

## TARGETED CHANGE DETECTION: A NOVEL SENSOR-INDEPENDENT PARTIALLY-SUPERVISED APPROACH

D. Fernández-Prieto<sup>a</sup>, M. Marconcini<sup>a,\*</sup>

<sup>a</sup> Dept. of Earth Observation Science, Applications and Future Technologies, European Space Agency, ESA-ESRIN,  
Via Galileo Galilei, 00044, Frascati, Rome, Italy - (diego.fernandez, mattia.marconcini)@esa.int

**KEY WORDS:** Change Detection, Land Cover, Satellite, Multitemporal, Multispectral, Hyperspectral

### ABSTRACT:

In several real-world applications (e.g., forestry, agriculture), the objective of change detection is actually limited to one (or few) specific “targeted” land-cover transition(s) affecting a certain area in a given time period. In such cases, ground-truth information is generally available for the only land-cover classes of interest at the two dates, which limits (or hinders) the possibility of successfully employing standard supervised approaches. Moreover, even unsupervised change-detection methods cannot be effectively used, as they allow identifying all the areas experiencing any type of change, but not discriminating where specific land-cover transitions of interest occur. In this paper, we present a novel technique capable of addressing this challenging issue (formulated in terms of a compound decision problem) by exploiting the only ground truth available for the targeted land-cover classes at the two dates. In particular, the proposed method relies on a partially-supervised approach and jointly exploits the Expectation-Maximization (EM) algorithm and an iterative labelling strategy based on Markov random fields (MRF) accounting for spatial and temporal correlation between the two images. Moreover, it also allows handling images acquired by different sensors at the two investigated times. Experimental results on different multi-temporal and multi-sensor data sets confirmed the effectiveness and the reliability of the proposed technique, which provided change-detection accuracies comparable with those obtained by fully-supervised methods.

### 1. INTRODUCTION

Detecting changes occurring on the Earth’s surface represents one of the main applications of satellite remote sensing. Indeed, in a variety of different fields and applications (e.g., urban planning, forestry, agriculture, disaster management, etc.) the employment of multi-temporal satellite data has become essential for identifying where and (when possible) which types of transitions have occurred between two given dates (Jensen, 2009).

Generally, change-detection methods are categorized as either supervised or unsupervised, depending on the availability of suitable prior information (Coppin *et al.*, 2004; Duda *et al.*, 2000; Lu *et al.*, 2004; Radke *et al.*, 2005; Singh, 1989).

When an exhaustive multi-temporal ground truth characterizing all the land-cover classes over the area of interest at both times is available, then, supervised approaches can be applied. These types of techniques are generally robust and effective, and allow identifying all the land-cover transitions occurred between the two considered dates. In this framework, three main approaches are generally employed: post-classification comparison (PCC), supervised direct multi-data classification (DMC) and compound classification (Duda *et al.*, 2000; Lu *et al.*, 2004; Singh, 1989).

When no ground truth is available, instead, unsupervised techniques must be used, which allow detecting areas experiencing changes (being even capable of separating land-cover transitions of different nature and characterizing their distribution), but are unable to provide information on the specific type of changes occurred. Several unsupervised change-detection techniques have been presented so far in the literature. Most of them are based on image differencing, image ratioing, image regression, change vector analysis (CVA) and principal component analysis (PCA), which all require the selection of proper thresholds for determining changed regions (Coppin *et al.*, 2004; Lu *et al.*, 2004; Radke *et al.*, 2005).

It is worth noting that, in the above described framework, su-

pervised methods represent an ideal approach to change-detection analysis, since they permit both to identify areas experiencing changes, as well as to reliably determine the associated land-cover transitions. Nevertheless, their range of applicability is significantly limited by the difficulties in gathering exhaustive and accurate ground-truth information for all the land-cover classes characterizing each date under analysis. Indeed, such a requirement is costly, time consuming and not always possible or feasible to satisfy.

However, in several operational change-detection problems the main objective is not to characterize all the land-cover transitions occurred in the investigated area, but rather to identify a single (or few) targeted land-cover transition(s) of interest. This is typical for instance in agriculture, urban planning, or forestry applications. In such circumstances, when only one (or few) specific land-cover transitions need to be identified, it is reasonable to assume that the collection of ground-truth information associated with the only (or few) land-cover class(es) of interest at the two considered dates is highly simplified. However, under this assumption, neither supervised nor unsupervised change-detection techniques can be effectively employed. Let us consider for instance the case of two images acquired over the same area at different times  $t_1$  and  $t_2$ , where the objective is to identify all the patterns experiencing the targeted land-cover transitions from class “A” (e.g., forest) to “B” (e.g., urban area) under the hypothesis that a ground truth for class “A” at  $t_1$  and for class “B” at  $t_2$  is available (or can be easily retrieved by an operator), respectively. In this context, on the one hand, supervised techniques cannot be used, since the lack of an exhaustive ground truth characterizing all the land-cover classes at the two dates under consideration will not allow a successful training of the classifiers. On the other hand, unsupervised techniques may allow identifying all the areas experiencing any type of change, but not discriminating where specific targeted land-cover transitions of interest occur. In this latter case, a comparison (e.g., through a significance-testing approach (Jeon and Landgrebe, 1999)) of the labelled samples available for the

classes of interest at the two dates against all those identified as changed pixels may provide some degree of information on the type of land-cover transition, but not with the accuracy and reliability provided by fully-supervised approaches).

In this paper, we formulate this complex issue in terms of a compound decision problem (Duda *et al.*, 2000) and propose a novel partially-supervised change-detection (PSCD) technique capable of exploiting the only prior knowledge available for the specific land-cover classes of interest at the two times (thus avoiding the need to rely on exhaustive ground-truth information for all the classes), while providing accuracies comparable with those of fully-supervised methods. Moreover, it has the great advantage of being sensor-independent, which allows selecting at each date the set of sensors and features most suitable for characterizing the targeted class(es) of interest.

