

© 2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Title: Classification of Time Series of Multispectral Images with Limited Training Data

This paper appears in: IEEE Transactions on Image Processing

Date of Publication: 2013

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

Volume:22, Issue: 8

Page(s): 3219 - 3233

DOI: 10.1109/TIP.2013.2259838

# Classification of Time Series of Multispectral Images with Limited Training Data

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

**Abstract**— Image classification usually requires the availability of reliable reference data collected for the considered image to train supervised classifiers. Unfortunately when time series of images are considered, this is seldom possible because of the costs associated with reference data collection. In most of the applications it is realistic to have reference data available for one or few images of a time series acquired on the area of interest. In this paper we present a novel system for automatically classifying image time series which takes advantage of image(s) with an associated reference information (i.e., the source domain) to classify image(s) for which reference information is not available (i.e., the target domain). The proposed system exploits the already available knowledge on the source domain and, when possible, integrates it with a minimum amount of new labeled data for the target domain. Moreover it is able to handle possible significant differences between statistical distributions of the source and target domains. Here the method is presented in the context of classification of remote sensing image time series, where ground reference data collection is a highly critical and demanding task. Experimental results show the effectiveness of the proposed technique. The method can work on multimodal (e.g., multispectral) images.

**Index Terms**— Time Series, Cascade Classification, Transfer Learning, Active Learning, Automatic Classification, Land-Cover Maps, Remote Sensing

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).  
Authors are with the Dept. of Information Engineering and Computer Science, University of Trento, Via Sommarive, 14, I-38123 Trento, Italy. e-mail: [demir@disi.unitn.it](mailto:demir@disi.unitn.it), [francesca.bovolo@disi.unitn.it](mailto:francesca.bovolo@disi.unitn.it), [lorenzo.bruzzone@ing.unitn.it](mailto:lorenzo.bruzzone@ing.unitn.it)

## 1 INTRODUCTION

Observing time evolution of phenomena is a hot topic in many fields, e.g., economy, ecology, medical science, geology, etc. The knowledge on phenomena evolution has strong spin-off for daily life. Most of the phenomena in the mentioned fields are described by 1-dimensional variables [e.g., electroencephalogram (EEG) signals; electrocardiogram (ECG) signals; river water level; temperature, humidity, economical trend variables, and so on]. In the last decades, another set of multitemporal signals became relevant in the scientific community, which are 2-dimensional signals (i.e., images). Biomedical images (e.g., images acquired by radiographs, functional Magnetic Resonance Imaging, etc.) and geoscience and remote sensing images (e.g., images acquired by active and passive sensors on board of airborne, spaceborne platforms, etc.) fall mainly into this group. The possibility to have image time series (temporal images taken from the same scene at different time periods) drove the interest of the scientific community to the development of effective methods for information extraction from this kind of data. Analysis of time series in both biomedical and remote sensing fields by automatic and reliable techniques allowed scientists to improve both their data comprehension and the accurate detection of possible anomalies in a short time. These achievements are mainly due to the capability of methods specifically developed for time series analysis to exploit temporal correlation. The outputs of such data processing are highly relevant in many anthropic- and environmental-related fields.

In biomedicine image time series play an important role for monitoring the disease evolution. In [1] time series of chest radiographs have been used for monitoring the pneumonia evolution. In [2] RGB color images have been used for analyzing the spreading of skin erythema. Breast magnetic resonance imaging has been exploited in [3] for investigating the evaluation of small mammographic lesions. Time series of magnetic resonance imaging has been used to study also changes in neuronal activities [4].

Color retinal fundus image sequences has been analyzed in [5] for the detection of possible lesions. Most of the cited works deals with the detection of changes occurred from one acquisition to the next one. In biomedicine literature only few works approach time series analysis from the classification perspective [1].

The remote sensing literature on the classification of image time series is more extended. This is mainly due to the repeat-pass nature of satellite orbits that results in intrinsic multitemporal remote sensing image sequences. The availability of huge amounts of such data pushed research into the development of data analysis techniques for image time series. Time series of remote sensing images are beneficial for detecting land-cover transitions occurred on the ground and updating land-cover maps, which are very important processes for regularly monitoring the Earth surface. Such information is highly relevant to support and improve environmental management policies. Because of the increasing number of images regularly acquired at different times on the same area, and their possible free availability (e.g., Landsat Thematic Mapper archive, future ESA Sentinel mission), it is important to develop methods to update land-cover maps with both high accuracy and low cost.

Classification maps can be updated by automatically classifying each image in the time series, which requires the availability of reference samples for each considered image to train the supervised classifier. However gathering such data is highly costly in terms of human time and effort, and thus also in terms of money. Moreover, often it is impossible from the practical viewpoint to collect reference training information for a given acquisition in the series. As an example, there are remote sensing applications that require the classification of a time series of historical images for which ground truth is not available and can be not collected in a retrospective way. Thus classifying a time series assuming that a sufficient number of ground reference samples is available for each acquisition is unfeasible.

In order to deal with this issue and generate a classification map for the desired image for which no prior information is available (target domain), there is a need to find out a way to use reference samples already available for an image acquired on the same scene at a different time (source domain). The simplest solution would be to use the classifier directly trained on the source domain to classify the target domain. However in most of the cases this does not provide reliable classification results, because source and target domains may differ from each other due to differences in the acquisition conditions (e.g., acquisition geometry and illumination differences) [20]. Thus reference samples of the source domain may not follow the same distribution of the samples in the target domain making the classifier trained on them unreliable for the target domain. To overcome this problem, domain adaptation (DA) methods in transfer learning (TL) have been recently introduced in the literature [6]-[8]. DA methods aim at classifying the target domain (for which no ground information is available) by exploiting the information available on the source domain, assuming that the two domains may have different, but strongly related, distributions. In the literature, DA is known also as partially supervised/unsupervised learning and is addressed with Semi-Supervised Learning (SSL) methods [9]-[13] or Active Learning (AL) methods [14]-[17]. On the one hand, SSL applies a classifier trained on the source domain to the target domain after tuning the parameters according to unlabeled data from the target domain [9]-[13]. In other words, the information of reference training samples from the source domain is improved by costless unlabeled samples from the target domain to obtain a reliable classifier for the target domain. On the other hand, AL methods aim at improving (from the target domain point of view) the information of the source domain reference samples by iteratively adding samples selected from the target domain [14]- [17]. Before inclusion in the training set these samples should be manually labeled by a human expert, thus these methods have associated a cost that SSL techniques do not have. However AL

methods try to reduce it by labeling the smallest possible number of unlabeled samples. This is achieved by selecting for labeling those samples that are the most informative for the target domain viewpoint, thus avoiding the huge cost of collecting large amount of labeled samples. In the literature examples of domain adaptation methods based on both SSL and AL are available. For example in [9] DA problems are addressed with SSL by updating on the basis of the distribution of the target domain the parameters of a parametric maximum-likelihood (ML) classifier already trained on the source domain. This method has been generalized in the context of the Bayes rule for cascade classification in [10] in order to exploit the temporal correlation between domains. Further improvements of this method are presented in [11] and [12]. A multiple cascade classifier system is proposed in [11] including ML and neural-network classifiers. In [12], the sets of classes of the target and the source domains are automatically analyzed in the DA step by the joint use of unsupervised change detection and Jeffreys-Matusita statistical distance measure. This process results in the detection of classes that appeared or disappeared between the domains. Despite the method presented in [12] can manage possible class differences, the values of the statistical parameters modeling classes propagated from the source to the target domain are still biased. SSL-DA-based methods have been developed also in the context of Support Vector Machine (SVM) classifiers. In [13], the initial discriminant function is estimated on the source domain labeled samples. The adaptation to the target domain is achieved by iteratively including in the training set unlabeled patterns of the target domain that have a high probability to be correctly classified. Simultaneously labeled samples of the source domain are gradually removed. In [14]- [17], DA problems are addressed with AL and thus, unlike SSL-based methods [9]-[13], a small number of labeled training samples is included in the target domain together with the labeled samples of the source one. In [14], the classification parameters are initialized by the distributions estimated on the labeled samples of the

source domain. Then the unlabeled samples of the target domain that have the maximum information gain (measured by the Kullback–Leibler divergence) are included in the training set of the target domain after manual labeling. In [15], the statistical parameters of a ML classifier are initialized by exploiting the labeled samples of the source domain, and the most informative samples are iteratively selected from the target domain by AL to be added to the training set after manual labeling. In this method, during the AL process, the source domain samples that do not fit with the distribution of the classes in the target domain are removed. In [16], the authors show how AL applied to the target domain can leverage information from the source domain. This is achieved based on two steps: i) applying any of the DA methods from the literature to adapt the source domain labeled samples to target domain, and ii) exploiting AL in the target domain. In this work, the domain separator hypothesis is introduced for the selection of target domain samples being queried in the AL step. Accordingly, the target domain samples that are not similar to source domain are selected to be queried for better modeling the target domain.

All the above mentioned DA methods initialize the parameters of the classifier to be applied to the target domain (i.e., the image to be classified) exploiting the labeled samples of the source domain. The adaptation to the target domain starts therefore from biased estimates. Thus if class statistical distributions in the target domain differ significantly from those of the source domain, the adaptation step cannot fully compensate for them. In these cases DA methods presented in the literature will provide low classification performance on the target domain. Therefore, it is necessary to develop new TL methods that are not significantly affected from the distribution differences in classes between the domains, thus providing good classification accuracy also in critical conditions. In order to alleviate the domain difference risk, in [17] a framework to actively transfer knowledge is presented, which assumes that the target domain also includes a small set of labeled training samples in addition to the source

domain. In this framework, initially a transfer classifier is constructed to be applied to the unlabeled samples of the target domain. The classifier is defined by using the training samples of both the source and the target domains [17]. Then, if an unlabeled sample is assigned to a class with low confidence, a domain expert will assign the correct class label to that sample. Even if this framework can reduce the domain difference risk, it has the drawback of assuming the availability of an initial training set for the target domain.

