Unsupervised Bayesian change detection for remotely sensed images

Walma Gharbi, Lotfi Chaari, Amel Benazza-Benyahia

Unsupervised Bayesian change detection for remotely sensed images

Walma Gharbi[1,2] · Lotfi Chaari[3] · Amel Benazza-Benyahia[1]

Abstract
The availability of remote sensing images with high spectral, spatial and temporal resolutions has motivated the design of new change detection (CD) methods for surveying changes in a studied area. The challenge of unsupervised CD is to develop flexible automatic models to estimate changes. In this paper, we propose a novel hierarchical Bayesian model for CD. Our main contribution lies in the application of Bernoulli-based models to change detection and transforming it to a denoising problem. The originality is related to the capacity of these models to act as implicit classifiers in addition to the denoising effect since even for changed pixels noise is also removed. The second originality lies in the way inference is conducted. Specifically, the hierarchical Bayesian model and Gibbs sampler ensure building an algorithm with secure convergence guarantees. Experiments performed on real data indicate the benefit that can be drawn from our approach.

Keywords Change detection · Multispectral satellites images · Remote sensing · Bayesian methods · MCMC

1 Introduction

The Earth surface is frequently submitted to numerous variations caused both by natural phenomena and human activities. Studying these variations for monitoring and risk assessing is of paramount importance in many application fields of remote sensing (RS) such as agriculture for land-cover or land-use [1,21]. This interest has motivated the design of change detection (CD) methods, namely methods that aim at identifying differences in the state of an object or a phenomenon by observing it at different times [1]. For the past few decades, a new generation of satellite sensors has enabled the acquisition of very high resolution (VHR) optical images such as Quickbird, GeoEye and IKONOS. Their spatial resolution is very often less than one meter in the panchromatic band and few meters in the spectral bands (approximately less than 5 m) which allows observing more subtle changes within "entities" (e.g., a forest parcel). However, for VHR images, the resort to CD methods developed for low or medium resolution images is less performing since they are tailored to only detect abrupt changes and not subtle ones [11]. Hence, the need to design efficient CD methods specifically dedicated to VHR images is clearly motivated for many applications such as natural disasters management and ecosystem [11] and forest monitoring [9]. In this paper, we propose a fully automatic hierarchical Bayesian method for multispectral CD which is handled as a denoising problem. The proposed hierarchical Bayesian model involves a mixture prior that allows to recover zero coefficients using a Bernoulli distribution and promotes the sparsity of the observed image using a Laplace one. The result is a clean version of the DI degraded by noise or any distortion where changed and unchanged pixels are separated. The remainder of this paper is organized as follows. Section 2 is dedicated to a brief review of CD in multispectral images. Section 3 introduces the proposed model and inference scheme. The proposed model is then validated using VHR images in Sect. 4. Conclusions and perspectives are outlined in Sect. 5.

✉ Walma Gharbi
gharbi.walma@gmail.com

Lotfi Chaari
lotfi.chaari@toulouse-inp.fr

Amel Benazza-Benyahia
benazza.amel@supcom.rnu.tn

[1] SUP'COM, LR11TIC04, COSIM Lab, University of Carthage, Tunis, Tunisia

[2] Digital Research Centre of Sfax, University of Sfax, Sfax, Tunisia

[3] IRIT-ENSEEIHT, University of Toulouse, Toulouse, France

2 A brief review multispectral change detection

CD methods could be divided into supervised or unsupervised, depending on the availability of labeled data [19]. Supervised CD methods have been employed to detect urban changes and land cover [30]. The work of [25,26] performed a multitemporal analysis using the support vector machine-based labeled graph to preserve the semantic information about changes (progressive or abrupt) and obtain temporal cartography of a studied region. Although supervised methods provide cover transition, prior knowledge using expert manual processing is required which can be time and effort consuming. In contrast, unsupervised CD methods do not require any expert action nor reference samples which importantly reduces the operational cost. As the detection is automatic, they are attractive from a practical point of view [3,6,18,20]. In this work, we focus on unsupervised CD for VHR images that allow to detect the presence of changes [6] or several kinds of change through a change vector analysis (CVA) [4,18]. The CD methods are mainly based on the comparison result of bi-temporal images which are either the difference image (DI) [28], its vectorized version CVA [1,4,6] or to the image ratio [23]. Analyzing RS data was initially carried out using empirical strategies with a manual trial error procedure to separate changes. This analysis can be carried out either in the original spectral domain, such as methods that were mentioned here-above, or in a transform domain. Principal component analysis aims at selecting the first few principal components which are considered relevant to change by achieving an overall percentage coverage [2]. However, this technique requires prior knowledge to select the most informative components which can be time-consuming and lead to an increase of the missed alarms rate due to unselected components. Generally, transformation-based techniques focus on specific kinds of change mainly related to the studied context but are not able to detect all changes related to a given area. To handle all variations (especially the subtle ones), more flexible models reflecting the multiple properties of the observed image are needed. Markov random fields have also been exploited due to their ability to integrate the spatial and temporal context information based on DI [22]. In [22], the authors have combined object level information with random field-based methods for binary CD. Object-based methods were proposed to handle the problem of radiometric differences and misregistration of bitemporal images in pixel-based methods [10,31]. However, they suffer from over or under-segmentation since different segmentation scales are required for different objects. This can highly affect the CD performance. In [14], the authors used the masks of clouds and shadows to improve the CD performance by removing pixels related to clouds and shadows [15]. Also, the Bayesian framework has been used to analyze the DI and produce binary CD maps. In this respect, the threshold could be estimated according to the Maximum A Posteriori (MAP) rule [5,6,20]. In [6], a Gaussian mixture model (GMM) of the two classes (change and no change) is considered for the threshold selection. A generalized GMM is retained in [20] to fit the conditional densities of the DI classes while assuming that these classes can follow a non-Gaussian distribution. In [5], the threshold is obtained by including all the spectral channels but with a high false alarm rate due to noisy components. Since the rate of false alarms is a highly important evaluation criterion of the CD methods effectiveness, the threshold has been set to minimize the number of false alarms in [17]. Even if most methods described here-above use the DI to extract changed pixels, the use of statistical features extracted from images is also possible for CD. In [13], an empirical Bayesian approach is designed with a false discovery rate (FDR) formulation. The aim is to control the false alarm rate using statistical features that have a given distribution under a no-change hypothesis while assuming that the change represents at most 25% of the image total size. Hence, the challenge is to define models that are flexible, robust to noise and suitable for analyzing VHR images for an automatic CD.

