Multimodal Change Detection in Remote Sensing Images Using an Unsupervised Pixel Pairwise-Based Markov Random Field Model

Redha Touati, Max Mignotte, and Mohamed Dahmane

***Abstract*— This work presents a Bayesian statistical approach to the multimodal change detection (CD) problem in remote sensing imagery. More precisely, we formulate the multimodal CD problem in the unsupervised Markovian framework. The main novelty of the proposed Markovian model lies in the use of an observation field built up from a pixel pairwise modeling and on the bitemporal heterogeneous satellite image pair. Such modeling allows us to rely instead on a robust visual cue, with the appealing property of being quasi-invariant to the imaging (multi-) modality. To use this observation cue as part of a stochastic likelihood model, we first rely on a preliminary iterative estimation technique that takes into account the variety of the laws in the distribution mixture and estimates the parameters of the Markovian mixture model. Once this estimation step is completed, the Maximum a posteriori (MAP) solution of the change detection map, based on the previously estimated parameters, is then computed with a stochastic optimization process. Experimental results and comparisons involving a mixture of different types of imaging modalities confirm the robustness of the proposed approach.**

***Index Terms*— Change detection, heterogeneous sensors, iterative conditional estimation (ICE), Markov random field (MRF), multimodal remote sensing, multisource data, multisensors, parameter estimation, pixel pairwise modeling, stochastic optimization, unsupervised Markovian segmentation.**

I. Introduction

MULTIMODAL Change Detection (CD) [1] is a procedure used to identify any land cover changes that occurred between two satellite images acquired at different times, in the same geographical area but by different kinds of sensors. Multimodal CD is a growing interest task which can be considered as a generalization of the basic and classic monomodal CD problem as it requires less stringent requirements about the characteristics and origin of the acquired data. It is also a challenging task since, such a procedure must be powerful and flexible enough to model any existing heterogeneous data types (thus sharing different statistics) in remote sensing imagery and to handle the same problems that have been already solved by monomodal CD techniques [2]–[5] such as anomaly and target detection (eventually in the presence of diurnal and seasonal variations), natural, land or environmental monitoring, damage monitoring (earthquake, flooding, landslides, etc.) or urban planning, to name a few.

Multimodal CD has recently aroused a growing interest, in the remote sensing community since this technique allows to relax the assumption of homogeneous and co-calibrated measurements and consequently to exploit the huge amount of heterogeneous data, we can now get from various archives or from different types of existing Earth observing satellites. In addition, the practical and technical advantages of such multimodal analysis procedure are obvious and are widely described in the literature, for instance [6]. Finally, let us add that the different imaging modalities may be complementary and this sensor fusion technique could potentially be exploited (not only in Geoscience imaging [7]) for further improving the change detection and analysis of land surfaces with complex properties subject to extreme conditions (e.g. temperature, fire, ice, etc.).

Despite its undeniable potential, there are relatively few research works that have been devoted to heterogeneous or multimodal CD using machine learning or image processing. Nevertheless, we can identify four main categories. First non-parametric based techniques such as learning machine algorithms (since these techniques do not assume explicitly a specific parametric distribution for the data) [8]–[13] or unsupervised non-parametric based procedures, that do not require supervised training step, such as the energy based model, in the least-squares, sense proposed in [6] and satisfying an overdetermined set of constraints, expressed for each pair of pixels existing in the before-and-after images. Secondly, algorithms relying on similarity measures with invariance properties according to the imaging modality [14]–[16]. Thirdly procedures mainly based on a transformation or projection of the two multimodal images to a common feature space, in

Manuscript received June 27, 2018; revised February 20, 2019 and May 22, 2019; accepted July 30, 2019. Date of publication August 13, 2019; date of current version October 9, 2019. This work was supported in part by the Computer Research Institute of Montreal (CRIM) and in part by the Ministry of Economic Science and Innovation (MESI) of the Government of Québec. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Trac D. Tran. *(Corresponding author: Redha Touati.)*

R. Touati is with the Vision Laboratory, Département d'Informatique et de Recherche Opérationnelle (DIRO), Faculté des Arts et des Sciences, Université de Montréal, Montréal, QC H3C 3J7, Canada, and also with the Research and Development Vision Département, Centre de Recherche Informatique de Montréal (CRIM), Montréal, QC H3N 1M3, Canada (e-mail: touatire @iro.umontreal.ca).

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

M. Dahmane is with the Research and Development Vision Département, Centre de Recherche Informatique de Montréal (CRIM), Montréal, QC H3N 1M3, Canada (e-mail: mohamed.dahmane@crim.ca).

Digital Object Identifier 10.1109/TIP.2019.2933747

which the two heterogeneous images share the same statistical properties and on which classical monomodal CD methods can then be applied [17]–[23]. Finally parametric models that we now describe in more details since the proposed model fits into this category. In parametric techniques, a set (or mixture) of multivariate distributions are generally used to model the joint statistics or the dependencies between the two imaging modalities. More precisely, local models of dependence between unchanged areas are modeled according to the copula theory in [24] and based on these models, Kullkack-Leibler-based comparisons on local statistical measures are then used to generate a similarity map which is subsequently binarized. An appealing two-step multivariate statistical approach has also been proposed in [25]–[27] where the first step aims to estimate a physical model, based on a mixture of multidimensional distributions (both taking into the noise model, the relationships between the sensor responses to the objects and their physical properties). A statistical test based on this model then allows to estimate the changes. In the same spirit, the authors in [28] also propose to first estimate a multidimensional distribution mixture estimation based on a new family of multivariate distributions with different shape parameters and especially well suited for detecting changes in SAR images with different numbers of looks.

Herein, we propose a different statistical approach, relying on an observation field built up from a pixel pairwise modeling on the bitemporal heterogeneous satellite image pair. This allows us to indirectly model the joint statistics or the dependencies between the two imaging modalities and to finally base our CD (or binary segmentation) model on a relevant imaging modality-invariant visual cue whose likelihood model parameters can be fully estimated within the standard ICE (Iterative conditional estimation) framework [29], [30] with ML (Maximum Likelihood) estimator in the Least Square (LS) sense. Once the estimation step is completed, the MAP (Maximum a posteriori) solution of the change detection map, based on the previously estimated parameters, is then computed with a stochastic optimization strategy.

The remainder of this paper is organized as follows: Section II describes proposed unsupervised Markovian CD model by first defining the ingredients of the proposed MRF model (likelihoods and priors), and the proposed strategy based on a two-step procedure; namely a parameter estimation step and a segmentation step. Section III presents a set of experimental results and comparisons with existing multimodal change detection algorithms. In this section, we describe the robustness assessment for our proposed technique. Finally, Section IV concludes the paper.

II. Unsupervised Markovian CD Model

Herein, we formulate the multimodal CD problem in the unsupervised Bayesian framework. To this end, a possible and interesting approach is a two-step process. First, a parameter estimation step is conducted to infer the likelihood model parameters (in the ML sense). Then a second step is devoted to the binary segmentation or change detection itself based on the value of estimated parameters [31].

Let y^{t_1} and y^{t_2}, a pair (co-registered) bi-temporal remote sensing (N pixel size) images acquired at two different times (before and after a given event), in the same geographical area, and from different sensors. We first consider $\mathbf{X} = \{X_s, s \in S\}$ the random label field located on the same rectangular lattice S of N sites s associated to the two input images, with each X_s taking its value in the discrete set $\Lambda_{label} = \{e_0 = \textit{no-change},\ e_1 = \textit{change}\}$.

A. Observation Field

In the classic monomodal (or homogeneous) CD case, the two coregistred images y^{t_1} and y^{t_2} are first compared pixel by pixel in order to generate a *difference image* by differencing or (log-)rationing (*i.e.,* by using a temporal gradient or a log temporal gradient operator). This latter *difference image* is such that the pixels associated with land cover changes present gray-level values significantly larger, compared to those associated with unchanged areas and this visual cue based on the norm of the temporal luminance gradient $|y^{t_1} - y^{t_2}|$ is a robust cue on which the *observation field* and the likelihood distributions of a MRF model can be built up. In the multimodal (or heterogeneous) case, this temporal gradient is not a robust and reliable cue. Indeed, the color or grey value of each pixel is not a useful information since the gray levels of the same region, in y^{t_1} and y^{t_2} may be radically different according to the characteristics of the two different imaging modalities. Conversely, $y_s^{t_1}$ and $y_s^{t_2}$ may be locally coded with the same (grey or color) value in the two imaging modalities but representing two completely different textures or regions.