To overcome the limitations of the DA methods available in the literature, we present a novel system for automatically updating classification maps by using image time series. The main idea of the proposed system is to classify the target domain by defining a method that: i) effectively reuses the available information on the source domain by mitigating possible bias effects; ii) labels the smallest possible number of unlabeled samples from the target domain for optimizing the classification accuracy; and iii) exploits the temporal correlation between the domains in the classification process. To this end, the proposed system is defined on the basis of two steps: i) low-cost definition of a training set for the target domain with transfer and active learning methods; and ii) target domain classification according to the Bayesian cascade decision rule. The first step properly uses the information available on the source domain to initialize the training set for the target domain without labeling any new sample. Source domain training samples with a high probability to have the same label also in the target domain are detected by applying unsupervised change detection to target and source images. Only class labels of detected unchanged training samples are propagated from the source to the target domain. Thanks to this choice, unlike the DA methods proposed in the literature, the classifier parameters of the target domain are estimated directly on the target domain samples that inherited a label from the source domain (i.e., statistics are not computed on samples from the source domain). Thus the proposed system is not

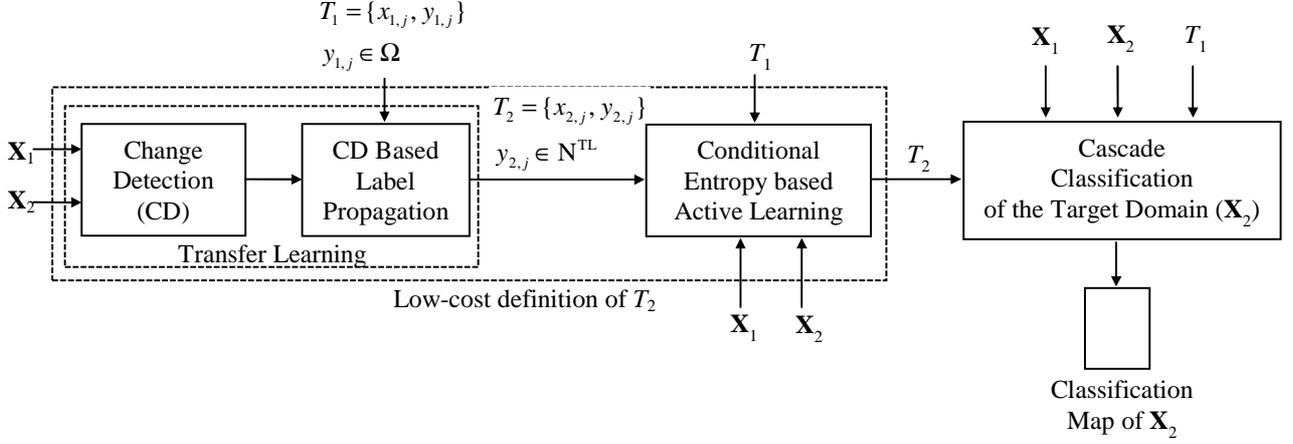
affected from the distribution differences between the domains. Then, if necessary, active learning is used to expand the initial target-domain training set by considering the temporal correlation between the domains. This is achieved by a novel AL technique defined by conditional entropy for the selection of most informative unlabeled samples in the image to be included in the training after manual labeling. In the second step, the target domain is classified by a cascade classifier defined in the framework of the Bayesian decision rule. This choice allows us to exploit the temporal correlation between the images (domains) also in the classification step.

The paper is organized into five sections. Section II defines the considered problem and describes the proposed system. Section III illustrates the considered data sets and the design of experiments. Section IV shows the experimental results. Finally, Section V draws the conclusion of this work.

## 2 A NOVEL SYSTEM TO CLASSIFICATION MAP UPDATING

Let  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_P]$  be a time series that consists of  $P$  co-registered images acquired at different times on the same scene. Let us assume that at least one of the images in the time-series has a reliable training set. It does not matter whether this is an earlier or older image. The goal of the proposed method is to classify the desired images in the time series even if no reference information is available for them. In order to simplify the mathematical treatment, we consider only a pair of images from the series (i.e., one for the source domain and one for the target domain). However, the proposed method can be easily applied to long time series by iteratively processing pairs of images from the series. Let  $\mathbf{X}_1 = \{x_{1,1}, x_{1,2}, \dots, x_{1,B}\}$  and  $\mathbf{X}_2 = \{x_{2,1}, x_{2,2}, \dots, x_{2,B}\}$  be two multimodal (multidimensional) images extracted from the time series  $\mathbf{X}$  acquired at times  $t_1$  and  $t_2$ , respectively. Both images include  $C$  channels and  $B$  pixels. Let  $(x_{1,j}, x_{2,j})$  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 us assume that the image  $\mathbf{X}_1$  is the source domain for which a reliable training set  $T_1 = \{x_{1,j}, y_{1,j}\}_{j=1}^M$  is available.  $x_{1,j} \in \mathbf{X}_1$  is the  $j$ -th training sample,  $y_{1,j} \in \Omega$  is the associated class label, and  $M < B$  is the number of training samples. The image  $\mathbf{X}_2$  is the target domain for which a training set  $T_2$  is not available. Let  $\Omega = \{\omega_1, \omega_2, \dots, \omega_R\}$  be the set of classes at time  $t_1$ , and  $\mathbf{N} = \{v_1, v_2, \dots, v_N\}$  be the set of classes at time  $t_2$ . We assume that different sets of classes may characterize the domains (i.e.,  $\Omega \neq \mathbf{N}$ ). The goal of the proposed system is to provide a classification map of the image  $\mathbf{X}_2$  by exploiting both the training samples available on the image  $\mathbf{X}_1$  and the temporal correlation between  $\mathbf{X}_2$  and  $\mathbf{X}_1$ , and by minimizing the labeling costs associated to the definition of  $T_2$  (which at the limit can be null). Here this is achieved by a novel system based on two steps: i) low-cost definition of  $T_2$ ; and ii) cascade-classification of the image  $\mathbf{X}_2$ . The first step takes advantage of both TL and AL methodologies, and includes a novel AL technique for the selection of the most uncertain samples. The second step classifies  $\mathbf{X}_2$  by taking into account the temporal correlation between the images in the considered pair thanks to the use of cascade classification. As the definition of  $T_2$  strictly depends on the selected cascade classifier, the detailed explanation of the proposed system starts from the last step and moves backward. Fig. 1 shows the block scheme of the proposed system.



**Fig. 1.** Block diagram of the proposed system.

#### A. Cascade Classification under Bayesian Decision Framework

The Bayesian decision rule for cascade classification identifies the class label to be assigned to each pixel  $x_{2,j} \in \mathbf{X}_2$  by considering the temporal dependence of  $\mathbf{X}_2$  from  $\mathbf{X}_1$  [10], *i.e.*,

$$x_{2,j} \in v_n \text{ if } v_n = \arg \max_{v_k \in N} \{P(v_k | x_{1,j}, x_{2,j})\} = \arg \max_{v_k \in N} \{p(x_{1,j}, x_{2,j} | v_k)P(v_k)\} \quad (1)$$

where  $P(v_k | x_{1,j}, x_{2,j})$  is the probability that the  $j$ -th pixel  $x_{2,j}$  at  $t_2$  belongs to class  $v_k$ , given the two observations  $x_{1,j}$  and  $x_{2,j}$ , and  $P(v_k)$  is the prior probability of having class  $v_k$  at  $t_2$ .  $p(x_{1,j}, x_{2,j} | v_k)$  is a mixture density that depends on the distributions of classes at  $t_1$ , and can not be estimated without making this dependence explicit. To overcome this problem,  $P(v_k | x_{1,j}, x_{2,j})$  can be rewritten by highlighting its dependence on the class labels at time  $t_1$ :

$$P(v_k | x_{1,j}, x_{2,j}) = \frac{\sum_{\omega_i \in \Omega} p(x_{1,j}, x_{2,j} | \omega_i, v_k)P(\omega_i, v_k)}{\sum_{\omega_s \in \Omega} \sum_{v_r \in N} p(x_{1,j}, x_{2,j} | \omega_s, v_r)P(\omega_s, v_r)} \quad (2)$$

where  $p(x_{1,j}, x_{2,j} | \omega_i, v_k)$  is the joint class-conditional density function. The estimation of the statistical quantities in (2) is a complex task due to the difficulty in collecting enough training samples for properly modeling the multitemporal correlation between all possible temporal combinations of classes [10]. Therefore, as often done in practical applications [10]-[11],[18],[19], we assume the conventional class-conditional independence in the time domain (*i.e.*,  $p(x_{1,j}, x_{2,j} | \omega_i, v_k) = p(x_{1,j} | \omega_i) p(x_{2,j} | v_k)$ ) to simplify the estimations process. Under this assumption (1) can be rewritten as

$$x_{2,j} \in v_n \text{ if } v_n = \arg \max_{v_k \in N} \left\{ \frac{\sum_{\omega_i \in \Omega} p(x_{1,j} | \omega_i) p(x_{2,j} | v_k) P(\omega_i, v_k)}{\sum_{\omega_s \in \Omega} \sum_{v_r \in N} p(x_{1,j} | \omega_s) p(x_{2,j} | v_r) P(\omega_s, v_r)} \right\} \quad (3)$$

where  $p(x_{1,j} | \omega_i)$  and  $p(x_{2,j} | v_k)$  are the single-date class-conditional density functions, and  $P(\omega_i, v_k)$  is the joint prior probability of having classes  $\omega_i$  at  $t_1$  and  $v_k$  at  $t_2$ . The class parameters at time  $t_1$  can be estimated using the available training samples in  $T_1$ . However class parameters at time  $t_2$  and joint prior probabilities  $P(\omega_i, v_k)$  cannot be estimated directly due to the unavailability of the training set  $T_2$ .

In the literature, the estimation of these terms has been achieved from the images under investigation by an iterative procedure defined on the basis of the Expectation-Maximization (EM) algorithm [11]. In the iterative EM algorithm, the values of the parameters for the density functions of classes at time  $t_2$  are initialized by considering the corresponding values estimated at time  $t_1$ , whereas joint prior probabilities are initialized by assigning equal joint prior probabilities to each pair of classes (including possible specific constraints on some transitions according to prior information). However, the EM algorithm may not provide accurate estimates when the statistical distributions of classes in the source and target domains significantly differ from each other. As mentioned before, this may happen

due to the different acquisition conditions. Such differences may cause the initialization values of the parameters of the density functions for the classes at time  $t_2$  considerably different from the accurate values. Therefore EM algorithm may show difficulties in converging. To overcome this drawback, in the next section we propose a low-cost approach to the definition of a training set  $T_2$  and thus to the estimation of the mentioned statistical terms.

### B. *Low-cost Generation of Training Samples for the Target Image*

As mentioned in the previous sub-section there is a need of defining a reliable training set  $T_2$  for  $\mathbf{X}_2$  with a low-cost procedure that takes into account the temporal correlation between  $\mathbf{X}_1$  and  $\mathbf{X}_2$ . Here we propose a 2-step procedure which takes advantage of transfer and active learning methods.