The proposed method aims at estimating at each date the probability density function (PDF) and the prior of both the class(es) of interest and the remaining unknown land-cover classes (for which no ground truth is available) represented as a single unknown information class. In particular, PDFs are approximated by a mixture of suitable basis functions whose free parameters are determined employing the iterative Expectation-Maximization (EM) algorithm (Dempster *et al.*, 1977). Changed pixels are then identified using an iterative labelling strategy based on Markov random fields (MRF) (Solberg *et al.*, 1996) which allows taking into account both spatial and temporal correlation between the two images, as well as properly constraining the probability estimates.

For demonstrating the capabilities of the proposed method, extensive experimental trials have been carried out with different combinations of multispectral, hyperspectral and SAR data. Obtained results confirmed the effectiveness and the reliability of the proposed technique, which provided very promising results. In particular, accuracies are comparable to those achieved with the PCC method in the presence of an exhaustive ground truth for each image both considering the maximum likelihood (ML) classifier (Richards and Jia, 2006), as well as support vector machines (SVM) (Cristianini and Shawe-Taylor, 2000).

## 2. PROBLEM FORMULATION AND ASSUMPTIONS

For the sake of simplicity we will describe the problem and the proposed method under the assumption of a single targeted land-cover transition of interest. The extension to the case of multiple transitions is straightforward.

Let us consider two  $I \times J$  co-registered remote-sensing images  $\mathcal{X}^1 = \{\mathbf{x}_{ij}^1\}_{i,j=1}^{I,J}$ ,  $\mathbf{x}_{ij}^1 \in \mathbb{R}^{D_1}$ , and  $\mathcal{X}^2 = \{\mathbf{x}_{ij}^2\}_{i,j=1}^{I,J}$ ,  $\mathbf{x}_{ij}^2 \in \mathbb{R}^{D_2}$ , referring the same geographical area at times  $t_1$  and  $t_2$ , respectively, where  $\mathbf{x}_{ij}^1$ ,  $\mathbf{x}_{ij}^2$  represent corresponding feature vectors (even derived from different sets of sensors at each date, respectively, and merged using a stacked vector approach (Richards and Jia, 2006)) associated with the pixel at position  $(i, j)$ , and  $D_1$ ,  $D_2$  define respective dimensionalities.

Let  $\Omega^1 = \{\omega^1, \dots, \omega_{N_1}^1\}$  and  $\Omega^2 = \{\omega^2, \dots, \omega_{N_2}^2\}$  be the set of land-cover classes characterizing  $\mathcal{X}^1$  and  $\mathcal{X}^2$ , respectively. In the following, we will denote as  $\omega_{m_1}^1 \in \Omega^1$  and  $\omega_{m_2}^2 \in \Omega^2$  the information classes of interest at  $t_1$  and  $t_2$ , for which  $N_1$  and  $N_2$  labelled training patterns are available, respectively. Hence,  $\omega_{m_1}^1 = \{\Omega^1 - \omega_{m_1}^1\}$  and  $\omega_{m_2}^2 = \{\Omega^2 - \omega_{m_2}^2\}$  will represent corresponding unknown classes (each consisting of the merger of remaining classes, for which no ground truth is available).

Let  $\mathcal{C}^1 = \{C_{ij}^1\}_{i,j=1}^{I,J}$  and  $\mathcal{C}^2 = \{C_{ij}^2\}_{i,j=1}^{I,J}$  denote two sets of labels for  $\mathcal{X}^1$  and  $\mathcal{X}^2$ , respectively, where  $C_{ij}^1 \in \{\omega_{m_1}^1, \omega_{m_1}^1\}$  and  $C_{ij}^2 \in \{\omega_{m_2}^2, \omega_{m_2}^2\}$  are associated with the pixel at position  $(i, j)$ . In this framework, our aim is to identify the two sets  $\tilde{\mathcal{C}}^1$ ,  $\tilde{\mathcal{C}}^2$

maximizing the posterior probability given the two images  $\mathcal{X}^1$ ,  $\mathcal{X}^2$  and, finally, to draw pixels experiencing the targeted land-cover transition from  $\omega_{m_1}^1$  to  $\omega_{m_2}^2$ . This can be formalized as a compound decision problem (Duda *et al.*, 2000):

$$\{\tilde{\mathcal{C}}^1, \tilde{\mathcal{C}}^2\} = \underset{\mathcal{C}^1, \mathcal{C}^2}{\operatorname{argmax}} \{P(\mathcal{C}^1, \mathcal{C}^2 | \mathcal{X}^1, \mathcal{X}^2)\} \quad (1)$$

According with the Bayes theory, finding a solution to (1) is equal to determine the sets of labels maximizing the likelihood  $\mathcal{L}(\mathcal{X}^1, \mathcal{X}^2 | \mathcal{C}^1, \mathcal{C}^2) = P(\mathcal{C}^1, \mathcal{C}^2) \cdot p(\mathcal{X}^1, \mathcal{X}^2 | \mathcal{C}^1, \mathcal{C}^2)$ .

In the reasonable hypothesis of time-conditional independence, the problem can be written as:

$$\{\tilde{\mathcal{C}}^1, \tilde{\mathcal{C}}^2\} = \underset{\mathcal{C}^1, \mathcal{C}^2}{\operatorname{argmax}} \{\mathcal{L}(\mathcal{X}^1, \mathcal{X}^2 | \mathcal{C}^1, \mathcal{C}^2) = P(\mathcal{C}^1, \mathcal{C}^2) \cdot p(\mathcal{X}^1 | \mathcal{C}^1) \cdot p(\mathcal{X}^2 | \mathcal{C}^2)\} \quad (2)$$

where  $p(\mathcal{X}^1 | \mathcal{C}^1)$  and  $p(\mathcal{X}^2 | \mathcal{C}^2)$  represent the conditional PDFs at  $t_1$  and  $t_2$  respectively.

