $\mathrm{N} = \{v_1, v_2, \ldots, v_N\}$ be the set of land-cover classes at time t_2. Differently from the standard TL methods [1]-[4], here we assume that the classes in Ω may be different from those in N. Let us assume that the image $\mathbf{X}_1$ is the source domain. Accordingly, we assume that a reliable training set $T_1 = \{x_{1,i}, y_{1,i}\}_{i=1}^{M}$ is available for it, where $x_{1,i} \in \mathbf{X}_1$ is the *i*-th training sample, $y_{1,i} \in \Omega$ is the associated class label, and $M < B$ is the number of training samples. In addition, we assume that a training set T_2 for the image $\mathbf{X}_2$ (target domain) is not available. The goal of the proposed method is to produce a classification map of image $\mathbf{X}_2$ by taking advantage of the previously available knowledge from $\mathbf{X}_1$. Here, this is achieved by a change-detection-driven transfer learning (CDTL) approach, which also includes an active learning strategy. The proposed approach is based on 3 steps: i) a novel change-detection-driven TL approach; ii) a novel change-detection-driven AL procedure; and iii) target domain classification. Fig. 1 shows the block scheme of the proposed CDTL approach. Each step of the proposed method is explained in detail in the following.

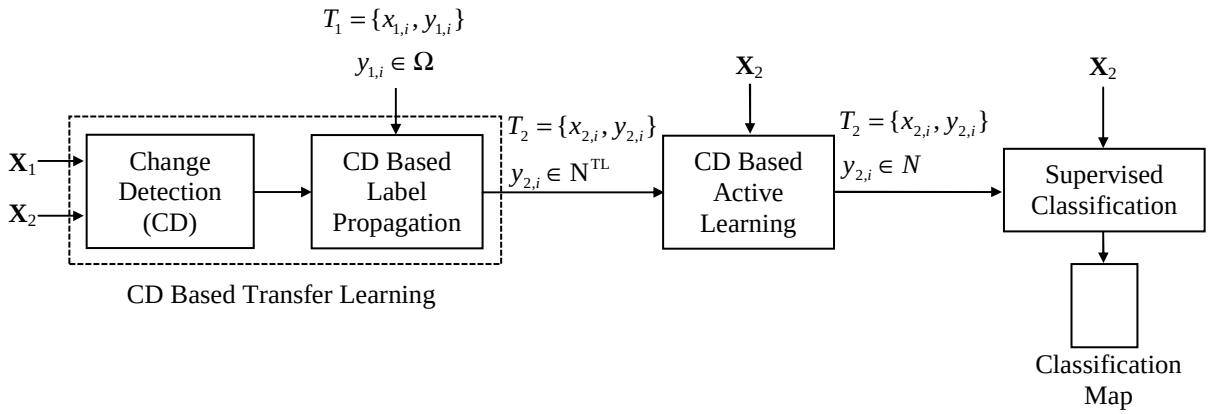


Fig. 1. Block diagram of the proposed change-detection-driven transfer learning (CDTL) approach.

Step 1: Change-Detection-Driven Transfer Learning

The first step aims at defining an initial training set for the image $\mathbf{X}_2$ (target domain) by taking advantage of the available knowledge from $\mathbf{X}_1$ (source domain). To this end, here we adopt a novel change-detection-based TL procedure. The basic idea behind this choice is that the labels of training samples in T_1 can be considered reliable for $\mathbf{X}_2$, and thus transferred, only if the related pixels did not change. Accordingly, at first there is a need of detecting whether changes occurred on the ground between $\mathbf{X}_1$ and $\mathbf{X}_2$, and if different land-cover transitions occurred. As no training set is assumed available for $\mathbf{X}_2$, an unsupervised change-detection method should be used. Here the Change Vector Analysis (CVA) technique [17] is adopted due to its capability in distinguishing among kinds of change associated to different land-cover transitions in an unsupervised way (which is not a common property of unsupervised change detection methods that usually just discriminate between changed and unchanged areas) and its simplicity and effectiveness. Note that the information about different kinds of change is required in the following step of the proposed approach (i.e., in the change-detection-driven AL procedure). In the CVA technique, temporally correlated (thus spatially corresponding) pixels $x_{1,j}$ and $x_{2,j}$ are subtracted to each other in order to build a multispectral difference image $\mathbf{X}_D$. To make CVA reliable, the basic assumption is that multitemporal images should be co-reregistered to each other. If after co-registration significant residual misregistration errors affect multitemporal data, CVA-based change detection methods robust to this kind of problem can be adopted [24]. The information present in the multispectral difference image is analyzed according to the theoretical framework for unsupervised change detection based on the CVA in polar domain proposed in [18]. According to [18], for simplicity two spectral bands out of *C* are selected such that the most informative features with respect to the specific considered problem are isolated excluding noisy and misleading spectral channels from the analysis. It is worth noting that, even if the assumption

of working with a couple of spectral channels is reasonable in many change-detection problems [17]-[19], the CVA can be also applied to all spectral channels. From the defined 2-dimensional feature space, the polar representation of the change-detection problem is built on the basis of the magnitude and the direction of spectral change vectors. In this feature space, unchanged pixels are concentrated close to the origin of the polar domain and fall within the *Circle of no-changed* (C_n) pixels, whereas changed pixels fall far from the origin within the *Annulus of changed* (A_c) pixels [18]. The threshold value T that separates C_n from A_c along the magnitude variable ρ can be computed according to any thresholding technique available in the literature [19], [20]. Changed pixels belonging to different land-cover transitions (*i.e.*, different kinds of change) show along the direction variable ϑ in A_c different preferred directions and fall therefore in different *Annular sectors of change* (S_h, h=1,…, H, where H is the number of detected kinds of change). Each sector is bounded by a pair of angular thresholds ϑ_{h_1} and ϑ_{h_2} that can be automatically detected according to the method described in [21] or interactively identified according to a visual analysis of the polar domain. Fig. 2 shows the decision regions according to the CVA framework.

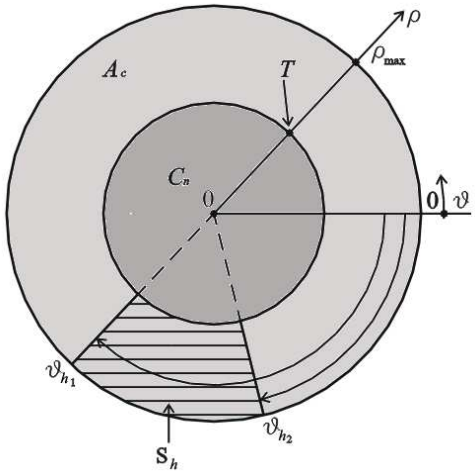


Fig. 2. Representation of the regions of interest for the CVA technique in the Polar coordinate system.

Once change detection has been performed, the knowledge related to $\mathbf{X}_1$ which has a high probability to be reliable also for $\mathbf{X}_2$ is transferred. Such information is represented by the labels

of pixels that are detected as being unchanged, i.e., the ones that fall in C_n. Let $T_1^{UC} = \{x_{1,i}, y_{1,i}\}_{i=1}^{R}$ be the set of unchanged training pixels at $\mathbf{X}_1$, where $x_{1,i} \in T_1^{UC}$ is the i-th training sample and R ($R < M$) is the number of unchanged training samples. The initial training set T_2 of $\mathbf{X}_2$ is defined as $T_2 = \{x_{2,i}, y_{1,i}\}_{i=1}^{R} = \{x_{2,i}, y_{2,i}\}_{i=1}^{R}$ where $x_{2,i} \in \mathbf{X}_2$ is the i-th initial training sample and $y_{1,i} \equiv y_{2,i}$ is its label transferred from T_1^{UC}. The classes that are represented in T_2 define the initial set of classes N^{TL} for $\mathbf{X}_2$ after TL. N^{TL} is by definition a subset of land-cover classes at t_1 (i.e., $\mathrm{N}^{TL} \subseteq \Omega$). The labels of all training samples detected as being changed, i.e., the ones that fall in S_h (h=1,…, H), are not transferred.

