# UNSUPERVISED CHANGE DETECTION IN MULTISPECTRAL SATELLITE IMAGES USING OPTIMAL THRESHOLDING TECHNIQUES

Anamaria RĂDOI[1]

*With the constant development of satellite sensors and the existence of huge Earth Observation data archives, unsupervised change detection techniques become important for the analysis of the Earth surface dynamics. In this paper, we present effective unsupervised change detection methods that use optimal thresholding techniques and reduce the number of singular detections by considering the degree of change in the proximity of each pixel. Additionally, we propose a simple solution to reduce the influence of clouds, which is a common problem when dealing with multispectral imagery. We also show that sample-based thresholding is efficient in the analysis of large images.*

**Keywords:** Unsupervised change detection, Expectation-Maximization algorithm, optimal discriminant thresholding, cloud-contaminated multispectral data, noise.

## 1. Introduction

The current progress of satellite sensors leads to a great volume of high-resolution optical images, for which precise, fast and, most important, unsupervised algorithms have to be designed. In this sense, machine learning techniques show promising results in this field, but, due to the specific properties of remote sensing data, many challenges still exist (i.e., heterogeneity of the data in time and space, absence of data, context-dependence, scarcity of land-cover changes) [1].

Usually, change detection implies the comparison of two remote sensing images acquired over the same geographical area at two different moments of time [2]. These changes might appear in different scenarios, such as modifications in land use (agriculture), appearance of new buildings (urban areas), natural disaster, flooding, forest monitoring, and so on. An unsupervised change detection algorithm has to be able to discover changes starting only from the two available, and, often, very large image files, without any external aid (e.g., user interaction) or additional information regarding the distribution of the pixels.

Many change detection techniques are based on the analysis of the difference image, which is computed as the subtraction between pixel intensities of the

[1]Lecturer, Faculty of Electronics, Telecommunications and Information Technology, University POLITEHNICA of Bucharest, Romania, e-mail: anamaria.radoi@upb.ro

two images [2]. For unchanged pixels, the corresponding values in the difference image are significantly smaller than the values associated to changed pixels [3]. Then, the delimitation of changed pixels is done by applying a threshold over the difference image. The threshold might be determined through an empirical search [4], but this strategy is usually time-demanding and prone to error. In a recent paper [5], we showed that the changes can be ranked depending on the intensity of change suffered by the analyzed areas using binary descriptors and the Lloyd Max's quantization algorithm. Another method to find the threshold delimiting the changes from non-changes is to minimize the overall change detection error probability. This can be done by using the Expectation-Maximization (EM) algorithm [3]. In the same paper, an optimization of this solution is considered by using Markov Random Fields (MRFs) to penalize those pixels that receive a label (change/no change) that is different by the neighboring ones. Although effective in grouping the changes together, the MRF technique requires a very large number of computations which makes it unsuitable for online change detection.

The difference image can be analyzed also from the perspective Principal Component Analysis (PCA) with the aim of finding the main directions (i.e., eigenvectors) of change [6]. The classification into *change* versus *no change* is performed by applying K-means over the features extracted by projecting the neighborhood of each pixel on these eigenvectors.

The aim of this paper is to build a simple, fast, effective, complete and unsupervised change detection framework that takes into consideration various aspects such as multispectral information, contextual change information, online change analysis, and cloud-contaminated data. The proposed approaches use thresholding techniques coming into two flavors: the Maximum Likelihood Estimation with Expectation Maximization (MLE-EM) and optimal discriminant thresholding (ODT). By inserting neighboring change information into the change vector, the algorithm is able to effectively distinguish the zones that were affected by change (i.e., grouped changes rather than isolated ones). As a consequence, singular pixels affected by noise, changes in illumination, digital image artifacts are not treated as change. Moreover, a solution for change detection over very large images is envisaged by considering sample-based versions of the proposed algorithms. This solution maintains a high performance level of change detection accuracy while significantly decreasing the average running time of the algorithm. Additionally, a pre-processing step for the extraction of cloud-contaminated areas is included in the analysis.

The remainder of the paper is structured as follows. Section 2 introduces the proposed methodologies and two optimizations for big data images and for cloud contaminated data, respectively. The next section presents a set of results obtained with the proposed methodology, along with the performances of state-of-the-art methods. Finally, the last section concludes the paper.

## 2. Proposed Change Detection Techniques

As stated already, this paper aims at providing a solution for the unsupervised change detection problem in multispectral remote sensing images. In this regard, both the multispectral and local information are taken into consideration for each pixel. The proposed method consists in building change vectors that, in the next stage, have to be separated in *change* versus *no change* classes. The class separation is done by finding the optimal threshold that divides the change vector space into two multidimensional subsets. In this sense, the optimality of the thresholding techniques is regarded from two different perspectives: the Maximum-likelihood (ML) criterion and the discriminant criterion, respectively. The main steps of the change detection procedure described in this paper are the following:

(1) Compute difference image (**D**)
(2) Apply cloud detection mask
(3) Build change vectors (**CV**)
(4) Find optimal thresholds that separate *change* from *no change*
(5) Retrieve final Change Detection map (**CDM**)