#### *Step 1: Change-Detection-driven Transfer Learning (CDTL)*

The first step aims at defining an initial training set for the image  $\mathbf{X}_2$  (target domain) by reusing the already available knowledge from  $\mathbf{X}_1$  (source domain) (i.e., without the need of a new labeling process) and taking advantage of the temporal correlation. This is achieved by adopting the Change-Detection-driven Transfer Learning (CDTL) approach proposed in [20]. According to this approach, the labels of training samples in  $T_1$  are considered reliable for  $\mathbf{X}_2$  if the related pixels did not experience any change. In such a situation the label  $y_{1,j}$  associated to the training sample  $x_{1,j} \in T_1$  is transferred from the source domain to the target domain and associated to the sample  $x_{2,j} \in \mathbf{X}_2$  in the corresponding spatial position as the unchanged pixel  $x_{1,j}$ . In order to detect changed and unchanged pixels, any change-detection method proposed in the literature can be applied [21]-[25]. As an example, assuming that multidimensional images are acquired by passive sensors and thus are corrupted by additive Gaussian noise, Change Vector Analysis (CVA) technique can be used [21]-[23]. CVA technique

applies to  $\mathbf{X}_1$  and  $\mathbf{X}_2$  by subtracting the feature vectors of temporally correlated pixels  $x_{1,j}$  and  $x_{2,j}$  from each other in order to build a multispectral difference image  $\mathbf{X}_D$ . CVA assumes that multitemporal images should be co-registered 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 applied [25]. Here,  $\mathbf{X}_D$  is analyzed according to the theoretical framework for unsupervised change detection based on the CVA in polar domain proposed in [21]. According to [21], for simplicity two channels out of  $C$  are selected such that the most informative features with respect to the specific considered problem are isolated excluding noisy and misleading channels from the analysis. It is worth noting that, even if the assumption of working with a pair of channels is reasonable in many change-detection problems [21]-[24] the CVA can be also applied to more than 2 channels. To highlight change information the magnitude and the direction of change vectors in  $\mathbf{X}_D$  is computed. The defined 2-dimensional feature space is referred in the literature such as the polar representation of the change-detection problem [21]. In this feature space, unchanged pixels are distributed close to the origin of the polar domain and fall within the circle of no-changed pixels (i.e., they show a low magnitude value) [21], whereas changed pixels are distributed far from the origin. According to this behavior, a proper threshold computed on the magnitude feature allows one to separate changed from unchanged pixels. The reader is referred to [24] for a systematic survey on thresholding methodologies. The direction feature can be used for distinguishing among different kinds of change [21]. Once unchanged pixels have been identified, the labels of unchanged training samples are transferred to the target domain. These labels will be associated to samples in the target domain that spatially correspond to the training sample in the source domain. Labels that do not satisfy this constraint are not transferred.

At the end of this step, the set of classes  $\mathbf{N}^{TL}$  that are represented in  $T_2$  is a subset of classes at  $t_1$ , i.e.,  $\mathbf{N}^{TL} \subseteq \Omega$ . Let  $T_1^{UC} = \{x_{1,j}, y_{1,j}\}_{j=1}^R$  be the set of unchanged training pixels at  $\mathbf{X}_1$ , where  $x_{1,j} \in T_1^{UC}$  is the  $j$ -th training sample and  $R$  ( $R < M$ ) is the number of detected unchanged training samples. The training set  $T_2$  of  $\mathbf{X}_2$  is initialized as  $T_2 = \{x_{2,j}, y_{1,j}\}_{j=1}^R = \{x_{2,j}, y_{2,j}\}_{j=1}^R$ , where  $x_{2,j} \in \mathbf{X}_2$  is the  $j$ -th initial training sample and  $y_{1,j} \equiv y_{2,j} \in \mathbf{N}^{TL}$  is its label transferred from  $T_1^{UC}$ . From the samples in the initial training set  $T_2$  unbiased estimates of class statistical parameters for the target domain can be computed. Thus, the proposed system becomes robust to the class statistical distribution differences between the source and target domains. In other words, this choice avoids the need of adapting the classifier parameters of the source domain to the target domain as it is necessary in the TL techniques present in the literature [10]-[15]. It is worth noting that this step implicitly handles the situation in which a class completely disappears between the two acquisitions with no additional effort: all the samples in  $T_1$  associated to a disappeared class fall into the set of changed pixels. Therefore their label will not be transferred and the associated class will not appear in  $\mathbf{N}^{TL}$ .

### *Step 2: Conditional-Entropy-based Active Learning (CEAL)*

The second step aims at optimizing the training set  $T_2$  (and thus the estimates of class statistical parameters) by active learning defined for Bayesian cascade classification. AL methods enrich the initial labeled training set by iteratively selecting the most informative samples from a pool of unlabeled samples and adding them to the current training set after manual labeling by a supervisor. The most informative unlabeled samples can be selected using an uncertainty criterion. This criterion should select the unlabeled samples that have the lowest probability to be correctly classified by the considered classifier and thus show the maximum uncertainty on their class labels. Such samples result to be the

most informative ones for the classifier itself. The supervisor is usually a human expert who gives a class label to the selected samples. The main benefit of AL is that it can significantly reduce the need for labeling samples, and thus the related cost as a result of avoiding redundant sampling.

Differently from the AL techniques proposed in the literature that are devoted to single-date image classification [26]-[30], here we re-define AL by generalizing its use to the classification of image time series in the context of Bayesian cascade decision rule. The proposed AL method models the uncertainty of unlabeled samples by taking into account the temporal correlation between  $\mathbf{X}_2$  and  $\mathbf{X}_1$ , and thus results in a conditional uncertainty criterion. To consider temporal dependence in modeling the uncertainty of samples, we introduce a novel AL method defined on the basis of conditional entropy. The concept of entropy has been already used in the literature for the definition of uncertainty in the context of AL [32]-[33]. In [32] marginal entropy is applied to select uncertain training samples for single-date image classification without considering any temporal information. In [33] the concept of multitemporal uncertainty is introduced. Joint entropy is used to select pairs of uncertain multitemporal training samples for the joint classification of multitemporal images in the context of compound classification [33]. The proposed conditional entropy based AL method significantly differs from [32]-[33] due to the fact that it models the uncertainty of unlabeled samples only in the target domain but it also considers the temporal correlation between the target and source domains.

Let  $H(x_{2,j}|x_{1,j})$  be the conditional entropy for the pixel  $x_{2,j}$ , given the observation  $x_{1,j}$ , *i.e.*,

$$H(x_{2,j}|x_{1,j}) = - \sum_{v_k \in \mathbf{N}} P(v_k|x_{1,j}, x_{2,j}) \log P(v_k|x_{1,j}, x_{2,j}) \quad (4)$$

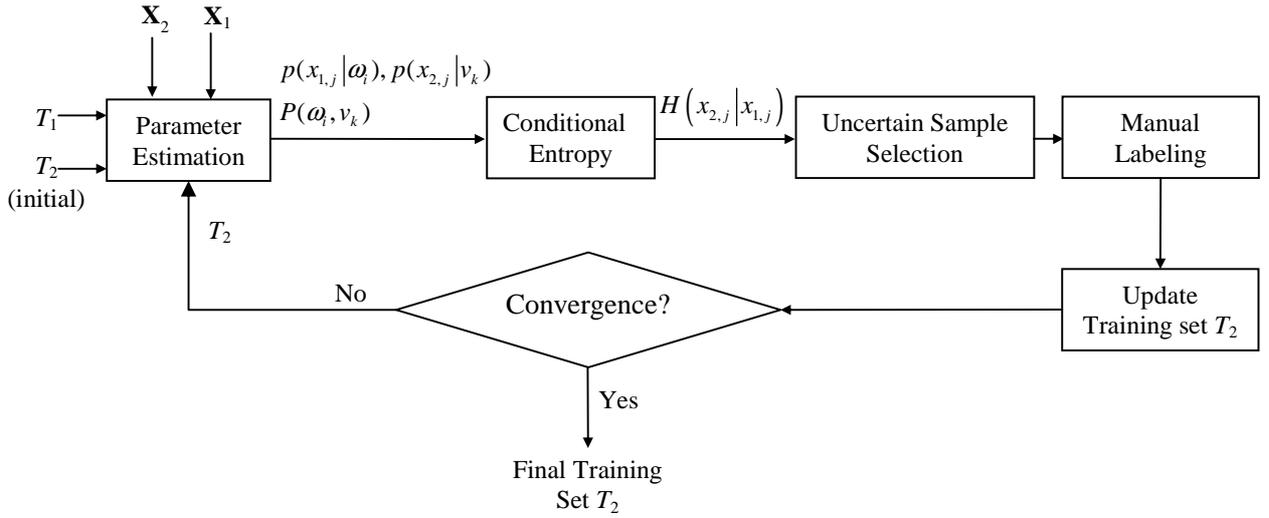
As mentioned before, the estimation of  $P(v_k | x_{1,j}, x_{2,j})$  is a complex task, and therefore we adopt the conventional assumption of class-conditional independence in the time domain to simplify the estimations as in [10],[11]. Under this assumption we can write:

$$H(x_{2,j} | x_{1,j}) = - \sum_{v_k \in N} \left[ \frac{\sum_{\omega_i \in \Omega} p(x_{1,j} | \omega_i) p(x_{2,j} | v_k) P(\omega_i, v_k)}{\sum_{\omega_s \in \Omega, v_r \in N} (p(x_{1,j} | \omega_s) p(x_{2,j} | v_r) P(\omega_s, v_r))} \log \left( \frac{p(x_{1,j} | \omega) p(x_{2,j} | v_k) P(\omega, v_k)}{\sum_{\omega_s \in \Omega, v_r \in N} (p(x_{1,j} | \omega_s) p(x_{2,j} | v_r) P(\omega_s, v_r))} \right) \right] \quad (5)$$

According to (5),  $P(\omega_i, v_k)$  is the only term that models the temporal correlation between the two domains. A small value of  $H(x_{2,j} | x_{1,j})$  shows that the decision of the cascade classifier on the pixel  $x_{2,j}$  is reliable, whereas a high value of  $H(x_{2,j} | x_{1,j})$  points out that the decision of the cascade classifier is not reliable, and thus the corresponding sample is uncertain (i.e., informative for the classifier).

At the first iteration of the proposed AL method conditional entropy is estimated according to class statistical parameters evaluated on  $T_1$  and on the initial  $T_2$  obtained by transfer learning. Then the training set is enlarged by adding a batch of unlabeled samples with maximum conditional entropy (i.e., those that have the maximum uncertainty among classes) after manual labeling by a supervisor. Once  $T_2$  has been enlarged, the class probability density functions associated to the image  $\mathbf{X}_2$  and the joint prior probability of classes should be updated accordingly. The joint prior probabilities of classes are derived from the images under investigation. To this end, each image is classified independently from the others by the standard Bayesian maximum posterior probability classifier [31], and then the class transitions are calculated. It is worth noting that the class-conditional density functions at time  $t_1$  are estimated from the available training set  $T_1$  and thus remain fixed during the AL iterations. Once parameters have been updated, the conditional entropy for each unlabeled sample present in the image  $\mathbf{X}_2$  is estimated by (5).