Due to this step, unlike other TL techniques presented in the remote sensing literature [1]-[8], the proposed approach does not require to adapt the classifier parameters of the source domain to the target domain. Thus the proposed approach results robust to the class statistical distribution differences between the source and target domains. It is worth noting that even if in this paper we use the CVA technique, this step can be implemented with any unsupervised change detection technique that can identify the presence of different land-cover transitions.

Step 2: Change- Detection-Based Active Learning

Step 1 completely removes from the representation of the problem at t_2 the information about changed pixels, since it does not transfer labels of training samples that fall in S_h (h=1,…, H) assuming that they are unreliable as possibly changed. However, changed pixels are highly important for the map-updating process since they may carry information about possible new classes appeared in $\mathbf{X}_2$ and/or about different statistical properties of spatially shifted classes (*i.e.*, classes that are already in N^{TL} but appear in a different spatial positions in $\mathbf{X}_2$ with respect to $\mathbf{X}_1$). Neglecting this information would lead to unreliable classification of $\mathbf{X}_2$. Thus we can state

that from the AL viewpoint unlabeled changed samples in $\mathbf{X}_2$ are potentially highly uncertain (and thus informative) with respect to the classification of $\mathbf{X}_2$. Accordingly in step 2 the initial training set T_2 is expanded by exploiting a novel change-detection-driven AL procedure. Let us focus the attention on changed pixels. Pixels among them associated to new classes appeared in $\mathbf{X}_2$ are the most informative, since they are not represented in the training set T_2, whereas pixels associated to spatially shifted classes are highly informative only if their spectral signature is different from that of the pixels of the same class already included in T_1^{UC}. On the opposite, changed pixels that are associated to spatially shifted classes and that have a spectral signature similar to each other are less interesting. The adopted unsupervised change-detection technique is able to identify the presence of different land-cover transitions. However, since it is unsupervised, it does not provide any information about the new labels assumed by changed pixels in $\mathbf{X}_2$. In other words, it is not able to distinguish among the above-mentioned three change cases. In order to deal with this problem and distinguish these cases, we adopt a method based on statistical distance measures recently proposed in [6]. Let $\mathrm{E}=\{\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_H\}$ be the set of unknown class labels that changed pixels assume in $\mathbf{X}_2$. Note that the number of transitions can be estimated on the basis of the different number of annular sectors of changes detected in the polar domain. In order to understand whether $\varepsilon_h \in \mathrm{E}$ is already present in the initial training set T_2 (i.e., $\varepsilon_h \in \mathrm{N}^{TL} \cap \mathrm{E}$) or not (*i.e.*, $\varepsilon_h \notin \mathrm{N}^{TL}$) the similarity between the statistical distribution of each $\varepsilon_h \in \mathrm{E}$ and that of each land-cover class $\omega_u \in \mathrm{N}^{TL} \subseteq \Omega$ present in the initial training set T_2 is computed. Class similarity is measured according to a pairwise Jeffreys-Matusita (*JM*) distance. Here *JM* distance is selected due to its asymptotic behavior. Unlike other distance measures that are unbounded, the *JM* distance reaches saturation to the square root of 2. This behavior makes it easy to establish a threshold value that defines a high distance (in terms of the Chernoff upper bound to the error

probability [27]) of classes [6]. However any other distance measures [25] can be also used. Thereby, the *JM* distance JM_{hu} between $\varepsilon_h \in \mathrm{E}$ and $\omega_u \in \mathrm{N}^{TL}$ can be calculated as

$$JM_{hu} = \sqrt{2\left(1 - e^{-B_{hu}}\right)} \tag{1}$$

where B_{hu} is the Bhattacharyya distance between the two considered classes and can be calculated as

$$B_{hu} = -\ln\left\{\int_{X_2} \sqrt{p(X_2 \mid \varepsilon_h)\, p(X_2 \mid \omega_u)}\, dX_2\right\} \tag{2}$$

where $p(X_2 \mid \varepsilon_h)$ and $p(X_2 \mid \omega_u)$ are the class conditional density functions of the random variable X_2 associated to the image $\mathbf{X}_2$. Under the assumption of Gaussian distributed classes, (2) can be rewritten as

$$B_{hu} = \frac{1}{8}\left(\mu_h - \mu_u\right)^T \left(\frac{\Sigma_h + \Sigma_u}{2}\right)^{-1} \left(\mu_h - \mu_u\right) + \frac{1}{2}\ln\left(\frac{1}{2}\frac{\left|\Sigma_h + \Sigma_u\right|}{\sqrt{\left|\Sigma_h\right|\left|\Sigma_u\right|}}\right) \tag{3}$$

where μ_h and μ_u are the mean vectors of the classes ε_h and ω_u, respectively, Σ_h and Σ_u are their covariance matrices, and *T* represents the transpose operator.

If for a given $\varepsilon_h \in \mathrm{E}$ all computed pairwise *JM* distances are higher than a user defined threshold value *Th*, we detect the presence of a new class (*i.e.*, $\varepsilon_h \notin \mathrm{N}^{TL}$) or of a spatially shifted class (*i.e.*, $\varepsilon_h \in \mathrm{N}^{TL}$) with a significantly different spectral signature compared to the same class already present in T_1^{UC}. These are the two situations in which changed pixels are highly uncertain and thus particularly informative for the AL step. If $\varepsilon_h \in \mathrm{E}$ exhibits a small *JM* distance with one of the $\omega_u \in \mathrm{N}^{TL} \subseteq \Omega$ classes, a spatially shifted class with spectral signatures similar to those of the same class already present in T_1^{UC} is detected. This procedure is applied until all classes in E have been analyzed. Once this information has been retrieved, the proposed approach applies AL

with a mechanism of priority. If classes $\varepsilon_h \in \mathrm{E}$ that show a high *JM* distance to all land-cover classes in N^{TL} have been detected, at the first iterations of the AL process the pool of unlabeled samples for $\mathbf{X}_2$ is made up only of pixels associated to these classes (i.e., changed samples). These pixels are candidate to be either a new class or a spatially shifted class which is not properly modeled by the samples in the training set T_2. This sub-step is called Priority AL Step. If there is only one unknown class with a high *JM* distance (i.e., H=1), the most uncertain *b* samples are selected from this class only and labeled manually to be included in the current training set. In the case there are more unknown classes with high *JM* distances (i.e., $H>1$), the most uncertain b/H samples are selected from each class, which results in the selection of *b* most uncertain samples in total. The manual labeling of the additional *b* samples solves the ambiguity between spatially shifted classes (i.e., $\varepsilon_h \in \mathrm{N}^{TL}$) that have different spectral signatures and new classes ($\varepsilon_h \notin \mathrm{N}^{TL}$). Accordingly, in case of $\varepsilon_h \notin \mathrm{N}^{TL}$ the set N of land-cover classes at time t_2 is defined as $\mathrm{N}=\mathrm{N}^{TL} \cup \{\varepsilon_h\}$, whereas if $\varepsilon_h \in \mathrm{N}^{TL}$, the set N of land-cover classes is equal to the initial set of classes N^{TL}, i.e., $\mathrm{N}=\mathrm{N}^{TL}$. In the second part of this step, priority is removed and standard AL is applied including all unlabeled samples in the pool, i.e., both unchanged and changed samples. It is worth nothing that if there are no changed pixels, the algorithm avoids the Priority AL Step. This is the case in which no new classes are detected and $\mathrm{N} \equiv \mathrm{N}^{TL}$. Moreover it may happen that the labels propagated from $\mathbf{X}_1$ are sufficient for a reliable classification of $\mathbf{X}_2$. In this case no additional labeled data will be collected for $\mathbf{X}_2$ and the step 2 can be avoided. Fig. 3 shows the block diagram of the change-detection-based active learning step, whereas Fig. 4 demonstrates the relationship between N and N^{TL} in the different cases. It is worth nothing that this step is independent from the adopted AL technique, and thus can be used with any AL technique presented in the literature [12]-[16].

