Automatic Analysis of the Difference Image for Unsupervised Change Detection

Lorenzo Bruzzone, *Member, IEEE*, and Diego Fernàndez Prieto, *Student Member, IEEE*,

***Abstract*—One of the main problems related to unsupervised change detection methods based on the "difference image" lies in the lack of efficient automatic techniques for discriminating between changed and unchanged pixels in the difference image. Such discrimination is usually performed by using empirical strategies or manual trial-and-error procedures, which affect both the accuracy and the reliability of the change-detection process. To overcome such drawbacks, in this paper, we propose two automatic techniques (based on the Bayes theory) for the analysis of the difference image. One allows an automatic selection of the decision threshold that minimizes the overall change detection error probability under the assumption that pixels in the difference image are independent of one another. The other analyzes the difference image by considering the spatial-contextual information included in the neighborhood of each pixel. In particular, an approach based on Markov Random Fields (MRF's) that exploits interpixel class dependency contexts is presented. Both proposed techniques require the knowledge of the statistical distributions of the changed and unchanged pixels in the difference image. To perform an unsupervised estimation of the statistical terms that characterize these distributions, we propose an iterative method based on the Expectation-Maximization (EM) algorithm. Experimental results confirm the effectiveness of both proposed techniques.**

***Index Terms*—Change detection, change vector analysis, difference image, multitemporal images, remote sensing.**

I. Introduction

In THE past few years, there has been a growing interest in the development of automatic change-detection techniques for the analysis of multitemporal remote sensing images [1]–[6]. This interest stems from the wide range of applications in which change detection methods can be used, like environmental monitoring [7], agricultural surveys [5], urban studies [1], forest monitoring [2], [8], [9], etc.

Usually, change detection involves the analysis of two registered multispectral remote sensing images acquired in the same geographical area at two different times. Such an analysis aims at identifying land cover changes that have occurred in the study area between the two times considered. In the remote sensing literature, two main approaches to the change-detection problem have been proposed: the supervised approach and the unsupervised approach [5], [10]. The former is based on supervised classification methods, which require the availability of a multitemporal ground truth in order to derive a suitable training set for the learning process of the classifiers. The latter performs change detection by making a direct comparison of the two multispectral images considered, without relying on any additional information. Although the supervised approach exhibits some advantages over the unsupervised one (e.g., capability to explicitly recognize the kinds of land cover transitions that have occurred, robustness to the different atmospheric and light conditions at the two acquisition times, ability to process multisensor/multisource images [5]), the generation of an appropriate multitemporal ground truth is usually a difficult and expensive task. Consequently, the use of effective unsupervised change-detection methods is fundamental in many applications in which a ground truth is not available.

In this paper, we focus on one of the most widely used types of unsupervised change-detection techniques, which are based on the so-called "difference image" [10], [11]. These techniques process the two multispectral images acquired at two different dates (or vegetation indexes [10], principal components [10], etc., derived from such images) in order to generate a further image. The computed difference image is such that the values of the pixels associated with land cover changes present values significantly different from those of the pixels associated with unchanged areas. Changes are then identified by analyzing (e.g., thresholding) the difference image. For example, the univariate image differencing technique [10], [11] generates the difference image by subtracting, pixel by pixel, a single spectral band of the two multispectral images under analysis. The choice of the spectral band depends on the specific type of change to be detected. An analogous concept is applied by the widely used change vector analysis (CVA) technique. In this case, several spectral channels are used at each time. For each pair of corresponding pixels, a "spectral change vector" is computed as the difference between the feature vectors at the two times. Then, the pixel values in the difference image are associated with the modules of the spectral change vectors. It follows that unchanged pixels present small gray-level values, whereas changed pixels present rather large values. Other techniques, like image ratioing, produce the difference image by computing the ratio, instead of the difference, between multitemporal images [10].

In spite of their relative simplicity and widespread use, the aforementioned change-detection methods exhibit a major drawback: a lack of automatic and nonheuristic techniques for the analysis of the difference image. In fact, in classical techniques, such an analysis is performed by thresholding the difference image according to empirical strategies [12] or manual trial-and-error procedures, which significantly affect the reliability and accuracy of the final change-detection map. In particular, the most widely used approach to the selection of

Manuscript received January 26, 1999; revised October 25, 1999. This work was supported by the Italian Space Agency (ASI), Ispra, Italy.

L. Bruzzone is with the University of Trento, Trento, Italy (e-mail: lorenzo.bruzzone@ing.unitn.it).

D. F. Prieto is with the Department of Biophysical and Electronic Engineering, University of Genoa, Genoa, Italy.

Publisher Item Identifier S 0196-2892(00)04130-9.

0196–2892/00$10.00 © 2000 IEEE

the decision threshold is based on the assumption (reasonable but not always verified) that only a few changes occurred in the study area between the two dates considered. Under this hypothesis, the density function of the pixel values in the difference image can be confused with the density function of the unchanged pixels. According to this assumption, pixels having gray-level values significantly different from the mean of the density function of the difference image are labeled as changed. In particular, the decision threshold is fixed at $n\sigma_n$ from the mean value of the difference image, σ_D being the standard deviation of the density function of the pixel values in the difference image and n being a real number derived by a trial-and-error procedure. In this context, the selection of the parameter n strongly depends on the end-user's subjective criteria, which may lead to unreliable change-detection results. In addition, such a selection usually requires several trials and hence, a nonnegligible computation time [13], [14].

In this paper, we define the problem of the analysis of the difference image for unsupervised change detection in terms of the Bayes decision theory. The application of this theory requires the estimations of the *a priori* probabilities and of the conditional density functions for the classes associated with the unchanged and changed pixels in the difference image. To this end, we present an approach (based on the Expectation-Maximization algorithm [15]–[17]) that allows such estimations to be performed in an unsupervised way. Within this framework, two automatic techniques for the analysis of the difference image are presented that overcome the main problems inherent in classical techniques. One assumes that the gray-level values of the pixels in the difference image are independent of one another. Under this assumption, the Bayes rule for minimum error is applied in order to select, in an automatic way, the decision threshold that minimizes the overall error probability in the change-detection process. The other technique considers the spatial-contextual information contained in the difference image in order to increase the accuracy of the final change-detection map. In particular, an approach based on Markov Random Fields (MRF's) is proposed that exploits the interpixel class dependence to model the prior probabilities of classes.

In order to assess the effectiveness of both proposed techniques, we carried out experiments on two different data sets. One was a real multitemporal data set composed of two multispectral images acquired by the Thematic Mapper sensor of the Landsat 5 satellite. The other was a synthetic data set generated to evaluate the robustness of the proposed techniques against different levels of noise.

This paper is organized into eight sections. The next section introduces the formulation of the unsupervised change-detection problem in terms of the Bayes theory. In particular, an automatic method for the unsupervised estimation of the statistical terms required by the Bayesian approach is presented. Section III addresses the automatic selection of the minimum error threshold, under the assumption that pixels in the difference image are independent of one another. Section IV deals with the context-based approach to the analysis of the difference image. The data sets used in the experiments are detailed in Section V, together with the experiments carried out. The results obtained in the real and synthetic data sets are reported in Sections VI and VII, respectively. Finally, conclusions are drawn in Section VIII.

II. Approach to the Unsupervised Estimation of Statistical Terms Associated with Classes in the Difference Image

Let us consider two multispectral images, X_1 and $\mathbf{X}_2$ of size $I \times J$ acquired in the same geographical area at two different times, t_1 and t_2. Let us assume that such images have been coregistered [18], [19] and that the possible differences in the light and atmospheric conditions at the two times have been corrected [20]. Let $\mathbf{X}$ be a random variable in the range $[0, 1, \cdots, G-1]$, and let it represent the values of the $I \times J$ pixels in the difference image $\mathbf{X}_D = \{X(i,j), 1 \le i \le I, 1 \le j \le J\}$ obtained by applying the CVA technique to $\mathbf{X}_1$ and $\mathbf{X}_2$. For the sake of simplicity, the two proposed techniques will be presented in the context of the CVA method. However, a generalization to other methods based on the difference image is straightforward.

