AGGREGATED PRIMARY DETECTORS
FOR GENERIC CHANGE DETECTION IN SATELLITE IMAGES

Vincent Vidal[1,2], Matthieu Limbert[2], Tugdual Ceillier[2], Lionel Moisan[1]

[1]MAP5, Université Paris Descartes, Paris, France
[2]Earthcube, Paris, France

ABSTRACT

Detecting changes between two satellite images of the same scene generally requires an accurate (and thus often uneasy to obtain) model discriminating relevant changes from irrelevant ones. We here present a generic method, based on the definition of four different a-contrario detection models (associated to arbitrary features), whose aggregation is then trained from specific examples with gradient boosting. The results we present are encouraging, and in particular the low false positive rate is noticeable.

Index Terms— change detection, a-contrario framework, very high resolution, optical, supervised learning, image comparison

1. INTRODUCTION

The last decade has seen a huge increase of the number of high resolution satellite images. This leads to an ever-growing use for many surveillance, detection and management tasks, such as deforestation control, urban growth monitoring or natural disasters damage assessment, where it is necessary to detect changes between two images taken at different times. Given two satellite images of the same geographic location, the change detection (CD) problem consists in detecting various generic changes while being robust to some others considered as irrelevant, such as noise, shadows, illumination or ground texture variation. One difficulty is that the notion of relevant change may depend on the use case; moreover, additional issues arise with misregistration errors and parallax, especially in very high resolution (VHR) satellites images [1]. In this paper, we are looking for a generic approach that can be easily adapted to multiple applications. We propose an hybrid solution for generic change detection, using a set of primary change detectors, whose supervised aggregation model can be trained on a small training set.

2. GENERIC CHANGE DETECTION PROBLEM

The difficulty of this task lies in the diversity of changes, illustrated in Fig. 1. On this example, the contrast variation and the shift of shadows would cause most basic models (in
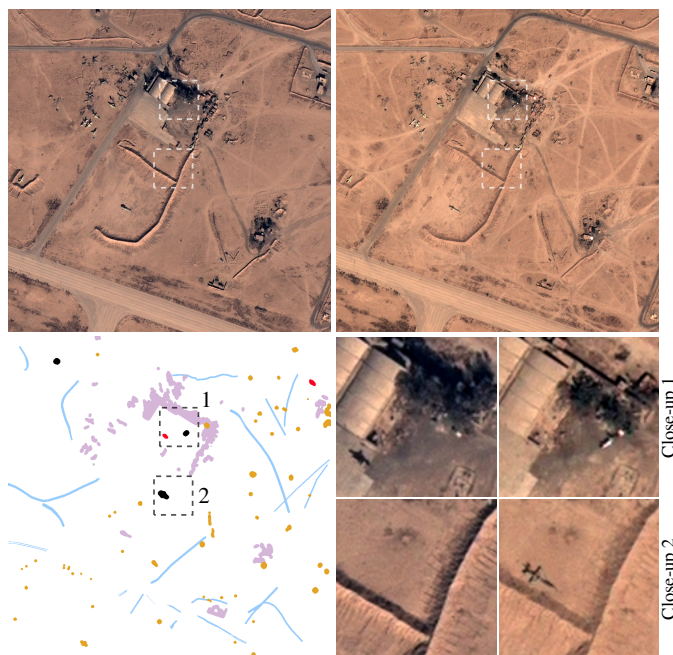

Fig. 1. **A diversity of changes. Top:** A pair of optical satellite images (Pleiades) of the same location, taken at different dates. **Bottom-left:** A colored representation of different types of changes: shift of shadows (in light purple), linear ground tracks modifications (in blue), appearances of small objects (in black) and white or black spots (in orange), and two of unknown type (in red). Notice other changes concerning global illumination, local contrast, slight texture variations and noise (not represented). **Bottom-right:** Close-ups of the two areas squared with dashes. We can see that changes are everywhere but only some of them would be considered as meaningful.

particular methods based on direct comparison) to produce too many false positives. A lot of other methods exist for CD, based for example on "Change Vector Analysis", Markov random fields or principal component analysis (see [1] and references therein) but most of them focus on a particular task from the very start, not really addressing the generic change detection problem. Some machine learning methods, like de-

cision trees or neural networks [2], have also been proposed for CD in a large number of tasks. However, these methods need a large training data set, which is generally impossible to gather, especially when it comes to "rare" changes, such as natural disasters. We here propose a generic method that postpone the need of a definition for relevant changes to the fusion process.

