

Iterative Classifiers Combination Model for Change Detection in Remote Sensing Imagery

Rachid Hedjam, Margaret Kalacska, Max Mignotte, Hossein Ziaei Nafchi, and Mohamed Cheriet

Abstract—In this paper, we propose a new unsupervised change detection method designed to analyze multispectral remotely sensed image pairs. It is formulated as a segmentation problem to discriminate the changed class from the unchanged class in the difference images. The proposed method is in the category of the committee machine learning model that utilizes an ensemble of classifiers (i.e., the set of segmentation results obtained by several thresholding methods) with a dynamic structure type. More specifically, in order to obtain the final “change/no-change” output, the responses of several classifiers are combined by means of a mechanism that involves the input data (the difference image) under an iterative Bayesian–Markovian framework. The proposed method is evaluated and compared to previously published results using satellite imagery.

Index Terms—Change detection, classifiers combination, image fusion, multispectral multitemporal image analysis, remote sensing imagery.

I. INTRODUCTION

WITH the increasing availability of remotely sensed data, there has been substantial interest in change detection, locating areas that may have changed in order to monitor the differences over time [1]–[3].

Change detection involves the analysis of bitemporal images acquired over a same geographical area at two different dates, or multitemporal images acquired at multiple dates (i.e., time series), in order to identify change that occurred over the time period considered. Change detection is often addressed through one of two approaches, namely, supervised approach and unsupervised approach. The former requires a temporal reference data for the training phase [5], [6], whereas the latter is based on a direct comparison of the images considered without additional prior information [2], [7]–[9]. Depending on the final goal of the application, there are methods for the detection of several types of changes (i.e., multiple-change detection) [7], [8], [10] and methods for the detection of one type of change (i.e., binary change detection, i.e., change or no change) [2], [11].

Manuscript received July 31, 2015; revised February 3, 2016 and May 30, 2016; accepted July 2, 2016. Date of publication August 24, 2016; date of current version September 30, 2016. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and in part by Fonds de Recherche du Québec—Nature et technologies (FRQNT).

R. Hedjam and M. Kalacska are with the Department of Geography, McGill University, Montréal, QC H3A 0B9, Canada (e-mail: rachid.hedjam@mcgill.ca; rhedjam@synchronmedia.ca; margaret.kalacska@mcgill.ca).

M. Mignotte is with the Department of Computer Science and Operations Research (DIRO), University of Montréal, Montréal, QC H3T 1J4, Canada.

H. Ziaei Nafchi and M. Cheriet are with École de Technologie Supérieure, Montréal, QC H3C 1K3, Canada.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2016.2593982

The majority of the unsupervised binary change detection techniques are based on the analysis of a “difference image” or “change index” generated from two images acquired at two different times [7] (see Fig. 1). There are different ways to generate the difference image according to the nature of the input images [e.g., optical, synthetic aperture radar (SAR) and longwave hyperspectral]. In the analysis of multispectral imagery, one of the most popular ways to compute the difference image is by change vector analysis [1], in which, for each pair of corresponding pixels, a “spectral change vector” is computed as the difference between the feature vectors at the two times. Then, the pixel values in the difference image are associated with the modules of the spectral change vectors. The result is that the gray-level values of the changed pixels are larger than those of the unchanged pixels [2]. Other techniques such as image ratioing based on the ratio between bitemporal images are also often used to generate the difference image as well [1]. Meanwhile, if the input images are SAR images, which are affected by multiplicative speckle noise, the log-ratio operator between the bitemporal images is preferred [12]. Binary change detection techniques aim to discriminate the pixels in the difference image into regions associated into “changed” and “unchanged” classes. These techniques are mostly based on statistical modeling and thresholding [13]–[15], clustering [16], fusion based [17], [18], and level set based [4], [19].

The method we propose in this paper belongs to the category of fusion-based methods dealing with the difference image generated from bitemporal multispectral images. The aim is to combine (fuse) several binary maps (obtained by performing simple thresholding techniques on the difference image) in order to generate a final more accurate binary map, which we expect to be more accurate than the various input maps individually. This is motivated by the challenge of determining an optimal threshold [13], [18], [20]: the first is the lack of training samples (ground truth), and therefore, no prior information is available to accurately guide the thresholding; the second is the local validity of the thresholds, due to the specific characteristics of the observed image, which makes it difficult to design generic thresholding algorithms specifically adapted even for one category of images. The concept of combining classifiers for the improvement of the performance of individual classifiers is known in image processing [21] and machine learning fields [22] as a committee machine or a mixture of experts. In the context of machine learning, Dietterich [22] has provided an accessible and informal reasoning, from statistical, computational, and representational viewpoints, of why ensembles can improve results. As stated in [23], despite the fact that the classifiers to be combined are indistinguishable from their training

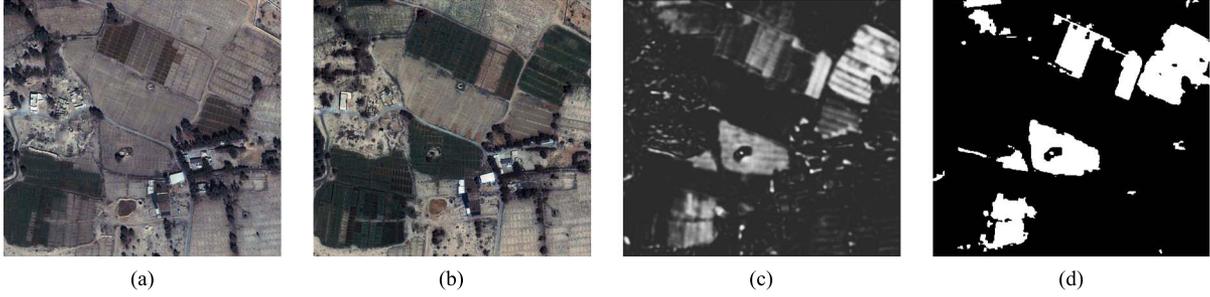


Fig. 1. Riyadh data set is a true-color composite image subset to a size of 430×401 pixels acquired over the region of Riyadh, Saudi Arabia, by the IKONOS sensor [4]. (a) Image acquired in February 2006. (b) Image acquired in October 2006. (c) Difference image. (d) Ground truth.

error (resubstitution), they may have different performance of generalization. Thus, instead of selecting a single classifier, it is beneficial to utilize all of them and then use the “average” of their outputs. It results that the obtained classifier could not be better than the best single classifier, but we reduce or eliminate the risk of selecting an inappropriate single classifier. It is accepted in machine learning to combine classifiers drawn from the same model family (i.e., homogeneous classifiers) [24] or from different (heterogeneous) models [25]–[27].

The proposed change detection method is formulated as a fusion problem of multiple segmentations generated from heterogeneous classifiers to discriminate the changed class from the unchanged class in the difference images. Therefore, since the single classifiers to combine are heterogeneous, some of them rely on the hypothesis of distribution and others do not; our attention is directed to the fusion model which is fed by their outputs (labels), rather than going into details about of each of them. Nevertheless, we limited ourselves to study the influence of the input classifiers all together (the average precision and their diversity) on the accuracy of the estimated change/no-change segmentation. Our fusion decision model is in the category of the committee machine learning model that utilizes an ensemble of classifiers (i.e., the set of segmentation results) with a dynamic structure type. In this class of committee machines, the responses of several classifiers are combined by means of a mechanism that involves the input data (or original image), contrary to the static structure-type-based mixture of experts, which only uses the set of segmentation maps, as the case in [18]. In other words, in order to empower the fusion accuracy, we propose a new iterative strategy of combining the results of different thresholding algorithms while taking into account the statistical characteristic of the input difference image. Henceforth, let us use the term segmentation (i.e., into two classes, namely, “change” and “no-change”) instead of thresholding result. The proposed method is based on the following hypothesis: the estimated segmentation is assumed to be more accurate than all other input segmentations individually, and more accurate input segmentations lead to a better estimated segmentation. Motivated by that, a new iterative scheme of segmentation is proposed. It consists of three iterative steps. The first step consists in stacking the input segmentations provided by different algorithms; in the second step, a new segmentation is estimated from the difference image and the input segmentation set, and the third step is to update the input segmentations. In the update step, the input segmentation

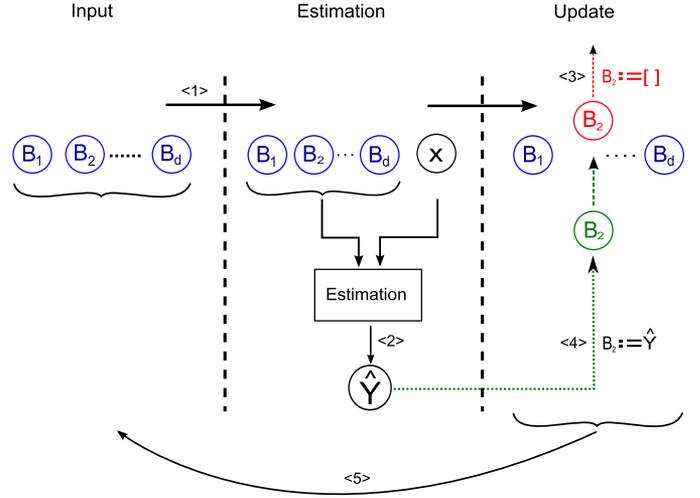


Fig. 2. Proposed iterative model of “change/no-change” segmentation method. The step’s numbers are between “<” and “>.” The sign “:=” means “replaced by.” “[]” means “empty.”

that contributes the least to the segmentation process will be replaced by the new estimated one. The three steps are carried out iteratively until a stopping criterion is reached, as illustrated in Fig. 2. The contribution of each input segmentation can be defined by its specificity and sensitivity by evaluating it to the estimated segmentation. More details will be given in Section II.

The remainder of this paper is organized as follows. Section II describes the mathematical formulation of the proposed iterative fusion model; the description of the data sets used for the evaluation of the proposed method and the experiment output and evaluation are given in Section III, and finally, our discussion and conclusion are provided in Section IV.