A. Basic Rationale

Unlike classical unsupervised methods used in remote sensing applications, our approach involves formulating the problem of the analysis of the difference image for change detection in terms of the Bayesian decision theory. Within this framework, we aim at discriminating between two opposite classes, ω_n and ω_c, associated with unchanged and changed pixels, respectively. In order to analyze the difference image on the basis of the Bayes theory, the main problems to be solved are the estimations of both the probability density functions $p(X/\omega_n)$ and $p(X/\omega_c)$ and the *a priori* probabilities $P(\omega_n)$ and $P(\omega_c)$ of the classes ω_n and ω_c, respectively [21]. Generally, these terms are estimated by using supervised approaches that require the availability of a multitemporal ground truth. However, as we deal with an unsupervised approach, the estimation process cannot be performed on the basis of a training set.

In this paper, we propose an unsupervised method for estimating the aforesaid statistical terms. In particular, the method assumes that the probability density function $p(X)$ computed on the pixel values in the difference image $\mathbf{X}_D$ can be modeled as a mixture density distribution consisting of two density components associated with the classes ω_n and ω_c, respectively, i.e.,

$$p(X) = p(X/\omega_n)P(\omega_n) + p(X/\omega_c)P(\omega_c). \tag{1}$$

Under this assumption, the unsupervised estimations of $p(X/\omega_n)$, $p(X/\omega_c)$, $P(\omega_n)$, and $P(\omega_c)$ can be performed by using the EM algorithm.

B. Estimations of $p(X/\omega_n)$, $p(X/\omega_c)$, $P(\omega_n)$, and $P(\omega_c)$ by the EM Algorithm

The EM algorithm is a general approach to maximum-likelihood (ML) estimation for incomplete data problems [15]–[17], [22]. It consists of an expectation step and a maximization step, which are iterated until convergence. The expectation step is computed with respect to the unknown underlying variables, using the current estimates of the parameters, and is conditioned

by observations. The maximization step provides new estimates of the parameters.

We propose using the EM algorithm to estimate the values of the *a priori* probabilities $P(\omega_n)$ and $P(\omega_c)$ and the values of the parameters that characterize the density functions $p(X/\omega_n)$ and $p(X/\omega_c)$. Let us assume that both $p(X/\omega_n)$ and $p(X/\omega_c)$ can be modeled by Gaussian distributions (this is a reasonable assumption for many applications involving images acquired by passive sensors). In this context, the density function associated with the class ω_n can be described by the mean μ_n and the variance σ_n^2. Analogously, the density function associated with the class ω_c can be described by the mean μ_c and the variance σ_c^2. It is possible to prove that the equations for estimating the aforementioned statistical terms for the class ω_n are the following [17]:

$$P^{t+1}(\omega_n) = \frac{\sum_{X(i,j)\in \mathbf{X_D}} \frac{P^t(\omega_n)p^t(X(i,j)/\omega_n)}{p^t(X(i,j))}}{IJ} \tag{2}$$

$$\mu_n^{t+1} = \frac{\sum_{X(i,j)\in \mathbf{X_D}} \frac{P^t(\omega_n)p^t(X(i,j)/\omega_n)}{p^t(X(i,j))} X(i,j)}{\sum_{X(i,j)\in \mathbf{X_D}} \frac{P^t(\omega_n)p^t(X(i,j)/\omega_n)}{p^t(X(i,j))}} \tag{3}$$

$$(\sigma_n^2)^{t+1} = \frac{\sum_{X(i,j)\in \mathbf{X_D}} \frac{P^t(\omega_n)p^t(X(i,j)/\omega_n)}{p^t(X(i,j))} [X(i,j)-\mu_n^t]^2}{\sum_{X(i,j)\in \mathbf{X_D}} \frac{P^t(\omega_n)p^t(X(i,j)/\omega_n)}{p^t(X(i,j))}} \tag{4}$$

where the superscripts t and $t+1$ denote the values of the parameters at the current and next iterations, respectively. Analogous equations are used to estimate the prior probability and the mean and variance values of the conditional density function associated with the class ω_c.

The estimates are obtained by starting from initial values of the considered statistical terms and by iterating the above equations until convergence. It is possible to prove that, at each iteration, the estimated parameters provide an increase in the log-likelihood function $L(\theta) = \ln\ p(\mathbf{X}_D|\theta)$, where $\theta = \{P(\omega_n), P(\omega_c), \mu_n, \mu_c, \sigma_n^2, \sigma_c^2\}$. At convergence, a local maximum of the log-likelihood function is reached [15], [16].

The initial values of the estimates can be determined by exploiting the intrinsic characteristics of the difference image obtained with the CVA technique. In particular, a subset $\mathbf{S}_n$ of pixels likely to belong to ω_n and a subset $\mathbf{S}_c$ of pixels likely to belong to ω_c can be obtained by applying two thresholds, T_n and T_c, to the right and left extremes of the histogram $h(X)$ of the difference image (see Fig. 1). We express T_n and T_c as $T_n = M_D(1-\alpha)$ and $T_c = M_D(1+\alpha)$, where M_D is the middle value of $h(X)$ (i.e., $M_D = [\max\{\mathbf{X}_D\} - \min\{\mathbf{X}_D\}]/2$), and $\alpha \in (0,1)$ is the initialization parameter that defines the range around M_D in which pixels cannot be easily identified as either changed or unchanged. Then the sets $\mathbf{S}_n = \{X(i,j)|X(i,j) < T_n\}$ and $\mathbf{S}_c = \{X(i,j)|X(i,j) > T_c\}$ are used to compute the initial estimates of the statistical parameters associated with the classes ω_n and ω_c, respectively.

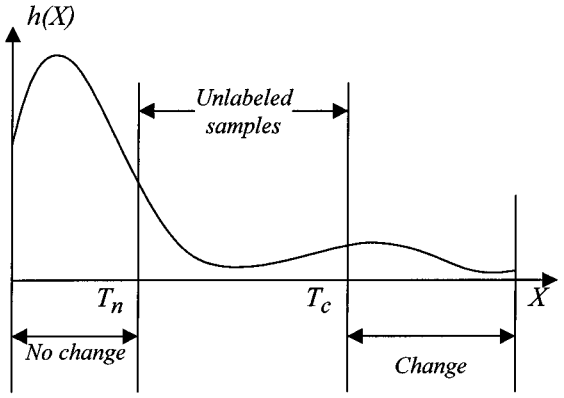


Fig. 1. Schematic representation of the thresholding strategy applied to the difference-image histogram $h(X)$ for the initialization of the EM algorithm.

The previously described formulation of the EM algorithm allows one to estimate the statistical parameters associated with both classes ω_n and ω_c under the assumption of Gaussian distributions. However, it is worth noting that more general approaches to estimating the mixture component parameters might be adopted. In particular, we recall the semiparametric and nonparametric approaches presented in [22] and the generalized mixture estimation technique proposed in [23].

The estimates obtained by the EM algorithm at convergence can be exploited to analyze the difference image with the techniques described in the next two sections.

III. Analysis of the Difference Image Under the Assumption of Independent Pixel Values

In this section, an automatic technique aimed at selecting the decision threshold that minimizes the error probability in the change-detection process is presented. This technique was developed under the assumption that pixel values are independent of one another.

Under the hypothesis of interpixel independence and according to the Bayes rule for minimum error, each pixel $X(i,j)$ in the difference image $\mathbf{X}_D$ should be assigned to the class ω_k that maximizes the posterior conditional probability, i.e.,

$$\begin{aligned}\omega_k &= \arg \max_{\omega_i \in \{\omega_n, \omega_c\}} \{P(\omega_i/X(i,j))\} \\ &= \arg \max_{\omega_i \in \{\omega_n, \omega_c\}} \{P(\omega_i)p(X(i,j)/\omega_i)\}.\end{aligned} \tag{5}$$

Applying this criterion to solve the change-detection problem is equivalent to thresholding the difference image at the ML boundary T_o between the classes ω_n and ω_c. Therefore, on the basis of the estimates of the statistical terms obtained by the EM algorithm, the optimum threshold value T_o can be estimated by solving the following equation with respect to the variable X:

$$\frac{P(\omega_c)}{P(\omega_n)} = \frac{p(X/\omega_n)}{p(X/\omega_c)} \tag{6}$$

which, in the Gaussian case, is equivalent to solving the following quadratic equation:

$$\begin{aligned}&(\sigma_n^2 - \sigma_c^2)\hat{T}_o^2 + 2(\mu_n\sigma_c^2 - \mu_c\sigma_n^2)\hat{T}_o \\ &\quad + \mu_c^2\sigma_n^2 - \mu_n^2\sigma_c^2 - 2\sigma_n^2\sigma_c^2 \ln\left[\frac{\sigma_c P(\omega_n)}{\sigma_n P(\omega_c)}\right] = 0.\end{aligned} \tag{7}$$

It is worth noting that the accuracy of the threshold value $\hat{T}_o$ was obtained and therefore, the precision of the final change-detection map depends on the accuracies of the estimates provided by the EM algorithm.