## 3. PROPOSED PARTIALLY-SUPERVISED CHANGE-DETECTION TECHNIQUE

For addressing the complex task described in Section 2, we propose a novel partially-supervised technique aimed at approximating the class-conditional densities  $p(\mathcal{X}^1 | \mathcal{C}^1)$ ,  $p(\mathcal{X}^2 | \mathcal{C}^2)$  as mixtures of suitable basis kernel functions and estimating the joint prior probability  $P(\mathcal{C}^1, \mathcal{C}^2)$  properly taking into consideration the spatio-temporal context.

The rationale is based on the observation that the PDF of an image can be always approximated by a mixture of suitable kernels (i.e., Parzen density estimation (Duda *et al.*, 2000)). Accordingly, similarly to what is commonly done in the context of Radial Basis Function Neural Networks (RBF-NN) (Bruzzone and Fernández-Prieto, 1999), we model for each pixel of both images the PDFs  $p(\mathbf{x}_{ij}^1)$ ,  $p(\mathbf{x}_{ij}^2)$  as a mixture of  $K$  circularly symmetric multivariate Gaussian functions. Kernel parameters (i.e., centres and variances) are initialized using the  $k$ -mean clustering algorithm (Bruzzone and Fernández-Prieto, 1999), whereas final estimates are obtained by using the EM algorithm (Dempster *et al.*, 1977). Then, class-conditional densities of the interest class  $p(\mathbf{x}_{ij}^1 | \omega_{m_1}^1)$ ,  $p(\mathbf{x}_{ij}^2 | \omega_{m_2}^2)$  are modelled by properly weighting the resulting set of kernels using again the EM algorithm over the training samples available for  $\omega_{m_1}^1$  and  $\omega_{m_2}^2$ . This is somewhat analogous to the training phase of RBF-NN which is generally carried out in two steps: i) selection of centres and variances of the kernel functions associated with hidden units on the basis of clustering techniques; and ii) computation of weights associated with the connections between the hidden and output layers on the basis of available training patterns.

The PDF of the entire image is itself a mixture of the interest and unknown class-conditional densities, weighted by corresponding prior probabilities. Accordingly, we obtain a first rough approximation for  $p(\mathbf{x}_{ij}^1 | \omega_{m_1}^1)$ ,  $p(\mathbf{x}_{ij}^2 | \omega_{m_2}^2)$  initializing priors to 0.5. Afterwards, estimates are refined using a novel MRF-based iterative labelling strategy accounting for spatio-temporal correlation, which permits to model  $P(\mathcal{C}_{ij}^1, \mathcal{C}_{ij}^2)$  mutually considering the local neighbourhood of each pixel in the two images. Finally, changed pixels are identified and associated with the targeted land-cover transition by minimizing a

proper energy function. In the following, we will first introduce the method adopted for modelling both the class-conditional densities and the joint prior probability; then, we will present the iterative strategy for identifying the sets of labels  $\tilde{C}^1, \tilde{C}^2$  maximizing the likelihood  $\mathcal{L}(\mathcal{X}^1, \mathcal{X}^2 | \mathcal{C}^1, \mathcal{C}^2)$ .

### 3.1 Conditional Density Modelling

Computing  $p(\mathcal{X}^t | \mathcal{C}^t)$ ,  $t=1,2$ , requires at the two considered dates the estimation of the class-conditional densities  $p(\mathbf{x}_{ij}^t | \omega_{int}^t)$  and  $p(\mathbf{x}_{ij}^t | \omega_{unk}^t)$ ,  $\forall \mathbf{x}_{ij}^t \in \mathcal{X}^t$ . The proposed approach is based on the observation that the PDF of each pixel can be modelled as a mixture of the conditional PDFs of both the interest and unknown classes:

$$p(\mathbf{x}_{ij}^t) = P(\omega_{int}^t) \cdot p(\mathbf{x}_{ij}^t | \omega_{int}^t) + P(\omega_{unk}^t) \cdot p(\mathbf{x}_{ij}^t | \omega_{unk}^t) \quad (3)$$

According with Parzen density estimation (Duda *et al.*, 2000), we aim at obtaining a reliable nonparametric estimate for  $p(\mathbf{x}_{ij}^t)$  as a mixture of a suitable set of kernel functions  $\Phi^t = \{\phi^k(\cdot)\}_{k=1}^K$ :

$$\hat{p}(\mathbf{x}_{ij}^t) = \sum_{k=1}^K w_k^t \cdot \phi^k(\mathbf{x}_{ij}^t) \quad (4)$$

where  $K$  denotes the number of kernels (a free parameter to be set by the user), and  $\mathcal{W}^t = \{w_k^t\}_{k=1}^K$  represent the weights regulating the contribution of each kernel.

However, since the density  $p(\mathbf{x}_{ij}^t)$  is given by a linear combination of  $p(\mathbf{x}_{ij}^t | \omega_{int}^t)$  and  $p(\mathbf{x}_{ij}^t | \omega_{unk}^t)$ , it is worth noting that, if the set of kernels  $\Phi^t$  allows obtaining a reliable estimate  $\hat{p}(\mathbf{x}_{ij}^t)$ , then also both the class-conditional densities can be reliably approximated as a linear combination of  $\Phi^t$ . Hence, they can be estimated as:

$$\hat{p}(\mathbf{x}_{ij}^t | \omega_{int}^t) = \sum_{k=1}^K w_{int\ k}^t \cdot \phi^k(\mathbf{x}_{ij}^t) \quad (5)$$

$$\hat{p}(\mathbf{x}_{ij}^t | \omega_{unk}^t) = \sum_{k=1}^K w_{unk\ k}^t \cdot \phi^k(\mathbf{x}_{ij}^t) \quad (6)$$

where  $\mathcal{W}_{int}^t = \{w_{int\ k}^t\}_{k=1}^K$ ,  $\mathcal{W}_{unk}^t = \{w_{unk\ k}^t\}_{k=1}^K$  are the weights regulating the contribution of each kernel for the interest and unknown classes, respectively.