3 Proposed Bayesian approach for CD

3.1 Problem formulation

Let us consider two co-registered multispectral images $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ in $\mathbb{R}^{M\times N\times B}$, acquired over the same geographical area at different dates t_1 and t_2, where B is the number of the spectral channels and $M \times N$ is the size of the b^{th} spectral component. Let $\boldsymbol{x}_1^b$ and $\boldsymbol{x}_2^b$ denote the vectorized version of the b^{th} channels of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, respectively. We define the difference vector $\boldsymbol{y}^b$ for the spectral channel b as

$$\boldsymbol{y}^b = \boldsymbol{x}_2^b - \boldsymbol{x}_1^b. \tag{1}$$

We propose to detect changed pixels from all the $\boldsymbol{y}^b$ ($b = 1, \ldots, B$) by formulating the CD as a denoising problem. Indeed, the observation $\boldsymbol{y}^b$ in a channel b is a noisy version of the clean DI $\boldsymbol{d}^b$. Thus, pixels of $\boldsymbol{d}^b$ corresponding to unchanged areas of the image should have a reflectance value close to zero, whereas the remaining ones have larger dynamic. The observation model in channel b reads to:

$$\boldsymbol{y}^b = \boldsymbol{d}^b + \boldsymbol{n}^b \tag{2}$$

where $\boldsymbol{y}^b$ and $\boldsymbol{d}^b$ are the noisy and clean difference vectors of dimension MN, respectively, and $\boldsymbol{n}^b \in \mathbb{R}^{MN}$ is an additive i.i.d. Gaussian noise vector with a covariance

matrix $\sigma_n^2 \boldsymbol{I}$. The CD problem amounts to derive an estimated version $\widehat{\boldsymbol{d}}^b$ of $\boldsymbol{d}^b$ from the observation $\boldsymbol{y}^b$. To this end, especially for ill-posed problems, two main competing regularization strategies are usually used by the signal and image processing community: variational and Bayesian techniques [7,24]. On the one hand, variational methods rely on the use of appropriate penalizations to formulate a cost function that has to be optimized with respect to the target variable. These techniques are known to be fast and efficient, especially for high-dimensional problems [32]. On the other hand, Bayesian techniques assume that all the target quantities are realizations of random variables. In fact, Bayesian models are known for their flexibility which allows them to handle complex problems as CD. In this paper, we adopt a Bayesian strategy (namely a MAP approach) in order to design a fully automatic CD algorithm that do not require any prior configuration. More precisely, we propose a hierarchical Bayesian model which consists in developing a statistical model in multiple levels with sub-models (hierarchical form) in order to estimate the parameters of the posterior distribution using the Bayesian method. In fact, the prior distribution of some of the model parameters depends on other parameters, which are also assigned to a prior. The advantage is, by combining these sub-models with the observed data, we are able to account for all the uncertainty that is present.

3.2 Hierarchical Bayesian model

We assume that $\boldsymbol{y}^b$ and $\boldsymbol{d}^b$ are realizations of random vectors $\boldsymbol{Y}^b$ and $\boldsymbol{D}^b$. In the following, we detail the retained likelihood and the different prior distributions.

3.2.1 Likelihood

Under the assumption of additive Gaussian noise of variance σ_n^2, the likelihood writes as

$$f\left(\boldsymbol{y}^b|\boldsymbol{d}^b, \sigma_n^2\right) = \left(\frac{1}{2\pi\sigma_n^2}\right)^{\frac{M\times N}{2}} \exp\left(-\frac{\|\boldsymbol{y}^b - \boldsymbol{d}^b\|^2}{2\sigma_n^2}\right) \quad (3)$$

where $\|\cdot\|$ denotes the Euclidean norm.

3.2.2 Priors

In our model, the unknown parameter vector to be estimated is denoted by $\theta = \left\{\boldsymbol{d}^b, \sigma_n^2\right\}$. In what follows, we introduce the prior distributions to be used for these two parameters.

Prior for $\boldsymbol{d}^b$
Since the $\boldsymbol{d}^b$ are either zero (no changed pixel) or with significant value (changed pixel), a mixture model is adopted as a prior for $\boldsymbol{d}^b$. Specifically, we model the zero coefficients by a Bernoulli distribution, while the nonzero ones by a Laplace distribution. This mixture prior promotes the sparsity of the target signal $\boldsymbol{d}^b$ in the original space and writes

$$f\left(d_i^b|\omega, \lambda\right) = (1-\omega)\delta\left(d_i^b\right) + \frac{\omega}{2\lambda}\exp\left(-\frac{|d_i^b|}{\lambda}\right) \quad (4)$$

where d_i^b is the ith component of $\boldsymbol{d}^b$ of the pixel i in the DI of the spectral channel b, ω is a weight belonging to $[0, 1]$ that reflects the rate of nonzero coefficients and $\lambda \geq 0$ is the hyperparameter of the Laplace distribution. Note that $\delta(.)$ refers to the Dirac delta function. The use of this prior reinforces the ability to separate changed pixels from unchanged ones. The latter are captured by the Bernoulli part of the prior. Assuming the independence between pixels, the prior distribution of the DI vector $\boldsymbol{d}^b$ can be expressed as

$$f\left(\boldsymbol{d}^b|\omega, \lambda\right) = \prod_{i=1}^{M\times N} f\left(d_i^b|\omega, \lambda\right). \quad (5)$$

