# An Energy-Based Model Encoding Nonlocal Pairwise Pixel Interactions for Multisensor Change Detection

Redha Touati<sup>®</sup> and Max Mignotte

Abstract—Image change detection (CD) is a challenging problem, particularly when images come from different sensors. In this paper, we present a novel and reliable CD model, which is first based on the estimation of a robust similarityfeature map generated from a pair of bitemporal heterogeneous remote sensing images. This similarity-feature map, which is supposed to represent the difference between the multitemporal multisensor images, is herein defined, by specifying a set of linear equality constraints, expressed for each pair of pixels existing in the before-and-after satellite images acquired through different modalities. An estimation of this overconstrained problem, also formulated as a nonlocal pairwise energy-based model, is then carried out, in the least square sense, by a fast linear-complexity algorithm based on a multidimensional scaling mapping technique. Finally, the fusion of different binary segmentation results, obtained from this similarity-feature map by different automatic thresholding algorithms, allows us to precisely and automatically classify the changed and unchanged regions. The proposed method is tested on satellite data sets acquired by real heterogeneous sensor, and the results obtained demonstrate the robustness of the proposed model compared with the best existing state-of-the-art multimodal CD methods recently proposed in the literature.

*Index Terms*—Change detection (CD), energy-based model, FastMap, fusion of binary segmentations, heterogeneous sensors, multidimensional scaling (MDS) mapping, MDS, multimodal remote sensing, multisensors, multisource data, pairwise pixel interactions.

## I. INTRODUCTION

**N** OWADAYS, change detection (CD) is a major application and also an active research topic in remote sensing image processing, since it plays an important role in various application domains, including environmental monitoring, deforestation, urban planning, and land or natural disaster/damage monitoring and management to name a few.

Until now, many CD approaches have been proposed for addressing the classical *monomodal* CD issue [1]–[5], which occurs when the pairs of images are obtained from the same sensor or, more generally, the same imaging modality. In this

Manuscript received October 25, 2016; revised March 17, 2017 and August 18, 2017; accepted September 20, 2017. Date of publication October 18, 2017; date of current version January 26, 2018. (*Corresponding author: Max Mignotte.*)

The authors are with the Vision Laboratory, Département d'Informatique et de Recherche Opérationnelle, Faculté des Arts et des Sciences, Université de Montréal, Montréal, QC H3C 3J7, Canada (e-mail: touatire@iro.umontreal.ca; mignotte@iro.umontreal.ca).

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.2017.2758359

*monomodal* case, the two images, recorded at two different times but under similar imaging conditions, are generally first coregistered and corrected (preprocessing) and then, most often, used to generate a difference image by differencing or (log-)rationing. Finally, the resulting difference image is then segmented into two classes to distinguish changes of interest of the land cover/land use.<sup>1</sup>

A less explored and more challenging problem is the so-called *multimodal* CD problem, which is based on pairs of images obtained from different imaging modalities. In this case, the two input (before-and-after change) images present radically different image statistics (along with possibly different spatial and spectral resolutions), which cannot be compared with traditional methods borrowed from *monomodal* CD approaches relying on a simple pixelwise difference model.

Multimodal CD is especially appealing for several reasons. In fact, in furthermore to generalize the *monomodal* case, this processing treatment has obviously less restrictive considerations about the formation of the input data pair, since it must adapt itself to the characteristics of data with different natures. As a consequence, it should be more robust to natural variations in environmental variables, such as soil moisture or phenological state (such as flowering, maturing, drying, senescence, and harvesting) that cannot be avoided and well taken into account and corrected in the preprocessing step of a classical monomodal CD approach. Another interest is its inherent practicality that it could bring in several emergency situations. For example, it is useful in the case when an optical image of a given area is provided by an available remote sensing image archive data, and only a new synthetic aperture radar (SAR) image can be acquired (for technical reasons, lack of time, availability, or atmospheric conditions) in an emergency situation for the same area. In addition to providing a wide variety of information and properties about the study area, let us stress out that the additional information provided by two different sensors could also be used to our advantage, to improve the accuracy of the final CD map. This can be efficiently achieved if one succeeds in modeling the complementary and supplementary information provided by the

0196-2892 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications\_standards/publications/rights/index.html for more information.

<sup>&</sup>lt;sup>1</sup>By changes of interest of the land cover, it must be understood that we do not seek to detect, in this paper, changes such as atmospheric effects, including haze, persistent cloud cover, phenological changes, thin snow or ice cover, soil moisture, and shadow. In this paper, we are just referring to land cover changes, such as major construction or excavations, flooding, earthquake, and deforestation.

two different imaging modalities, with modeling techniques borrowed, for example, from the data fusion-based classification theory. Finally, let us also mention that this *multimodal* approach may be useful and sometimes indispensable in some specific cases, such as forest monitoring in tropical or boreal areas for which SAR, thanks to its ability to penetrate heavy clouds and fog, is often used as a complement to optical data. Another example, where SAR and optical sensors are complementary, is the case of frequently snow-covered regions of high altitudes, since the SAR is also able to penetrate a thin snow layer.

Up to now, a relatively few research works have been developed in CD using heterogeneous remote sensing images. Among the few existing models proposed in the literature, we can, however, mention the theoretical approach proposed in [6]. In this paper, the model is based on the assumption that some dependence indeed exists between the two images in unchanged areas and more precisely relies on the estimation of the local statistics of the first image through the point of view (in a statistical sense) of the second one (and vice versa). This dependence is modeled by quantile regression applied according to the copula theory, and Kullkack-Leiblerbased comparisons of these above-mentioned local statistics are applied to define a change measure, which is then finally analyzed by thresholding, in order to detect between change and no change areas. However, this method remains supervised, since it requires to learn the cumulative distribution function of the pixel intensity in the after image, conditioned to "no-change" hypothesis (i.e., the so-called copula) by using a manually selected (carefully chosen) training set of samples in the after image. Let us note that the model is also not easily generalizable in the case when more than one image, before and after a given event, is available and also not well suited (in terms of modeling, speed, and efficiency) to be used for images acquired with homogeneous sensors.