As commonly done in the literature we consider normalized isotropic Gaussian kernels, i.e.

$$\phi^k(\mathbf{x}_{ij}^t) = \frac{1}{(\sqrt{2\pi\sigma_k^{2t}})^{D_t}} \exp\left[-\frac{\|\mathbf{x}_{ij}^t - \boldsymbol{\mu}_k^t\|^2}{2\sigma_k^{2t}}\right] \quad (7)$$

where  $\boldsymbol{\mu}_k^t$  is the centre and  $\sigma_k^{2t}$  is the variance (which tunes the smoothness of the estimate).

In the following, we describe into details the procedures adopted for estimating  $\hat{p}(\mathbf{x}_{ij}^t)$ ,  $\hat{p}(\mathbf{x}_{ij}^t | \omega_{int}^t)$  and  $\hat{p}(\mathbf{x}_{ij}^t | \omega_{unk}^t)$ , respectively.

#### 3.1.1 Estimation of $\hat{p}(\mathbf{x}_{ij}^t)$

For computing both the centres  $M^t = \{\boldsymbol{\mu}_k^t\}_{k=1}^K$  and variances  $\Sigma^t = \{\sigma_k^{2t}\}_{k=1}^K$  of all the kernels, as well as the set of weights  $\mathcal{W}^t$  defining  $\hat{p}(\mathbf{x}_{ij}^t)$ , we employ the EM algorithm over all the pixels of  $\mathcal{X}^t$ . EM allows determining the maximum likelihood (ML) estimator of the parameters characterizing a certain distribution in the presence of incomplete observations (Dempster *et al.*, 1977). Indeed, our objective is to identify the ML estimate for the set of parameters  $\theta^t = \{M^t, \Sigma^t, \mathcal{W}^t\} = \{\boldsymbol{\mu}_k^t, \sigma_k^{2t}, w_k^t\}_{k=1}^K$  that allows maximizing the log-likelihood of  $\mathcal{X}^t$ , i.e.

$$\ln \mathcal{L}(\theta^t) = \ln [p(\mathcal{X}^t | \theta^t)] = \sum_{i,j=1}^{I,J} \ln [\hat{p}(\mathbf{x}_{ij}^t)] \quad (8)$$

At each iteration  $l$ , the set of estimated parameters  $[\theta^t]^{(l)}$  provides an increase in the log-likelihood until a local maximum is reached, i.e.  $\ln \mathcal{L}([\theta^t]^{(l)}) \geq \ln \mathcal{L}([\theta^t]^{(l-1)})$ .

For simplicity, weights are initially set to  $1/K$ , whereas kernel parameters are initialized using the  $k$ -means clustering algorithm (Bruzzone and Fernández-Prieto, 1999) fixing the number of clusters equal to  $K$ . In particular, centres and variances of the Gaussians are initialized to the centres and variances of the resulting clusters. Then, according with Dempster *et al.*, 1977, the updated estimates for the unknown parameters are given by:

$$[w_k^t]^{(l)} = \frac{\sum_{i,j=1}^{I,J} [w_k^t]^{(l-1)} \cdot [\phi^k(\mathbf{x}_{ij}^t)]^{(l-1)}}{I \cdot J \cdot [\hat{p}(\mathbf{x}_{ij}^t)]^{(l-1)}} \quad (9)$$

$$[\boldsymbol{\mu}_k^t]^{(l)} = \frac{\sum_{i,j=1}^{I,J} [w_k^t]^{(l-1)} \cdot [\phi^k(\mathbf{x}_{ij}^t)]^{(l-1)} \cdot \mathbf{x}_{ij}^t}{\sum_{i,j=1}^{I,J} [w_k^t]^{(l-1)} \cdot [\phi^k(\mathbf{x}_{ij}^t)]^{(l-1)}} \quad (10)$$

$$[\sigma_k^{2t}]^{(l)} = \frac{\sum_{i,j=1}^{I,J} [w_k^t]^{(l-1)} \cdot [\phi^k(\mathbf{x}_{ij}^t)]^{(l-1)} \cdot \|\mathbf{x}_{ij}^t - [\boldsymbol{\mu}_k^t]^{(l)}\|^2}{D_t \cdot \sum_{i,j=1}^{I,J} [w_k^t]^{(l-1)} \cdot [\phi^k(\mathbf{x}_{ij}^t)]^{(l-1)}} \quad (11)$$

Reasonably, we assume that convergence is reached when the relative increase in the log-likelihood is lower than a prefixed threshold  $\varepsilon$ .

#### 3.1.2 Estimation of $\hat{p}(\mathbf{x}_{ij}^t | \omega_{int}^t)$

Once  $M^t$  and  $\Sigma^t$  have been determined (and hence the set of kernels  $\Phi^t$  properly defined), we exploit the available ground truth for the class of interest  $\omega_{int}^t$  at each date for deriving the estimate of the corresponding conditional density  $\hat{p}(\mathbf{x}_{ij}^t | \omega_{int}^t)$ . In particular, the set of weights  $\mathcal{W}_{int}^t$  associated with  $\omega_{int}^t$  is determined using again the EM algorithm, but solely on the available training samples  $\mathcal{T}^t = \{\mathbf{x}_{ij}^t \in \mathcal{X}^t | y_{ij} = \omega_{int}^t\}$ ,  $|\mathcal{T}^t| = N^t$ , where  $y_{ij}$  denotes the true label for pixel at position  $(i, j)$ .

