

MULTISCALE UNSUPERVISED CHANGE DETECTION ON OPTICAL IMAGES BASED ON MARKOV RANDOM FIELDS AND WAVELETS

Gabriele Moser, Elena Angiati, and Sebastiano B. Serpico

Dept. of Biophysical and Electronic Engineering (DIBE), University of Genoa
Via Opera Pia 11a, I-16145, Genova, Italy; e-mail: gabriele.moser@unige.it

ABSTRACT

In the context of environmental monitoring and natural or industrial disaster management, change-detection methods represent powerful tools for studying and mapping the evolution of the Earth's surface. In order to optimize the accuracy of the change maps, a multiscale approach can be adopted, that jointly exploits observations at coarser scales (to globally identify changed areas and gain robustness to noise) and finer scales (to improve the detection of details). In this paper, a multiscale contextual unsupervised change-detection method is proposed for optical images. It is based on discrete wavelet transforms and Markov random fields. The image-differencing approach to change-detection with optical data is adopted, wavelets are applied to the difference image to extract multiscale features, and Markovian data fusion is used to integrate both these features and the spatial contextual information in the change-detection process. Expectation-maximization and Besag's algorithms are used to estimate the model parameters. Experiments on real optical images point out the effectiveness of the method, also as compared with state-of-the-art techniques.

Index Terms— Multiscale change detection, unsupervised change detection, discrete wavelet transforms, Markov random fields, expectation-maximization, Besag's algorithm.

1. INTRODUCTION

Multitemporal remote sensing represents a powerful source of information for monitoring the evolution of the Earth's surface, for instance, in applications such as environmental monitoring or disaster management. A relevant task is the identification of the changes that occurred in a given area between two observation dates [16]. When adopting an unsupervised approach, i.e., when assuming no training data to be available at any acquisition date, image differencing and image (log-)ratioing are usually applied to address this task with optical and synthetic aperture radar (SAR) images, respectively [16]. They consist in generating a difference and (log-)ratio image, respectively, by computing pixel by pixel the difference or (logarithmic) ratio of the pixel intensities in the images acquired at the two observation dates. Thresholding can then

be used to distinguish changed and unchanged areas in these difference and (log-)ratio images, either by manual trial-and-error procedures [16] or by more advanced automatic techniques based on parametric Bayesian approaches [2, 11, 12].

In order to improve the accuracy of the change maps, a multiscale strategy can be adopted, in which transformed images at different scales are jointly used [1, 8, 9]. The images at the finest scales are likely to highlight many geometrical details, but also to be more affected by noise. Data at coarser scales exhibit less precise details, but a stronger immunity to noise. A multiscale approach, exploiting coarser scales to globally identify changed areas and finer scales to improve the detection of details, may represent an effective choice. This approach is expected to be particularly promising when dealing with very high resolution images acquired by either optical (e.g., QuickBird or IKONOS) or SAR (e.g., COSMO/SkyMed, TerraSAR-X) sensors. Multiscale change detection methods for SAR images have been proposed in [9] and [5], by using a log-cumulant-based information-theoretic change measure and the combination of undecimated stationary wavelet transforms (SWTs) with adaptive optimal scale-selection schemes, respectively. In [1] an unsupervised multiscale change-detection method was proposed by applying a previous unsupervised multiresolution classifier to wavelet transforms of the difference image. Object-oriented analysis and watershed segmentation were used in [8] for semi-interactive multiscale change detection on optical data. A semi-interactive approach was also proposed in [13] by combining wavelets, anisotropic diffusion filtering, and an empirical procedure for optimal-scale selection. In [3] a heuristic multiscale change measure for optical data, based on a normalized product of several low-pass filtered versions of the difference image, was introduced and tested by MonteCarlo simulations.

In this paper, a multiscale contextual unsupervised change-detection method is proposed for optical images, based on discrete wavelet transform (DWT) [6] and Markov random fields (MRFs) [4]. DWT is applied to the difference image to extract multiscale features. It provides a two-dimensional multiresolution decomposition of the input (difference) image as the superposition of several components, each highlighting the image content at a given scale [6]. MRFs jointly

allows spatial context to be introduced into pixel labeling problems and different information sources to be fused by suitably defining “energy functions” [17]. Here, an MRF-based approach is proposed by modifying the method introduced in [18] for multiresolution image classification and by combining it with DWTs and image differencing in order to extend it to multiscale change detection. The related MRF model is characterized by internal parameters that are estimated by the expectation-maximization (EM) [15] and Besag’s [4] algorithms [18]. The main novelties of the paper lie in the extension of the method in [18] from multiresolution classification to multiscale change-detection and in its combination with wavelets. A similar approach was proposed in our previous work [1] by using a more heuristic combination of DWT and of the multiresolution classifier in [18]. Unlike [1], here a modified formulation of this classifier is developed in order to optimally integrate it with DWT. The proposed method is experimentally validated on optical satellite images and compared with previous single-scale and multiscale approaches.

The paper is organized as follows. Sec. 2 provides a methodological description of the proposed technique and of the related parameter-estimation issues. Sec. 3 presents the results of the experimental validation and of the comparisons with state-of-the-art techniques. Conclusions are drawn in Sec. 4.

