2009 Ninth International Conference on Intelligent Systems Design and Applications

A contextual multiscale unsupervised method for change detection with multitemporal remote-sensing images

Gabriele Moser, Elena Angiati, and Sebastiano B. Serpico
Dept. of Biophysical and Electronic Eng. (DIBE)
University of Genoa
Genova, Italy
Email: sebastiano.serpico@unige.it

***Abstract*—Change-detection represents a powerful tool for monitoring the evolution of the Earth's surface by multitemporal remote-sensing imagery. Here, a multiscale approach is proposed, in which observations at coarser and finer scales are jointly exploited, and a multiscale contextual unsupervised change-detection method is developed for optical images. Discrete wavelet transforms are applied to extract multiscale features that discriminate changed and unchanged areas and Markovian data fusion is used to integrate both these features and the spatial contextual information in the change-detection process. Unsupervised statistical learning methods (expectation-maximization and Besag's algorithms) are used to estimate the model parameters. Experiments on burnt-forest area detection in multitemporal Landsat TM images are presented.**

***Keywords*-Multiscale change detection, unsupervised change detection, discrete wavelet transforms, Markov random fields, expectation-maximization, Besag's algorithm.**

I. Introduction

Multitemporal remote sensing represents a powerful source of information in applications such as environmental monitoring or environmental disaster management. A key problem in such applications is the identification of the changes that occurred in a given area between two observation dates [1]. When adopting an unsupervised approach, i.e., when assuming no training data to be available at any acquisition date, image differencing and image ratioing are usually applied to address this task with optical and synthetic aperture radar (SAR) images, respectively [1]. Manual [1] or automatic (e.g., Bayesian) [2]–[4] thresholding algorithms can then be used to distinguish changed and unchanged areas in the difference and ratio images. In order to improve the accuracy of the change maps, a multiscale approach can be adopted, in which transformed images at different scales are exploited [5], [6]. Data at finer scales are likely to highlight many geometrical details, but also to be more affected by noise; coarser-scale transforms 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 strategy. Multiscale change-detection methods for remote-sensing images have been proposed in [7], [6], and [5] by using stationary wavelet transforms, information-theoretic change measures, and object-oriented analysis, respectively.

In the present paper, a multiscale contextual unsupervised change-detection method is proposed for optical images, based on discrete wavelet transforms (DWTs) [8] and Markov random fields (MRFs) [9]. DWTs are applied to the difference image to extract multiscale features. They provide two-dimensional multiresolution signal decompositions that allow the input (difference) image to be expanded as the superposition of several components, each highlighting the image content at a given scale [8]. MRFs jointly permit spatial context to be introduced into pixel labeling problems and different information sources to be fused by suitably defining "energy functions" [10]. Here, an MRF-based approach is proposed by modifying the method introduced in [11] for multiresolution image classification and by combining it with DWTs and image-differencing. The related MRF model is characterized by several internal parameters that are estimated by an unsupervised statistical learning strategy based on the expectation-maximization (EM) [12] and Besag's [9] algorithms. Experimental results and comparisons with a previous single-scale method are presented on Landsat TM images.

II. Methodology

A. Overview of the proposed method

Let us denote by $\mathcal{I}_0$ and $\mathcal{I}_1$ two single-channel accurately coregistered optical images, composed of N pixels each and acquired over the same area at times t_0 and t_1, respectively ($t_1 > t_0$; the reformulation in the multichannel case is straightforward). The change-detection problem is expressed as a binary hypothesis testing problem [13], by marking the "change" and "no-change" hypotheses as H_1 and H_0, respectively. The image-differencing approach, which generates a difference image $\mathcal{D}$ by subtracting pixel-by-pixel the pixel intensities in $\mathcal{I}_0$ by those in $\mathcal{I}_1$, is used [1]. The key idea of the proposed method lies in generating a finite set $\{\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_S\}$ of multiscale features by applying DWT to $\mathcal{D}$, and to fuse this multiscale information by an MRF

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.102

approach. MRFs also allow the spatial contextual information to be exploited, thus gaining robustness against noise. The MRF classifier proposed in [11] for images acquired at different spatial resolutions is generalized here to multiscale change detection. It adopts the "iterated conditional mode" (ICM) approach to MRF-based classification, which usually represents a good tradeoff between classification accuracy and computational burden [10]. 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 [2] and based on the Kittler and Illingworth (K&I) unsupervised Bayesian thresholding algorithm [14].