II. PROPOSED METHODOLOGY

A. Problem Formulation

Given \mathbf{X}_1 and \mathbf{X}_2 , i.e., two d -dimensional coregistered/calibrated multispectral remote sensing images of m pixels each (we use a vector form for a simple notation) acquired over the same geographical area at two different dates. Let \mathbf{X}_D of dimension d be the multispectral difference image obtained from \mathbf{X}_1 and \mathbf{X}_2 as follows [8]: $\mathbf{X}_D^i = \mathbf{X}_1^i - \mathbf{X}_2^i$, $i = 1, \dots, d$. The final image used for the analysis, which is denoted by \mathbf{X} , is, thus, the modulus image computed from \mathbf{X}_D as

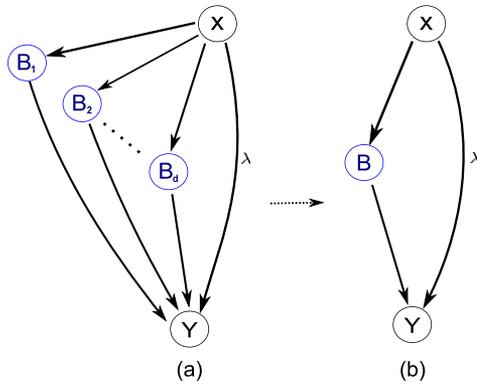


Fig. 3. Graphical illustration of the probabilistic model of “change/no-change” segmentation. (a) Coarse illustration. (b) Compact illustration.

follows: $\mathbf{X} = [(X_D^1)^2 + (X_D^2)^2 + \dots + (X_D^d)^2]^{1/2}$. Let $\mathbf{B} = \{\mathbf{B}_j | \mathbf{B}_j = \{B_{ij}; i = 1 \dots m; j = 1 \dots s\}\}$ be a set of s input segmentations estimated by performing the different thresholding algorithms on \mathbf{X} . Therefore, \mathbf{B} can be considered as a cube of input segmentations of size $m \times s$, in which $B_i = (b_{i1}, \dots, b_{is})$ denotes the i th binary vector. The aim of the proposed method is to estimate, from \mathbf{X} and \mathbf{B} , a new binary image, i.e., $\mathbf{Y} = \{Y_i; i = 1 \dots m\}$, where each pixel Y_i is assigned to one of two information classes, i.e., $\{1, 0\}$ = “change,” “no-change”. The class “change” represents the changes that occurred between two acquisition dates of the two multitemporal images, and “no-change” represents the unchanged areas. Fig. 3 graphically illustrates the proposed model.

Let us define the joint probability linking \mathbf{X} , \mathbf{Y} , and \mathbf{B} as follows:

$$\begin{aligned} p(\mathbf{Y}, \mathbf{X}, \mathbf{B}) &= p(\mathbf{Y}|\mathbf{X}, \mathbf{B})p(\mathbf{B}|\mathbf{X})p(\mathbf{X}) \\ &= \prod_j^s p(\mathbf{Y}|\mathbf{X}, \mathbf{B}_j) \prod_j^s p(\mathbf{B}_j|\mathbf{X})p(\mathbf{X}) \\ &\propto \prod_j^s p(\mathbf{X}|\mathbf{Y})p(\mathbf{Y}|\mathbf{B}_j) \prod_j^s p(\mathbf{B}_j|\mathbf{X})p(\mathbf{X}) \\ &\propto p(\mathbf{X}|\mathbf{Y})p(\mathbf{X}) \prod_j^s p(\mathbf{Y}|\mathbf{B}_j) \prod_j^s p(\mathbf{B}_j|\mathbf{X}). \quad (1) \end{aligned}$$

Without loss of generality, let us associate a real-valued weight (λ) to the conditional probability $p(\mathbf{X}|\mathbf{Y})$ [28]. Through this, we want to encourage, with a certain weight, the fidelity of \mathbf{Y} to the difference image \mathbf{X} , whereas the estimation process of \mathbf{Y} is made from the input segmentations $\{\mathbf{B}_j, \forall j\}$. Hence, (1) can be developed as follows:

$$p(\mathbf{Y}, \mathbf{X}, \mathbf{B}) \propto p(\mathbf{X}|\mathbf{Y})^\lambda p(\mathbf{X}) \prod_j^s p(\mathbf{Y}|\mathbf{B}_j) \prod_j^s p(\mathbf{B}_j|\mathbf{X}). \quad (2)$$

As \mathbf{X} is known and $\{\mathbf{B}_j; \forall j\}$ are already obtained by different thresholding algorithms, the quantity $p(\mathbf{X}) \prod_j^s p(\mathbf{B}_j|\mathbf{X})$ is a constant. Consequently, (2) can be written as

$$\begin{aligned} p(\mathbf{Y}, \mathbf{X}, \mathbf{B}) &\propto p(\mathbf{X}|\mathbf{Y})^\lambda \prod_j^s p(\mathbf{Y}|\mathbf{B}_j) \\ &\propto p(\mathbf{X}|\mathbf{Y})^\lambda p(\mathbf{Y}|\mathbf{B}). \quad (3) \end{aligned}$$

What we are looking for is to find out $\hat{\mathbf{Y}}$, which maximizes the joint probability $p(\mathbf{Y}, \mathbf{X}, \mathbf{B})$, or its logarithm, i.e.,

$$\begin{aligned} \hat{\mathbf{Y}} &= \arg \max_{\mathbf{Y}} \log p(\mathbf{Y}, \mathbf{X}, \mathbf{B}) \\ &= \arg \max_{\mathbf{Y}} \log p(\mathbf{X}|\mathbf{Y})^\lambda p(\mathbf{Y}|\mathbf{B}) \\ &= \arg \max_{\mathbf{Y}} \{\log p(\mathbf{X}|\mathbf{Y})^\lambda + \log p(\mathbf{Y}|\mathbf{B})\} \\ &= \arg \max_{\mathbf{Y}} \{\lambda \log p(\mathbf{X}|\mathbf{Y}) + \log p(\mathbf{Y}|\mathbf{B})\} \quad (4) \end{aligned}$$

where $p(\mathbf{X}|\mathbf{Y})$ stands for the likelihood or conditional probability of a pixel given the parameters of the class, and $p(\mathbf{Y}|\mathbf{B})$ stands for the *posterior* probability estimation of \mathbf{Y} given the thresholding maps.

It is of interest to show that the proposed model consists of estimating the final segmentation “change” and “no-change” from both the set of thresholding maps $\{\mathbf{B}_j; \forall j\}$ and the difference image \mathbf{X} with a certain weight.

B. Estimation of $p(\mathbf{Y}|\mathbf{B})$

A new binary image \mathbf{Y} can be estimated from a set of threshold maps $\mathbf{B} = \{\mathbf{B}_j\}$, and it is derived as [29]

$$p(\mathbf{Y}|\mathbf{B}) \propto p(\mathbf{B}|\mathbf{Y})p(\mathbf{Y}) = \prod_i \left(\prod_j p(B_{ij}|Y_i)p(Y_i) \right) \quad (5)$$

where $p(Y_i)$ is the prior probability of Y_i . For a given pixel i , we have

$$p(Y_i|\mathbf{B}_i) \propto \prod_j p(B_{ij}|Y_i)p(Y_i) \quad (6)$$

which is the conditional probability of the pixel i to be assigned a label Y_i (1 or 0). Since the change detection problem is treated as a binary random variable, thus

$$\begin{aligned} p(Y_i = 1|\mathbf{B}_i) &= \frac{p(Y_i = 1, \mathbf{B}_i)}{p(\mathbf{B}_i)} \\ &= \frac{\alpha}{\alpha + \beta} = \omega_i \quad (7) \\ p(Y_i = 0|\mathbf{B}_i) &= 1 - p(Y_i = 1|\mathbf{B}_i) \\ &= 1 - \frac{\alpha}{\alpha + \beta} \\ &= \frac{\beta}{\alpha + \beta} = 1 - \omega_i \quad (8) \end{aligned}$$

where α and β are derived under the hypothesis that the thresholding maps are generated independently and are defined as follows:

$$\begin{aligned} \alpha &= p(\mathbf{B}_i|Y_i = 1)p(Y_i = 1) \\ &= \prod_j p(B_{ij}|Y_i = 1)p(Y_i = 1) \quad (9) \end{aligned}$$

$$\begin{aligned} \beta &= p(\mathbf{B}_i|Y_i = 0)p(Y_i = 0) \\ &= \prod_j p(B_{ij}|Y_i = 0)p(Y_i = 0). \quad (10) \end{aligned}$$

The weight ω_i stands for the conditional probability that the pixel i belongs to class 1. Similarly to the simultaneous

truth and performance level estimation algorithm [29], each thresholding map \mathbf{B}_j contributes to the final estimation (i.e., likelihood) of “change/no-change” detection based on its performance in terms of *sensitivity* or “true positive (TP) fraction,” which is denoted by p_j , and *specificity* or “true negative (TN) fraction,” which is denoted by q_j . Therefore, their likelihood terms are respectively defined as

$$p(B_{ij}|Y_i=1) = p_j\delta(B_{ij}, 1) + (1-p_j)(1-\delta(B_{ij}, 1)) \quad (11)$$

$$p(B_{ij}|Y_i=0) = q_j(1-\delta(B_{ij}, 0)) + (1-q_j)\delta(B_{ij}, 0) \quad (12)$$

where $\delta(a, b)$ is equal to 1 if $a = b$; otherwise, it is 0. The sensitivity and the specificity are respectively defined by

$$p_j = \frac{\sum_i \delta(B_{ij}, 1) \cdot \delta(Y_i, 1)}{\sum_i \delta(Y_i, 1)} \quad (13)$$

$$q_j = \frac{\sum_i \delta(B_{ij}, 0) \cdot \delta(Y_i, 0)}{\sum_i \delta(Y_i, 0)}. \quad (14)$$

Therefore, ω_i is a normalized product of $p(Y_i = 1)$, which is the prior probability that a pixel belongs to class 1, the sensitivity of each thresholding map that leads to a “change” pixel and the product of the specificity (i.e., 1-*sensitivity*) of each thresholding map that leads to a “no-change” pixel.