2. METHODOLOGY

2.1. Overview of the proposed method

Let \mathcal{I}_0 and \mathcal{I}_1 be two coregistered optical images, composed of N pixels each and acquired over the same area at times t_0 and t_1 , respectively ($t_1 > t_0$). We assume \mathcal{I}_0 and \mathcal{I}_1 to be single-channel images; the reformulation in the multichannel case is straightforward. Change detection is formalized as a binary hypothesis testing problem [7], by marking the “change” and “no-change” hypotheses as H_1 and H_0 , respectively. Image differencing, which generates a difference image \mathcal{D} by subtracting pixel-by-pixel the pixel intensities in \mathcal{I}_0 by the pixel intensities in \mathcal{I}_1 , is adopted [16].

The key idea of the proposed method lies in generating a finite set $\{\mathcal{W}_1, \mathcal{W}_2, \dots, \mathcal{W}_S\}$ of multiscale features by applying DWT to \mathcal{D} (see Sec. 2.2), and to fuse this multiscale information by an MRF approach. MRFs also allow the spatial contextual information to be exploited, thus gaining robustness against noise. The MRF classifier proposed in [18] for images acquired at different spatial resolutions is generalized here to multiscale change detection (see Sec. 2.3). It adopts the “iterated conditional mode” (ICM) approach to MRF-based classification, which usually represents a good tradeoff between accuracy and computational burden [17]. The proposed method is iterative and is initialized by using the change map generated by applying to \mathcal{D} the (single-scale)

change-detection technique proposed in [11] and based on the Kittler and Illingworth (K&I) unsupervised thresholding algorithm (see Sec. 2.4) [10, 11].

2.2. Wavelet transform stage

An S -scale multiresolution decomposition of \mathcal{D} (S being a predefined number of scales) is obtained by applying a dyadic DWT. \mathcal{D} is decomposed in terms of a low-pass transformed image and of three transformed images conveying high-pass information about fine-scale details along the horizontal image axis, the vertical axis, or both axes. Then, the procedure is recursively applied S times to the low-pass component. $\mathcal{W}_1, \mathcal{W}_2, \dots, \mathcal{W}_S$ are obtained by applying S times the inverse transform (IDWT) while neglecting all high-pass components and keeping only the low-pass terms [6]. Thus, as s increases in $[1, S]$, coarser-scale approximations \mathcal{W}_s of \mathcal{D} are obtained, while the finest scale is \mathcal{D} itself (which will also be denoted by \mathcal{W}_0 hereafter). More precisely, \mathcal{W}_s allows appreciating spatial details that are (2^s) -times coarser than \mathcal{W}_0 ($s = 1, 2, \dots, S$).

2.3. MRF-based classification stage

\mathcal{W}_s is modeled as a finite set $\mathcal{W}_s = \{u_{sk}\}_{k=1}^N$ of random variables, where u_{sk} denotes the intensity of the k -th pixel at the s -th scale ($s = 0, 1, \dots, S; k = 1, 2, \dots, N$). Similar to the multiresolution approach in [18], the statistical relationships between images at different scales are modeled in terms of “linear mixtures.” In [18] coregistered multispectral images with different resolutions are used to generate a classification map at the finest resolution. For each coarser-resolution image, a set of “virtual” pixel intensities is assumed to exist at the finest resolution such that the pixel intensities at the coarser resolution are modeled by mosaic averaging these virtual intensities [18]. Here, this approach is generalized to multiscale analysis by postulating that, for each \mathcal{W}_s ($s = 1, 2, \dots, S$), a set $\{\tilde{u}_{sk}\}_{k=1}^N$ of virtual pixel intensities exists at the 0-th scale, such that:

$$u_{sk} = \frac{1}{4^s} \sum_{r \in Q_{sk}} \tilde{u}_{sr}, \quad (1)$$

where Q_{sk} is a $2^s \times 2^s$ window centered on the k -th pixel ($k = 1, 2, \dots, N$). If $s = 0$, the identity $\tilde{u}_{0k} = u_{0k}$ holds.

Since \mathcal{W}_s is obtained from \mathcal{W}_0 through DWT/IDWT, the pixel values in \mathcal{W}_s could be deterministically expressed as linear combinations of the pixel values in \mathcal{W}_0 , provided a given DWT operator (e.g., Daubechies, biorthogonal) is chosen. In this perspective, the virtual intensities \tilde{u}_{sk} may not be needed. Indeed we propose to adopt the formalization based on them, first, because this allows extending the Markovian formulation and the parameter-optimization procedures already developed in [18] for the multiresolution case. Furthermore, this choice also prevents the need to predefine a

specific DWT at each scale and to plug the analytical expressions of the impulse responses of the related filters in the parameter-estimation process. This allows any DWT to be used in the proposed MRF framework and ensures a higher flexibility.

For each k -th pixel in \mathcal{W}_0 , let $\ell_k \in \{H_0, H_1\}$ be the related hypothesis label and x_k be the multiscale vector of all corresponding pixel intensities u_{sk} ($s = 0, 1, \dots, S; k = 1, 2, \dots, N$). Given the label configuration $\{\ell_k\}_{k=1}^N$, the variables \tilde{u}_{sk} ($k = 1, 2, \dots, N$) are assumed to be conditionally independent and identically distributed. The probability density function (PDF) of \tilde{u}_{sk} , given $\ell_k = H_i$, is modeled as a Gaussian with mean μ_{si} and variance σ_{si}^2 ($s = 0, 1, \dots, S; i = 0, 1$). Let the vector θ collect all μ_{si} and σ_{si} parameters. Thanks to the linearity of DWT/IDWT and of the model in Eq. (1), also the PDF of u_{sk} , given $\ell_k = H_i$ and the labels ℓ_r of the pixels r such that $r \in Q_{sk}$, is easily proved to be Gaussian (which is a usually accepted model for the statistics of \mathcal{D} in the case of input optical images [11]), with mean and variance parameterized by θ and given by [18]:

$$\begin{aligned} \mu_{sik}(\theta) &= E\{u_{sk} | \ell_k = H_i; \ell_r, r \in Q_{sk}\} = \\ &= 4^{-s} [n_{sik} \mu_{si} + (4^s - n_{sik}) \mu_{s,1-i}] \end{aligned} \quad (2)$$

and

$$\begin{aligned} \sigma_{sik}^2(\theta) &= \text{Var}\{u_{sk} | \ell_k = H_i; \ell_r, r \in Q_{sk}\} = \\ &= 16^{-s} [n_{sik} \sigma_{si}^2 + (4^s - n_{sik}) \sigma_{s,1-i}^2], \end{aligned} \quad (3)$$

where n_{sik} is the number of pixels r such that $r \in Q_{sk}$ and $\ell_r = H_i$ ($k = 1, 2, \dots, N; s = 0, 1, \dots, S; i = 0, 1$).

The label configuration $\{\ell_k\}_{k=1}^N$ is assumed to be an MRF [4]. An isotropic 2nd-order Potts MRF model is adopted. Therefore, the distribution of ℓ_k , given x_k and all the other image labels is expressed by [18]:

$$P\{\ell_k = H_i | x_k; \ell_r, r \neq k\} = \frac{\exp[-U_k(H_i | \lambda, \theta)]}{\sum_{j=0}^1 \exp[-U_k(H_j | \lambda, \theta)]}, \quad (4)$$

where:

$$U_k(H_i | \lambda, \theta) = \sum_{s=0}^S \left\{ \frac{[u_{sk} - \mu_{sik}(\theta)]^2}{\sigma_{sik}^2(\theta)} + \ln \sigma_{sik}^2(\theta) \right\} - \lambda m_{ik} \quad (5)$$

is the energy function of the MRF model, m_{ik} is the number of labels equal to H_i in the 2nd-order neighborhood of the k -th pixel ($k = 1, 2, \dots, N; i = 0, 1$), and λ is a positive parameter. The energy is a linear combination of $(S+2)$ contributions: the last term is related to the spatial contextual information, represented by m_{ik} according to the Potts model, and the other terms are related to the information conveyed by the H_i -conditional distributions of the pixel values at all considered scales. The Potts energy contribution encourages the generation of connected regions whose pixels are assigned to the same hypothesis, thus essentially representing a spatial

regularization term. The parameter λ tunes the relative importance between the spatial energy term and the multiscale contributions. Coherently with the Markovianity property, the distribution in Eq. (4) does not depend on all image labels ℓ_r with $r \neq k$, but only on the labels of the neighbors of the k -th pixel (through m_{ik}) and on the labels ℓ_r such that $r \in Q_{sk}$ for some $s \in \{1, 2, \dots, S\}$ (through $\mu_{sik}(\theta)$ and $\sigma_{sik}(\theta)$).

We recall that the MRF-based method in [18] assumes the single-resolution images to be independent of one another, when conditioned to the label configuration. Here, this corresponds to the fact that single-scale energy terms are additively combined in Eq. (5) without modeling the correlations among features at different scales. Indeed, since all \mathcal{W}_s features ($s = 1, 2, \dots, S$) are obtained from \mathcal{W}_0 through DWT/IDWT, the conditional independence assumption cannot be rigorously fulfilled in the present case. We will accept it as a simplifying analytical tractability assumption.

As in [18], the EM and the Besag's methods are iteratively used to estimate θ and λ , respectively. EM addresses estimation problems characterized by data incompleteness and converges, under mild assumptions, to a local maximum of the log-likelihood function (even though convergence to a global maximum is not ensured, usually a good solution is obtained) [15]. The Besag's method estimates the spatial regularization parameters of an MRF (here, λ) by maximizing a pseudolikelihood function, obtained by an approximation of the joint likelihood function of all image labels [4].

The proposed method is iterative and is initialized with the label map given by K&I (see Sec. 2.4). At the t -th iteration, denoting by the superscript " t " the current parameter estimates and pixel labels, the following operations are performed ($t = 0, 1, 2, \dots$):

1. compute the updated estimate λ^t , given the current labels $\{\ell_k^t\}_{k=1}^N$, by numerically maximizing the following pseudolikelihood function (Besag's method) [4]:

$$\Phi^t(\lambda) = \prod_{i=0}^1 \prod_{k: \ell_k^t = H_i} \frac{\exp(\lambda m_{ik}^t)}{\exp(\lambda m_{0k}^t) + \exp(\lambda m_{1k}^t)}; \quad (6)$$

2. compute the updated parameter estimate θ^t by running EM until convergence (see below);
3. update the label of each k -th pixel according to the MRF minimum-energy rule, by setting ℓ_k^{t+1} as the label H_i that corresponds to the lowest value of $U_k^t(H_i | \lambda^t, \theta^t)$ ($k = 1, 2, \dots, N; i = 0, 1$) [17].