Another recent study was proposed by Prendes et al. [7] to overcome multisensor variability problems in CD. The authors propose an interesting multivariate statistical approach aiming to estimate a physical model, based on a mixture of multidimensional distributions, which both takes into account the relationships between the sensor responses to the objects contained in the observed scene, the physical properties of these objects, and the statistical properties of the noise corrupting the images. The parameters of this multidimensional mixture model are estimated by the expectation-maximization (EM) algorithm [8], which are then subsequently used to infer the relationships between the sensor physical properties involved through manifold learning. A statistical test based on this model allows to estimate the changes. An extension of this model, taking advantage of the correlations between adjacent pixels via a Markov random field model, has also been proposed by Prendes et al. [9]. However, this method also assumes a training set and more precisely that two training images associated with an unchanged area are available. Also, the method has been designed for heterogeneous multisensor in the case of optical/SAR data and is not easily generalizable for another pair of different sensors. Besides, it requires a lot of EM estimations (nearly one for each pixel), each one

relying on a good unsupervised estimation of the optimal number of existing components. Another CD approach for heterogeneous multisensor SAR data based on the multidimensional distribution mixture estimation has also been proposed in [10]. In particular, the authors have studied a new family of multivariate distributions whose margins are univariate gamma distributions with different shape parameters referred to as multisensor multivariate gamma distributions (MuMGDs), which are well suited for detecting changes in SAR images acquired by different sensors having different numbers of looks. The parameters of this multidimensional mixture model are estimated by the maximum likelihood (ML) or inference function for margins algorithm. Also the method, and more precisely the family of MuMGDs, has been especially designed for heterogeneous SAR sensors and cannot be easily generalized for other or different sensors. Finally, [11] proposes to use a methodology borrowed from coregistration (used in the field of medical imagery), based on the use of similarity measures (such as correlation ratio and mutual information) and to use the correspondence between the same points in the two images to detect eventual changes existing between the two data acquisitions. A comparison between the results of the performance of tested similarity measures is reported, which indicates that the mutual information and the cluster reward algorithm (CRA) seem the best indicator for multimodal (optical/SAR) CD. The CRA measure is built from the joint and the marginal probabilities, as the mutual information, and has a large value when the joint histogram has little dispersion (thus indicating a good correlation score). Nevertheless, these two measures are sensitive to the dimension of the estimation windows used for the pixel statistics and the similarity measure calculation.

Contrary to CD techniques based on a classical pixelwise modeling approach, we propose, as first and main contribution of this paper, a new change detector relying on the set of all pairs of (possibly nonlocal) pixels existing in the beforeand-after remote sensing images. This allows us to build a robust similarity feature map, especially well suited to estimate the difference between heterogeneous sensors exhibiting radically different image statistics. In our model, a set of linear equality constraints is expressed for each pair of pixels (in terms of gray levels or local statistics difference), and this overconstrained problem is then embedded or formulated into a final energy-based model encoding all the local pairwise pixel interactions. The quadratic complexity in the number of pixels of this resulting energy-based model is reduced to a linear complexity procedure, thanks to the FastMap-based optimization procedure proposed by Faloutsos and Lin [12]. This technique acts as an efficient and fast global minimizer of the cost function, integrating all the pairwise constraints, of our model by performing geometric linear projections (using the cosine law) in an *n*-dimensional space over an axis defined by a pair of pixels from the image (or in our application, from a pair of images) called pivots. Conceptually, the FastMap treats each distance or constraints between a pair of pixels (in terms of gray level difference) as a spring between the pixels, and tries to rearrange the gray-level values of each pixel to minimize the stress of the springs (also called the

*stress function*) or equivalently to satisfy all the constraints in the least square (LSQ) sense. Moreover, as second contribution, changed and unchanged areas are then finally identified, from this latter similarity feature map, by fusing the results of different automatic thresholding algorithms. In this way, we efficiently combine the intrinsic properties and criteria related to the different automatic thresholding algorithms in order to further increase the robustness and reliability of our *multimodal* CD strategy.

Let us note that, within the FastMap-based optimization and energy-based model framework encoding the nonlocal pairwise pixel interactions, we can mention the recent gait analysis model proposed in [13], which allows us to convert a video sequence of depth images of a human gait (on a treadmill) into an informative color map providing a quick overview of asymmetry existing in a given gait cycle for a rapid clinical diagnosis. In this model that uses a video data cube of the human gait, the pairwise interactions are defined to encode the degree of similarity existing between two gait movements (represented by two temporal depth signals) taken on two different locations on a human's body surface walking on a treadmill and such that the (pairwise) distance is defined as zero if the two motions are either pointwise similar or in perfect phase opposition (i.e., with a phase difference of half a gait cycle as it is normally the case for legs and arms during the gait cycle of a healthy subject). The set of distances between each pair of pixels is then used by the FastMap algorithm to generate a final mapping in which these distances should then code (as constraint) the L2-norm of color difference existing between this pair of pixels. By this means, two pixels (or two points located on the human's body surface) that share the same color on this mapping have to be considered as symmetric (and conversely, all the more antisymmetric as their color difference is high).

The remainder of this paper is organized as follows. Sections II and III describe, respectively, the proposed CD technique and the optimization procedure related to this model which allows us to estimate the similarity-feature map, from which changed and unchanged areas are then identified in Section IV by combining the results of different automatic thresholding algorithms. Section V presents a set of experimental results and comparisons with existing *multimodal* and *monomodal* CD algorithms. Finally, Section VI concludes this paper.

#### II. PROPOSED CHANGE DETECTION MODEL

Let us consider two (coregistered) bitemporal remote sensing (N pixel size) images,  $y^{t_1}$  and  $y^{t_2}$  acquired at two times (before and after a given event), in the same geographical area, from different sensors or from the same sensor but without the correction step, in terms of radiometric, atmospheric, and distortion consistencies and characteristics.

In order to estimate  $y^D$ , the similarity feature map, which is supposed to represent the difference between the multitemporal (multisensor) images, we rely on an improved version of the model introduced in [14] for the *monomodal* CD problem. In this model, first, we have to specify an overdetermined set of constraints to be satisfied (for  $y^D$ ) and expressed for each pair of pixels  $\langle s, t \rangle$  existing in each of the two multitemporal images  $y^{t_1}$  and  $y^{t_2}$ . The similarity map  $y^D$  is then seen as a solution to this set of constraints *via* the following nonlocal pairwise cost function to be optimized:

$$\hat{y}^{D} = \arg\min_{y^{D}} \sum_{\langle s,t \rangle_{s \neq t}} \left( \beta_{s,t} - \| y_{s}^{D} - y_{t}^{D} \|_{2} \right)^{2}$$
(1)

where the summation is done over all the pairs of pixels existing in the similarity feature image  $y^D$  to be estimated and  $\|.\|_2$  is the Euclidean distance. In (1), the set of  $\beta_{s,t}$  represents the set of N(N-1)/2 equality constraints expressed for each pair of pixels  $\langle s, t \rangle$ , in terms of difference of gray levels (or local statistics), in order to obtain a reliable similarity feature image  $y^D$  in which unchanged pixels will be associated with small gray-level values, whereas changed pixels will present rather large values.<sup>2</sup> These constraints are the following.

First, let us assume that two distinct pixels at locations s and t belong to the class *urban* at time  $t_1$  and still belong to the same class (*urban*), at time  $t_2$ . In this case, these two pixels should both belong to the (same) class label *unchanged area* in the binary segmentation of  $y^D$ . Let us consider another scenario: let us assume that two distinct pixels at locations s and t belong to the class *urban* at time  $t_1$  and both belong to the class *river*, at time  $t_2$  (due to a flooding event). In this case, these two pixels should both belong to the (same) class label *changed area* in the binary segmentation of  $y^D$ . These two scenarios can be summarized, as first constraint, as follows.

Constraint #1: Two distinct pixels  $\langle s, t \rangle$  should belong to the class label unchanged pixels or belong to the class label changed pixels, in the binary segmentation of  $y^D$ , if  $y_s^{t_1}$  and  $y_t^{t_1}$  have a similar gray level (or similar local statistics), and if  $y_s^{t_2}$  and  $y_t^{t_2}$  also have a similar gray level (or similar local statistics).