In our application, in order to rely on a robust visual cue with the specific property to be (nearly) invariant to the imaging modality, we have considered a pixel pairwise modeling, estimated from (y^{t_1}, y^{t_2}) and for each pixel pairs $\langle s, t\rangle$ existing in S, with the following symmetric relation:

$$y_{\langle s,t\rangle} = \left| |\mathbf{y}_s^{t_1} - \mathbf{y}_t^{t_1}|_1 - |\mathbf{y}_s^{t_2} - \mathbf{y}_t^{t_2}|_1 \right| \tag{1}$$

where $|.|_1$ is the L_1 norm and $\mathbf{y}_s^{t_1}$ and $\mathbf{y}_s^{t_2}$ represents a local statistics vector at pixel s (that will be made explicit in the following) in the *before* and *after* image.

This visual cue $y_{\langle s,t\rangle}$ already proposed, in a simplified version without texture in [6],[1] is defined as a function of the pixel pair $\langle s, t\rangle$ and (y^{t_1}, y^{t_2}). This is discriminant in our application since, whatever the imaging modality, $y_{\langle s,t\rangle}$ will give a high value for two pixels at sites s and t that must belong to two different class labels (*no-change*/*change* in our case) in the CD binary map (to be estimated) and conversely, will give a low value, for two pixels at sites s and t that must share the same class label (see Fig. 1 and its caption).

To use this cue in our Bayesian framework, we first consider that the set of $y_{\langle s,t\rangle}$ values are a realization of a random variable vector $\mathbf{Y}_{\langle s,t\rangle} = \{Y_{\langle s,t\rangle}, Y_{\langle s,u\rangle}, ..., Y_{\langle u,v\rangle}, ...\}$ gathering the $N(N-1)$ random variables associated to each site pair, that we herein call the random (pixel pairwise) observation field and secondly that $\mathbf{X}_{\langle s,t\rangle}$ is its corresponding random

[1] in which the authors define a set of constraints which will be satisfied, in the least squares (LS) sense, by a multidimensional scaling-based constraint model aiming to generate a soft CD map that is then binarized.

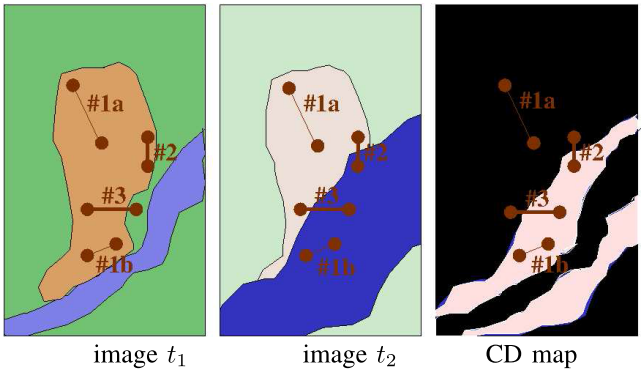

Fig. 1. In lexicographic order; (synthetic) image before a flooding event, with an *urban* region at the center, a *vegetation* region all around the image and a *river* crossing the image from right to left (bottom); image of the same area (and obtained by another imaging modality, thus with different colored textures) after a flooding event, and ground truth CD map (with the white region corresponding to the *changed area*). Illustration of the four pixel pair locations $\langle s,t\rangle$ leading to the four possible cases (#1*a* & #1*b*: low value for $y_{\langle s,t\rangle}$ implying that $\langle s,t\rangle$ must share the same class label in the CD map x, #2 & #3: high value for $y_{\langle s,t\rangle}$ implying that $\langle s,t\rangle$ must share a different class label between s and t in the final CD map x to be estimated. The link (between each pair of pixels considered) is drawn from such way that its thickness is proportional to the value that Eq. (1) could give.

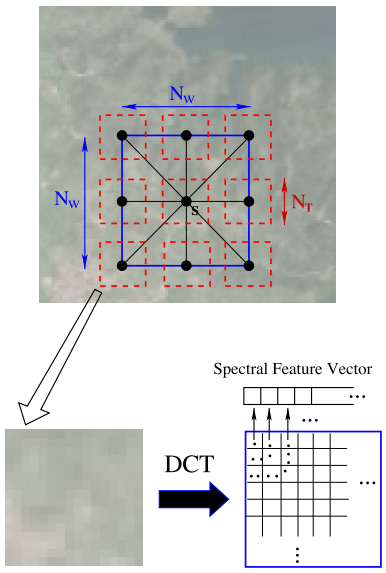

Fig. 2. We consider, for each pixel s, a sub-sample $\mathcal{G}_s$ of 8 pairs of pixels $\langle s,t\rangle$ in which the pixel t is regularly distributed around a squared window of size $N_w \times N_w$ (with $N_w = 41$ in our application). Besides $\mathbf{y}_s$ and $\mathbf{y}_t$ (see Eq. (1)) is in fact a radially-integrated (DCT) spectral feature vector encoding the textural and structural information existing around each local squared region of size $N_T \times N_T$ ($N_T = 16$) centered at the considered pixel.

(pairwise) label field taking its value in $\Lambda_{label_{\langle s,t\rangle}} = \{id, di\}$. The pixel-pairwise label id means that the pixel at location s an t must share the same (**id***entical*) class label in the final CD map $\hat{x}$ to be estimated (leading to the configuration $\langle x_s = change, x_t = change\rangle$ or $\langle x_s = no-change, x_t = no-change\rangle$). Conversely, $x_{\langle s,t\rangle} = di$ means that we have a **di***fferent* configuration, *i.e.*, either the configuration $\langle x_s = change, x_t = no-change\rangle$ or $\langle x_s = no-change, x_t = change\rangle$.

In our application, in order to decrease the computational load of our algorithm and to keep a quasi-linear complexity with respect to the number of image pixels, we consider for each pixel, a sub-sample $\mathcal{G}_s$ of 8 pairs of pixels regularly distributed around a squared window of size $N_w \times N_w$ centered around the pixel s (see Fig. 2). Besides, we consider at site s or t (image before t_1 or after t_2) a feature vector $\mathbf{y}$ (see Eq. (1)) encoding the textural and structural information existing around each local squared region of size $N_T = 16 \times N_T = 16$ centered at the considered pixel (see Fig. 2). To this end, in our application, we first estimate the Discrete Cosine Transform (DCT) of each local squared window, compute its module (*i.e.,* its absolute value since DCT is real) and then apply a half circular or Radial Integration Transform (RIT) (using a bi-linear interpolation) to estimate a spectral descriptor vector of size $N_T/2$. Since this texture descriptor is obtained from the compressed domain, this has the ability to be both, robust to noise (several denoisers are built from a filtering in this DCT domain [32], [33]), be strongly reduced in size, while combining the properties to encode a texture with rotation and translation invariance. In addition, compared to a Discrete Fourier Transform (DFT), the DCT has a higher compression efficiency and above all, its spectrum is less biased than the DFT spectrum (especially when this one is computed on small images) due to the even-symmetric extension properties of DCT that avoids the generation of artifacts or spurious spectral components created by edge effects caused by the inherent periodic nature of the DFT. Also, DCT uses real computations, unlike the complex computations used in DFT. This makes the computation of DCT extremely fast.[2]

B. Likelihood Distributions

To use the observation measure $y_{\langle s,t\rangle}$ (see Eq. (1)) in a Bayesian settings, we must, before all, estimate the (marginal / conditional) likelihood distributions of $Y_{\langle s,t\rangle}$ in the two possible cases; identical pixel-pairwise label $x_{\langle s,t\rangle} = id$ or not $x_{\langle s,t\rangle} = di$.

1) Identical Pixel-Pairwise Label Distribution: In our experiments, we have noticed that, if $x_{\langle s,t\rangle} = id$, $P_{Y_{\langle s,t\rangle}|X_{\langle s,t\rangle}}$ is well approximated, for a given s, by an exponential distribution $p_{id} = \mathcal{E}(.;\lambda)$ with shape (or inverse rate) parameter λ, *i.e.*:

$$p_{id}(y_{\langle s,t\rangle}) = P_{Y_{\langle s,t\rangle}|X_{\langle s,t\rangle}}(y_{\langle s,t\rangle}|x_{\langle s,t\rangle}=id) = \frac{\exp\left(-y_{\langle s,t\rangle}/\lambda\right)}{\lambda} \cdot H(y_{\langle s,t\rangle}) \quad (2)$$

with the right-continuous Heaviside step function, $H(x)$ where $H(0) = 1$ and $\lambda > 0$ (which makes the distribution supported on the interval $[0\ \ \infty]$).

