



# An a-contrario approach for sub-pixel change detection in satellite imagery

Amandine Robin, Lionel Moisan, Sylvie Le Hégarat-Mascle

## ► To cite this version:

Amandine Robin, Lionel Moisan, Sylvie Le Hégarat-Mascle. An a-contrario approach for sub-pixel change detection in satellite imagery. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32 (11), pp. 1977-1993. 10.1109/TPAMI.2010.37 . hal-00399698

**HAL Id: hal-00399698**

**<https://hal.science/hal-00399698>**

Submitted on 28 Jun 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# An a-contrario approach for sub-pixel change detection in satellite imagery

Amandine Robin, Lionel Moisan, and Sylvie Le Hégarat-Masclé

**Abstract**—This paper presents a new method for unsupervised sub-pixel change detection using image series. The method is based on the definition of a probabilistic criterion capable of assessing the level of coherence of an image series relatively to a reference classification with a finer resolution. In opposition to approaches based on an a priori model of the data, the model developed here is based on the rejection of a *non-structured* model — called a-contrario model — by the observation of *structured* data. This coherence measure is the core of a stochastic algorithm which selects automatically the image subdomain representing the most likely changes. A theoretical analysis of this model is led to predict its performances, in particular regarding the contrast level of the image as well as the number of change pixels in the image. Numerical simulations are also presented, that confirm the high robustness of the method and its capacity to detect changes impacting more than 25% of a considered pixel under average conditions. An application to land-cover change detection is then provided using time series of satellite images.

**Index Terms**—change detection, a-contrario modeling, significance test, sub-pixel, mixture model, image series

## I. INTRODUCTION

THE detection of change areas using an image series is an important issue in image processing due to the large number of impacted applications, including in particular remote sensing [1]–[3], medical diagnosis [4], [5] and video surveillance [6], [7]. Even though more and more sensors are specialized in order to provide fine information dedicated to a specific range of applications, the intrinsic sensors characteristics generally result in a trade off between fine spatial resolution and fine spectral resolution or high time frequency. Therefore when a fine spectrum or a high time frequency is required to discriminate the target of interest (and preferred to a fine spatial resolution), this latter may be smaller than the pixel size. In such cases, providing efficient solutions for sub-pixel change detection is of crucial interest.

This work has been initially motivated by the detection of land cover changes using remote sensed images. In this context and in particular for emergency applications such as floods or fires, the use of data acquired with a high time frequency is mandatory but the changes of interest may impact only a fraction of the observed pixel. Apart from emergency

applications, changes can typically consist in natural land-cover transformations (intra or inter-annual variability) or human-induced changes (*e.g.* forest cuts, crop rotation). The approach presented here applies as well to multitemporal change detection using image time series as to bi-date change detection using single or multi-spectral images.

In the literature, a wide range of change detection methods has been proposed to analyze image series (with all images to the same spatial resolution). Most of them are based either on classification processes (post-classification comparison, joint classification [8] or fusion of classifications [9]), on change vector analysis and thresholding [10]–[12], or on predictive models [13], [14]. Some methods use Markov Random Fields in order to take into account the spatial neighborhood in the difference image [15], [16] and, more recently, a Hopfield-type neural network was proposed to model spatial correlations [17]. For more details, good overviews are given by [18], [19]. All these methods generally lead to limited results due to misregistration errors or illumination variations, which introduce some alterations of the signal that should not be detected as true changes (*e.g.* camera motion, sensor noise, shadows, atmospheric absorption, etc.). Even if pre-processing (*e.g.* geometric or radiometric corrections) are generally performed in order to avoid such limiting factors, they are only partly weakened and hence need to be taken into account. Moreover, some typical issues such as the definition of an *a priori* model or the choice of a detection threshold remain. In addition, very few works focused directly on the sub-pixel change detection problem. An early reference is [20], restricted to edge detection using zero-crossings of the Laplacian. More recently, [21] provided a supervised method restricted to change detection in pairs of images, [22] used an unmixing model for target detection in hyperspectral images and [23] estimated alternately class means and proportions in order to determine changes by comparing a CR time series to a HR classification.

Here, given the methodological concerns previously pointed out, we propose a novel, efficient and fully unsupervised approach to perform sub-pixel change detection by comparing a coarse resolution (CR) time series to a former high resolution (HR) reference classification. Such a comparison let us follow objects which are observable in the HR classification but often impossible to distinguish at a coarse resolution, meanwhile avoiding the frequent inter-calibration problems encountered in the literature. Compared to existing methods, this choice also highly improves robustness in terms of noise and lightning conditions, as the comparison refers to a minimal description of the initial state. Moreover, taking a classification

A. Robin is now with the department of Computational and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa. The main part of this research has been done when she was with the Laboratory of Applied Mathematics (MAP5), University of Paris Descartes e-mail: amandine.robin@wits.ac.za.

L. Moisan is with the Laboratory of Applied Mathematics (MAP5), University of Paris Descartes, Paris, France e-mail: lionel.moisan@parisdescartes.fr.

S. Le Hégarat-Masclé is with the Fundamental Electronic Institute (IEF), University Paris Sud XI, 91405 Orsay, France e-mail: le-hegarat@ief.psud.fr.

for reference rather than an image or an image series enables us to reduce prior hypotheses on the expected evolution.

The theoretical framework of the approach we propose is the so-called a-contrario detection, that was first proposed in [24] for the detection of alignments in images, and then developed for various tasks such as the analysis of histogram modes [25], the recovery of stereo point matches [26], motion detection [27], the detection of spots in textured background [28], etc. The main principle driving a-contrario detection is a perception principle due to Helmholtz, first formulated in the context of image analysis by Lowe [29]. It basically says that we perceive a structure in a set of objects (say, an alignment in a set of black dots drawn on a white sheet of paper) when their configuration is too unlikely to happen by chance. This principle is formalized in the a-contrario framework using two ingredients: first, a naive (*non-structured*) model, which describe the random distribution we expect in case no structure is present (the classical  $H_0$  hypothesis in statistical decision theory); second, one or several measurements related to the structure we want to detect. With these two ingredients, the a-contrario framework allows to combine all measurements into a single real-valued function (called NFA, for *Number of False Alarms*), that can be thresholded to select salient structures. Compared to the classical statistical decision theory, the a-contrario framework has several advantages:

- 1) No  $H_1$  hypothesis has to be formulated, which make a-contrario models easier to build and less sensitive to questionable modeling choices;
- 2) The NFA function is easy to compute, because it does not aim at the control of the probability of a false alarm (which is often impossible to compute in the case of non-independent tests), but at the control of the expectation of the number of false alarms, whose computation can be performed by considering each test independently;
- 3) It guarantees that the number of (wrongly) detected structures under  $H_0$  by thresholding the NFA function to  $\varepsilon$  is, in average, less than  $\varepsilon$ .

The last property is responsible for the name “NFA”, which can be defined in a general setting [28]:

*Definition 1.1 (Number of false alarms):* Let  $(X_i)_{1 \leq i \leq N}$  be a set of random variables. A function  $F(i, x)$  is a NFA (Number of False Alarms) for the random variables  $(X_i)$  if

$$\forall \varepsilon > 0, \quad \mathbb{E}[|\{i, F(i, X_i) \leq \varepsilon\}|] \leq \varepsilon. \quad (1)$$

Assume that the  $X_i$  random variables represent the measurements that characterize the structures we want to detect (that is, the measurement  $x_i$  of  $X_i$  will be all the higher than the  $i$ -th structure is salient). Then, according to [28] (Proposition 2 p. 318), the function

$$NFA(i, x_i) = n_i \cdot \mathbb{P}(X_i \geq x_i) \quad (2)$$

is a NFA as soon as

$$\sum_i \frac{1}{n_i} \leq 1 \quad (3)$$

(and in particular if  $n_i = N$  for all  $i$ ).

One important property of the a-contrario framework is that it reduces all detection parameters to one unique parameter:

the expectation of the number of false alarms  $\varepsilon$ . This is a key point dealing with unsupervised detection as, in practice, fixing the number of false alarms (*e.g.* to 1) makes the detection fully automatic. Moreover, this criterion used jointly with a random sampling algorithm leads to a very robust change detection method, in particular with regard to the proportion of change pixels in the image (which is a common limiting factor in the literature).

This paper is organized as follows. Section II describes the residual error on which the change detection is based. Section III details the a-contrario detection model. Section IV provides a theoretical analysis of the model performances when the main image parameters vary. In Section V, we propose an algorithm based on a random sampling strategy, in order to provide a robust estimation without considering all image subdomains. Section VI presents particular issues occurring in the multitemporal case and the solution adopted. Section VII provides experimental results and illustrates the performance of the approach in the case of land-cover change detection using remote sensing data. We end with concluding remarks in Section VIII.

## II. RESIDUAL ERROR

The first step in building a change detection method is to define a relevant error measure regarding the application. Here, we assume that a HR label map is available and describes the reference state. The label map is considered as an application that assigns a label  $l \in \mathcal{L}$  to each pixel of the HR image domain  $\mathcal{D}_{HR}$ . We aim at localizing the changes as areas where the HR label map no longer corresponds to the CR image time series. Hence, the HR label map acts like a *mask* through which the spatial coherence of the image time series can be studied, and the detected changes will consist in parts of the time series that are not coherent with the reference label map. Such formulation *a priori* enables the detection of sub-pixel changes. Moreover, using a classification as a reference rather than an image time series is interesting as it puts a weaker prior on the expected evolution. It also enables to overcome typical intercalibration problems between dates. In Section VIII, we also show how a radiometric image, after quantization, can be considered as a reference in an application to video surveillance.

Denote  $\mathcal{D}_{CR}$  the CR image domain,  $\mathcal{T}$  the set of acquisition dates and assume that all images of the time series are well co-registered. The observed CR time series is denoted  $v = (v_1, \dots, v_{|\mathcal{T}|})$  where for each date  $t \in \mathcal{T}$ , the grey-level image  $v_t$  is a real-valued function defined on  $\mathcal{D}_{CR}$ . In order to establish the link between the HR label map and the CR time series, let us assume that each CR image corresponds to the block average of a HR image of the same scene. Let  $N = |\mathcal{D}_{HR}|/|\mathcal{D}_{CR}|$  be the number of HR pixels that are contained in a CR pixel and, for any pixel  $y \in \mathcal{D}_{CR}$ ,  $N_l(y)$  be the number of HR pixels with label  $l$  represented within pixel  $y$ . The proportion of label  $l$  within pixel  $y$  is then  $\alpha_l(y) = N_l(y)/N$  and by definition,  $\sum_{l \in \mathcal{L}} \alpha_l(y) = 1$ . For sake of simplicity, the model will be from now on considered in the monotemporal case, and the multitemporal case will be

discussed in section VI. If  $\mu = (\mu(l))_{l \in \mathcal{L}}$  stands for the mean intensity characterizing labels  $l$ , then the intensity of a pixel  $y \in \mathcal{D}_{\text{CR}}$  is estimated by

$$\hat{v}(y, \mu) = \sum_{l \in \mathcal{L}} \alpha_l(y) \mu(l). \quad (4)$$

Practically, this block-average hypothesis may be only roughly satisfied. In general, CR values are obtained by a modulation transfer function which is not an indicator function as supposed by the block-average hypothesis. However, Equation 4 boils down to the *linear mixture model* [30] which is commonly used in the remote sensing community since numbers of applications showed its relevance even when the hypothesis of block-average is not valid (see [31], [32]).

Then, knowing the label proportions, an optimal HR reconstruction  $\hat{v}_D$  may be obtained by using the label map and choosing label means  $(\mu(l))_{l \in \mathcal{L}}$  that minimize the reconstruction error, equal to

$$\delta(v_D) = \sqrt{\min_{\mu \in \mathbb{R}^{\mathcal{L}}} \|v_D - \hat{v}_D\|_2^2}, \quad (5)$$

where  $v_D$  is the CR time series restricted to the subdomain  $D$  of  $\mathcal{D}_{\text{CR}}$ .

At this stage, the main detection issue is the definition of an *a priori* threshold on the residual error  $\delta_D$ , in order to decide between changes and no-changes. This threshold should be an appropriate combination of the residual error  $\delta_D$  and the size of the subdomain  $D$ . Indeed, even without changes, larger subdomains are expected to yield larger residual errors since they involve more pixels. The method we present here estimates jointly the label means and the change pixels using a random sampling strategy (see [33]), thus ensuring a high robustness to the amount of change pixels or outliers in the image.

In section III, we build an a-contrario detection model that enables to combine these parameters into a single probabilistic criterion allowing the detection of the most coherent subdomain with a given classification (namely a given high-resolution label map) meanwhile controlling the expected number of false alarms. The complementary of this subdomain is then considered as the set of pixels presenting some changes.

### III. A-CONTRARIO CHANGE DETECTION MODEL

#### A. Definition of a coherence measure

In the framework of a-contrario modeling, the idea is to detect an image subdomain as a large deviation from a naive model. To that aim, we introduce a measurement of the *coherence* between the HR label map describing the state of reference at a given date  $t_0$  and the observed CR image at date  $t$  based on the degree of contradiction which it implies referring to a *non-structured* model (the naive model).