To satisfied this constraint,  $y_s^D$  and  $y_t^D$  should be assigned to a small gray-level value in  $y^D$  or should be assigned to a high gray-level value in  $y^D$  (since *s* and *t* should finally share the same label in the binary segmentation of  $y^D$ ), or equivalently, this constraint requires that the gray level difference between  $y_s^D$  and  $y_t^D$  is small.

If two pixels at locations *s* and *t* belong to a same class at time  $t_1$  (for example *urban*) and a different class at  $t_2$  (for example *urban* for pixel *s* and *river* for pixel *t*) or conversely. In this case, these two pixels should belong to a different class label in the binary segmentation of  $y^D$ , i.e., *unchanged pixels* for one of the two pixels and *changed pixels* for the other. This belonging to different class labels, in the binary segmentation of  $y^D$ , requires that two different (gray-level) values to be assigned to these two pixels in  $y^D$  (so that the binary segmentation of  $y^D$  correctly assigns two different

<sup>&</sup>lt;sup>2</sup>Let us note that our model can handle separately the individual channels or bands of a multispectral or hyperspectral sensor system, since, in our energy-based model, the difference between each pair of pixels can be formulated as a Euclidean distance between two *d*-dimensional spectral vectors with *d* being the number of spectral bands. By handling the bands separately, the similarity-feature map  $\hat{y}^D$  is estimated according to a similar (but opposite) criterion (i.e., a difference of "preservation of spectral distance") as the one often used as a criterion in the compression of hyperspectral images [15].



Fig. 1. Illustration of the four constraints (#1*a*, #1*b*, #2, and #3) corresponding to the scenario described in Section II. (From left to right) Image at time  $t_1$ before a flooding event (with the *urban* region at the center, the *vegetation* region all around the image, and the *river* region represented by a narrow, elongated region at the bottom right of the image), image at time  $t_2$  after a flooding event, and (ideal binarized) similarity map  $y^D$  (with the white region corresponding to the *changed area*) with the link (between each pair of pixels considered) drawn in such a way that its thickness is proportional to the associated distance defined by (2) between the gray levels (or local statistics vector) of each considered pair of pixels.

classes to these two pixels). This leads us to the Constraint #2.

The third and last cases, which lead us to the Constraint #3, involve a situation in which two pixels  $\langle s, t \rangle$  belong to a pair of different classes at time  $t_1$  (for example, *urban* for *s* and *vegetation* for *t*) and also belong to a pair of different classes, different from the first pair, at time  $t_2$  (for example, *urban* for *s* and *river* for *t*) or conversely. In this case,  $\langle s, t \rangle$  should also belong to a different class label in the binary segmentation of  $y_s^D$  and  $y_t^D$  is high (see Fig. 1).

In summary, the three above-specified constraints, in terms of pairwise gray-level difference in  $y^D$ , for each pair of locations  $\langle s, t \rangle$ , can be quite well satisfied by using [in (1)] the following pairwise distance between pixels  $\langle s, t \rangle$  at time  $t_1$  and  $t_2$  (which was empirically found and inspired from the max-Symmetric  $\chi^2$  distance combined with the city block distance [16]):

$$\beta_{s,t} = \left| \max\left( \frac{|y_s^{t_1} - y_t^{t_1}|}{y_s^{t_1}}, \frac{|y_s^{t_1} - y_t^{t_1}|}{y_t^{t_1}} \right) - \max\left( \frac{|y_s^{t_2} - y_t^{t_2}|}{y_s^{t_2}}, \frac{|y_s^{t_2} - y_t^{t_2}|}{y_t^{t_2}} \right) \right| \quad (2)$$

where we recall that  $y_{s^1}^{t_1}$  and  $y_{s^2}^{t_2}$  are, respectively, the gray level (or a local statistics vector) at pixel *s* in, respectively, the *before* and *after* image (i.e., at time  $t_1$  and  $t_2$ ). In our model, (1) thus becomes a composite cost function encoding our N(N-1)/2 constraints given by the *observed data* composed of all the pairwise pixels existing in  $y^{t_1}$  and  $y^{t_2}$ . Optimization of (1) will ensure a robust similarity feature map  $\hat{y}^D$  with land cover changes presenting significantly different values from those associated with the pixels belonging to unchanged areas. Nevertheless, it is important to note that the estimation of  $\hat{y}^D$ , according to (1), does not necessarily ensure that the low gray-level value is assigned for the pixel belonging to the *unchanged area* and conversely that the high gray-level values are assigned for the *changed area*. It could be the opposite. Nevertheless, let us mention that this latter case can be easily and automatically detected with a correlation metric or more simply by assuming that the land cover change is often much smaller than the unchanged area, and once detected, we can easily remedy it by simply inverting the gray-level values of the estimated image  $\hat{y}^D$  before its (binary) segmentation (see Section IV).

Let us note that the major advantage of the proposed model lies in its flexibility given by its LSQ criterion. Indeed, contrary to maximum a posteriori and ML approaches [7], [10], the proposed model does not require an explicit knowledge of the data distribution and also an (ML) parameter estimation step of these distribution laws, which can be complex and/or of very different natures, in the multimodal CD case, since the images (before and after the change) are from different modalities. Besides, contrary to machine learning-based approaches, the proposed scheme does not also require a large and representative and supervised training set. In addition, contrary to recent methods that seek to transform the original pair of temporal images into a new feature space or representation that significantly highlights the changes and which may be somewhat regarded as a CD method which could be invariant to imaging modality [17], the proposed method also has the advantage of not requiring the same number of spectral bands for the two (temporal) satellite images (as it is most often the case in practice for multimodal CD, since the two imaging modalities are assumed to be different).

## III. FASTMAP-BASED MODEL OPTIMIZATION

Let us note that the function to be minimized (1) is also the so-called *stress* function used as a criterion in the mapping based on the multidimensional scaling (MDS) technique [21], [22]. MDS has already been successfully used in a number of practical applications, such as color image segmentation [23], [24], hyperspectral compression [15], asymmetry detection [13], human action recognition [25], and database browsing and visualization [26] to name a few.

In our case, MDS is able to estimate a mapping, i.e., a gray-level similarity image  $y^D$ , such that the distances between each pair of gray-level values associated with pixels *s* and *t* are close to  $\beta_{s,t}$  as faithfully as possible (in the least square sense). Nevertheless, the originally proposed MDS algorithm (called *metric MDS* [21], [22]) is not appropriate in our application (and more generally for all large-scale applications), because it requires an entire  $N \times N$  distance matrix to be stored in memory with an  $O(N^2)$  complexity (*N* being the number of pixels). Instead, we have herein used a fast alternative called FastMap [12] whose main advantage is its linear complexity (thanks to a Nyström [27] approximation of the estimation of the eigenvectors and eigenvalues of the distance matrix) compared with the other MDS procedures.