This process is iterated until convergence, which is achieved when either the values of class parameters do not change from an iteration to the next one, or the upper bound of the cost for labeling samples is achieved (i.e., the maximum possible number of samples is labeled). Fig. 2 shows the architecture of the proposed Conditional-Entropy-based Active Learning (CEAL) method. When the AL process is completed, the training set includes a small number of most informative samples for the considered classifier. Finally, image  $\mathbf{X}_2$  is classified by using the Bayesian decision rule for cascade classification given in Section II.A using the estimated parameters for the class-conditional density functions at time  $t_2$ , as well as the joint prior probabilities of classes and the class-conditional density functions at time  $t_1$ .



**Fig. 2.** Architecture of the proposed Conditional-Entropy-based Active Learning (CEAL) method.

It is worth noting that labels assigned to uncertain samples during the AL step may not belong to the initial set of classes  $N^{TL}$ . Due to the expert labeling process, the proposed AL step allows one to detect new classes that may be appeared between the acquisition dates of  $\mathbf{X}_1$  and  $\mathbf{X}_2$ . Let  $\varepsilon$  be the generic class label associated by the human expert to a specific sample at a given iteration of the AL

process. If  $\varepsilon$  does not appear in the initial set  $N^{TL}$  (i.e.,  $\varepsilon \notin N^{TL}$ ) at time  $t_2$ , the final set  $N$  of classes at time  $t_2$  is defined as  $N=N^{TL} \cup \{\varepsilon\}$ , whereas if  $\varepsilon \in N^{TL}$ , the final set  $N$  is equal to the initial set  $N^{TL}$ , i.e.,  $N=N^{TL}$ .

### 3 DATA SET DESCRIPTION AND DESIGN OF EXPERIMENTS

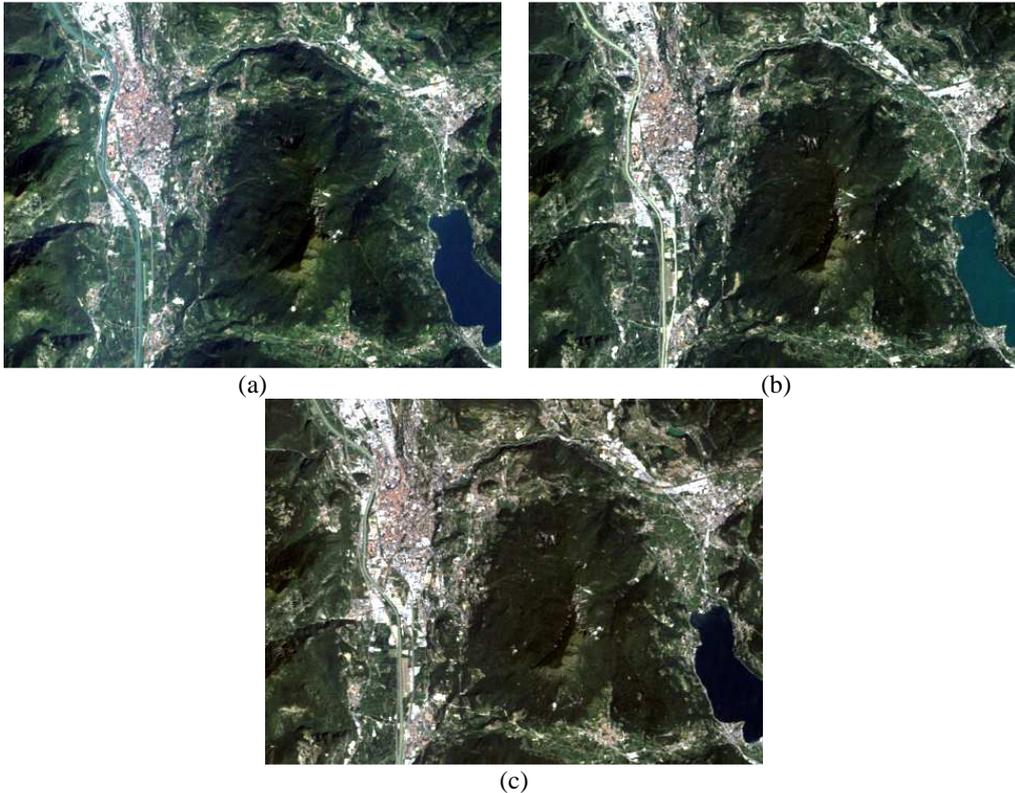
#### A. Data Set Description

Experiments were carried out by exploiting two time series of remote sensing images aiming at land-cover maps updating. However the proposed system can be effectively employed in other domains that require the analysis of image time series (e.g., biomedicine). The first time series is made up of three co-registered multispectral images acquired on the surrounding of the city of Trento, Italy, in September 1999, July 2003 and July 2007. The selected test site is a section of  $470 \times 351$  pixels with a spatial resolution of 30 m. True color composites of the images are shown in Fig. 3. The July 2003 and July 2007 images were acquired by the Thematic Mapper (TM) sensor of the Landsat-5 satellite, whereas the September 1999 image was acquired by the Enhanced Thematic Mapper Plus (ETM+) sensor of the Landsat-7 satellite. Both multispectral scanners acquire images in the same seven spectral channels. However, only six ( $C = 6$ ) were used in the experiments. The thermal infrared band was neglected due to its lower geometrical resolution. Several land-cover classes were considered for experiments: water, forest, fields, urban area and industrial area. Between the acquisitions no additional land-cover classes are observed. In the experiments, two scenarios were considered that share the same source domain, i.e., the September 1999 image,  $\mathbf{X}_1$ . However they differ because of the considered target domain (i.e.,  $\mathbf{X}_2$ ). In the first scenario (Scenario 1) the target domain is represented by the July 2003 image, whereas in the second one (Scenario 2), it is represented by the July 2007 image (see TABLE I). The considered scenarios

simulate challenging conditions, due to the fact that: i) different sensors were used for the data acquisition, and ii) class statistical distributions in the target domain differ

TABLE I. SCENARIOS CONSIDERED IN THE EXPERIMENTS.

Scenario	$X_1$	$X_2$
1	September 1999 (Landsat-7 ETM+)	July 2003 (Landsat-5 TM)
2	September 1999 (Landsat-7 ETM+)	July 2007 (Landsat-5 TM)



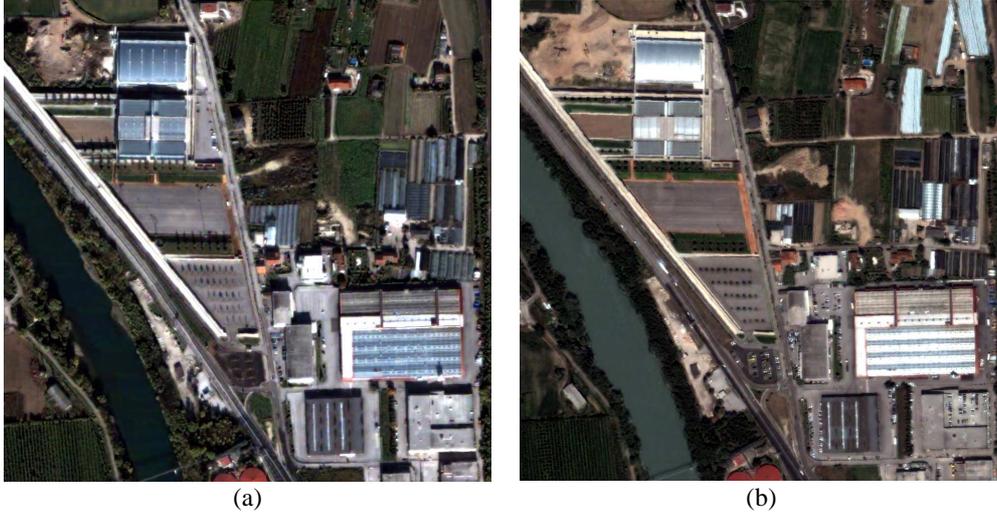
**Fig. 3.** True color composite of the Trento Landsat data set: (a) image acquired in September 1999 by Landsat-7 ETM+ ( $X_1$ );(b) image acquired in July 2003 by Landsat-5 TM ( $X_2$ ); and (c) image acquired in July 2007 by Landsat-5 TM ( $X_2$ ) courtesy of the U.S. Geological Survey.

significantly from those of the source domain. The latter condition was quantitatively assessed by computing the class similarity according to the pairwise Jeffreys-Matusita (JM) statistical distance [12].

JM values between the source and the target distributions are high for all land-cover classes (i.e., all the

pairwise JM distances are close to the saturation value  $\sqrt{2}$ ). Therefore standard DA methods proposed in the literature can not work properly on these scenarios, *i.e.*, they fail in effectively classifying the target domain due to the significantly different behaviors of the land-cover class statistical distributions. This is further confirmed by the overall accuracy achieved by a Bayesian classifier trained on  $T_1$  and applied to  $\mathbf{X}_2$ , which is very poor (around 12%) for both scenarios.

The second time series is made up of two co-registered multispectral images acquired by the QuickBird satellite on the south part of the city of Trento (Italy), in October 2005 and July 2006, respectively (see Fig. 4). The 4 QuickBird spectral bands were considered in the experiments. The selected test site is a section of  $1520 \times 1504$  pixels with a spatial resolution of 0.7 m after pansharpening. The October 2005 image is considered as the source domain (*i.e.*,  $\mathbf{X}_1$ ) and the July 2006 image is considered as target domain (*i.e.*,  $\mathbf{X}_2$ ). 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 the appearance of a new class (*i.e.*, plastic-mulched fields) is observed. Also for this data set, the considered scenario models challenging conditions, due to the fact that: i) a new class is present in the target domain (*i.e.*, in the July 2006 image) and, ii) class statistical distributions in the target domain differ significantly from those of the source domain. This was quantitatively assessed by computing the class similarity according to the pairwise JM statistical distance [12]. JM values between the source and the target distributions are high (*i.e.*, close to the saturation value) for all land-cover classes. As a further confirmation the overall accuracy achieved by a Bayesian classifier trained on  $T_1$  and applied to  $\mathbf{X}_2$  is very poor (around 40%).



**Fig. 4.** True color composite of the Trento Quickbird data set: (a) image acquired in October 2005 ( $\mathbf{X}_1$ );(b) image acquired in July 2006 ( $\mathbf{X}_2$ ).