### 2.1. Building Change Vectors from the Difference Image

In the following, let us consider two multispectral images, $\mathbf{I}_1$ and $\mathbf{I}_2$ of size $W \times H$ pixels and $B$ spectral bands. The difference image between the two original images, normalized between $[0,1]$, is defined as:

$$\mathbf{D} = |\mathbf{I}_2 - \mathbf{I}_1| \tag{1}$$

Starting from this multispectral difference image **D**, a three-dimensional matrix $W \times H \times B$ of change vectors **CV** can be easily derived. At each pixel location, each change vector $\mathbf{CV}_{i,j}$ contains $B$ elements that correspond to the $B$ spectral bands ($1 \leq i \leq W, 1 \leq j \leq H$). In order to take into consideration the change occurring in the $p \times p$ neighborhood of each pixel, each element of the change vector is computed as:

$$CV_{i,j,b} = \frac{1}{p^2} \sum_{i'=-\lfloor p/2 \rfloor}^{\lfloor p/2 \rfloor} \sum_{j'=-\lfloor p/2 \rfloor}^{\lfloor p/2 \rfloor} D_{i+i',j+j',b} \tag{2}$$

Generally speaking, the averaging operation recalls the moving average filter, which is the simplest and among the most widely used solutions to reduce the level of noise in an image. Here, we employ this type of image transformation in order to take into account the degree of change in the immediate neighborhood of a pixel with almost no additional computational resources. After local averaging, if change occurs, the observed pixel and those surrounding it will likely be characterized by higher values in the difference image than the pixels in unchanged areas or isolated changed pixels. Thus, a compact changed area is represented by high values in the change vectors $\mathbf{CV}_{i,j}$.

## 2.2. The MLE–EM Approach

For the beginning, let us consider only one band of the multispectral **CV** matrix, namely $\mathbf{CV}_b = [CV_{i,j,b}]_{i,j}$. The change detection algorithm is applied on each $\mathbf{CV}_b$ and the resulting change maps are merged into a final change detection map **CDM**, as detailed below.

Following the formalism initially presented in [7] and, then, particularized in [3], the optimal threshold that separates the values in $\mathbf{CV}_b$ into *change* and *no change* classes, can be determined under the framework of the Bayesian decision theory. In the change detection context, the MLE-EM approach can be regarded as a method of determining an optimal separation threshold with respect to the Maximum-Likelihood (ML) criterion.

As already known from the literature, the statistical distribution of classes in images acquired by optical sensors follow, in most cases, a Gaussian distribution, or it can be modeled as a mixture of Gaussian distributions [3]. In the change detection context, this means that the values of the $\mathbf{CV}_b$ are drawn from a mixture of two Gaussian random variables, i.e. $\mathbf{CV}_b|\Omega_c \sim \mathcal{N}(\mu_c, \sigma_c^2)$ and $\mathbf{CV}_b|\Omega_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$, where $(\mu_n, \sigma_n)$ and $(\mu_c, \sigma_c)$ are the mean and variance pairs of the density components that correspond to the *no change* class $\Omega_n$ and *change* class $\Omega_c$, respectively. More precisely, under these assumptions, the overall posterior probability for $x \in \mathbf{CV}_b$ is modeled as:

$$p_\theta(x) = \pi_n \mathcal{N}(x|\mu_n, \sigma_n^2) + \pi_c \mathcal{N}(x|\mu_c, \sigma_c^2) \tag{3}$$

where $\theta = \left[\pi_n, \mu_n, \sigma_n^2, \pi_c, \mu_c, \sigma_c^2\right]$ is the set of the model parameters, whilst $\pi_n$ and $\pi_c$ are the unknown prior probabilities (or, mixture probabilities [8]) of classes $\Omega_n$ and $\Omega_c$, respectively, such that $\pi_n + \pi_c = 1$. Each component of the mixture is represented by a Gaussian density:

$$\mathcal{N}(x|\mu_k, \sigma_k^2) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left\{-\frac{1}{2\sigma_k^2}(x-\mu_k)^2\right\} \tag{4}$$

with $k \in \{n, c\}$. The set of statistical parameters $\theta$ is determined iteratively using the Expectation-Maximization (EM) algorithm under the Maximum Likelihood Estimation (MLE) framework. Each iteration $t \geq 0$ consists of two basic steps:

(1) *Expectation step (E-step).* Compute the log-likelihood (i.e., the logarithm of the posterior probability) with respect to the current values of the parameters $\theta_t$:

$$L(\theta_t) = \ln p_{\theta_t}(x) \tag{5}$$

(2) *Maximization step (M-step).* The model parameters are updated such that the log-likelihood approaches its maximum [8]. According to [7] and following the notations in [8], the mixture probabilities, the means and the covariances for each class $k \in \{n, c\}$ can be re-estimated using the following mapping between

the new model parameters, $\theta_{t+1}$, and the previous ones, $\theta_t$:

$$\pi_{k,t+1} = \frac{\sum\limits_{x \in \mathbf{CV}_b} \zeta_{k,t}(x)}{N} \tag{6}$$

$$\mu_{k,t+1} = \frac{\sum\limits_{x \in \mathbf{CV}_b} \zeta_{k,t}(x) \cdot x}{\sum\limits_{x \in \mathbf{CV}_b} \zeta_{k,t}(x)} \qquad \sigma^2_{k,t+1} = \frac{\sum\limits_{x \in \mathbf{CV}_b} \zeta_{k,t}(x) \cdot \left(x - \mu_{k,t}\right)^2}{\sum\limits_{x \in \mathbf{CV}_b} \zeta_{k,t}(x)} \tag{7}$$

where $W \times H$ is the total number of pixels in the image and $\zeta_{k,t}(x)$ is:

$$\zeta_{k,t}(x) = \frac{\pi_{k,t}\mathcal{N}(x|\mu_{k,t}, \sigma^2_{k,t})}{\sum_{k'} \pi_{k',t}\mathcal{N}(x|\mu_{k',t}, \sigma^2_{k',t})} \tag{8}$$