C. Estimation of $p(\mathbf{X}|\mathbf{Y})$

$p(\mathbf{X}|\mathbf{Y})$ stands for the likelihood of \mathbf{X} given \mathbf{Y} i.e., the intensity distribution of the classes “change” and “no-change.” The latter are defined by an initial segmentation provided by the majority voting (MV) algorithm applied on the input segmentation set $\{\mathbf{B}_j, \forall j\}$. These two classes can be modeled as random processes, which are, in turn, modeled by fitting a mixture of suitable distribution functions over the histogram of the difference image \mathbf{X} . We have noticed that, in general, each difference image \mathbf{X} follows a specific probability distribution function, as will be detailed in Section III.

D. Regularization of Segmentation

As expected, some difference images have substantial noise. Therefore, small isolated regions can be generated within each segmented area due to a misclassification error caused by the influence of the noise. Morphological operations can be applied to remove those isolated small regions (sieving). As an alternative solution, in this work, we propose to incorporate prior information about the neighboring of the labeled pixels using Markov random fields (MRF) [30]. The latter aims to model the spatial interaction between the labeled pixels to encourage the neighboring pixels to have the same label.

First, let us introduce “theoretically” a segmentation \mathbf{Y}_0 (which, in practice, designates exactly \mathbf{Y} , as illustrated in Fig. 4). The updated segmentation model can be reformulated as follows:

$$\begin{aligned} p(\mathbf{Y}, \mathbf{Y}_0, \mathbf{B}, \mathbf{X}) &= p(\mathbf{Y}|\mathbf{Y}_0, \mathbf{B}, \mathbf{X})p(\mathbf{B}|\mathbf{X})p(\mathbf{X}) \\ &= p(\mathbf{Y}_0|\mathbf{Y})p(\mathbf{Y}|\mathbf{B}, \mathbf{X})p(\mathbf{B}|\mathbf{X})p(\mathbf{X}) \\ &= p(\mathbf{Y}_0|\mathbf{Y})p(\mathbf{X}|\mathbf{Y})^\lambda p(\mathbf{Y}|\mathbf{B})p(\mathbf{B}|\mathbf{X})p(\mathbf{X}). \end{aligned} \quad (15)$$

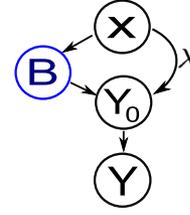


Fig. 4. Graphical illustration of the probabilistic model of “change/no-change” segmentation.

Since \mathbf{X} and \mathbf{B} are known at the beginning of each iteration and do not depend on \mathbf{Y} , $p(\mathbf{X})$ and $p(\mathbf{B}|\mathbf{X})$ can be neglected (because their product results in a constant). Thus, the model (15) can be reduced to

$$p(\mathbf{Y}, \mathbf{Y}_0, \mathbf{B}, \mathbf{X}) \propto p(\mathbf{Y}_0|\mathbf{Y})p(\mathbf{X}|\mathbf{Y})^\lambda p(\mathbf{Y}|\mathbf{B}) \quad (16)$$

where $p(\mathbf{Y}|\mathbf{B})$ and $p(\mathbf{X}|\mathbf{Y})^\lambda$ are defined earlier in Sections II-B and C, respectively. Assume that, in some situations, the posterior probabilities will not deviate dramatically from the prior probabilities [31], particularly when dealing with noisy data. In such a situation, $p(\mathbf{Y}_0|\mathbf{Y})$ can be defined as

$$p(\mathbf{Y}_0|\mathbf{Y}) = (1 + \delta_{\mathbf{Y}_0\mathbf{Y}})p(\mathbf{Y}_0) \quad (17)$$

where $\delta_{\mathbf{Y}_0\mathbf{Y}}$ satisfies $\delta_{\mathbf{Y}_0\mathbf{Y}} \ll 1$. Substituting (17) into (15)

$$p(\mathbf{Y}, \mathbf{Y}_0, \mathbf{B}, \mathbf{X}) = (1 + \delta_{\mathbf{Y}_0\mathbf{Y}})p(\mathbf{Y}_0)p(\mathbf{X}|\mathbf{Y})^\lambda p(\mathbf{Y}|\mathbf{B}). \quad (18)$$

Similarly to what has been previously said, we are searching for $\hat{\mathbf{Y}}$, which maximizes the joint probability $p(\mathbf{Y}, \mathbf{Y}_0, \mathbf{B}, \mathbf{X})$, or its logarithm, i.e.,

$$\begin{aligned} \hat{\mathbf{Y}} &= \arg \max_{\mathbf{Y}} \log p(\mathbf{Y}, \mathbf{Y}_0, \mathbf{B}, \mathbf{X}) \\ &= \arg \max_{\mathbf{Y}} \log (1 + \delta_{\mathbf{Y}_0\mathbf{Y}})p(\mathbf{Y}_0)p(\mathbf{X}|\mathbf{Y})^\lambda p(\mathbf{Y}|\mathbf{B}) \\ &= \arg \max_{\mathbf{Y}} \{ \log (1 + \delta_{\mathbf{Y}_0\mathbf{Y}}) + \log p(\mathbf{Y}_0) \\ &\quad + \log p(\mathbf{X}|\mathbf{Y})^\lambda + \log p(\mathbf{Y}|\mathbf{B}) \} \\ &= \arg \max_{\mathbf{Y}} \{ \log p(\mathbf{Y}_0) + \lambda \log p(\mathbf{X}|\mathbf{Y}) + \log p(\mathbf{Y}|\mathbf{B}) \}. \end{aligned} \quad (19)$$

It is worth noting that the factor $(1 + \delta_{\mathbf{Y}_0\mathbf{Y}})$ is removed from (19) because it can be assumed as a constant. As aforementioned, in practice, \mathbf{Y}_0 means exactly \mathbf{Y} (see Fig. 4). Therefore, (19) can be written as follows:

$$\begin{aligned} \hat{\mathbf{Y}} &= \arg \max_{\mathbf{Y}} \log p(\mathbf{Y}, \mathbf{B}, \mathbf{X}) \\ &= \arg \max_{\mathbf{Y}} \{ \log p(\mathbf{Y}) + \lambda \log p(\mathbf{X}|\mathbf{Y}) + \log p(\mathbf{Y}|\mathbf{B}) \}. \end{aligned} \quad (20)$$

E. Prior Model $P(\mathbf{Y})$

In order to favor homogeneous regions in the segmentation output, contextual spatial information of the segmented image is integrated in our model. In this paper, we resort to an MRF-based isotropic prior $p(\mathbf{Y})$, which consists of modeling the class label \mathbf{Y} and biasing the neighboring pixels to have the

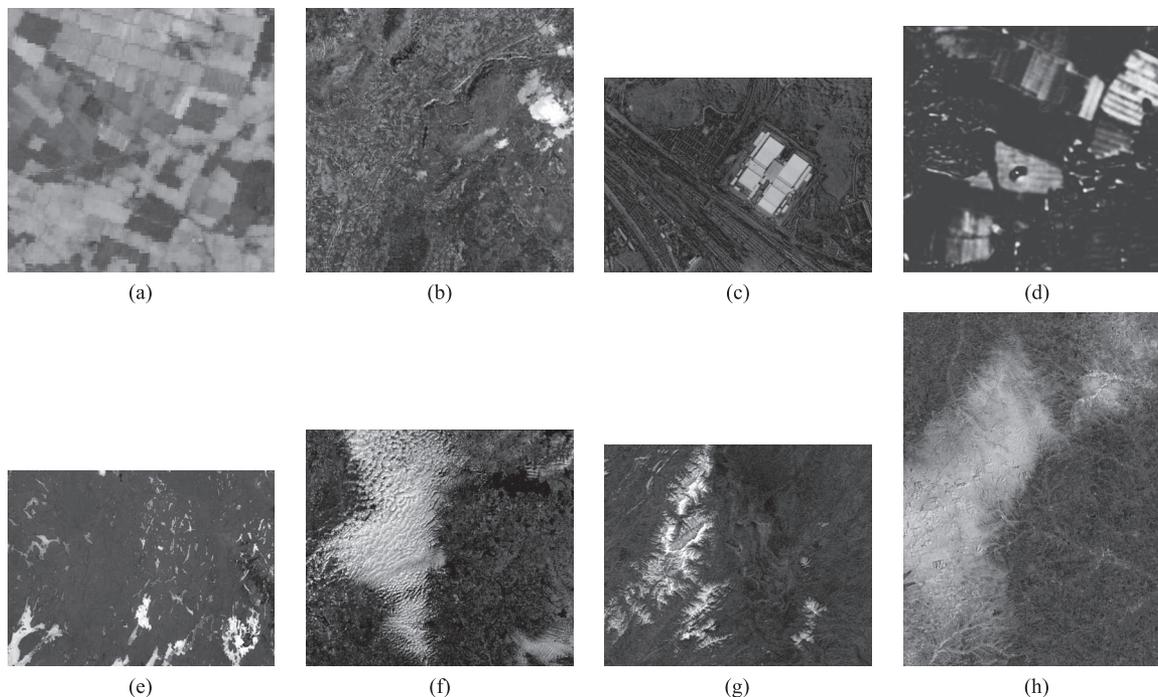


Fig. 5. Eight difference images used for change detection. (a) Po. (b) Trentino. (c) Mina. (d) Riyadh. (e) Knee. (f) Merzières-en-Brenne. (g) Dashte Nawur. (h) El Dorado.