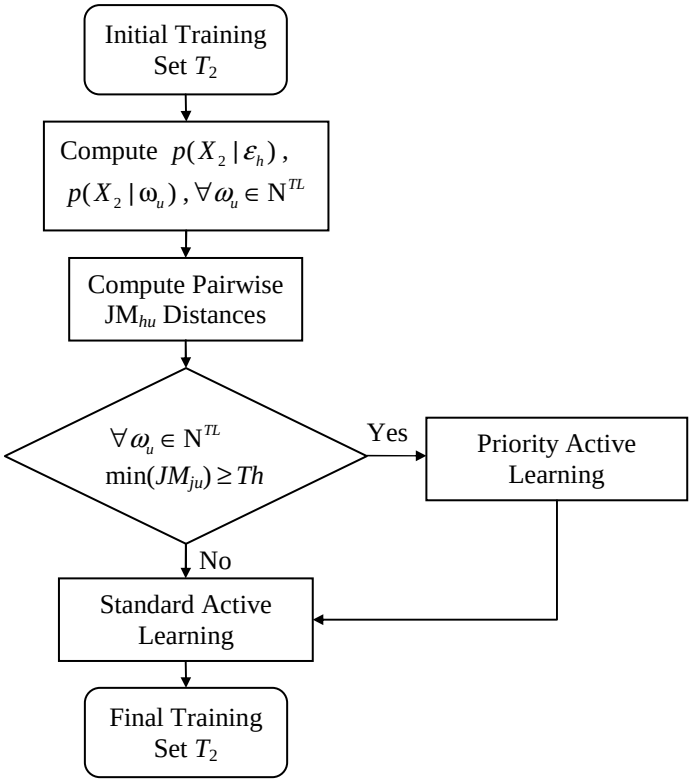


Fig. 3. Block diagram of the change-detection-based active learning step (second step).

As a final remark, it is worth pointing out that detected kinds of change may interest all the pixels associated to a given class at t_1, i.e., classes in Ω may disappear. This may happen in the case of detecting both a new class or a spatial shift of classes. However this situation is not critical since it is implicitly managed by the proposed method by not transferring labels associated to changed training samples in Step 1. Thus, if a given class in Ω is no longer present in N, all the related pixels will be changed and their labels will not be represented in N^{TL}.

Step 3: Target Domain Classification

In the last step, when the AL process is completed, the image $\mathbf{X}_2$ is classified. This is done by training the classifier with the training set T_2 obtained at the convergence of the step 2. It is worth nothing that the proposed approach is independent from the classification method, and therefore can be used with any classification technique presented in the literature [25],[26].

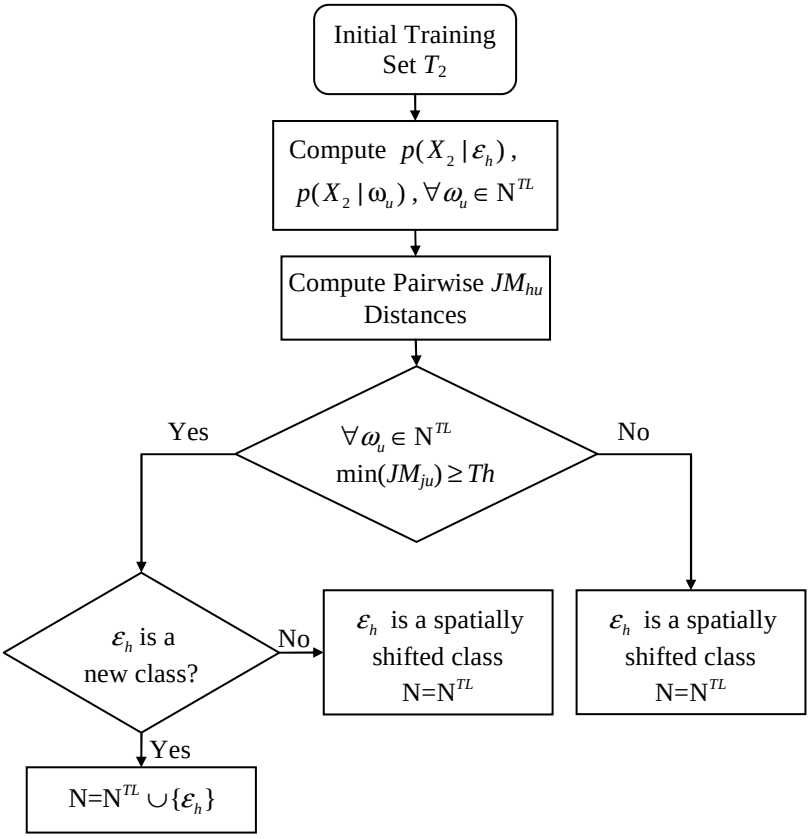


Fig. 4. The relationship between the set of classes N obtained at the convergence of the step 2 and the set of classes N^{TL} obtained at the end of step 1.