B. Multiscale feature extraction

Let S be a predefined number of scales; an S-scale multiresolution decomposition of $\mathcal{D}$ is obtained by applying a dyadic DWT to decompose $\mathcal{D}$ in terms of a low-pass transformed image and of three high-pass transformed images conveying information about fine-scale details along the horizontal image axis, the vertical axis, or both axes. The procedure is recursively applied S times to the resulting low-pass components. Then, in order to generate the set $\{\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_S\}$, the inverse transform is applied S times by neglecting all high-pass components and keeping only the low-pass terms [8]. Thus, as s increases in $[1, S]$, coarser-scale approximations $\mathcal{W}_s$ of $\mathcal{D}$ are obtained, while the finest scale is represented by $\mathcal{D}$ itself (which will also be denoted by $\mathcal{W}_0$ in the following). More precisely, $\mathcal{W}_s$ allows appreciating spatial details that are (2^s)-times coarser than $\mathcal{D}$ $(s = 1, 2, \ldots, S)$.

C. Unsupervised MRF-based learning

For each k-th pixel, let $\ell_k \in \{H_0, H_1\}$ be the related hypothesis label, u_{sk} be the corresponding pixel intensity in $\mathcal{W}_s$, and x_k be the multiscale vector of all u_{sk} variables $(s = 0, 1, \ldots, S; k = 1, 2, \ldots, N)$. Similar to the multiresolution approach in [11], the statistical relationships between images at different scales are modeled in terms of "linear mixtures." In [11] a set of 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 [11]. Here, this approach is generalized to multiscale analysis by postulating that, for the image $\mathcal{W}_s$ at each s-th scale, a set $\{\tilde{u}_{sk}\}_{k=1}^{N}$ of virtual pixel intensities exist at the 0-th scale, such that:

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

where $Q(s, k)$ is a $2^s \times 2^s$ window centered on the k-th pixel $(k = 1, 2, \ldots, N; s = 0, 1, \ldots, S)$. If $s = 0$, the identity $\tilde{u}_{0k} = u_{0k}$ holds. Since $\mathcal{W}_s$ is obtained from $\mathcal{W}_0$ through a cascade of linear filters, the pixel values in $\mathcal{W}_s$ could be deterministically expressed as suitable linear combinations of the pixel values in $\mathcal{W}_0$, provided a specific DWT operator is chosen. In this perspective, introducing the random variables $\tilde{u}_{sk}$ may not be needed. Indeed we propose to adopt the formalization based on such virtual intensities, first, because this allows extending the Markovian formulation and the learning procedures developed in [11] for the multiresolution case. Then, this choice prevents the need to predefine a specific DWT operator at each scale and plug the analytical expressions of the impulse responses of the related filters in the parameter-estimation process, thus allowing any DWT operator to be used in the proposed MRF framework and ensuring a higher flexibility.

Given the label configuration $\{\ell_k\}_{k=1}^{N}$, the variables $\tilde{u}_{sk}$ $(k = 1, 2, \ldots, N)$ are assumed to be conditionally independent and identically distributed, i.e., the joint probability density function (PDF) of all $\tilde{u}_{sk}$ random variables, conditioned to all label variables ℓ_k $(k = 1, 2, \ldots, N)$ can be factorized as follows:

$$p(\tilde{u}_{s1}, \tilde{u}_{s2}, \ldots, \tilde{u}_{sN} | \ell_1, \ell_2, \ldots, \ell_N) = \prod_{k=1}^{N} p(\tilde{u}_{sk} | \ell_k), \tag{2}$$

and the pixelwise conditional PDF $p(\tilde{u}_{sk}|\ell_k = H_i)$ of $\tilde{u}_{sk}$, given $\ell_k = H_i$ $(i = 0, 1)$ is independent of the pixel position $k \in \{1, 2, \ldots, N\}$. This pixelwise PDF is specifically modeled as a Gaussian with mean μ_{si} and variance σ_{si}^2 $(s = 0, 1, \ldots, S; i = 0, 1)$. Let the vector θ collect all μ_{si} and σ_{si} parameters. Thanks to the linearity of DWT and of the model in Eq. (1), also the PDF of u_{sk}, given $\ell_k = H_i$ and the labels ℓ_r of all other pixels r such that $r \in Q(s, k)$, 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 [2]), with mean and variance parameterized by θ and given by [11]:

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

and

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

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

The label configuration $\{\ell_k\}_{k=1}^{N}$ is assumed to be an MRF [9]. Specifically, 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 [11]:

$$P\{\ell_k = H_i | x_k, \ell_r, r \neq k\} = \frac{\exp[-U_k(H_i|\lambda, \theta)]}{\sum_{j=0}^{1} \exp[-U_k(H_j|\lambda, \theta)]}, \tag{5}$$

where:

$$U_k(H_i|\lambda,\theta)=\sum_{s=0}^{S}\left\{\frac{[u_{sk}-\mu_{sik}(\theta)]^2}{\sigma^2_{sik}(\theta)}+\ln\sigma^2_{sik}(\theta)\right\}-\lambda m_{ik} \quad (6)$$

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 ($k=1,2,\ldots,N; i=0,1$), and λ is a positive parameter. The energy function 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 intensities at all scales. 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. (5) 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 $r\in Q(s,k)$ for some $s=1,2,\ldots,S$ (through $\mu_{sik}(\theta)$ and $\sigma_{sik}(\theta)$). The Potts model has been proved to be an effective contextual MRF model for medium-resolution images [11]; accordingly, the proposed method is expected to be well-tailored to change-detection with this kind of remote-sensing data. The extension of the model in Eqs. (5) and (6) to very high resolution (VHR) images could be accomplished by replacing this simple contextual model by more sophisticated MRFs, incorporating geometrical information related, for instance, to edge [15], [16] or object/regions [17] in the image.

We recall that the MRF-based method in [11] assumes the single-resolution images to be independent of one another, when conditioned to the label configuration. In fact, single-scale energy contributions are combined additively in Eq. (6) without modeling the correlations among transforms at different scales. Since $\mathcal{W}_{s+1}$ is obtained by filtering $\mathcal{W}_s$ ($s=0,1,\ldots,S-1$), the conditional independence assumption cannot be rigorously fulfilled in the present case. We will accept it as a simplifying analytical tractability assumption.

As in [11], the EM and the Besag's methods are iteratively used to learn optimal values for the parameters (λ,θ) from the collections $\{\mathcal{W}_s\}_{s=0}^{S}$ of multiscale images. 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) [12]. The Besag's method estimates the spatial regularization parameters of an MRF (here, λ) by maximizing a pseudolikelihood function [9]. The proposed method is iterative and is initialized with the label map generated by the single-scale K&I algorithm. At the t-th iteration, denoting by the superscript "t" the current parameter estimates and pixel labels, the following operations are performed ($t=0,1,2,\ldots$):

1) compute the updated estimate λ^t, given the current labels $\{\ell_k^t\}_{k=1}^N$, by maximizing the following pseudo-likelihood function (Besag's method) [9]:
$$\Phi^t(\lambda)=\prod_{i=0}^{1}\prod_{k:\ell_k^t=H_i}\frac{\exp(\lambda m_{ik}^t)}{\exp(\lambda m_{0k}^t)+\exp(\lambda m_{1k}^t)}; \quad (7)$$
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 [10], by setting ℓ_k^{t+1} as the label H_i that corresponds to the lowest value of $U_k^t(H_i|\lambda^t,\theta^t)$ ($k=1,2,\ldots,N; i=0,1$).

At each t-th iteration of the method ($t=0,1,\ldots$), the numerical maximization in step 1 is solved by Newton-Raphson's algorithm. 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 $\{\bar\theta^{qt}\}_{q=0,1,\ldots}$ of estimates of θ. Thanks to the convergence properties of EM, $\bar\theta^{qt}$ converges for $q\to+\infty$ 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 $\bar\theta^{0t}=\theta^{t-1}$ ($t=1,2,\ldots$). Thanks to the close similarity between the proposed multiscale linear-mixture model of Eq. (1) and the multiresolution one used in [11], one can prove through the same analytical calculations as in [11] that the components of $\bar\theta^{qt}$ are updated as follows ($i=0,1; s=0,1,\ldots,S$):

$$\bar\mu_{si}^{q+1,t}=\frac{1}{N_i^t}\sum_{k:\ell_k^t=H_i}\eta_{sk}^{qt}$$
$$\bar\sigma_{si}^{q+1,t}=\sqrt{\frac{1}{N_i^t}\sum_{k:\ell_k^t=H_i}[\xi_{sk}^{qt}+(\eta_{sk}^{qt}-\bar\mu_{si}^{q+1,t})^2]}, \quad (8)$$