This approximation can be justified and understood if we notice that, for a pixel pair $\langle s,t\rangle$ located in a spatially and temporally homogeneous region (*e.g.*, cases #1*a* & #1*b*

[2]For the implementation of this step, we have used the very fast 16×16 (FFT2D) DCT package implemented in C code by Takuya Euro (functions DDCT16X16S tested in program SHRTDCT.C) and available online at http address given in [34].

illustrated in Fig. 1), *i.e.* for $x_{\langle s,t\rangle} = id$, $y_{\langle s,t\rangle}$ is in fact related to the norm of a first order temporal gradient over a n-order (n is the distance in pixel between s and t) spatial gradient and the gradient norm of the intensity image is known to be well approximated by a simple exponential distribution [35] or its numerous variant (such as its truncated [36], [37], generalized [38] or long-tail version with a shape and scale factor [39], [40]).

2) Different Pixel-Pairwise Label Distribution: In the case of $x_{\langle s,t\rangle} = di$ (different pixel-pairwise labels), we have empirically noticed that the Gaussian law $p_{di} = \mathcal{N}(.; \mu, \sigma^2)$ is well adapted to describe the measure $y_{\langle s,t\rangle}$:

$$\begin{aligned} P_{di}(y_{\langle s,t\rangle}) &= P_{Y_{\langle s,t\rangle}|X_{\langle s,t\rangle}}(y_{\langle s,t\rangle}|x_{\langle s,t\rangle}=di) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_{\langle s,t\rangle}-\mu)^2}{2\sigma^2}\right) \end{aligned} \tag{3}$$

Let us note that, in the case of a heterogeneous pair of images and two heterogeneous temporal regions ($x_{\langle s,t\rangle} = di$), this distribution is consistent with the central limit theorem and the fact that this results from the addition of lots of different phenomena (*i.e.,* lot of numerical differences achieved between many possible different textural feature vectors, coded by different imaging modality with possibly different scales, etc.).

3) Data Likelihood and Posterior Distribution: Now, if we assume that the pairwise data $Y_{\langle s,t\rangle}$ are independent conditionally on the pairwise labeling process $X_{\langle s,t\rangle}$, and take into consideration the sub-sample $\mathcal{G}_s$ of pairs of pixels defined in Section II-A (and shown in Fig. 2), one gets:

$$P_{\mathbf{Y}_{\langle s,t\rangle}|\mathbf{X}_{\langle s,t\rangle}}(.) = \prod_{s\in S} \prod_{\substack{\langle s,t\rangle \\ t\in\mathcal{G}_s}} P_{Y_{\langle s,t\rangle}|X_{\langle s,t\rangle}}(y_{\langle s,t\rangle}|x_{\langle s,t\rangle}) \tag{4}$$

In addition, if we consider that the distribution of $\mathbf{X}$ is stationary and Markovian and choose a standard prior for the distribution of the labeling process $\mathbf{X}$ and that the CD map x defines $x_{\langle s,t\rangle}$ without ambiguity, one gets for the posterior distribution[3]:

$$P_{\mathbf{X}|\mathbf{Y}_{\langle s,t\rangle}}(.) \propto \prod_{s\in S} \prod_{\substack{\langle s,t\rangle \\ t\in\mathcal{G}_s}} P_{Y_{\langle s,t\rangle}|X_{\langle s,t\rangle}}(.) \cdot P_X(x) \tag{5}$$

If we consider a standard isotropic Pott-type prior model relative to the second-order neighborhood system η_s, with identical potential value β for the different (horizontal vertical, right diagonal or left diagonal) *cliques* $\langle s,t\rangle$ of η_s, thus a model favoring for $\hat{x}$, homogeneous regions of the same class *no-change* or *change*; *i.e.*, $P_X(x) \propto -\beta \exp\{\sum_{\langle s,t\rangle\in\eta_s}[1 - \delta(x_s, x_t)]\}$ [41], (where δ is the delta Kronecker function) $\hat{x}$, the CD map to be estimated becomes the global maxima of the following corresponding posterior probability:

$$\begin{aligned} \hat{x} &\propto \arg\max_x \prod_{s\in S} P_{X_s|Y_{\langle s,t\rangle}}(.) \\ &\propto \arg\max_x \prod_{s\in S} \Bigg\{ \prod_{\substack{\langle s,t\rangle \\ t\in\mathcal{G}_s}} P_{Y_{\langle s,t\rangle}|X_{\langle s,t\rangle}}(.) \\ &\quad \cdot \underbrace{\exp -\left\{\beta \sum_{\langle s,t\rangle\in\eta_s} [1-\delta(x_s,x_t)]\right\}}_{P_{X_s}(x_s)} \Bigg\} \end{aligned} \tag{6}$$

In this context, the corresponding posterior energy to be minimized is:

$$\begin{aligned} U(x,y) = &\sum_{s\in S} \sum_{\substack{\langle s,t\rangle \\ t\in\mathcal{G}_s}} -\ln P_{Y_{\langle s,t\rangle}|X_{\langle s,t\rangle}}(y_{\langle s,t\rangle}|x_{\langle s,t\rangle}) \\ &+ \sum_{\langle s,t\rangle\in\eta_s} \beta[1-\delta(x_s,x_t)] \end{aligned} \tag{7}$$

and $\hat{x}_{\text{MAP}} = \arg\min_x\{U(x, y)\}$.

C. Iterative Conditional Estimation

1) Principle: In our unsupervised Markovian segmentation case, we have to estimate in a first step (*estimation step*), the parameter vector $\Phi_{y_{\langle s,t\rangle}}$ which defines respectively the likelihood distributions $p_{id}(y_{\langle s,t\rangle})$ and $P_{di}(y_{\langle s,t\rangle})$ (or $P_{Y_{\langle s,t\rangle}|X_{\langle s,t\rangle}}(y_{\langle s,t\rangle}|x_{\langle s,t\rangle})$ for each two classes $x_{\langle s,t\rangle}$ of $y_{\langle s,t\rangle}$. (see Equations (2)-(3)), *i.e.*, the parameter vector $\Phi_{y_{\langle s,t\rangle}}(\lambda, \mu, \sigma)$ gathering the scale parameter of the exponential law $p_{id}(y_{\langle s,t\rangle})$ and the mean μ and σ parameters of the Gaussian distribution $p_{di}(y_{\langle s,t\rangle})$.

In our case, this estimation step is particularly challenging for three reasons; first, one has to deal with a mixture of different distributions (exponential and Gaussian) which are also strongly mixed (see Fig. 3) and which also exhibits different mixing proportions (generally the class di is under weighted (<15%) because this class is related to the fewer pixel-pairwise labels, or transitions, existing between the class *change* and the class *no-change* (see Fig. 1).

To this end, we resort to the ICE [29], [30] iterative procedure which is able to cope with different distributions and which experimentally turned out to be more efficient than the classical Expectation Maximization (EM) [42] algorithm or its stochastic version; the Stochastic EM (SEM) [43]. This efficiency can be explained by the fact that the ICE [29], [30] procedure can also be viewed as the stochastic and Markovian version of the EM procedure and thus is also constrained by the distribution of $\mathbf{X}$ defined as stationary and Markovian.

The ICE procedure is a fixed point algorithm which first requires to find an estimator $\hat{\Phi}_{y_{\langle s,t\rangle}} = \Phi(x_{\langle s,t\rangle}, y_{\langle s,t\rangle})$ providing an estimate of $\Phi_{y_{\langle s,t\rangle}}$ based on the complete data configuration $(x_{\langle s,t\rangle}, y_{\langle s,t\rangle})$ (see Appendix).

2) ICE-Based ML Estimator: For the Gaussian law, a ML estimate of (μ, σ^2), based on the complete data configuration, can be easily given by the empirical mean and empirical

[3]Using likelihood (Eq. (4)), and since the CD map x defines $x_{\langle s,t\rangle}$ without ambiguity, the joint distribution of $(\mathbf{X}, \mathbf{X}_{\langle s,t\rangle}, \mathbf{Y}_{\langle s,t\rangle})$ writes: $P_{\mathbf{X},\mathbf{X}_{\langle s,t\rangle},\mathbf{Y}_{\langle s,t\rangle}}(.) = P_{\mathbf{X}}(.) \cdot P_{\mathbf{Y}_{\langle s,t\rangle}|\mathbf{X}_{\langle s,t\rangle}}(.)$ and the posterior distribution gets: $P_{\mathbf{X},\mathbf{X}_{\langle s,t\rangle}|\mathbf{Y}_{\langle s,t\rangle}} = P_{\mathbf{X}|\mathbf{Y}_{\langle s,t\rangle}} / P_{\mathbf{Y}_{\langle s,t\rangle}} \propto P_{\mathbf{X}|\mathbf{Y}_{\langle s,t\rangle}}$ since the pixel pairwise observation field $y_{\langle s,t\rangle}$ is known ($P(y_{\langle s,t\rangle}$ is equal to a constant) and does not depend on the unobserved labeling process x (or the unobserved pairwise labeling process $x_{\langle s,t\rangle}$).