### B. Design of Experiments

In the experiments 2 batch size values for the AL technique were selected, namely  $h=5$  and  $h = 10$ . Experimental results are reported as the average accuracy on ten trials defined according to ten initial randomly selected training sets for  $T_1$ . For each trial the Trento Landsat data set includes a training set  $T_1$  of 338 samples selected randomly among 1224 available labeled samples in the image  $\mathbf{X}_1$ , whereas the Trento Quickbird data set includes a training set  $T_1$  of 574 samples selected randomly among 3850 available labeled samples in the image  $\mathbf{X}_1$ . A training set  $T_2$  is assumed initially not available for the target domain  $\mathbf{X}_2$  for both data sets. In order to build the initial training set  $T_2$ , CDTL is applied. To select the unchanged samples whose class labels will be transferred to  $\mathbf{X}_2$ , a conservative criterion is employed such that only those training samples with a high probability of being unchanged are selected. Therefore the risk of propagating the class labels of uncertain samples (i.e., training samples that changed their label) is reduced as much as possible. For sake of consistency among the 10 trials the same

number of labels is transferred. The available ground reference samples of  $\mathbf{X}_2$  are used to derive a pool set and a test set. The pool set includes samples that will be queried during the AL iterations, whereas the test set consists of samples that are only used for accuracy assessment. Note that the samples of the pool set are different from those of the test set. The class-conditional density functions required for the estimation of the Bayesian cascade decision rule [see (3)] and the conditional entropy [see (5)] are assumed to follow a parametric Gaussian distribution. As we are considering multispectral images acquired by optical passive sensor, this is a reasonable and common assumption [10]-[11],[18],[19].

We applied the proposed system which is based on: i) Change-Detection-driven Transfer Learning (CDTL) and Conditional-Entropy-based Active Learning (CEAL) for the generation of the target domain training set; and ii) the Bayes decision rule for cascade classification. In the all experiments the proposed system, which takes advantage of temporal correlation between target and source domain in both steps, is denoted as CDTL-CEAL. Thus, in order to assess its effectiveness, the results obtained by it were compared with those achieved by two methods that neglect or only partially consider the temporal correlation.

The first method is based on single-date classification. Thus temporal correlation is ignored both in the training set definition as well as in the classification step. The goal is to classify the target domain (i.e.,  $\mathbf{X}_2$ ) without considering the existence of the source domain  $\mathbf{X}_1$ . In this experiment the training set for  $\mathbf{X}_2$  is built according to the standard Marginal-Entropy-based Active Learning (MEAL) technique [32], whereas classification is performed by applying the single date Bayesian maximum posterior probability classifier [31]. The MEAL technique selects the most uncertain unlabeled samples from a single-date image on the basis of the standard marginal entropy (*i.e.*, ignoring the temporal dependence

between images) [32]. As for applying the Bayesian decision rule an initial training set  $T_2$  for  $\mathbf{X}_2$  is required (otherwise it is not possible to estimate the parameters of the class distributions), there is a need to build an initial training set including enough labeled samples to train the classifier before starting with AL iterations. To this end, Random Sampling (RS) is applied to the samples in  $\mathbf{X}_2$ . Selected samples are labeled to define the initial  $T_2$ . The number of samples being labeled depends on the data sets and on the number of image features given as input to the classifier. Here, in order to make the comparison with the proposed system reliable, the number of randomly selected samples is set equal to that of samples transferred from source to target domain according to the CDTL step of the proposed system. It is worth noting that the initial estimates might be obtained even with a slightly smaller number of samples. However, the RS step requires anyway an additional cost of samples for the initialization of the training set  $T_2$ , whereas the proposed system does not. This method is denoted in the following as RS-MEAL.

The second method selected for comparison purposes, neglecting the temporal correlation only in the active learning step for the cascade classification. Thus a system similar to the proposed one is considered where Conditional-Entropy-based Active Learning (CEAL) is replaced by RS, i.e., after the CDTL step, unlabeled samples to be added to the initial  $T_2$  are randomly selected. In this way temporal correlation is neglected as RS is performed without involving the source domain. This choice allows one to assess the usefulness of the CEAL step of the proposed system. After performing RS, the source domain is classified by a cascade classifier. This method is denoted in the following as CDTL-RS.

## 4 EXPERIMENTAL RESULTS

### A. *Results for Trento Landsat Data Set: Scenario 1*

In order to use the proposed system, the CVA technique has been applied to images  $\mathbf{X}_1$  (i.e., the

September 1999 image) and  $\mathbf{X}_2$  (i.e., the July 2003 image). Here the thresholding method presented in [34] is applied; however other methods can be used. The class labels of training pixels in  $\mathbf{X}_1$  detected as unchanged are transferred to  $\mathbf{X}_2$ . For each of the ten trials,  $T_2$  is built by transferring the class labels of 95 unchanged training samples from  $\mathbf{X}_1$  to  $\mathbf{X}_2$ . The number of samples for each class in  $T_2$  is given in Table II. The effectiveness of this step is validated by the fact that JM distances between the distributions in the target domain and the ones estimated on the initial training samples (i.e., after CDTL) are smaller than the distances between class distributions in the target domain and source domain. Once label propagation is completed, the AL step is applied to enrich the initial training set by selecting the most informative samples from a set of unlabeled samples (i.e., the pool set). Table II together with the numbers of initial training samples obtained by the CDTL, reports unlabeled samples in the pool set, and of test samples (which are used for accuracy assessment) available for each land-cover class in  $\mathbf{X}_2$ .

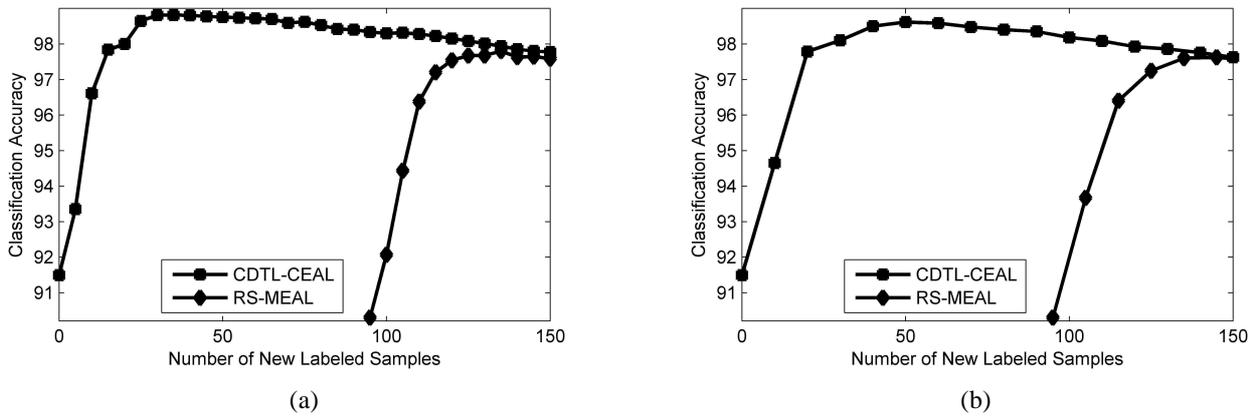
TABLE II. INITIAL TRAINING, POOL AND TEST SETS FOR IMAGE (SCENARIO 1).

Land-cover classes	Initial Training Set	Pool Set	Test Set
<b>Water</b>	21	168	420
<b>Forest</b>	25	200	402
<b>Fields</b>	15	115	207
<b>Industrial Area</b>	17	135	159
<b>Urban Area</b>	17	136	172
<b>Total</b>	95	754	1360

In the first set of trials, we compare the effectiveness of the proposed system (defined for the cascade classifier) with the RS-MEAL method (defined for the standard Bayesian maximum posterior probability classifier). Fig. 5 shows the average (on 10 trials) classification accuracies of  $\mathbf{X}_2$  versus the number of new labeled samples obtained by the proposed system and the RS-MEAL method. The new labeled

samples are the patterns labeled from the expert during the considered iteration of AL process. The reason of obtaining a good accuracy with the proposed system in case of labeling no samples (i.e., the number of new labeled samples is zero) is due to the CDTL step. In other words, training set without labeling any sample is not empty in the case of the proposed system, whereas it is empty in case of the RS-MEAL method (thus the related accuracy is zero). To implement RS-MEAL, 95 samples (that corresponds to the number of samples included in  $T_2$  after the CDTL step of the proposed system) are randomly selected and labeled to initialize the training set. From Fig. 5, one can observe that the accuracies obtained by the proposed system are significantly higher than those yielded by the RS-MEAL method for both values of  $h$ . As an example (see Fig. 5.a that refers to  $h=5$ ), the proposed system yields an accuracy of 98.33% by labeling 100 unlabeled samples from the pool, whereas the RS-MEAL provides an accuracy of 92.06% with the same amount of labeled samples. Table III shows the confusion matrices resulted by using the proposed CDTL-CEAL and RS-MEAL when 100 new samples are labeled in the case of  $h=5$ . From the table one can see that the most critical classes are “Fields” and “Urban Area” that are confused mainly with “Forest” and “Industrial area” in the case of the RS-MEAL. CDTL-CEAL effectively reduces misclassified samples in the mentioned critical situations and solves also the minor problems related to “Forest” and “Water” classes. The results reported in Fig. 5 demonstrate that, due to the CDTL step, the proposed system significantly reduces the labeling cost with respect to the initial definition of  $T_2$ . It is worth noting that the effectiveness of proposed system compared to the RS-MEAL is due to its ability in: i) defining the initial training set  $T_2$  without any labeling cost (i.e., the CDTL step), ii) expanding it according to the novel AL method that exploits the temporal correlation between the images (i.e., the CEAL step), and iii) classifying the image by considering the temporal correlation between the images. The results clearly show the importance of information conveyed by

temporal correlation between multitemporal images for both optimizing the definition of the training sets and classifying the image. By analyzing figure, one can observe that the proposed system provides a fast improvement on the accuracy at early iterations, whereas at the late iterations a small reduction on the accuracy is observed. This is due to the fact that after a number of iterations the few outliers included in the pool that do not properly model the distribution of test samples are added to the training set. This is not the case for early iterations where outliers are not selected due to their certain (wrong) class label [33]. As the number of labeled samples increases, the CDTL-CEAL method (as well as other AL methods) definitely converges to the accuracy that is obtained when all the unlabeled samples are added to the training set.



**Fig. 5.** Average (on 10 trials) overall classification accuracy versus the number of new labeled samples at  $t_2$  obtained by the proposed CDTL-CEAL and the RS-MEAL methods when (a)  $h=5$  and (b)  $h=10$  (Scenario 1).