Prior for σ_n^2
Since σ_n^2 is a real positive scalar, we use an inverse gamma (IG) prior given by

$$f\left(\sigma_n^2|\alpha, \beta\right) = \frac{\beta^\alpha}{\Gamma(\alpha)}\sigma_n^{-2(\alpha+1)}\exp\left(-\frac{\beta}{\sigma_n^2}\right) \quad (6)$$

where $\Gamma(.)$ is the standard gamma function and the positive reals α and β refer to the shape and scale hyperparameters, respectively. This is a common choice for σ_n^2 since the inverse gamma distribution is the conjugate distribution of the normal distribution [29]. A non-informative prior distribution is chosen to reflect uncertainty when no or little prior information on σ_n^2 is available. The hyperparameters α and β can be either manually fixed or estimated (e.g., according to the maximum likelihood criterion).

3.2.3 Hyperparameter priors

This work defines an Hierarchical Bayesian model that allows to estimate the hyperparameters from the observed data. We denote by $\Phi = \{\lambda, \omega\}$, the hyperparameter vector associated with the aforementioned priors.

Hyperprior for ω
Since ω is a weight reflecting the rate of nonzero coefficients (sparsity factor), we use a uniform distribution on $[0, 1]$ if no further information about ω is available. However, it is worth noting that a more informative version could be used if further information on the rate of nonzero coefficients is available.

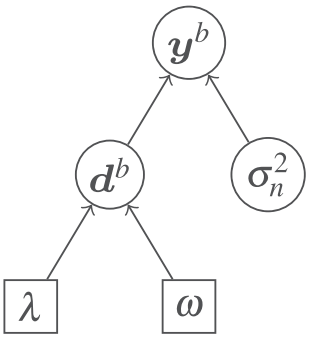

Fig. 1 Direct acyclic graph of the proposed hierarchical Bayesian model

Hyperprior for λ

We adopt an non-informative prior for λ by assuming $\lambda \sim IG(\lambda|a, b)$ as $\lambda \in [0, +\infty[$. The hyperparameters a and b are set to 10^{-3}. Assuming that the hyperparameters λ and ω are statistically independent, the full prior distribution for $\Phi = \{\lambda, \omega\}$ can be written as $f(\Phi|a, b) = f(\lambda|a, b) f(\omega)$. Figure 1 illustrates the hierarchical structure deployed using a direct acyclic graph where the hyperparameters are indicated by squares and the parameters by circles.

3.3 Bayesian inference scheme

Our goal is to derive an estimation of the unknown parameter vector θ as well as the hyperparameter vector Φ according to the MAP criterion by combining the adopted likelihood and prior models. It is straightforward to show that thanks to the Bayes rule the joint posterior distribution is expressed as

$$f\left(\theta, \Phi|\boldsymbol{y}^b, \alpha, \beta, a, b\right) \propto f\left(\boldsymbol{y}^b|\theta\right) f(\theta|\Phi) f(\Phi|a, b). \quad (7)$$

Figure 2 describes the inputs and outputs of the MAP estimator. Based on the joint distribution in Eq. (7), the principle is to derive the conditional distribution for each unknown parameter (in θ and Φ) by integrating with respect to the other variables. These conditional distributions are used to derive estimators, either by performing analytic calculations or by resorting to numerical simulation techniques [27].

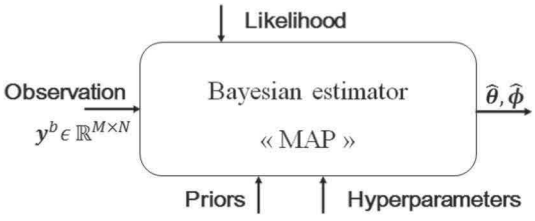

Fig. 2 Scheme of the designed Bayesian MAP estimator

$$\begin{aligned} f\left(\theta, \Phi|\boldsymbol{y}^b, \alpha, \beta, a, b\right) \propto & \left(\frac{1}{2\pi\sigma_n^2}\right)^{\frac{M\times N}{2}} \exp\left(-\frac{\|\boldsymbol{y}^b - \boldsymbol{d}^b\|^2}{2\sigma_n^2}\right) \\ & \times \prod_{i=1}^{M\ N}\left[(1-\omega)\delta(d_i^b) + \frac{\omega}{2\lambda}\exp\left(-\frac{|d_i^b|}{\lambda}\right)\right] \times \mathscr{U}_{[0,1]} \\ & \times \frac{\beta^\alpha}{\Gamma(\alpha)}\sigma_n^{-2(\alpha+1)}\exp\left(-\frac{\beta}{\sigma_n^2}\right) \times \frac{b^a}{\Gamma(a)}\lambda^{-a-1}\exp\left(-\frac{b}{\lambda}\right). \end{aligned} \quad (8)$$

Due to the complexity of the posterior distribution in Eq. (8), it is difficult to analytically derive a simple closed-form expression of the estimators related to θ and Φ. Thus, resorting to Markov Chain Monte Carlo (MCMC) sampling techniques is a common strategy to obtain a numerical approximation of the target posterior distribution [27]. To this respect, a Gibbs sampler (GS) [27] is designed to sample from the joint posterior distribution. More precisely, in an iterative manner, Markov Chain samples are generated for each variable by sampling from its conditional distribution with the remaining variables fixed to their current values until convergence. Generally, the convergence is attained when the Markov chain reaches a steady state that asymptotically characterizes the posterior distribution. After applying the GS algorithm, the obtained samples will not all be considered. For this purpose, the burn-in period is of great significance, since the samples generated during this period have to be discarded because they are not yet asymptotically sampled according to the target distribution. Consequently, the rest of the samples is used to calculate the target estimators. The iterative procedure is used to generate samples, at each iteration, according to the following conditional posteriors: $f(\boldsymbol{d}^b|\boldsymbol{y}^b, \omega, \lambda, \sigma_n^2)$, $f(\sigma_n^2|\boldsymbol{y}^b, \boldsymbol{d}^b, \alpha, \beta)$, $f(\lambda|\boldsymbol{d}^b, a, b)$ and $f(\omega|\boldsymbol{d}^b)$. The main steps of the proposed sampling algorithm are summarized in Algorithm 1 where S indicates the number of iterations fixed properly, to be large enough, in order to ensure reaching the convergence.