IV. Analysis of the Difference Image by Considering Spatial-Contextual Information

In this section, we describe an automatic technique that, unlike the most widely used approaches to change detection, explicitly considers spatial-contextual information for the analysis of the difference image. Such a technique is based on the assumption that the changes to be identified are large enough to be detected by the sensor used. Under this hypothesis, a pixel belonging to the class ω_k is likely to be surrounded by pixels belonging to the same class. Therefore, an efficient use of this interpixel class dependence may yield more reliable and accurate change-detection results.

Let the set $\mathbf{C} = \{\mathbf{C}_l, 1 \le l \le L\}$ with $L = 2^{IJ}$ be composed of all the possible sets of labels in the difference image $\mathbf{X}_D$, where $\mathbf{C}_l = \{C_l(i,j), 1 \le i \le I, 1 \le j \le J\}$ with $C_l(i,j) \in \{\omega_n, \omega_c\}$ is a generic set of labels in $\mathbf{X}_D$. By taking into account the spatial-contextual information, the Bayes rule for minimum error, as defined in (5), can be rewritten as the selection of a set $\mathbf{C}_k$ that maximizes the following rule:

$$\begin{aligned}\mathbf{C}_k &= \arg\max_{\mathbf{C}_l \in \mathbf{C}} \{P(\mathbf{C}_l/\mathbf{X}_D)\} \\ &= \arg\max_{\mathbf{C}_l \in \mathbf{C}} \{P(\mathbf{C}_l)p(\mathbf{X}_D/\mathbf{C}_l)\} \end{aligned} \tag{8}$$

where $P(\mathbf{C}_l)$ is the prior model for the class labels, and $p(\mathbf{X}_D/\mathbf{C}_l)$ is the joint density function of the pixel values in the difference image given the set of labels $\mathbf{C}_l$. The maximization of (8) requires the estimations of both $P(\mathbf{C}_l)$ and $p(\mathbf{X}_D/\mathbf{C}_l)$, which are very complex tasks. A simplification of the problem can be achieved if we model the spatial-contextual information in a local spatial neighborhood. This is rather a reasonable approach if we consider the interpixel class dependence as the interactions between pixel classes decrease rapidly as the distances between pixels increase. In this context, we propose the use of an MRF approach to model the spatial context in the prior model for the class labels $P(\mathbf{C}_l)$. In fact, MRF's provide a methodological framework that allows the interpixel class dependence to be fully exploited. As a further simplification of the problem, we assume the following conditional independence:

$$p(\mathbf{X_D}/\mathbf{C_l}) = \prod_{X(i,j) \in \mathbf{X}_D} p(X(i,j)/C_l(i,j)). \tag{9}$$

A. Description of the Considered MRF Model

In order to formulate the problem by using MRF's, it is necessary to introduce the concept of a spatial neighborhood system defining set $\mathbf{N}$. Let us define the neighbor system of the pixel with coordinates (i,j) as $\mathbf{N}(i,j) = \{(i,j)+(\nu,\varsigma), (\nu,\varsigma) \in \mathbf{N}\}$. Although it is possible to use various spatial neighborhood systems, in this paper, we consider a second-order spatial neighborhood system (see Fig. 2). Therefore, in our case, $\mathbf{N} = \{(\pm 1, 0), (0, \pm 1), (1, \pm 1)$, and $(-1, \pm 1)\}$.

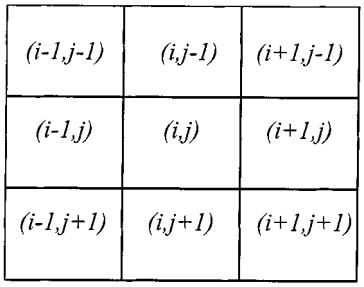


Fig. 2. Second-order neighborhood system defining set used by the considered MRF approach.

The Markov modeling of the conditional distribution of the pixel label $C_l(i,j)$ given the pixel labels elsewhere, is expressed as [24]–[26]

$$\begin{aligned}&P(C_l(i,j)/\{C_l(g,h), (g,h) \ne (i,j)\}) \\ &\quad = P(C_l(i,j)/\{C_l(g,h), (g,h) \in \mathbf{N}(i,j)\}) \\ &\quad = \frac{1}{Z}\ \exp[-U(C_l(i,j)/\{C_l(g,h), (g,h) \in \mathbf{N}(i,j)\})]\end{aligned} \tag{10}$$

where $U(\cdot)$ is the Gibbs energy function, and Z is a normalizing factor. $U(C_l(i,j)/\{C_l(g,h), (g,h) \in \mathbf{N}(i,j)\})$ is given by [24]–[26]

$$\begin{aligned}&U(C_l(i,j)/\{C_l(g,h), (g,h) \in \mathbf{N}(i,j)\}) \\ &\quad = \sum_{(g,h) \in \mathbf{N}(i,j)} \beta \delta_k(C_l(i,j), C_l(g,h))\end{aligned} \tag{11}$$

where δ_k is the Kronecker delta function, which can be expressed as

$$\delta_k(C_l(i,j), C_l(g,h)) = \begin{cases} -1, & \text{if } C_l(i,j) = C_l(g,h) \\ 0, & \text{if } C_l(i,j) \ne C_l(g,h) \end{cases} \tag{12}$$

and β is a constant that tunes the influence of the spatial-contextual information on the change-detection process. It is worth noting that (11) can be regarded as a simplification of the more general clique potential notation adopted by many authors [27]–[29]. For more detailed descriptions of MRF's and of the specific model adopted in this paper, we refer the reader to [24]–[29].

B. Generation of the Change-Detection Map

According to (8), the generation of the final change-detection map involves the labeling of all the pixels in the difference image so that, under the aforementioned assumptions, the posterior probability is maximized. In terms of the Markovian approach, this is equivalent to the minimization of the following energy function [24]–[26]:

$$\begin{aligned}&U(\mathbf{X}_D, \mathbf{C}_l) \\ &\quad = \sum_{1 \le i \le I} \sum_{1 \le j \le J} [U_{data}(X(i,j)/C_l(i,j)) \\ &\qquad + U_{context}(C_l(i,j)/\{C_l(g,h), (g,h) \in \mathbf{N}(i,j)\})].\end{aligned} \tag{13}$$

On the one hand, the energy term $U_{context}(\cdot)$ describes the interpixel class dependence, which is determined according to (11). On the other hand, the term $U_{data}(\cdot)$ represents the statistics

of the gray levels in the difference image under the assumption of conditional independence, as defined in (9). In the Gaussian case, the energy term $U_{data}(\cdot)$ can be written as

$$U_{data}(X(i,j), C_l(i,j)) = \tfrac{1}{2} \ln|2\pi\sigma^2_{C_l(i,j)}| + \tfrac{1}{2} (X(i,j) - \mu_{C_l(i,j)})^2[\sigma^2_{C_l(i,j)}]^{-1} \quad (14)$$

where $\sigma^2_{C_l(i,j)} \in \{\sigma_n^2, \sigma_c^2\}$ and $\mu_{C_l(i,j)} \in \{\mu_n, \mu_c\}$ are the estimates obtained by the EM algorithm under the assumption of independence.

Generally, the minimization of (13) is carried out by using an iterative algorithm (e.g., the simulated annealing algorithm [27]). In this paper, we suggest using a simple and fast approach based on Besag's iterated conditional modes (ICM) algorithm, which has been proved to converge to a local minimum of the energy function [30]. According to this strategy, the $\mathbf{C_l}$ that minimizes (13) is obtained by the following algorithm.

1) For all pixels $X(i,j) \in \boldsymbol{X_D}$, initialize $C_l(i,j)$ with the class that minimizes the noncontextual energy function $U_{data}(X(i,j), C_l(i,j))$.
2) For all pixels $X(i,j) \in \boldsymbol{X_D}$, update $C_l(i,j)$ to the class that minimizes (13).
3) Repeat step 2 until convergence is reached.

V. Descriptions of Data Sets and Experiments

In order to assess the effectiveness of the proposed techniques for the analysis of the difference image, we considered two different data sets: a real multitemporal data set corresponding to the geographical area of the Island of Elba, Italy, and a synthetic data set artificially generated to evaluate the robustness of the proposed techniques to noise. In the following, both the data sets and the carried out experiments are detailed.