Steps 1-3 are iterated until the percentage of pixels with different labels in successive iterations goes below a predefined threshold (here, equal to 0.1%). At each t -th iteration ($t = 0, 1, \dots$), the numerical maximization in step 1 is solved by the Newton-Raphson algorithm, which can be applied to this end because $\Phi^t(\cdot)$ is a twice continuously differentiable function [14]. Given the resulting estimate λ^t of

λ and the current label configuration $\{\ell_k^t\}_{k=1}^N$, EM computes in step 2 a sequence $\{\bar{\theta}^{qt}\}_{q=0,1,\dots}$ of estimates of θ . Thanks to the convergence properties of EM, $\bar{\theta}^{qt}$ converges for $q \rightarrow +\infty$ to a limit θ^t which is the updated estimate of θ resulting from step 2. Operatively, EM is iterated until the distance between $\bar{\theta}^{q+1,t}$ and $\bar{\theta}^{qt}$ goes below a predefined threshold (here, equal to 0.001) and is initialized by setting $\bar{\theta}^{0t} = \theta^{t-1}$ ($t = 1, 2, \dots$).

Let us denote by η_{sk} and ξ_{sk} the mean and variance of the virtual intensity \tilde{u}_{sk} , given all multiscale observations and all image labels, i.e. ($s = 0, 1, \dots, S; k = 1, 2, \dots, N$):

$$\begin{aligned}\eta_{sk} &= E\{\tilde{u}_{sk} | \ell_r, x_r, r = 1, 2, \dots, N\} \\ \xi_{sk} &= \text{Var}\{\tilde{u}_{sk} | \ell_r, x_r, r = 1, 2, \dots, N\}\end{aligned}\quad (7)$$

According to the identity $\tilde{u}_{0k} = u_{0k}$ and to the fact that u_{0k} is among the components of x_k , one has $\eta_{0k} = u_{0k}$ and $\xi_{0k} = 0$. Thanks to the close similarity between the linear-mixture model of Eq. (1) and the similar model used in [18], one can prove through the same analytical calculations as in [18] that, at each q -th EM iteration, the components $\bar{\mu}_{si}^{qt}$ and $\bar{\sigma}_{si}^{qt}$ of $\bar{\theta}^{qt}$ as well as estimates η_{sk}^{qt} and ξ_{sk}^{qt} of η_{sk} and ξ_{sk} , respectively, are updated as follows ($i = 0, 1; s = 0, 1, \dots, S; k = 1, 2, \dots, N$):

$$\begin{aligned}\eta_{sk}^{qt} &= \begin{cases} \bar{\mu}_{sj}^{qt} + \frac{(\bar{\sigma}_{sj}^{qt})^2}{4^s} \cdot \frac{u_{sk} - \mu_{sjk}(\bar{\theta}^{qt})}{\sigma_{sjk}^2(\bar{\theta}^{qt})} & \text{for } s \geq 1 \\ u_{0k} & \text{for } s = 0 \end{cases} \\ \bar{\mu}_{si}^{q+1,t} &= \frac{\sum_{r=1}^N \delta(\ell_r^t, H_i) \eta_{sr}^{qt}}{\sum_{r=1}^N \delta(\ell_r^t, H_i)} \\ \xi_{sk}^{qt} &= \begin{cases} (\bar{\sigma}_{sj}^{qt})^2 - \frac{(\bar{\sigma}_{sj}^{qt})^4}{16^s \cdot \sigma_{sjk}^2(\bar{\theta}^{qt})} & \text{for } s \geq 1 \\ 0 & \text{for } s = 0 \end{cases} \\ (\bar{\sigma}_{si}^{q+1,t})^2 &= \frac{\sum_{r=1}^N \delta(\ell_r^t, H_i) [\xi_{sr}^{qt} + (\eta_{sr}^{qt} - \bar{\mu}_{si}^{q+1,t})^2]}{\sum_{r=1}^N \delta(\ell_r^t, H_i)},\end{aligned}\quad (8)$$

where $\delta(\cdot, \cdot)$ is the Kronecker symbol (i.e., $\delta(a, b) = 1$ if $a = b$ and $\delta(a, b) = 0$ otherwise) and j is the index of the hypothesis H_j such that $\ell_k^t = H_j$.

2.4. Initialization by the K&I method

K&I is used to generate the initial change map. It automatically computes the optimal threshold to be applied to \mathcal{D} (in order to distinguish between changed and unchanged areas) by minimizing a ‘‘criterion function’’ related to the probability of error of a Bayesian binary classifier [10, 11]. For

initialization purposes, the initial map is used to define the initial context of each k -th pixel for the computation of m_{ik} ($i = 0, 1; k = 1, 2, \dots, N$) and the initial estimate of the parameter vector θ . Specifically, a desired choice would be to compute initial estimates of μ_{si} and σ_{si}^2 as the sample mean and sample variance of \tilde{u}_{sk} on the set of samples assigned to H_i in the initial map ($i = 0, 1; s = 0, 1, \dots, S$). However, the values of the virtual intensities \tilde{u}_{sk} are available only for $s = 0$. So, as a feasible simple approach, initial estimates of μ_{si} and σ_{si}^2 for $s = 1, 2, \dots, S$ are calculated as the sample mean and sample variance of u_{sk} on the set of samples assigned to H_i in the initial map ($i = 0, 1$).

3. EXPERIMENTAL RESULTS

3.1. Data set and experimental set-up

A multitemporal data set consisting of a pair of coregistered 256×256 multispectral Landsat-5 TM images acquired over the western part of the Island of Elba (Italy) in August and September 1994, was used for experiments (Figs. 1(a)-1(b)). During the period between the acquisition dates, a fire occurred in a forested area in the image. TM bands 4 and 7 were available. Experiments were separately performed with each band in order to focus on a more difficult classification problem, as compared with the two-band case, and to consequently better highlight the possible advantages of multiscale processing as compared to single-scale approaches.