TABLE III. THE CONFUSION MATRICES OBTAINED AT THE SECOND ITERATION OF (A) THE PROPOSED CDTL-CEAL AND (B) THE RS-MEAL WHEN  $H=5$

Reference Data \ Classified Data	Water	Forest	Fields	Industrial Area	Urban Area
Water	419	-	-	-	-
Forest	-	399	34	-	-
Fields	-	2	169	-	-
Industrial Area	1	1	4	157	67
Urban Area	-	-	-	2	105

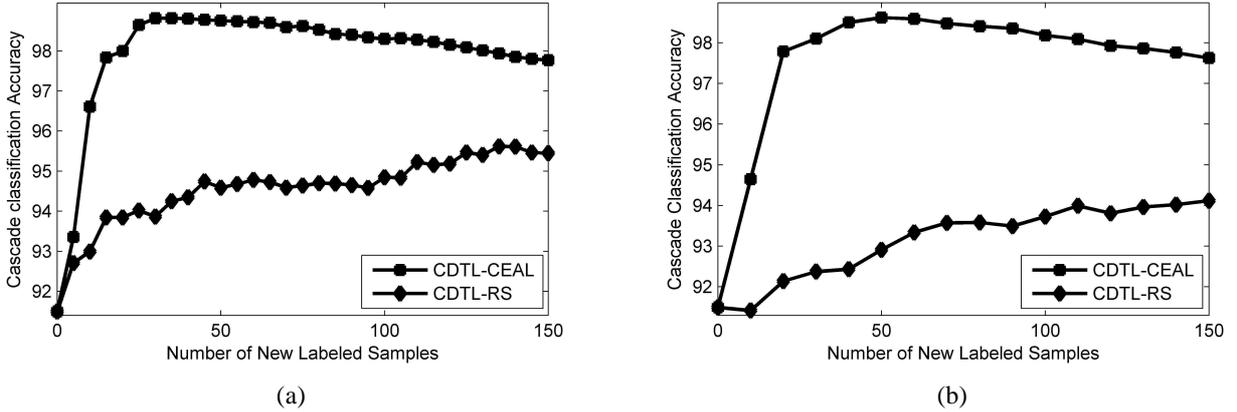
(a)

Reference Data \ Classified Data	Water	Forest	Fields	Industrial Area	Urban Area
Water	420	-	-	-	-
Forest	-	402	3	-	-
Fields	-	-	204	-	-
Industrial Area	-	-	-	158	17
Urban Area	-	-	-	1	155

(b)

The second set of experiments, aims at assessing the effectiveness of the AL step for the proposed system. This is done by comparing the proposed system with the CDTL-RS technique. Fig. 6 shows the average (on 10 trials) classification accuracies of  $\mathbf{X}_2$  versus the number of new labeled samples obtained by both the proposed system and the CDTL-RS method defined in the framework of cascade classification for two values of  $h$ . Due to the CDTL step, both methods result in a good accuracy when the number of new labeled samples is zero. From Fig. 6, one can see that the proposed system provides always higher accuracies than the CDTL-RS and reaches convergence with a smaller number of new labeled samples. As an example, by analyzing Fig. 6.b (which refers to  $h=10$ ) the accuracy of CDTL-CEAL is 94.64% with only 10 new labeled samples, whereas the accuracy of CDTL-RS is only 91.41 %. These results confirm the importance of temporal dependence information conveyed by the proposed Conditional-Entropy-based Active Learning in optimizing the training set  $T_2$  for cascade classification. Note that also in this case the reason of the small reduction on the accuracy with the proposed system at

the late iterations is related to the addition of noisy samples to the training set during the active-learning process.



**Fig. 6.** Average (on 10 trials) overall classification accuracy versus the number of new labeled samples at  $t_2$  obtained by the proposed CDTL-CEAL and the CDTL-RS methods when (a)  $h=5$  and (b)  $h=10$  (Scenario 1).

### B. Results for Trento Landsat Data Set: Scenario 2

Similarly as before, the CVA has been applied to images  $\mathbf{X}_1$  (i.e., the September 1999 image) and  $\mathbf{X}_2$  (i.e., the July 2007 image). The class labels of unchanged pixels in  $\mathbf{X}_1$  are propagated to  $\mathbf{X}_2$ . For each of the ten trials,  $T_2$  is built by transferring the class labels of 94 unchanged training samples from  $\mathbf{X}_1$  to  $\mathbf{X}_2$ . The number of samples for each class in  $T_2$  is given in Table IV. The effectiveness of this step is confirmed by the fact that JM distances between the distributions in the target domain and the ones estimated on the initial training samples (i.e., after CDTL) are smaller than the distances between class distributions in the target and source domains. Afterward, the initial training set is expanded by applying the AL step. Table IV together with the number of initial training samples obtained on the basis of the CDTL, shows unlabeled samples in the pool set, and test samples (which are used for accuracy assessment) available for each land-cover class in  $\mathbf{X}_2$ .

TABLE IV. INITIAL TRAINING, POOL AND TEST SETS FOR IMAGE (SCENARIO 2).

Land-cover classes	Initial Training Set	Pool Set	Test Set
<b>Water</b>	32	388	217
<b>Forest</b>	31	371	252
<b>Fields</b>	8	97	154
<b>Industrial Area</b>	12	147	172
<b>Urban Area</b>	11	127	155
<b>Total</b>	94	1130	950

Fig. 7 shows the behavior of the average (on 10 trials) overall accuracies obtained by the CDTL-CEAL and the RS-MEAL (which ignores the CDTL step and therefore assumes that an initial training set is populated by selecting and labeling samples randomly) techniques in the cases of  $h=5$  (Fig. 7.a) and  $h=10$  (Fig. 7.b). To implement RS-MEAL, 94 samples, which correspond to the number of samples included in  $T_2$  after the CDTL step of the proposed system, are randomly selected and labeled to initialize the training set. By analyzing Fig. 7, one can observe that the proposed system results in the highest accuracies at all the iterations for both values of  $h$ . Moreover, it again reaches convergence with a smaller number of new labeled samples, due to: i) the CDTL (which provides initial training samples without any labeling cost), ii) the AL step (which exploits the temporal correlation between images), and iii) the cascade classification (which also uses the temporal dependence information between the images). As an example, the proposed system yields an accuracy of 95.12% with only 10 new labeled samples at  $t_2$ , whereas the RS-MEAL reaches a similar accuracy with 110 samples (see Fig. 7.a). Table V shows the confusion matrices obtained by using the proposed CDTL-CEAL and RS-MEAL when almost 100 new samples are labeled in the case of  $h=5$ . From the table one can see that the most critical classes are “Fields” and “Urban Area” that are confused mainly with “Forest” and “Industrial area” in the case of the RS-MEAL. CDTL-CEAL effectively reduces misclassified samples in the mentioned critical situations. Note that the classification of the image  $\mathbf{X}_2$  using the classifier directly trained with

the training set  $T_1$  provides an accuracy of only 21.34%. This proves the complexity of the considered problem.

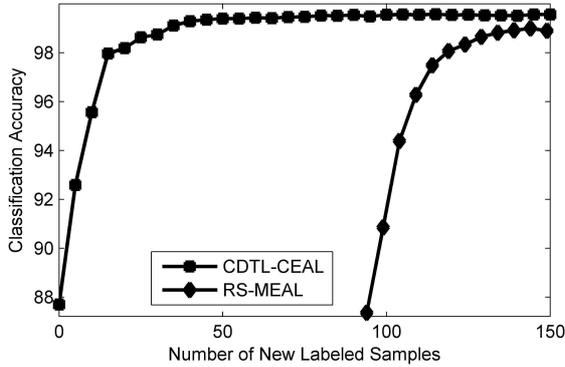
TABLE V. THE CONFUSION MATRICES OBTAINED AT THE SECOND ITERATION OF (A) THE PROPOSED CDTL-CEAL AND (B) THE RS-MEAL WHEN  $H=5$

Reference Data \ Classified Data	Water	Forest	Fields	Industrial Area	Urban Area
Water	216	-	-	-	-
Forest	1	250	41	-	-
Fields	-	2	105	-	-
Industrial Area	-	-	-	169	33
Urban Area	-	-	8	3	122

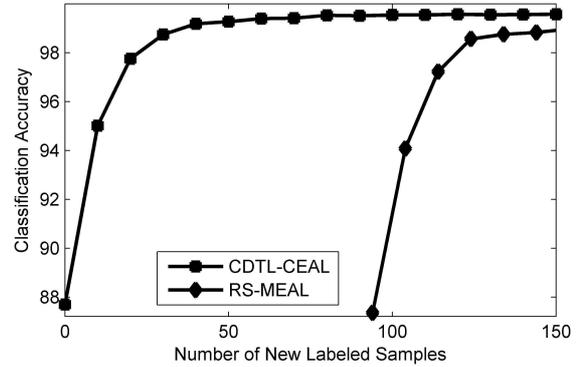
(a)

Reference Data \ Classified Data	Water	Forest	Fields	Industrial Area	Urban Area
Water	217	-	-	-	-
Forest	-	252	1	-	-
Fields	-	-	152	-	-
Industrial Area	-	-	1	170	2
Urban Area	-	-	-	2	153

(b)



(a)

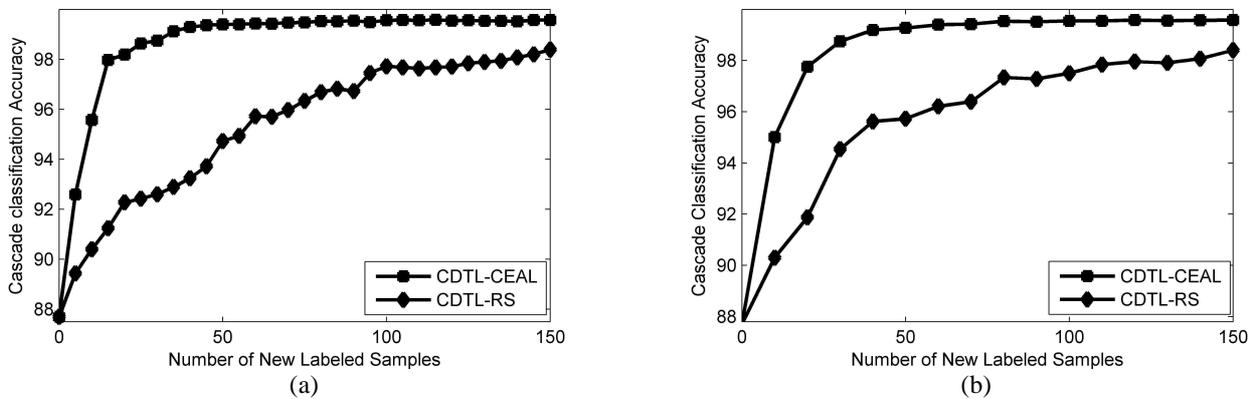


(b)

Fig. 7. Average (on 10 trials) overall classification accuracy versus the number of new labeled samples at  $t_2$  obtained by the proposed CDTL-CEAL and the RS-MEAL methods when (a)  $h=5$  and (b)  $h=10$  (Scenario 2).