A. Data Set Related to the Island of Elba

The first of the two data sets used in the experiments consisted of two multispectral images acquired by the Landsat-5 Thematic Mapper (TM) sensor in the western part of the Island of Elba in August 1994 and September 1994. The area selected for the experiments was a section (414 × 326 pixels) of the two scenes acquired by the TM sensor. As an example of the images used, Fig. 3(a) and (b) show channel 4 of the August and September images, respectively. As is readily apparent (see the upper left parts of the images), a wildfire destroyed a notable portion of the vegetation in the aforesaid area between the two dates considered. The available ground truth concerning the location of the wildfire was used to prepare a "reference map" [see Fig. 3(c)] useful to assess change-detection errors. Such a map was refined by a manual analysis of the remote sensing images considered.

The September image was registered to the August one. The analyses of the histograms of both images did not reveal any significant difference in the light conditions at the times of the two acquisitions. Therefore, no correction algorithms were applied. The noise affecting the intensity values of the images was reduced by applying a simple running mean filtering (3 × 3 window size) to both images.

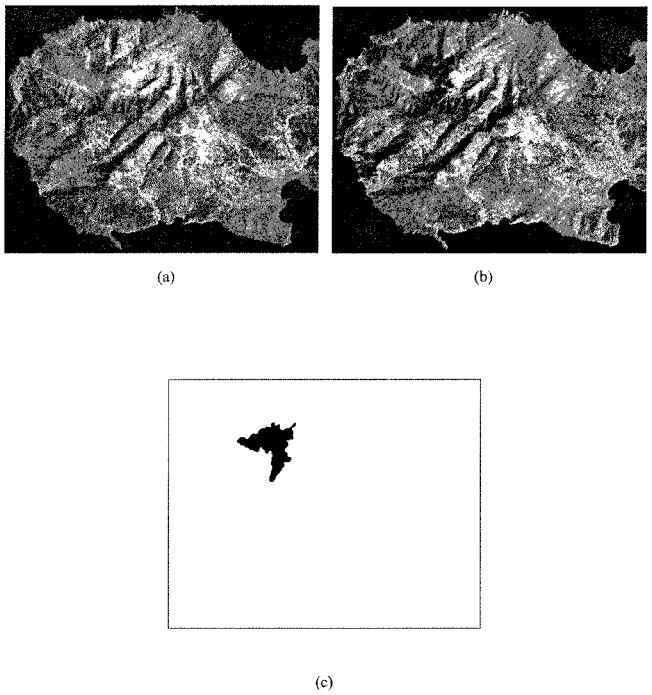


Fig. 3. Images of the Island of Elba, Italy, utilized in the experiments. (a) Band 4 of the Landsat TM image acquired in August 1994, (b) band 4 of the Landsat TM image acquired in September 1994, and (c) ground-truth map of the changed area used as a reference map in the experiments.

B. Synthetic Data Set

The second data set was artificially generated in order to control the noise affecting the difference image. This allowed us to assess more accurately the robustness of the proposed techniques against different levels of noise. The data set was obtained by the procedure described in the following. An image acquired by the Daedalus 1268 Airborne Thematic Mapper (ATM) multispectral scanner [31] was used as the reference image. In particular, a section (250 × 350 pixels) of a scene acquired in an agricultural area near the village of Feltwell, U.K., was selected (for the sake of simplicity, only band 5 of the ATM was considered). This image was assumed to be the t_1 image of the data set (i.e., the image acquired at time t_1). The t_2 image was artificially generated from the reference one. In particular, a first version of the t_2 image was obtained by inserting some changes in the t_1 image in order to simulate land cover variations. Then the histogram of the resulting image was slightly shifted to simulate different light conditions in the two images. Finally, five versions of the t_2 image were generated by adding different realizations of zero-mean Gaussian noise to the t_2 image (the five values of the SNR used were 10, 5, 2, 1, and 0 dB). For simplicity, we assumed the spatial independence of the noise components in the images.

According to the previously described procedure, we obtained five pairs of images. Each pair was composed of the t_1 image and one of the t_2 images (characterized by a specific value of the SNR). As an example, Fig. 4(a) and (b) show the t_1 image and the t_2 image for an SNR = 0 dB, respectively. The map of the areas with simulated changes is presented in Fig. 4(c).

TABLE I
TRUE VALUES OF THE STATISTICAL TERMS OF THE DIFFERENCE IMAGE AND ESTIMATES PROVIDED BY THE PROPOSED APPROACH FOR DIFFERENT VALUES OF THE INITIALIZATION PARAMETER α (DATA SET RELATED TO THE ISLAND OF ELBA)

	α	μ_n	σ_n	$P(\omega_n)$	μ_c	σ_c	$P(\omega_c)$
Proposed approach	0.3	25.31	18.44	0.98	109.62	37.22	0.02
	0.4	25.31	18.44	0.98	109.78	37.17	0.02
	0.5	25.31	18.44	0.98	109.63	37.21	0.02
	0.6	25.30	18.44	0.98	109.31	37.36	0.02
	0.7	25.32	18.46	0.98	110.65	36.83	0.02
True value	-	25.47	18.64	0.98	120.73	30.31	0.02

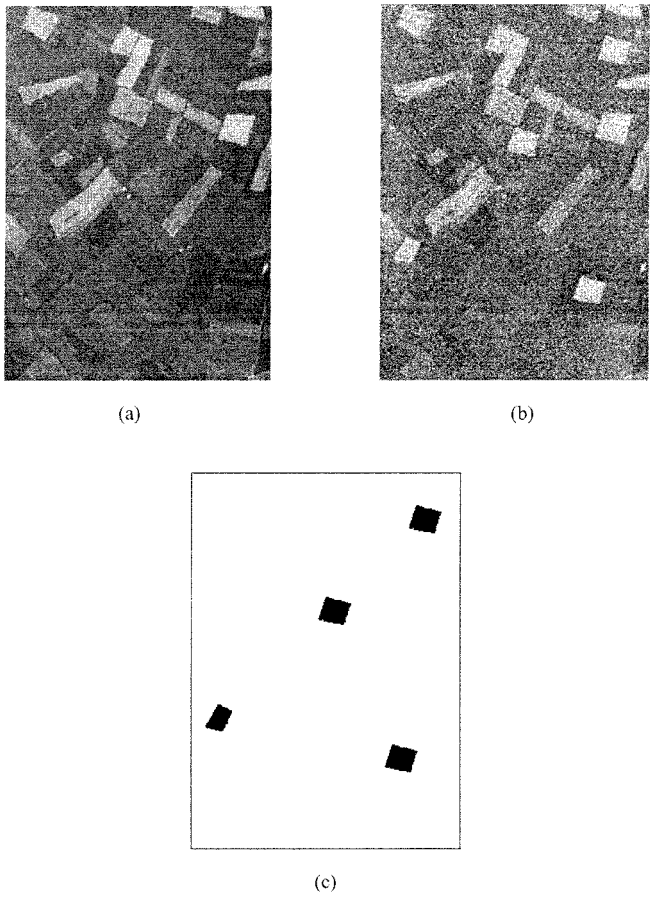


Fig. 4. Synthetic data set utilized in the experiments. (a) t_1 image, (b) t_2 image (for SNR = 0 dB), (c) map of the areas with simulated changes used as the reference map in the experiments.

C. Description of the Experiments

Three different experiments were carried out to test the validity of the proposed techniques.

The first experiment allowed an evaluation of the accuracy and stability of the proposed approach based on the EM algorithm for the estimation of the statistical terms involved in (1). To this end, the true values of the *a priori* probabilities $P(\omega_n)$ and $P(\omega_c)$, as well as the means and standard deviations of the density functions $p(X/\omega_n)$ and $p(X/\omega_c)$, were computed by using the information available in the reference maps. These values were then compared with the estimates obtained by the proposed approach.

The second experiment aimed at assessing the effectiveness of the technique for the analysis of the difference image under the assumption of independent pixel values (see Section III). In particular, the value $\hat{T}_o$ of the decision threshold derived by using the proposed technique was compared with the threshold value T_o that provided the minimum overall change-detection error. The minimum-error threshold T_o was obtained by performing a nonautomatic evaluation of the change-detection errors versus all possible values of the decision threshold. The comparison was made in terms of both the overall change-detection error and the number of false and missed alarms.