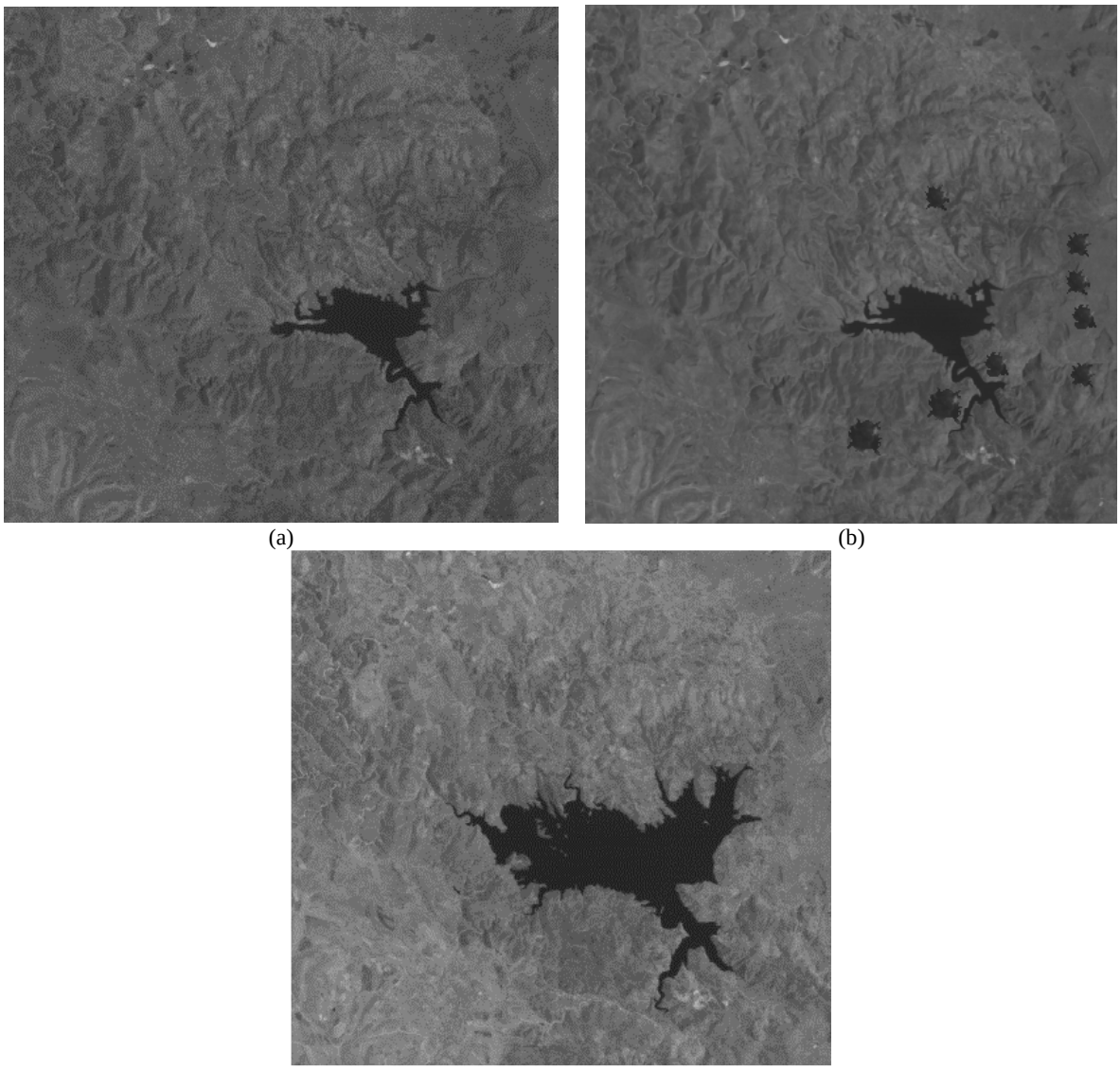
IV. Data Set Description and Design of Experiments

A. Data Set Description

In our experiments, we used two multitemporal data sets. The first one is made up of two co-registered multispectral images acquired by the Landsat-5 satellite on the Island of Sardinia, Italy, in September 1995 and July 1996 (see Fig. 5.a and Fig. 5.c). The two images share five land-cover classes (*i.e.*, pasture, forest, urban area, water, vineyard). The image acquired in September 1995 also includes one additional class, *i.e.*, bare soil. In order to properly demonstrate the effectiveness of the proposed method in our experiments we considered the July 1996 image as $\mathbf{X}_1$ and the September 1995 image as $\mathbf{X}_2$. This choice allows us to test the proposed method in the case in which a new class appeared in the time interval between $\mathbf{X}_1$ and $\mathbf{X}_2$. In order to simulate more complex situations with multiple new appeared classes, an additional class of burned area related

to a forest fire was included in the September 1995 image (see Fig. 5.b). To this end, a data set made up of two multispectral images acquired by the Landsat-5 TM multispectral sensor on the Island of Elba (Italy) in August 1994 and September 1994 [17] was utilized. Using the available ground truth for Elba image, the spectral signatures of burned areas taken from the Elba image were inserted in the forest areas of the September 1995 image. Thus, the modified September 1995 image includes two additional classes, i.e., bare soil and burned area, with respect to the July 1996 image. Accordingly, for the Sardinia data set we have two scenarios. In the first scenario (Scenario 1) the original September 1995 image (which includes the appearance of only one new class, i.e., bare soil) is considered as $\mathbf{X}_2$, whereas in the second one (Scenario 2) the modified September 1995 image (in which two new classes appeared, i.e., bare soil and burned area) is considered as $\mathbf{X}_2$ fixing the July 1996 image as $\mathbf{X}_1$ (see Table I).

The second data set is made up of two co-registered and pansharpened multispectral Very High geometrical Resolution images acquired by the QuickBird satellite on the city of Trento, Italy, in October 2005 and July 2006, respectively (see Fig. 6). Here we considered the October 2005 image as $\mathbf{X}_1$ and the July 2006 image as $\mathbf{X}_2$. The images share five land-cover classes (*i.e.*, water, red roof, asphalt, fields, and bare soil). One additional class is present in the July 2006 image, thus between the two acquisitions an addition of one new class (*i.e.*, plastic-mulched fields) is registered.

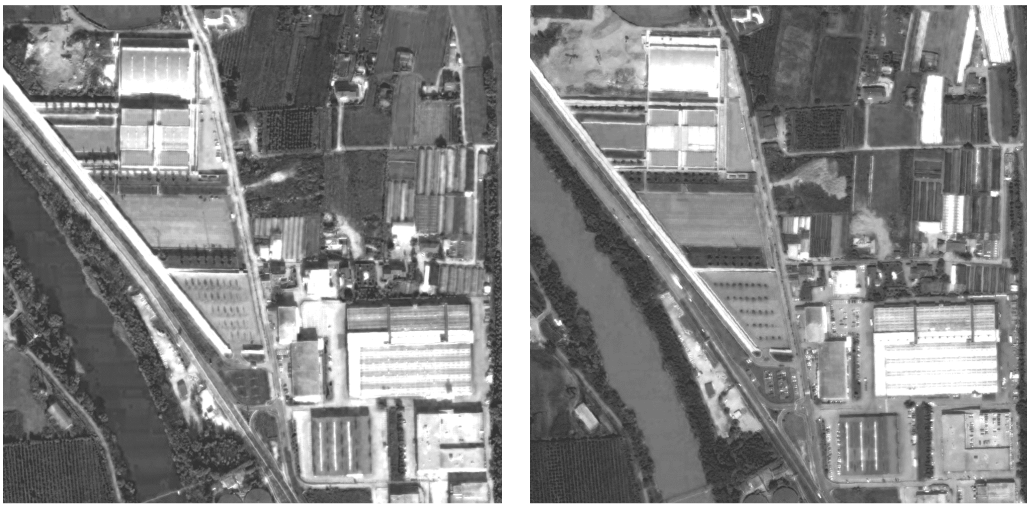


(a) (b) (c)

Fig. 5. Spectral band 4 of the Sardinia data set: (a) image acquired in September 1995 ($\mathbf{X}_2$, Scenario 1); (b) modified image acquired in September 1995 ($\mathbf{X}_2$, Scenario 2); and (c) image acquired in July 1996 ($\mathbf{X}_1$).

TABLE I. SCENARIOS CONSIDERED IN THE EXPERIMENTS FOR THE SARDINIA DATA SET

Scenario	$\mathbf{X}_1$	$\mathbf{X}_2$
1	July 1996	September 1995
2	July 1996	Modified September 1995



(a) (b)

Fig. 6. Spectral band 1 of the Trento data set: (a) image acquired in October 2005 ($\mathbf{X}_1$); and (b) image acquired in July 2006 ($\mathbf{X}_2$).

B. Design of Experiments

For both data sets, all the reported experimental results are referred to the average accuracy obtained on ten trials according to ten initial randomly selected training sets for T_1 that correspond to ten different initial training sets for T_2. For each trial, the Sardinia data set includes a training set T_1 of 378 samples for the $\mathbf{X}_1$ image selected randomly among 2565 available labeled samples, whereas the Trento data set consists of a training set T_1 of 470 samples for the $\mathbf{X}_1$ image selected randomly among 3885 available labeled samples. For both data sets, a training set T_2 for the image $\mathbf{X}_2$ is assumed initially not available.