The proposed method is based on the fusion of multiscale and contextual information. Its performances were first evaluated as a function of the choice of the DWT. Then experimental comparisons were conducted with: (i) the technique developed in [11] (which is based on the application of K&I to the moving-average filtered difference image and which will be denoted hereafter by K&I-Filt) to assess the advantages of the proposed method as compared to a single-scale one; (ii) a variant of the proposed technique based on an MRF model with no spatial energy (i.e., $\lambda = 0$) to focus on the relative role of multiscale and contextual information; (iii) a version adapted to optical data of the ‘‘fusion at feature level – all reliable scales’’ (FFL-ARS) method in [5]. The behavior of the proposed technique as a function of the number of scales was also analyzed.

Given the exhaustive test map in Fig. 2(a), all results were quantitatively assessed by computing the detection accuracy (i.e., the percentage of changed test pixels that were correctly labeled as changed), the false-alarm rate (i.e., the percentage of unchanged test pixels that were erroneously labeled as changed) and the overall error rate (i.e., the percentage of erroneously labeled pixels). For each of the two bands, up to four lower-scale images were generated. As a preliminary experiment, aimed at assessing the behavior of the method with a very simple multiscale feature extraction, a ‘‘neighbor averaging’’ decomposition was used that generates \mathcal{W}_{s+1} by triv-

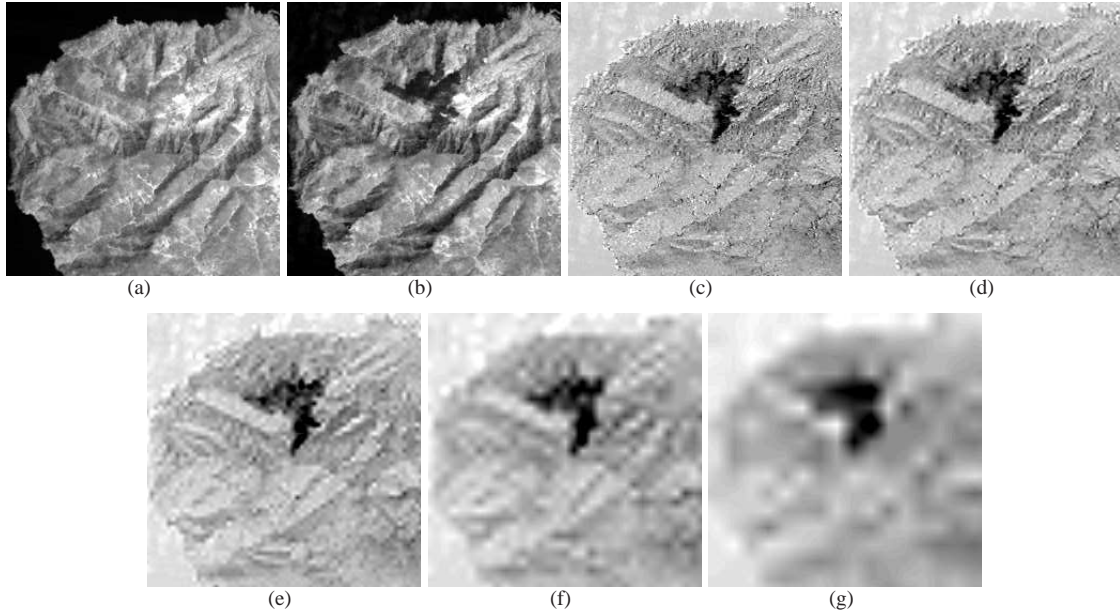


Fig. 1. Band 4 of the considered data set: images acquired in August (a) and September 1994 (b), difference image (c), and multiscale features extracted by a (2, 8)-order biorthogonal DWT at scales $1 \div 4$ (d)-(g) (after histogram stretching).

ially averaging the pixel intensities of four neighboring pixels in \mathcal{W}_s ($s = 0, 1, 2, 3$). Then, several DWTs were considered, namely, Haar, Daubechies of order ranging in the interval $[2, 30]$, symlet of order ranging in $[2, 20]$, biorthogonal and reverse biorthogonal with the two order parameters ranging in $[1, 6]$ and $[1, 8]$, respectively, discrete Meyer, and coiflet (details about these DWTs can be found in [6]). As an example, the multiscale features obtained by applying a (2, 8)-order biorthogonal DWT to band 4 are shown in Figs. 1(d)-1(g), together with the related difference image (Fig. 1(c)).

3.2. Change-detection results

As shown in Table 1, a high detection accuracy and a low error rate, equal to 91.84% and 0.47%, respectively, were obtained by the proposed method when applied with neighbor averaging to band 4. A large 19.35% increase in the detection accuracy and a 0.63% reduction in the error rate were achieved as compared with K&I-Filt. This improvement is interpreted as a consequence of the use of multiscale and contextual information. The same comments are suggested by a visual comparison of the related change maps (Figs. 2(b)-2(c)). However, Fig. 2(c) also points out the slight presence of undesired blocky artifacts, due to neighbor averaging. Lower accuracies, even though better than the ones of K&I-Filt, were obtained by the proposed technique when applied to band 7 with neighbor-averaging.