where $\eta_{0k}^{qt}=u_{0k}$ and $\xi_{0k}^{qt}=0$ for $s=0$,

$$\eta_{sk}^{qt}=\bar\mu_{sj}^{qt}+4^{-s}[\sigma_{sjk}^{-1}(\bar\theta^{qt})\bar\sigma_{sj}^{qt}]^2[u_{sk}-\mu_{sjk}(\bar\theta^{qt})]$$
$$\xi_{sk}^{qt}=(\bar\sigma_{sj}^{qt})^2\{1-16^{-s}[\sigma_{sjk}^{-1}(\bar\theta^{qt})\bar\sigma_{sj}^{qt}]^2\}, \quad (9)$$

for $s\geq 1$, j is the index of the hypothesis H_j such that $\ell_k^t=H_j$, and N_i^t is the number of pixels assigned to H_i at the current iteration. Analytical details can be found in [11].

III. Experimental results

A pair of coregistered 256×256 multispectral Landsat-5 Thematic Mapper (TM) images of the western part of the Island of Elba (Italy), acquired 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 used in the experiments, as they are known from the literature to be effective for burnt forest-area detection [2], [18], [19].

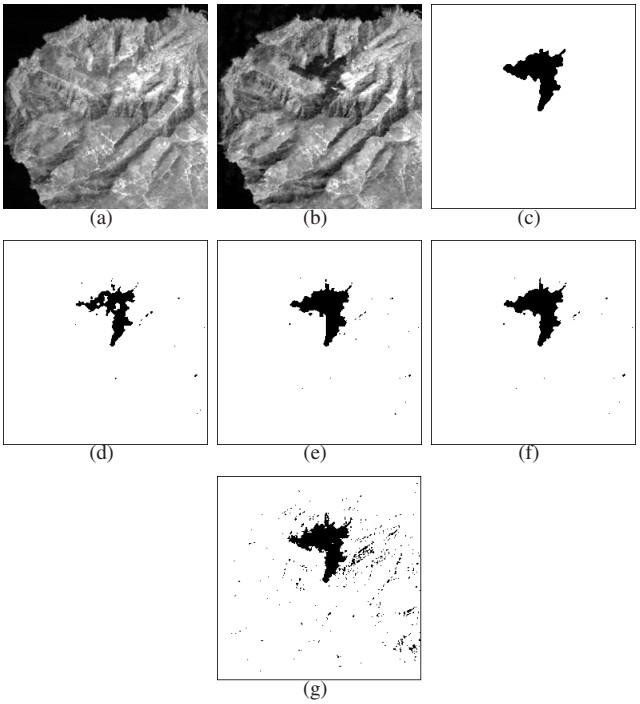

Figure 1. Band TM-4: images acquired in August (a) and September 1994 (b) (after histogram stretching); test map (c); change maps obtained by the technique in [2] (d) and by the proposed method applied with neighbor averaging (e) and with the biorthogonal transform of order $(2, 8)$ (with (f) or without (g) the spatial energy term). Legend: black = "change," white = "no-change."

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 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. Hence, its performances were first evaluated as a function of the choice of the DWT operator. Then experimental comparisons were conducted with: (i) the technique proposed in [2] (which is based on the application of K&I to the moving-average filtered difference image) to assess the overall advantages of the proposed method as compared to a single-scale classical one; (ii) a variant of the proposed technique based on an MRF model with no spatial energy contribution (i.e., $\lambda = 0$) to focus on the relative role of multiscale and contextual information; (iii) the behavior of the technique as a function of the number of scales.

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) with respect to the exhaustive test map in Fig. 1(c). Each TM band was transformed to generate up to four lower-scale images. As a preliminary experiment, aimed at assessing the behavior of the method with a very simple multiscale feature extraction, a "neighbor averaging" pyramid decomposition was used that generates $\mathcal{W}_{s+1}$ by trivially mosaic averaging the pixel intensities of neighboring pixels in $\mathcal{W}_s$ ($s = 0, 1, 2, 3$). Then, several DWTs were considered, namely, Haar, Daubechies of order ranging in $[2, 30]$, symlet of order ranging in $[2, 20]$, biorhogonal 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 [8]).