We carried out the experiments with batch size values b=4 and b=10 which show the total number of samples being added to the current training set T_2 at each iteration of the AL step. For the classification step an SVM classifier [22], [23] with Radial Basis Function (RBF) kernel is used. The choice of the SVM is done as it is one of the most widely used classifiers in remote sensing due to both its theoretical properties and its proven empirical effectiveness on many different kinds of data and applications. However, as also mentioned before, the proposed method is general and can be used with any classifier. The values of the regularization parameter of the

SVM classifier and the γ parameter of the RBF kernel are obtained by a five-fold cross validation. The regularization parameter is tested between [10-1000] with a step size increment of 20, and the γ parameter of the RBF kernel function is tested between [0.1-2] with a step size increment of 0.1. The value of threshold *Th* is set to 0.99 due to the fact that if a pair of classes have a *JM* distance smaller than 0.99 they are assumed to be similar [6].

In this paper, we have exploited the Multiclass-Level Uncertainty with Enhanced Clustering Based Diversity (MCLU-ECBD) technique, which is an effective AL method defined on the basis of uncertainty and diversity criteria [13]. The uncertainty criterion aims at selecting the unlabeled samples that have maximum uncertainty about their correct label among all samples in the unlabeled sample pool (and thus are the most useful to be included in the training set), whereas the diversity criterion aims at selecting a set of unlabeled uncertain samples that are as more diverse as possible to reduce the redundancy among the samples selected by the uncertainty criterion. In the MCLU-ECBD technique, the most uncertain samples are initially selected by the Multiclass-Level Uncertainty (MCLU) strategy that assesses the uncertainty of unlabeled samples on the basis of their functional distances to the decision boundaries of the binary SVM classifiers [13]. After the MCLU step, the most uncertain samples are analyzed by the *k*-means clustering method applied in the kernel space. Since the samples within the same cluster are correlated and provide similar information, one sample (the most uncertain one) from each cluster is selected and added to the training set at each iteration of AL.

We carried out two kinds of experiments in order to assess the effectiveness of the proposed CDTL approach. In the first set of trials the usefulness of the change-detection-driven TL step, which automatically defines an initial training set T_2 for $\mathbf{X}_2$ without any need for manual labeling of samples, was assessed. To this end the results obtained by the proposed CDTL approach are compared with those obtained with an AL technique that neglects the TL step. Since in this case

no labels are propagated from $\mathbf{X}_1$, the initial training set T_2 for $\mathbf{X}_2$ is empty. However, MCLU-ECBD technique needs an initial training set including labeled samples to train the classifier before starting with AL iterations. To this end, random sampling (RS) is applied to the samples of the image $\mathbf{X}_2$ and two samples related to each land-cover class are randomly selected and labeled to define an initial training set T_2. It is worth nothing that the RS step required for the initialization of the training set T_2 implies an additional cost of sample labeling whereas the proposed system does not.

In the second set of trials, we tested the effectiveness of the change-detection-based AL step by analyzing the improvements introduced by the Priority AL mechanism. This is done by assessing the CDTL approach with the priority rule and without it. For both data sets, in the experiments the Priority AL step is only performed at the first iteration of the AL process.

V. EXPERIMENTAL RESULTS

A. Results on the Sardinia Data Set (Scenario 1)

In order to apply the proposed procedure, first CVA has been applied to Sardinia images $\mathbf{X}_1$ and $\mathbf{X}_2$ (see Table I, Scenario 1). In this case, the changed area is associated to a reduction of the lake surface, and thus to a transition from water to bare soil. Note that the class transition is not known a priori, but automatically identified by the proposed approach according to the AL step. After performing change detection, the class labels of unchanged pixels are propagated to $\mathbf{X}_2$, providing 128 initial training samples for T_2. Once label propagation is completed the statistical analysis of the changed area is performed. This is done by measuring all the pairwise *JM* distances between the changed pixels in $\mathbf{X}_2$ and all the classes already present in T_2. Since all the pairwise *JM* distances are found higher than the threshold $Th = 0.99$, a new class (or a spatially shifted class with a different spectral signature compared to the same class already present in T_2) is detected.

Accordingly, the Priority AL step is performed at the first iteration and the human expert assigned "bare soil" label to the selected unlabeled samples. Thus the high values of the *JM* distances were associated to the presence of a new class. Starting from the second iteration, the Standard AL step is performed including both changed and unchanged samples in the pool. Table II shows the number of initial training samples (which are obtained on the basis of the CVA driven TL), and of changed and unchanged unlabeled samples in the pool set available for each land-cover class in $\mathbf{X}_2$. In addition also the number of test samples, which are used for accuracy assessment, is given in the table.

TABLE II. INITIAL TRAINING, POOL AND TEST SETS FOR $\mathbf{X}_2$ IMAGE (SARDINIA DATA SET, SCENARIO 1).

Land-cover classes	Initial Training Set	Pool Set		Test Set
		Changed Samples	Unchanged Samples	
Pasture (ω_1)	28	-	526	589
Forest (ω_2)	15	-	289	274
Urban area (ω_3)	20	-	388	418
Water body(ω_4)	56	-	748	551
Vineyard (ω_5)	9	-	170	117
Bare soil (ω_6)	-	316	-	316
Total	128	316	2121	2265

In the first set of experiments, we assess the effectiveness of change-detection-driven TL step by comparing the results of proposed approach with those obtained when T_2 is empty and populated according to only an AL procedure neglecting TL. Fig. 7 shows the average (on 10 trials) classification accuracies on $\mathbf{X}_2$ versus the number of new labeled samples obtained by i) the proposed CDTL approach, and ii) the RS method. In both cases the MCLU-ECBD technique is used in the AL step. Note that the new labeled samples are the patterns labeled by the human expert during the considered iteration of the AL process. From the figure, one can see that without labeling any new sample, the accuracy is zero in the case of using RS, whereas that is 77.64% in the case of using the proposed CDTL approach. This is of course due to the fact that the initial training set without labeling any new sample is not empty if the CDTL approach is used.

Moreover, by analyzing the figure one can observe that the CDTL provides higher accuracies than the RS and reaches convergence with a smaller number of new labeled samples. As an example, the accuracy of CDTL is 88.57% with only 12 new labeled samples, whereas that of RS is only 73.35% (see Fig. 7.a that refers to *b*=4). These results demonstrate that the proposed approach significantly reduces the labeling cost, thanks to the label-propagation step of the proposed algorithm defined on the basis of change-detection-driven TL.

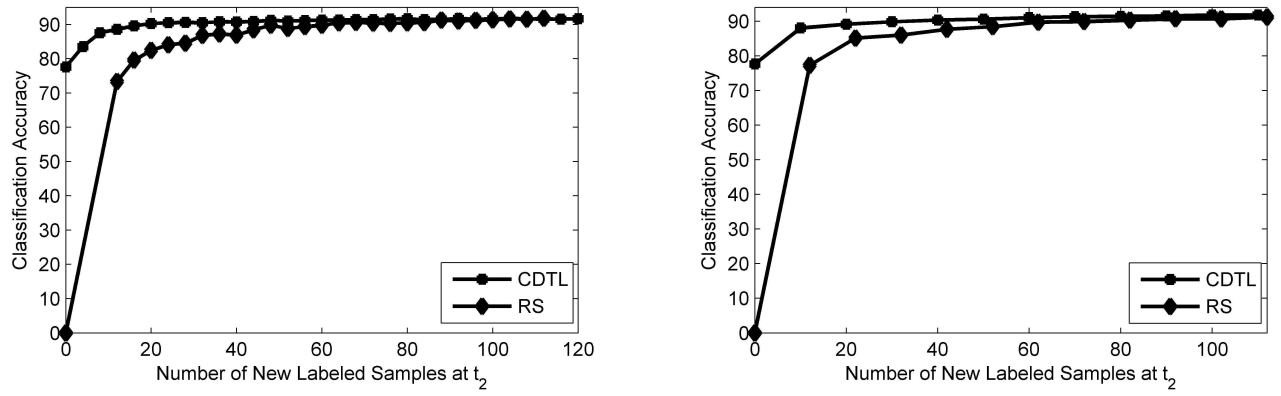


Fig. 7. Average (on 10 trials) overall classification accuracy versus the number of new labeled samples at t_2 obtained by the CDTL and the RS approaches when (a) *b*=4 and (b) *b*=10 (Sardinia data set, Scenario 1).