Weights are initialized to  $1/K$ , and then updated (according with Dempster *et al.*, 1977) using the following equation:

$$[W_{\text{int } k}^t]^{(l)} = \frac{\sum_{i,j=1}^{I,J} \alpha_{ij} \cdot \frac{[W_{\text{int } k}^t]^{(l-1)} \cdot \phi_k^t(\mathbf{x}_{ij}^t)}{[\hat{p}(\mathbf{x}_{ij}^t | \omega_{\text{int}}^t)]^{(l-1)}}}{N_t}, \alpha_{ij} = \begin{cases} 1 & \text{if } \mathbf{x}_{ij}^t \in \mathcal{T}' \\ 0 & \text{if } \mathbf{x}_{ij}^t \notin \mathcal{T}' \end{cases} \quad (12)$$

The corresponding log-likelihood is given by:

$$\ln \mathcal{L}(\theta' \equiv \mathcal{W}_{\text{int}}^t) = \sum_{i,j=1}^{I,J} \alpha_{ij} \cdot \ln[\hat{p}(\mathbf{x}_{ij}^t | \omega_{\text{int}}^t)] \quad (13)$$

Even in this case we assume that convergence is reached if the relative increase in the log-likelihood is lower than  $\varepsilon$ .

### 3.1.3 Estimation of $\hat{p}(\mathbf{x}_{ij}^t | \omega_{\text{int}}^t)$

It is worth noting that Eq. (3) can be re-written as:

$$P(\omega_{\text{int}}^t) \cdot \sum_{k=1}^K W_{\text{int } k}^t \cdot \phi_k^t(\mathbf{x}_{ij}^t) = \sum_{k=1}^K w_k^t \cdot \phi_k^t(\mathbf{x}_{ij}^t) - P(\omega_{\text{int}}^t) \cdot \sum_{k=1}^K W_{\text{int } k}^t \cdot \phi_k^t(\mathbf{x}_{ij}^t) \quad (14)$$

Hence,  $\forall k=1, \dots, K$  it holds  $P(\omega_{\text{int}}^t) \cdot W_{\text{int } k}^t = w_k^t - P(\omega_{\text{int}}^t) \cdot W_{\text{int } k}^t$ . Since  $P(\omega_{\text{int}}^t) + P(\omega_{\text{int}}^t) = 1$ , we have:

$$W_{\text{int } k}^t = \frac{w_k^t - P(\omega_{\text{int}}^t) \cdot W_{\text{int } k}^t}{P(\omega_{\text{int}}^t)} = \frac{w_k^t - P(\omega_{\text{int}}^t) \cdot W_{\text{int } k}^t}{1 - P(\omega_{\text{int}}^t)} \quad (15)$$

Then, as  $M^t$ ,  $\Sigma^t$ ,  $\mathcal{W}^t$  and  $\mathcal{W}_{\text{int}}^t$  have been determined, we can compute  $\hat{p}(\mathbf{x}_{ij}^t | \omega_{\text{int}}^t)$  substituting (15) into (6), upon it is possible to obtain a reliable estimate  $\hat{P}(\omega_{\text{int}}^t)$  for the prior probability of the class of interest. This can be properly accomplished throughout the iterative labelling phase presented below.

## 3.2 Joint Prior Modelling

For modelling the joint prior  $P(\mathcal{C}^1, \mathcal{C}^2)$ , we propose an approach based on MRF (Solberg *et al.*, 1996). In particular, we assume that the couple of labels  $C_{ij}^1$ ,  $C_{ij}^2$  associated with pixel at position  $(i, j)$  at times  $t_1$  and  $t_2$  depends on the couples of labels associated with pixels belonging to the spatial neighbourhood  $\mathcal{G}_{ij}$  of  $(i, j)$  at the two dates (we always considered first order neighbourhoods).

In other words, the higher the number of its spatial neighbours experiencing a certain land-cover transition is, the higher the probability for a given pixel of experiencing the same transition is. In this hypothesis, according with the MRF theory, it holds the equivalence:

$$P(C_{ij}^1, C_{ij}^2 | C_{gh}^1, C_{gh}^2; (g, h) \in \mathcal{G}_{ij}) = Z^{-1} \cdot \exp[-U_{\text{context}}(C_{ij}^1, C_{ij}^2)] \quad (16)$$

where  $Z = Z(\mathcal{G}_{ij})$  is a normalizing constant called *partition function*, while  $U_{\text{context}}$  is a *Gibbs energy function* (accounting for the spatio-temporal context) of the form:

$$U_{\text{context}}(C_{ij}^1, C_{ij}^2) = - \sum_{(g, h) \in \mathcal{G}_{ij}} \beta \cdot \delta[(C_{ij}^1, C_{ij}^2), (C_{gh}^1, C_{gh}^2)] \quad (17)$$

where  $\beta > 0$  tunes the influence of the context and  $\delta$  is the Kronecker delta function defined as:

$$\delta[(C_{ij}^1, C_{ij}^2), (C_{gh}^1, C_{gh}^2)] = \begin{cases} 1 & \text{if } (C_{ij}^1, C_{ij}^2) = (C_{gh}^1, C_{gh}^2) \\ 0 & \text{if } (C_{ij}^1, C_{ij}^2) \neq (C_{gh}^1, C_{gh}^2) \end{cases} \quad (18)$$

## 3.3 Iterative Labelling

Solving Eq. (2) is equivalent to maximize the log-likelihood  $\ln \mathcal{L}(\mathcal{X}^1, \mathcal{X}^2 | \mathcal{C}^1, \mathcal{C}^2)$ , which can be written as:

$$\ln \mathcal{L}(\mathcal{X}^1, \mathcal{X}^2 | \mathcal{C}^1, \mathcal{C}^2) = - \sum_{t=1}^2 U_{\text{data}}^t(\mathcal{X}^t, \mathcal{C}^t) - U_{\text{context}}(\mathcal{C}^1, \mathcal{C}^2) - \ln(Z) \quad (19)$$

where  $U_{\text{data}}^t(\mathcal{X}^t, \mathcal{C}^t) = -\ln[p(\mathcal{X}^t | \mathcal{C}^t)]$ , represents the class-conditional energy function at date  $t=1, 2$ , while  $U_{\text{context}}$  is given by (17). Since  $Z$  solely depends on the selected type of neighbourhood, the final problem becomes solving:

$$\{\tilde{\mathcal{C}}^1, \tilde{\mathcal{C}}^2\} = \underset{\mathcal{C}^1, \mathcal{C}^2}{\text{argmin}} \{U_{\text{context}}(\mathcal{C}^1, \mathcal{C}^2) + U_{\text{data}}^1(\mathcal{X}^1, \mathcal{C}^1) + U_{\text{data}}^2(\mathcal{X}^2, \mathcal{C}^2)\} \quad (20)$$