**Definition 3.1 (A-contrario (naive) model for CR):** The a-contrario model ( $H_0(m)$ ) for the CR image is a random Gaussian vector  $V \sim \mathcal{N}(m, \sigma^2 I_{|\mathcal{D}_{\text{CR}}|})$ , where  $m \in \mathbb{R}^{|\mathcal{D}_{\text{CR}}|}$  and  $\sigma > 0$  are fixed, and  $I_{|\mathcal{D}_{\text{CR}}|}$  is the identity matrix in dimension  $|\mathcal{D}_{\text{CR}}|$ .

The choice of the parameters of the naive model is discussed in Section III-B.

Following Helmholtz principle, we focus on image subdomains for which the quadratic error measured between the image and its estimation is too small to be reasonably explained by randomness. Actually, we consider the level of *surprise* of observing an intensity map  $v_D$  on a subdomain  $D$  of  $\mathcal{D}_{\text{CR}}$  which, for a given choice of  $\mu$ , is particularly close to the intensity estimated from  $\mu$  and label proportions within pixels. A detection threshold  $\delta_D$  on the error  $\delta(V_D)$  could then be chosen such as ensuring that

$$\mathbb{P}(\exists D, \delta(V_D) \leq \delta_D) \leq \varepsilon, \quad (6)$$

where  $\varepsilon$  is a fixed parameter (e.g.  $10^{-3}$ ) and  $\delta(V_D)$  is the quadratic error obtained considering the random field  $V$  described by the a-contrario model (Gaussian white noise image). The parameter  $\varepsilon$  enables us to supervise the test reliability. Indeed, the smaller  $\varepsilon$  is, the more demanding and reliable the test. Now, dependences between the random variables  $(\delta(V_D))_{D \in \mathcal{D}_{\text{CR}}}$  are very difficult to estimate, which makes the explicit calculation of  $\mathbb{P}(\exists D, \delta(V_D) \leq \delta_D)$  impossible.

Following the framework introduced in [24], we suggest to measure the expectation of the number of false detections rather than controlling the probability of having at least one false detection. This measure is defined in order to guarantee a number of false detections on random data as low as desired, using a quantization of the number of tested subdomains (through a weighting coefficient called “number of tests”).

**Proposition 3.1 (Number of Falses Alarms):** Consider a function  $\eta : \mathbb{N} \rightarrow (0, +\infty)$  and, for any subdomain  $D \subset \mathcal{D}_{\text{CR}}$ , the quantity

$$NFA(D, \delta(v_D), \sigma, m) = \eta(|D|) \cdot \mathbb{P}_{H_0}[\delta(V_D) \leq \delta(v_D)], \quad (7)$$

where  $\mathbb{P}_{H_0}[\delta(V_D) \leq \delta(v_D)]$  stands for the probability of observing a minimal error  $\delta(V_D)$  less than  $\delta(v_D)$  under Hypothesis  $H_0(m)$ . Then, the *NFA* function is a number of falses alarms provided that

$$\sum_{D \in \mathcal{D}_{\text{CR}}} \frac{1}{\eta(|D|)} \leq 1. \quad (8)$$

For any  $\varepsilon > 0$ , we shall say that a subdomain  $D$  of  $\mathcal{D}_{\text{CR}}$  is  $\varepsilon$ -meaningful if  $NFA(D, \delta(v_D), \sigma, m) \leq \varepsilon$ .

**Proof** — Let us set  $N = 2^{|\mathcal{D}_{\text{CR}}|}$  and define, for  $i \in \{1, \dots, N\}$ ,  $X_i = -\delta(V_{D_i})$ , where  $(D_i)_{1 \leq i \leq N}$  represents all subsets of  $\mathcal{D}_{\text{CR}}$ . Taking  $n_i = \eta(|D_i|)$ , we then have

$$NFA(D_i, \delta(v_{D_i}), \sigma, m) = NFA(i, x_i)$$

(the right term being defined in Equation (2)), hence we know thanks to [28] (Proposition 2 p. 318) that (7) defines a Number of False Alarms as soon as Equation (3) — or equivalently, Equation (8) — holds.  $\square$

Choosing a function  $\eta$  that satisfies (8) is then sufficient to guarantee a number of false alarms less than  $\varepsilon$ . In this study, the first natural choice for the  $\eta$  function is a uniform weight on all subdomains of  $\mathcal{D}_{\text{CR}}$ , i.e.  $\eta(|D|) = |\{D, D \subset \mathcal{D}_{\text{CR}}\}| = 2^{|\mathcal{D}_{\text{CR}}|}$ . This choice ensures less than  $\varepsilon$  false detections (on the average) as  $\sum_{D \subset \mathcal{D}_{\text{CR}}} \frac{1}{2^{|\mathcal{D}_{\text{CR}}|}} = 1$ . It distributes the risk uniformly upon all subdomains of  $\mathcal{D}_{\text{CR}}$  (whatever their size

and even if they overlap) as, for all subdomain  $D \in \mathcal{D}_{\text{CR}}$ , the probability of mistakenly detecting  $D$  is

$$\mathbb{P}_{H_0}(\delta(V_D) \leq \delta(v_D)) = \frac{\varepsilon}{2^{|\mathcal{D}_{\text{CR}}|}}, \quad (9)$$

where the right term does not depend on  $D$ . With such a choice, a subdomain with medium size is more likely to be detected by chance, as such subdomains are more numerous.

Now, notice that choosing  $\eta$  is an opportunity to introduce some *a priori* for the detection. In practice, changes generally affect a minor part of the image ( $\leq 50\%$ ) but in order to maintain the genericity of the method, we suggest an alternate function  $\eta$  that equally considers subdomains with any given size.

More precisely, we suggest to distribute the risk uniformly with respect to the subdomain size, by taking for any  $D \subset \mathcal{D}_{\text{CR}}$ ,

$$\eta(|D|) = |\mathcal{D}_{\text{CR}}| \binom{|\mathcal{D}_{\text{CR}}|}{|D|}. \quad (10)$$

This is a valid choice for the weight function  $\eta$ , since it satisfies Equation (8):

$$\begin{aligned} \sum_{D \in \mathcal{D}_{\text{CR}}} \frac{1}{\eta(|D|)} &= \sum_{D \in \mathcal{D}_{\text{CR}}} \frac{1}{|\mathcal{D}_{\text{CR}}| \binom{|\mathcal{D}_{\text{CR}}|}{|D|}} \\ &= \sum_{k=1}^{|\mathcal{D}_{\text{CR}}|} \sum_{|D|=k} \frac{1}{|\mathcal{D}_{\text{CR}}| \binom{|\mathcal{D}_{\text{CR}}|}{|D|}} \\ &= \sum_{k=1}^{|\mathcal{D}_{\text{CR}}|} \binom{|\mathcal{D}_{\text{CR}}|}{k} \frac{1}{|\mathcal{D}_{\text{CR}}| \binom{|\mathcal{D}_{\text{CR}}|}{k}} \\ &= |\mathcal{D}_{\text{CR}}| \frac{1}{|\mathcal{D}_{\text{CR}}|} \\ &= 1. \end{aligned}$$

In the following sections, this choice of  $\eta$  will be considered and we will set  $\varepsilon = 1$  in order to ensure less than one false detection on average while making the method fully automatic.

Using the a-contrario hypothesis given by Definition 3.1, the *NFA* can be computed explicitly.

**Theorem 3.1:** Using the a-contrario hypothesis denoted by  $H_0(m)$ , for all  $m \in \mathbb{R}^{|\mathcal{D}_{\text{CR}}|}$ , the number of false alarms associated to a subdomain  $D$  of an image  $v$  is determined by

$$NFA(D, \delta(v_D), \sigma, m) = \eta(|D|) \cdot f(q, \delta(v_D), \sigma, m), \quad (11)$$

where  $\eta(|D|) = |\mathcal{D}_{\text{CR}}| \binom{|\mathcal{D}_{\text{CR}}|}{|D|}$ ,  $q = |D| - |\mathcal{L}|$  ( $|\mathcal{L}|$  being the number of labels) and, for all  $q \in \mathbb{N}^+$

$$f(q, \delta, \sigma, m) = \frac{1}{\sigma^q (2\pi)^{\frac{q}{2}}} \int_{\mathcal{B}_q(\delta)} e^{-\frac{1}{2} \sum_{i=1}^q \left(\frac{x_i - m_i}{\sigma}\right)^2} dx_1 \dots dx_q,$$

$\mathcal{B}_q(\delta)$  being the ball of  $\mathbb{R}^q$  with center 0 and radius  $\delta$ .

*Proof* — The proof is detailed in Appendix A, it is a consequence of Proposition 3.1 and the explicit computation of the *NFA* function for the chosen hypotheses.  $\square$

## B. Discussion on the parameters

The naive model  $H_0$  (see begining of Section III) used for the definition of this criterion depends on two parameters: the mean  $m$  and the variance  $\sigma^2$ . In this section, the choice of these parameters is discussed. It is mainly guided by the condition of no detection in a noise image (*i.e.* following the naive model) but also by the fact that the *NFA* needs to be minimized.

First, we consider the naive model according to which the mean CR image is a constant image.

**Hypothesis  $H_0^a$ :** The mean vector  $m$  of the naive model  $H_0(m)$  is  $(\theta, \theta, \dots, \theta)^T$  (for a given  $\theta \in \mathbb{R}$ ) and the standard deviation  $\sigma$  is *a priori* fixed.

**Corollary 3.1:** Under hypothesis  $H_0^a$ , the number of false alarms is

$$NFA_a(D, \delta(v_D), \sigma) = |\mathcal{D}_{\text{CR}}| \binom{|\mathcal{D}_{\text{CR}}|}{|D|} \cdot f(q, \delta(v_D), \sigma, 0), \quad (12)$$

with  $q = |D| - |\mathcal{L}|$ .

*Proof* — Let  $A_D = (\alpha_l(y))_{y \in D, l \in \mathcal{L}}$  be the matrix of label proportions restricted to the subdomain  $D$ . The proof comes from Theorem 3.1, remarking that as the sum per line of  $A_D$  is 1, the vector space  $\mathbb{R}(1, 1, \dots, 1)^T$  is included in the image space of  $A_D$ , denoted  $\text{Im} A_D$ . In particular, since the mean vector  $m$  belongs to this set, it satisfies  $A_D m = m$ . Hence, if the matrix  $P$  denotes the orthogonal projection onto the space  $(\text{Im} A_D)^\perp$ , the mean vector projected by  $P$  boils down to the null vector ( $Pm = 0$ ) and

$$\mathbb{P}_{H_0^a}[\delta(V_D) \leq \delta(v_D)] = f(|D| - |\mathcal{L}|, \delta(v_D), \sigma, 0)$$

thanks to (27) (see Appendix A, proof of Theorem 3.1).  $\square$

An alternate hypothesis for the parameters of the  $H_0$  model is to assume any mean CR image and a fixed variance for each pixel.

**Hypothesis  $H_0^b$ :** The mean vector  $m$  of the naive model  $H_0(m)$  is any unknown vector of  $\mathbb{R}^{|\mathcal{D}_{\text{CR}}|}$  and the standard-deviation  $\sigma$  is *a priori* fixed.

As  $m$  is an unknown parameter of the *NFA* model, we are looking for the model which best enables to control the number of false alarms for all  $H_0(m)$  models, where  $m \in \mathbb{R}^{|\mathcal{D}_{\text{CR}}|}$ . Hence, we need to consider the most pessimistic *NFA* given by

$$\arg \max_{m \in \mathbb{R}^{|\mathcal{D}_{\text{CR}}|}} NFA(D, \delta(v_D), \sigma, m).$$

This maximum happens to be reached when the mean of the  $H_0$  model is null, which leads to the following result.

**Proposition 3.2:** Under  $H_0^a$  or  $H_0^b$  hypotheses,

$$NFA(|D|, \delta(v_D), \sigma) = |\mathcal{D}_{\text{CR}}| \binom{|\mathcal{D}_{\text{CR}}|}{|D|} \bar{f}(q, \delta(v_D), \sigma), \quad (13)$$

where,

$$\begin{aligned} \bar{f}(q, \delta, \sigma) &\equiv f(q, \delta, \sigma, 0) \\ &= \frac{1}{\sigma^q (2\pi)^{\frac{q}{2}}} \int_{x_1, q \in \mathcal{B}_q(\delta)} e^{-\frac{1}{2} \sum_{i=1}^q \left(\frac{x_i}{\sigma}\right)^2} dx_1 \dots dx_q, \end{aligned}$$

and  $q = |D| - |\mathcal{L}|$ .