same labels [30]. The Hammersly–Clifford theorem shows that \mathbf{Y} is an MRF (associated with a neighborhood system) if $P(\mathbf{Y})$ follows a Gibbs distribution [30], which is defined by

$$P(\mathbf{Y}) = \frac{1}{Z} \exp[-E(\mathbf{Y})]. \quad (21)$$

The choice of $p(\mathbf{Y})$ is always crucial, mainly to ensure a good reconstruction of the region boundaries. Simplest MRFs use the Ising model [30] with the following form:

$$E(\mathbf{Y}) = \sum_i \sum_{j \in \mathcal{N}_i} \beta (1 - \delta(Y_i, Y_j)) \quad (22)$$

where Y_i and \mathcal{N}_i are the label and the set of neighboring pixels of the pixel i , respectively; $\delta(a, b)$ is equal to 1 if $a = b$; otherwise, it is 0; and β is a constant. It is worth noting that the model expressed by (22) is isotropic, and therefore, it wrongly assumes an isotropic regularity on the region boundaries. In order to favor homogeneous regions while preserving their boundaries, the energy term in (22) is made dependent on the local gradient (intensity change) of the input difference image as follows:

$$E(\mathbf{Y}) = \sum_i \sum_{j \in \mathcal{N}_i} \beta \phi(X_i) (1 - \delta(Y_i, Y_j)) \quad (23)$$

where $\phi(X_i)$ consists of controlling the effect of the Ising regularization term dependently on the degree of gradient magnitude at pixel i . To achieve this, the following formula is used [32]:

$$\phi(X_i) = \frac{1}{1 + \left(\frac{\|\nabla_i\|}{k}\right)^2} \quad (24)$$

where $\|\nabla_i\|$ is the magnitude of the gradient at pixel i , and k is a positive factor controlling the magnitude of this effect.

Finally, the proposed problem in (20) is handled with the deterministic iterative conditional mode algorithm [33], which converges to a local minimum depending on the initialization. In this paper, the initial configuration is provided by the totality voting rule applied on the set of initial segmentation images.

III. DATA SETS AND EXPERIMENTAL RESULTS

A. Data Set Description

Eight satellite images are used for the validation of the proposed iterative change detection model (see Fig. 5). These data sets are acquired by different sensors and present different kinds of change. The description of each data set is given in Table I.

The difference images of Po and Trentino with the corresponding ground-truth images have been provided by Melgani and Bazi [18], [34] (the original bands were not provided; thus, our algorithm is performed directly on the provided difference image). The Mina and Riyadh data sets, which are true-color composites, and their ground-truth images are provided by Hichri *et al.* [4]. They are coregistered after having been downscaled to 8 bits. Their difference images are computed as indicated in Section II-A (i.e., all the channels R, G, and B are considered). The remaining data sets (Knee, Merzières-en-Brenne, Dashte Nawur, and El Dorado) contain eight bands each. The coastal aerosol band (band 1) is not used because it is not present in the other sensors examined here (i.e., Landsat TM5, ETM+, and IKONOS). These data sets were first radiometrically calibrated (to radiance) in ENVI 5.3 and atmospherically corrected using the FLAASH module of ENVI 5.3. Based on the spectral signatures of specific land covers/materials, before calculating the difference image, a

TABLE I
BRIEF DESCRIPTION OF THE DIFFERENT DATA SETS CONSIDERED

Dataset N.	Name	Size (pixels)	Dates of acquisition	Location	Sensor	Change type
1	Po	300 x 300	April 1994 - May 1994	Po river, North of Italy	Landsat TM 5	Agriculture
2	Trentino	300 x 300	May 2000 - July 2000	Northern Italy	Landsat-7 ETM+	Clouds
3	Mina	950 x 700	Jan. 2007 - Dec. 2007	Mina, Saudi Arabia	IKONOS	Construction
4	Riyadh	430 x 401	Feb. 2006 - Oct. 2006	Riyadh, Saudi Arabia	IKONOS	Agriculture
5	Knee	2313 x 1706	April 2015 - May 2015	Manitoba, Canada	Landsat-8 OLI	Ice
6	Mezieres-en-Brenne	2125 x 1897	Oct. 2014 - Dec. 2014	Indre dep. France	Landsat-8 OLI	Clouds
7	Dashte Nawur	2261 x 1890	Oct. 11 & 17, 2015	Ghanzi prov. Afghanistan	Landsat-8 OLI	Snow
8	El-Dorado	1918 x 2570	Dec. 2014 - Jan. 2015	Flint Hills, Kansas, USA	Landsat-8 OLI	Snow

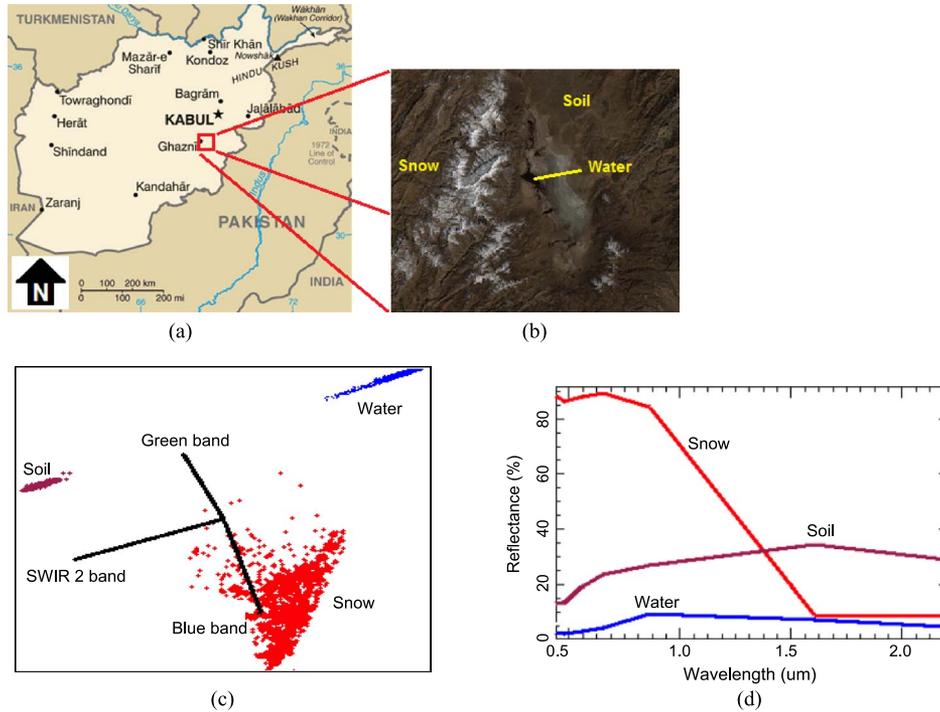


Fig. 6. Dashte Nawur data set. (a) Map showing the location of Dashte Nawur in Ghazni Province in Afghanistan. (b) Its true-color composite. (c) Visualization of the 3-D space of spectral features of three Regions of interests (ROIs): snow, soil, and water. Each point in the feature space represents an ROI pixel, and the axis is in units reflectance. (d) Plot of the ROIs' mean reflectance.

forward feature selection technique (in MATLAB 2014a) is used to select a subset of bands from each data set according to some training samples picked from the classes cloud, snow/ice, soil, and vegetation (an example is shown in Fig. 6). The selected bands for each data set are as follows: Knee {Blue (band 2), Green (band 3)}, El Dorado {Red (band 4), SWIR 1 (band 6)}, Dashte Nawur {Blue, Green, SWIR 2 (band 7)}, and Mezières-en-Brenne {Red, SWIR 1, SWIR 2}. Finally, the difference images are computed as indicated in Section II-A according to the selected bands.

The ground truths are used to measure the performance of the segmentation output in terms of percentage of error rate P_E , percentage of false alarm rate P_F , and percentage of missed alarm rate P_M . P_E is the most important measure compared with P_F and P_M [18]. Let us define the following: N_0 and N_1 are the number of “change” and “no-change” pixels in the ground truth, respectively. TP (true positive) is defined to occur when the pixel is labeled as “change” and the ground truth as well. FP occurs if the pixel is labeled “change” when the ground truth is “no-change.” FN occurs when the pixel is labeled “no-change” but the ground truth is “change.” Therefore, $P_E =$

$$(TP + TN)/(N_1 + N_0), \quad P_F = (FP/N_1) \times 100, \quad \text{and} \quad P_M = [FN/N_0] \times 100.$$

In addition to those measures, we have used two other important measures well known in the domain of remote sensing and information retrieval, which are Fm (F-measure) [35] and Kappa coefficient, i.e., Kap [36]. $Fm = (2 \times R \times P)/(R + P)$, where $R = TP/(TP + FN)$, and $P = TP/(TP + FP)$. The Kappa coefficient [36] is well known in the domain of remotely sensed hyperspectral image classification. Its purpose is to give the reader a quantitative measure of the magnitude of agreement between observers. The calculation is based on the difference between how much agreement is actually present (“observed” agreement P_o) compared with how much agreement would be expected to be present by chance alone (“expected” agreement P_e): $Kappa = (P_o - P_e)/(1 - P_e)$, where $P_o = (TP + TN)/(N_0 + N_1)$, and $P_e = [(n_1/n) \times (m_1/n)] + [(n_0/n) \times (m_0/n)]$, with $n = N_0 + N_1$, $n_1 = TP + FP$, $n_0 = FN + TN$, $m_1 = TP + FN$, and $m_0 = FP + TN$. In contrast to P_E , P_F , and P_M , higher values of F-measure and Kappa mean better performance.