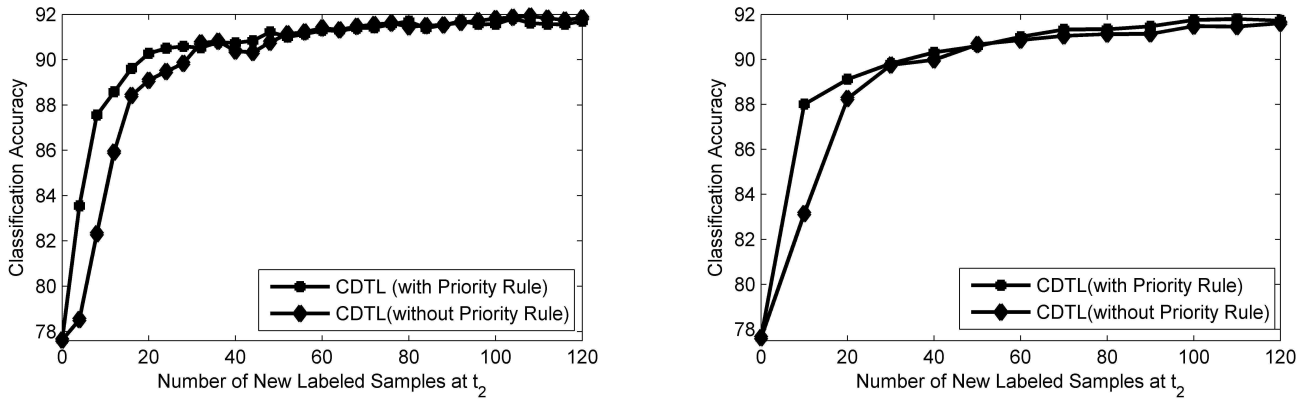


Fig. 8. Average (on 10 trials) overall classification accuracy versus the number of new labeled samples at t_2 obtained by the proposed CDTL with and without the Priority AL step in the cases of (a) *b*=4 and (b) *b*=10 (Sardinia data set, Scenario 1).

In the second set of experiments, we assess the effectiveness of the Priority AL procedure in the step 2 of the proposed method. Fig. 8 shows the average (on 10 trials) classification accuracies

of $\mathbf{X}_2$ versus the number of new labeled samples obtained by the proposed CDTL approach with and without the Priority AL step. From the results one can observe that the use of the Priority AL step provides higher accuracies at the early iterations, whereas it leads to similar accuracies at the remaining iterations. This is due to the ability of the MCLU-ECBD technique in selecting the most informative samples from the new class. However, at the early iterations, the Priority AL provides a faster improvement in the accuracy.

B. Results on the Sardinia Data Set (Scenario 2)

In order to better assess the effectiveness of the proposed approach, a more complex situation in which two new classes are present in the $\mathbf{X}_2$ (Scenario 2) is considered. Similarly as before, the CVA has been applied to images $\mathbf{X}_1$ and $\mathbf{X}_2$, and resulted in detecting two kinds of change. The first kind of change is related to the transition between bare soil and water on the lake area (as in Scenario 1), whereas the second one is related to the presence of the burned area on a forest area. Note that this information is assumed to be unknown. After performing CVA, the class labels of unchanged pixels in $\mathbf{X}_1$ are propagated to $\mathbf{X}_2$, providing 103 initial training samples for T_2. Then, the *JM* distances between the detected changed pixels in $\mathbf{X}_2$ and all the classes already present in T_2 are calculated and found higher than the threshold $Th = 0.99$. The process has been carried out independently for each detected kind of change. Based on these results, the presence of two new classes (or of spatially shifted classes) is identified. Accordingly, initial training set is expanded by applying the Priority AL step at the first iteration and applying Standard AL at the remaining ones. Since there are two kinds of change, in the Priority AL step (i.e., at the first iteration of AL) $b/2$ most uncertain unlabeled samples are chosen from one unknown class and $b/2$ most uncertain unlabeled samples are selected from the other one looking at the two annular sectors in the CVA polar domain. In the labeling process of AL, the human expert assigned “bare soil” and “burned area” labels to the selected unlabeled samples. Thus the high values of the *JM* distances were

associated to the presence of new classes. Table III shows the number of initial training samples obtained on the basis of the CVA driven TL, changed and unchanged unlabeled samples in the pool set, and test samples available for each land-cover class in $\mathbf{X}_2$.

TABLE III. INITIAL TRAINING, POOL AND TEST SETS FOR $\mathbf{X}_2$ IMAGE (SARDINIA DATA SET, SCENARIO 2).

Land-cover classes	Initial Training Set	Pool Set		Test Set
		Changed Samples	Unchanged Samples	
Pasture (ω_1)	28	-	526	589
Forest (ω_2)	6	-	122	159
Urban area (ω_3)	20	-	388	418
Water body(ω_4)	40	-	764	551
Vineyard (ω_5)	9	-	170	117
Bare soil (ω_6)	-	316	-	316
Burned area (ω_7)	-	176	-	115
Total	103	492	1970	2265

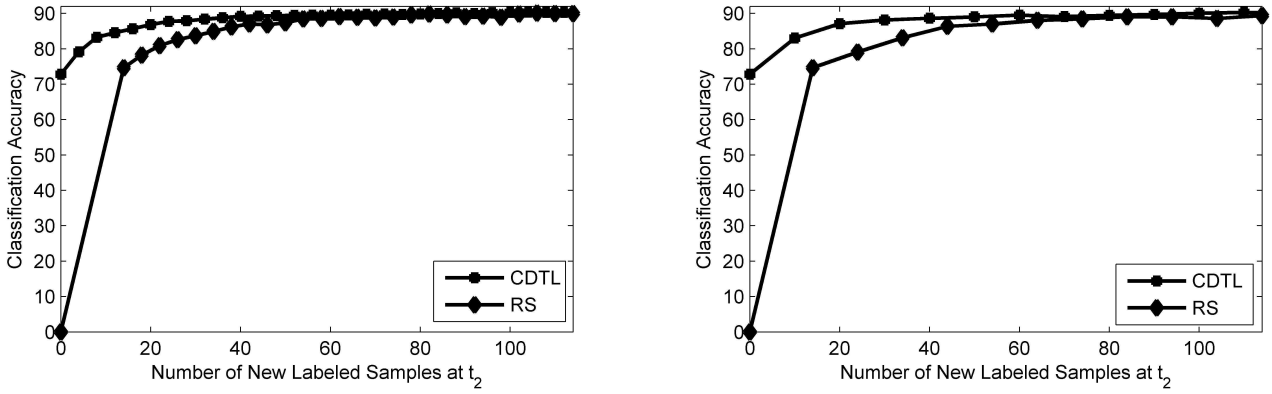
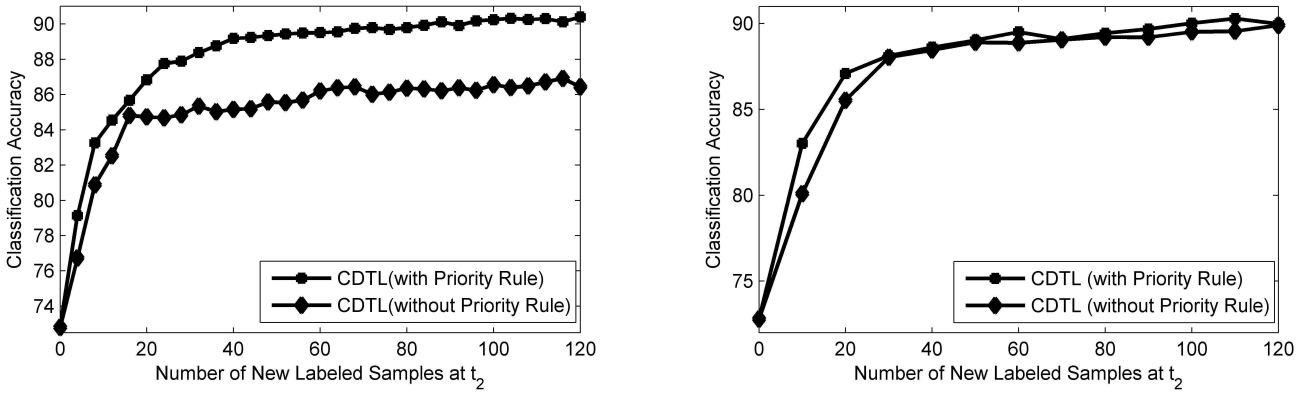