To this aim, we propose a strategy based on the Iterated Conditional Modes (ICM) algorithm (Besag, 1996) which allows maximizing local conditional probabilities sequentially. In particular, at each iteration  $l$  we update the estimated prior probabilities for the class of interest at each date  $\hat{P}(\omega_{\text{int}}^t)$  and, accordingly, also the class-conditional densities of the unknown classes  $\hat{p}(\mathbf{x}_{ij}^t | \omega_{\text{int}}^t)$ . The algorithm works as follows:

- Step 1.* After estimating  $\hat{p}(\mathbf{x}_{ij}^t | \omega_{\text{int}}^t)$ ,  $\forall \mathbf{x}_{ij}^t \in \mathcal{X}^t$ ,  $t=1, 2$  following the approach described in the previous paragraphs, set  $\hat{P}(\omega_{\text{int}}^t) = 0.5$  (no prior knowledge is assumed to be available about the true prior  $P(\omega_{\text{int}}^t)$ ) and compute the conditional density of the unknown classes  $\hat{p}(\mathbf{x}_{ij}^t | \omega_{\text{int}}^t)$  accordingly;
- Step 2.* Derive the initial sets of labels  $\tilde{\mathcal{C}}^1$ ,  $\tilde{\mathcal{C}}^2$  by solely minimizing the non-contextual terms of Eq. (20), i.e.  $\{\tilde{\mathcal{C}}^1, \tilde{\mathcal{C}}^2\} = \underset{\mathcal{C}^1, \mathcal{C}^2}{\text{argmin}} \{U_{\text{data}}^1(\mathcal{X}^1, \mathcal{C}^1) + U_{\text{data}}^2(\mathcal{X}^2, \mathcal{C}^2)\}$ ;
- Step 3.* On the basis of current  $\tilde{\mathcal{C}}^1$  and  $\tilde{\mathcal{C}}^2$ , compute the new estimated prior probabilities ratioing the number of pixels associated with the class of interest over the whole number of pixels, i.e.  $\hat{P}(\omega_{\text{int}}^t) = |\tilde{\mathcal{C}}_{\text{int}}^t| / (I \cdot J)$ ,  $\tilde{\mathcal{C}}_{\text{int}}^t = \{C_{ij}^t | C_{ij}^t \in \tilde{\mathcal{C}}^t, C_{ij}^t = \omega_{\text{int}}^t\}_{i,j=1}^{I,J}$ ; then, update the class-conditional densities for the unknown classes  $\hat{p}(\mathbf{x}_{ij}^t | \omega_{\text{int}}^t)$  accordingly;
- Step 4.* Update  $\tilde{\mathcal{C}}^1$  and  $\tilde{\mathcal{C}}^2$  according with Eq. (20);
- Step 5.* Repeat Step 3 and Step 4 until no changes occur between successive iterations.

At the end of the process, the final targeted change-detection map  $\mathcal{C}^*$  is defined as:

$$\mathcal{C}^* = \{C_{ij}^* \}_{i,j=1}^{I,J}, C_{ij}^* = \begin{cases} 1 & \text{if } C_{ij}^1 = \omega_{\text{int}}^1 \text{ and } C_{ij}^2 = \omega_{\text{int}}^2 \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

## 4. EXPERIMENTAL RESULTS

In order to assess the effectiveness of the proposed technique, we carried out several experiments with different combinations

of multispectral, hyperspectral and SAR data. Nevertheless, due to space constraints, we will focus the attention solely on a representative change-detection problem referring to an intense farming area experiencing several land-cover transitions located in Barrax, a village close to Albacete (Castilla-La Mancha, Spain). In particular, available data consist in: two Landsat-5 Thematic Mapper (TM) images acquired on 15<sup>th</sup> July 2003 and 17<sup>th</sup> July 2004, respectively (as generally done in the literature among the 7 spectral bands we did not consider the low-resolution band associated with the thermal infrared channel); one PROBA CHRIS image (composed by 63 spectral bands with centre wavelengths from 400 to 1050 nm) acquired on 16<sup>th</sup> July 2004; and one Envisat ASAR alternate polarization (VV, HH) image (despeckled using a 3×3 Gamma filter) acquired on 18<sup>th</sup> July 2004. All of them have been properly co-registered to a common spatial geometry of 30 meters and a study area of 512 × 512 pixels has been selected. July 2003 will be referred to as  $t_1$ , whereas July 2004 will be referred to as  $t_2$  (no significant changes indeed occurred between 16<sup>th</sup> and 18<sup>th</sup> July 2004). From the original available images we derived the following three datasets (using the stacked vector approach for multi-source data fusion (Richards and Jia, 2006)) composed by: i) the 6 Landsat TM bands at both  $t_1$  and  $t_2$  (i.e.,  $D_1=D_2=6$ ) [Dataset I]; ii) the 6 Landsat TM bands at  $t_1$  and the merger of the 2 Envisat ASAR backscattering intensity images with the 6 Landsat TM bands at  $t_2$  (i.e.,  $D_1=6$  and  $D_2=8$ ) [Dataset II]; and iii) the 6 Landsat TM bands at  $t_1$  and the 63 PROBA CHRIS bands at  $t_2$  (i.e.,  $D_1=6$  and  $D_2=63$ ) [Dataset III]. However, it is worth noting that our objective is not to seek for a set of features at both dates which could be more effective for solving the investigated problem, but rather to demonstrate that the presented method is even capable of effectively handling different types of data at the two times.