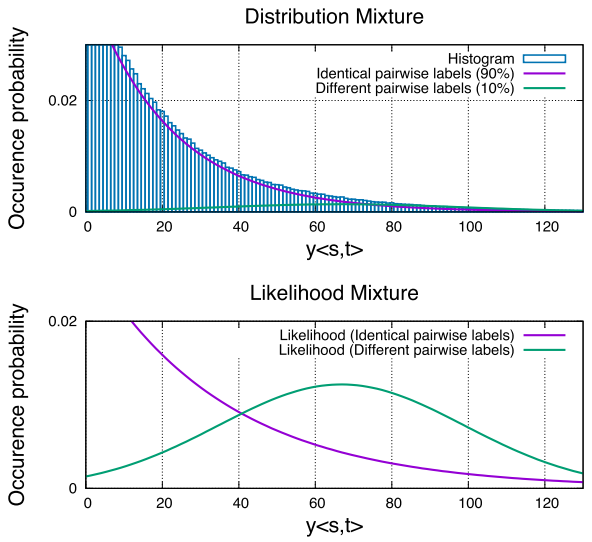

Fig. 3. From top to bottom; Distribution mixture: Histogram of $y_{\langle s,t\rangle}$ associated to the heterogeneous image pair Dataset-3 and the two weighted (90% of identical pairwise labels and 10% of different pairwise label) mixture components that are estimated by the ICE procedure (see Section II-C). Likelihood mixture: the two preceding likelihood distributions (without proportion priors) that are estimated by the ICE procedure.

variance. If $N_{di} \stackrel{\triangle}{=} \#\{x_{\langle s,t\rangle} = di\}$, one gets:

$$\hat{\mu}(x_{\langle s,t\rangle}, y_{\langle s,t\rangle}) = \hat{\mu}(x, y_{\langle s,t\rangle}) = \frac{\sum_{x_{\langle s,t\rangle}=di} y_{\langle s,t\rangle}}{N_{di}} \quad (8)$$

$$\hat{\sigma}^2(x, y_{\langle s,t\rangle}) = \frac{\sum_{x_{\langle s,t\rangle}=di} (y_{\langle s,t\rangle} - \hat{\mu})^2}{(N_{di} - 1)} \quad (9)$$

For the exponential law, if $N_{id} \stackrel{\triangle}{=} \#\{x_{\langle s,t\rangle} = id\}$, a ML estimate of the shape parameter is:

$$\hat{\lambda}(x, y_{\langle s,t\rangle}) = \frac{\sum_{x_{\langle s,t\rangle}=id} y_{\langle s,t\rangle}}{N_{id}} \quad (10)$$

In our Bayesian CD framework, we do not need to estimate the proportion of each class. Nevertheless, the mixing proportion can be easily estimated within this procedure with the empirical frequency estimator; $\pi_{id} = N_{id}/(N_{id}+N_{di})$ and $\pi_{di} = N_{di}/(N_{id}+N_{di})$.

3) ICE Algorithm: $\Phi_{y_{\langle s,t\rangle}}(\lambda, \mu, \sigma^2)$ are thus estimated with the ICE procedure in the following way:

- *Parameter Initialization*: we start from a CD map x randomly sampled from two classes (*change* / *no-change*) and start from $\Phi^{[0]}_{y_{\langle s,t\rangle}} = (\lambda^{[0]}, \mu^{[0]}, \sigma^{2\,[0]})$.
- **ICE procedure**: $\Phi^{[k+1]}_{y_{\langle s,t\rangle}}$ is then computed from $\Phi^{[k]}_{y_{\langle s,t\rangle}}$ as follows:
  1) *Stochastic Step*: using the Gibbs sampler, one realization x of the CD map is simulated according to the posterior distribution $P_{X/Y_{\langle s,t\rangle}}(x/y_{\langle s,t\rangle})$, with parameter vector $\Phi^{[k]}_{y_{\langle s,t\rangle}}$.
     More precisely, for each site s (lexicographically), we sample x_s with the local version of Eq. (5), *i.e.*,
     $$P_{X_s|Y_{\langle s,t\rangle}}(.) \propto \prod_{\substack{\langle s,t\rangle \\ t\in\mathcal{G}_s}} P_{Y_{\langle s,t\rangle}|X_{\langle s,t\rangle}}(.) \cdot P_{X_s}(x_s) \quad (11)$$
     a) with $P_{Y_{\langle s,t\rangle}|X_{\langle s,t\rangle}}$ an Exponential law for $x_{\langle s,t\rangle} = id$ (see Section II-B1).
     b) with $P_{Y_{\langle s,t\rangle}|X_{\langle s,t\rangle}}$ a Gaussian law for $x_{\langle s,t\rangle} = di$ (see Section II-B2).
  2) *Estimation Step*: the parameter vector $\Phi^{[k+1]}_{y_{\langle s,t\rangle}}$ is estimated with the ML estimator of the "complete data" (see Eqs (8), (9), (10)).
  3) Repeat until convergence is achieved;
     i.e., if $\Phi^{[k+1]}_{y_{\langle s,t\rangle}} \not\approx \Phi^{[k]}_{y_{\langle s,t\rangle}}$, we return to Stochastic Step.

In our application, one has to deal with a mixture of different distributions which are strongly mixed with unbalanced mixing proportions (see Fig. 1). This makes the convergence of the ICE procedure still difficult in some cases. Thus, it is necessary to add an additional hard constraint. In our application, we can capture the fact that the shape parameter λ of the exponential distribution $p_{id}(y_{\langle s,t\rangle})$ is in fact not too far from its shape parameter computed from the set of $y_{\langle s,t\rangle}$ regardless of its label $x_{\langle s,t\rangle}$ (id or di) since there are generally fewer labels di (let $\lambda^\star$ this parameter value). In fact, since the true shape parameter λ of the exponential distribution $p_{id}(y_{\langle s,t\rangle})$ is computed from $y_{\langle s,t\rangle}$ given $x_{\langle s,t\rangle} = id$, λ is thus computed from a subset of smaller values of $\{y_{\langle s,t\rangle}\}$, or equivalently, we can surely assert that a reliable estimation for λ is necessarily a value inferior to $\lambda^\star$. We model this by imposing the hard constraint $\lambda = \lambda^\star/\alpha$ for the different iteration of the ICE procedure.

In order to further help the iterative ICE procedure, we start, at iteration $^{[0]}$ with $\Phi^{[0]}_{\langle s,t\rangle} = (\lambda^{[0]}, \mu^{[0]}, \sigma^{[0]})$, with $\mu^{[0]} = 2\,\lambda^{[0]}$ (with $\lambda^{[0]} = \lambda^\star$) and $(\sigma^2)^{[0]} = 1000$ to model the fact that the *mean* of the Gaussian is generally greater to the λ parameter and that the variance of the Gaussian is generally around 1000. We finally use the *Stochastic Step* with a Gibbs sampler with a temperature equals to 0.25 in order to allow a fast convergence and to reduce the number of explored solutions around the initialization values.

D. Segmentation Step

Once the estimation step is completed, the MAP (Maximum a posteriori) solution of the CD map x, based on the previously estimated parameters, is then computed. In our application, the energy function (see Eq. (7)) is complex and the MAP solution is difficult to estimate (essentially due to the strongly mixed likelihood mixture model which is possibly of slightly different shapes according to the type of multimodality). In order to avoid local minima we must resort to a simulated annealing (SA) procedure [41] with a sufficient number of iterations about 200000 iterations in our application), or equivalently by varying the temperature of a Gibbs sampler (see Eq. (11)) from the initial temperature $T_o = 1.25$ to $T_{final} = 0.01$ with a slow geometric decreasing schedule such as $T = T_o \times (0.999975)^k$.

Once $\hat{x}_{\text{MAP}}$ is estimated, it is important to note that, due to the pixel (label) pairwise modeling, there are two global minima to the optimization problem defined in Eq. (7). One for the solution ("1" for *change* class and "0" for *no-change* class) and the second one corresponding to its binary inverse (*i.e.*, its binary complement, with "0" for *change* class and

"1" for *no-change* class). In our case, this ambiguity can be easily resolved with a correlation metric or more simply by assuming that the land cover change is generally smaller than the unchanged area.