Fig. 4 summarizes the detection accuracies obtained by the proposed method, when applied to band 4 with all above-mentioned DWTs. The false-alarm rates were below 0.5%

in all cases. A similar plot was obtained for band 7 and is not presented for brevity. The performance was affected by the choice of the DWT operator. However, for all considered DWTs, higher detection accuracies and lower error rates were always obtained, as compared with K&I-Filt. The results suggested that the biorthogonal and reverse biorthogonal families provided the highest detection accuracies. Table 1 presents, for each band, the results of three out of the DWTs yielding the most accurate change maps. A high 92.71% detection accuracy and a 0.41% error rate were obtained on band 4 by using a (2, 8)-order biorthogonal DWT, thus gaining a 20.22% detection accuracy as compared with K&I-Filt. A visual analysis of the resulting change maps (see, for instance, Fig. 2(d)) confirms the accurate identification of the burnt area with a low number of false alarms, and points out that the result is not affected by blocky artifacts. Less accurate results were obtained with band 7, reaching an 82.23% detection accuracy and a 1.15% error rate with a (3, 7)-order reverse biorthogonal transform. This confirms a more limited effectiveness of band 7 than of band 4 in the considered data set, with respect to the problem of burnt forest-area detection.

Focusing, for each band, on the DWT providing the highest detection accuracy, the proposed method was applied with all five scales, while removing the Potts spatial energy component (i.e., while setting $\lambda = 0$). As compared with the case with the spatial energy, increases in the detection accuracy and especially in the false-alarm rate are remarked (Table 1). This yields a much worse overall error rate than in the case with the spatial energy. A visual analysis of the related change map (Fig. 2(e)) confirms the presence of many false

Table 1. Change-detection performances of the proposed method applied with several DWTs (the values of possible order parameters are in parentheses), its variant with no spatial energy ($\lambda = 0$), K&I-Filt, and FFL-ARS.

TM band	Method	False-alarm rate	Detection accuracy	Overall error rate
4	K&I-Filt	0.08%	72.49%	1.09%
	neighbor averaging	0.17%	91.84%	0.47%
	biorthogonal (2, 8)	0.15%	92.71%	0.41%
	reverse biorthogonal (6, 8)	0.15%	92.29%	0.43%
	symlet (5)	0.19%	91.30%	0.50%
	biorthogonal (2, 8) ($\lambda = 0$)	1.79%	95.53%	1.89%
	FFL-ARS	0.70%	86.87%	1.15%
7	K&I-Filt	0.92%	71.29%	1.94%
	neighbor averaging	0.66%	77.92%	1.45%
	reverse biorthogonal (3, 7)	0.51%	82.23%	1.15%
	symlet (4)	0.54%	78.75%	1.31%
	Daubechies (12)	0.54%	78.62%	1.30%
	reverse biorthogonal (3, 7) ($\lambda = 0$)	1.59%	81.11%	2.23%
	FFL-ARS	0.84%	63.79%	2.14%

alarms when no spatial energy term is used, and points out that spatially more regular maps (especially at the borders between changed and unchanged areas) are obtained when including this term. These results can be explained by noting that the proposed method assumes a Gaussian monomodal PDF for each hypothesis, whereas the difference image statistics is actually the mixture of three populations related to sea, non-burnt vegetation, and burnt vegetation, respectively. EM converges to a solution that correctly detects burnt vegetation as “change” and sea as “no-change” but erroneously splits the non-burnt vegetation mode, thus generating many sparse false alarms. On the other hand, the spatial regularization yielded by the Potts energy term penalizes such false alarms. This confirms the importance of the MRF approach as a contextual data-fusion tool.

FFL-ARS was proposed in [5] for unsupervised multiscale change detection on SAR data. It extracts coarser-scale features by applying a 4-th order Daubechies SWT to the log-ratio image. Then, for each pixel, the “reliable scales,” i.e., the scales at which possible edges or spatial details can still be appreciated, are adaptively identified based on a moving-window computation of a local coefficient of variation on the ratio image. The features at the reliable scales are averaged together and the resulting image is (manually or automatically) thresholded to distinguish changed and unchanged areas. Here, FFL-ARS was adapted to optical data by applying both SWT (with four lower-scale transforms) and the reliable-scale selection to the difference image, and by using K&I to identify the optimal threshold. This approach was used for comparison purposes because no unsupervised multiscale change-detection techniques have been proposed so far for optical data (an exception could be our previous method in [1], but it may be considered as a more heuristic variant

of the technique proposed here). The results characterized by the lowest overall error rate, as the size of the moving window was varied between 5×5 and 51×51 , are shown in Table 1. FFL-ARS provided quite an accurate change map for band 4, with a higher detection accuracy but also a slightly higher overall error rate than K&I-Filt. However, the method proposed here yielded a lower overall error rate and a higher detection accuracy than FFL-ARS. A visual analysis of the related change maps points out an improved spatial regularity of the result given by the proposed technique (see Fig. 2(d)) as compared to the map generated by FFL-ARS (see Fig. 2(f)). FFL-ARS gave a slightly lower error rate than the variant of the proposed method with no spatial energy. This suggests a higher effectiveness of the MRF-based multiscale approach as compared to the adaptive multiscale fusion strategy used by FFL-ARS and further confirms the role of the MRF model (similar comments hold for band 7). Note that the extension of FFL-ARS to optical data may be partially heuristic: the local coefficient of variation is related to speckle statistics in SAR data [5]; when computed on the difference of optical images, it does not have this data-specific interpretation, even though it still represents a local homogeneity index.