As described in Section 3, the user is required to set the number of Gaussian kernels  $K$  to be employed for approximating the PDFs. Hence, in order to understand how significant the selection of this free parameter is, we performed a series of experiments varying  $K$  from 20 to 120 with steps of 10.

According with a variety of experiments on toy and real datasets we fixed  $\varepsilon=10^{-4}$  and  $\beta=10^2$ .

In all the trials we employed the  $k$ -means clustering for initializing both centres and variances of kernel functions. Nevertheless, the very high complexity of the algorithm (i.e., approximately  $O(N^{D(K+1)} \log N)$ , where  $N$  and  $D$  represent the number of samples to be clustered and their dimensionality, respectively (Inaba *et al.*, 1994)), prevented us from using all the patterns of each investigated image at a time, as this would have required a very high computational burden. In order to overcome this limitation, for each image we ran the  $k$ -means algorithm on a random subset containing one third of the total amount of samples. However, as this might affect the final change-detection accuracies of the proposed technique, for each number of considered kernels  $K$ , we performed 10 different trials running each time the  $k$ -means clustering on a different random subset. Moreover, we finally also combined the 10 resulting change-detection maps through a majority voting ensemble.

For validating the potentialities of the presented method, we compared the results with those obtained by supervised PCC. In particular, we considered ML and SVM fully-supervised classifiers trained by exploiting a complete ground truth for all the land-cover classes characterizing each considered date. ML is a simple yet generally rather effective statistical classifier, which does not require the user to set any free parameter (Richards and Jia, 2006). SVM are advanced state-of-art classifiers, which proved capable of outperforming other traditional approaches

(Cristianini and Shawe-Taylor, 2000). For the selection of the two free parameters (i.e., a penalization parameter and the variance of considered Gaussian kernels) we employed a 10-fold cross-validation strategy (Duda *et al.*, 2000).

Available prior knowledge has been used for defining regions of interest composed on the whole by 21941 pixels whose ground truth was known at the two times, respectively. 10 land-cover classes have been considered at  $t_1$ , whereas 9 have been taken into consideration at  $t_2$ . At both dates, from all the available labeled samples we defined two spatially-disjoint training sets (see Table 1). This means that there is no overlapping between training samples at  $t_1$  and  $t_2$ . All of them have been used for training both the ML and SVM supervised classifiers at each time.

| Land-cover class | $t_1$ (July 2003) | $t_2$ (July 2004) |
|------------------|-------------------|-------------------|
| alfalfa          | 2031              | 634               |
| bare soil        | 2585              | –                 |
| corn             | 1737              | 2664              |
| garlic           | 101               | 302               |
| grasslands       | –                 | 42                |
| onions           | 213               | 220               |
| poppy            | 336               | –                 |
| potatoes         | 208               | 283               |
| spring crops     | –                 | 3318              |
| stubble          | 2416              | 2247              |
| sugar beet       | 365               | –                 |
| sunflower        | –                 | 449               |
| wheat            | 369               | –                 |
| total            | 10361             | 10159             |

Table 1. Number of spatially-disjoint training samples considered at the two dates.

Change-detection results have been evaluated (over those samples whose land-cover class is known at both dates) in terms of: percentage overall accuracy  $OA\%$  (i.e., the percentage of samples correctly identified as both changed or unchanged over the whole number of samples), and kappa coefficient of accuracy (which also takes into consideration errors and their type) (Richards and Jia, 2006).

Among the different land-cover transitions occurred between the two dates, here we take into consideration (one at a time) two of them, namely “bare soil to spring crops” and “alfalfa to corn” (experienced by 4583 and 1035 pixels, respectively, over the whole available 21941 whose ground truth was known at both times).

In our trials, we empirically experienced that a common range for  $K$  resulting in average high detection accuracies spans from 60 to 80. When instead nearing the lower or the upper bound of the considered interval, performances tend to vary depending on the specific land-cover transition of interest. Accordingly, in Table 2 and Table 3 we show the results obtained with the proposed technique for  $K=40, 60, 80, 100$ . In particular, we report the median over the 10 realizations with different  $k$ -means clustering initialization. Moreover, also accuracies finally obtained with the majority voting ensemble (denoted as  $PSCD_{MV}$ ), as well as those obtained by supervised PCC using ML and SVM (denoted as  $PCC_{ML}$  and  $PCC_{SVM}$ , respectively) are presented.

While investigating the “bare soil to spring crops” transition with the proposed PSCD technique, from all the training pixels reported in Table 1, we considered the only 2585 available for bare soil at  $t_1$  and the only 3318 spatially-disjoint available for

spring crops at  $t_2$ . Obtained results are very satisfactory, as confirmed by both the high kappa and  $OA\%$  values reported in Table 2 (always higher than 0.79 and 92, respectively). Moreover, by employing the majority voting ensemble it is possible to further improve the performances and obtaining for both indexes accuracies even closer to those obtained by  $PCC_{ML}$  and  $PCC_{SVM}$  with a fully-supervised training at both dates.

| $K$         | Dataset I |        | Dataset II |        | Dataset III |        |
|-------------|-----------|--------|------------|--------|-------------|--------|
|             | kappa     | $OA\%$ | kappa      | $OA\%$ | kappa       | $OA\%$ |
| 40          | 0.8925    | 96.59  | 0.8403     | 95.09  | 0.7955      | 92.61  |
| 60          | 0.8903    | 96.53  | 0.8458     | 95.21  | 0.8488      | 94.97  |
| 80          | 0.8822    | 96.25  | 0.8412     | 95.11  | 0.8441      | 94.80  |
| 100         | 0.8878    | 96.47  | 0.8470     | 95.31  | 0.8502      | 94.93  |
| $PSCD_{MV}$ | 0.9076    | 97.08  | 0.8649     | 95.82  | 0.8622      | 95.32  |
| $PCC_{ML}$  | 0.9339    | 97.88  | 0.9323     | 97.83  | 0.9330      | 97.86  |
| $PCC_{SVM}$ | 0.9501    | 98.36  | 0.9359     | 97.90  | 0.9703      | 99.02  |

Table 2. kappa coefficient of accuracy and  $OA\%$  obtained for the “bare soil to spring crops” land-cover transition.