It takes between 30 and 70 minutes to perform a SA (depending on the image size) with so many iterations for a non-optimized C++ code running on Linux on a $i7-930$ Intel CPU, 2.8 GHz. Nevertheless, by considering a Jacobi-type version of the Gauss-Seidel based SA procedure [44], the final energy-based minimization procedure can be efficiently implemented by using the parallel abilities of a graphics processor unit (GPU) with a speed gain up to (about) 200 [44].

The overall unsupervised Markovian CD proposed model is outlined in pseudo-code in Algorithm 1. The C++ code running on Linux, data, and all that is necessary for reproduction of the results shown in this paper is freely accessible at http://www.iro.umontreal.ca/∼mignotte/ResearchMaterial

III. Experimental Results

A. Heterogeneous Dataset Description

To validate our approach, we present in this section a series of tests conducted on four real heterogeneous (multimodal) datasets, reflecting different change detection conditions in multimodal case (see Table I); Namely, (#1 and #4) two multisensor optical datasets (*i.e.*, same sensor type but with two different optical sensors or same satellite sensor but with different specifications), (#2-#3) two multisource datasets (*i.e.*, different sensor types), respectively optical/SAR and SAR /optical. This allows us to compare the performance of the proposed method with different state-of-the-art multimodal change detection algorithms recently proposed in this field [6], [16], [23]–[25] in different multimodal CD conditions, and also for a wide variety of changed event when the resolution varies from 0.52 to 30 meters. In this benchmark, all the ground-truth images (change detection mask) was provided by an expert photo interpreter.

B. Results & Evaluation

In all the experimental results, we have considered the simple grey level of the image (and thus converted, when necessary, the optical color image to grayscale), reduce the size of the image such that its maximal size (length or width) is around 500 pixels with a bilinear interpolation (this simplification compared to a more elaborate interpolation or the use of all of the gray levels of the image has no impact on the detection performance results mainly because our MRF modeling uses the distribution law of these gray levels which remains little different in all these cases) and use a double histogram matching.

The internal parameter of our Markovian model are for, from decreasing order of importance, the parameter α of the data likelihood (see Subsection II-C2), the parameter β of the prior model (see Subsection II-B3) and the length N_w of the graph $\mathcal{G}_s$ (see Subsection II-A and Fig. 2) for which the

Algorithm 1 M3CD (Markov Model for Multimodal Change Detection) Algorithm

M3CD Algorithm

Input: Pair of bi-temporal satellite images: (y^{t_1}, y^{t_2})
Output: A binary CD segmentation map: x

α Hard constraint param. of the likelihood model
β Regularization term of the Markov. prior model
N_w Length of the graph $\mathcal{G}_s$
r Cooling rate of the simulated annealing (SA)
$T_{o,f}$ Initial and final Temp. of the SA

$\mathbf{\Phi}^{[k]}_{y_{<s,t>}}$ Parameter vector $(\lambda^{[k]}, \mu^{[k]}, \sigma^{[k]})$ gathering the scale, mean and variance of the likelihood mixture at iteration k
$x^{[k]}$ CD (binary) map at iteration k

1. Initialization Step
∘ Co-Registration of the image pair (y^{t_1}, y^{t_2})
∘ Conversion of (y^{t_1}, y^{t_2}) into grayscale (if necessary)
∘ Image size reduction of (y^{t_1}, y^{t_2}) until the maximum (length or width) side is ≈ 500 pixels
∘ Double histogram matching on (y^{t_1}, y^{t_2})

2. Parameter Estimation Step
▷ Initialization:
. $x^{[0]} \leftarrow$ Rand. sampling from 2 classes (*change* / *no*)
. $\lambda^{\star} \leftarrow \sum y_{<s,t>} / (N_{id} + N_{di})$ (shape param. of $\{y_{<s,t>}\}$)
. $\mathbf{\Phi}^{[0]}_{y_{<s,t>}} \leftarrow (\lambda^{[0]}, \mu^{[0]}, \sigma^{[0]})$
with $\mu^{[0]} = 2\lambda^{[0]}$, $\lambda^{[0]} = \lambda^{\star}$ and $(\sigma^2)^{[0]} = 1000$
. $k \leftarrow 0$

▷ ICE Procedure:
while $\mathbf{\Phi}^{[k+1]}_{y_{<s,t>}} \not\approx \mathbf{\Phi}^{[k]}_{y_{<s,t>}}$ **do**
  for *each pixel at site s (lexicographically)* **do**
    Sample $x_s^{[k]}$ with Posterior dist. $P_{X_s|Y_{<s,t>}}(.)$ based on graph $\mathcal{G}_s$, temperature $T = 0.25$ and $\mathbf{\Phi}^{[k]}_{y_{<s,t>}}$ [see Eq. (11)]
  ∘ $\mathbf{\Phi}^{[k+1]}_{y_{<s,t>}}$ is estimated with the ML estimator of the "complete data" $(y_{<s,t>}, x^{[k]})$ [see Eqs (8-10)]
  ∘ Hard constraint: $\lambda^{[k+1]} = \lambda^{\star}/\alpha$
  . $k \leftarrow k + 1$
$\mathbf{\Phi}_{y_{<s,t>}} \leftarrow \mathbf{\Phi}^{[k+1]}_{y_{<s,t>}}$

3. CD Segmentation Step
▷ Initialization: $k \leftarrow 0$ and $T \leftarrow T_o$

▷ Simulated Annealing Procedure:
while $T > T_f$ **do**
  for *each pixel at site s (lexicographically)* **do**
    Sample $x_s^{[k]}$ according to the Gibbs Posterior distribution $P_{X_s|Y_{<s,t>}}(.)$ based on graph $\mathcal{G}_s$, temperature T and $\mathbf{\Phi}_{y_{<s,t>}}$ [see Eq. (11)]
  . $k \leftarrow k + 1$ and $T \leftarrow T_0 \cdot \mathrm{r}^k$

TABLE I
DESCRIPTION OF THE FOUR HETEROGENEOUS DATASETS

Dataset	Date	Location	Size (pixels)	Event (& Spatial resolution)	Sensor
1	Sept. 1995 - Jul. 1996	Sardinia, It	412×300	Lake overflow (30 m.)	Landsat-5 Thermic (NIR band) / Optical
2	July 2006 - July 2007	Gloucester, UK	2325×4135	Flooding (0.65 m.)	TerraSAR-X / QuickBird 02
3	Feb. 2009 - July 2013	Toulouse, Fr	4404×2604	Construction (2 m.)	TerraSAR-X / Pleiades
4	May 2012 - July 2013	Toulouse, Fr	2000×2000	Construction (0.52 m.)	Pleiades / WorldView 2