Finally, focusing again on the DWT with the highest detection accuracy, the number of scales was varied in the range [1, 5] (see Fig. 4). When just one scale is used, the method degenerates to a fairly standard ICM-based unsupervised classification of the difference image. Higher detection accuracies were obtained, as expected, when using all five scales than in this degenerate case. The differences between the detection accuracies obtained in the two cases were 1.37% and 9.49% with bands 4 and 7, respectively (in both cases the false-alarm rates did not exceed 0.8% and 1.5% for bands 4 and 7, respectively). Even slightly higher accuracies were obtained with

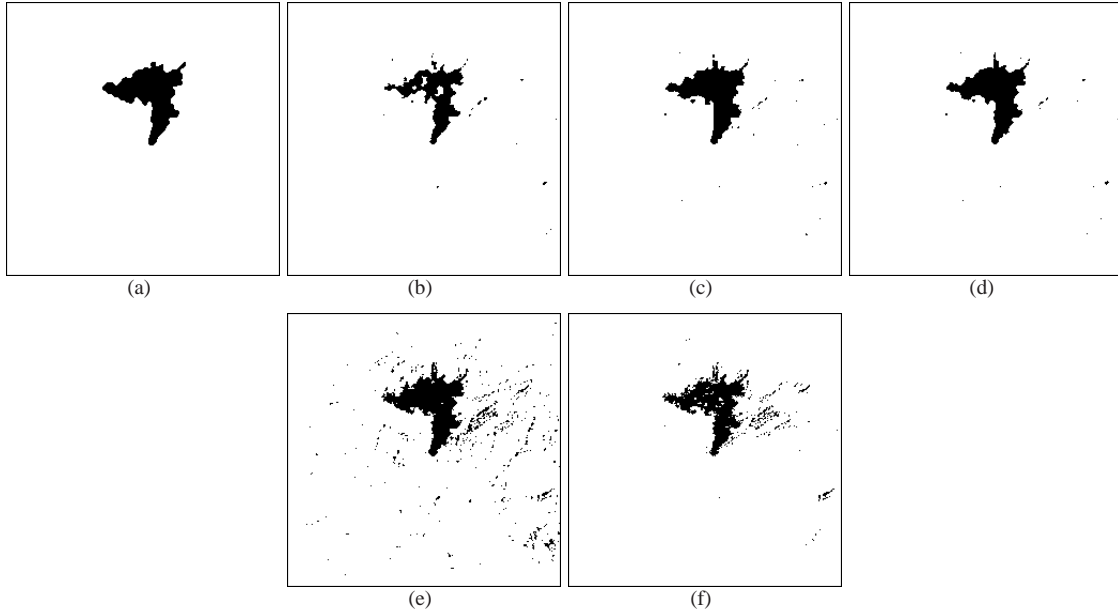


Fig. 2. Band 4 of the considered data set: test map (a) and change maps obtained by K&I-Filt (b), by the proposed method applied with neighbor averaging (c) and with the biorthogonal transform of order (2, 8) (with (d) or without (e) the spatial energy contribution), and by FFL-ARS (f). Legend: black = “change,” white = “no-change.”

two scales on this data set. This confirms the importance of multiscale information (with, at least, two scales) in the proposed method and also suggests quite a limited sensitivity to the number of scales.

4. CONCLUSIONS

An unsupervised multiscale contextual change-detection method has been proposed in this paper for optical multi-temporal images by extracting multiscale wavelet features, and by combining them through Markovian data fusion. The approach proposed in [18] for multiresolution classification has been modified and extended here to multiscale change detection. Experiments on real optical images suggested the effectiveness of the proposed technique, which generated accurate change maps, outperforming a single-scale approach based on the application of unsupervised thresholding to the noise-filtered difference image [11]. This improvement was obtained even when using a trivial neighbor-averaging to extract the multiscale features; however, blocky artifacts were generated in this case. On the contrary, no artifacts were remarked when using smoother wavelets, which also allowed a further significant change-detection accuracy increase to be obtained as compared with neighbor averaging. These results suggest the usefulness of multiscale information for change-detection purposes, in order to jointly exploit the higher robustness to noise at coarser scales and the presence of more precise geometrical details at finer scales. The proposed technique has also been compared to a previous

multiscale method developed for SAR change detection [5] and adapted here to optical data. The method is based on stationary wavelet transforms and on an adaptive selection of the optimal scale according to the presence of details and borders. The proposed approach provided more accurate results than this technique, essentially thanks to the improved spatial regularization granted by the adopted Markovian multiscale fusion approach.

With all considered wavelets, more accurate change maps were obtained, as compared with the single-scale approach in [11]. However, the performances were sensitive to the choice of the DWT and of its possible order parameters (symlets, biorthogonal, and reverse biorthogonal wavelets of suitable order provided the most accurate change maps on the considered data set). In this perspective, an interesting future extension of this work is the automatic optimization of the number of scales and of the choice of the DWT.

The method has been tested on single-channel images, but it can also be generalized to multispectral images. From an analytical viewpoint, the extension is straightforward [18]. Operatively, if several multiscale features are extracted from each band in a multispectral image, the resulting overall number of features may be quite large, which could involve dimensionality issues [7]. This would be an interesting aspect worth being investigated. A further extension of the method could consist in the extraction of textures as spatial-contextual input features and in the corresponding integration of a non-parametric estimation algorithm for the nongaussian texture statistics.