To the best of our knowledge, there is no optimal distribution for each class of the difference image histogram (see Fig. 7) for

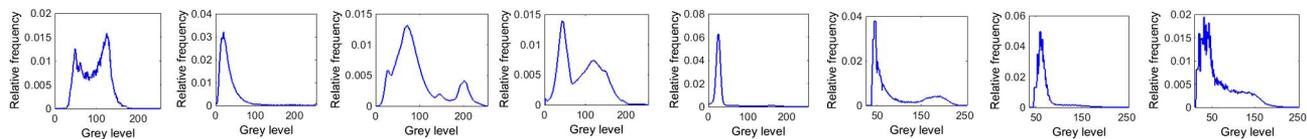


Fig. 7. Histogram distributions of the eight difference images used in the experiment. (Left to right) Po, Trentino, Mina, Knee Lake, Merzières-en-Brenne, Dashte Nawur, and El Dorado Lake.

the optical data. The image distribution modeling and fitting problem has been addressed in the literature. The study in [15] proposed to use the Gamma distribution to model the image classes for change detection in optical remote sensing images; the study in [11] proposed to use a mixture of Rayleigh–Rice for change detection in optical remote sensing; the study in [37] proposed to use a mixture of Gaussian–Weibull for sonar image segmentation, among others. In our case, the generalized extreme value (GEV) model has been used because of its flexibility. It generalizes and combines the Gumbel, Frchet, Weibull, Rayleigh, and exponential distribution families. In order to improve the segmentation accuracy, it is better to fit the histogram of each difference image to the GEV distribution, which is a suitable statistical distribution.

Therefore, the joint probability $P(\mathbf{X}|\mathbf{Y})$, which stands for the likelihood term, is computed deceptively to the input difference image. Experimentally, the two classes “change” and “no-change” can be modeled as a random process, which is, in turn, modeled by fitting a mixture of GEV distribution function over the histogram of the difference image \mathbf{X} . In addition, we would like to recall that our fusion model is in the category of the committee machine learning model combining a set of segmentation results with a dynamic structure type. It means that the initial input segmentations are combined by means of a mechanism that involves the input data (or original image), in our case, with a likelihood fitting model using GEV model. Thus, the GEV model is a part of our fusion model. The interest of using a fusion model, with a dynamic structure type, and a likelihood fitting model using GEV model is (empirically) demonstrated, in terms of classification result (by comparison with other methods).

B. Initial Input Segmentations

Six thresholding algorithms are used to produce six initial input segmentations. These algorithms are as follows: 1) Abutaleb [38], where the optimal threshold is selected by maximizing the sum of the posterior entropies of two classes; 2) “Intermodes” [39], where the histogram is iteratively smoothed until only two peaks remain (the threshold being the midpoint of the two peaks); 3) Kapur [40], where a threshold is chosen such that the entropies of distribution are maximized; 4) Kittler [41], which assumes a Gaussian mixture model in order to determine the optimal threshold; 5) Shanbhag [42], which is also based on an entropy measure; and 6) Yen [43], where the optimal threshold is chosen so that a correlation criterion is maximized. We have found that, in general, for each image, there is one or two input segmentations that are very different from others. These input segmentations are considered as outliers and must be discarded from the ensemble because

TABLE II
RELATIONSHIP BETWEEN λ AND THE MEAN SIMILARITY OF THE INPUT SEGMENTATION SET

Dataset	Mean Similarity	λ
Po	93.0	1
Mina	88.0	3
Kenn Lake	78.0	3
Trentino	73.7	5
Mezières-en-Bernne	71.0	5
Dashte Nawur	55.0	9
Riyadh	50.0	9
El-Dorado lake	41.0	11

they may impede the achievement of good results. The sign “/” in Table III means that the corresponding input segmentations are not considered in the segmentation process. In this paper, we have proposed a simple automatic technique to detect only one outlier image (but nothing prevents discarding more than one). This technique consists in comparing, in terms of Kappa (Fm can be used as well), all the input segmentations to that produced by the MV algorithm, and then discarding the one with the lowest Kappa. As shown in Table III, the Kittler and Illingworth thresholding algorithm, for example, is the one that produces the least similar binary input segmentation of Po; therefore, it will not be taken into account in the segmentation process.

C. Parameter Sensitivity

The proposed segmentation model is guided by two parameters: 1) λ , which is related to the likelihood term [cf., (20)] and thus controls the contribution of the difference image \mathbf{X} in the estimation of \mathbf{Y} , and 2) β , which is linked to the prior term [cf., (23)] and therefore controls the smoothing of the segmentation output. In order to show the sensitivity of the proposed segmentation model to these two parameters, a series of experiments are carried out by changing their values. The optimal value of β is equal to 1. In this paper, we consider the parameter λ as the most important as it concerns the contribution of the difference image in the segmentation process. We made from λ a factor that can be learned dependently from the input segmentation set. In this paper, we have attempted to study the relationship between λ and the cross correlation (or the total mean similarity) between the input segmentations provided by the different thresholding algorithms. In this paper, we have used the Kappa measure (F-measure can be also used) to compute the mean similarity between all the possible couples of input segmentations set. We have found that the optimal parameters of λ strongly correlate negatively (-0.98) with the mean similarity measure. In other terms, λ is inversely proportional to the mean similarity of the input segmentations, and this means that the difference image \mathbf{X} is considered strongly in the segmentation process if the similarity of input segmentations is low, and vice versa. Table II demonstrates

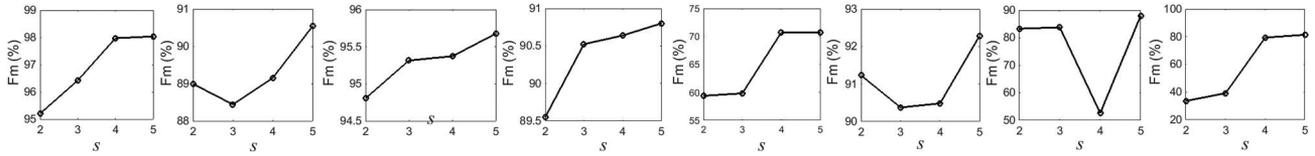


Fig. 8. Evolution of the F-measure (Fm) as a function of the number of segmentations (s) to be fused. (Left to right) Po, Trentino, Mina, Riyadh, Knee Lake, Merzières-en-Brenne, Dashte Nawur, and El Dorado Lake.

TABLE III
PERFORMANCE (IN %) COMPARISON IN TERMS OF FM, KAPPA (KAP), P_E , P_F , AND P_M

Methods	Dataset 1 Po					Dataset 2 Trentino					Dataset 3 Mina					Dataset 4 Riyadh				
	F_m	Kap	P_E	P_F	P_M	F_m	Kap	P_E	P_F	P_M	F_m	Kap	P_E	P_F	P_M	F_m	Kap	P_E	P_F	P_M
Abutaleb [38]	92.6	86.9	6.4	10.1	1.0	69.5	68.0	3.1	3.2	1.3	90.4	89.8	1.1	0.9	4.4	59.5	54.8	9.7	0.6	56.2
Intermodes [39]	97.7	96.1	1.8	1.5	2.3	77.1	76.4	1.4	0.3	31.0	92.0	91.5	0.9	0.6	6.9	87.6	85.3	3.9	1.6	15.7
Kapur [40]	95.0	91.3	5.1	8.0	0.8	78.6	77.8	1.6	0.8	21.6	94.7	94.4	0.6	0.4	3.8	61.0	56.4	9.3	0.5	55.0
Kitler & Illingworth[41]	/	/	/	/	/	63.8	62.0	3.7	3.7	4.9	85.9	84.9	1.8	1.8	2.0	44.3	25.6	40.7	48.5	0.1
Shanbhag [42]	96.3	93.6	3.1	4.2	1.3	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Yen [43]	94.3	90.2	4.8	7.0	0.8	78.6	77.8	1.5	0.8	21.6	86.6	85.7	1.7	1.7	1.5	89.6	87.6	3.4	2.1	101.0
MV [44]	95.4	92.0	3.8	5.8	0.9	78.6	77.8	1.6	0.8	21.6	92.8	92.3	0.8	0.7	2.2	88.2	86.1	3.6	1.3	15.7
Bazi2005 [45]	-	-	3.4	0.9	7.2	-	-	-	4.2	4.2	4.3	-	-	-	-	-	-	-	-	-
Bruzzone and Prieto [2]	-	-	5.7	0.4	13.4	-	-	-	3.7	3.8	0.9	-	-	-	-	-	-	-	-	-
Chan-Vese [46]	73.0	44.2	29.9	50.3	0.2	77.1	76.4	1.4	0.0	35.6	88.4	87.6	1.4	1.3	3.1	80.7	76.3	7.3	7.4	6.5
MLSK [19]	-	-	2.5	3.9	0.9	-	-	-	-	-	-	-	0.5	4.0	2.3	-	-	4.2	2.1	12.7
SVM-MLS [4]	-	-	2.0	3.0	1.2	-	-	-	-	-	-	-	0.4	0.2	4.7	-	-	3.7	1.9	10.9
ML-MRF [15]	-	-	-	-	-	-	-	-	-	-	73.0	71.1	3.6	3.1	12.9	1.1	0.0	20.8	5.6	99.3
MRF_FUSION [18]	-	-	2.8	2.4	3.4	-	-	0.7	0.5	5.4	-	-	-	-	-	-	-	-	-	-
FCR [47]	95.3	90.8	3.9	5.1	2.1	86.4	85.9	1.0	0.9	5.7	78.2	76.6	3.1	3.2	1.2	87.3	84.9	4.0	2.2	13.4
FMBFM [48]	88.2	78.6	10.7	17.5	0.8	74.9	74.0	1.9	1.3	18.5	81.2	79.6	2.6	2.6	1.3	86.4	83.8	4.3	2.4	14.4
PRIF [49]	95.5	92.2	3.7	5.6	1.0	78.5	77.8	1.5	0.8	21.6	89.8	89.3	1.2	1.2	2.0	67.6	63.3	8.2	0.7	47.4
VOIBFM [21]	90.4	82.8	8.5	14.1	0.4	78.9	78.1	1.5	0.8	19.6	94.5	94.5	0.6	0.3	4.9	61.5	56.8	9.3	0.6	54.1
Our	98.1	96.7	1.5	1.1	2.1	90.6	88.8	0.6	0.2	11.7	95.7	96.0	0.4	0.2	5.0	90.8	89.1	2.9	1.7	9.3
Methods	Dataset 5 Knee Lake					Dataset 6 Merizeme-en-Bennne					Dataset 7 Dashte Nawur					Dataset 8 El-Dorado lake				
	F_m	Kap	P_E	P_F	P_M	F_m	Kap	P_E	P_F	P_M	F_m	Kap	P_E	P_F	P_M	F_m	Kap	P_E	P_F	P_M
Abutaleb [38]	49.0	46.7	5.8	5.9	2.4	/	/	/	/	/	82.4	84.2	3.5	0.5	27.1	19.0	16.7	13.2	0.0	89.5
Intermodes [39]	69.4	68.3	2.3	2.1	9.1	88.1	84.2	5.9	1.2	18.6	27.8	63.4	9.5	0.0	83.8	16.5	14.4	13.4	0.0	91.0
Kapur [40]	57.7	55.9	4.0	4.0	4.9	70.5	63.4	12.4	0.1	45.2	/	/	/	/	/	15.6	13.6	13.5	0.0	91.3
Kitler & Illingworth[41]	41.0	38.2	8.1	8.3	2.1	84.2	77.1	9.9	12.8	2.3	87.2	77.1	2.8	1.1	15.6	/	/	/	/	/
Shanbhag [42]	/	/	/	/	/	74.0	67.4	11.3	0.1	40.9	48.7	67.4	7.7	0.0	67.7	4.0	3.4	14.4	0.0	97.9
Yen [43]	56.4	54.5	4.2	4.2	4.7	90.6	87.2	4.9	2.2	12.2	85.7	87.2	3.5	3.2	5.8	82.7	79.2	5.9	6.5	2.9
MV [44]	56.3	54.5	4.2	4.2	4.6	88.1	84.2	5.9	1.2	18.6	82.8	81.0	3.4	0.3	27.5	17.0	14.8	13.3	0.0	90.7
Bazi2005 [45]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bruzzone and Prieto [2]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chan-Vese [46]	51.7	49.7	4.6	4.3	14.6	88.4	83.5	6.8	7.8	4.3	77.9	75.0	5.1	3.2	19.7	48.1	33.2	31.6	37.1	0.3
MLSK [19]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SVM-MLS [4]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ML-MRF [15]	46.1	43.5	6.8	7.0	0.1	91.8	88.4	4.8	5.7	2.3	73.3	70.5	5.2	1.2	36.8	70.1	66.1	7.4	1.5	41.4
MRF_FUSION [18]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FCR [47]	52.7	50.6	5.0	5.1	2.4	89.8	50.0	5.2	1.1	15.9	86.1	82.9	3.0	1.2	17.0	60.3	50.0	19.2	22.5	0.5
FMBFM [48]	55.8	53.9	4.5	4.6	0.6	90.8	86.3	4.9	2.1	12.0	87.0	85.4	2.8	1.3	15.0	16.9	14.8	13.3	0.0	90.7
PRIF [49]	45.4	55.6	4.1	4.1	4.1	84.6	87.5	7.3	0.2	26.1	73.2	70.6	5.0	0.6	39.3	11.0	9.6	13.8	0.0	94.1
VOIBFM [21]	57.3	54.5	4.2	4.3	0.7	76.8	80.0	10.4	0.3	37.0	24.9	22.7	9.7	0.0	85.7	15.0	13.0	13.5	0.0	91.9
Our	70.7	69.6	2.1	1.9	9.3	92.3	89.6	4.0	1.0	11.9	88.0	86.6	2.5	0.8	15.9	82.8	79.7	5.2	3.6	14.5