*Proof* — The proof relies on the Lemma B.1 (see appendix B) as its application to  $f$  implies directly

$$\begin{aligned}
& NFA_b(|D|, \delta(v_D), \sigma) \\
&= \max_m |\mathcal{D}_{CR}| \binom{|\mathcal{D}_{CR}|}{|D|} \cdot f(|D| - |\mathcal{L}|, \delta(v_D), \sigma, m) \\
&= |\mathcal{D}_{CR}| \binom{|\mathcal{D}_{CR}|}{|D|} \cdot f(|D| - |\mathcal{L}|, \delta(v_D), \sigma, 0) \\
&= NFA_a(|D|, \delta(v_D), \sigma). \tag{14}
\end{aligned}$$

Hypotheses  $H_0^a$  and  $H_0^b$  lead to the same number of false alarms, denoted  $NFA$ , which proves Proposition 3.2.  $\square$

According to this result, the naive model  $H_0(0)$  is the most *contrary* model in the family of models  $H_0(m)$ , where  $m \in \mathbb{R}^{|\mathcal{D}_{CR}|}$ . Then, considering the mean vector  $m = 0$ , rejecting  $H_0(0)$  implies the reject of all the naive models of the family  $H_0(m)$  and it enables to free from the mean parameter.

Finally, note that numerically speaking the number of false alarms is easier to compute from the following form.

*Corollary 3.2:* Under  $H_0^a$  or  $H_0^b$  hypotheses,

$$NFA(|D|, \delta(v_D), \sigma) = |\mathcal{D}_{CR}| \binom{|\mathcal{D}_{CR}|}{|D|} \Gamma_{inc} \left( \frac{q}{2}, \frac{\delta(v_D)^2}{2\sigma^2} \right), \tag{15}$$

where  $q = |D| - |\mathcal{L}|$  and, for all  $x \geq 0$  and  $a > 0$ ,

$$\Gamma_{inc}(a, x) = \frac{1}{\Gamma(a)} \int_0^x e^{-t} t^{a-1} dt.$$

The obtained criterion hence depends only on the size of the subdomain, on the error which is computed on this subdomain and on the standard deviation of the naive model  $H_0$ . As these first two parameters are directly determined from the studied images for a given sub-domain, the standard deviation of the naive model  $\sigma$  is the only parameter to set *a priori*. Here, we suggest to set it to the empirical variance of the observed image. Such choice is motivated by the fact that it ensures the absence of any detection in a white noise image on average.

### C. Comparison with Multiple testing procedures

The calculation of the number of false alarms is related to the statistical problem of multiple testing [34] as it takes the number of performed tests into account in its definition. Multiple testing procedures have been developped in order to process large amounts of data without increasing the average number of erroneous detections. A classical statistical approach for testing  $N$  random variables while controlling the number of false alarms (also called false positives) would be to threshold the  $N$  p-values individually to a level  $\alpha$ , thus ensuring an average number of false positives less than  $\alpha N$ , *i.e.* non-decreasing according to the number of tests. Assuming that the  $N$  tests are independent, the probability of occurrence of one error of the first order is then  $1 - (1 - \alpha)^N$ , which is *e.g.* 40% for 10 tests with a level 5%.

As far as the change detection issue is concerned, it could be considered as an adequacy test, assuming a theoretical model for the CR image. For instance, a model such as those where the random variables corresponding to each pixel are assumed to be Gaussian, with parameters related to the occupation

rate of each class represented within the pixel and to class characteristics, can be considered as an *a priori* model. The  $2^{|\mathcal{D}_{CR}|}$  tests of all subdomains of  $\mathcal{D}_{CR}$  are then considered with the following hypotheses:

- $H_0(D)$ : “the CR image restricted to the subdomain  $D$  is not structured according to the reference classification.”
- $H_1(D)$ : “the CR image restricted to the subdomain  $D$  is structured according to the reference classification.”

Hence, pixels that are detected as changes are those which satisfy the null hypothesis  $H_0(D)$ . This test can be performed, for instance, with a  $\chi^2$  statistic [35]. Using previous notations, the decision rule consists in rejecting  $H_0(D)$  when  $\delta(V_D) \leq \delta(v_D)$ . If each test of  $H_0(D)$  against  $H_1(D)$  is to the level  $\alpha_D$ , then by definition the thresholds  $\delta(v_D)$  ensure that the probability of mistakenly rejecting  $H_0(D)$  satisfies

$$\mathbb{P}(\delta(V_D) \leq \delta(v_D) | H_0(D)) \leq \alpha_D. \tag{16}$$

After performing  $N$  tests of hypothesis  $H_0 = \cup_{D \in \mathcal{D}_{CR}} H_0(D)$  against  $H_1 = H_0^c$  (so as to consider all sub-domains  $D$ ), it seems natural to reject  $H_0$  for  $H_1$  if there exist a subdomain  $D \in \mathcal{D}_{CR}$  such as the hypothesis  $H_0(D)$  is rejected for  $H_1(D)$ , *i.e.* such as  $\delta(V_D) \leq \delta(v_D)$ . Let us remark that the probability of mistakenly rejecting  $H_0$  then satisfies

$$\begin{aligned}
& \mathbb{P}(\exists D \in \mathcal{D}_{CR}; \delta(V_D) \leq \delta(v_D) | H_0) \\
& \leq \sum_{D \in \mathcal{D}_{CR}} \mathbb{P}(\delta(V_D) \leq \delta(v_D) | H_0^D) \\
& \leq \sum_{D \in \mathcal{D}_{CR}} \alpha_D. \tag{17}
\end{aligned}$$

The naive approach previously mentionned would be to consider that each test is performed to the level  $\alpha_D = \alpha$ , independently from the tested subdomain and, consequently, that the probability of mistakenly rejecting  $H_0$  is less than  $N\alpha$  (*e.g.*  $2^{10000} \times 0.05$  for a  $100 \times 100$  image and a significance level  $\alpha = 0.05$ ). To avoid such an increase of first type errors, the Bonferroni procedure [36] aims at maintaining the probability of false alarms less than  $\alpha > 0$  while testing, for instance, all subdomains to a level  $\alpha_D \leq \frac{\alpha}{N}$ . The  $H_0$  hypothesis is then rejected if  $\alpha_D \leq \frac{\alpha}{N}$  for all  $D \in \mathcal{D}_{CR}$ . This strategy is very restrictive as each individual test is then maintained to a demanding threshold so as to control the rate of false alarms, even though errors of type 2 then increase. Hence, some detections tend to be missed by such procedure. Formally, the calculation of the number of false alarms (defined for  $N = 2^{|\mathcal{D}_{CR}|}$  tests) is equivalent to Bonferroni procedure as it simply boils down to test if  $N\alpha_D \leq \alpha$ . However, their concept is different and, in particular, the  $NFA$  enables to directly control the number of false alarms (false positives). Indeed, in opposition to Bonferroni procedure, it has an intrinsic meaning allowing, for instance, to decide the acceptation of exactly 10 errors out of 1 million tests on average (while a “test of level 10” has no meaning). The  $NFA$  definition that we use in practice is a variant of this procedure as the number of tests ( $N = |\mathcal{D}_{CR}| \binom{|\mathcal{D}_{CR}|}{|D|}$ ) considered for the  $NFA$  penalizes the tested subdomain according to their size, favouring those with large and with small size. Formally, it means testing all subdomains  $D$  with the level  $\alpha_D$  such as  $\alpha_D \leq \frac{\alpha}{|\mathcal{D}_{CR}| \binom{|\mathcal{D}_{CR}|}{|D|}}$ .

This choice enables to ensure an error of type 1 less than  $\alpha$  as

$$\sum_{D \in \mathcal{D}_{CR}} \alpha_D \leq \sum_{D \in \mathcal{D}_{CR}} \frac{\alpha}{|\mathcal{D}_{CR}| \binom{|\mathcal{D}_{CR}|}{|D|}} \leq \alpha. \quad (18)$$

Numerous procedures have been proposed in the literature as alternatives to Bonferroni procedure [37], [38]. They are based on controlling the different error types such as, for instance, the FWER (Family Wise Error Rate) which controls the probability of obtaining at least one false positive, the PCER (Per Comparison Error Rate) which controls the expectation of the proportion of false positives among all tests or the FDR (False Discovery Rate) to control the expected proportion of mistakenly rejected hypotheses. In particular, the widely used Benjamini-Hochberg procedure [38] enables to controlling the FDR meanwhile considering all tests. Then, decision is taken by considering a significant test set instead of using each test individually. More precisely, as a first step, this procedure consist of sorting the p-values

$$p_{(1)} \leq \dots \leq p_{(i)} \leq \dots \leq p_{(N)}$$

then, if  $k = \max\{i : p_{(i)} \leq \frac{i}{N}\alpha\}$  exists, of rejecting  $H_0(i)$  for  $i = 1 \dots k$  and thus ensuring a false discovery rate  $FDR \leq \alpha$ . This procedure is less conservative than Bonferroni's as it accepts all subdomains which p-values are less than  $p_k$  whereas Bonferroni correction does not allow the acceptation of subdomains which p-value is less than the threshold  $\alpha/N$ . Finally, the Benjamini-Hochberg procedure globally detects a set of subdomains that do not contain any change meanwhile accepting some percentage of errors, but no subdomain is detected individually. In our context, considering the union of significant subdomains seems difficult as the detection of a collection of significant subdomains is hard to interpret. Hence, such procedure does not really seem to fit the considered problem. Moreover, the NFA values correspond typically to very weak probabilities (about  $10^{-1000}$ ). Hence, a threshold such as  $\alpha/N$  on the NFA value would not be restrictive.

The relevance of the NFA definition relies on the fact that the expected number of false alarms of  $\varepsilon$ -meaningful events is less than  $\varepsilon$ . This property enables to automatically set the decision thresholds in order to ensure an *a priori* fixed number of false alarms (e.g.  $\varepsilon = 1$ ). Generally, the most meaningful domain is retained, that is the one which minimizes the NFA (as the most coherent subdomain with the label map).

#### IV. THEORETICAL PERFORMANCES

##### A. Image model

The simple image model described here is used in order to consider separately the main detection parameters and, then, to study the sensitivity of the theoretical model according to each parameter. As a first step, let us consider an image  $u$  of the same size and resolution than the label map. Assume that any image  $u$  can be written as  $u = I + b$ , where  $I$  is a piecewise constant image and  $b$  a Gaussian noise image. Given a label map with  $|\mathcal{L}|$  labels, the image  $I$  is assumed to be perfectly superimposable to the label map, with the corresponding label characteristic value mapped to each pixel (i.e. typically, a mean

label value image). The empiric image variance  $\sigma^2$  of  $u$  can then be obtained directly from  $\sigma^2 = \sigma_I^2 + \sigma_b^2$ . Moreover, the estimation of the label characteristics family  $\mu$  is assumed to be accurate in order to analyse the sensitivity of the number of false alarms. The quadratic residual error measured on average for each pixel is then  $\sigma_b^2$ , which implies an average cumulative quadratic error on a subdomain  $D$  of

$$\mathbb{E}[\delta_D^2] = |D| \times \sigma_b^2.$$

Let us consider that the mean square error obtained on a subdomain of size  $|D|$  is  $\delta_{|D|}^2 = |D| \sigma_b^2$ . Referring to Equation (15), the *number of false alarms* associated to a subdomain of cardinal  $|D|$  which is part of a domain of cardinal  $|\mathcal{D}_{CR}|$  can be written as

$$\begin{aligned} & NFA(|D|, \delta_{|D|}, \sigma) \\ &= |\mathcal{D}_{CR}| \binom{|\mathcal{D}_{CR}|}{|D|} \Gamma_{inc} \left( \frac{|D| - |\mathcal{L}|}{2}, \frac{|D|}{2((\frac{\sigma_I}{\sigma_b})^2 + 1)} \right) \end{aligned} \quad (19)$$

This expression highlights the fact that the number of false alarms mainly depends on the following parameters : the considered subdomain size, the number of labels represented in the label map and the variances ratio  $\sigma_I/\sigma_b$ . As this ratio can be interpreted as an image contrast measure, let us denote the image contrast by  $c = \frac{\sigma_I}{\sigma_b}$  (with  $c \geq 0$ ) and introduce the parameter  $\gamma = \frac{1}{c^2+1}$ ,  $\gamma \in (0, 1]$ . Moreover, the considered subdomain cardinal can be expressed as a proportion  $p \in [0, 1]$  of the whole domain cardinal:  $p = |D|/|\mathcal{D}_{CR}|$ . The number of false alarms of an image with contrast  $c$ , associated to a subdomain concerning a proportion  $p$  of the image, can then be expressed as

$$\begin{aligned} & \mathcal{F}(p, |\mathcal{D}_{CR}|, \gamma) \\ &= |\mathcal{D}_{CR}| \binom{|\mathcal{D}_{CR}|}{p|\mathcal{D}_{CR}|} \Gamma_{inc} \left( \frac{p|\mathcal{D}_{CR}| - |\mathcal{L}|}{2}, \gamma \frac{p|\mathcal{D}_{CR}|}{2} \right) \end{aligned} \quad (20)$$

for  $p \in (\frac{|\mathcal{L}|}{|\mathcal{D}_{CR}|}, 1]$  and  $\gamma = \frac{1}{c^2+1} \in (0, 1]$ . This new expression is used for the performance and sensitivity analysis which is developed in the three next sections.

##### B. Sensitivity to the contrast level

In general, image comparison is improved when images have the same dynamic. As satellite image intensity values vary a lot with luminance conditions, the robustness to the contrast level is an important criterion for a change detection method relevancy.