Algorithm 1: Gibbs Sampler (GS).

- Initialize with some $\boldsymbol{d}^{b(0)}$.
for $s = 1 \ldots S$ **do**
  Sample σ_n^2 according to Eq.(10);
  Sample λ according to Eq.(11);
  Sample ω according to Eq.(12);
  for $i = 1$ *to* $M \times N$ **do**
    Sample d_i according to Eq.(13);
  end
end

The value of S was empirically set to 200. The MAP is the retained estimation criterion to ensure the selection of the most probable value sampled to approximate d_i^b. For the proposed model, this estimator operates in two steps. First,

we determine if d_i^b is equal to zero or not. If it belongs to the nonzero class of coefficients, the approximated value of d_i^b is computed. The hierarchical MAP allows to properly recover zero coefficients promoted by the Bernoulli distribution, which enforces the sparsity of the target image. We estimate the variance of noise and the hyperparameters according to the minimum mean square error principle. Hence, the estimators $\widehat{\sigma_n^2}$, $\widehat{\lambda}$ and $\widehat{\omega}$ are obtained by minimizing an expected quadratic error. More specifically, we recall that the minimum mean square error estimation is based on a quadratic loss function. For example, a quadratic loss function associated to an estimator, $\widehat{\omega}$, can be formally denoted by $L(\widehat{\omega}, \omega) = (\widehat{\omega}-\omega)^2$, where $L(\widehat{\omega}, \omega) = 0$ if $\widehat{\omega}$ is the exact estimate of ω. Thus, the estimators $\widehat{\sigma_n^2}$, $\widehat{\lambda}$ and $\widehat{\omega}$ can be computed by averaging the samples obtained after discarding those generated during the burn-in period. The conditional distributions used in the GS (namely Algorithm 1) are detailed in the following.

3.3.1 Sampling according to $f(\sigma_n^2|\boldsymbol{y}^b, d^b, \boldsymbol{\alpha}, \boldsymbol{\beta})$

It consists of drawing σ_n^2 according to its conditional distribution which can be expressed as

$$f\left(\sigma_n^2|\boldsymbol{d}^b, \boldsymbol{y}^b, \alpha, \beta\right) \sim f\left(\boldsymbol{y}^b|\boldsymbol{d}^b, \sigma_n^2\right) f\left(\sigma_n^2|\alpha, \beta\right). \tag{9}$$

After integration of Eq. (7) with respect to all the other variables, straightforward calculations lead to the following expression of the conditional distribution for σ_n^2

$$\sigma_n^2|\boldsymbol{d}^b, \boldsymbol{y}^b, \alpha, \beta \sim \texttt{IG}\left(\alpha + \frac{M \times N}{2}, \beta + \frac{\|\boldsymbol{y}^b - \boldsymbol{d}^b\|^2}{2}\right). \tag{10}$$

3.3.2 Sampling according to $f(\boldsymbol{\lambda}|d^b, \boldsymbol{a}, \boldsymbol{b})$

Calculations following the same principle as in Sect. 3.3.1 lead to the following expression

$$\lambda|\boldsymbol{d}^b, a, b \sim \texttt{IG}\left(a + \|\boldsymbol{d}^b\|_0, b + \|\boldsymbol{d}^b\|_1\right) \tag{11}$$

where the number of nonzero coefficients is computed using the l_0 pseudo-norm denoted as $\|.\|_0$, while $\|.\|_1$ denotes the l_1 norm defined as $\|\boldsymbol{d}^b\|_1 = \sum_{i=1}^{M\times N} |d_i^b|$.

3.3.3 Sampling according to $f(\boldsymbol{\omega}|d^b)$

Straightforward calculations similar to that [12] show that the posterior of ω is a Beta distribution

$$\omega|\boldsymbol{d}^b \sim \mathscr{B}\left(1 + \|\boldsymbol{d}^b\|_0, 1 + M \times N - \|\boldsymbol{d}^b\|_0\right), \tag{12}$$

according to which it is easy to sample.

3.3.4 Sampling according to $f(d^b|\boldsymbol{y}^b, \boldsymbol{\omega}, \boldsymbol{\lambda}, \boldsymbol{\sigma_n^2})$

Since the signal coefficients are assumed to be *a priori* independent, by doing similarly to [8], we can derive the following expression of the conditional distribution

$$\begin{aligned} f\left(d_i^b|\boldsymbol{y}^b, \omega, \lambda, \sigma_n^2\right) = \omega_{1,i}\delta\left(d_i^b\right) + \omega_{2,i}\mathscr{N}^+\left(\mu_{i+}, \sigma_n^2\right) \\ + \omega_{3,i}\mathscr{N}^-\left(\mu_{i-}, \sigma_n^2\right) \end{aligned} \tag{13}$$

where $\mathscr{N}^+$ and $\mathscr{N}^-$ denote the truncated Gaussian distributions on $\mathbb{R}^+$ and $\mathbb{R}^-$, respectively. The weights $(\omega_{l,i})_{1\leq l\leq 3}$ in Eq. (13) can be expressed as $\omega_{l,i} = \frac{\mu_{l,i}}{\sum_{l=1}^{3}\mu_{l,i}}$ where