The third experiment made it possible to assess the capability of the proposed technique that exploits the spatial-contextual information (see Section IV) to improve the change-detection accuracies provided by the classical thresholding approach. For this evaluation, the results yielded by the application of the presented context-based technique were compared with the results obtained by thresholding the difference image with the minimum error threshold value T_o.

VI. EXPERIMENTAL RESULTS ON THE DATA SET RELATED TO THE ISLAND OF ELBA

Preliminary trials were carried out in order to determine the most effective spectral bands for detecting the burned area in the considered data set. On the basis of the results of these trials and in accordance with the literature [32], we applied the CVA technique to spectral bands 4 and 7 of the images. In fact, such bands turned out to be very effective in locating the burned area.

A. First Experiment: Estimation of the Statistical Terms Associated with the Classes ω_n and ω_c

In order to assess both the accuracy and the stability of the proposed approach to estimating $P(\omega_n), P(\omega_c), \mu_n, \mu_c, \sigma_n^2$, and σ_c^2, different trials were carried out for different values of the initialization parameter α. In particular, we used values of α ranging from 0.3 to 0.7. The estimates obtained are presented in Table I, in which the true values computed on the reference map are also given to allow a comparison. From the analysis of the table, one can deduce that the proposed technique provided quite accurate estimates of the above statistical terms. In particular, despite the estimates of the mean μ_c and of the standard deviation, σ_c are slightly different from the real values, and the estimates of the prior probabilities $P(\omega_n)$ and $P(\omega_c)$ as well as those of the mean μ_n and of the standard deviation σ_n are very close to the corresponding true values. In addition, it is important to point out the high stability of the estimates versus the various values of the initialization parameter α. This confirms

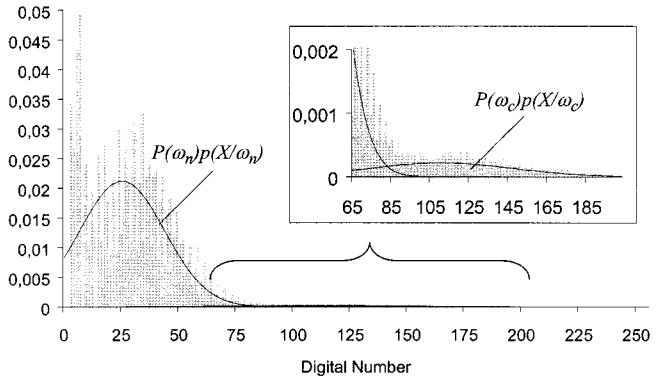


Fig. 5. Histogram of the difference image corresponding to the data set related to the Island of Elba. For the sake of comparison, the estimates of the distributions of both classes ω_n and ω_c obtained by the EM algorithm are superimposed ($\alpha = 0.7$). The histogram was normalized in order to permit a direct comparison with the estimated distributions of the classes.

that the initialization phase is not critical for the proposed approach.

Additional information about the accuracies of the estimates obtained by the EM algorithm can be derived from the analysis of Fig. 5, which shows both the histogram of the difference image and the density functions of the classes ω_n and ω_c for $\alpha = 0.7$ (the histogram was normalized to allow a direct comparison with the estimated density functions). As can be seen, the density functions of the classes derived by the EM algorithm resulted in a fairly reasonable approximation of the distribution of the difference image. This confirms the effectiveness of the proposed technique. It is worth noting that a more accurate estimate of the density function of the class ω_n could be obtained with nonparametric or semiparametric mixture density estimation techniques [22], [23]. However, the levels of complexity inherent in these approaches do not seem to be justified in the present case.

A deeper insight into the behavior of the EM algorithm is made possible by Fig. 6, which presents the trend of each estimate provided by the proposed technique (for $\alpha = 0.7$) versus the number of algorithm iterations for the class ω_n. As can be seen, the estimates evolve from wrong initial values to accurate final ones in only six iterations (similar behaviors were shown by the estimates of the class ω_c).

B. Second Experiment: Analysis of the Difference Image Under the Assumption of Independent Pixel Values

In this experiment, the decision threshold $\hat{T}_o$ estimated by the technique described in Section III was compared with the minimum-error threshold T_o derived by a manual trial-and-error procedure. Thanks to the stability of the EM algorithm, the estimates given in Table I for different α values resulted in the same value (i.e., 82) of the decision threshold $\hat{T}_o$. This value was very close to the minimum-error threshold T_o which, in this case, was equal to 84. As a consequence, our technique involved an overall change-detection error (i.e., 438 pixels) that was very close to the minimum one (i.e., 424 pixels).

A better understanding of the results obtained can be gained by the analysis of Fig. 7. In this figure, the trends of the overall error, false alarms, and missed alarms versus the

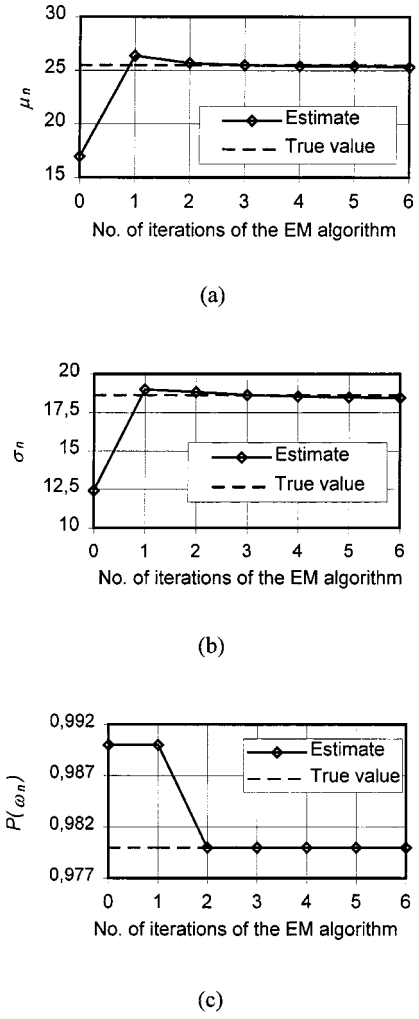


Fig. 6. Estimates of the statistical terms provided by the proposed approach (for $\alpha = 0.7$) versus the number of iterations of the EM algorithm for the class ω_n (data set related to the Island of Elba).

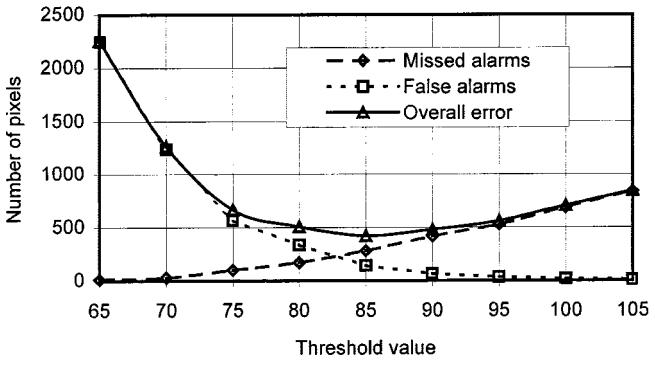


Fig. 7. Behaviors of the change-detection errors (overall error, missed alarms, and false alarms) versus the decision threshold for the data set related to the Island of Elba. The minimum-error threshold T_o was found for a gray-level value equal to 84.

decision-threshold value are plotted. The optimum threshold value corresponds to the point at which the curve of the overall error reaches the minimum value. As one can see, the decision threshold derived with the proposed technique corresponds to a value on the overall-error curve that is very close to the

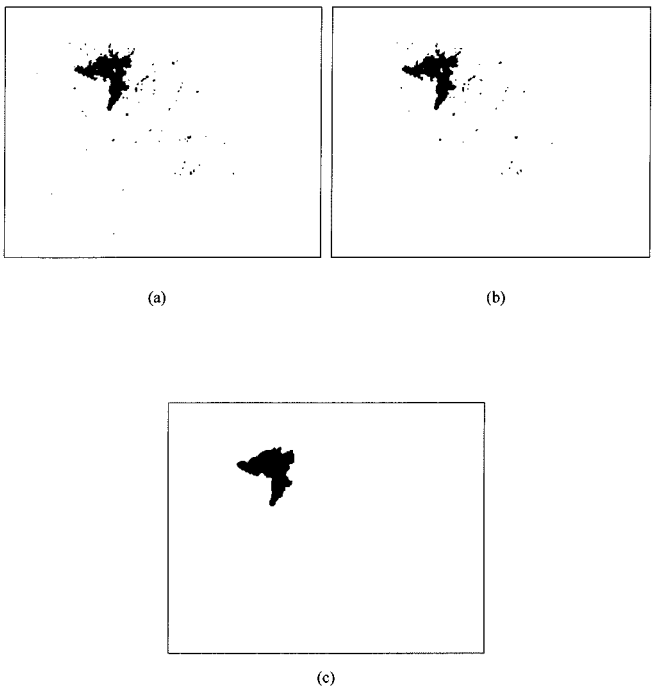


Fig. 8. Change-detection map obtained for the data set related to the Island of Elba by using (a) the proposed technique based on the assumption of independent pixel values, (b) the optimal threshold value provided by a manual trial-and-error procedure under the hypothesis of independent pixel values, and (c) the proposed technique, which exploits the spatial context ($\beta = 1.6$).

minimum one. Concerning the error typology, the proposed technique resulted in 218 false alarms and 220 missed alarms, and the minimum overall-error threshold involved 142 false alarms and 282 missed alarms.