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 the neighbor-averaging approach to band 4 (see Table I). A large 19.35% increase in the detection accuracy and a 0.63% reduction in the error rate were achieved as compared with the single-scale approach in [2]. This improved effectiveness 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. 1(d)-1(e)). However, Fig. 1(e) also points out the slight presence of undesired blocky artifacts, due to neighbor-averaging. Lower accuracies, even though better than the ones of the method in [2], were obtained by the proposed technique when applied to band 7 with neighbor-averaging.

The detection accuracies of the change maps obtained by the proposed method, when applied to band 4 with all above-mentioned DWTs, are summarized in Fig. 2 (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 the method in [2]. The numerical results suggested that the biorthogonal and reverse biorthogonal families provided the highest detection accuracies. Table I 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 an order-$(2, 8)$ biorthogonal transform, thus gaining a 20.22% detection accuracy as compared with the method in [2]. A visual analysis of the resulting change maps (see Fig. 1(f)) 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 on 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.

Considering, for each band, the DWT providing the

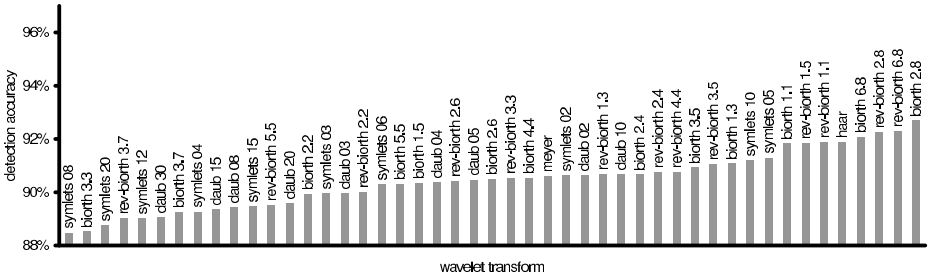

Figure 2. Band 4 of the considered data set: detection accuracy of the proposed method as a function of the choice of the wavelet transform. For wavelets depending on order parameters, the parameter values are also reported.

highest detection accuracy, the proposed method was applied with all five scales, while removing the Potts spatial energy term (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 I). 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. 1(g)) confirms the presence of many false alarms when no spatial energy term is used and also points out that spatially more regular maps are obtained when including this term, especially at the borders between changed and unchanged areas. 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 areas, respectively. EM converges to a solution that correctly detects the burnt vegetation as "change" and the 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 the generation of such false alarms. This further confirms the importance of the adopted MRF approach as a contextual data-fusion tool.

Finally, focusing again on the DWT giving the highest detection accuracy, the number of scales was varied in the range $[1, 5]$ (see Fig. 3). 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 slighlty higher accuracies were obtained with two scales on this data set. This confirms the importance of multiscale information (with, at least, two scales) in the proposed change-detection method and also suggests quite a limited sensitivity to the number of scales.

Table I
CHANGE-DETECTION PERFORMANCES OF THE PROPOSED METHOD APPLIED WITH SEVERAL DWTS (THE VALUES OF POSSIBLE ORDER PARAMETERS ARE IN PARENTHESES), OF THE VARIANT WITH NO SPATIAL ENERGY ($\lambda = 0$), AND OF THE METHOD IN [2].

TM band	Method	False-alarm rate	Detection accuracy	Overall error rate
4	method in [2]	0.08%	72.49%	1.09%
	neighbor averaging	0.17%	91.84%	0.47%
	biorth. $(2, 8)$	0.15%	92.71%	0.41%
	rev. biorth. $(6, 8)$	0.15%	92.29%	0.43%
	symlet (5)	0.19%	91.30%	0.50%
	biorth. $(2, 8)$, $\lambda = 0$	1.79%	95.53%	1.89%
7	method in [2]	0.92%	71.29%	1.94%
	neighbor averaging	0.66%	77.92%	1.45%
	rev. biorth. $(3, 7)$	0.51%	82.23%	1.15%
	symlet (4)	0.54%	78.75%	1.31%
	Daubechies (12)	0.54%	78.62%	1.30%
	rev. biorth. $(3, 7)$, $\lambda = 0$	1.59%	81.11%	2.23%

IV. CONCLUSIONS

The problem of change detection with optical multitemporal images has been addressed in this paper by proposing an unsupervised multiscale contextual method based on wavelet multiscale feature extraction and Markovian data fusion and learning. The approach proposed in [11] for multiresolution classification has been extended to multiscale change detection. Experiments on real optical images acquired before and after a forest fire suggested a good effectiveness of the proposed technique, which generated accurate change maps, outperforming a previously proposed single-scale approach based on the application of unsupervised thresholding to the noise-filtered difference image [2]. 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.