$$\mu_{1,i} = 1 - \omega, \tag{14}$$

$$\mu_{2,i} = \frac{\omega}{2\lambda}\exp\left(\frac{\mu_{i+}^2}{2\sigma_n^2}\right)\sqrt{2\pi\sigma_n^2}C(\mu_{i+}, \sigma_n^2), \tag{15}$$

$$\mu_{3,i} = \frac{\omega}{2\lambda}\exp\left(\frac{\mu_{i-}^2}{2\sigma_n^2}\right)\sqrt{2\pi\sigma_n^2}C(-\mu_{i-}, \sigma_n^2), \tag{16}$$

$$\mu_{i+} = \left(y_i - \frac{\sigma_n^2}{\lambda}\right), \tag{17}$$

$$\mu_{i-} = \left(y_i + \frac{\sigma_n^2}{\lambda}\right), C(\mu, \sigma^2) = \sqrt{\frac{\sigma^2\pi}{2}}\left(1 + \text{erf}\left(\frac{\mu}{\sqrt{2\sigma^2}}\right)\right). \tag{18}$$

To sample from Eq. (13), a two-step procedure is adopted. First of all, a discrete variable of choice γ_i is introduced which will be equal to 0 if the candidate coefficient has to be sampled according to the Dirac function (equal to zero). If $\gamma_i = 1$, the candidate coefficient belongs to the nonzero class of coefficients. At this point, an additional discrete variable κ_i has to be introduced to determine, based on the weights $\omega_{2,i}$ and $\omega_{3,i}$, whether the candidate coefficient have to be sampled according to $\mathscr{N}^+(\mu_{i+}, \sigma_n^2)$ or $\mathscr{N}^-(\mu_{i-}, \sigma_n^2)$.

4 Experiments and analysis

4.1 Experimental description

In this section, we evaluate the performance of the proposed CD model via three experiments. The first experiment is conducted on synthetic data, whereas the second and the third experiments handle real VHR images with different spatial resolutions. Both experiments are performed with MATLAB R2016a using a laptop with processor Intel(R) Core(TM) i5-4200U CPU @ 1.60 GHz and RAM 6 GB on 64 bits Ubuntu operating system. The proposed model is compared

to the work of [18] where the authors handled the CD problem as a three-step procedure. We only focus here on the first step, also called a pseudobinary CD since it results in changed, unchanged and uncertain pixels (considered at the end according to an uncertainty interval). This step is conducted using a thresholding technique detailed in [6]. This step aims to separate changed pixels from unchanged ones by automatically estimating the threshold on the magnitude of the DI and using an Expectation-Maximization algorithm.

4.2 Experiment 1: synthetic dataset

In this round of experiments, we perform a CD on two multispectral synthetic images ($B = 4$). Based on a realistic image $\boldsymbol{X}_1$ of size 100×100, a simulated image $\boldsymbol{X}_2$ is artificially generated by adding changes affecting two major areas in $\boldsymbol{X}_1$ figuring as an additional building structure and water surface appearance. Next, a DI $\boldsymbol{D}$ has been computed and disturbed by a Gaussian noise with variance σ_n^2 according to the observation model introduced in Eq. (2). The images $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are displayed in Fig. 3a, b, respectively. The generated DI $\boldsymbol{D}$ according to the observation model in Eq. (2) is also depicted in Fig. 3c. The proposed model is applied to estimate changes for each DI $\boldsymbol{d}^b$ where $b \in \{1..4\}$. The estimated binary Change Map (CM) is obtained by taking into consideration all the spectral bands. The GS has been run with 200 iterations (981 seconds), including 100 burn-in iterations. Changed areas are outlined by yellow squares in Fig. 3b. The CMs of the proposed method and the thresholding technique in S. Liu et al. [18] are illustrated in Fig. 4b, c, respectively. Figure 4a presents the ground truth. These binary CMs clearly indicate that the proposed model provides a more accurate estimation of change especially since the estimation of unchanged pixels was not affected by noise, which is due to the used prior. Although the thresholding technique adopted by S. Liu et al. succeeded to detect the main changing pixels, a severe salt and pepper noise is present. This problem is caused by only considering the spectral change magnitude of the DI. An objective evaluation in terms of overall accuracy, precision, average accuracy, recall, intersection over union, missed and false alarms is conducted. Table 1 presents the values of the retained accuracy measures obtained over the synthetic dataset. Based on the quantitative results, it is worth noting that the proposed approach outperforms the one of S. Liu et al. Both the precision and IoU values of the proposed method largely exceed the ones of one of S. Liu et al. The choice of a global threshold over the magnitude DI can cause high rates of both false alarms and missed alarms. This is due to the spectral variability in VHR images, i.e., the same object can have different spectral characteristics. Thus, it is not evident that pixels with high magnitude are changed and those with low values are not.

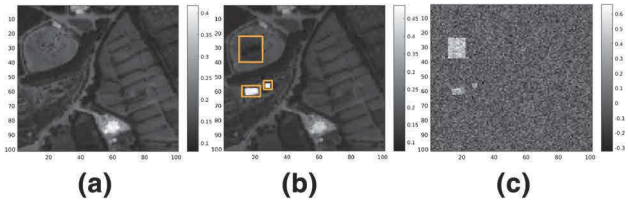

Fig. 3 **a** Image $\boldsymbol{X}_1$, **b** simulated image $\boldsymbol{X}_2$, **c** the DI $\boldsymbol{D}$, where $b = 2$

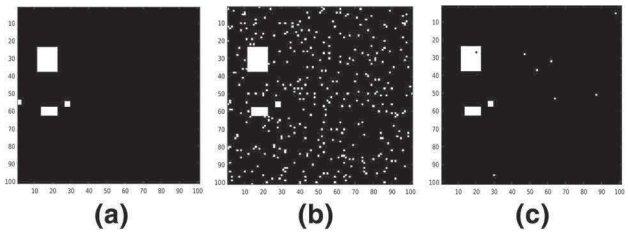

Fig. 4 **a** Ground truth and binary CMs using (**b**) the thresholding technique of S. Liu et al. and **c** the proposed model