Fig. 9. Average (on 10 trials) overall classification accuracy versus the number of new labeled samples at t_2 obtained by the CDTL and the AL approaches when (a) *b*=4 and (b) *b*=10 (Sardinia data set, Scenario 2).

Fig. 9 shows the behavior of the average (on 10 trials) overall accuracies obtained by the CDTL and the RS (which ignores the TL step and therefore assume that an initial training set is not available) in the cases of *b*=4 (see Fig. 9.a) and *b*=10 (see Fig. 9.b). By analyzing the figure, one can observe that the CDTL results in higher accuracies for both values of *b*. Moreover, it reaches convergence with a smaller number of new labeled samples, thanks to the change-detection-driven TL step. As an example, from the Fig. 9.a one can observe that the CDTL yields

an accuracy of 79.13% with only 4 new labeled samples at t_2, whereas the RS reaches a similar accuracy with 22 new labeled samples. It is worth nothing that the effectiveness of the proposed approach compared to RS is significant due to its ability in defining the initial training set T_2.

Fig. 10 shows the comparison of the average overall accuracy obtained by the CDTL approach with and without the Priority AL step for both values of *b*. From these graphs, one can see that the CDTL with priority rule, provides again the selection of more informative samples compared to the CDTL without priority rule, and achieves higher accuracies for the same number of labeled samples (thus the same accuracy with less samples). As an example, the CDTL with priority rule provides the highest accuracies at each iteration and also converges with a small number of labeled samples when *b*=4 (see Fig. 10.b). In the case of *b*=10, the CDTL with priority rule results in better accuracies at the early iterations resulting in a faster improvement of the accuracy, whereas the results are similar to those of the CDTL without priority rule at the remaining iterations. This is due to the fact that MCLU-ECBD is a very effective AL technique for the selection of the most informative samples, and thus after some iterations the new appeared classes are sufficiently represented in the training set.



(a) (b)

Fig. 10. Average (on 10 trials) overall classification accuracy versus the number of new labeled samples at t_2 obtained by the proposed CDTL with and without the Priority AL step when using RS in the AL step in the cases of (a) *b*=4 and (b) *b*=10 (Sardinia data set, Scenario 2).

C. Results on the Trento Data Set

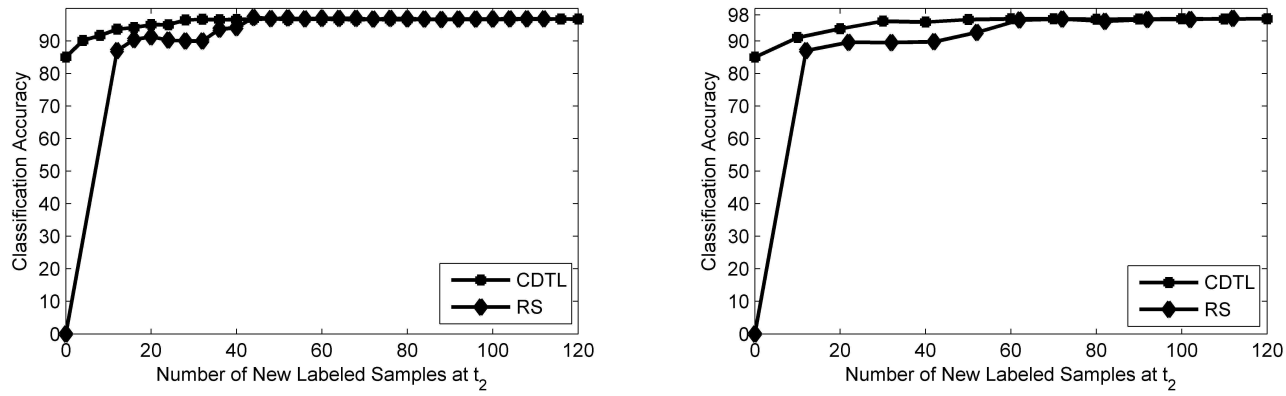
Also for the Trento data set the CVA technique has been applied to images $\mathbf{X}_1$ and $\mathbf{X}_2$, and the class labels of training pixels in $\mathbf{X}_1$ detected as unchanged are transferred to $\mathbf{X}_2$ providing 170 initial training samples for T_2. Once label propagation is completed, the statistical analysis of the changed areas is performed. On the basis of this analysis, it is observed that the detected changed pixels in $\mathbf{X}_2$ have high *JM* distances to all the classes present in T_2. This shows the presence of a new class (or of a spatially shifted class with a significantly different spectral signature from that modeled in the initial T_2). Accordingly, the Priority AL step is applied and in the labeling process of AL the human expert assigned "plastic-mulched fields" label to the selected unlabeled samples. Thus the high values of the *JM* distances are associated to the presence of a new class. From the second iteration, Standard AL is performed involving both changed and unchanged samples. Table IV lists the number of initial training samples obtained by the CVA driven TL, the changed and unchanged unlabeled samples in the pool set, and the test samples available for each land-cover class in $\mathbf{X}_2$.

TABLE IV. INITIAL TRAINING, POOL AND TEST SETS FOR $\mathbf{X}_2$ IMAGE (TRENTO DATA SET).

Land-Cover Classes	Initial Training Set	Pool Set		Test Set
		Changed Samples	Unchanged Samples	
Water (ω_1)	52	-	986	1104
Red Roof (ω_2)	22	-	427	469
Asphalt (ω_3)	34	-	438	474
Fields (ω_4)	32	-	615	534
Bare soil in field (ω_5)	30	-	570	483
Plastic-mulched field(ω_6)	-	471	-	290
Total	170	417	3236	3354

Fig. 11 shows the comparison of the average overall accuracy obtained by the CDTL and the RS for both values of *b*. Also for the Trento data set the CDTL resulted in higher accuracies at most of the iterations, and also converged with a smaller number of labeled samples. For example,

the CDTL yields an accuracy of 93.55% by adding only 12 labeled samples to T_2, whereas the RS provides an accuracy of 87.% with the same number of labeled samples.