While addressing the “alfalfa to corn” transition with the proposed PSCD technique, from all the training pixels reported in Table 1, we considered the only 2031 available for alfalfa at  $t_1$  and the only 2664 spatially-disjoint available for corn at  $t_2$ . Such a transition is rather difficult to characterize, as only experienced by few fields in the considered area. Indeed, according with the results in Table 3, this is confirmed by the very low accuracies obtained by  $PCC_{ML}$  despite fully-supervised training. Instead, in the light of the high complexity of the problem, performances exhibited by the proposed method are very promising, especially for Dataset II and Dataset III. Moreover, with the majority voting ensemble the gap with respect to  $PCC_{SVM}$  becomes very small (Dataset II) or it is even possible to outperform results obtained with SVMs (Dataset III).

| $K$         | Dataset I |        | Dataset II |        | Dataset III |        |
|-------------|-----------|--------|------------|--------|-------------|--------|
|             | kappa     | $OA\%$ | kappa      | $OA\%$ | kappa       | $OA\%$ |
| 40          | 0.6725    | 97.36  | 0.8549     | 98.79  | 0.8117      | 98.40  |
| 60          | 0.7110    | 97.79  | 0.8190     | 98.49  | 0.8343      | 98.60  |
| 80          | 0.6530    | 97.25  | 0.8172     | 98.45  | 0.8206      | 98.45  |
| 100         | 0.6553    | 97.32  | 0.7233     | 97.92  | 0.8367      | 98.62  |
| $PSCD_{MV}$ | 0.7524    | 98.16  | 0.8896     | 99.06  | 0.9079      | 99.22  |
| $PCC_{ML}$  | 0.5163    | 96.87  | 0.5134     | 96.86  | 0.5353      | 97.05  |
| $PCC_{SVM}$ | 0.9003    | 99.08  | 0.9244     | 99.32  | 0.8969      | 99.04  |

Table 3. kappa coefficient of accuracy and  $OA\%$  obtained for the “alfalfa to corn” land-cover transition.

## 5. CONCLUSIONS

In this paper we presented a novel partially-supervised change-detection (PSCD) technique capable of addressing targeted change-detection problems where the objective is to identify one (or few) targeted land-cover transitions, under the assumption that ground-truth information is available for the only (few) class(es) of interests at the two investigated dates.

In this context, either supervised or unsupervised standard approaches cannot be effectively employed. The proposed method, instead, allows exploiting the only prior knowledge available for the specific land-cover classes of interest at the

two times, while providing accuracies comparable with those of fully-supervised methods. In particular, the PSCD technique relies on a partially-supervised approach and jointly exploits the Expectation-Maximization (EM) algorithm and an iterative labelling strategy based on Markov random fields (MRF) accounting for spatial and temporal correlation between the two images. Moreover, it also allows handling images acquired by different sensors at the two considered times. Experimental results on multi-sensor datasets derived from multispectral, hyperspectral and SAR data confirmed the effectiveness and the reliability of the proposed technique

## REFERENCES

- Besag, J., 1986. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B*, 48, pp. 259-302.
- Bruzzone, L., and Fernández-Prieto, D., 1999. A Technique for the Selection of Kernel-Function Parameters in RBF Neural Networks for Classification of Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.*, 37(2), pp. 1179-1184.
- Coppin, P., Jonckheere, I., Nackaerts, K., and Muys, B., 2004. Digital change detection methods in ecosystem monitoring: a review. *Int. J. Remote Sens.*, 25(9), pp. 1565-1596.
- Cristianini N., and Shawe-Taylor, J., 2000. *An Introduction to Support Vector Machines*. Cambridge University Press.
- Dempster, A., Laird, N., and Rubin, D., 1977. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *The Royal Statistical Society, Series B*, 39(1), pp. 1-38.
- Duda, R. O., Hart, P. E., and Stork, D. G., 2000. *Pattern Classification*. Wiley, New York.
- Inaba, M., Katoh, N., and Imai, H., 1994. Applications of weighted voronoi diagrams and randomization to variance-based k-clustering. *Proceedings of the 10<sup>th</sup> Annual Symposium on Computational Geometry*, pp. 332-339.
- Jensen, J. R., 2009. *Remote Sensing Of The Environment*. Pearson, Upper Saddle River.
- Jeon, B., and Landgrebe, D. A., 1999. Partially supervised classification using weighted unsupervised clustering. *IEEE Trans. Geosci. Remote Sens.*, 37(3), pp. 1073-1079.
- Lu, D., Mausel, P., Brondizio, E., and Moran, E., 2004. Change detection techniques. *Int. J. Remote Sens.*, 25(12), pp. 2365-2407.
- Radke, R. J., Andra, S., Al-Kofahi, O., and Roysam, B., 2005. Image Change Detection Algorithms: A systematic Survey. *IEEE Trans. Geosci. Remote Sens.*, 14(3), pp. 294-307, 2005.
- Richards, J. A., and Jia, X., 2006. *Remote Sensing Digital Image Analysis. An Introduction*. Springer, Berlin.
- Singh, A., 1989. Digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.*, 10(6), pp. 989-1003.
- Solberg, A. H. S., Taxt, T., and Jain, A. K., 1996. A Markov Random Field Model for Classification of Multisource Satellite Imagery. *IEEE Trans. Geosci. Remote Sens.*, 34(1), pp. 100-113.