First, let us focus on the particular case of an image with a null contrast ( $c = 0$ , i.e.  $\sigma_I/\sigma_b = 0$ ). Such a case would occur if the image is so noisy that its geometrical structure is swallowed up in the noise (i.e. any  $\sigma_I$  and  $\sigma_b = +\infty$ ) or with any level of noise, if the image is not geometrically structured ( $\sigma_I = 0$ ). As mentionned in Section III-B, the a-contrario model has been chosen so as to ensure the absence of any detection in such images. Proposition 4.1 formally verifies this property as, when the image contrast is null,  $\gamma = 1$ .

*Proposition 4.1:* If  $\gamma = 1$ , for any fixed  $|\mathcal{D}_{CR}| > 1$  and for all  $p \in (\frac{|\mathcal{L}|}{|\mathcal{D}_{CR}|}, 1]$ ,  $\mathcal{F}(p, |\mathcal{D}_{CR}|, 1) \geq 1$ . The proof is given in Appendix C.

More generally, let us focus on the evolution of the *number of false alarms* when the image contrast varies (Proposition 4.2) or when it reaches limit values (Proposition 4.3).

**Proposition 4.2:** For all fixed  $|\mathcal{D}_{\text{CR}}| \in \mathbb{N}^*$  and  $p \in (\frac{|\mathcal{L}|}{|\mathcal{D}_{\text{CR}}|}, 1]$ , the function  $\gamma \mapsto \mathcal{F}(p, |\mathcal{D}_{\text{CR}}|, \gamma)$  is non-decreasing.

*Proof* — For all fixed  $|\mathcal{D}_{\text{CR}}| > 0$  and  $p \in (\frac{|\mathcal{L}|}{|\mathcal{D}_{\text{CR}}|}, 1]$ , the derivative of  $\mathcal{F}$  with respect to  $\gamma$  is

$$\begin{aligned} & \frac{\partial \mathcal{F}}{\partial \gamma}(p, |\mathcal{D}_{\text{CR}}|, \gamma) \\ &= \frac{|\mathcal{D}_{\text{CR}}|^2 p \binom{|\mathcal{D}_{\text{CR}}|}{p|\mathcal{D}_{\text{CR}}|}}{2\Gamma(\frac{p|\mathcal{D}_{\text{CR}}| - |\mathcal{L}|}{2})} e^{-\gamma \frac{p|\mathcal{D}_{\text{CR}}|}{2}} \left( \frac{\gamma p |\mathcal{D}_{\text{CR}}|}{2} \right)^{\frac{p|\mathcal{D}_{\text{CR}}| - |\mathcal{L}|}{2} - 1}, \end{aligned}$$

which is non-negative.  $\square$

As the parameter  $\gamma = \frac{1}{c^2 + 1}$  is inversely proportional to the image contrast, this result can be interpreted for the *NFA* by the fact that any subdomain representing a proportion  $p$  of the image domain is all the more meaningful that the image is contrasted. An image subdomain is hence all the better validated by the *NFA* that its contrast level is high.

Conversely, let us study the behavior of the *NFA* when the image noise is very low or when the geometrical structure of the image widely prevail upon its noise. In such case, the image contrast  $c$  tends to infinity and  $\gamma$  tends to 0.

**Proposition 4.3:** For all fixed  $|\mathcal{D}_{\text{CR}}| > 0$  and  $p \in (\frac{|\mathcal{L}|}{|\mathcal{D}_{\text{CR}}|}, 1]$ ,  $\lim_{\gamma \rightarrow 0} \mathcal{F}(p, |\mathcal{D}_{\text{CR}}|, \gamma) = 0$ .

*Proof* — As the function  $\gamma \mapsto \mathcal{F}(p, |\mathcal{D}_{\text{CR}}|, \gamma)$  is  $\mathcal{C}^\infty$ ,  $\lim_{\gamma \rightarrow 0} \mathcal{F}(p, |\mathcal{D}_{\text{CR}}|, \gamma) = \mathcal{F}(p, |\mathcal{D}_{\text{CR}}|, 0) = 0$ .  $\square$

From the point of view of the *NFA*, this property points out that any subdomain is detectable as soon as the image contrast is high enough. Even though intuitive, this result proves in particular that images containing a large proportion of changes can be analyzed as soon as their contrast is high enough.

### C. Impact of the image size

Here, the image contrast level is assumed fixed and the behavior of the *NFA* is analysed with respect to the image size.

**Proposition 4.4:** For all fixed  $\gamma$  within the interval  $(0, 1)$ , there exist  $p^*(\gamma)$  such that, for all  $p \in [p^*(\gamma), 1]$ ,

$$\lim_{|\mathcal{D}_{\text{CR}}| \rightarrow +\infty} \mathcal{F}(p, |\mathcal{D}_{\text{CR}}|, \gamma) = 0.$$

The proof is given in Appendix D.

This result can be interpreted as follows: for a given contrast level (any fixed  $\gamma$ ), the a-contrario method enables to detect all change pixels when they represent a proportion of the image domain which is smaller than  $1 - p^*(\gamma)$ , provided that the image is large enough. Figure 1 shows how this value  $p^*$  evolves when  $\gamma$  is increasing (that is, when the contrast level is decreasing). Even though this result is asymptotic, practically it suggests to use as much data as possible when dealing with weak contrast images, rather than restricting to an extract of the area of interest.

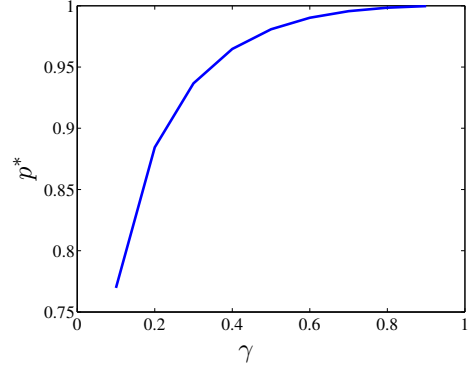


Fig. 1. Plot of the  $p^*$  values obtained numerically for various values of  $\gamma \in (0, 1]$ .

## V. ALGORITHM

From the definition of the *NFA* (Equation (11)), a subdomain is detected as coherent with a given label map if it has a small *NFA* (and the smaller the *NFA*, the more coherent the subdomain). This latter depends on the size of the CR image domain, the number of labels, the size of the considered subdomain  $D$ , the standard-deviation of the naive model and the quadratic residue cumulated on this subdomain. Note that all *NFA* parameters are obtained directly from the data except the cumulated quadratic residue  $\delta^2(v_D)$  which depends on the class means (*a priori* unknown). It is hence necessary to estimate the mean of each class before being able to compute the quadratic residues considered subdomain and then the corresponding *NFA*. The mean estimation and the detection itself are two linked problems as the quality of the estimation has a strong impact on the performance of the detection.

In general, the goal of usual estimation methods such as the mean square method is to maximize, for a given measure, the adequacy between a defined model and the data. Such methods are very sensitive to outliers as they aim at getting as close as possible to the whole sample. While dealing with change detection, change pixels play the role of outliers and, when numerous, they may dramatically bias the mean estimation. Robust methods such as M-estimators, LMedS or RANSAC have been introduced in order to face this problem. For instance, the M-estimators [39] enable a good estimation even when 50% of the pixels are outliers, by weighting the considered distance. However, the choice of the weight function is delicate.

The RANSAC (Random Sampling Consensus) method, proposed by [33], is based on the idea of using a sub-sample as small as possible and to fill it out with consistent data when possible, rather than using as many data as possible. More precisely, let us assume that a subsample of data is selected randomly and that parameters are estimated from this subsample. By chance, the considered subsample might happen not to contain any outlier, the parameter estimation is then correct and classifies the whole sample between correct and incorrect pixels. The RANSAC strategy is based on the idea that repeating this random sampling process a large amount of times must lead to a satisfactory solution. This approach, introduced in image analysis by [33], enables a



good parameter estimation even when outliers are numerous (about 50%). However, it is limited by the arbitrary choice of a threshold from which deciding that a subsample is in adequacy with the considered model.

In addition, minimizing the  $NFA$  ideally requires a combinatory exploration of all image subdomains. However, such a thorough research is not conceivable in practice as, even for a very small image of 256 pixels,  $2^{256} \simeq 1.15 \cdot 10^{77}$  subdomains should be analyzed. By noting that the best subdomains for the mean estimation are those minimizing the residues, we suggest to limit the exploration to the subdomains selected using a RANSAC-like strategy.

Hence, the detection algorithm we propose is based both on the random sampling strategy (cf. [33]) and on the  $NFA$  probabilistic criterion defined in Section III in order to free from the choice of an arbitrary decision threshold during the parameter estimation step meanwhile reducing the number of test subdomains. Such a collaboration has already been performed in [26] for rigidity detection and estimation of the fundamental matrix in stereoscopic vision.

The research of the minimum  $NFA$  can be accelerated using the following result.

*Proposition 5.1:* Given a fixed vector  $\mu \in \mathbb{R}^{\mathcal{L}}$ , define for any  $D \subset \mathcal{D}_{CR}$  the square error  $\delta_{\mu}^2(v_D) = \sum_{y \in D} r(y)$ , with  $r(y) = (v(y) - \sum_{l \in \mathcal{L}} \alpha_l(y) \mu(l))^2$ . If  $\mathcal{D}_{CR} = \{y_i\}_{i=1 \dots |\mathcal{D}_{CR}|}$  and the function  $i \mapsto r(y_i)$  is nondecreasing, then

$$\arg \min_{\substack{D \subset \mathcal{D}_{CR} \\ |D|=q}} NFA(|D|, \delta_{\mu}(v_D), \sigma) = \{y_1, \dots, y_q\}. \quad (21)$$

*Proof* — Since  $r$  is sorted, for all  $D \in \mathcal{D}_{CR}$  such that  $|D| = q$  we have  $\delta_{\mu}^2(v_D) \geq \delta_{\mu}^2(v_{\{y_1, y_2, \dots, y_q\}})$ . As the  $NFA$  function is non-decreasing with respect to  $\delta$ , its minimum value is obtained for the subdomain  $\{y_1, \dots, y_q\}$ .  $\square$

Hence, if the quadratic residues are sorted with a non-decreasing order, considering each subdomain following the order of its residues is sufficient to minimize the  $NFA$  over all subdomains, for a given vector  $\mu$  of label means.

Finally, in the case where a single CR image  $v$  is used, the algorithm is:

- Compute  $\sigma^2$ , the CR image variance.
- Initialize  $\delta_{min}^2$ ,  $NFA$  and  $NFA_{min}$  to  $+\infty$ .
- Repeat  $N$  times
  - 1) draw randomly  $|\mathcal{L}|$  CR pixels, denoted by  $I = (x_1, \dots, x_{|\mathcal{L}|})$ ;
  - 2) estimate the label mean vector  $\mu$  from equations

$$v(x) = \sum_l \alpha_l(x) \mu_l,$$

defined for  $x \in I$ ;

- 3) compute  $r(x) = (v(x) - \sum_l \alpha_l(y) \mu_l)^2$ , for  $x \in \mathcal{D}_{CR}$ ;
- 4) sort  $\mathcal{D}_{CR}$  into a vector  $(x_i)_{1 \leq i \leq |\mathcal{D}_{CR}|}$  by increasing error  $r(x_i)$ ;
- 5) initialize  $\delta^2 = \sum_{i=0}^{|\mathcal{L}|} r(x_i)$ ;
- 6) for each index  $i \in \{|\mathcal{L}| + 1, \dots, |\mathcal{D}_{CR}|\}$ ,
  - set  $\delta^2 = \delta^2 + r(x_i)$ ;

- if  $\delta^2 < \delta_{min}^2$  then
  - \* set  $\delta_{min}^2 = \delta^2$ ;
  - \* compute the corresponding  $NFA[i]$  value;
  - \* if  $NFA[i] \leq NFA_{min}$ , then
    - set  $NFA_{min} = NFA[i]$ ;
    - set  $D = \{x_k\}_{k=1 \dots i}$ ;
  - \* end if
- end if
- 7) end for
- end repeat

This algorithm uses a HR label map and a CR image as inputs and returns the subdomain minimizing the  $NFA$  and the corresponding class means. The only parameter of the algorithm is the number of iterations ( $N$ ). Due to RANSAC strategy, convergence requires a very high number of iterations ( $N = 100\,000$  in our experiments). However, this is not really a limiting factor as the computation time of each iteration is very fast (for instance, 100 000 iterations for change detection considering a HR classification of size  $256 \times 256$  and a CR image of size  $16 \times 16$  takes about 10s on a laptop).

## VI. THE MULTITEMPORAL CASE

In the multitemporal case, different approaches may be chosen depending whether the application requires the detection of a spatio-temporal subdomain or a spatial subdomain. For instance, a sequential approach can be considered, comparing the minimum  $NFA$  values obtained for each image separately. Such an approach can be used in order to find the image of a time series which is the most coherent with the classification but it does not permit to detect a spatio-temporal subdomain. In practice, time series are rather used to analyse the temporal evolution of intensities and to enable the detection of spatio-temporal domains, which may be useful for applications where changes can occur at some dates without impacting other ones.