In Section 3, we introduce the a-contrario statistical framework, develop the four primary models ("Direct comparison", "Histogram", "Context" and "Common fate"), and eventually specify the fusion process. In Section 4, we present and discuss the results of these methods on several pairs of images, given a particular detection problem.

3. PRIMARY DETECTORS AND FUSION

From now on, we consider two images on a domain $D \subset \mathbb{Z}^2$ with c channels, seen as vectors in $\mathbb{R}^{|D|c}$. The observed images will be noted u et v and we will use V to denote a random model of v.

Each of the following models will take as input some precomputed features, suitable for a particular use case. In practice, we will use 9 different features: the gray scale value, the gradient, its norm, nine wavelet coefficients, steerable Gaussian filters and four features based on the second moment matrix (the full matrix, its eigen values, its cornerness and its coherence) [4]. Each choice of one of the 4 models and one of the 9 features will be seen as a "primary detector". Excluding the "Common fate and wavelet coefficients" detector for computational reasons, it will give us a total of 35 primary detectors.

3.1. A-contrario framework

In order to define change models, we use the a-contrario framework, introduced in [3]. Given $\mathcal{H}_0$, a model of V corresponding to its expected distribution without change, and a set $\mathcal{I}$ of tests (e.g. the set of evaluated pixels), it defines a Number of False Alarms (NFA) as a function $\mathrm{NFA}(p, v)$ satisfying

$$\forall \varepsilon > 0, \quad \mathbb{E}\Big[\mathrm{Card} \left\{ p \in \mathcal{I}, \mathrm{NFA}(p, V) \leq \varepsilon \right\} \Big] \leq \varepsilon. \quad (1)$$

This property guarantees us that, if v were drawn according to $\mathcal{H}_0$, the number of (wrongly) detected structures (that is, tests p such that $\mathrm{NFA}(p, v) < \varepsilon$) would be on average negligible (less than ε). This framework gives an easy way to build such NFA functions: for every function $\mu : \mathcal{I} \times \mathbb{R}^{|D|c} \to \mathbb{R}$, called "structure measure", the NFA property (1) is satisfied by

$$\mathrm{NFA}\,(p, v) = |\mathcal{I}| \cdot \mathbb{P}_{V \sim \mathcal{H}_0}\Big(\mu(p, V) \geq \mu(p, v) \Big). \quad (2)$$

3.2. A-contrario change detection from linear features

The first model is based on the idea that a change would correspond to a significant variation in the features between the two images. Following the idea of the most commonly used model, the "Direct Comparison Change Detection" [1], we here compare the two images pixelwise.

We consider K extracted features, linear and invariant by translation, given by convolutional kernels $(W^k)_{1 \leq k \leq K}$, and a function f concatenating all those features such as

$$\forall k \in \{1, \ldots, K\}, \ \forall p \in D, \quad f(u)_{k,p} = \sum_{q \in D, \ p-q \in D} u_{p-q} W_q^k.$$

We define $\mathcal{H}_0$ as the Gaussian model $V = u + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_{|D|c})$. Following the idea behind the Mahalanobis distance, we decorrelate the features by considering

$$\Delta = \Sigma_0^{-1/2}\big(f(u) - f(V)\big),$$

with $\Sigma_0^{-1/2}$ the inverse square root of the correlation matrix $(\Sigma_0)_{k,l} = \sum_{p \in D} W_p^k W_p^l$. Then, since $\Delta \sim \mathcal{N}(0, \sigma^2 I_K)$, by taking the structure measure $\mu(p, u, v) = \|\Delta\|_2$, Equation (2) gives

$$\mathrm{NFA}_1(p, v) = \frac{|D|}{\Gamma\left(\frac{K}{2}\right)} \Gamma\left(\frac{K}{2}, \frac{1}{2\sigma^2} \left\| \Sigma_0^{-1/2}\big(f(u) - f(v)\big) \right\|_2^2 \right),$$

where $\Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt$ is the incomplete gamma function [5].

Because this model is impacted by any significant variation, it will roughly detect all changes noticeable by the given features. This may become a serious issue in case of spatial inconsistencies, such as misregistration, parallax or small object translations, which are more than usual in VHR images.