In the proposed application, the FastMap allows us to find a mapping  $y^D$  with a linear complexity, such that the distances between each pair of gray-level values associated with pixels *s* and *t* are close to  $\beta_{s,t}$  as much as possible. To this end, we recall that the first step, and an essential element of the

FastMap algorithm, is to select two objects (pixels in our case) to form the projection line. These two pixels, also called pair of anchor nodes or pivots (or pivot line), are selected, such that the distance ( $\beta_{s,t}$  in our application) is maximal. To accomplish such a task, Faloutsos and Lin [12] proposed a linear heuristic algorithm based on a deterministic procedure called "choose distant objects." The second step is to project any other object (pixels) onto this orthogonal axis (pivot line) by employing the cosine rule.

However, the price paid for the low linear complexity of the FastMap is its sensitivity to outliers and nonlinearities. In our case, this characteristic may give a poor or noisy estimation of the similarity image  $y^D$ . In order to get a more reliable estimation, an interesting solution is obtained by averaging the estimations from different pivot lines. To this end, the linear heuristic and deterministic procedure proposed by Faloutsos and Lin [12] can be easily modified in order to propose more than one pivot line.

## **IV. FUSION-BASED SEGMENTATION STEP**

Finally, in order to achieve more robustness, changes are then identified, from the (previously estimated) similarity image  $y^D$ , by combining the results of T = 5 different automatic thresholding algorithms<sup>3</sup> (namely [28]–[32]). In this way, this strategy (already been used in [33]) allows us to synergistically integrate multiple different criteria, for which these binary segmentation algorithms have been designed to be optimal in order to further increase the robustness and reliability of our proposed segmentation scheme. In our application, this binary fusion process is simply achieved by using a majority vote filter using a 3-D window  $W \times W \times T$ whose the first two dimensions are spatial and the third dimension indexes the different binary thresholded maps to be fused. In our application, this majority vote is achieved with a 3-D window, which is spatially centered on the pixel to be classified, and that collects the binary class labels of the different binary thresholded maps and, finally by assigning to that central pixel, the class label that has the majority vote. This strategy ensures both the spatial regularization of the final fused (detection) map result and also a reliable decision fusion between the results obtained by different thresholding strategies.

## V. EXPERIMENTAL RESULTS

## A. Results on Multimodal Data Sets

To evaluate the efficiency of our proposed model, we validate our approach on three real pairs of heterogeneous remote sensing images (see Fig. 2), provided by the CNES center (French National Centre for Space Studies), and already used in [7] and [18] and for which the different change mask constructions were provided by a photointerpreter. Besides, we have at our disposal one pair of heterogeneous SAR images given in [10]. This allows us to compare the performance of our model with the four existing state-of-the-art *multimodal* CD algorithms in this field, namely, the one introduced in [7] and [18] (and its improved version proposed in [9]), the multidimensional EM-based model proposed in [6], and the method proposed in [10] for heterogeneous multisensor SAR data. Besides, we have also compared our result with change detector traditionally used in *monomodal* approaches provided by the ORFEO Toolbox [20].

- The first *multimodal* data set is a pair of SAR/optical satellite images (Toulouse, France), with size 4404 × 2604 pixels, before and after a construction. The SAR image was taken by the TerraSAR-X satellite (February 2009) and the optical image by the Pleiades (High-Resolution Optical Imaging Constellation of CNES) satellite (July 2013). The TSX image was coregistered and resampled in [19] with a pixel resolution of 2 m to match the optical image.
- 2) The second one is a pair of optical/SAR satellite images (Gloucestershire region, in southwest England, near Gloucester), with size  $2325 \times 4135$  pixels, before and after a flooding (on a mixture of urban and rural areas). The optical image is a screenshot from Google Earth and comes from the Quick Bird 02 (QB02) VHR satellite (July 15, 2006), and the SAR image was acquired by the TerraSAR-X satellite (July 2007). The TSX image presents a resolution of 7.3 m, and the QB02 image (with resolution of 0.65 m and 0% cloud cover) was coregistered and resampled in [19] to match this resolution.
- 3) The third data set is a pair of different optical images, with different sensor specifications (i.e., spectral bands), with size 2000 × 2000 pixels (with the same resolution of 0.52 m and 0% cloud cover), before and after the construction of a building (in the urban area of Toulouse, France). The first optical image is captured by the Pleiades sensor (May 2012), and the second optical image is a screenshot from Google Earth and is acquired by WorldView2 satellite from three (red, green, and blue) spectral bands (July 11, 2013). The WorldView2 VHR-image was coregistered in [19] to match the Pleiades image.
- 4) The fourth *multimodal* data set [10] is a pair of SAR/SAR satellite images (Gloucester, U.K.) before and during a flood, with size 762 × 292 pixels, acquired by the RADARSAT satellite. The numbers of looks for the before SAR image are 1-look image (September 2000), and the numbers of looks for the after image are 5-looks (October 2000). These two SAR images have a resolution of about 40 m.

We have considered the pairwise distance formula given by (2), where  $y_s^{t_1}$  corresponds to the simple gray level of the image (and not a local statistics vector around a neighborhood of *s*). In the case of an optical image, this also requires the conversion of the possible color image to a grayscale image. We have finally considered the final majority vote with a squared window spatial size set to  $W = 3 \times 3$ .

<sup>&</sup>lt;sup>3</sup>Let us note that the concept of combining classifiers for the improvement of the performance of individual classifiers is known, in machine learning field, as a committee machine, ensemble classifiers, ensemble methods, or mixture of experts [34], [35]. In this context, Dietterich [35] has provided an accessible and informal reasoning, from statistical, computational, and representational viewpoints, of why ensembles can improve results.



Fig. 2. (From left to right) Image  $t_1$  (before event), image  $t_2$  (after event), ground truth, estimated similarity feature map  $\hat{y}^D$ , final binary map result, and confusion map (white region: TN, red region: TP, blue region: FP, and cyan region: FN). (From top to bottom) *Multimodal* image pair: SAR/optical (image from TerraSAR-X/Pleiades satellite of Toulouse, France), optical/SAR (image from QB02/TerraSAR-X satellite of Gloucester, U.K.), heterogeneous optical/optical (image from Pleiades/WorldView2 of Toulouse, France), and heterogeneous SAR/SAR (image from SAR 1-look/SAR 5-looks of Gloucester, U.K.).

We have summarized, respectively, in Tables I and II, the accuracy rates and the confusion matrix obtained by our approach, compared with the four existing *multimodal* CD methods (see also Fig. 2), and some classical CD methods borrowed from *monomodal* techniques. We can notice that the proposed model outperforms quantitatively the four existing state-of-the-art approaches recently published in this field.