Here, a vectorial approach is considered for the detection of a spatio-temporal subdomain of changes, assuming that all images of the time series are accurately registered. Denote  $\mathcal{T}$  the set of acquisition dates of a time series. The a-contrario detection model presented Section III can be easily extended to time series, considering a spatio-temporal subdomain  $\omega \in \mathcal{D}_{CR} \times \mathcal{T}$ . As a naive model, the CR time series is assumed to be a random field of  $|\mathcal{D}_{CR}| \times |\mathcal{T}|$  independent Gaussian random variables of zero-mean and variance  $\sigma^2$ . From there, the  $NFA$  is defined as in the monotemporal case by

$$NFA(|\omega|, \delta(v_{\omega}), \sigma) = \eta(|\omega|) \cdot \Gamma_{inc} \left( \frac{q}{2}, \frac{\delta(v_{\omega})^2}{2\sigma^2} \right), \quad (22)$$

where  $\eta(|\omega|) = |\mathcal{D}_{CR}| \cdot |\mathcal{T}| \cdot \binom{|\mathcal{D}_{CR}| \times |\mathcal{T}|}{|\omega|}$  and  $q = |\omega| - |\mathcal{L}| \times |\mathcal{T}|$ . Concerning the choice of the variance of the naive model, let us recall that, in the monotemporal case, taking the CR image variance as the variance of the naive model was justified by the fact that nothing should be detected in a white noise image. In the multitemporal case, setting the variance of the naive model to the variance of the CR time series does not ensure this property anymore as, in the case of high variance differences between dates within the time series, such a naive

model could detect noisy pixels as coherent. To avoid the detection of irrelevant subdomains, the intensity values are normalized by the image variance within each image and the variance of the naive model is set to 1. The multitemporal algorithm is very close to the monotemporal one but the fact that time series may contain missing data at different locations for each date must be taken into account in the research of the subdomain of changes. A simple possibility is to restrict the study to the set of pixels that are validated for all dates but this might considerably reduce the analysed domain. Instead, we propose to consider a restriction of  $\mathcal{D}_{CR} \times \mathcal{T}$  to the set of valid pixels. More precisely, each pixel of the spatial domain is considered as a vector whose coordinates correspond to each of its valid dates. If a pixel is impacted by some changes at a given date, it will then be rejected for the whole series. In order to allow the comparison of subdomains of different size for each date, the cumulated residues are normalized by the number of valid dates, leading to a mean cumulated residue. Notice that with such an exploration, for a given pixel, the mean residue value is duplicated as many times as there are valid dates for this pixel thus enabling the detection of a spatial subdomain for which all valid dates are meaningfully coherent with the reference classification. From there, the same algorithm as in Section V can be used.

## VII. RESULTS

In this section, the unique parameter of the a-contrario change detection method,  $\varepsilon$ , was set to 1 for all experiments so as to ensure less than 1 false alarms on average.

### A. Experimental performances

Some experiments have been conducted in order to evaluate the performance and the limits of the method when the amount of change pixels varies in the image and when the objects of interest are small relative to the CR pixel size (*i.e.* when the resolution ratio is important). Simulated data have been used in order to enable a quantitative estimation of the performance while controlling various impacting parameters.

First, let us focus on the robustness of the method when the amount of change pixels (or outliers) varies in the CR image. To that aim, a fixed resolution ratio of  $16 \times 16$  has been considered and CR images containing 256 pixels have been simulated with an average contrast level. Change pixels have been introduced by corrupting the CR images with an impulse noise (random intensity values assigned to random pixels) impacting 0 to 100% of the image. The method has then been run for each image of the data set (500 tests) and the different types or detection error have been counted:

- the set of **true positives** is the set of all pixels which are rightly detected as coherent with the reference classification. Conversely, the set of false positives is the set of all pixels which are wrongly detected as coherent with the reference classification (errors of type 1).
- the set of **true negatives** is the set of all pixels which are rightly not detected as coherent with the reference classification (*i.e.* considered as *changes*). Conversely, the set of false negatives is the set of all pixels which are not

detected as coherent but should have been (errors of type 2).

Figure 2 shows the different types of errors obtained versus the number of change pixels. On both plots, each dot represents a test. Figure 2(a) represents the number of true positives (in green), the number of false negatives (in blue) and the number of pixels that effectively correspond to no change (in red, plotted just to increase the readability) versus the actual number of change pixels in the image (the line  $y = 1 - x$ ). Thus, the closer the true positives dots are to the no-change pixels dots, the more competitive is the detection.

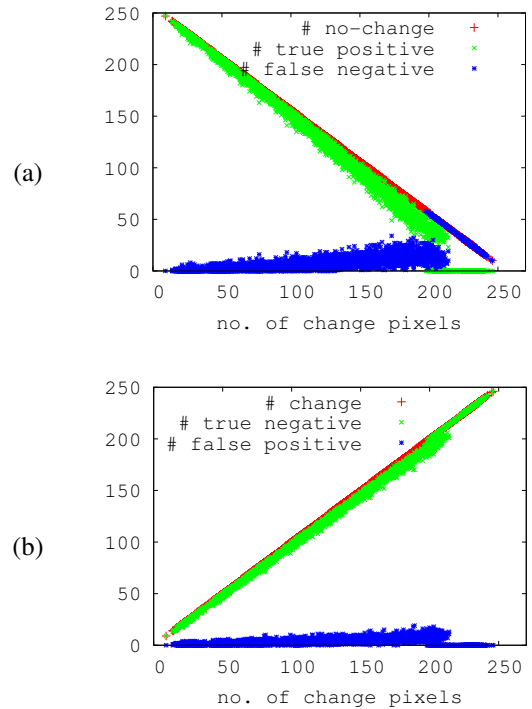


Fig. 2. Evolution of the performances versus the number of change pixels in an image with 256 pixels. On the left, the red dots on the diagonal stand for the actual no-change pixels in the test image, the number of pixels detected erroneously as changes (false negatives) is represented in blue, and the number of pixels accurately detected as no-change (true positives) is represented in green. On the right, the red dots on the diagonal stand for the change pixels, the number of pixels erroneously detected as no-change is represented in blue cross and, in green, the number of pixels detected as changes. The quality of the detection is thus shown as the closer the set of true positives gets to the set of no change pixels, the better is the detection performance.

Figure 2(b) represents the number of true negatives (in green), of false positives (in blue) and of actual change pixels (in red) versus the number of actual change pixels. It appears that the closer the true negatives dots are to the  $x = y$  line, the better are the detection performances. Besides, these two plots show the absence of detection only when the number of change pixels in the image represents more than about 80% of the CR image. Such performance is particularly high compared to the usual limitation threshold of 25% or 30% of change pixels found in the literature (*cf.* [23]). This result confirms the asymptotic theoretical result mentioned Section IV. Besides, the fact that the method is based on the control of the average number of false alarms (false positives) appears comparing

Figures 2(a) and 2(b) (plots in blue). Indeed, the number of false positives (Figure 2(b)) barely increases when the number of change pixels increases, whereas the number of false negatives (Figure 2(a)) increases more clearly with the number of change pixels.

Finally, the performance of the method has been studied when the change areas only partially impact a CR pixel. Such case occurs whenever the change object of interest is smaller than the CR pixel. It is one of the main challenge addressed by the method. To that aim, a fixed average contrast level and a fixed global amount of change pixels of 20% are considered for the data set, and changes are simulated with a size varying from 0 to 100% of the CR pixel. Figure 3 presents the global error percentage obtained as a function of the percentage of changes within the CR pixel. It appears that the method is capable of detecting changes with less than 5% errors (median) as soon as their impacted surface represents less than 13% of the CR pixel, and with less than 3% errors when changes represents more than 25% of the CR pixel. The performance of the detection is barely increasing after this threshold.

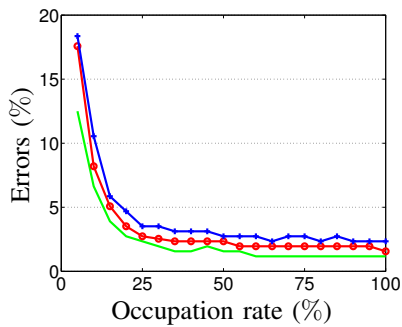


Fig. 3. Errors of detection (% of the number of pixels in the image) obtained versus the occupation rate of changes within a CR pixel : 25, 50 and 75 percentiles (respectively in green, red and blue). Performances obtained for images simulated with 20% of change pixels and a resolution ratio of  $16 \times 16$ , showing that changes impacting more than 13% of a CR pixel are detected with less than 5% of error (in median).

### B. Application to land cover change detection

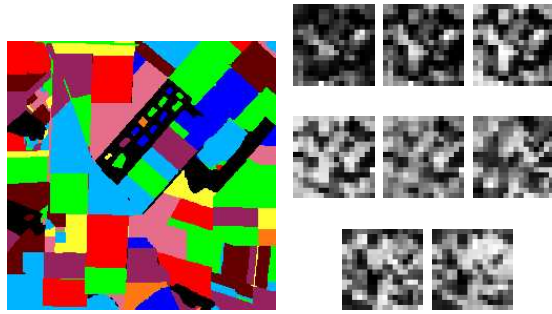
In this section, a case of application to remotely sensed imagery is presented. Using Earth observation data, changes of interest are, typically, natural phenomena such as vegetation growth, flooding or fires and phenomena related to human activities such as urbanism, forest cuts, crop rotation or pollution. Here, a real time series of HR (SPOT 4) images of an area of intensive agricultural practice in the Danubian plain (Rumania, ADAM database<sup>1</sup>) has been considered. As no groundtruth on changes was available, a HR classification has been derived from the HR time series and changes have been simulated either on the HR classification or on the CR image (obtained by averaging the real HR time series by blocks  $16 \times 16$ ).

Figure 4 (a) shows the HR classification derived from the 8 HR real images (Fig. 4 (b)) and containing 10 labels

(using [40]). The change detection method applied to this classification and the corresponding CR time series with 8 dates enabled to validate the whole domain except the pixel in red Figure 4 (c). However this latter pixel shall not be considered as a false detection as it involves the quality of the classification rather than the change detection method in itself. We then artificially introduced different changes in the HR classification. First, in Figure 4, changes were introduced in the reference classification by replacing a random selection of segments label with another existing label. Figure 4(d) to (f) present several cases of such simulated changes and pixels detected as changes are presented in red in the CR domain. On the same image, the boundary of each segment is plotted in black, pixels corresponding to simulated changes are represented in green and those that were already detected in (c) are in pink. The small number of segments enables to visualize the impact area of segments of interest within CR pixels. Changes are well detected by the method, even when they impact a very small area of a CR pixel (Figure 4(d) and (f)).

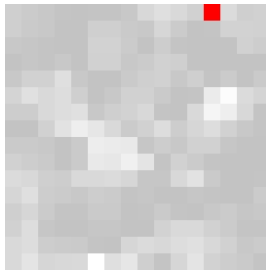
We now aim at evaluating the interest of our approach relatively to state of the art methods. In the literature, most change detection methods apply to images having the same spatial resolution (see [18], [19]). In general, they assume a Markov Random Field model for the change detection image (as in [15], [16]), which is adapted only to the detection of objects with a size greater than the pixel size (typically the field surface in our remote sensing application), or they boil down to an automatic thresholding of the difference image (*e.g.* [41]–[43]). As the spatial resolution we consider in this paper is generally too coarse relatively to the objects of interest, we focused on the sub-pixel change detection problem which cannot be compared directly to the previously mentioned methods. Alternately, few works are dedicated to the sub-pixel problem. We choose to compare our method to the IE method described in [23] as it considers a very similar problem (comparison of a CR series with a HR reference label map). In [23], change detection is performed through an alternate estimation of class means (pixel disaggregation problem) and pixel composition (supervised subpixel classification problem) removing pixels causing largest errors (assumed to correspond to changes). The weak points of such an approach are as follows. First, during the pixel disaggregation step, the number of considered pixels should be greater than the number of classes (in a ratio greater than two in order to get good estimations). Then, during the pixel composition estimation, the number of considered bands (*e.g.* the number of dates times the number of spectral bands) should be greater than the number of classes (also in a ratio greater than two if possible). Lastly, the number of change pixels should not be too important since they are present at the beginning of the iterative process (practically no more than 30% of the total number of pixels). Figure 5 shows the obtained results in the case of changes introduced the reference classification 5(a), as for Fig. 4. More precisely, our aim is to detect changes occurring in the CR 8-dates time series 4(c) referring to the HR 5-classes classification Fig. 5(a). Indeed, as the iterative estimation (IE) method needs more dates in the time series

<sup>1</sup>The ADAM database (<http://kalideos.cnes.fr>) has been constituted in the context of the ADAM project in order to assess the potential of spatial data assimilation techniques with agronomic models.