The initialization of the EM algorithm is done by dividing the pixels of $\mathbf{CV}_b$ into two subsets, $\mathcal{S}_{n,0}$ and $\mathcal{S}_{c,0}$, through a $K$-means pre-clustering. A rearrangement of the initial classes might be necessary such that the centroid value of the *no change* class is smaller than the centroid value of the *change class* (i.e., the difference pixels have smaller values if there is no change). Then, the initial values for the means and squared standard deviations are computed over these two subsets, whilst the prior probabilities are derived as $\pi_{k,0} = \text{card}(\mathcal{S}_{k,0})/(W \cdot H)$. Compared to the initialization proposed in [3], the K-means initialization might speed up the convergence of the EM algorithm [8].

After determining the estimate values for the $\theta$ parameters, the optimal threshold $T_o$ that separates the two classes (i.e., *change* and *no change*) can be easily retrieved from the condition below:

$$\pi_n\mathcal{N}(T_o|\mu_n, \sigma^2_n) = \pi_c\mathcal{N}(T_o|\mu_c, \sigma^2_c) \tag{9}$$

that naturally follows from the maximum likelihood rule:

$$\pi_n\mathcal{N}(x|\mu_n, \sigma^2_n) \underset{c}{\overset{n}{\gtrless}} \pi_c\mathcal{N}(x|\mu_c, \sigma^2_c). \tag{10}$$

Using equation (4) and taking the logarithm of both sides of equation (9) yields a quadratic equation in $T_o$ [3]:

$$(\sigma^2_n - \sigma^2_c)T_o^2 + 2\left(\mu_n\sigma^2_c - \mu_c\sigma^2_n\right)T_o + \mu^2_c\sigma^2_n - \mu^2_n\sigma^2_c - 2\sigma^2_c\sigma^2_n \ln\left[\frac{\sigma_n\pi_c}{\sigma_c\pi_n}\right] = 0 \tag{11}$$

that has two possible solutions, from which only the one in the interval $[\mu_n, \mu_c]$ correctly determines the optimal threshold value.

The algorithm presented above yields one change map $\mathbf{CDM}_b$ per band, with each element given by:

$$CDM_{i,j,b} = \mathbb{1}_{\{CV_{i,j,b} > T_{o,b}\}} \tag{12}$$

where $T_{o,b}$ represents the optimal EM-derived threshold that corresponds to band $b$ and $\mathbb{1}_A$ is the indicator function that maps the elements to 1 if statement $A$ is true and 0, otherwise. The algorithm can be easily extended to multidimensional data by computing each element of the final change map **CDM** using a majority voting rule:

$$CDM_{i,j} = \mathbb{1}_{\{\sum_{b=1}^{B} CDM_{i,j,b} > \frac{B}{2}\}} \tag{13}$$

### 2.3. The Optimal Discriminant Thresholding (ODT) Approach

The optimal threshold that separates *change* from *no change* can be found using discriminant analysis. From this viewpoint, Otsu's thresholding method [9] performs a pool-based search to find the optimal threshold. For each possible threshold, the method evaluates a class separability measure (i.e., the between-class variance). Thus, the optimal threshold with respect to the discriminant criterion is the one that maximizes the class separability measure.

For the beginning, let us consider single-band change vectors (i.e., each $\mathbf{CV}_b$ is composed of only one element) that can take only $L$ evenly-spaced values, $v_1 \leq v_2 \leq \ldots \leq v_L$. For each level $u$, the probability $p_u$ can be computed as the proportion of elements equal to $v_u$. If the threshold is fixed at level $l$, the class probabilities are:

$$\omega_n = \sum_{u=1}^{l} p_u \qquad \omega_c = \sum_{u=l+1}^{L} p_u \tag{14}$$

whereas the class mean levels are:

$$\mu_n = \sum_{u=1}^{l} v_u \frac{p_u}{\omega_n} \qquad \mu_c = \sum_{u=l+1}^{L} v_u \frac{p_u}{\omega_c}. \tag{15}$$

Denoting by $\mu_T = \sum_{u=1}^{L} v_u p_u$ the total mean level, the between-class variance can be written as [9]:

$$\sigma_B^2 = \omega_n(\mu_n - \mu_T)^2 + \omega_c(\mu_c - \mu_T)^2 \tag{16}$$

The optimal threshold $v_{opt}$ is the $l^{th}$ level value for which the maximum between-class variance is achieved. Following the steps presented for the MLE-EM approach, the ODT algorithm can be easily extended for multispectral data.