the relationship between λ and the mean similarity within the input segmentation set of each data set. Further analysis on this issue will be conducted in future works. We can also notice that, in general, a higher number of initial segmentations (s) leads to better performance. This is shown in Fig. 8 (here, only the F-measure is shown due to space limitations). This experiment demonstrates the validity of our fusion procedure and the performance measures obtained by the proposed change detection model.

D. Results and Evaluation

The proposed method is compared with the following: 1) the ensemble of thresholding algorithms used to produce the initial input segmentations; 2) some similar fusion-based methods: the well-known majority voting rule (MV) [44]; Fusion-MRF [18], which is based on fusing multiple thresholding outputs; Fusion of clustering Results (FCR) [47], which is

based on the within-cluster fusion; Probabilistic Rand index based Fusion (PRIF) [49], where the fusion of segmentations is achieved in the probabilistic random index sense; variation of information based fusion model (VOIBFM) [21], where the fusion of segmentation is performed according to the variation of information criterion; and F-measure based fusion model (FMBFM) [48], where the segmentation is performed based on F-measure; and 3) some non-fusion-based methods as well: Bazi *et al.* [45], Bruzzone and Prieto [2], Level-Set Chan-Vese [46], Multiresolution level set (MLS) [19], SVM-MLS [4], and Multilevel Markov random fields (ML-MRF) [15].

Table III presents, in percentage, the performance of the different methods. The sign “/” means that the corresponding input segmentation is not considered in the combination process, and the sign “-” means that the image has not been tested by the corresponding algorithm for two reasons: either the original bands of that image are not available (we have only the difference image), and therefore the algorithm is not

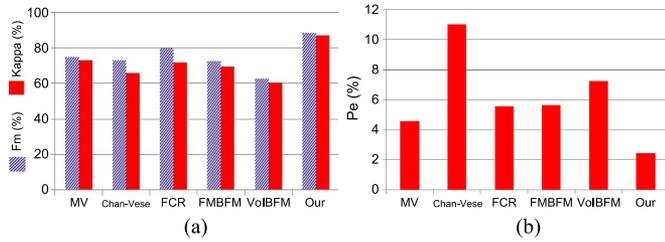


Fig. 9. Average performance over eight images processed by the different methods. (a) (hashed) Mean Fm and (solid) mean Kappa. (b) Mean P_E . (Other methods are not considered because of the two reasons aforementioned in the beginning of this subsection).

applicable, or the code of the algorithm is not available (for example, Po and Trentino are not processed by the algorithm ML-MRF [15] because ML-MRF is designed to perform on the original spectral bands, which are not available). From this table, two points can be emphasized. First, the proposed method outperforms, in general, all the ensemble of thresholding algorithms (used to generate the initial segmentations to be fused), particularly in terms of Fm, Kap, and P_E , and this confirms the validity of our hypothesis postulating that the segmentation estimated by the proposed method should be more accurate than all the initial segmentations individually. Second, the proposed method outperforms all the other fusion-based and non-fusion-based (level-set-based and MRF-based) methods, particularly in terms of Fm, Kap, and P_E (the most important metrics). Since the proposed method is better in terms of P_E and P_F than in terms of P_M , this means that it is able to avoid the detection of wrong change (i.e., false positive), but it fails to detect some real change (i.e., true positive). A possible reason is that some isolated changed pixels are affected by the MRF smoothing effect. On some data sets, some simple thresholding algorithms can perform better than recent methods. For instance, the Inter-modes algorithm outperforms other literature methods on the Po image in terms of P_E (i.e., $P_E = 1.8$; this is not strange since there are some indications in the literature that very simple rules may achieve surprisingly high accuracy on many data sets [50]). In terms of generalization power, the proposed methods outperforms the literature methods (see Fig. 9).

It is also important to report the performance of the segmentation model over the iterations. From Fig. 10, we can remark that, over the iterations, Fm (for instance) increases as expected. This means that the segmentation quality is improving from one iteration to the other. The question that may arise is how many iterations are needed to reach the maximum performance of segmentation. Theoretically, the accuracy of the estimated segmentation is always increasing as long as the model is fed by good segmentations at each iteration. In this work, we have set, empirically, the number of iterations to be equal to half the total number of initial thresholding images plus one, i.e., $6/2 + 1 = 4$. We have found that, beyond that number of iterations, the performance does not change significantly, as shown in Fig. 10.

From the aforementioned results, we can conclude that the hypothesis on which the proposed model is based is experimentally validated, i.e., almost, the estimated segmentation is better than all initial segmentations, and a good initial set of segmentations leads to better final segmentation. This can

be achieved due to the contribution of several factors: 1) the outliers detection and rejection from the initial segmentation set; 2) the fitness of the intensity distribution of difference images to their appropriate statistical functions; 3) the incorporation of the MRF-based regularization; 4) the contribution of the difference images in the segmentation is based on their consistency and self-similarity; and 5) the benefit provided by the iterative scheme of the segmentation model.

E. Influence of the Quality of the Input Segmentation Set

We have attempted to analyze the accuracy of change detection output, i.e., \mathbf{Y} (estimated segmentation of the difference image), based on the quality of the set of input segmentations, i.e., $\{\mathbf{B}_{j:1\dots 6}\}$. We have investigated experimentally how the performance, in terms of Fm, for example, of \mathbf{Y} , can be influenced by the performance of $\{\mathbf{B}_{j:1\dots 6}\}$ all together over the eight data sets considered. First, we studied the influence of the mean performance of $\{\mathbf{B}_{j:1\dots 6}\}$, and second, we studied the influence of its standard deviation (variability or diversity within the set). Let Fm_k , $k = 1 \dots 8$, be the performance measure of \mathbf{Y} generated from the k th data set and MFm_k and SFm_k be the mean and standard deviation of the performance of the set $\{\mathbf{B}_{j:1\dots 6}\}$ generated from the k th data set, respectively. Fig. 11(a) illustrates the change of Fm_k according to MFm_k , whereas Fig. 11(b) illustrates the change of Fm_k according to SFm_k . The dashed curves are the regression lines (trend lines) used for the overall interpretation. In general, from Fig. 11(a), we can notice that Fm_k is highly proportional to MFm_k , $\forall k$, with a correlation coefficient equal to 0.77, and Fm_k is always better than MFm_k , $\forall k$. Furthermore, it is clear that the proposed method can achieve good results (i.e., $Fm_8 = 82.8\%$) from less accurate input segmentations (i.e., $MFm_8 = 27.5\%$). Nevertheless, the gap between Fm_k and MFm_k diminishes when the mean performance approaches saturation ($MFm_k \rightarrow 100\%$), which is indeed normal. From Fig. 11(b), Fm_k seems to be inversely proportional to SFm_k as expected. This means that the smaller the variability (higher similarity) within $\{\mathbf{B}_{j:1\dots 6}\}$ is, the higher the performance of \mathbf{Y} is. It is worth noting that a high number of less accurate segmentations in the set $\{\mathbf{B}_{j:1\dots 6}\}$ may lead to estimating a better \mathbf{Y} , but not necessarily that it is much better than any other input individually. The El Dorado data set is a typical example (see Table III), in which five over six input segmentations have very low performances. Despite this, the method is able to estimate a good result, but is still very close to the Yen method [43] ($Fm_8 = 82.8\%$ versus 82.7%). Based on this simple investigation, it can be concluded that the estimation of more accurate segmentation needs to have at least a subset of input segmentations with reasonable accuracy. The issue of defining/selecting such a subset will be investigated more thoroughly in future work.