(a) (b)

Fig. 11. Average (on 10 trials) overall classification accuracy versus the number of new labeled samples at t_2 obtained by the CDTL and the RS approaches when (a) *b*=4 and (b) *b*=10 (Trento data set).

The comparison of the average overall accuracy provided by the CDTL with and without the Priority AL step when the MCLU-ECBD technique is used in the AL step is not reported here as it is very similar to what already presented on the previous data sets.

VI. DISCUSSION AND CONCLUSION

In this paper, we have presented a novel change-detection-driven transfer learning approach (CDTL) for updating land-cover maps by classifying image time series. The proposed approach classifies an image for which no ground truth information is available (target domain) by using the knowledge available for an image acquired on the same area of interest at a different time (source domain). The CDTL approach overcomes the two main drawbacks of other techniques presented in the remote sensing literature: i) it is not affected from the possible significant differences between the land-cover class distributions of the source and the target domains, and ii) it is able to handle situations in which different sets of land-cover classes may characterize the two domains. Moreover, the proposed approach allows one to significantly reduce the complexity and the cost of the collection of labeled samples for the target domain.

The CDTL approach is defined on the basis of three steps. The first step is devoted to define an initial training set for the target domain. This is done by firstly applying unsupervised change detection to source and target domains, and then transferring the class-labels of unchanged training samples from source domain to target domain. Thus, the initial training set for target domain is inherited from the source domain. Because of the label propagation of source training patterns, the classifier is directly trained on the original samples of the target domain, and thus there is no need to adapt the classification parameters of the source domain to target domain. Therefore, the dissimilarity between the land-cover class distributions of source and target domains does not affect the proposed method. Moreover this step significantly reduces the number of new labeled samples to be collected at target domain for optimizing the classification results. In addition, this step intrinsically rejects possible disappeared classes in the target domain as it does not transfer labels associated to changed training samples. The second step aims at optimizing the initial training set by Active Learning (AL) giving a priority to samples detected as being changed at the first iterations. This is because they have a high probability to be most informative among all the other unlabeled samples. Therefore this choice allows one to increase rapidly the classification accuracy by labeling few specific unlabeled samples of the target domain, and thus to optimize the classification accuracy. In the remaining iterations priority is removed, and AL is applied in a standard way. Thanks to the AL step, the proposed approach does not have limitations on the sets of land-cover classes that characterize the two domains. When the AL process is completed, in the third step the target image is classified to obtain the classification map.

Experimental results obtained on two multitemporal data sets show that: i) the change-detection-driven TL step generates reliable initial training set for the target domain reducing the cost of labeled sample collection to the minimum; and ii) the change-detection-driven AL with Priority AL rule provides a faster convergence to the desired accuracy, with respect to standard AL. In greater details, in most of the cases we achieved high accuracy on the target image by

adding few (10-20) labeled samples for the target domain. From a different perspective, if no changes occurred between the two considered images, the proposed method can provide accurate classification results also without labeling new samples for the target domain.

It is worth nothing that the proposed approach is very promising for possible operational applications due to both its general properties and its simplicity in the implementation. As a final remark, we point out that the proposed method can be easily applied to long time series by extracting different pairs of source and target images from the series and applying an iterative pairwise analysis.

REFERENCES

[1] L. Bruzzone, and D. Fernandez Prieto, "Unsupervised retraining of a maximum-likelihood classifier for the analysis of multitemporal remote-sensing images", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 456–460, 2001.

[2] L. Bruzzone, and D. Fernandez Prieto, "A partially unsupervised cascade classifier for the analysis of multitemporal remote-sensing images", *Pattern Recognition Letters*, vol. 23, no.9, pp. 1063-1071, 2002.

[3] L. Bruzzone and R. Cossu, "A multiple cascade-classifier system for a robust a partially unsupervised updating of land-cover maps", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 1984-1996, 2002.

[4] L. Bruzzone, M. Marconcini, "Domain Adaptation Problems: a DASVM Classification Technique and a Circular Validation Strategy," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 770-787, 2010.

[5] S. Rajan, J Ghosh, and M Crawford, "Exploiting class hierarchies for knowledge transfer in hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol.44, no. 11, pp. 3408-3417, Jan 2006.

[6] K. Bahirat, F. Bovolo, L. Bruzzone and S. Chaudhuri, "A novel domain adaptation maximum likelihood classifier for updating land-cover maps in complex scenarios," *SPIE Conference on Image and Signal Processing for Remote Sensing*, Toulouse, France, Sept. 2010.

[7] S. Rajan, J. Ghosh, and M. M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 4, pp. 1231-1242, Apr. 2008.

[8] C. Persello and L. Bruzzone, "A novel active learning strategy for domain adaptation in the classification of remote sensing images," *IEEE International Geoscience and Remote Sensing Symposium*, Vancouver, Canada, 2011.

[9] S. J. Pan, and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no.10, pp.1345-1359, Oct. 2010.

[10] E. W. Xiang, B. Cao, D. H. Hu, and Q. Yang, "Bridging domains using worldwide knowledge for transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 6, June 2010

[11] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang "Domain Adaptation via Transfer Component Analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, Feb. 2011.

[12] P. Mitra, B. U. Shankar, and S. K. Pal, "Segmentation of multispectral remote sensing images using active support vector machines," *Pattern Recognit. Lett*., vol. 25, no. 9, pp. 1067–1074, Jul. 2004.

[13] B. Demir, C. Persello, and L. Bruzzone, "Batch mode active learning methods for the interactive classification of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no.3, pp. 1014-1031, March 2011.

[14] S. Patra and L. Bruzzone, "A fast cluster-based active learning technique for classification of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no.5, pp.1617-1626, 2011.

[15] A. Liu, G. Jun and J. Ghosh, "Spatially cost-sensitive active learning," *In SIAM International Conference on Data Mining (SDM)*, Sparks, Nevada, USA, pp.814-825, 2009.

[16] A. Liu, G. Jun, and J. Ghosh, "Active learning of hyperspectral data with spatially dependent label acquisition costs," *IEEE International Geoscience and Remote Sensing Symposium*, Cape Town, South Africa, pp. V-256 - V-259, 2009.

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

[18] F. Bovolo, L. Bruzzone, "A Theoretical Framework for Unsupervised Change Detection Based on Change Vector Analysis in the Polar Domain," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, No.1, pp. 218–236, Jan. 2007.

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

[20] L. Bruzzone, and D. Fernandez Prieto, "Automatic analysis of the difference image for unsupervised change detection", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no.3, pp. 1171-1182, 2000.

[21] F. Bovolo, S. Marchesi, L. Bruzzone, "A Framework for Automatic and Unsupervised Detection of Multiple Changes in Multitemporal Images", *IEEE Transactions on Geoscience and Remote Sensing*, in press.

[22] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[23] F.Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[24] S. Marchesi, F. Bovolo, and L. Bruzzone, "A Context-Sensitive Technique Robust to Registration Noise for Change Detection in VHR Multispectral Images," *IEEE Transaction on Image Processing*, vol. 19, no. 7, pp. 1877-1889, 2010.

[25] J. A. Richards and X. Jia, Remote Sensing Digital Image Analysis, 4th ed. Berlin, Germany: Springer-Verlag, 2006.

[26] E. Alpaydın, *Introduction to Machine Learning*, MIT Press, Cambridge, Massachusetts, London, England, 2004.

[27] L. Bruzzone, F. Roli, and S.B. Serpico, "An extension of the Jeffreys-Matusita distance to multiclass cases for feature selection", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 6, pp. 1318-1321, Nov. 1995.