### 2.4. Optimization of the Approaches for Large Data Sets.

From a complexity point of view, computing the MLE's parameter estimates over very large images might prove to be difficult and time-consuming. A sample-based approach can represent the solution towards the estimation of the model parameters and, thus, towards a lower computational complexity. More precisely, the parameters can be found using a smaller subset of randomly selected pixels and, then, the parameters can be applied over the whole dataset to compute the optimal

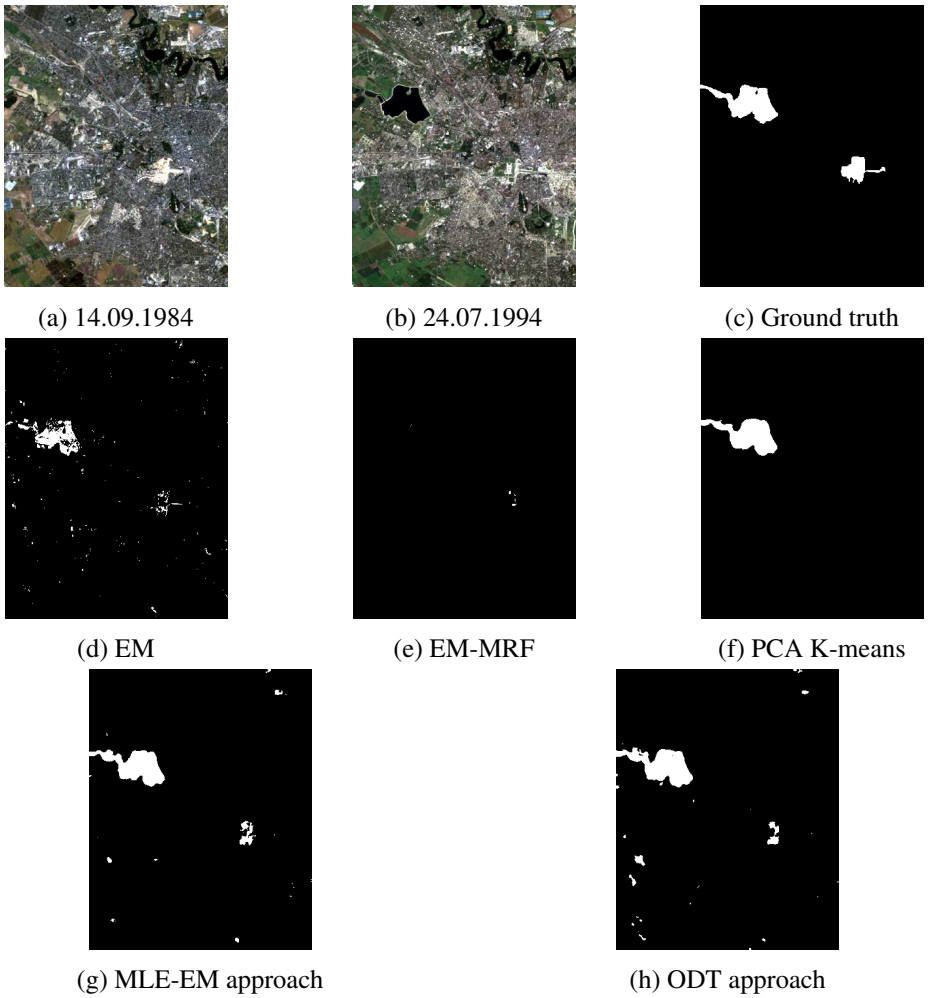
threshold and generate the change detection map. According to [10], the statistical guarantees of a sample-based EM algorithm can be achieved after a relatively small number of EM steps. Moreover, even under noise conditions (i.e., Gaussian mixtures with low Signal-to-Noise Ratios), Balakrishnan et al. [10] prove that the convergence of the EM algorithm can still be attained. However, estimating the model parameters from too few samples might drive the algorithm to a non-global MLE [10]. In this regard, a compromise between speed and performance has to be assessed beforehand. In the following, $\Delta$ represents the percentage of the change vectors selected through random sampling and used to estimate the MLE model parameters (i.e., $\Delta = 100\%$ means that all pixels are used). The same reasoning of sample-based threshold derivation can be applied for the ODT method.

### 2.5. Removing the Influence of Cloud-contaminated Data.

Cloud contamination represents a major issue for the exploitation of optical remote sensing images and often conducts to misleading results in change detection because the algorithms tend to include the appearance or disappearance of a cloud in the change class. For these reasons, the cloudy pixels need to be found and removed from the analyzed data before the change detection algorithm is applied. In order to perform cloud detection, we adopt the Fmask algorithm which has been recently proposed in [11] and is currently one of the most successful object-based approaches in determining cloud masks and cloud-shadow masks for multispectral images. The Fmask algorithm is based on the computation of similarity measures between cloud and cloud-shadow at different cloud heights, whilst the final cloud mask is derived using the spectral variability probability and the brightness probability. However, the method assumes that the view angle of the satellite sensor, the solar zenith angle and the solar azimuth angle are known, but this information can be easily extracted from the metadata accompanying the satellite images. For a more detailed description of the Fmask algorithm, we refer the reader to [11].

## 3. Experiments

In order to measure the performance achieved by the proposed methods, several experiments were carried on pairs of Landsat and IKONOS images. Landsat images are captured over Bucharest, Romania, at a spatial resolution of 30 meters. The first pair of Landsat images ($342 \times 432$ pixels) is captured on 14.09.1984 and 24.07.1994, respectively (Fig. 1). The second pair of Landsat images, captured on 14.09.1984 and 03.10.1994, is cloud contaminated (Fig. 2). Two major changes occurred in this period of time, namely the construction of the Morii Lake and the Palace of Parliament. IKONOS images ($759 \times 734$ pixels) are captured before and after the 2004 Indian Ocean earthquake that affected part of Thailand. Some additional tests were performed on other Landsat pairs of images and synthetic data from [5].