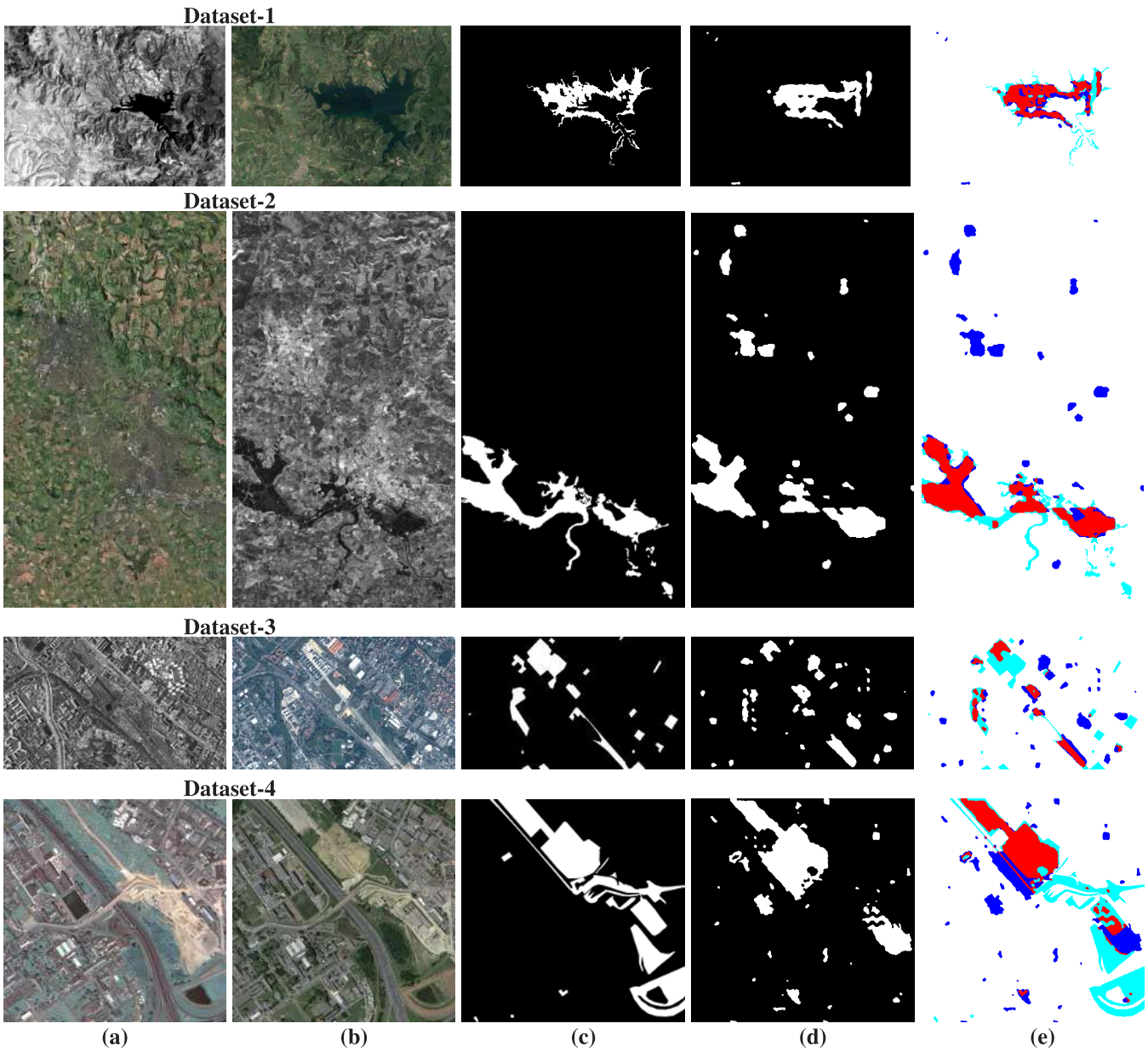

Fig. 4. Heterogeneous datasets (see Table I). (a-c) image t_1, t_2, ground truth; (d) final (changed-unchanged) segmentation result and (e) confusion map (white: TN, red: TP, blue: FP, Cyan: FN) obtained by the proposed approach.

sensitivity is not important. We do not consider the parameter N_T as an important internal parameter; in fact, we have taken $N_T = 16$ in order to use the very fast (since 16 is a power of 2) DCT package implemented in C code by [34].[4] In our application, the DCT is thus applied on the grey-scale band of the image or the gray-level band resulting from the grayscale conversion of the three color bands (for a color image). For all the experimental results, we use $\alpha = 1.5, \beta = 0.1$, $N_{to} = 41$.

In order to discuss and compare obtained results, a quantitative study is realized by computing the classification rate

[4]We have also tested $N_T = 8$ and noticed that the classification results was slightly altered in our application.

TABLE II

CONFUSION MATRIX IN TERMS OF NUMBER OF PIXELS AND PERCENTAGE FOR THE FOUR HETEROGENEOUS DATASETS *i.e.*, [TM/TM], [TSX/QB02], [TERRASAR-X/PLEIADES], [PLEIADES/WORLDVIEW 2] (SEE TABLE I)

Multimodal pair	TP	TN	FP	FN
Thermic/Optical (Landsat-5)	5189 (67.3%)	114007 (98.4%)	1884 (1.6%)	2520 (32.7%)
QB02/TerraSAR-X	5272 (69.3%)	10782 (78.3%)	2990 (21.7%)	2337 (30.7%)
TerraSAR-X/Pleiades	4025 (35.8%)	124468 (95.7%)	5611 (4.3%)	7217 (64.2%)
Pleiades/WorldView2	15904 (41.8%)	199794 (94.3%)	12164 (5.7%)	22138 (58.2%)

TABLE III

ACCURACY RATE OF CHANGE DETECTION ON THE FOUR HETEROGENEOUS DATASETS OBTAINED BY THE PROPOSED METHOD AND THE STATE-OF-THE-ART MULTIMODAL CHANGE DETECTORS (FIRST UPPER PART OF EACH TABLE) AND MONO-MODAL CHANGE DETECTORS (SECOND LOWER PART OF EACH TABLE)

Optical(NIR band)/Optical [#1]	Accuracy
Proposed method	**0.964**
Touati *et al.* [23]	0.942
Zhang *et al.* [10]	0.975
PCC [10]	0.882

SAR/Optical [#2]	Accuracy
Proposed method	**0.955**
Touati *et al.* [6]	0.949
Touati *et al.* [16]	0.932
Touati *et al.* [23]	0.943
Prendes *et al.* [26]	0.844
Correlation [26]	0.670
Mutual Inf. [26]	0.580

Optical/SAR [#3]	Accuracy
Proposed method	**0.909**
Touati *et al.* [6]	0.867
Touati *et al.* [23]	0.878
Prendes *et al.* [27][45]	0.918
Prendes *et al.* [25]	0.854
Copulas [24][25]	0.760
Correlation [24][25]	0.688
Mutual Inf. [24][25]	0.768
Pixel Dif. [25][46]	0.782
Pixel Ratio [25][46]	0.813

Optical/Optical [#4]	Accuracy
Proposed method	**0.862**
Touati *et al.* [6]	0.853
Touati *et al.* [16]	0.870
Touati *et al.* [23]	0.877
Prendes *et al.* [26][27]	0.844
Correlation [26][27]	0.679
Mutual Inf. [26][27]	0.759
Pixel Dif. [27][46]	0.708
Pixel Ratio [27][46]	0.661

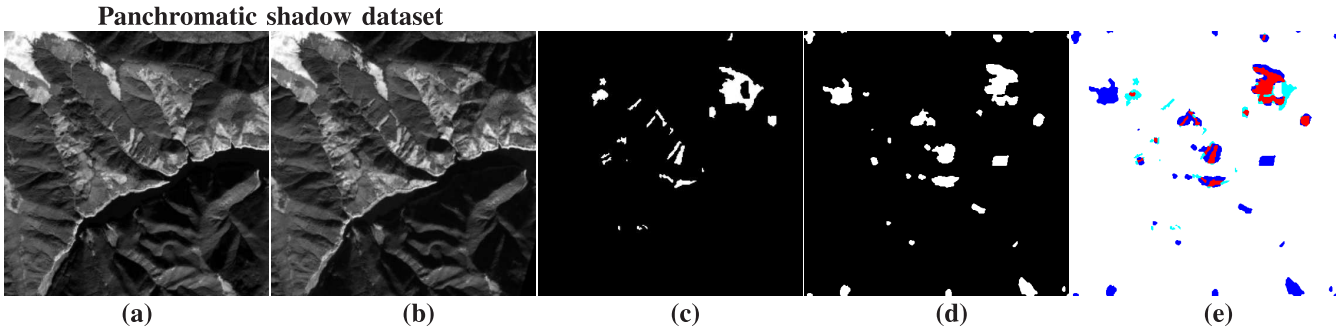

Fig. 5. Panchromatic data set: (a-c) image t_1, t_2, ground truth; (d) final (changed-unchanged) segmentation result and (e) confusion map (white: TN, red: TP, blue: FP, Cyan: FN) obtained by the proposed approach.

accuracy that measures the percentage of the correct changed and unchanged pixels: PCC = (TP+TN)/(TP+TN+FN+FP) where TP, TN, FN, FP designate classically the true positives, negatives, and false negatives and positives.

A comparison with different state of the art approaches [6], [16], [23]–[25] is summarized in Table III. We have also summarized in Table II the confusion matrix obtained by our proposed Markovian CD model. From Table III, we can see that the rate accuracy of our method performs very well and outperforms in average the other state-of-the-art approaches.

The average accuracy rate obtained on the four multimodal dataset based on our Markovian CD approach is 92.3% with well balanced confusion matrices (see Table II).