The change-detection maps obtained by using the threshold $\hat{T}_o$ selected by the proposed technique and the minimum-error threshold T_o are shown in Fig. 8(a) and (b), respectively. A comparison of such maps highlights the ability of our technique to generate, in an automatic way, a change-detection map very similar to the best one that can be achieved by a manual trial-and-error procedure.

C. Third Experiment: Analysis of the Difference Image by Considering Spatial Contextual Information

The proposed technique that takes into account the spatial contextual information in the analysis of the difference image was tested by carrying out trials for different values of the parameter β (11). The results obtained are summarized in Table II. As one can see, for each trial carried out, our technique provided a sharp reduction in the overall change-detection error, as compared with the error resulting from the minimum-error threshold T_o (e.g., 176 versus 424 for $\beta = 1.6$). In particular, the numbers of missed and false alarms were significantly reduced. As an example, for $\beta = 1.6$, the number of missed alarms decreased from 282 to only 38, and the number of false alarms reduced from 142 to 138.

Fig. 8(c) shows the change-detection map obtained with the proposed technique for $\beta = 1.6$. A comparison of Fig. 8(c) with Fig. 8(b) confirms that our technique based on the spatial context provided, in an automatic way, a change-detection map that was significantly more accurate (see Fig. 3(c) for a comparison) than the one achieved by using the minimum-error threshold value T_o.

TABLE II
OVERALL ERROR, FALSE ALARMS, AND MISSED ALARMS RESULTING FROM THE PROPOSED CONTEXT-BASED TECHNIQUE FOR DIFFERENT VALUES OF THE PARAMETER β (DATA SET RELATED TO THE ISLAND OF ELBA). FOR THE SAKE OF COMPARISON, THE TABLE ALSO GIVES THE OVERALL ERROR, THE FALSE ALARMS, AND THE MISSED ALARMS ASSOCIATED WITH THE MINIMUM ERROR THRESHOLD FOUND BY A NONAUTOMATIC PROCEDURE UNDER THE ASSUMPTION OF PIXEL INDEPENDENCE

Technique	β	False alarms	Missed alarms	Overall error
Pixel independence (threshold To)	-	142	282	424
Spatial-class dependence	1.6	138	38	176
	1.7	133	40	173
	1.8	124	80	204

VII. EXPERIMENTAL RESULTS ON THE SYNTHETIC DATA SET

The experiments described in Section VI were repeated on the synthetic data set in order to evaluate the performances of the presented techniques versus variations in the level of noise. To this end, for the five pairs of synthetic images considered, the corresponding difference images were obtained by applying the CVA technique. For all the trials carried out on this data set, the initialization parameter α was fixed at 0.5. In addition, when the context-based technique was used, the parameter β was set to 1.3.

A. First Experiment: Estimation of the Statistical Terms Associated With the Classes ω_n and ω_c

Table III shows the results obtained in this experiment for the five SNR values selected. As one can see, in all the trials, the estimates provided by the proposed technique accurately approximate the true values of the considered statistical parameters. In particular, even in the cases characterized by high levels of noise (i.e., SNR $= 1$ dB and SNR $= 0$ dB), the obtained estimates turned out to be very close to the corresponding true values (the largest error concerns the estimate of μ_c for SNR $= 1$ dB, for which a value of about 168.00 was obtained, instead of the true value of 163.78).

In Fig. 9, the histogram of the difference image is compared with the estimates of the density functions of the classes derived by the EM algorithm for SNR = 0 dB. As one can see, in spite of the relative complexity of the problem (the histogram of the difference image does not present two well-separated modes), the estimates achieved with the proposed technique provide an accurate approximation for the density function of the difference image.

This experiment also gives some information about the convergence capabilities of the EM algorithm. In particular, as one can see in Fig. 10, the number of iterations necessary to reach convergence increases with the level of noise affecting the images. In greater detail, for SNR $= 10$ dB, the algorithm reached

TABLE III
TRUE VALUES OF THE STATISTICAL TERMS OF THE DIFFERENCE IMAGES AND ESTIMATES PROVIDED BY THE PROPOSED APPROACH FOR DIFFERENT SNR VALUES (SYNTHETIC DATA SET): (a) CLASS ω_c PARAMETERS AND (b) CLASS ω_n PARAMETERS

SNR (dB)	μ_c		σ_c		$P(\omega_c)$	
	Estimated	True	Estimated	True	Estimated	True
10	224.54	224.10	22.77	23.71	0.03	0.03
5	189.54	189.58	25.36	25.61	0.03	0.03
2	166.92	166.50	26.52	27.44	0.03	0.03
1	168.00	163.78	26.16	28.92	0.03	0.03
0	157.75	154.31	27.28	29.43	0.03	0.03

(a)

SNR (dB)	μ_n		σ_n		$P(\omega_n)$	
	Estimated	True	Estimated	True	Estimated	True
10	99.06	99.06	12.31	12.31	0.97	0.97
5	85.93	85.94	17.96	17.98	0.97	0.97
2	76.85	76.87	21.99	22.02	0.97	0.97
1	75.68	75.61	24.12	24.05	0.97	0.97
0	71.47	71.40	25.53	25.49	0.97	0.97

(b)

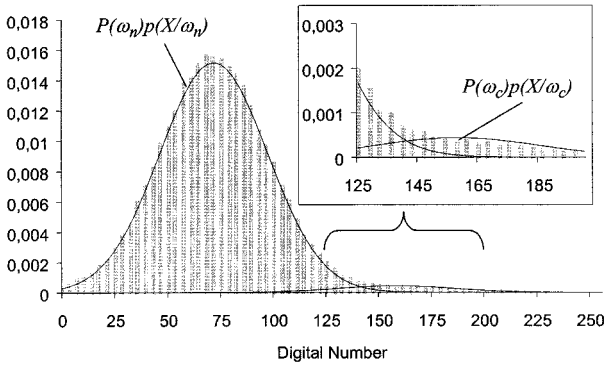


Fig. 9. Histogram of the difference image corresponding to the synthetic data set (SNR = 0 dB). For the sake of comparison, the estimates of the distributions of both classes ω_n and ω_c obtained by the EM algorithm are superimposed. The histogram was normalized in order to permit a direct comparison with the estimated distributions of the classes.

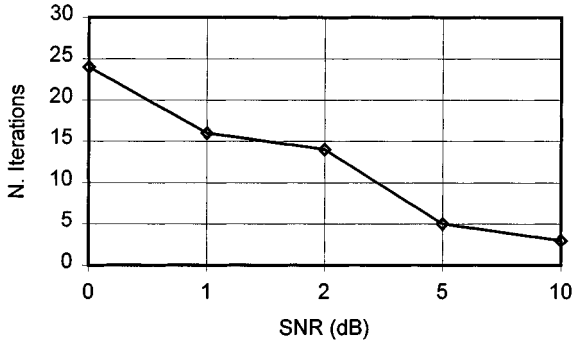


Fig. 10. Number of iterations of the EM algorithm to reach convergence versus the SNR.

convergence in only three iterations, whereas the number of iterations increased up to 26 for SNR $= 0$ dB.