(a) 14.09.1984 (b) 24.07.1994 (c) Ground truth

(d) EM (e) EM-MRF (f) PCA K-means

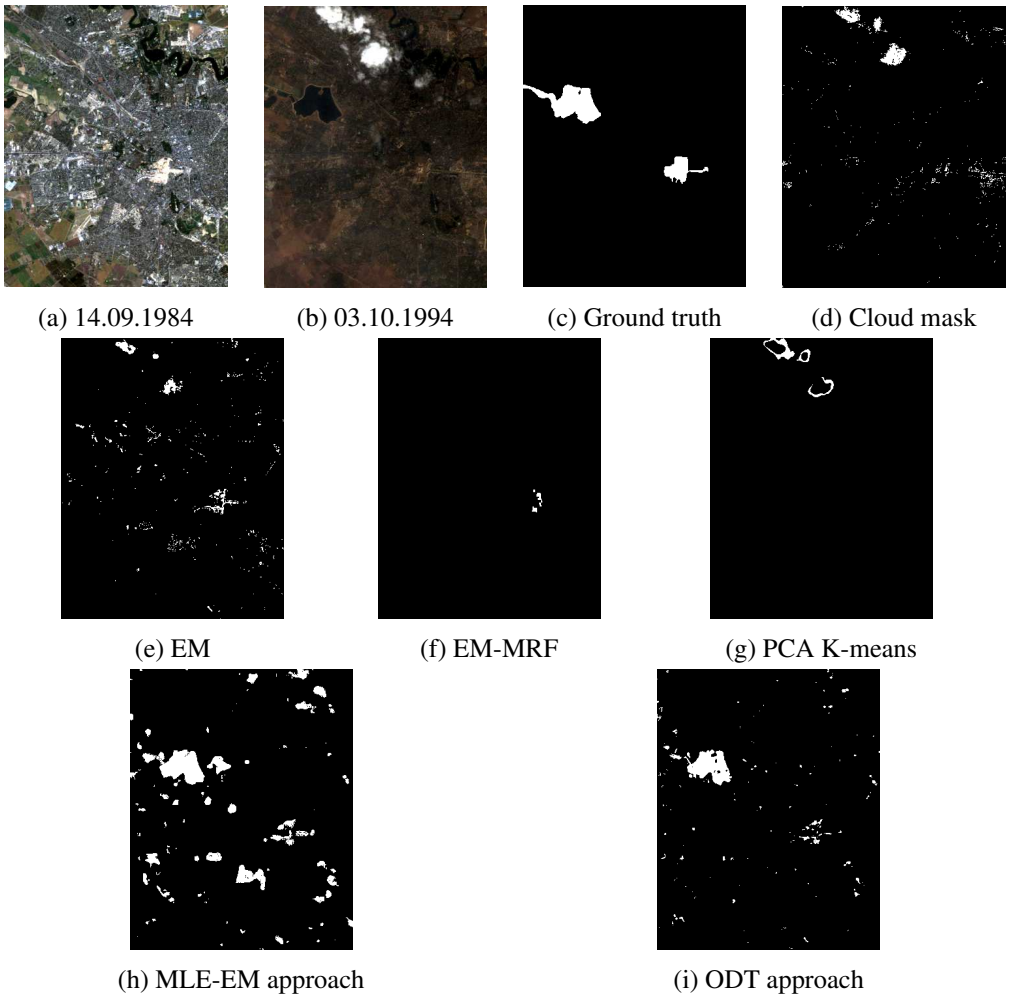
(g) MLE-EM approach (h) ODT approach

Fig. 1. Change detection results for the Landsat image pair – no clouds.

The results of the proposed methods (MLE-EM and ODT) are compared against state-of-the-art methods such as EM and EM-MRF [3], PCA K-means [6] and BRIEF [5]. In Figs. 1 and 2, several results are shown, whilst numerical results for Landsat image pairs are provided in Table 2.

### 3.1. Stability of Sample-based Versions of the Proposed Techniques

In order to check the stability of the optimal thresholds when the sampled-based versions of the proposed algorithms, 10 tests were performed for $\Delta$ varying between 10% and 100% (Table 1). As expected, the standard deviation around the optimal threshold increases when the size of the subset decreases. Compared to the MLE-EM approach, ODT achieves a significantly higher stability because of



(a) 14.09.1984 (b) 03.10.1994 (c) Ground truth (d) Cloud mask

(e) EM (f) EM-MRF (g) PCA K-means

(h) MLE-EM approach (i) ODT approach

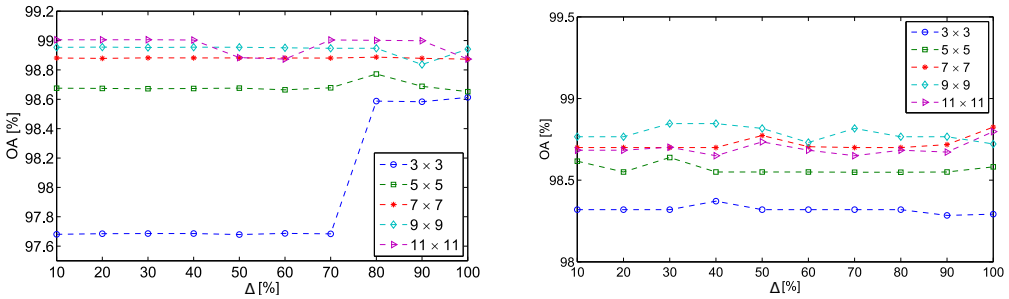
Fig. 2. Change detection results for the Landsat image pair – with clouds.