Fig. 3 shows the binary maps obtained by the Prewitt [28], Kapur [29], Zack [30], Yen [31], and Shanbhag [32] binarizers on the feature similarity map generated by the FastMap in the case of the second and the fourth *multimodal* data set and the fusion results obtained by the proposed fusion strategy based on a 3-D ( $3 \times 3 \times 5$ ) majority vote filter. We can notice that the different binarizers estimate a different optimal

#### TABLE I

ACCURACY RATE OF CD ON THE FOUR HETEROGENEOUS DATA SETS OBTAINED BY THE PROPOSED METHOD AND THE STATE-OF-THE-ART *Multimodal* CHANGE DETECTORS (FIRST TOP PART OF EACH TABLE) AND *Monomodal* CHANGE DETECTORS (SECOND BOTTOM PART OF EACH TABLE)

		QB02 / TSX	Accuracy				
		Proposed method	0.949	Pleiades / WorldView2	Accuracy		
TSX / Pleiades	Accuracy	Prendes et al. [9], [19]	0.918	Proposed method	0.853	SAR 1-look / SAR 5-looks	Accuracy
Proposed method	0.867	Prendes et al. [7]	0.854	Prendes et al. [18], [19]	0.844	Proposed method	0.781
Prendes et al. [18]	0.844	Copulas [6],[7]	0.760	Correlation [18], [19]	0.679	Chatelain et al. [10]	0.732
Correlation [18]	0.670	Correlation [6], [7]	0.688	Mutual Inf. [18], [19]	0.759	Correlation [10]	0.521
Mutual Inf. [18]	0.580	Mutual Inf. [6], [7]	0.768	Pixel Dif. [19], [20]	0.708	Ratio edge [10]	0.382
		Pixel Dif. [7], [20]	0.782	Pixel Ratio [19], [20]	0.661		
		Pixel Ratio [7], [20]	0.813				

TABLE II CONFUSION MATRIX FOR THE FOUR *Multimodal* DATA SETS, i.e., [TSX/PLEIADES] (4404 × 2604 PIXELS), [QB02/TSX] (2325 × 4135 PIXELS), [PLEIADES/WORLDVIEW2] (2000 × 2000 PIXELS), AND [SAR 1-LOOK/ SAR 5-LOOKS] (762 × 292 PIXELS)

Multi-modal pair	TP	TN	FP	FN
TSX / Pleiades	4.83%	81.96%	10.08%	3.13%
QB02 / TSX	4.36%	90.58%	3.00%	2.06%
Pleiades / WorldView2	9.08%	76.30%	8.54%	6.08%
SAR 1-look / SAR 5-looks	9.88%	68.22%	14.24%	7.66%



Fig. 3. Individual binary CD maps given by, respectively, the Prewitt [28], Kapur [29], Zack [30], Yen [31], and Shanbhag [32] binarizers on the similarity-feature map generated by the FastMap (see Fig. 2) and fusion results using a majority vote filter using a 3-D ( $3 \times 3 \times 5$ ) window.

threshold leading to a different binary map, since different criteria are used. Nevertheless, the proposed fusion strategy ensures both an efficient spatial and consensus regularization, even if the statistical distribution of the feature similarity map is not clearly bimodal (see Fig. 4).

We can notice that some histograms of the similarity map are not bimodal. In our case, this is not a problem, since four of the five binarizers, used in our procedure, do not necessarily assume that the histogram is bimodal. For example, the socalled triangle method presented in [30] proposes to construct a line between the histogram peak and the farthest end of the



Fig. 4. Histogram of the four similarity-feature maps of the four *multimodal* image pairs generated by the FastMap (see Fig. 2).

histogram, and the threshold is the point of maximum distance between the line and the histogram. Another binarization method, which is applicable, even if the histogram is not bimodal, is the binarizer proposed in [29], which uses the entropy concept. In this case, the threshold is estimated, such that the entropies of distributions above and below are maximized. In the same spirit, [31] uses the maximum correlation criterion as a more computationally efficient alternative to entropy measures. Finally, [32] proposes an extension of the method proposed in [29]. Only the binarizer proposed in [28] seeks two modes in the histogram and thus relies on the presence of a bimodal shape of the histogram. The method consists in iteratively smoothing the histogram (using a running average of size 3) until two peaks remain; the threshold is then the minimum or midpoint between the two peaks. Nevertheless, algorithmically, if a bimodality in the histogram is not detected after a maximum number of iterations, the threshold is generally the gray value corresponding to the highest peak. All these different binarizers generally ensure the diversity, which is then needed for a reliable subsequent fusion process.

Let us stress out that the proposed model can also be easily generalized in the case where more than one image, before and after a given event, is available. Indeed, this can be easily done by considering the following averaged pairwise distance:

$$\overline{\beta}_{s,t} = \frac{1}{N_p} \sum_{\langle b,a \rangle} \beta_{s,t}(y^b, y^a)$$
(3)

with  $\beta_{s,t}(y^a, y^b)$  the distance expressed in (2) for an image, respectively, belonging to the before (and after) event set  $y^b$   $(y^a)$  and where the averaging is done over all possible pairs of images available before and after a given event  $(\langle b, a \rangle)$ 



Fig. 5. Comparison of the similarity-feature maps obtained by Prendes *et al.*'s method [19] and the proposed method on the first three *multimodal* data sets. From lexicographic order, ground truths, similarity-feature map obtained by Prendes *et al.*'s method in false colors (the red areas represent high similarity between the two images, while the blue areas correspond to low similarity) and similarity-feature maps obtained by the FastMap-based proposed method for, from top to bottom, the TSX/Pleiades, QB02/TSX, and Pleiades/WorldView2 data sets.

(and  $N_p$  is the number of averaging pairs). Let us note, however, that this technique can be applied only if the date of change event is known beforehand. This averaging procedure could even improve the estimation of  $\beta_{s,t}$ , since the averaging procedure is a reliable strategy to reduce the noise of any estimation procedure. In addition, it would be interesting to study, in this multiple before-and-after image case, the effect of a median, harmonic, or geometric mean operator instead of this arithmetic mean-based operator.

Fig. 5 shows a comparison of the similarity-feature maps obtained by Prendes et al.'s method [19] and the proposed method on the first three multimodal data sets. By comparison with Prendes's method, the proposed CD method seems to visually produce more distinctly two clustering structures (modeling the unchanged and changed areas) a bit more separated and more compacted (with lower internal variance within a cluster) and with less overlap. Besides, our method yields to more spatially and properly regularized (or less noisy) similarity-feature maps. It is interesting to note that our multimodal CD strategy is able to detect very thin structure in the changed area class, such as the thin S-shaped region in the middle bottom of the middle image (contrary to the Prendes et al.'s method). We can also notice that some false positives are detected in the same locations in the two methods (see the rectangular shape at the top-right of the bottom-left quadrant of the third image). Let us note that, in our case, the similarity-feature maps closely depend on the pairwise

distance used [see (2)]. A clever and more discriminative pairwise distance metric would allow us to obtain a better similarity-feature map. In addition, it is worth mentioning that the proposed method still remains perfectible if a better binarization strategy is found.

## B. Results on Monomodal Data Sets

In order to demonstrate that our approach is flexible enough to also be efficiently used in *monomodal* CD (i.e., with homogeneous sensor), we present a set of experimental results obtained on different real, publicly available, monomodal optical, multitemporal, multispectral, airborne SAR, or radar data sets with available ground truth. In this case, we use the following and simple pairwise distance  $\beta_{s,t}$ :

$$\beta_{s,t} = \left| \left| y_s^{t_1} - y_t^{t_1} \right| - \left| y_s^{t_2} - y_t^{t_2} \right| \right| \tag{4}$$

which turned out a bit more efficient that the distance used in *multimodal* case. In addition, for the *monomodal* case, we have considered the final majority vote filter with a squared window size set to  $W = 3 \times 3$ .