(a) classification

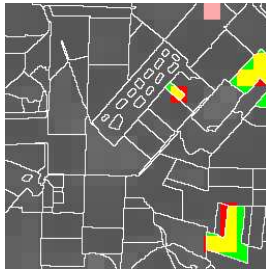
(b) CR time series



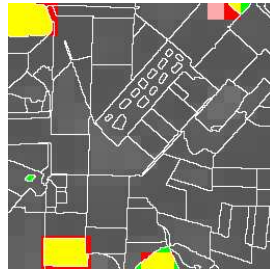
(c) Detected changes



(d) 2 changes



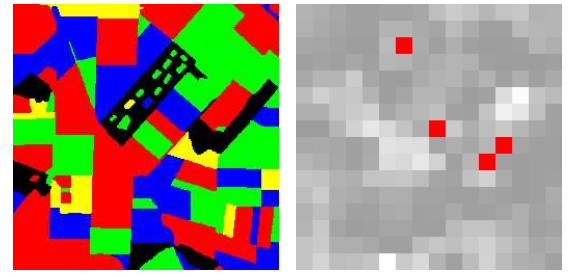
(e) 3 changes



(f) 4 changes

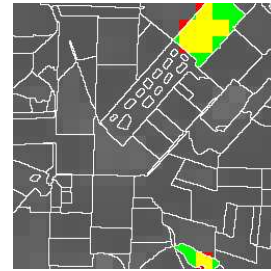
Fig. 4. Results obtained for changes introduced in the classification by a random choice of 3, 4 or 5 segments, to which are assigned random labels between 1 and  $L$ . Changes that have been simulated in the classification are represented in yellow when they are detected, green otherwise. Detected pixels that do not correspond to changes are represented in pink if they were already detected before the simulation of changes (cf. (c)) and in red otherwise. Globally, remark that simulated changes are well detected even when they impact a weak proportion of a CR pixel. On Figure (e), missed detections can be observed (to the right, in green).

than the number of classes in the classification, we had to reduce the number of classes to 5 so that the IE method performs well. The colour code for change and/or detected pixels is the same as in Fig. 4. This time, no pixel has been detected running our method (referred as NFA) for the CR time series Fig. 4(b) and the corresponding classification Fig. 5(a) before introducing some changes, showing that all CR pixels are coherent with the reference classification. Conversely, Fig. 5(b) shows in red the pixels which are detected using IE for the same experiment. Hence, these four pixels

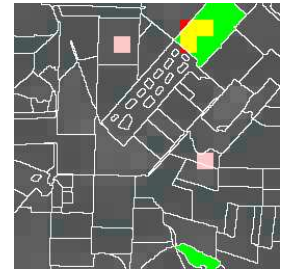


(a) classification

(b) Detected changes



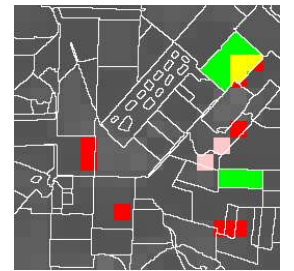
(c) 2 changes (NFA)



(d) 2 changes (IE)



(e) 3 changes (NFA)



(f) 3 changes (IE)

Fig. 5. Comparison of NFA and IE methods for changes introduced in the classification by a random choice of 2 and 3 segments, and a new label between 1 and  $L$  assigned to each chosen segment. Almost all changes are detected and almost no false alarm is obtained using the NFA whereas missed detections and false alarms are obtained with IE, thus confirming the high performance of the NFA.

are considered as *not valid* for the next change detection experiments using IE. Figures 5(c) and (d) present the results obtained respectively using the NFA and the IE methods when 2 changes have been introduced in the classification (in green): using the NFA method, we obtain 7 true alarms (in yellow), 0 false alarm and 8 missed alarms (only one of which impacts the whole pixel); using the IE method, 3 true detections, 12 missed alarms have been obtained. Note that the two pixels in pink are not considered as false alarms as they were already detected in Fig.5(b). Figures 5(e) and (f) show the results of the NFA and IE for 3 other changes: 10 true detections, 1 false alarm and 3 missed detections can be observed using the NFA whereas only 3 true detections can be observed using the IE (and 6 false alarms, 10 missed detections). Through these



results, the better performance of the NFA compared to the IE appears clearly. In addition, as mentioned before, the case of application of the IE is much more limited than the NFA. Besides, the good control of the number of false alarms using the NFA is again proven.

Next, changes are simulated on the CR image, standing for wide area changes, whereas the previous experiments validated sub-pixel land cover change detection. To that aim, the HR classification Figure 6 (a) is considered as a reference and changes are simulated on the CR image Figure 6 (b) replacing the pixel values within two rectangular areas (see in the white areas Figure 6(c)) by the pixel values obtained on another part of the image, for another date. The method applied to Fig. 6 (a) and (b) enabled us to detect all red pixels Figure (c). The pixels in pink correspond to pixels that were already detected before the simulation of changes in Fig. 6 (b).

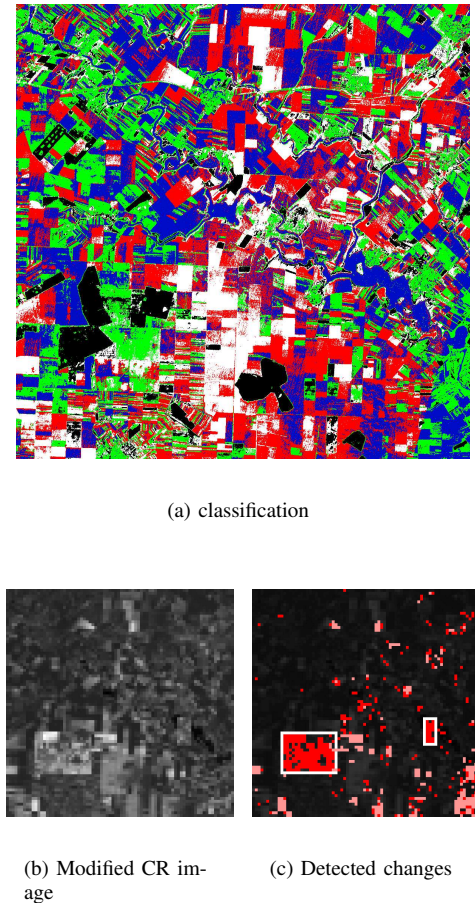


Fig. 6. Detection of changes introduced in the CR image (b) corresponding to the HR classification (a): detected changes are represented in red Figure (c) and the boundary of introduced changes is represented in white in the same image. Changes concern 5.46% of the CR pixels and 96.2% of the image had been validated before introducing changes. Detected pixels concerns 89.3% of the pixels, which is close to the expected 90.7%.

An important aspect of this method is the resolution ratio between HR and CR. The comparison of the results presented Figure 7 shows the robustness of the method with respect to the resolution ratio. Indeed, in a monotemporal context, the change detection method has been applied to the classification

shown on Figure 6 (a) and a corresponding HR image averaged by blocks of size  $5 \times 5$  (Figure 7(b)),  $15 \times 15$  (Figure 7(c)),  $30 \times 30$  (Figure 7(d)) and  $50 \times 50$  (Figure 7(e)). Even though we cannot discuss the accuracy of the detected changes (as no groundtruth were available), notice that in these four cases, about 4.5% of the pixels are detected and that the spatial location of the detected pixels as non-coherent (in red) is stable, showing the good robustness of the method with respect to the resolution ratio. Moreover, as most methods in the literature are based on the analysis of the difference image, Figure 7(a) presents the difference image between the radiometric images acquired at the two dates which are considered for change detection Figure 7(b) to (e) thus illustrating the interest of using a classification instead of a radiometric image as a reference: the difference grey levels are very spread (difficult choice of a threshold) and many areas are characterized by large difference values but do not necessarily correspond to changes of interest.

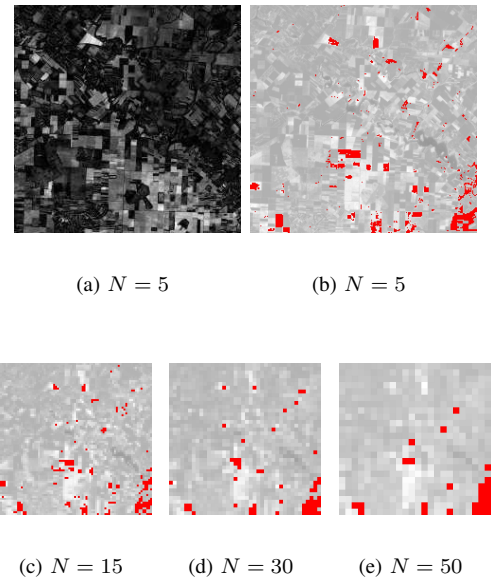


Fig. 7. Change detection using the HR classification of Figure 6 (a) and a CR image with a resolution ratio ( $N$ ) of  $5 \times 5$ ,  $15 \times 15$ ,  $30 \times 30$  and  $50 \times 50$ . Detected pixels are presented in red, superimposed on the CR image used: the same areas are detected when the resolution ratio increases.

## VIII. CONCLUSION

In this paper, a fully unsupervised method has been provided for sub-pixel change detection. The change detection problem has been considered comparing a CR time series to a HR classification at a reference date. This formulation minimizes the required prior on the expected evolution, taking into account a reference classification rather than an image. Moreover, it allows the detection of sub-pixel changes.

The method we proposed is based on the definition of a coherence measure between a CR time series and the HR classification, using an *a-contrario* framework which does not require any *a priori* information. Indeed, rather than providing an *a priori* model for the data, the method is based on the

reject of a naive model (*non-structured*) by the observation of structured data. A theoretical analysis of this model underlined properties announcing its performances as functions of the image contrast level or of the amount of change pixels in the image.

In practice, utilizing this model requires the estimation of class means and the research of the spatio-temporal subdomain minimizing the *NFA*. To that aim, a stochastic algorithm using a RANSAC strategy has been used. It enables a robust estimation even when numerous pixels are outliers. Indeed, whereas existing methods are generally limited to less than 20% outliers in the image, this approach enables to override these limits showing good performances with up to 70% outliers. Besides, simulations enabled an experimental assessment of the performances. In particular, it appeared that for an average contrast level, changes impacting more than 25% of a CR pixel are accurately detected as soon as less than 65% of the image is impacted.

The multitemporal case has been discussed in particular as it implies to deal with the chronic issue of missing data. An adapted extension of the algorithm has been proposed, taking into account the fact that a time series often shows high variabilities between two dates. However, recall that this approach theoretically applies to any type of image series. Further studies and experiments will concern other cases of application using hyperspectral images or video sequences. As an instance, Fig. 8 shows some results obtained by testing our approach on an Infra-Red sequence acquired in the framework of video surveillance application (Safearound project). In this application, the CR image is the background image both for memory save and low-pass filtering of noise, and the HR image is the radiometric image just acquired. First, this latter HR image is quantized (in 10 grey levels Fig. 8(b)). Then the proposed algorithm can be applied directly. In Fig. 8(c), the intruder clambering up the wire fence is well detected (no missed alarm, no false detection) even if he impacts only partly two CR pixels. As a comparison, Fig. 8(d) corresponds to the result finding the best threshold on the difference image (between HR images): the intruder is correctly detected but several false alarms (cars and hut roof) can also be observed, thus illustrating the interest of our method.

Lastly, the results obtained using pseudo-real data showed very good performance and robustness to the resolution ratio used. However, further validation on real time series with known changes are still to be performed, in order to analyse in particular the sensitivity of the model to other departures from the block average hypothesis. Moreover, this approach is based on the assumption of perfect image registration. Further work should focus on a registration sensitivity analysis as, in reality, registration is not perfect and the use of misregistered time series would lead to cumulated errors.

#### APPENDIX A THEOREM 3.1

Using the a-contrario hypothesis denoted by  $H_0(m)$ , for all  $m \in \mathbb{R}^{|\mathcal{D}_{CR}|}$ , the number of false alarms associated to a

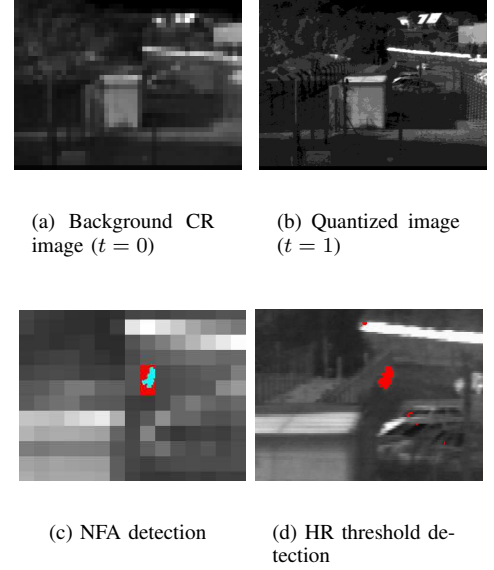


Fig. 8. Video surveillance using Infra-Red images. The background CR image (a) is an image of the sequence without any event of interest. In image (b), an intruder appears (quantized into 10 grey levels). The result of the NFA method (Image (c)) using (a) and (b) shows in (c) in red the CR detected pixels and in blue the track of the intruder (no false alarm, no missed detection). Image (d) presents in red the detected pixels using a thresholding method with HR images: the intruder is well detected but several false alarms can also be observed.

subdomain  $D$  of an image  $v$  is determined by

$$NFA(D, \delta(v_D), \sigma, m) = \eta(|D|) \cdot f(q, \delta(v_D), \sigma, m), \quad (23)$$

where  $\eta(|D|) = |\mathcal{D}_{CR}| \binom{|\mathcal{D}_{CR}|}{|D|}$ ,  $q = |D| - |\mathcal{L}|$  ( $|\mathcal{L}|$  being the number of labels) and, for all  $q \in \mathbb{N}^+$

$$f(q, \delta, \sigma, m) = \frac{1}{\sigma^q (2\pi)^{\frac{q}{2}}} \int_{\mathcal{B}_q(\delta)} e^{-\frac{1}{2} \sum_{i=1}^q \left( \frac{x_i - m_i}{\sigma} \right)^2} dx_1 \dots dx_q,$$

$\mathcal{B}_q(\delta)$  being the ball of  $\mathbb{R}^q$  with center 0 and radius  $\delta$ .