B. Second Experiment: Analysis of the Difference Image Under the Assumption of Independent Pixel Values

Thanks to the accuracy of the estimates provided by the proposed technique in the previous experiment, all the trials carried out for the different SNR values considered resulted in a decision threshold value $\hat{T}_o$ very close to the corresponding minimum-error threshold T_o. In particular, the largest error was incurred for SNR $= 0$ dB, for which the estimated threshold was equal to 144, whereas the minimum-error threshold was found to be equal to 142. As a consequence, the overall change-detection error involved in the proposed technique (i.e., 940 pixels) was comparable to the minimum one (i.e., 899 pixels). In greater detail, the proposed technique (for SNR $= 0$ dB) resulted in 839 and 101 missed and false alarms, respectively, whereas the minimum overall error involved 733 missed alarms and 166 false alarms.

Fig. 11(a) and (b) show the change-detection maps achieved by using the decision threshold $\hat{T}_o$ (selected with the proposed technique) and the minimum-error threshold T_o, respectively, for SNR $= 0$ dB. A comparative analysis of these images points out that, as occurred for the data set related to the Island of Elba, the two maps are very similar also for this data set. This confirms the reliability of the proposed automatic technique.

C. Third Experiment: Analysis of the Difference Image by Considering Spatial Contextual Information

The results obtained in this experiment (see Table IV) point out the validity of the presented context-based technique. In particular, also on this data set and in all the trials carried out, the overall change-detection error was reduced, as compared with

TABLE IV
OVERALL ERROR, FALSE ALARMS, AND MISSED ALARMS RESULTING FROM THE PROPOSED CONTEXT-BASED TECHNIQUE FOR DIFFERENT SNR VALUES (SYNTHETIC DATA SET). FOR THE SAKE OF COMPARISON, THE TABLE ALSO GIVES THE OVERALL ERROR, THE FALSE ALARMS, AND THE MISSED ALARMS ASSOCIATED WITH THE MINIMUM-ERROR THRESHOLD FOUND BY A NONAUTOMATIC PROCEDURE UNDER THE ASSUMPTION OF PIXEL INDEPENDENCE

SNR (dB)	Spatial-class dependence			Pixel independence (threshold To)		
	False alarms	Missed alarms	Overall error	False alarms	Missed alarms	Overall error
10	0	5	5	1	9	10
5	0	9	9	42	95	137
2	3	26	29	174	328	502
1	1	89	90	154	531	685
0	5	152	157	166	733	899

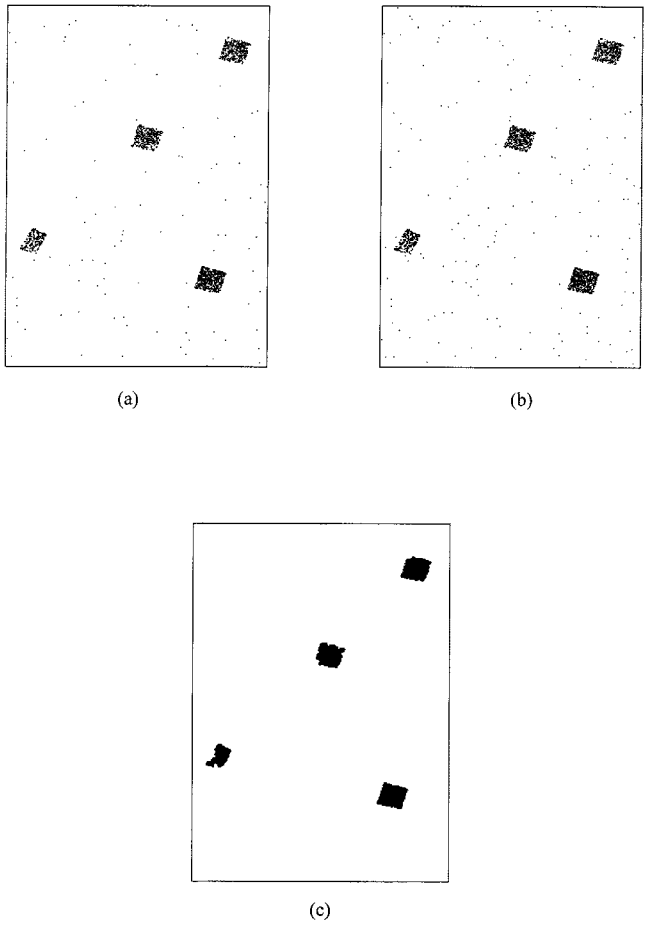


Fig. 11. Change-detection map obtained for the synthetic data set (SNR = 0 dB) by using (a) the proposed technique based on the assumption of independent pixel values, (b) the optimal threshold value provided by a manual trial-and-error procedure under the hypothesis of independent pixel values, and (c) the proposed technique that exploits the spatial context.

the one incurred when using the corresponding minimum-error threshold T_o under the pixel-independence assumption (the reduction is sharper for an increasing level of noise). For example, for SNR = 0 dB, the overall error made with the context-based technique was equal to 157 pixels, whereas the error made for the minimum error threshold T_o was found to be equal to 889 pixels. In greater detail, the number of missed alarms decreased from 733 to 152, and the number of false alarms was reduced from 166 to 5.

Fig. 11(c) shows the change-detection map resulting from the application of the proposed context-based technique in the case of SNR = 0 dB. A comparative analysis of Fig. 11(b) and (c) and Fig. 4(c) confirms the effectiveness of this technique, which provides a change-detection map that is more accurate and significantly less noisy than the one achieved for the minimum-error threshold T_o.

VIII. CONCLUSIONS

In this paper, two techniques for the analysis of the difference image in unsupervised change-detection problems have been proposed. Such techniques, unlike classical ones, perform an automatic analysis of the difference image by exploiting theoretically well-founded methods.

From a methodological viewpoint, the main innovation of this paper lies in the formulation of the unsupervised change-detection problem in terms of the Bayesian decision theory. In particular, we have proposed an iterative technique (based on the EM algorithm) that allows unsupervised estimations of the *a priori* probabilities and density functions associated with changed and unchanged pixels in the difference image. Such estimates make it possible to apply supervised methods in the context of unsupervised change detection. Within this framework, two automatic techniques for the analysis of the difference image have been presented.

The first technique is based on the assumption that the pixels in the difference image are independent of one another. Under this assumption, it allows the automatic selection of the decision-threshold value that minimizes the overall change-detection error probability. It is worth noting that, thanks to the availability of the estimates provided by the EM algorithm, other decision strategies could also be adopted for the selection of the threshold (e.g., the Bayes rule for minimum cost [33], [34]).

The second technique performs the analysis of the difference image by using an MRF approach that exploits the interpixel class dependency context in order to improve the accuracy of the final change-detection map. For the sake of simplicity, a simple method for the MRF modeling has been used, even though more complex MRF models might be exploited (e.g., hierarchical MRF's [28] and detail-preserving MRF's [35]). In addition, more sophisticated MRF approaches might be adopted in order to further increase the accuracy and the degree of automation (e.g., automatic selection of the parameter β) of the

presented technique [36], [37]. Further research should be conducted to test the potential improvements associated with such approaches.

The experimental results reported in this paper confirm the effectiveness of both presented techniques. Such effectiveness depends mainly on the accuracy and the stability provided by the EM algorithm in estimating the statistical terms of the difference image. Thanks to this accuracy, the decision-threshold values provided by the proposed technique based on the pixel-independence assumption turned out to be very close to the optimum ones for both considered data sets. This resulted in change-detection maps comparable to those provided by the corresponding minimum-error threshold T_o. Further improvements in the change-detection accuracies were obtained by the proposed method, which exploits spatial-contextual information in the change-detection process. This method proved very effective even on images affected by high levels of noise.

As a final remark, it is worth noting that we have formulated the EM algorithm under the assumption that the conditional density functions of classes can be modeled by Gaussian distributions. In several change-detection applications involving the use of images acquired by passive sensors, this assumption seems to be a reasonable approximation. However, when the number of different typologies of land cover changes to be identified increases or when active sensors are used instead of passive ones, the Gaussian model might turn out to be inappropriate. In these cases, more general approaches to the mixture density estimation problem [22], [23], [38] may represent powerful tools in obtaining accurate estimates of the density functions associated with changed and unchanged pixels.

ACKNOWLEDGMENT

The authors are grateful to the anonymous referees for their constructive criticism.

REFERENCES

[1] K. R. Merril and L. Jiajun, "A comparison of four algorithms for change detection in an urban environment," *Remote Sens. Environ.*, vol. 63, pp. 95–100, 1998.

[2] T. Hame, I. Heiler, and J. San Miguel-Ayanz, "An unsupervised change detection and recognition system for forestry," *Int. J. Remote Sensing*, vol. 19, no. 6, pp. 1079–1099, 1998.