Table 1 Quantitative results of the synthetic dataset

Accuracy measures	S. Liu et al.	The proposed method
Overall accuracy (OA%)	95.57	99.69
Precision (%)	31.68	98.33
Average accuracy (AA%)	95.68	98.88
Recall (%)	97.79	97.92
Intersection over union (IoU%)	31.25	96.31
Missed alarms (MA%)	4.2	2.3
False alarms (FA%)	4.42	0.15

4.3 Experiment 2: quickbird real dataset

This dataset covers the region of Zaghouane in Tunisia, acquired by QUICKBIRD[1] on June 21, 2005 and June 27, 2010, respectively. The images have a spatial resolution of 2.4 m with $B = 4$ (red, green, blue, near infrared) spectral channels. The images are georeferenced, both radiometrically and geometrically corrected. In this experiment, ground truth is not available. Figure 5a, b shows the images $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ of size 500×500, respectively. The white arrows indicate the major changes observed when visually analyzing the scene by an expert in RS. The studied area is divided into 25 non-overlapped sub-images of size 100×100 as shown in Fig. 5. This division allows handling homogeneous areas to avoid confusing different change types. The binary CMs are obtained for each of the underlying 25 patches. For the sake of conciseness, two patches were selected (indicated by the white arrows in Fig. 5). The obtained results are illustrated in

[1] A commercial Earth observation satellite, owned by DigitalGlobe: https://www.digitalglobe.com/resources/satellite-information.

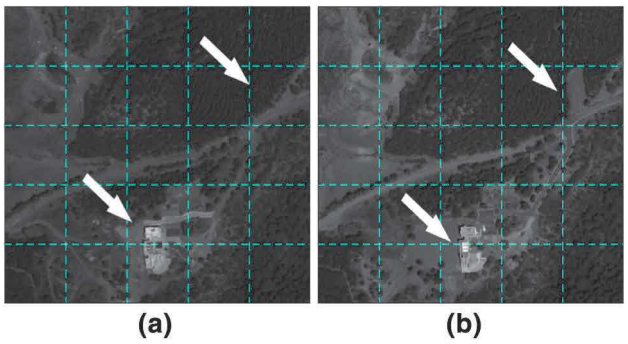

Fig. 5 **a** Image X_1 and **b** image X_2 acquired on June 21, 2005 and June 27, 2010, respectively

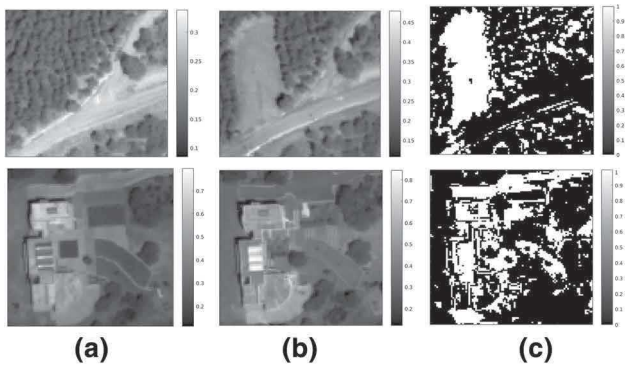

Fig. 6 First and second patches from **a** X_1 and **b** X_2 and their corresponding binary CMs (**c**)

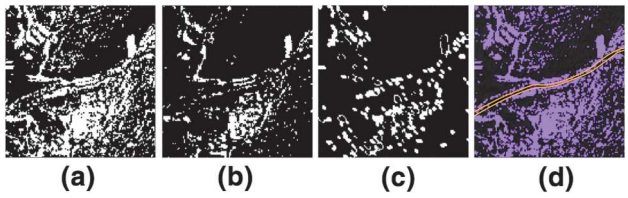

Fig. 7 Binary CMs of: **a** the thresholding technique in S. Liu et al., **b** the proposed model, **c** DI between **a** and **b** and **d** an overlay of **a** over the DI in **c**

Fig. 6, one patch per row. The original patches are displayed in Fig. 6a, b, respectively. The binary CMs are displayed in Fig. 6c where white pixels indicate changed areas. The proposed method was able to detect the main changes, which confirms the obtained results on synthetic data. Note that the same burn-in period and iteration number as for synthetic data have been considered. Figure 7a, b depicts the obtained CM using thresholding technique used in S. Liu et al. and the proposed model, respectively. Figure 7c compares the differences between the two CMs. The overlay of the result of Fig. 7a and the initial DI between the bi-temporal images is illustrated in Fig. 7d. It is worth noting that all changed pixels detected by the proposed method have also been detected as changed by the thresholding technique of Liu et al. [18]. The only difference is that the latter detected more changed pixels counted to 2048 pixels (11.93% of the image size). In Fig. 7d, the purple pixels correspond to those detected by the thresholding technique of Liu et al. [18]. The yellow trajectory illustrates the shape of the road where many pixels have been detected as changed by the method of S. Liu et al. However, the road was not submitted to any change which suggests that these pixels can be considered as false alarms due to the shade. Since the ground truth is not available, the percentage of changed and unchanged are presented in Table 2 for a numerical evaluation.