More accurate change maps were obtained with all consid-

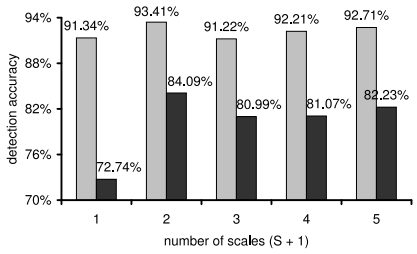

Figure 3. Behavior of the detection accuracy as a function of the number of considered scales (grey: band 4; black: band 7).

ered wavelet transforms, as compared with the single-scale approach in [2]. 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 multiscale transform. A further extension will consist in validating the method with other data sets related to different typologies of land-cover change and different applications (e.g., flooded-area mapping or urban-development monitoring).

The method has been tested on single-channel images. The generalization of the method to multispectral images is analytically straightforward [11]. 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 [13]. Application-specific band combinations [19] or more general feature-reduction methods [13] can be used to address these issues. This would be an interesting aspect worth being investigated. A further development of this work could regard its generalization to change detection with VHR optical images by suitably extending the related MRF model, for instance, through line processes [15], [16] and/or object-based analysis [20]. Extension to VHR SAR (e.g., COSMO/SkyMed or TerraSAR-X) could also be pursued by incorporating parametric PDF estimation methods developed for SAR change detection [4] into the proposed multiscale MRF framework.

REFERENCES

[1] A. Singh, "Digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, pp. 989–1003, 1989.

[2] F. Melgani, G. Moser, and S. B. Serpico, "Unsupervised change detection methods for remote sensing images," *Opt. Eng.*, vol. 41, no. 12, pp. 3288–3297, 2002.

[3] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 4, pp. 874–887, 2005.

[4] G. Moser and S. B. Serpico, "Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 40, no. 10, pp. 2972–2982, 2006.

[5] O. Hall and G. J. Hay, "A multiscale object-specific approach to digital change detection," *Int. J. Appl. Earth Obs.*, vol. 4, no. 4, pp. 311–327, 2003.

[6] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Trans. Geosci. Remote Sensing*, vol. 45, no. 5, 2007.

[7] F. Bovolo and L. Bruzzone, "A detail-preserving scale-driven approach to change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 12, 2005.

[8] I. Daubechies, *Ten lectures on wavelets*. SIAM, 1992.

[9] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Statist. Soc.*, vol. 68, pp. 259–302, 1986.

[10] A. H. S. Solberg, T. Taxt, and A. K. Jain, "A Markov random field model for classification of multisource satellite imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, no. 1, pp. 100–113, 1996.

[11] G. Storvik, R. Fjortoft, and A. H. S. Solberg, "A bayesian approach to classification of multiresolution remote sensing data," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 3, pp. 539–547, 2005.

[12] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data and the EM algorithm," *J. R. Statist. Soc.*, no. 39, pp. 1–38, 1977.

[13] K. Fukunaga, *Introduction to statistical pattern recognition*. Academic Press, 1990.

[14] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recogn.*, vol. 19, pp. 41–47, 1986.

[15] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 721–741, 1984.

[16] G. Moser and S. B. Serpico, "Edge-preserving classification of high-resolution remote-sensing images by Markovian data fusion," in *Proc. of IGARSS-2009, 12-17 July 2009, Cape Town, South Africa (in print)*, 2009.

[17] ——, "Classification of high-resolution images based on MRF fusion and multiscale segmentation," in *Proc. of IGARSS-2008, 6-11 July 2008, Boston, (MA, USA)*, 2008, pp. II.277–II.280.

[18] J. Richards and X. Jia, *Remote sensing digital image analysis*. Springer, 2005.

[19] O. Zammit, X. Descombes, and J. Zerubia, "Combining one-class support vector machines and hysteresis thresholding: application to burnt area mapping," in *Proc. EUSIPCO-2008, Lausanne (Switzerland)*, 2008.

[20] G. G. Hazel, "Multivariate Gaussian MRF for multispectral scene segmentation and anomaly detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 3, pp. 1199–1211, 2000.