[3] R. D. Johnson and E. S. Kasischke, "Change vector analysis: A technique for the multispectral monitoring of land cover condition," *Int. J. Remote Sensing*, vol. 19, no. 3, pp. 411–426, 1998.

[4] X. Dai and S. Khorram, "The effects of image misregistration on the accuracy of remotely sensed change detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 1566–1577, Sept. 1998.

[5] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, pp. 858–867, July 1997.

[6] L. Bruzzone and D. F. Prieto, "An adaptive parcel-based technique for unsupervised change detection," *Int. J. Remote Sensing*, vol. 21, no. 4, pp. 817–822, 2000.

[7] P. S. Chavez, Jr. and D. J. MacKinnon, "Automatic detection of vegetation changes in the southwestern united states using remotely sensed images," *Photogramm. Eng. Remote Sensing*, vol. 60, no. 5, pp. 1285–1294, 1994.

[8] M. E. Jakubauskas, K. P. Lulla, and P. W. Mausel, "Assessment of vegetation change in a fire-altered landscape," *Photogramm. Eng. Remote Sensing*, vol. 56, no. 3, pp. 371–377, 1990.

[9] S. Gopal and C. Woodcock, "Remote sensing of forest change using artificial neural networks," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 398–404, 1996.

[10] A. Singh, "Digital change detection techniques using remotely-sensed data," *Int. J. Remote Sensing*, vol. 10, no. 6, pp. 989–1003, 1989.

[11] T. Fung, "An assessment of TM imagery for land-cover change detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, no. 12, pp. 681–684, 1990.

[12] T. Fung and E. LeDrew, "The determination of optimal threshold levels for change detection using various accuracy indices," *Photogramm. Eng. Remote Sensing*, vol. 54, no. 10, pp. 1449–1454, 1988.

[13] M. L. Stauffer and R. L. McKinney, Landsat Image Differencing as an Automated Land Cover Change Detection Technique, Computer Science Corp., Silver Springs, MD, 1978.

[14] A. Singh, "Tropical forest monitoring using digital Landsat data in northeastern India," Ph.D. dissertation, University of Reading, Reading, U.K., 1984.

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

[16] T. K. Moon, "The expectation-maximization algorithm," *Signal Processing Mag.*, vol. 13, no. 6, pp. 47–60, 1996.

[17] A. P. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, vol. 26, no. 2, pp. 195–239, 1984.

[18] J. R. G. Townshend, C. O. Justice, and C. Gurney, "The impact of misregistration on change detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 30, pp. 1054–1060, Sept. 1992.

[19] L. Bruzzone and S. B. Serpico, "Detection of changes in remotely-sensed images by the selective use of multi-spectral information," *Int. J. Remote Sensing*, vol. 18, no. 18, pp. 3883–3888, 1997.

[20] P. S. Chavez, Jr., "Radiometric calibration of Landsat Thematic Mapper multispectral images," *Photogramm. Eng. Remote Sensing*, vol. 55, no. 9, pp. 1285–1294, 1989.

[21] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. London, U.K.: Academic, 1990.

[22] B. M. Shahshahani and D. A. Landgrebe, "The effect of unlabeled samples in reducing the small size problem and mitigating the Hughes phenomenon," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 1087–1095, Sept. 1994.

[23] Y. Delignon, A. Marzouki, and W. Pieczynski, "Estimation of generalized mixture and its application in image segmentation," *IEEE Trans. Image Processing*, vol. 6, pp. 1364–1375, Oct. 1997.

[24] R. Chellapa and A. Jain, *Markov Random Fields: Theory and Applications*. New York: Academic, 1993.

[25] R. C. Dubes and A. K. Jain, "Random field models in image analysis," *J. Appl. Statist.*, vol. 16, pp. 131–163, 1989.

[26] E. Rignot and R. Chellappa, "Segmentation of polarimetric synthetic aperture radar data," *IEEE Trans. Image Processing*, vol. 1, pp. 281–300, Mar. 1992.

[27] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, no. 6, pp. 721–741, 1984.

[28] C. S. Regazzoni, F. Arduini, and G. Vernazza, "A multilevel GMRF-based approach to image segmentation and restoration," *Signal Processing*, vol. 34, pp. 43–67, 1993.

[29] A. H. S. Solberg, A. K. Jain, and T. Taxt, "Fusion of multitemporal satellite images and GIS data for land-use classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 768–778, July 1994.

[30] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B*, vol. 48, pp. 259–302, 1986.

[31] L. Bruzzone and D. F. Prieto, "A technique for the selection of kernel-function parameters in RBF neural networks for classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 1179–1184, Mar. 1999.

[32] E. M. Pereira and A. W. Setzer, "Spectral characteristics of fire scars in Landsat-5 TM images of Amazonia," *Int. J. Remote Sensing*, vol. 14, no. 11, 1993.

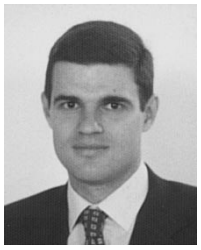
[33] L. Bruzzone and D. F. Prieto, "A Bayesian approach to automatic change detection," in *Proc. IEEE 1999 Int. Geoscience and Remote Sensing Symp. (IGARSS'99)*, Hamburg, Germany, July 1999, pp. 1816–1818.

[34] L. Bruzzone, "An approach based on the Bayes rule for minimum cost to feature selection and classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 429–438, Jan. 2000.

[35] P. C. Smits and S. G. Dellepiane, "Synthetic aperture radar image segmentation by a Detail Preserving Markov Random Field Approach," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, no. 4, pp. 844–857, 1997.

[36] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs Random Fields using Simulated Annealing," *IEEE Trans. Geosci. Remote Sensing*, vol. 11, no. 8, pp. 799–813, 1989.

[37] B. Chalmond, "An iterative Gibbsian technique for reconstruction of m-ary images," *Pattern Recognit.*, vol. 22, no. 6, pp. 747–762, 1989.
[38] C. E. Priebe and D. J. Marchette, "Adaptive mixture density estimation," *Pattern Recognit.*, vol. 26, no. 5, pp. 771–785, 1993.



Lorenzo Bruzzone (S'95–M'99) received the "Laurea" (M.S.) degree in electronic engineering and the Ph.D. degree in telecommunications, both from the University of Genoa, Genoa, Italy, in November 1993 and June 1998, respectively.

Since June 1998, he has been a Postdoctoral Researcher at the University of Genoa. He was the Scientific Coordinator of the activites on remote sensing image analysis carried out by the Signal Processing and Telecommunications Group, Department of Biophysical and Electronic Engineering, University of Genoa. Since February 2000, he has been an Assistant Professor of telecommunications with the University of Trento, Trento, Italy, where he teaches electrical telecommunications. His main research contributions are in the area of remote sensing image processing and recognition. In particular, his interests include feature extraction and selection, classification, change detection, and data fusion. He conducts and supervises research on these topics within the frameworks of several national and international projects. Since 1999, he has been Evaluator of Project Proposals within the Fifth Framework Programme of the European Commission. He is the author or co-author of more than 60 scientific publications and a referee for the international journals IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, *International Journal of Remote Sensing*, and *Signal Processing*.

Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium. He received the recognition of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING as Best Reviewer in 1999. Since 1998, he has served on the Scientific Committee of the EUROPTO International Conferences on "Signal and Image Processing for Remote Sensing." He is a member of the International Association for Pattern Recognition (IAPR).

Diego Fernàndez Prieto (S'99) received the B.S. degree in physics from the University of Santiago de Compostela, Spain, in 1994. He received the M.B.A. degree in business administration from the University of Deusto, Spain, and the M.E.B.A from the University of Kent, U.K., both in 1997. He is currently pursuing the Ph.D. degree in the Department of Biophysical and Electronic Engineering, University of Genoa, Genoa, Italy.

In 1994 and 1996, he was involved in the activities of the Istituto per la Matematica Applicata (I.M.A), National Research Council (C.N.R), Genoa, Italy, where he studied parallel-computing applications for the Geographic Information System (G.I.S). His main research activity is in the area of the remote sensing image processing and recognition. In particular, his current interests include neural networks for classification purposes, incremental learning, data fusion, and change detection techniques.

Mr. Prieto is a Member of the International Association of Pattern Recognition (IAPR) and the Asociaciòn Española de Teledetecciòn (AET). He is a Reviewer for the journal IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.