4.4 Experiment 3: GoogleEarth real dataset

This dataset and ground truth was made available in [16]. It is composed of 7 different scenes acquired from Google earth (DigitalGlobe) with different spatial resolutions from 3 to 100 cm. In [16], a deep learning technique for CD was proposed based on a convolutional neural network to derive the threshold that separates the changed pixels from the unchanged ones. This work only considers differences related to the appearance or disappearance of objects in a scene, rather than differences due to seasonal object changes like brightness variations and other factors. Therefore, some inconsistencies related to the fact that we take into consideration all the observed change including the spectral changes (e.g., changing canopies) can be observed. These VHR images cover the same region of size 4725×2700 with three spectral channels (red, green and blue). Figure 8 displays the studied scene in this experiment. In this experiment, we validate the proposed method on more complex change scenarios. The studied scene was divided into patches of larger size 1000×1000 where different changes types are simultaneously present. For performance analysis, two patches that depict numerous changes were selected. These patches are illustrated by red boxes in Fig. 8. The observed changes include the appearance of building and objects as well as changes in land cover such as tree appearance. The result of the proposed model is compared to the previously described thresholding technique and the ground truth employed in [16]. The overall result obtained over the whole studied scene of size 4725×2700 is depicted in Fig. 9. These CMs clearly depict the ability of the proposed method to detect relevant changes. In the first patch, the upper left red box changes such as the appearance of vehicles and even the change of the tree shape in the upper left corner were detected. For the second patch (bottom right), the binary CM was not successful to detect the appeared building due to the fact that the scenes before and after have very similar textures. However, the proposed method succeeds to detect all the observed changes surrounding the building since it was affected by change in luminosity and texture. The benefit that can be

Table 2 Quantitative results of the Quickbird dataset

	S. Liu et al.	The proposed method
Changed pixels (%)	11.93	4.78
Unchanged pixels (%)	88.07	95.22

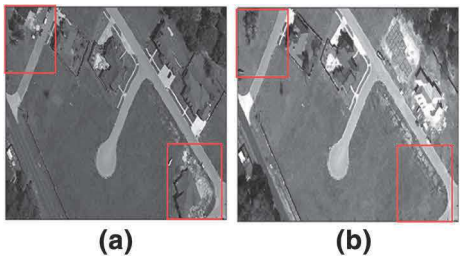

Fig. 8 X_1 (**a**) and X_2 (**b**) images of the studied scene

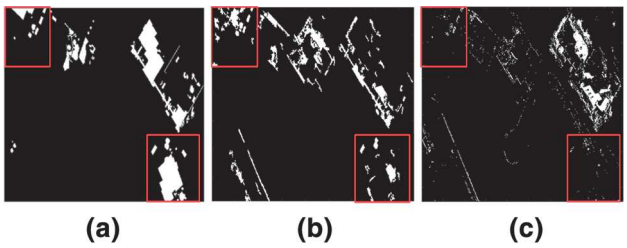

Fig. 9 **a** The ground truth in [16], **b** the binary CM of the proposed model and **c** the binary CM of S. Liu et al.

drawn from the proposed model that we also consider spectral changes observed in vegetation canopies or land cover. In fact, the growth of trees, changes post-construction in uncultivated land and also additional entryway for new houses have been detected. The visual comparison with the work of S. Liu et al. indicates that the proposed method outperforms the thresholding technique. Both of them have successfully detected the spectral changes such as the changing canopies. However, as depicted in Fig. 9c, the road have been mistakenly detected as changed. Also, there are some objects such as vehicles in the upper left that are partially missed to be detected. In Table 3, S. Liu et al. scored higher rates of missed and false alarms which is related to the fact that the thresholding method is based on the magnitude image. Because of the impact of the spectral variability, high magnitude values do not always refer to changed pixels. Besides, it is very sensitive to noise and acquisition conditions. In contrary, the key strength of the proposed method is being robust to noise which explains the difference in precision values. The quantitative assessment in Table 3 indicates that the results are consistent with the visual results in Fig. 9 as well as the quantitative results of Table 1. It is obvious that the proposed method obtained high accuracy rate in detection and exhibited stability and reproductibility of the results.

Table 3 Quantitative results of the GoogleEarth dataset

Accuracy measures	S. Liu et al.	The proposed method
Overall accuracy (OA%)	90.02	92.64
Precision (%)	14.28	52.2
Average accuracy (AA%)	58.69	71.27
Recall (%)	24.25	46.05
Intersection over union (IoU%)	9.88	32.39
Missed alarms (MA%)	6.5	4.12
False alarms (FA%)	3.5	3.22

5 Conclusions

In this paper, we have tackled the problem of CD in a Bayesian framework. The two key strengths of the proposed method are Bernoulli-Laplace model and the proposed algorithm based on the hierarchical Bayesian model for CD. Removing noise from changed pixels allows to detect critical changes and to minimize the rate of false alarms. Besides, the high performance and secure convergence of the proposed algorithm enables to account for almost all changed pixels. Several change scenarios with different images sizes and spatial resolutions were considered. Multiple types of changes are simultaneously present such as changing tree canopies, new buildings and vehicles. The tolerance to such scenarios was demonstrated for both simulated and real data. As future work, we aim at considering a context-sensitive CD by taking into account the relationship between pixels. We also are in the process of generalizing the proposed model to consider hyperspectral RS data. Besides, the applicability of the proposed method will be investigated in a transformed domain (e.g., wavelet domain).

Acknowledgements This work is supported by the Tunisian program "Projets de Recherche Fédérés" of the Ministry of Higher Education and Scientific Research under the project "Supervision Sensitive de lieux Sensibles multi-capteurs : Super-Sense".