*Table 1*

**The Optimal Thresholds (average $\pm$ standard deviation)**

| Percentage of data | $T_o$ – MLE-EM approach | $T_o$ – ODT approach |
|---|---|---|
| $\Delta = 100\%$ | 0.1410 | 0.1137 |
| $\Delta = 70\%$ | $0.1413 \pm 0.00035$ | $0.1137 \pm 1.39 \cdot 10^{-17}$ |
| $\Delta = 40\%$ | $0.1409 \pm 0.00065$ | $0.1137 \pm 1.39 \cdot 10^{-17}$ |
| $\Delta = 10\%$ | $0.1431 \pm 0.00253$ | $0.1120 \pm 0.0045$ |

the smaller number of parameters needed to be estimated. For most of the following experiments, we consider $\Delta = 40\%$ as a good compromise between speed and stability of threshold computation.




(a) MLE-EM approach (b) ODT approach

Fig. 3. Average overall change detection accuracy of the proposed methods for different window sizes and various sample rates Δ.

### 3.2. Performance Assessment

We denote by $N$ the number of pixels that are marked as unchanged in the ground truth map (i.e., negative samples), whilst the number of changed pixels is given by $P$ (i.e., positive samples). Correct classifications are given by $TP$ (true positives) and $TN$ (true negatives), whilst type I and II errors are denoted as $FP$ (false positives or false alarms) and $FN$ (false negatives or missed detections). The performance of the algorithm is assessed using a set of numerical measures that are usually used in binary classification problems:

1) correct classification rate (overall accuracy): $OA = (TP+TN)/(P+N) \times 100$;
2) true positive rate (hit rate): $HR = TP/P \times 100$;
3) missed rate : $MR = FN/P \times 100$;
4) probability of false alarm : $PFA = FP/N \times 100$.

Following the reasoning from the previous paragraph, the settings that are used for MLE-EM and ODT are $\Delta = 40\%$ and a window size of $7 \times 7$ pixels. In general, the compactness of changed area grows with the analysis window size. However, if the size of the window is too large, the changed area starts to lose its sharp edges and geometry details, and spatial information might be destroyed. A detailed analysis with respect to the selection of the window size and of the percentage of data used is presented in Fig. 3. The overall accuracy remains, in almost all cases, relatively stable for these settings, representing a good compromise in terms of compactness of the changed areas and reported accuracy.

In almost all cases presented in Table 2, the proposed methodologies achieve high hit rate and overall accuracies, whilst keeping the probability of false alarms at a very low level. This is explained by the inclusion of neighboring change information which leads to a better grouping of changes than the EM method. The results over the Morii Lake and Palace of Parliament pairs from [5] show that MLE-EM and ODT focus more on the extraction of significant (or, major) changes, whilst