*Proof* — From Proposition 3.1, specifying explicitly the *NFA* requires the calculation, for a given  $\delta(v_D)$ , of the probability

$$\mathbb{P}_{H_0}[\delta(V_D) \leq \delta(v_D)]. \quad (24)$$

Let  $A_D = (\alpha_l(y))_{y \in D, l \in \mathcal{L}}$  be the matrix of label proportion restricted to the subdomain  $D$  (matrix of size  $|D| \times |\mathcal{L}|$ ). The residual error

$$\delta(V_D) = \min_{\mu \in \mathbb{R}^{\mathcal{L}}} \sqrt{\sum_{y \in D} \left( V(y) - \sum_{l \in \mathcal{L}} \alpha_l(y) \mu(l) \right)^2} \quad (25)$$

can be interpreted as a distance from the vector  $V_D$  to the matrix  $A_D$  image space denoted by  $\text{Im} A_D$ . The minimal distance is hence the orthogonal projection of  $V_D$  on the space  $(\text{Im} A_D)^\perp$ , i.e.

$$\delta(V_D) = \min_{\mu \in \mathbb{R}^{\mathcal{L}}} \|V_D - A_D \mu\| = \|p_{(\text{Im} A_D)^\perp}(V_D)\|. \quad (26)$$

According to Hypothesis  $H_0$ , the vector  $V_D$  follows a Gaussian law  $\mathcal{N}(m_D, \sigma^2 I_{|D|})$ , where  $m_D$  is the mean vector  $m$

restricted to the subdomain  $D$  and  $I_D$  is the identity matrix in dimension  $|D|$ . Let  $P$  be the orthogonal projection matrix of  $p(\text{Im}A_D)^\perp$ , then

$$\mathbb{P}[\delta(V_D) \leq \delta(v_D)] = \mathbb{P}[\|PV_D\| \leq \delta(v_D)]. \quad (27)$$

As  $\mathbb{R}^{|D|} = \text{Im}A_D \oplus (\text{Im}A_D)^\perp$ , there is a basis  $e = (e_1, \dots, e_{|D|})$  in which the projection matrix  $P$  is such as

$$\text{Mat}_e(p(\text{Im}A_D)^\perp) = \begin{pmatrix} I_q & 0 \\ 0 & 0_{|D|-q} \end{pmatrix}$$

where  $q = \dim(\text{Im}A_D)$  et  $|D| - q = \dim(\text{Im}A_D)^\perp$ . As

$$\mathbb{P}(PV_D \in \mathcal{B}^q(\delta)) = \mathbb{P}(V_D \in P^{-1}(\mathcal{B}^q(\delta))) \quad (28)$$

and knowing that  $P^{-1}(\mathcal{B}^q(\delta)) = \mathcal{B}^q(\delta) \times \mathbb{R}^{|D|-q}$ , we get

$$\begin{aligned} & \mathbb{P}(V_D \in P^{-1}(\mathcal{B}^q(\delta))) \\ &= \frac{1}{(2\pi)^{\frac{|D|}{2}} \sqrt{\det(\sigma^2 I_{|D|})}} \\ & \quad \times \int_{x_1^q \in \mathcal{B}^q(\delta)} \int_{x_{q+1}^{|D|} \in \mathbb{R}^{|D|-q}} e^{-\frac{1}{2} \sum_{i=1}^{|D|} \left(\frac{x_i - m_i}{\sigma}\right)^2} dx_1 \dots dx_{|D|} \\ &= \frac{1}{\sigma^{|D|} (2\pi)^{\frac{|D|}{2}}} \int_{x_1^q \in \mathcal{B}^q(\delta)} e^{-\frac{1}{2} \sum_{i=1}^q \left(\frac{x_i - m_i}{\sigma}\right)^2} dx_1 \dots dx_q \\ & \quad \times \int_{x_{q+1}^{|D|} \in \mathbb{R}^{|D|-q}} e^{-\frac{1}{2} \sum_{i=q+1}^{|D|} \left(\frac{x_i - m_i}{\sigma}\right)^2} dx_{q+1} \dots dx_{|D|}, \end{aligned}$$

with  $x_1^q = (x_1, \dots, x_q)$  and  $x_{q+1}^{|D|} = (x_{q+1}, \dots, x_{|D|})$ . Remark-  
ing that

$$\frac{1}{\sigma^{|D|} (2\pi)^{\frac{|D|}{2}}} = \frac{1}{\sigma^q (2\pi)^{\frac{q}{2}}} \times \frac{1}{\sigma^{|D|-q} (2\pi)^{\frac{|D|-q}{2}}} \quad (29)$$

and that

$$\begin{aligned} & \int_{x_{q+1}^{|D|} \in \mathbb{R}^{|D|-q}} e^{-\frac{1}{2} \sum_{i=q+1}^{|D|} \left(\frac{x_i - m_i}{\sigma}\right)^2} dx_{q+1} \dots dx_{|D|} \\ &= \sigma^{|D|-q} (2\pi)^{\frac{|D|-q}{2}}, \end{aligned} \quad (30)$$

finally the random vector  $PV_D$  is a Gaussian vector with mean  $(m_1, \dots, m_q)$  and covariance  $\sigma^2 I_q$  on the space  $\text{Im}P$  and the probability  $\mathbb{P}_{H_0}[\delta(V_D) \leq \delta(v_D)]$  is determined by the function

$$f(q, \delta, \sigma, m) = \frac{1}{\sigma^q (2\pi)^{\frac{q}{2}}} \int_{\mathcal{B}^q(\delta)} e^{-\frac{1}{2} \sum_{i=1}^q \left(\frac{x_i - m_i}{\sigma}\right)^2} dx_1 \dots dx_q. \quad (31)$$

Here,  $\dim(\text{Im}A_D) = |\mathcal{L}| = |D| - q$ , i.e.  $q = |D| - |\mathcal{L}|$   $\square$

#### APPENDIX B LEMMA B.1

*Lemma B.1:* Let  $t > 0$  be fixed. The function defined for all vector  $y \in \mathbb{R}^q$  by

$$g(y) = \int_{\|x\| \leq t, x \in \mathbb{R}^q} e^{-\|x-y\|^2} dx$$

is maximal when  $y = 0$ .

*Proof*—Let  $x_i$  and  $y_i$  be the  $i$ th elements of, respectively, the vectors  $x$  and  $y$ .

$$\frac{\partial g}{\partial y_i} = \int_{\|x\| < t, x \in \mathbb{R}^q} -2(y_i - x_i) e^{-\sum_{i=1}^q (y_i - x_i)^2} dx$$

If  $a(x) = \sqrt{t^2 - \sum_{j \neq i} x_j^2}$ , then

$$\begin{aligned} & \int_{x_i^2 < a(x)^2} -2(y_i - x_i) e^{-(y_i - x_i)^2} dx_i \\ &= [e^{-(y_i - x_i)^2}]_{-a(x)}^{a(x)}, \end{aligned}$$

and hence

$$\begin{aligned} \frac{\partial g}{\partial y_i} &= \int_{\sum_{j \neq i} x_j^2 < t^2} e^{-\sum_{j \neq i} (y_i - x_i)^2} \\ & \quad \times [e^{-(y_i - x_i)^2}]_{-a(x)}^{a(x)} dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_q. \end{aligned}$$

The function  $g$  is continuous, non-negative and null at infinity. Hence, by compactness, it reaches its maximum. Moreover, it is  $\mathcal{C}^1$  hence this maximum stands for a critical point. But

$$\begin{aligned} \frac{\partial g}{\partial y_i} &= 0 \\ \Rightarrow \forall x, e^{-(y_i - a(x))^2} - e^{-(y_i + a(x))^2} &= 0 \\ \Rightarrow \forall x, (y_i - a(x))^2 &= (y_i + a(x))^2 \\ \Rightarrow y_i &= 0, \end{aligned}$$

hence the unique critical point of  $g$  is  $y = 0$ . Consequently, the unique global maximum of  $g$  on  $\mathbb{R}^q$  is the unique critical point  $y = 0$ .  $\square$

#### APPENDIX C PROPOSITION 4.1

If  $\gamma = 1$ , for any fixed  $n > 1$  and for all  $p \in (\frac{1}{n}, 1]$ ,  $\mathcal{F}(p, n, 1) \geq 1$ .

*Proof*—Let us substitute 1 for  $\gamma$  in Equation (20) and set  $x = \frac{pn}{2}$  and  $a = \frac{1}{2}$ . We have to prove that for all  $a > 0$  and for all  $x \geq a$ ,

$$\mathcal{F}(p, n, 1) \equiv n \binom{pn}{n} \frac{1}{\Gamma(x-a)} \int_0^x e^{-t} t^{x-a-1} dt \geq 1.$$

Using repeated integrations by parts, it is not difficult to show that

$$\begin{aligned} & \int_0^x e^{-t} t^{x-a-1} dt \\ &= \left[ - \sum_{q=1}^{x-a-1} (x-a-1) \dots (x-a-q+1) t^{x-a-q} e^{-t} \right]_0^x \\ & \quad + (x-a-1)! \int_0^x e^{-t} dt \\ &= (x-a-1)! \left( e^{-x} \sum_{q=x-a}^{+\infty} \frac{x^q}{q!} \right). \end{aligned} \quad (32)$$

Now, noticing that

$$\sum_{q=0}^{x-1} \frac{x^q}{q!} \leq \sum_{q=x}^{2x-1} \frac{x^q}{q!} \leq \sum_{q=x}^{+\infty} \frac{x^q}{q!} \quad (33)$$

(because  $\frac{x^{x-k-1}}{(x-k-1)!} \leq \frac{x^{x+k}}{(x+k)!}$  for all  $k$  such that  $0 \leq k \leq x-1$ ), and using the fact that

$$\sum_{q=0}^{x-1} \frac{x^q}{q!} + \sum_{q=x}^{+\infty} \frac{x^q}{q!} = e^x, \quad (34)$$

it follows

$$e^{-x} \sum_{q=x}^{+\infty} \frac{x^q}{q!} \geq \frac{1}{2}. \quad (35)$$

The integral (32) thus satisfies

$$\int_0^x e^{-t} t^{x-a-1} dt \geq \frac{\Gamma(x-a)}{2} \quad (36)$$

and it follows that

$$\mathcal{F}(p, n, 1) \geq \frac{n}{2} \binom{n}{pn} \quad (37)$$

and that  $\mathcal{F}(p, n, 1) \geq 1$  as soon as  $n > 1$  and  $p > \frac{L}{n}$ .  $\square$

#### APPENDIX D PROPOSITION 4.4

For all fixed  $\gamma$  within the interval  $(0, 1)$ , there exist  $p^*(\gamma)$  such that, for all  $p \in [p^*(\gamma), 1)$ ,

$$\lim_{n \rightarrow +\infty} \mathcal{F}(p, n, \gamma) = 0.$$

*Proof* — Denoting  $a = \frac{pn-L}{2} - 1$  and using Stirling formula, it holds that for all  $n > 1$ ,  $p \in [p^*(\gamma), 1)$  and  $\gamma \in (0, 1]$ ,

$$\mathcal{F}(p, n, \gamma) \underset{a \rightarrow \infty}{\sim} \binom{n}{pn} \int_0^{\gamma(a+1)+\frac{L}{2}} e^{-t} t^a \frac{2(a+1+\frac{L}{2})}{p\sqrt{2\pi a}} \left(\frac{e}{a}\right)^a dt.$$

Using the variable change  $u = \frac{t}{a}$ , it turns out that

$$\mathcal{F}(p, n, \gamma) = \binom{n}{pn} \frac{2a(a+1+\frac{L}{2})}{p\sqrt{2\pi a}} \int_0^{\gamma(1+\frac{1}{a}+\frac{L}{2a})} (e^{1-u}u)^a du.$$

Remarking on the one hand that there exists a rank  $A$  such that for all  $a > A$ ,  $\frac{a}{a+1+\frac{L}{2}} > \gamma$ , and on the other hand that the function  $u \mapsto e^{1-u}u$  is non-decreasing on  $[0, 1)$ , we can write

$$\mathcal{F}(p, n, \gamma) \leq \binom{n}{pn} \frac{2\gamma(a+1+\frac{L}{2})^2 e^{a(1-\gamma+\log \gamma)+(1-\gamma)(1+\frac{L}{2})+o(1)}}{p\sqrt{2\pi a}}$$

as  $\log(1 + \frac{1}{a} + \frac{L}{2a}) = \frac{1}{a} + \frac{L}{2a} + o(\frac{1}{a})$  when  $a$  tends to  $+\infty$ .