Fig. 8 shows the comparison of the average overall accuracy obtained by the proposed system with and CDTL-RS for both values of  $h$ . From these plots, one can see that the proposed system provides

again the selection of more informative samples than the CDTL-RS, and achieves higher accuracies for the same number of labeled samples. As an example, in the case of  $h=10$  the accuracy yielded by proposed system is 95.11% with only 10 labeled samples, whereas it is 90.39% with the same number of labeled samples if the CDTL-RS is used. These results confirm the effectiveness of the proposed CEAL method.



**Fig. 8.** Average (on 10 trials) overall classification accuracy versus the number of new labeled samples at  $t_2$  obtained by the proposed CDTL-CEAL and the CDTL-RS methods when (a)  $h=5$  and (b)  $h=10$  (Scenario 2).

### C. Results for Trento Quickbird Data Set

In order to run the proposed system also on the Trento Quickbird data set, the CVA technique has been applied to images  $\mathbf{X}_1$  (i.e., the October 2005 image) and  $\mathbf{X}_2$  (i.e., the July 2006 image). Also in this case the thresholding method presented in [34] is applied. Class labels of training pixels in  $\mathbf{X}_1$  detected as unchanged are transferred to  $\mathbf{X}_2$  providing 112 initial training samples for  $T_2$ . The JM statistical distances between the distributions in the target domain and the ones obtained from the initial training samples (obtained by CDTL) are significantly smaller than the distances between distributions in the target domain and the ones in the source domain. Once label propagation is completed, the AL step is

applied to enrich the initial training set  $T_2$ . In the labeling process of AL, the human expert detects a new land-cover class “Plastic-mulched fields”. Table VI shows the number of initial training samples (which are obtained on the basis of the CDTL) and unlabeled samples in the pool set available for each land-cover class in  $X_2$ . In addition, the number of test samples that are used for accuracy assessment is given.

TABLE VI. INITIAL TRAINING, POOL AND TEST SETS FOR  $X_2$  IMAGE (TRENTO QUICKBIRD DATA SET).

Land-cover classes	Initial Training Set	Pool Set	Test Set
Water	34	1004	1032
Red Roof	15	434	469
Asphalt	22	650	474
Fields	21	626	534
Bare soil	20	670	483
Plastic-mulched field	-	354	290
<b>Total</b>	<b>112</b>	<b>3738</b>	<b>3282</b>

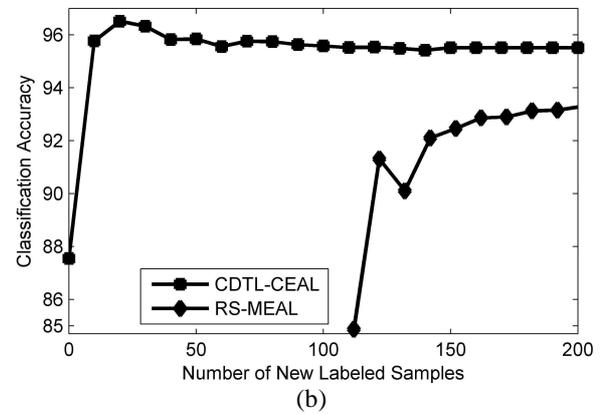
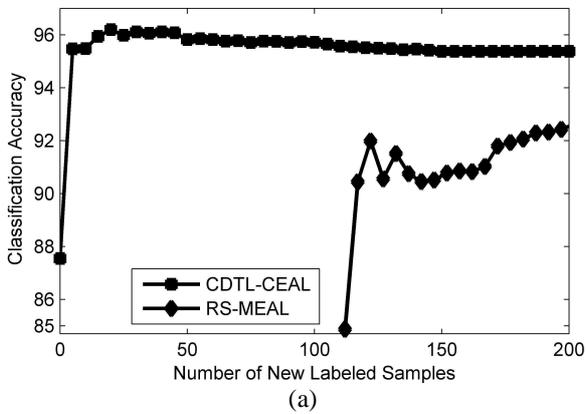


Fig. 9. Average (on 10 trials) overall classification accuracy versus the number of new labeled samples at  $t_2$  obtained by the CDTL-CEAL system and the RS-MEAL method when (a)  $h=5$  and (b)  $h=10$  (Trento Quickbird data set).

Fig. 9 shows the average (on 10 trials) classification accuracies of  $X_2$  versus the number of new labeled samples obtained by the proposed CDTL-CEAL and RS-MEAL in the cases of  $h=5$  (see Fig. 9.a) and  $h=10$  (see Fig. 9.b). To implement RS-MEAL, 112 samples (which correspond to the number of samples included in  $T_2$  after the CDTL step of the proposed system) are randomly selected and labeled

to initialize the training set. From the figures, one can observe that the CDTL-CEAL, again, provides significantly higher accuracies for both values of  $h$ , and reaches convergence with a smaller number of new labeled samples. Also for this data set, the results show the importance of information conveyed by temporal correlation between multitemporal images for both optimizing the definition of the training set and the classification accuracy. As an example, the CDTL-CEAL yields an accuracy of 95.82% with 130 new labeled samples at  $t_2$ , whereas the RS-MEAL provides an accuracy of 91.98% with a similar number of new labeled samples (see Fig. 9.a). Table VII shows the confusion matrices obtained by using the proposed CDTL-CEAL and RS-MEAL when almost 120 new samples are labeled in the case of  $h=5$ . From the table one can see that RS-MEAL results in misclassifications for most of the classes. CDTL-CEAL effectively reduces misclassified samples in the critical situations and shows only few minor problems between “Fields” and “Bare soil” classes.

TABLE VII. THE CONFUSION MATRICES OBTAINED AT THE SECOND ITERATION OF (A) THE PROPOSED CDTL-CEAL AND (B) THE RS-MEAL WHEN  $H=5$

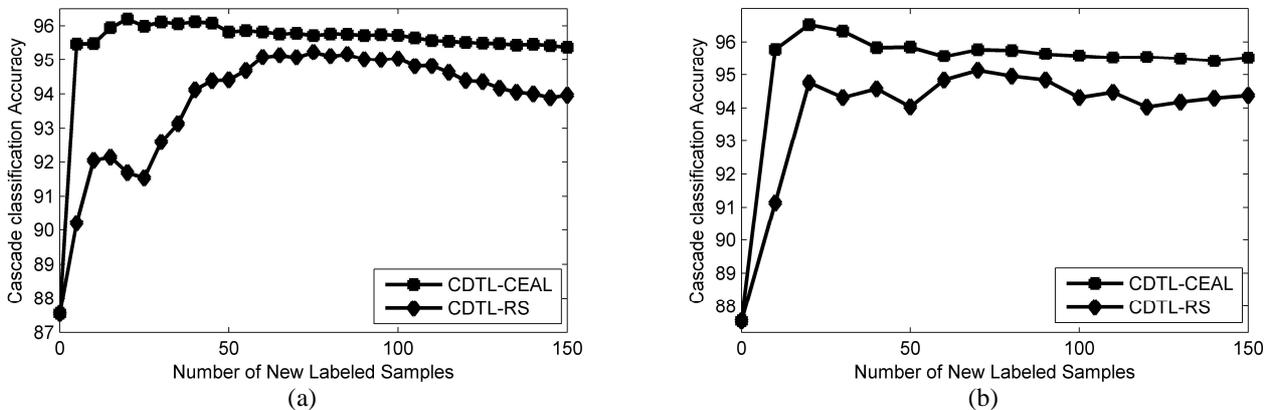
Reference Data \ Classified Data	Water	Red Roof	Asphalt	Fields	Bare soil	Plastic-mulched field
Water	1032	-	-	-	-	-
Red Roof	-	443	5	3	91	2
Asphalt	-	-	408	-	-	5
Fields	-	-	-	524	4	-
Bare soil	-	26	52	1	386	59
Plastic-mulched field	-	-	9	6	2	225

(a)

Reference Data \ Classified Data	Water	Red Roof	Asphalt	Fields	Bare soil	Plastic-mulched field
Water	1032	-	-	-	-	-
Red Roof	-	468	2	1	5	-
Asphalt	-	1	471	-	-	-
Fields	-	-	-	505	-	-
Bare soil	-	-	-	-	368	-
Plastic-mulched field	-	-	1	28	10	290

(b)

Fig. 10 shows the average (on 10 trials) classification accuracies of  $\mathbf{X}_2$  versus the number of new labeled samples obtained by the CDTL-CEAL and CDTL-RS in the cases of  $h=5$  (Fig. 10.a) and  $h=10$  (Fig. 10.b). From the figures one can observe that the CDTL-CEAL provides significantly higher accuracies than the CDTL-RS for both values of  $h$  and leads to a fast improvement of the classification accuracy. This is due to the fact that  $\mathbf{X}_2$  includes a new class (namely “Plastic-mulched filed”), and thus the role of the AL step is crucial to identify and label the new class at the early iterations. As an example (see Fig. 10.a that refers to  $h=5$ ), the CDTL-CEAL yields an accuracy of 96.19% by labeling only 20 samples from the pool, whereas the CDTL-RS provides an accuracy of 91.68% with the same amount of labeled samples. These results show the effectiveness of the proposed CEAL method in the framework of Bayesian rule for cascade classification.



**Fig. 10.** Average (on 10 trials) overall classification accuracy versus the number of new labeled samples at  $t_2$  obtained by the CDTL-CEAL and the CDTL-RS approaches when (a)  $h=5$  and (b)  $h=10$  (Trento Quickbird data set).

## 5 DISCUSSION AND CONCLUSION

In this paper a novel system for updating classification maps by classifying image time series has been presented. The proposed system aims at classifying an image for which no reference data are

available (target domain) by exploiting another image acquired on the same area at a different time for which reference data are available. This is done by a 2-step procedure. The first step is devoted to the low-cost definition of a training set for the target domain with transfer and active learning methods. To this end, firstly a Change-Detection-driven Transfer Learning method, which propagates the class labels of unchanged samples from the source domain to target domain, is applied to define an initial training set for the target domain. This step offers two main advantages: i) the sample labeling cost in order to populate the initial training set is zero due to transferring the class labels of the unchanged source training patterns, and ii) the dissimilarity between the class distributions of source and target domains does not affect the proposed method since the original samples of the target domain are directly used to estimate the related classifier parameters (i.e., there is no need to adapt the classification parameters of the source domain to target domain, as only the label of samples is transferred). After defining the initial training set, active learning is applied to optimize the training set of the target domain by labeling a small number of most informative unlabeled samples. To this end, we have presented a novel active learning method in the framework of Bayes rule for cascade classification. The proposed active learning method is based on the conditional entropy associated with the cascade-classification decision rule, and evaluates the uncertainty of samples taking into account the temporal dependence modeled by the joint prior probabilities of classes. The proposed active learning method significantly reduces the number of new labeled samples to be collected at the target domain for optimizing the classification results, and therefore minimizes the related sample labeling cost. Moreover accurate classification accuracy is obtained due to improved class models on the basis of the cascade classification. The second step of the proposed system is devoted to the classification of the target domain. This is done by cascade classification exploiting the temporal correlation between the domains.