In this paper, a new approach of change detection from bitemporal remotely sensed images is proposed. The objective is to segment into two classes (“change/no-change”) the difference image that is generated from two bitemporal images of the same area acquired at two different dates. In contrast to the computationally costly and complex segmentation models proposed in the recent state of the art [3], [51], the idea of fusion

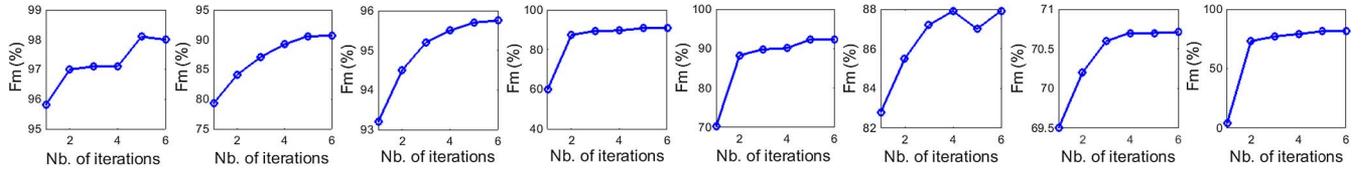


Fig. 10. Improvement of the performance in terms of Fm of the estimated segmentations over the iterations. (Left to right) Po, Trentino, Mina, Riyadh, El Dorado Lake, Dashte Nawur, Knee Lake, and Merzières-en-Brenne.

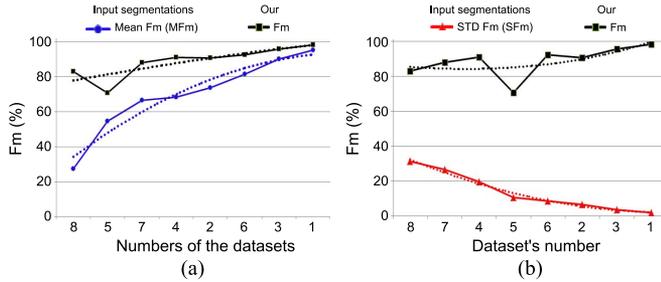


Fig. 11. Influence of the mean performance (MFm) and variability (SFm) of the input segmentation set on the performance (Fm) of the estimated segmentation (i.e., change detection output). (a) MFm versus Fm. The performance values are sorted by MFm from smallest to largest (the order of the data sets numbers is changed accordingly). (b) SFm versus Fm. The performance values are sorted by SFm from largest to smallest. The dashed curves are the regression lines used for the overall interpretation.

low-cost thresholding (or classifiers) outputs, to estimate a new more accurate one, has been shown to be very competitive [18]. It was motivated in machine learning [23] that, even when the input classifiers are indistinguishable with respect to their (resubstitution) training error, they may have different generalization performances. Instead of picking only one classifier, a safer option would be to use them all and average their outputs. The new classifier might not be better than the single best classifier but would diminish or eliminate the risk of picking an inadequate single classifier. The proposed fusion model is in the category of the committee machine learning model combining a set of segmentation, resulting from different parametric and/or nonparametric models, with a dynamic structure type. It means that the initial input segmentations are combined by means of a mechanism that involves the input data (or original image), in our case, with a likelihood fitting model using GEV model. Thus, the GEV model is a part of our fusion model. The interest of using a fusion model, with a dynamic structure type, and a likelihood fitting model using GEV model is (empirically) demonstrated, in terms of classification result by comparison with other methods.

IV. DISCUSSION AND CONCLUSION

The proposed method is compared, in the first part of Table III, to the ensemble of thresholding algorithms used to produce the initial input segmentations (and used in our fusion procedure) and, in the second part of Table III, to some other methods (based on a fusion procedure [18], [21], [44], [47]–[49] or not [2], [4], [19], [45], [46], [15]) with which we could compare our approach. The classification result obtained by each individual thresholding algorithm is important in order to quantify the gain obtained by our fusion procedure and the comparison with the other methods, in order to evaluate the

accuracy of our method against state-of-the-art ones. The thresholding algorithms, which are listed in Table III, are used as such. In fact, they cannot be rederived for the GEV distribution since these methods do not (generally) use a distribution mixture model to determine their threshold from the histogram of the image, except the Kittler method, which assumes a Gaussian mixture model in order to determine the optimal threshold, but this method cannot be easily derived in the case of a mixture of GEV distributions. For the methods in the second part of Table III, the authors use sometimes their own statistical distribution model, for the difference image [2], [45] (in fact, the distribution model that they believe to be the most appropriate for this problem), and sometimes, they do not use a distribution model [18], [46], [19]. To empower the fusion accuracy, we have proposed, in this paper, a new iterative strategy of fusing the results of different thresholding algorithms while taking into account the statistical characteristic of the input difference image. The proposed method is based on the following hypothesis: the estimated output segmentation is assumed to be more accurate than all other input segmentations individually, and more accurate input segmentations lead almost to better output segmentation. This gives rise to the idea of an energy-based fusion model in which the outlier segmentation is detected and rejected to further estimate a more accurate segmentation.

Mathematically speaking, the proposed model aims to maximize the joint probability between the difference image, the ensemble of input segmentations, and the output segmentation under a Bayesian–Markovian stochastic framework. In this work, the difference image contributes in the segmentation process by a certain weight defined automatically from the degree of similarity within the input segmentation set. It has been found empirically that, when the similarity of the input segmentations is low, the contribution of the difference image is really important to compensate the disagreement between the input segmentations. In addition, it was found experimentally that if the majority of input segmentations are largely different from each other and have “very low” accuracy (the case of all the El Dorado data set outputs), the proposed fusion model is, somehow, able to get a more accurate output, but not necessarily the one that it is better than any other input individually. Therefore, it is advisable to select a subset of input segmentations with reasonable accuracy to achieve better performances. The MRF is also incorporated in the segmentation model in order to regularize and smooth the segmentation output by encouraging the neighboring pixels to have the same label. This has a positive impact on the final segmentation, particularly when the noise level is high. Objective evaluation of the proposed approach is conducted on the basis of ten data sets. A numerical comparison with other methods is also realized. The experiment

results show that the proposed method has a power of generalization compared with the state of the art.

We believe that the proposed pixel-based method may still be used in the future for time-series images from same or different sensors that meet its basic assumption [3]. Furthermore, the proposed method will be further investigated to deal with perturbing factors, which may affect the image pair. For instance, in some cases, the cloud cannot be preprocessed (i.e., removed) and appear as unwanted change, which may harm the real change. Nevertheless, based on the difference between the motion frequencies of the unwanted change (e.g., due to cloud) and the real change, which may be both computed using multiple short-temporal images, it may be feasible to differentiate between these two types of changes.

The advantages of the proposed method are that it is simple, automatic, easy to implement, and has a positive impact on the fusion accuracy. In addition, the Bayesian framework has the advantage of being open to including other terms, such as MRF, which is responsible for homogenizing the outputs by removing the very small regions. However, the MRF must be taken with caution, because it can lead to undesirable results if the small regions represent areas of interest. Thus, the user must have sufficient prior knowledge about these regions and weaken the weight of the prior probability by diminishing the value of his factor. Unfortunately, similar to other methods, it has a disadvantage that resides in setting the parameter λ , which we want to be learned automatically. This point will be investigated in future work. We would like to mention that the proposed method was tested on subsets trimmed from their original remote sensing images for demonstration purposes, but it is applicable across the entire images.

ACKNOWLEDGMENT

The authors would like to thank Prof. F. Melgani for providing the Po and Trentino data sets and Prof. H. S. Hichri for providing the Mina and Riyadh data sets. They would also like to thank Prof. C. Benedek for the time spent in performing the ML-MRF method to detect the changes from the Mina, Riyadh, Knee, Merzières-en-Brenne, Dashte Nawur, and El Dorado data sets.