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

References

1. Ashbindu, S.: Digital change detection techniques using remotely-sensed data. Int. J. Remote Sens. **10**(6), 989–1003 (1989)
2. Baisantry, M., Negi, D.S., Manocha, O.P.: Change vector analysis using enhanced PCA and inverse triangular function-based thresholding. Def. Sci. J. **62**(4), 236–242 (2012)

3. Bovolo, F., Bruzzone, L.: A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain. IEEE Trans. Geosci. Remote Sens. **45**(1), 218–236 (2007)
4. Bovolo, F., Marchesi, S., Bruzzone, L.: A framework for automatic and unsupervised detection of multiple changes in multitemporal images. IEEE Trans. Geosci. Remote Sens. **50**(6), 2196–2212 (2012)
5. Bruzzone, L., Bovolo, F.: A novel framework for the design of change-detection systems for very-high-resolution remote sensing images. Proc. IEEE **101**(3), 609–630 (2013)
6. Bruzzone, L., Prieto, D.F.: Automatic analysis of the difference image for unsupervised change detection. IEEE Trans. Geosci. Remote Sens. **38**(3), 1171–1182 (2000)
7. Chaari, L., Pesquet, J.C., Tourneret, J.Y., Ciuciu, P., Benazza-Benyahia, A.: A hierarchical bayesian model for frame representation. IEEE Trans. Signal Process. **58**(11), 5560–5571 (2010)
8. Chaari, L., Tourneret, J.Y., Batatia, H.: Sparse Bayesian regularization using Bernoulli–Laplacian priors. In: Signal Process. Conference, pp. 1–5. IEEE (2013)
9. Chen, G., Hay, G.J.: An airborne lidar sampling strategy to model forest canopy height from quickbird imagery and geobia. Remote Sens. Environ. **115**(6), 1532–1542 (2011)
10. Chen, M.D.H.D.: Segmentation for object-based image analysis: a review of algorithms and challenges from remote sensing perspective. J. Photogramm. Remote Sens. **150**, 115–134 (2019)
11. Coppin, P., Jonckheere, I., Lambin, E.: Digital change detection methods in ecosystem monitoring: a review. Int. J. Remote Sens. **25**, 1565–1596 (2004)
12. Dobigeon, N., Hero, A.O., Tourneret, J.Y.: Hierarchical Bayesian sparse image reconstruction with application to MRFM. IEEE Trans. Image Process. **18**(9), 2059–2070 (2009)
13. Krylov, V.A., Moser, G., Serpico, S.B., Zerubia, J.: False discovery rate approach to unsupervised image change detection. IEEE Trans. Image Process. **25**(10), 4704–4718 (2016)
14. Kwan, C. et al.: Change detection using original and fused landsat and worldview images. In: Proceedings of the IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference, New York, NY, USA, pp. 10–12 (2019)
15. Kwan, C., et al.: Simple and effective cloud-and shadow-detection algorithms for landsat and worldview images. Signal Image Video Process. **14**(1), 125–133 (2020)
16. Lebedev, M.A., et al.: Change detection in remote sensing images using conditional adversarial networks. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. (2018). https://doi.org/10.5194/isprs-archives-XLII-2-565-2018
17. Liu, G., Delon, J., Gousseau, Y., Tupin, F.: Unsupervised change detection between multi-sensor high resolution satellite images. In: EUSIPCO, pp. 2435–2439. IEEE (2016)
18. Liu, S., Bovolo, F., Bruzzone, L., Du, P.: Hierarchical unsupervised change detection in multitemporal hyperspectral images. IEEE Trans. Geosci. Remote Sens. **53**(1), 244–260 (2015)
19. Liu, S., Marinelli, D., Bruzzone, L., Bovolo, F.: A review of change detection in multitemporal hyperspectral images: current techniques, applications, and challenges. IEE Trans. Geosci. Remote Sens. Mag. **7**(2), 140–158 (2019)
20. Liu, W., Yang, J., Zhao, J., Yang, L.: A novel method of unsupervised change detection using multi-temporal polsar images. Remote Sens. **9**(11), 1135 (2017)
21. Lu, D.: Change detection techniques. Int. J. Remote Sens. **25**(12), 2365–2401 (2004)
22. Lv, P., Zhong, Y., Zhao, J., Zhang, L.: Unsupervised change detection based on hybrid conditional random field model for high spatial resolution remote sensing imagery. IEEE Trans. Geosci. Remote Sens. **56**(7), 4002–4015 (2018)
23. Prakash, A., Gupta, R.P.: Land-use mapping and change detection in a coal mining area: a case study in the jharia coalfield, India. Int. J. Remote Sens. **19**(3), 391–410 (1998)
24. Pustelnik, N., Benazza-Benhayia, A., Zheng, Y., Pesquet, J.C.: Wavelet-based image deconvolution and reconstruction. Wiley Encyclopedia of Electrical and Electronics Engineering, pp 1–34 (1999)
25. Réjichi, S., Chaabane, F.: Satellite image time series classification and analysis using an adapted graph labeling. In: 8th International Workshop on the Analysis of Multitemporal Remote Sensing Images, MultiTemp, Annecy, France, July 22–24, pp. 1–4 (2015)
26. Réjichi, S., Chaabane, F.: Spatio-temporal regions' similarity framework for VHR satellite image time series analysis. In: IEEE International Geoscience and Remote Sensing Symposium, Fort Worth, TX, USA, July 23–28, pp. 2845–2848 (2017)
27. Robert, C., Castella, G.: Monte Carlo Statistical Methods. Springer, Berlin (2004)
28. Sohl, T.L.: Change analysis in the united arab emirates: an investigation of techniques. Photogramm. Eng. Remote Sens. **65**(4), 475–484 (1999)
29. Tiao, G.C., Tan, W.Y.: Bayesian analysis of random-effect models in the analysis of variance. I. Posterior distribution of variance-components. Biometrika **52**(1/2), 37–53 (1965)
30. Volpi, M., et al.: Supervised change detection in VHR images using contextual information and support vector machines. Int. J. Appl. Earth Observ. Geoinf. **20**, 77–85 (2013)
31. Wang, X., et al.: Object-based change detection in urban areas from high spatial resolution images based on multiple features and ensemble learning. Remote Sens. **10**(2), 276 (2018)
32. Zheng, Y., Fraysse, A., Rodet, T.: Efficient variational Bayesian approximation method based on subspace optimization. IEEE Trans. Image Process. **24**(2), 681–693 (2015)