- The first data set<sup>4</sup> (see Fig. 7) is a pair of optical satellite images produced by the EROS data center in southwest U.S., corresponding to a part of Reno-Lake Tahoe area of Nevada (acquired on August 5, 1986, and August 5, 1992), with size 200 × 200 pixels, captured by the Landsat Multispectral Scanner. The Burn images show a change that results from forest fire phenomena. The Cuts images show a change described by a decrease in the surface area of the lake that results from drought effects. The Dray Lake images show a change that corresponds to the beginning and culmination of drought conditions in the western U.S. The Surface Disturbance images show increased surface disturbance due to construction or excavations for construction, including road resurfacing or paving.
- 2) The second data set [4], [36] [37], [39] (see Fig. 6) is provided by the Defence Research and Development Canada, Ottawa (Canada), and is two multitemporal SAR images relating to Ottawa, with size 290 × 350 pixels, acquired by the RADARSAT SAR sensor, respectively, in July 1997 during the summer flooding, and August 1997 after the summer flooding.
- The third data set [4], [36]–[38] (see Fig. 6) is a pair of two multitemporal SAR images with size 301 × 301 pixels (the pixel resolution is 12.5 m), acquired by ERS-2 satellite (the European Remote Sensing satellite). It presents a natural phenomenon, generally occurring during the rainy season in the Switzerland area, near

<sup>4</sup>1) The first monomodal data set: Burn, Cuts, and Dray Lake images and their ground truths have been downloaded from http://geochange.er.usgs.gov/sw/changes/natural/reno-tahoe/ Surface disturbance and its ground truth has been downloaded from:

https://geochange.er.usgs.gov/sw/changes/anthropogenic/vegas/const.html. 2) Images of the second, third, and fourth monomodal data sets and their

ground truths have been provided by Xiong *et al.* [40] and Li *et al.* [41].

3) Images of the fifth data set have been provided by UMD-NASA and downloaded from http://glcf.umd.edu/data/landsatTreecover/ and their ground truths from http://www.landcover.org/.

TABLE III Accuracy Rate of CD Obtained by Different State-of-the-Art Methods, on Bern (ERS-2), Ottawa (RADARSAT), and Beijing (Airborne SAR) Data Sets

Dataset	Proposed method	[4]	[36]	[37]	[38]	[39]	[40]
Bern (ERS-2)	.993	.996	.997	.996	.996	-	-
OTTAWA (RADARSAT)	.943	.972	.965	.974	-	.988	-
BEIJING (SAR AIRBORNE)	.986	-	-	-	-	-	.997
Nb. of images tested	17	3	2	3	2	5	2

Ottawa, CA [SAR] Ottawa, CA [SAR] integrading a base of the second s

Fig. 6. Experimental results on *monomodal* SAR (second and third) and airborne SAR (fourth) data set: Ottawa, Bern, and Beijing. (From left to right) Image acquired at time  $t_1$  and  $t_2$ , ground truth, similarity feature map, final (changed/unchanged) binary segmentation result, and confusion map (white region: TN, red region: TP, blue region: FP, and cyan region: FN) obtained by our approach.

the city of Bern, in April 1999 before the flooding and in May 1999 after the flooding.

- 4) The fourth data set [40] (see Fig. 6) shows a pair of X-band airborne SAR (intensity) images with size  $900 \times 900$  pixels (the pixel resolution is 0.5 m), acquired over a field in Beijing, China, on April 4 and 6, 2004. It shows the number and positions of the vehicle on the field which were different during the two data acquisition dates.
- 5) The fifth data set (see Fig. 8) is a collection of images with size 7660 × 7402 pixels (the pixel resolution is 30 m), provided by the NASA/USGS Global Land Survey [42], captured by the multispectral scanner Landsat-5 (TM) and Landsat-7 Enhanced Thematic Mapper Plus and showing various change phenomena in landscape, in different areas, between 2000 and 2005. For each pair of images of the same area, this data set proposes a ground truth image containing the different evolutions undergone by the area for five years (thrust drills and loss of trees).

Table III summarizes the different CD accuracy rates obtained by our approach with a comparison with other *monomodal* "state-of-the-art" approaches [4], [36]–[39], [40] for different data sets with different imaging modalities (with the total number of images tested in each case). We can see that the different changed–unchanged detection binary map results match fairly the different regions present in the ground truth, and that the most changed regions for the different imagery modalities are well recognized by our strategy (see Figs. 6–8).

# C. Shadow Effects

In this paper, <sup>1</sup> our goal is to detect changes of interest in the land cover or land use. So far, we have considered, in our *mul-timodal* experiments, a major or a localized minor construction and two types of flooding (Fig. 2) and in the monomodal case; a deforestation (due to a forest fire), two examples of decrease of a given lake's surface area (resulting from drought effects), a surface disturbance (i.e., an excavations/construction for road paving) (Fig. 7), two different floodings, the detection of vehicles in an agricultural field (Fig. 6) and various change phenomena in the landscape such as thrust drills, loss of trees, changes in tree cover over time, etc. (Fig. 8).



Fig. 7. Experimental results on *monomodal* optical Eros center (first) data set: Burn, Cuts, Dray Lake, and Surface Disturbance. (From left to right) Image acquired at time  $t_1$  and  $t_2$ , ground truth, similarity feature map, final (changed/unchanged) segmentation result, and confusion map (white region: TN, red region: TP, blue region: FP, and cyan region: FN) obtained by our approach.

As an additional experiment, it would also be interesting to see how the proposed CD model behaves when one of the two images has glow and shadow effects. To this end, for the homogeneous CD detection case, we have considered a stereo panchromatic data set provided in [43], with size  $900 \times 900$  pixels (the pixel resolution is 5 m) and captured by Cartosat-1 satellite sensor. This pair of panchromatic images is acquired over Arges region (Romania near Piatra Craiului national park), on October 2008 and November 2009 and shows a forest changes caused by storms, and containing many shadow areas caused by steep terrain due to the mountainous forest area [43].

We have applied our CD model with and without any preprocessing step on the image pair. As preprocessing, we use a simple (double) histogram matching method [44]. More precisely, the *before* image is histogram matched to the *after* image to give the preprocessed *before* image, and the *after* image is then histogram matched to the latter (preprocessed *before*) image. We show in Fig. 9 the obtained results with a comparison in Table IV with other state-of-the-art *monomodal* change detectors studied in [43]. The result shows that our

method is also robust in this *monomodal* case. Nevertheless, it would have also been interesting to evaluate how our model behaves in the *multimodal* case involving shadow effects, especially between SAR and optical images, since the shadow is a quite different phenomenon between these two imagery modalities that cannot be corrected with a simple preprocessing scheme as a simple histogram matching method. This special case still remains to be studied.