Experiments carried out on 2 multitemporal data sets of remote sensing images show that: i) the proposed system is robust to the class statistical distribution differences between the source and target domains, due to the Change-Detection-driven Transfer Learning step, ii) exploiting the temporal dependence in the definition of active learning for cascade classification problems results in higher accuracies for land-cover map updating than the other algorithms when the same number of labeled samples is considered, and iii) the use of cascade classifier for multitemporal images improves the classification performance with respect to standard single date Bayesian maximum posterior classifier. These results are very important as obtained on a heterogeneous area working on images acquired in different seasons (i.e., September and July as well as October and July) and by different sensors (i.e., Landsat-5 TM and Landsat-7 ETM+). The achievements point out the flexibility of the proposed technique that can obtain high accuracy with few labeled samples for the target domain also in critical conditions.

It is worth noting that the proposed system does not have any limitation on the set of land-cover classes that characterize the two domains. This is due to the fact that the initial training set for the target domain (which only contains land-cover classes shared by the target and source domains due to class-label propagation of unchanged samples in the CDTL step) is enriched with possible new classes during the AL step. We can state that from the AL viewpoint the unlabeled samples associated to new classes that may appear in the target domain are expected to be highly informative, as they are not represented in the training set of the target domain yet. Thus, if a new class appears in the target domain, the AL step is very likely to select samples that belong to it at the first iterations. These samples will be thus represented in the training set after manual labeling by the supervisor. The case of class disappearance in the target domain is implicitly managed in the CDTL step by not transferring labels associated to

changed training samples. In other words, if a given class in source domain is no longer present in the target domain, all of its pixels will be changed and their labels will not be represented in the training set of the target domain.

It is worth emphasizing that updating classification maps in a cost-effective way is becoming more and more important in real applications. This is due to the increased number of time series of images and their possible free availability. In this context, the proposed system is very promising as it generates a classification map for a generic image in the time series for which no prior information is available, decreasing significantly the cost and effort required for reference data collection. As a future development of this work, we plan to extend the test of the proposed system to longer time series of images and to data acquired by different sensors. Moreover, we also plan to consider the modeling of the spatial context information in the process of active learning. As a final remark, we would like to point out that the proposed system in general and can be adopted for different application domains in the framework of analysis of image time-series.

#### REFERENCES

- [1] D.K. Iakovidis, S. Tsevas, M.A. Savelonas, G.Papamichalis, "Image analysis framework for infection monitoring", *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 1135 – 1144, 2011.
- [2] M. Nischik, C. Forster, "Analysis of skin erythema using true-color images", *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 711-716, 1997.
- [3] A. Meyer-Baese, O. Lange, T. Schlossbauer, A. Wismuller, "Computer-aided diagnosis and visualization based on clustering and independent component analysis for breast MRI", *15th IEEE International Conference on Image Processing*, San Diego, California, USA, 2008, pp.3000-3003.
- [4] J. Patriarche and B. Erickson, "A review of the automatic detection of change in serial imaging studies of the brain," *Journal of Digital Imaging*, vol. 17, pp. 158–174, 2004.
- [5] H. Narasimha-Iyer, A. Can, B. Roysam, H. L. Tanenbaum, A. Majerovics "Integrated analysis of vascular and nonvascular changes from color retinal fundus image sequences," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 8, pp. 1436-1445, 2007.
- [6] 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.
- [7] 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.
- [8] 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, pp. 199-210, Feb. 2011.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] K. Bahirat, F. Bovolo, L. Bruzzone, S. Chaudhuri, “A Novel Domain Adaptation Bayesian Classifier for Updating Land-Cover Maps with Class Differences in Source and Target Domains”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 7, 2012, pp. 2810-2826.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] P. Rai, A. Saha, H. Daume III, and S. Venkatasubramanian, “Domain adaptation meets active learning”, *Workshop on Active Learning for Natural Language Processing*, pp. 27-32, Los Angeles, California, USA, 2010.
- [17] X. Shi, W. Fan, and J. Ren, “Actively transfer domain knowledge”, *European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases*, pp. 342-357, Antwerp, Belgium, 2008.
- [18] L. Bruzzone, and S.B., Serpico, “An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 858-867,1997
- [19] L. Bruzzone, F. D., Prieto, S.B.Serpico, “A neural-statistical approach to multitemporal and multisource remote-sensing image classification”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1350-1359,1999.
- [20] B. Demir, F. Bovolo, L. Bruzzone, "Updating Land-Cover Maps by Classification of Image Time Series: A Novel Change-Detection-Driven Transfer Learning Approach", *IEEE Trans. Geo. Rem. Sens.*, to appear.
- [21] F. Bovolo and 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, 2007.

- [22] L. Bruzzone, F. Bovolo, A Conceptual Framework for Change Detection in Very High Resolution Remote Sensing Images, Proceedings of the IEEE, 2013, in press.
- [23] A., Singh, "Digital change detection techniques using remotely-sensed data", *Int. J. Remote Sens.*, vol. 10, 989-1003, 1989.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] E., Alpaydm, Introduction to Machine Learning, MIT Press, Cambridge, Massachusetts, London, England, 2004.
- [32] D. Tuia, F. Ratle, F. Pacifici, M. Kanevski, and W. J. Emery, "Active learning methods for remote sensing image classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2218 -2232, Jul. 2009.
- [33] B. Demir, F. Bovolo, L. Bruzzone "Detection of land-cover transitions in multitemporal remote sensing images with active learning based compound classification", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 5, pp. 1930-1941, 2012.
- [34] L. Bruzzone, and Fernandez Prieto, D., "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.



**Begüm Demir** received the B.S. degree in 2005, the M.Sc. degree in 2007, and the Ph.D. degree in 2010, all in Electronic and Telecommunication Engineering from Kocaeli University, Turkey. In 2005, she joined KULIS (Kocaeli University Laboratory of Image and Signal Processing) as a researcher, and she was a research assistant at the same department of Kocaeli University during 2006-2010. From March 2009 to September 2009, she was a visiting researcher at the Remote Sensing Laboratory, University of Trento, Italy supported by the International Research Fellowship Program of The Scientific and Technological Research Council of Turkey (TUBITAK). Begüm Demir is currently a Post Doc member of the Remote Sensing Laboratory at the Department of Information Engineering and Computer Science, University of Trento. Her main research interests include image processing and machine learning with applications to remote sensing image analysis. She is the organizer of the Special Sessions on Remote Sensing Image Analysis in the IEEE 19th and 20th Conferences on Signal Processing and Communications Applications. She is a referee for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTER, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of Selected Topics in Signal Processing, IEEE Transactions on Image Processing, ISPRS Journal of Photogrammetry and Remote Sensing, International Journal of Remote Sensing, and Several International Conferences.



**Francesca Bovolo** (S'05–M'07) received the “Laurea” (B.S.), the “Laurea Specialistica” (M.S.) degrees in telecommunication engineering (*summa cum laude*) and the Ph.D. in Communication and Information Technologies from the University of Trento, Italy, in 2001, 2003 and 2006, respectively. She is presently a research fellow at the Remote Sensing Laboratory of the 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 multitemporal remote sensing image analysis and change detection in multispectral and SAR images, and very high resolution images. She conducts research on these topics within the frameworks of several national and international projects. Francesca Bovolo is a referee for the *IEEE Transaction on Geoscience and Remote Sensing*; *IEEE Geoscience and Remote Sensing Letters*; *IEEE Transactions on Image Processing*; *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*; *International Journal of Remote Sensing*; *Pattern Recognition*; *Pattern Recognition Letters*; *Remote Sensing of Environment*; *Photogrammetric Engineering and Remote Sensing*; *Photogrammetry and Remote Sensing*; *IEEE Transactions on Aerospace and Electronic Systems*; *Sensors*. Francesca Bovolo ranked first place in the Student Prize Paper Competition of the 2006 *IEEE International Geoscience and Remote Sensing Symposium* (Denver, August 2006). Since January 2011 she is an associate editor of the *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. She is the Technical Chair of the Sixth International Workshop on the Analysis of Multi-temporal Remote-Sensing Images (MultiTemp 2011). Since 2006, she has served on the Scientific Committee of the SPIE International Conference on “Signal and Image Processing for Remote Sensing”. She has served on the Scientific Committee of the *IEEE Fourth and Fifth International Workshop on the Analysis of Multi-Temporal Remote Sensing Images (MultiTemp 2007 and 2009)*.



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

He is currently a Full Professor of telecommunications at the University of Trento, Italy, where he teaches remote sensing, radar, pattern recognition, and electrical communications. Dr. Bruzzone is the founder and the director of the Remote Sensing Laboratory in the Department of Information Engineering and Computer Science, University of Trento. His current research interests are in the areas of remote sensing, radar and SAR, signal processing, and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects. Among the others, he is the Principal Investigator of the *Radar for icy Moon exploration (RIME)* instrument in the framework of the *JU*pter *ICy moons Explorer (JUICE)* mission of the European Space Agency. He is the author (or coauthor) of 137 scientific publications in referred international journals (93 in IEEE journals), more than 190 papers in conference proceedings, and 16 book chapters. He is editor/co-editor of 11 books/conference proceedings and 1 scientific book. His papers are highly cited, as proven from the total number of citations (more than 7490) and the value of the h-index (44) (source: Google Scholar). He was invited as keynote speaker in 24 international conferences and workshops. Since 2009 he is a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society.

Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (Seattle, July 1998). Since that time he was recipient of many international and national honors and awards. Dr. Bruzzone was a Guest Co-Editor of different Special Issues of international journals. He is the co-founder of the IEEE International Workshop on the Analysis of Multi-Temporal Remote-Sensing Images (MultiTemp) series 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. Since 2013 he has been the founder Editor-in-Chief of the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE. Currently he is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the CANADIAN JOURNAL OF REMOTE SENSING. Since 2012 he has been appointed *Distinguished Speaker* of the IEEE Geoscience and Remote Sensing Society.