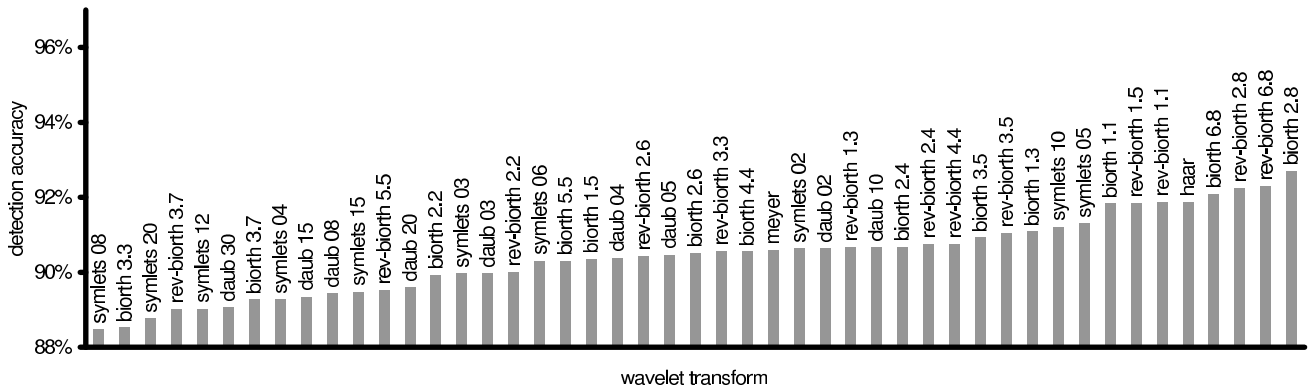


Fig. 3. Detection accuracy of the proposed method, when applied to band 4, as a function of the DWT operator (the values of possible order parameters of each DWT are also reported).

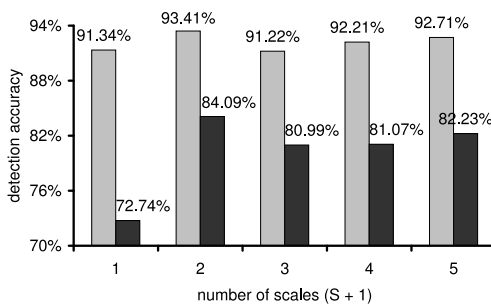


Fig. 4. Behavior of the detection accuracy as a function of the number of considered scales, for bands 4 (grey) and 7 (black).

5. REFERENCES

- [1] ANGIATI, E., MOSER, G., AND SERPICO, S. B. Multiscale unsupervised change detection by markov random fields and wavelet transforms. In *Proc. of SPIE-ISPRS XIII, Florence, Italy, 17-20 September 2007* (2007).
- [2] BAZI, Y., BRUZZONE, L., AND MELGANI, F. An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. *IEEE Trans. Geosci. Remote Sensing* 43, 4 (2005), 874–887.
- [3] BEAUCHEMIN, M., AND FUNG, K. B. Investigation of multiscale product for change detection in difference images. In *Proc. of IGARSS-2004, Anchorage, USA* (2004), pp. 3853–3856.
- [4] BESAG, J. On the statistical analysis of dirty pictures. *J. R. Statist. Soc.* 68 (1986), 259–302.
- [5] BOVOLO, F., AND BRUZZONE, L. A detail-preserving scale-driven approach to change detection in multitemporal SAR images. *IEEE Trans. Geosci. Remote Sensing* 43, 12 (2005).
- [6] DAUBECHIES, I. *Ten lectures on wavelets*. SIAM, 1992.
- [7] FUKUNAGA, K. *Introduction to statistical pattern recognition*. Academic Press, 1990.
- [8] HALL, O., AND HAY, G. J. A multiscale object-specific approach to digital change detection. *Int. J. Appl. Earth Obs.* 4, 4 (2003), 311–327.
- [9] INGLADA, J., AND MERCIER, G. A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis. *IEEE Trans. Geosci. Remote Sensing* 45, 5 (2007).
- [10] KITTLER, J., AND ILLINGWORTH, J. Minimum error thresholding. *Pattern Recogn.* 19 (1986), 41–47.
- [11] MELGANI, F., MOSER, G., AND SERPICO, S. B. Unsupervised change detection methods for remote sensing images. *Opt. Eng.* 41, 12 (2002), 3288–3297.
- [12] MOSER, G., AND SERPICO, S. B. Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery. *IEEE Trans. Geosci. Remote Sensing* 40, 10 (2006), 2972–2982.
- [13] OUMA, Y. O., JOSAPHAT, S. S., AND TATEISHI, R. Multi-scale remote sensing data segmentation and post-segmentation change detection based on logical modeling: Theoretical exposition and experimental results for forestland cover change analysis. *Computers and Geosciences* 34 (2008), 715–737.
- [14] PRESS, W. H., TEUKOLSKY, S. A., WETTERLING, W. T., AND FLANNERY, B. P. *Numerical recipes in C*. Cambridge University Press, Cambridge (UK), 2002.
- [15] REDNER, R. A., AND WALKER, H. F. Mixture densities, maximum likelihood, and the EM algorithm. *SIAM Review* 26, 2 (1984), 195–239.
- [16] SINGH, A. Digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* 10 (1989), 989–1003.
- [17] SOLBERG, A. H. S., TAXT, T., AND JAIN, A. K. A Markov random field model for classification of multisource satellite imagery. *IEEE Trans. Geosci. Remote Sensing* 34, 1 (1996), 100–113.
- [18] STORVIK, G., FJORTOFT, R., AND SOLBERG, A. H. S. A Bayesian approach to classification of multiresolution remote sensing data. *IEEE Trans. Geosci. Remote Sensing* 43, 3 (2005), 539–547.