# D. Discussion

We can also notice that the rate accuracy of our method remains comparable, although slightly lower than the other *monomodal* "state-of-the-art" approaches but above all that the strength of the proposed model is its ability to process a wide variety of satellite imaging modalities (i.e., multitemporal, multispectral, airborne SAR, or radar data) potentially degraded by different noise types and different noise levels (see, for example, Fig. 6 where the SAR images are corrupted by different speckle noise levels). This peculiarity certainly comes from the fact that our model is, before all, designed to



Fig. 8. Experimental results on *monomodal* UMD-NASA (fifth) data set. (From left to right) Image acquired at time  $t_1$  and  $t_2$ , ground truth, similarity feature map, final (changed/unchanged) binary segmentation result, and confusion map (white region: TN, red region: TP, blue region: FP, and cyan region: FN) obtained by our approach.

Arges, ROM [Panchromatic]



Fig. 9. Panchromatic data set: image  $t_1$  and  $t_2$ , ground truth, similarity feature map, final (changed/unchanged) segmentation result, and confusion map (white region: TN, red region: TP, blue region: FN) obtained by the proposed approach. First row presents the results obtained without any preprocessing step. Second row presents the results obtained with a double histogram matching method-based preprocessing step.

be used for the *multimodal* CD case. The average accuracy rate obtained by our CD approach over 17 image pairs stemming from this five different *monomodal* data sets with the distance expressed by (4) is  $\rho = 0.94$  (94%). With the distance expressed by (2), especially well suited for the *multimodal* CD case, the average accuracy rate obtained on these five different *monomodal* data sets is  $\rho = 0.92$  (92%).

Consequently, we can say that the proposed method also has the defect of its main quality. Its ability to process a wide variety of imaging modalities (with different noise types and levels) explains why it will also be less accurate than a specific monomodal CD model only dealing with a specific type of noise and for which the similarity map, obtained by some local operations, follows a particular mixture of distributions whose each distribution's shape may be theoretically estimated and for which the parameters of the finite distribution mixture can then be efficiently estimated with an EM-like algorithm to finally obtain a reliable binary CD map.

#### TABLE IV

KAPPA STATISTIC OF CD ON THE PANCHROMATIC SHADOW DATA SET OBTAINED BY THE PROPOSED METHOD AND OTHER UNSUPERVISED (FIRST TOP PART OF THE TABLE) AND SUPERVISED (SECOND PART OF THE TABLE) STATE-OF-THE-ART MONOMODAL CHANGE DETECTORS [43]

Method	Kappa				
Proposed method (with preprocessing)	0.513				
Proposed method (without preprocessing)	0.281				
kMNF OPTI <sup>‡</sup> [43]	0.487 - <b>0.509</b> - 0.506 - 0.501 - 0.487 - 0.475				
Height Difference <sup>‡</sup> [43]	0.127 - 0.316 - 0.469 - <b>0.526</b> - 0.0 - 0.0				
CVA <sup>‡</sup> [43]	0.07 - 0.242 - 0.403 - <b>0.457</b> - 0.0 - 0.0				
k-Means [43]	0.472				
ICDA [43]	0.495				
OSVM [43]	0.478				
Random Forests [43]	0.432				

<sup>‡</sup>based on different threshold levels given in increasing order

We can also notice that the proposed model has, comparatively, more difficulties to separate the changed and unchanged areas when the SAR imaging modality is involved (see Fig. 2). This behavior can be probably explained by the inherent multiplicative speckle noise degrading the quality of any SAR images and creating, for each land cover class, a kind of macrotexture with grainy patterns (and referring to variations in radar brightness that are larger than many resolution cells). More precisely, this can be explained by the fact that the pairwise distances, used in our energy-based model [see (1) and (2)], with  $y_p^t$  corresponding to the simple gray level at site v, cannot fully model a coarse texture. In a multimodality case involving SAR imaging, a more appropriate model would have been to consider local statistics around the pixel and, therefore, a distance computed between two feature vectors instead of two scalars. Nevertheless, experimentally, it would seem that a complex distance (i.e., a more complex, realistic model) also leads to a harder optimization problem and, finally, a more approximated solution given by the FastMap optimization procedure. In our case, a good solution of a simpler, approximate model seems preferable than an approximate solution of a complex (and maybe more realistic) model.

## VI. CONCLUSION

In this paper, we have proposed a new model for CD in heterogeneous remote sensing images. Our method is mainly based on the estimation of a robust similarity feature map, containing the difference caused by the event between the bitemporal multisensor images involved, and which is formulated as the solution of a set of constraints expressed for each pixel pair via a global cost function. A FastMap-based optimization and then a simple fusion step, used to combine a set of binary segmentation maps generated by several automatic thresholding algorithms on this similarity feature map, allow us to identify between the changed and unchanged areas. The proposed method is unsupervised and does not require a training data set or the estimation of an important parameter and can be used for any pairs of heterogeneous sensors. Besides, the proposed method is flexible, since it can also be efficiently used in monomodal CD (i.e., with homogeneous sensor). It can be easily generalized in the case where more than one image, before and after a given event, is available or be used to handle separately the individual bands of a multispectral or hyperspectral image (with d spectral bands),

by simply formulating the constraint or difference between each pair of pixels, as the distance existing between two *d*dimensional spectral vectors. Finally, the model is perfectible by identifying a better pairwise distance or a better binarization strategy and its time complexity is linear with the total pixel number.

#### ACKNOWLEDGMENT

The authors would like to thank Dr. J. Prendes and CNES (French National Center for Space) for sharing the *multimodal* data set in order to validate our model, and Dr. Y. Li [41], Dr. B. Xiong [40], and UMD-NASA for having provided them, respectively, the second and third, fourth, and fifth monomodal data sets [42]. They would like to thank Dr. J. Tian who put at our disposal the change detection Panchromatic shadow data set [43] and also for the time spent in providing the comparison results. They would also like to thank all the other researchers who kindly made their databases available for our study and the comparisons made in this paper, and the four anonymous reviewers for their numerous comments and suggestions that helped improve both the scientific content and the presentation of this paper.

#### REFERENCES

- L. Bruzzone and D. F. 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.
- [2] L. Bruzzone and D. F. Prieto, "An adaptive semiparametric and contextbased approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 452–466, Apr. 2002.
- [3] J. Liu, M. Gong, Q. Miao, L. Su, and H. Li, "Change detection in synthetic aperture radar images based on unsupervised artificial immune systems," *Appl. Soft Comput.*, vol. 34, pp. 151–163, Sep. 2015.
- [4] 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.
- [5] R. Hedjam, M. Kalacska, M. Mignotte, H. Z. Nafchi, and M. Cheriet, "Iterative classifiers combination model for change detection in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 6997–7008, Dec. 2016.
- [6] G. Mercier, G. Moser, and S. B. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1428–1441, May 2008.
- [7] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J. Y. Tourneret, "A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 799–812, Mar. 2015.
- [8] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B (Methodol.)*, vol. 39, no. 1, pp. 1–38, 1976.