*Table 2*

**Comparisons of binary change detection methods**

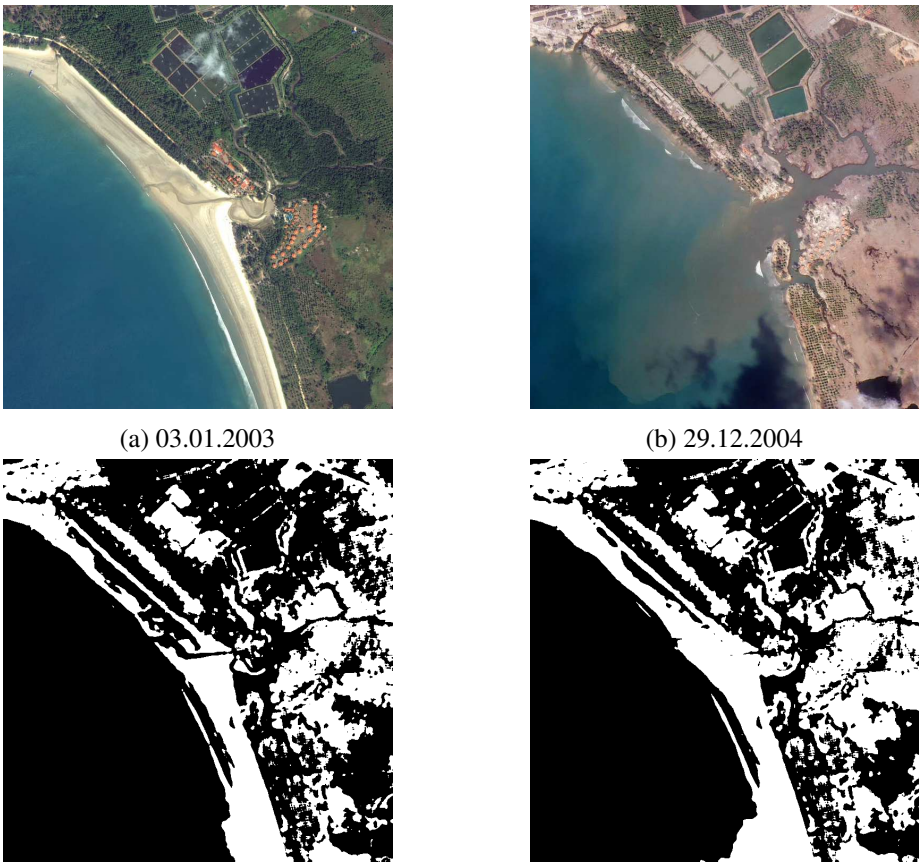
| Dataset | Method | Overall Accuracy | Hit Rate | Missed Rate | Prob. of False Alarms |
|---|---|---|---|---|---|
| Landsat pair no clouds | EM | 98.12 | 47.83 | 52.17 | 0.37 |
| | EM-MRF | 97.11 | 0.97 | 99.03 | 0 |
| | PCA K-Means | 98.90 | 69.57 | 30.43 | 0.22 |
| | BRIEF | 75.22 | 91.37 | 8.63 | 25.57 |
| | MLE-EM approach | 98.88 | 75.36 | 24.63 | 0.41 |
| | ODT approach | 98.76 | 77.17 | 22.82 | 0.58 |
| Landsat pair with clouds | EM | 96.31 | 12.27 | 87.72 | 1.15 |
| | EM-MRF | 97.18 | 3.41 | 96.59 | 0 |
| | PCA K-Means | 96.44 | 0 | 100 | 0.65 |
| | BRIEF | 74.54 | 91.00 | 9.00 | 26.25 |
| | MLE-EM approach | 95.52 | 68.93 | 31.07 | 3.68 |
| | ODT approach | 98.45 | 59.66 | 40.33 | 0.38 |
| Morii Lake pair from [5] | EM | 85.34 | 50.91 | 49.09 | 3.95 |
| | EM-MRF | 83.26 | 31.84 | 69.16 | 0.75 |
| | PCA K-Means | 82.97 | 29.38 | 70.62 | 0.36 |
| | BRIEF | 86.80 | 87.38 | 12.62 | 13.39 |
| | MLE-EM approach | 80.50 | 28.23 | 71.77 | 0.13 |
| | ODT approach | 84.38 | 45.84 | 54.16 | 0.53 |
| Palace of Parliament pair from [5] | EM | 67.73 | 20.28 | 79.72 | 6.08 |
| | EM-MRF | 70.67 | 34.19 | 65.81 | 9.51 |
| | PCA K-Means | 84.76 | 64.97 | 35.03 | 19.31 |
| | BRIEF | 85.03 | 79.73 | 20.27 | 33.10 |
| | MLE-EM approach | 84.50 | 68.93 | 31.07 | 5.53 |
| | ODT approach | 84.34 | 62.75 | 37.25 | 3.34 |
| Synthetic changes from [5] | EM | 87.09 | 52.48 | 47.52 | 3.86 |
| | EM-MRF | 82.61 | 26.07 | 73.93 | 2.60 |
| | PCA K-Means | 86.11 | 35.43 | 64.57 | 0.62 |
| | BRIEF | 94.60 | 90.37 | 9.63 | 4.29 |
| | MLE-EM approach | 92.39 | 100 | 0 | 9.58 |
| | ODT approach | 92.20 | 76.71 | 23.29 | 3.74 |

BRIEF with two levels of quantization ($q = 2$ in [5]) divides the changes into two groups, namely medium-to-major and minor-to-medium changes (e.g., high *PFA* of BRIEF in the first two cases). In addition, MLE-EM and ODT disregard seasonal changes, which is desirable in a binary change detection application.

In general, ODT provides more restrictive thresholds (Table 1) because the aim is to minimize the overlap of the two clusters by minimizing their combined spread. For this reason, MLE-EM may register more false alarms than ODT.

The cloud detection pre-processing step was applied on all images for the proposed approaches. In the cases with no clouds, the cloud detection did not signal any cloud to remove. In cloud contaminated images as in Fig. 2, if no cloud detection is performed, EM and PCA K-means methods mark the appearance of the cloud as change. Removing the cloud-contaminated pixels from the change analysis and extracting context-based change vectors leads to better qualitative results (Fig. 1 and Table 2).

Another use case scenario is presented in Fig. 4, which represent images captured before and after the earthquake that devastated part of Thailand in 2004. In this case, MLE-EM and ODT methods are able to automatically assess the percentage of the shore that practically vanished and also to detect the portions of land



(a) 03.01.2003 (b) 29.12.2004

(c) MLE-EM approach ($OA = 86.53\%$) (d) ODT approach ($OA = 82.34\%$)

Fig. 4. Change detection results for the Thailand earthquake image pair.

that were damaged by the floods. This is one of the application scenarios in which these methods can be successfully applied.