REFERENCES

- [1] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989.
- [2] L. Bruzzone and D. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.
- [3] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogramm. Remote Sens.*, vol. 80, pp. 91–106, 2013.
- [4] H. Hichri, Y. Bazi, N. Alajlan, and S. Malek, "Interactive segmentation for change detection in multispectral remote-sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 298–302, Mar. 2013.
- [5] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 858–867, Jul. 1997.
- [6] B. Demir, F. Bovolo, and L. Bruzzone, "Classification of time series of multispectral images with limited training data," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3219–3233, Aug. 2013.
- [7] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 218–236, Jan. 2007.
- [8] F. Bovolo, S. Marchesi, and L. Bruzzone, "A framework for automatic and unsupervised detection of multiple changes in multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2196–2212, Jun. 2012.
- [9] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Unsupervised multitemporal spectral unmixing for detecting multiple changes in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2733–2748, May 2016. [Online]. Available: <http://dx.doi.org/10.1109/TGRS.2015.2505183>
- [10] V. Ortiz-Rivera, M. Vélez-Reyes, and B. Roysam, "Change detection in hyperspectral imagery using temporal principal components," in *Proc. Def. Security Symp.*, 2006, pp. 1–12.
- [11] M. Zanetti, F. Bovolo, and L. Bruzzone, "Rayleigh–Rice mixture parameter estimation via EM algorithm for change detection in multispectral images," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5004–5016, Dec. 2015.
- [12] M. Gong, Z. Zhou, and J. Ma, "Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2141–2151, Apr. 2012.
- [13] V. Tramutoli, "Robust Satellite Techniques (RST) for natural and environmental hazards monitoring and mitigation: Theory and applications," in *Proc. IEEE Int. Workshop Anal. MultiTemp.*, 2007, pp. 1–6.
- [14] P. He, W. Shi, Z. Miao, H. Zhang, and L. Cai, "Advanced Markov random field model based on local uncertainty for unsupervised change detection," *Remote Sens. Lett.*, vol. 6, no. 9, pp. 667–676, 2015. [Online]. Available: <http://dx.doi.org/10.1080/2150704X.2015.1054045>
- [15] C. Benedek, M. Shadaydeh, Z. Kato, T. Szirányi, and J. Zerubia, "Multi-layer Markov random field models for change detection in optical remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, p. 18, 2015. [Online]. Available: <https://hal.inria.fr/hal-01116609>
- [16] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Hierarchical unsupervised change detection in multitemporal hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 244–260, Jan. 2015.
- [17] S. L. Hégarat-Masclé and R. Seltz, "Automatic change detection by evidential fusion of change indices," *Remote Sens. Environ.*, vol. 91, no. 3, pp. 390–404, 2004. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0034425704001099>
- [18] F. Melgani and Y. Bazi, "Markovian fusion approach to robust unsupervised change detection in remotely sensed imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 457–461, Oct. 2006.
- [19] Y. Bazi, F. Melgani, and H. D. Al-Sharari, "Unsupervised change detection in multispectral remotely sensed imagery with level set methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 8, pp. 3178–3187, Aug. 2010. [Online]. Available: <http://dx.doi.org/10.1109/TGRS.2010.2045506>
- [20] V. Tramutoli, "Robust AVHRR Techniques (RAT) for environmental monitoring: Theory and applications," in *Remote Sensing*. Bellingham, WA, USA: Int. Soc. Opt. Photon., 1998, pp. 101–113.
- [21] M. Mignotte, "A label field fusion model with a variation of information estimator for image segmentation," *Inf. Fus.*, vol. 20, pp. 7–20, 2014.
- [22] T. Dietterich, "Ensemble methods in machine learning," in *Proc. 1st Int. Workshop Multiple Classifier Syst.*, 2000, vol. 1857, pp. 1–15.
- [23] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, NJ, USA: Wiley, 2004.
- [24] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 993–1001, Oct. 1990.
- [25] G. Tsoumakas, L. Angelis, and I. Vlahavas, "Selective fusion of heterogeneous classifiers," *Intell. Data Anal.*, vol. 9, no. 6, pp. 511–525, 2005.
- [26] P. Yang, Y. Hwa Yang, B. B. Zhou, and A. Y. Zomaya, "A review of ensemble methods in bioinformatics," *Current Bioinform.*, vol. 5, no. 4, pp. 296–308, 2010.
- [27] S. Ye, D. Chen, and J. Yu, "A targeted change-detection procedure by combining change vector analysis and post-classification approach," *J. Photogramm. Remote Sens.*, vol. 114, pp. 115–124, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0924271616000356>
- [28] Y. Zhou and T. S. Huang, "Weighted Bayesian network for visual tracking," in *Proc. IEEE 18th ICPR*, 2006, vol. 1, pp. 523–526.
- [29] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous Truth and Performance Level Estimation (STAPLE): An algorithm for the validation of image segmentation," *IEEE Trans. Med. Imag.*, vol. 23, no. 7, pp. 903–921, Jul. 2004.
- [30] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

- [31] J. Kittler, M. Hatef, R. P. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.
- [32] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
- [33] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Statist. Soc. B Methodol.*, vol. 48, no. 3, pp. 259–302, 1986.
- [34] F. Melgani and Y. Bazi, "Robust unsupervised change detection with Markov random fields," in *Proc. IEEE IGARSS*, Jul. 2006, pp. 208–211.
- [35] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, 2009.
- [36] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: The kappa statistic" *Fam Med.*, vol. 37, no. 5, pp. 360–363, 2005.
- [37] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Sonar image segmentation using an unsupervised hierarchical MRF model," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1216–1231, 2000. [Online]. Available: <http://dx.doi.org/10.1109/83.847834>
- [38] A. S. Abutaleb, "Automatic thresholding of gray-level pictures using two-dimensional entropy," *Comput. Vis., Graph., Image Process.*, vol. 47, no. 1, pp. 22–32, 1989.
- [39] J. M. Prewitt and M. L. Mendelsohn, "The analysis of cell images," *Ann. NY Acad. Sci.*, vol. 128, no. 3, pp. 1035–1053, Jan. 1966.
- [40] J. N. Kapur, P. K. Sahoo, and A. K. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Comput. Vis., Graph., Image Process.*, vol. 29, no. 3, pp. 273–285, 1985.
- [41] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognit.*, vol. 19, no. 1, pp. 41–47, 1986.
- [42] A. G. Shanbhag, "Utilization of information measure as a means of image thresholding," *CVGIP: Graph. Models Image Process.*, vol. 56, no. 5, pp. 414–419, 1994.
- [43] J.-C. Yen, F.-J. Chang, and S. Chang, "A new criterion for automatic multilevel thresholding," *IEEE Trans. Image Process.*, vol. 4, no. 3, pp. 370–378, Mar. 1995.
- [44] L. Lam and C. Y. Suen, "Application of majority voting to pattern recognition: An analysis of its behavior and performance," *IEEE Trans. Syst., Man Cybern., A, Syst. Humans*, vol. 27, no. 5, pp. 553–568, Sep. 1997.
- [45] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874–887, Apr. 2005. [Online]. Available: <http://dx.doi.org/10.1109/TGRS.2004.842441>
- [46] T. F. Chan *et al.*, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
- [47] M. Mignotte, "Segmentation by fusion of histogram-based K-means clusters in different color spaces," *IEEE Trans. Image Process.*, vol. 17, no. 5, pp. 780–787, May 2008.
- [48] C. Helou and M. Mignotte, "A precision-recall criterion based consensus model for fusing multiple segmentations," *Int. J. Signal Image Process.*, vol. 7, pp. 61–82, 2014.
- [49] M. Mignotte, "A label field fusion Bayesian model and its penalized maximum Rand estimator for image segmentation," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1610–1624, Jun. 2010.
- [50] R. C. Holte, "Very simple classification rules perform well on most commonly used datasets," *Mach. Learn.*, vol. 11, no. 1, pp. 63–90, 1993.
- [51] K. Karantzas, "Recent advances on 2D and 3D change detection in urban environments from remote sensing data," in *Computational Approaches for Urban Environments*, vol. 13, ser. Geotechnologies and the Environment, M. Helbich, J. Jokar Arsanjani, and M. Leitner, Eds. New York, NY, USA: Springer-Verlag, 2015, pp. 237–272. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-11469-9_10



Rachid Hedjam received the M.Sc. degree in computer science and operation research from the University of Montreal, Montreal, QC, Canada, in 2009 and the Ph.D. degree in computer engineering from the University of Quebec, Quebec, QC, Canada, in 2013.

He is currently a Postdoctoral Fellow with the Remote Sensing Laboratory (Department of Geography), McGill University, Montreal, where he is working on multispectral image processing and pattern recognition for cultural heritage preservation, forensic applications, and remote sensing.



Margaret Kalacska received the B.Sc., M.Sc., and Ph.D. degrees from the University of Alberta, Edmonton, AB, Canada, in 2001, 2003, and 2006, respectively, all in earth and atmospheric sciences.

She was a Postdoctoral Fellow with the School of Criminology, Simon Fraser University, Burnaby, Canada, until 2008. She is currently an Associate Professor of geography with McGill University, Montreal, QC, Canada. Her research interests include *in situ* and airborne hyperspectral remote sensing, signal processing, pattern recognition, and

modeling and satellite image analysis, with applications in environmental and forensic sciences. She has conducted field research in Costa Rica, Mexico, Panama, and Tanzania.



Max Mignotte received the D.E.A. degree in digital signal, image, and speech processing from the Grenoble Institute of Technology, Grenoble, France, in 1993, and the Ph.D. degree in electronics and computer engineering from the University of Bretagne Occidental, Brest, France, and the French Naval Academy, Brest, in 1998.

From 1998 to 1999, he was an INRIA Postdoctoral Fellow with the Department of Computer Science and Operations Research (DIRO), University of Montreal, QC, Canada, where he is currently an Associate Professor. His research interests include statistical methods, Bayesian inference, and hierarchical models for high-dimensional inverse problems.



Hossein Ziaei Nafchi received the B.Sc. and M.Sc. degrees in computer engineering from the Islamic Azad University, Tehran, Iran. He is currently working toward the Ph.D. degree in the Synchronmedia Laboratory for Multimedia Communication in Telepresence, École de Technologie Supérieure (University of Quebec), Montreal, QC, Canada.

His research interests include low-level vision, image quality assessment, and document image processing.



Mohamed Cheriet received the M.Sc. and Ph.D. degrees in computer science from the University of Pierre and Marie Curie (Paris VI), Paris, France, in 1985 and 1988, respectively.

Since 1992, he has been a Professor with the Department of Automation Engineering, École de Technologie Supérieure (University of Quebec), Montreal, QC, Canada, where he was appointed a Full Professor in 1998. He is the Founder and has been the Director since 1998 of the Synchronmedia Laboratory at École de Technologie Supérieure,

which targets multimedia communication in telepresence applications. He is an expert in computational intelligence, pattern recognition, mathematical modeling for image processing, and cognitive and machine learning approaches and perception. In addition, he has authored or coauthored over 350 technical papers in the field.

Dr. Cheriet is the founder and a former Chair of the IEEE Montreal Chapter of Computational Intelligent Systems.