- [9] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "Change detection for optical and radar images using a Bayesian nonparametric model coupled with a Markov random field," in *Proc. IEEE Int. Conf. Acoustic, Speech, Signal Process. (ICASSP)*, Brisbane, QLD, Australia, Apr. 2015, pp. 1513–1517.
- [10] F. Chatelain, J.-Y. Tourneret, and J. Inglada, "Change detection in multisensor sar images using bivariate gamma distributions," *IEEE Trans. Image Process.*, vol. 17, no. 3, pp. 249–258, Mar. 2008.
- [11] V. Alberga, "Similarity measures of remotely sensed multi-sensor images for change detection applications," *Remote Sens.*, vol. 1, no. 3, pp. 122–143, 2009.
- [12] C. Faloutsos and K.-I. Lin, "FastMap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, San Jose, CA, USA, Jun. 1995, pp. 163–174.
- [13] A. Moevus, M. Mignotte, J. de Guise, and J. Meunier, "A perceptual map for gait symmetry quantification and pathology detection," *Biomed. Eng. OnLine*, vol. 14, no. 1, p. 99, 2015.
- [14] R. Touati and M. Mignotte, "A multidimensional scaling optimization and fusion approach for the unsupervised change detection problem in remote sensing images," in *Proc. 6th IEEE Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Oulu, Finland, Dec. 2016, pp. 1–6.
- [15] M. Mignotte, "A bicriteria-optimization-approach-based dimensionalityreduction model for the color display of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 501–513, Feb. 2012.
- [16] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *Int. J. Math. Models Methods Appl. Sci.*, vol. 1, no. 4, pp. 300–307, 2007.
- [17] C. Wu, B. Du, and L. Zhang, "Slow feature analysis for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2858–2874, May 2014.
- [18] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J. Tourneret, "Performance assessment of a recent change detection method for homogeneous and heterogeneous images," in *Revue Française de Photogrammétrie et de Télédétection*, France, vol. 209. 2015, pp. 23–29.
- [19] J. Prendes, "New statistical modeling of multi-sensor images with application to change detection," Ph.D. dissertation, Dept. Institut de Recherche en Informatique de Toulouse et Laboratoire de Télécommunications Spatiales et Aéronautiques, Toulouse, France, 2015.
- [20] OTB Development Team. (2014). The ORFEO Toolbox Software Guide. [Online]. Available: http://orfeo-toolbox.org/
- [21] W. S. Torgerson, "Multidimensional scaling: I. Theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, Dec. 1952.
- [22] T. Cox and M. Cox, *Multidimensional Scaling*. London, U.K.: Chapman & Hall, 1994.
- [23] M. Mignotte, "MDS-based multiresolution nonlinear dimensionality reduction model for color image segmentation," *IEEE Trans. Neural Netw.*, vol. 22, no. 3, pp. 447–460, Mar. 2011.
- [24] M. Mignotte, "MDS-based segmentation model for the fusion of contour and texture cues in natural images," *Comput. Vis. Image Understand.*, vol. 116, no. 9, pp. 981–990, 2012.
- [25] R. Touati and M. Mignotte, "MDS-based multi-axial dimensionality reduction model for human action recognition," in *Proc. 11th Conf. Comput. Robot Vis. (CRV)*, Montreal, QC, Canada, May 2014, pp. 262–267.
- [26] A. Khlif and M. Mignotte, "Segmentation data visualizing and clustering," *Multimedia Tools Appl.*, vol. 76, no. 1, pp. 1531–1552, 2017.
- [27] E. J. Nyström, "Über die praktische auflösung von integralgleichungen mit anwendungen auf randwertaufgaben," *Acta Math.*, vol. 54, no. 1, pp. 185–204, 1930.
- [28] J. Prewitt and M. Mendelsohn, "The analysis of cell images," Ann. New York Acad. Sci., vol. 128, no. 3, pp. 1035–1053, 1996.
- [29] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for graylevel picture thresholding using the entropy of the histogram," *Comput. Vis., Graph., Image Process.*, vol. 29, no. 3, pp. 273–285, 1985.
- [30] G. W. Zack, W. E. Rogers, and S. A. Latt, "Automatic measurement of sister chromatid exchange frequency," *J. Histochem. Cytochem.*, vol. 25, no. 7, pp. 741–753, 1977.
- [31] 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.
- [32] 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.

- [33] F. Melgani and Y. Bazi, "Robust unsupervised change detection with Markov random fields," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Denver, CO, USA, Jul./Aug. 2006, pp. 208–211.
- [34] A. Sharkey, Combining Artificial Neural Nets Ensemble and Modular Multi-Net Systems. New York, NY, USA: Springer-Verlag, 1999.
- [35] T. Dietterich, "Ensemble methods in machine learning," in Proc. 1st Int. Workshop Multiple Classifier Syst., vol. 1857. 2000, pp. 1–15.
- [36] J. Lu, J. Li, G. Chen, L. Zhao, B. Xiong, and G. Kuang, "Improving pixel-based change detection accuracy using an object-based approach in multitemporal SAR flood images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3486–3496, Jul. 2015.
- [37] H. Li, M. Gong, and J. Liu, "A local statistical fuzzy active contour model for change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 582–586, Mar. 2015.
- [38] J. Ma, M. Gong, and Z. Zhou, "Wavelet fusion on ratio images for change detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 6, pp. 1122–1126, Nov. 2012.
- [39] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 125–138, Jan. 2016.
- [40] B. Xiong, Q. Chen, Y. Jiang, and G. Kuang, "A threshold selection method using two SAR change detection measures based on the Markov random field model," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 287–291, Mar. 2012.
- [41] Y. Li, M. Gong, L. Jiao, L. Li, and R. Stolkin, "Change-detection map learning using matching pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4712–4723, Aug. 2015.
- [42] J. Sexton *et al.*, "Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error," *Int. J. Digit. Earth*, vol. 6, no. 5, pp. 427–448, 2013.
- [43] J. Tian, A. A. Nielsen, and P. Reinartz, "Improving change detection in forest areas based on stereo panchromatic imagery using kernel MNF," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7130–7139, Nov. 2014.
- [44] D. Shapira, S. Avidan, and Y. Hel-Or, "Multiple histogram matching," in *Proc. 20th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2013, pp. 2269–2273.



**Redha Touati** received the M.Sc. degree in computer science from the University of Montreal, Montreal, QC, Canada, in 2014, where he is currently pursuing the Ph.D. degree with the Vision Laboratory, Department of Computer Science and Operations Research, in collaboration with the Imaging and Vision Department, Computer Research Institute of Montréal, Montréal, QC, Canada.

His research interests include statistical methods and applied mathematics in video imaging and remote sensing imagery.



Max Mignotte received the DEA 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 Occidentale, Brest, France, and the Digital Signal Laboratory, French Naval academy, Brest, in 1998.

He was an INRIA Post-Doctoral Fellow with the Department of Computer Science and Operations Research (DIRO), University of Montreal, Montreal,

QC, Canada, from 1998 to 1999. He is currently a Professor with the Computer Vision and Geometric Modeling Laboratory, DIRO, University of Montreal. He is also a member with the Laboratoire de Recherche en Imagerie et Orthopedie, Centre de Recherche du CHUM, Hopital Notre-Dame, Montreal, QC, Canada, where he is also a Researcher. His research interests include statistical methods, Bayesian inference, and energy-based models (especially encoding non-local pairwise pixel interactions) for solving diverse large-scale high-dimensional ill-posed inverse problems in imaging.