### 3.3. Robustness to Noise

The robustness of the proposed algorithms with respect to noise is tested under the assumption of additive zero-mean Gaussian noise at different Peak Signal-to-Noise Ratios (PSNRs). The PSNR is defined as:

$$PSNR = 10\log_{10}\left(\frac{H \cdot W}{\sum_{i=1}^{H}\sum_{j=1}^{W}\left(X_{i,j} - \hat{X}_{i,j}\right)^2}\right) \tag{17}$$

where we considered that the maximum intensity value of the original image is 1 and that $\hat{\mathbf{X}}$ is the noisy correspondent of the original image $\mathbf{X}$. The PSNR was varied between 20 dB and 50 dB and the hit rate was plotted in Fig. 5. Being block-based data analysis methods, MLE-EM and ODT outperform other pixel-based methods

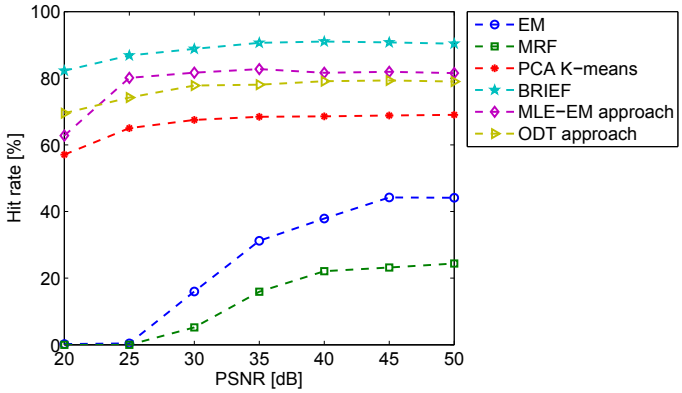



Fig. 5. Performance of the proposed methods versus other methods for different Peak Signal-to-Noise Ratios (PSNRs)

*Table 3*

**Average running time comparisons for large image files**

| Change detection method | Average running time [seconds] |
|---|---|
| EM | 20.83 |
| EM - MRF | 8768.1 |
| PCA K-Means | 225.74 |
| BRIEF | 42.84 |
| MLE-EM approach | 15.84 |
| ODT approach | 1.49 |

(EM). In the case of PCA K-means, the noise affects the directions of change, and thus the change detection.

### 3.4. Average Running Time

In order to speed up the computation of optimal thresholds, the estimates of the MLE-EM and ODT parameters can be obtained by analyzing a smaller subset of randomly chosen change vectors. Compared to the other methods, MLE-EM and ODT achieve the lowest execution time. By contrary, the optimization performed over undirected graphical models in the case of EM-MRF or the extraction of eigenvectors for multivariate analysis in the case of PCA K-means are both time-consuming tasks. Nevertheless, the ODT method does not lead to a significantly decreased computational complexity – this is due to the fact that ODT is based on a simple computation of zero-order and first-order moments, whereas the MLE-EM requires several iterations to compute the estimates of the model parameters. The average running times are shown in Table 3 for pairs of Landsat images of $1000 \times 1000$ pixels. All the experiments were carried on a Desktop PC Intel (R) Core (TM) i7-4770 CPU @ 3.40 Ghz having a RAM memory of 16 GB, with no parallel processing included.

## 4. Conclusions

In this paper, we have proposed two unsupervised change detection algorithms based on optimal thresholding techniques, regarded from two different perspectives, namely Maximum-Likelihood and discriminant analysis. Incorporating change information of the neighborhood of each pixel, the proposed methods achieve a high degree of compactness of the changed areas and also robustness against noise interference. Moreover, two optimizations are proposed, namely a pre-processing phase to diminish the effect of clouds and a sample-based estimation of parameters that decreases the computational time. With high change detection accuracy and low execution time compared to other state-of-the art methods, MLE-EM and ODT are powerful candidates for online change detection applications.

### Acknowledgment

The work was partially funded through the European Space Agency project DAMATS. The author would like to thank SIC for providing the samples of IKONOS imagery for the research experiment.

### REFERENCES

[1] *A. Karpatne and Z. Jiang*, "Monitoring land-cover changes: A machine-learning perspective," IEEE Geoscience and Remote Sensing Magazine, Vol. **4**, no. 2, 2016, pp. 8–21.

[2] *J. Théau*, *Change detection*, Springer handbook of geographic information, 2012, pp. 75–94.

[3] *L. Bruzzone and D. F. Prieto*, "Automatic analysis of the difference image for unsupervised change detection," Geoscience and Remote Sensing, IEEE Transactions on, Vol. **38**, no. 3, 2000, pp. 1171–1182.

[4] *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, 1988, pp. 1449–1454.

[5] *A. Rădoi and M. Datcu*, "Automatic change analysis in satellite images using binary descriptors and Lloyd's Max quantization," IEEE Geoscience and Remote Sensing Letters, Vol. **12**, no. 6, 2015, pp. 1223–1227.

[6] *T. Celik*, "Unsupervised Change Detection in Satellite Images Using Principal Component Analysis and k-Means Clustering," Geoscience and Remote Sensing Letters, IEEE, Vol. **6**, no. 4, 2009, pp. 772–776.

[7] *R. A. Redner and H. F. Walker*, "Mixture Densities, Maximum Likelihood and the EM Algorithm," SIAM Review, Vol. **26**, no. 2, 1984, pp. 195–239.

[8] *Christopher M. Bishop*, *Pattern recognition and machine learning*, Springer-Verlag, 2006.

[9] *N. Otsu*, "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, Man, and Cybernetics, Vol. **9**, no. 1, 1979, pp. 62–66.

[10] *S. Balakrishnan, M. J. Wainwright B. and Yu*, "Statistical guarantees for the EM algorithm: From population to sample-based analysis," Annals of Statistics, arXiv:1408.2156, [to appear].

[11] *Z. Zhu, S. Wang and C. E. Woodcock*, "Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection," Remote Sensing of Environment, Vol. **159**, 2015, pp. 269–277.