C. Results on Homogeneous Dataset With Shadow Effects

As an additional experiment, it is also interesting to see how the proposed unsupervised Markovian CD model behaves and adapts in the presence of homogeneous images (see Fig. 5) when one of the two images has glow and shadow effects. To this end, for this (non-trivial) homogeneous CD detection case, we have considered a stereo panchromatic

TABLE IV

KAPPA STATISTIC [47] $(po - pe)/(1 - pe)$ (WITH PO = OBSERVED ACCURACY = $(TP + TN)/(TP + FP + FN + TN)$ AND PE = EXPECTED ACCURACY = $[(TP + FP)(TP + FN) + (FN + TN)(FP + TN)]/[(TP + FP + FN + TN)^2]$) OF CHANGE DETECTION ON THE PANCHROMATIC SHADOW DATASET OBTAINED BY THE PROPOSED METHOD AND COMPARISONS OTHER UNSUPERVISED (FIRST UPPER PART OF THE TABLE) AND SUPERVISED (SECOND PART OF THE TABLE) STATE-OF-THE-ART MONOMODAL CHANGE DETECTORS [47]

Method	Kappa
Proposed method	**0.403**
Touati *et al.* [6] (with preprocessing)	**0.513**
Touati *et al.* [6] (without preprocessing)	0.281
kMNF OPTI* [47]	0.487 - **0.509** - 0.506 - 0.501 - 0.487 - 0.475
Height Difference* [47]	0.127 - 0.316 - 0.469 - **0.526** - 0.0 - 0.0
CVA* [47]	0.07 - 0.242 - 0.403 - **0.457** - 0.0 - 0.0
k-Means [47]	0.472
ICDA [47]	0.495
OSVM [47]	0.478
Random Forests [47]	0.432

*based on different threshold levels given in increasing order

data set provided by [47], with size 900 × 900 pixels (pixel resolution is 5 meters) and captured by the Cartosat-1 satellite sensor. This pair of panchromatic images is acquired over the Arges region (Roumania near Piatra Craiului national park), on Oct. 2008 and Nov. 2009 and shows a forest changes caused by storms, and containing many shadow areas caused by steep terrain due to the mountainous forest area [47]. From Table IV, we can see that the kappa coefficient of our method is correct and quite comparable to others state-of-the-art homogeneous CD approaches, though slightly less good (than the methods purely dedicated and optimized for the monomodal case). In fact, our model remains ideally and best suited for the multimodal CD case with a mixture of distributions specifically chosen to take into account a (quite large) number of pairs of rather different imaging modalities usually observed in remote sensing.

D. Discussion

Concerning the technical specifications of the proposed model, we have noticed that the L1 norm, for the pixel pairwise spatio-temporal difference (used as visual cue) in Eq. (1), is slightly more robust than the L2 norm for which we obtain an average accuracy rate (obtained on the four multimodal datasets) of 89.3% (versus 92.3% for the L1 norm). Besides, it is important to mention that our choice concerning the likelihood distributions was made after a pre-study where we empirically tried different mixtures of statistical laws. More precisely, we have successively tried different law combinations including, for identical pixel-pairwise labels (in addition to the exponential law that was finally used), an half Gaussian, Rayleigh and Gaussian laws along with for different pixel-pairwise labels, (in addition to the Gaussian law that was finally used); a Rayleigh, an exponential and finally an uniform distribution. The best combination was the mixture of Exponential/Gaussian likelihood distributions used in our model and presented in Section II-B.

From the experiment, we can notice that the CD result in multisensor optical Dataset-4 is the least accurate of the four examples given. We think that this can be explained by several reasons. The first one is due to the macro texture generated by the (high-resolution) satellite view of the (Toulouse) urban area. DCT features have more difficulties to model such macro textural patterns and is in fact better suited to model micro-textural features usually present in a lower resolution satellite image (as datasets #1-3). The second reason is due to the nature of change. In this image, the change (*i.e.*, an area under construction) can be subtle and light and thus difficult to distinguish even with a trained eye. Thirdly, the different colors between the two optical images give, after grey-level conversion, different grey levels which may further complicate the CD result.

It is interesting to notice that, in a way, the proposed herein model can be viewed as the Markovian version, thus in the ML sense of the LS model, based on the Multidimensional scaling (MDS) mapping proposed in [6] (however, we herein consider a slightly different observation field including texture information).

Let us also note that, in the ML criterion sense, we try to maximize the posterior probability of a given (pair of observation(s) and consequently this one is thus closely related both to the choice of the observation field (in our case $y_{\langle s,t\rangle}$ and also, above all, the choice of the mixture of distributions (in our case Exponential/Gaussian). We think that more flexible (or generalized) distribution laws would be perhaps more suited to the heterogeneous remote sensing imagery (*i.e.,* thus leading to a better model) but this flexibility would be at the cost of a more complicated (already very complex and computational demanding) final optimization procedure.

IV. CONCLUSION

In this paper, we have addressed the problem of change detection in heterogeneous remote sensing. Although this issue has become important, due to the huge amount of heterogeneous data, we can now get from various archives or from existing (and different types of) Earth observing satellites, it has only received little attention in the literature. In addition, this issue has really been very little discussed in the statistical field and, to our knowledge, no Bayesian or Markovian-based multimodal CD method has been proposed until now. This paper fills the gap by proposing a complete unsupervised Markovian approach which has been validated

on a number of real multimodal bitemporal satellite image pairs and whose the main novelty, and not only in Geoscience imaging, lies in the use of an observation field built up from a pixel pairwise modeling. In fact, in our application, in order to decrease the computational load of our algorithm, we consider for each pixel, a sub-sample of pairs of pixels. Nevertheless, the proposed MRF model turns out to iteratively propagate information *via* this sub-sample of pairs of pixels very efficiently during the estimation or segmentation step, while keeping a quasi-linear complexity with respect to the number of image pixels. We also think that the concept of pixel pairwise modeling can be interesting for other issues in traditional digital image processing, not only in Geoscience imaging, since the underlying framework based on pixel-pairwise affinity can really model complex statistical phenomena with possibly important invariance properties.

APPENDIX

The ICE procedure first requires to find an estimator $\hat{\Phi}_{y_{\langle s,t\rangle}} = \Phi(x_{\langle s,t\rangle}, y_{\langle s,t\rangle})$ providing an estimate of $\Phi_{y_{\langle s,t\rangle}}$ based on the complete data configuration $(x_{\langle s,t\rangle}, y_{\langle s,t\rangle})$. Random field $\mathbf{X}_{\langle s,t\rangle}$ being un-observable, the iterative ICE procedure thus defines the parameter $\Phi^{[k+1]}_{y_{\langle s,t\rangle}}$, at step $[k+1]$, as the conditional expectations of $\hat{\Phi}_{y_{\langle s,t\rangle}}$ given $Y_{\langle s,t\rangle} = y_{\langle s,t\rangle}$ and the current parameter $\Phi^{[k]}_{y_{\langle s,t\rangle}}$. The fixed point of this iteration corresponds to the best approximations of $\Phi_{y_{\langle s,t\rangle}}$ in terms of the mean squared error [29]. By denoting E_k the conditional expectation based on $\Phi^{[k]}_{y_{\langle s,t\rangle}}$, this iterative procedure is defined as follows:

- One takes an initial value $\Phi^{[0]}_{y_{\langle s,t\rangle}}$
- $\Phi^{[k+1]}_{y_{\langle s,t\rangle}}$ is computed from $\Phi^{[k]}_{y_{\langle s,t\rangle}}$ and from $y_{\langle s,t\rangle}$ using:

$$\Phi^{[k+1]}_{y_{\langle s,t\rangle}} = E_k\big[\hat{\Phi}_{y_{\langle s,t\rangle}}(x, y) \,|\, \mathbf{Y}_{\langle s,t\rangle} = y_{\langle s,t\rangle}\big]$$

The computation of this expectation is impossible in practice, but we can approach it thanks to the law of large numbers [29]:

$$\Phi^{[k+1]}_{y_{\langle s,t\rangle}} = \frac{1}{n}\big[\hat{\Phi}_{y_{\langle s,t\rangle}}(x^{(1)}_{\langle s,t\rangle}, y_{\langle s,t\rangle}) + \cdots + \hat{\Phi}_{y_{\langle s,t\rangle}}(x^{(n)}_{\langle s,t\rangle}, y)\big]$$

where $x^{(i)}_{\langle s,t\rangle}, i = 1, \ldots, n$ are realizations drawn from the posterior distribution: $P_{X_{\langle s,t\rangle}|Y_{\langle s,t\rangle},\Phi}(x_{\langle s,t\rangle}|y_{\langle s,t\rangle}, \Phi^{[k]}_{y_{\langle s,t\rangle}})$.

In our application, since x completely defines $x_{\langle s,t\rangle}$ without ambiguity (but the inverse is not true), these realizations can be drawn from the posterior distribution $P_{X|Y_{\langle s,t\rangle},\Phi}(x|y_{\langle s,t\rangle}, \Phi^{[k]}_{y_{\langle s,t\rangle}})$ (see Section II-B3 and Eq. (5)). As it turns out, $n = 1$ is sometimes found sufficient (or even better) to get good estimates [29]. It is the case in our unsupervised Markovian CD model, and we actually chose $n = 1$ in our experiments.

ACKNOWLEDGMENT

The authors would like to acknowledge all other researchers that made at our disposal the change detection dataset in order to validate the proposed change detector.

REFERENCES

[1] N. Longbotham *et al.*, "Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009–2010 data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 5, no. 1, pp. 331–342, Feb. 2012.

[2] A. Schaum, "Local covariance equalization of hyperspectral imagery: Advantages and limitations for target detection," in *Proc. IEEE Aerosp. Conf.*, Mar. 2005, pp. 2001–2011.