- Case  $p = 1$  : From the fact that  $\phi : \gamma \mapsto 1 - \gamma + \log \gamma$  is non-decreasing and strictly negative on  $(0, 1)$ ,

$$\lim_{a \rightarrow \infty} \frac{2\gamma(a+1)^2}{\sqrt{2\pi a}} e^{a(1-\gamma+\log \gamma)+(1-\gamma)(1+\frac{L}{2})+o(1)} = 0.$$

Finally as  $a = \frac{pn-L}{2} - 1$ ,  $\lim_{n \rightarrow \infty} \mathcal{F}(p, n, \gamma) = 0$ .

- Case  $p \neq 1$  : Using the following logarithm properties:

$$\log \binom{n}{pn} = \log(n!) - \log((pn)!) - \log((n-pn)!),$$

and

$$\log(n!) = n(\log n - 1) + o(n),$$

remark that

$$\log \binom{n}{pn} = -pn \log p - n(1-p) \log(1-p) + o(n).$$

Then, returning to initial notations,

$$\mathcal{F}(p, n, \gamma) \leq \frac{\gamma p n^2 e^{n(g(\lambda, p)+o(1)) - \frac{L+2}{2} \log \gamma + o(1)}}{2\sqrt{\pi(pn-L-2)}},$$

where

$$g(\gamma, p) = p \left( \frac{1}{2} (1 - \gamma + \log \gamma) - \log p \right) - (1-p) \log(1-p).$$

Now the domains of  $\gamma$  and  $p$  such as  $g(\gamma, p) \neq 0$  can be studied. An analysis of the variations of  $g$  shows that it is non-decreasing according to parameter  $p$  and non-negative for all  $p \in [0, \frac{1}{1+e^{-\frac{1}{2}(1-\gamma+\log \gamma)}}]$ . Moreover, as for

any fixed  $\gamma \in (0, 1)$ ,  $\lim_{p \rightarrow 1} g(\gamma, p) = \frac{1}{2} (1 - \gamma + \log \gamma) < 0$  and  $g$  is non-increasing on  $[\frac{1}{1+e^{-\frac{1}{2}(1-\gamma+\log \gamma)}}, 1]$ , the function  $g$  necessarily vanishes on the interval  $(0, 1)$ . Hence, for any fixed  $\gamma$  in  $(0, 1)$ , there is a value  $p^*(\gamma)$  such as for all  $p \in [p^*(\gamma), 1]$ ,  $g(\gamma, p) < 0$  which leads to the conclusion considering that

$$\lim_{n \rightarrow +\infty} \frac{\gamma p n^2 e^{n(g(\lambda, p)+o(1)) - \frac{L+2}{2} \log \gamma + o(1)}}{2\sqrt{\pi(pn-L-2)}} = 0.$$

$\square$

#### ACKNOWLEDGMENT

The authors would like to thank the Centre National d'Études Spatiales (CNES) and EADS/ASTRIUM for funding a part of this work. In particular, they are grateful to J. Inglada for fruitful discussions and to the CNES for providing the ADAM database.

#### REFERENCES

- [1] S. Bontemps, P. Bogaert, N. Titeux, and P. Defourny, "An object-based change detection method accounting for temporal dependences in time series with medium to coarse spatial resolution," *Rem. Sens. of Env.*, vol. 112, pp. 3181–3191, 2008.
- [2] K. Conradsen, A. Nielsen, J. Schou, and H. Skriver, "A test statistic in the complex wishart distribution and its application to change detection in polarimetric sar data," *IEEE Trans. on Geosc. and Rem. Sens.*, vol. 41, no. 1, pp. 4–19, 2003.
- [3] L. Bruzzone and D. Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Trans. on Image Processing*, vol. 11, no. 4, pp. 452–466, 2002.
- [4] M. Bosc, F. Heitz, J. Armspach, I. Namer, D. Gounot, and L. Rumbach, "Automatic change detection in multimodal serial mri: application to multiple sclerosis lesion evolution," *Neuroimage*, vol. 20, pp. 643–656, 2003.
- [5] D. Rey, G. Subsol, H. Delingette, and N. Ayache, "Automatic detection and segmentation of evolving processes in 3d medical images: Application to multiple sclerosis," *Med. Image Anal.*, vol. 6, no. 2, pp. 163–179, Jun. 2002.
- [6] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," *IEEE Trans. on Pattern Analysis and Mach. Intell.*, vol. 30, no. 3, pp. 555–560, 2008.
- [7] R. Collins, A. Lipton, and T. Kanade, "Introduction to the special section on video surveillance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 745–746, Aug. 2000.
- [8] P. Agouris, S. Gyftakis, and A. Stefanidis, "Uncertainty in image-based change detection," in *Proc. of Int. Symposium on Spatial Accuracy*, Amsterdam (Netherlands), 2000, p. 18.
- [9] S. Le Hégarat-Masclé and R. Seltz, "Automatic change detection by evidential fusion of change indices," *Rem. Sens. Env.*, vol. 91, pp. 390–404, 2004.
- [10] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Trans. on Geosc. and Rem. Sens.*, vol. 45, no. 1, pp. 218–236, 2007.



- [11] L. Bruzzone and D. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. on Geosc. and Rem. Sens.*, vol. 38, no. 3, pp. 1171–1182, 2000.
- [12] W. Malila, "Change vector analysis: an approach for detecting forest changes with landsat," in *Proc. of the Annual Symposium on Machine Processing of Remotely Sensed Data*, 1980, pp. 326–335.
- [13] C. Clifton, "Change detection in overhead imagery using neural networks," *Applied Intelligence*, vol. 18, pp. 215–234, 2003.
- [14] A. Elfshawy, S. Kesler, and A. Abutaleb, "Adaptive algorithms for change detection in image sequence," *Signal Processing*, vol. 23, no. 2, pp. 179–191, 1991.
- [15] T. Kasetkasem and P. Varshney, "An image change detection algorithm based on markov random field models," *IEEE Trans. on Geosc. and Rem. Sens.*, vol. 40, no. 8, pp. 1815–1823, 2002.
- [16] L. Bruzzone and D. Prieto, "An adaptive parcel-based technique for unsupervised change detection," *Int. J. of Rem. Sens.*, vol. 21, no. 4, pp. 817–822, 2000.
- [17] S. Ghosh, L. Bruzzone, S. Patra, F. Bovolo, and A. Ghosh, "A context-sensitive technique for unsupervised change detection based on hopfield-type neural networks," *IEEE Trans. on Geosc. and Rem. Sens.*, vol. 45, no. 3, pp. 778–788, 2007.
- [18] R. Radke, S. Andra, O. Al-Kohafi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. on Image Processing*, vol. 14, no. 3, pp. 294–307, 2005.
- [19] F. Melgani, G. Moser, and S. Serpico, "Unsupervised change detection methods for remote sensing images," *Optical Engineering*, vol. 41, no. 12, pp. 81–90, 2002.
- [20] A. Huertas and G. Medioni, "Detection of intensity changes with subpixel accuracy using laplacian-gaussian masks," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 8, no. 5, pp. 651–664, 1986.
- [21] Y. Limin, X. G., K. J., and B. Deal, "Urban land-cover change detection through sub-pixel imperviousness mapping using remotely sensed data," *Photogrammetric engineering and remote sensing*, vol. 69, no. 9, pp. 1003–1010, 2003.
- [22] D. Manolakis, C. Siracusa, and S. G., "Hyperspectral subpixel target detection using the linear mixing model: Analysis of hyperspectral image data," *IEEE Trans. on Geoscience and Rem. Sens.*, vol. 39, no. 7, pp. 1392–1409, 2001.
- [23] S. Le Hégarat-Masclé, C. Ottlé, and C. Guérin, "Land cover change detection at coarse spatial scales based on iterative estimation and previous state information," *Rem. Sens. of Env.*, vol. 95, pp. 464–479, 2005.
- [24] A. Desolneux, L. Moisan, and J. Morel, "Meaningful alignments," *Int. J. Comp. Vision*, vol. 40, no. 1, pp. 7–23, 2000.
- [25] —, "A grouping principle and four applications," *IEEE Trans. on Pattern and Machine Intelligence*, vol. 25, no. 4, pp. 508–513, 2003.
- [26] L. Moisan and B. Stival, "A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix," *Int. J. Comp. Vision*, vol. 57, no. 3, pp. 201–218, 2004.
- [27] F. Cao, T. Veit, and P. Bouthemy, "Image comparison and motion detection by a contrario methods," in *Computational Vision in Neural and Machine Systems*, L. Harris and M. Jenkin, Eds. Cambridge University Press, 2005.
- [28] B. Grosjean and L. Moisan, "A-contrario detectability of spots in textured backgrounds," *J. of Mathematical Imaging and Vision*, vol. 33, no. 3, pp. 313–337, 2009.
- [29] D. Lowe, *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, 1985.
- [30] H. Horwitz, R. Nalepka, P. Hyde, and J. Morgenstern, "Estimating the proportions of objects within a single resolution element of a multispectral scanner," in *Proc. of the 7th Int. Symposium on Rem. Sens. of Env.*, Ann Arbor, Michigan, 1971, pp. 1307–1320.
- [31] R. Faivre and A. Fischer, "Predicting crop reflectances using satellite data observing mixed pixels," *J. of Agricultural, Biological and Env. Statistics*, vol. 2, pp. 87–107, 1997.
- [32] H. Cardot, R. Faivre, and M. Goulard, "Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data," *J. of Applied Statistics*, vol. 30, pp. 1185–1199, 2003.
- [33] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, pp. 381–385, 1981.
- [34] Y. Hochberg and A. Tamhane, *Multiple comparison procedures*. New-York: John Wiley & Sons, 1987.
- [35] G. Saporta, *Probabilités, Analyse des Données et Statistique*. Editions TECHNIP, 1990.
- [36] C. Bonferroni, "Teoria statistica delle classi et calcolo delle probabilita." *Pubblicazioni del Instituto Superiore de Scienze Economiche e Commerciali di Firenze*, vol. 8, pp. 3–62, 1936.
- [37] Y. Hochberg, "A sharper bonferroni procedure for multiple tests of significance," *Biometrika*, vol. 75, pp. 800–803, 1988.
- [38] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. of the Royal Statistical Society*, vol. 57, no. 1, pp. 289–300, 1995.
- [39] Z. Zhang, R. Deriche, J. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *AI J.*, no. 78, pp. 87–119, 1994.
- [40] A. Robin, S. Le Hégarat-Masclé, and L. Moisan, "Unsupervised sub-pixelic classification using coarse resolution time series and structural information," *IEEE Trans. on Geosc. and Rem. Sens.*, vol. 46, no. 5, pp. 1359–1374, 2008.
- [41] G. Moser and S. Serpico, "Generalized minimum-error thresholding for unsupervised change detection from sar amplitude imagery," *IEEE Trans. on Geosc. and Remote Sens.*, vol. 44, no. 10, pp. 2972–2982, 2006.
- [42] L. Bruzzone and D. Prieto, "A minimum cost thresholding technique for unsupervised change detection," *Int. J. of Rem. Sens.*, vol. 21, no. 18, pp. 3539–3544, 2000.
- [43] T. Fung and E. Le Drew, "The determination of optimal threshold levels for change detection using various accuracy indices," *Photogrammetric Eng. & Rem. Sens.*, vol. 54, no. 10, pp. 1449–1454, 1988.



**Amandine Robin** received the PhD degree in Applied Mathematics from Paris Descartes University (France) in 2007. She is currently a lecturer at the School of Computational and Applied Mathematics, University of the Witwatersrand (South Africa). Her research interests concerns image processing, a-contrario statistical modeling for image analysis, data fusion and remote sensing. Her work is mostly dedicated to change detection and classification using remote sensed images or video sequences.



**Lionel Moisan** studied mathematics and computer science at the Ecole Normale Supérieure in Paris. He received the PhD degree in 1997 from Paris-Dauphine University and the HDR degree in 2003 from Paris XI University. He was a CNRS researcher at Ecole Normale Supérieure de Cachan from 1998 to 2003, and is currently Professor of Applied Mathematics at Paris Descartes University. His research activities are mainly focused on variational models for image processing (restoration, smoothing) and on a-contrario statistical models for image analysis

(gestalt and structure detection).



**Sylvie Le Hégarat-Masclé** received the PhD degree in Signal and Image from ENST Paris (France) in 1996 and the "Habilitation à Diriger des Recherches" at Versailles University in 2006. She was associated professor at Versailles University from 1998 to 2005. Since 2006, she is professor at Paris XI University. Her interests include statistical pattern recognition, image analysis, and data fusion. She works with the Bayes theory and the Dempster-Shafer evidence theory. In remote sensing domain (SAR and optical sensor), concerned applications

are: land cover / land-use classification, vegetation and soil moisture monitoring.