[3] M. T. Eismann, J. Meola, and R. C. Hardie, "Hyperspectral change detection in the presenceof diurnal and seasonal variations," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 237–249, Jan. 2008.

[4] J. Zhou, C. Kwan, B. Ayhan, and M. T. Eismann, "A novel cluster kernel RX algorithm for anomaly and change detection using hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6497–6504, Nov. 2016.

[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] R. Touati and M. Mignotte, "An energy-based model encoding nonlocal pairwise pixel interactions for multisensor change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 1046–1058, Feb. 2018.

[7] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.

[8] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. L. Rojo-Alvarez, and M. Martinez-Ramon, "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1822–1835, Jun. 2008.

[9] P. Du, S. Liu, J. Xia, and Y. Zhao, "Information fusion techniques for change detection from multi-temporal remote sensing images," *Inf. Fusion*, vol. 14, no. 1, pp. 19–27, Jan. 2013.

[10] P. Zhang, M. Gong, L. Su, J. Liu, and Z. Li, "Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images," *Photogram. Remote Sens.*, vol. 116, pp. 24–41, Sep. 2016.

[11] N. Merkle, P. Fischer, S. Auer, and R. Müller, "On the possibility of conditional adversarial networks for multi-sensor image matching," in *Proc. IGARSS*, Fort Worth, TX, USA, Jul. 2017, pp. 2633–2636.

[12] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 545–559, Mar. 2018.

[13] Z. Liu, G. Li, G. Mercier, Y. He, and Q. Pan, "Change detection in heterogenous remote sensing images via homogeneous pixel transformation," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1822–1834, Apr. 2018.

[14] V. Alberga, "Similarity measures of remotely sensed multi-sensor images for change detection applications," *Remote Sens.*, vol. 1, no. 3, pp. 122–143, Jul. 2009.

[15] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using VHR optical and SAR imagery," *IEEE Trans. Geosc. Remote Sens.*, vol. 48, no. 5, pp. 2403–2420, May 2010.

[16] R. Touati, M. Mignotte, and M. Dahmane, "A new change detector in heterogeneous remote sensing imagery," in *Proc. 7th IEEE Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Montreal, QC, Canada, Dec. 2017, pp. 1–6.

[17] M. Xu *et al.*, "Change detection of the Tangjiashan barrier lake based on multi-source remote sensing data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 4, Jul. 2009, pp. IV-303–IV-306.

[18] M. Volpi, F. de Morsier, G. Camps-Valls, M. Kanevski, and D. Tuia, "Multi-sensor change detection based on nonlinear canonical correlations," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2013, pp. 1944–1947.

[19] 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.

[20] Z.-G. Liu, G. Mercier, J. Dezert, and Q. Pan, "Change detection in heterogeneous remote sensing images based on multidimensional evidential reasoning," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 168–172, Jan. 2014.

[21] X. Chen, J. Li, Y. Zhang, and L. Tao, "Change detection with multi-source defective remote sensing images based on evidential fusion," *ISPRS Ann. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 3, no. 7, pp. 125–132, Jun. 2016.

[22] L. T. Luppino *et al.*, "A clustering approach to heterogeneous change detection," in *Proc. Scand. Conf. Image Anal.*, Jun. 2017, pp. 181–192.
[23] R. Touati, M. Mignotte, and M. Dahmane, "Change detection in heterogeneous remote sensing images based on an imaging modality-invariant mds representation," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Athens, Greece, Oct. 2018, pp. 3998–4002.
[24] 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.
[25] 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.
[26] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J. Tourneret, *Performance Assessment of a Recent Change Detection Method for Homogeneous and Heterogeneous Images*, vol. 209. France: Société Française de Photogrammétrie et de Télédétection, 2015, pp. 23–29.
[27] J. Prendes, "New statistical modeling of multi-sensor images with application to change detection," Ph.D. dissertation, TeSA Lab., Paris-Saclay Univ., Toulouse, France, 2015.
[28] 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.
[29] F. Salzenstein and W. Pieczynski, "Parameter estimation in hidden fuzzy Markov random fields and image segmentation," *Graph. Model Image Process.*, vol. 59, no. 4, pp. 205–220, Jul. 1997.
[30] W. Pieczynski, "Convergence of the iterative conditional estimation and application to mixture proportion identification," in *Proc. IEEE/SP 14th Workshop Stat. Signal Process.*, Aug. 2007, pp. 49–53.
[31] M. Mignotte, C. Collet, P. Perez, and P. Bouthemy, "Sonar image segmentation using an unsupervised hierarchical MRF model," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1216–1231, Jul. 2000.
[32] M. Mignotte, J. Meunier, and J.-P. Soucy, "DCT-based complexity regularization for EM tomographic reconstruction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 2, pp. 801–805, Feb. 2008.
[33] M. Mignotte, "Fusion of regularization terms for image restoration," *Proc. SPIE*, vol. 19, no. 3, Jul. 2010, Art. no. 033004.
[34] T. Ooura. *General Purpose FFT (Fast Fourier/Cosine/Sine Transform) Package*. Accessed: 2001. [Online]. Available: http://momonga.t.u-tokyo.ac.jp/~ooura/fft.html
[35] P. Perez, A. Blake, and M. Gangnet, "JetStream: Probabilistic contour extraction with particles," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, Vancouver, BC, Canada, vol. 2, Jul. 2001, pp. 524–531.
[36] F. Destrempes and M. Mignotte, "A statistical model for contours in images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 626–638, May 2004.
[37] F. Destrempes, M. Mignotte, and J. F. Angers, "Localization of shapes using statistical models and stochastic optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, pp. 1603–1615, Sep. 2007.
[38] N. Widynski and M. Mignotte, "A multiscale particle filter framework for contour detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 1922–1935, Oct. 2014.
[39] M. Mignotte, "An energy-based model for the image edge-histogram specification problem," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 379–386, Jan. 2012.
[40] M. Mignotte, "Non-local pairwise energy based model for the hdr image compression problem," *J. Electron. Imag.*, vol. 21, no. 1, pp. 1–12, Jan. 2012.
[41] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc.*, vol. 48, no. 3, pp. 259–279, Jul. 1986.
[42] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.*, vol. 39, no. 1, pp. 1–22, Sep. 1977.
[43] P. Masson and W. Pieczynski, "SEM algorithm and unsupervised statistical segmentation of satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 3, pp. 618–633, May 1993.
[44] P.-M. Jodoin and M. Mignotte, "Markovian segmentation and parameter estimation on graphics hardware," *J. Electron. Imag.*, vol. 15, no. 3, 2006, Art. no. 033005.
[45] 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. Acoust., Speech Signal Process. (ICASSP)*, Brisbane, QLD, Australia, Apr. 2015, pp. 1513–1517.
[46] O. D. Team, (2014). *The Orfeo Toolbox Software Guide*. [Online]. Available: http://orfeo-toolbox.org/
[47] 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.

Redha Touati received the M.Sc. degree in computer science from the Department of Computer Science and Operations Research (DIRO), University of Montreal, Montreal, QC, Canada, in 2014. He is currently pursuing the Ph.D. degree with the Vision Laboratory, DIRO, University of Montreal, in collaboration with the Imaging and Vision Department, Computer Research Institute of Montreal (CRIM). His research interests include statistical methods and applied mathematics in video imaging and remote sensing imagery.

Max Mignotte received the DEA (Postgraduate degree) degree in digital signal, image, and speech processing from INPG University, France, Grenoble, in 1993, and the Ph.D. degree in electronics and computer engineering from the University of Bretagne Occidentale (UBO) and the Digital Signal Laboratory (GTS), French Naval Academy, France, in 1998. He was an INRIA Postdoctoral Fellow with the University of Montreal (DIRO), QC, Canada, from 1998 to 1999. He is currently a Professor with the Computer Vision & Geometric Modeling Laboratory, DIRO, University of Montreal. He is also a member of the Laboratoire de recherche en imagerie et orthopedie (LIO), Centre de recherche du CHUM, Hopital Notre-Dame, and a Researcher with CHUM. His current research interests include statistical methods, Bayesian inference, and energy-based models for solving diverse large-scale high-dimensional ill-posed inverse problems in imaging.

Mohamed Dahmane received the master's degree in satellite imagery and the master's degree in video surveillance from the Université de Montréal. In 2012, he joined CRIM as an expert in the development of algorithms for facial expression recognition, the main subject of his doctoral thesis. He is currently a Researcher with CRIM, in 2012. He is also an Associate Professor with the École de Technologie Supérieure (ÉTS), Université du Québec.