3.3. "Histogram" model

In order to deal with spatial inconsistencies, we propose a second model based on a comparison of local distributions. We use the Kolmogorov–Smirnov test, which satisfies the following property: Given $(X_i)_{i \in \mathbb{N}}$ a set of i.i.d. random variables, their cumulative distribution function F, and the empirical cumulative distribution function

$$F_n : x \mapsto \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{x \geq X_i\},$$

for every $c > 0$, we have

$$\mathbb{P}\left(\|F_n - F\|_\infty > \frac{c}{\sqrt{n}} \right) \xrightarrow[n \to \infty]{} 2 \sum_{k=1}^\infty (-1)^{k-1} e^{-2c^2 k^2}.$$

Thus, by considering $\mu(p, u, v) = \|F_u(p) - F_v(p)\|_\infty$, with $F_u(p)$ the empirical cumulative distribution function of u in a neighborhood of p, we use, as an approximation,

$$\mathrm{NFA}_2(p, v) = 2 |D| \sum_{k=1}^\infty (-1)^{k-1} e^{-2n_p \|F_u - F_v\|_\infty^2 k^2}.$$

This model discards spatial information, and is thus robust to parallax and misregistration. This increases the scale of analysis, which is an issue for the precise localization of changes. Note also that, because this model is based on cumulative distribution functions, it can only be applied to a 1-dimensional feature. In case of multidimensional features, it will be used on the first component only.

3.4. Local context model based on Gaussian Mixtures

We now propose another model based on local distributions. To keep an accurate result, we only discard the spatial information in one of the two images, and compare the feature values of each pixel of one image with their distributions estimated from the other image. This idea is used on single images for anomaly detection in the RX algorithm [6]. Using an Expectation-Maximization algorithm, we learn the local distribution around a pixel p as a mixture of R Gaussian densities g_r of weight α_r, with probability distribution function $\sum_{r=1}^{R} \alpha_r g_r(x)$. To build the structure measure $\mu(x)$, we use the approximation of the log-likelihood

$$-\log \sum_{r=1}^{R} \alpha_r g_r(x) \simeq \min_{1 \leq r \leq R} m_r + d_{\Sigma_r, \mu_r}(x)^2 \stackrel{\text{def}}{=} \mu(x),$$

with $m_r = \log |\Sigma_r| - 2\log \alpha_r$ and $d_{\Sigma_r, \mu_r}(x)$ the Mahalanobis distance. One can prove that

$$\mathbb{P}\big(\mu(V_p) \geq \delta\big) \leq \frac{1}{\Gamma\left(\frac{K}{2}\right)} \Gamma\left(\frac{K}{2}, \frac{\delta - \max_r m_r}{2}\right),$$

and thus, Equation (2) yields

$$\text{NFA}_3(p, v) = \frac{|D|}{\Gamma\left(\frac{K}{2}\right)} \Gamma\left(\frac{K}{2}, \frac{\mu(v_p) - \max_r m_r}{2}\right).$$

While still being robust to misregistration, this model provides more spatial accuracy. However, like the previous ones, it highly suffers from global changes, such as contrast shifts or illumination variations. If some of these changes can be reduced (for example with histogram registration), others like shadow appearances may substantially restrict the use of this model.

3.5. "Common fate" change detection model

This model was build to take global changes into account. It is inspired by the Gestalt theory of visual perception [7]. The idea is that the transition of a pixel feature between the two images should not be considered as a change if a significant number of other pixels share the same "common fate" (that is, have this same feature transition). Because modeling the distribution of those transitions in $\mathbb{R}^{2c}$ might be too complex, we quantize feature values by considering variable subsets F of $\mathbb{R}^c$ and the associated sub-domain

$$D_u^F = \{p \in D, u_p \in F\} \quad \text{and} \quad D_v^F = \{p \in D, v_p \in F\}.$$

Given a spatial domain $E \subset D$ and two subsets F and G of $\mathbb{R}^c$, the structure measure is chosen to be the number of pixels in E sharing the same "fate" from F to G, that is,

$$\mu(E, F, G; v) = \text{Card}\big(D_u^F \cap D_v^G \cap E\big).$$

Assuming pixels are independent, Equation (2) gives

$$\text{NFA}_4(\phi, v) = N_\phi . \mathcal{B}\left(\left|E \cap D_u^F\right|, \left|E \cap D_u^F \cap D_v^G\right|, \frac{\left|D_u^F \cap D_v^G\right|}{\left|D_u^F\right|}\right),$$

with $\mathcal{B}(N, n, p) = \sum_{k=n}^{N} \binom{N}{k} p^k (1-p)^{(N-k)}$ the tail of the binomial distribution, $\phi = (E, F, G)$ the evaluated test and N_ϕ the total number of ϕ triplets. Unlike the previous ones, this model is robust to contrast shifts and most global changes, provided they each occur on enough pixels. However, it is sensitive to spatial inconsistencies.

3.6. Fusion

The final detection algorithm is obtained by fusing the primary detectors detailed above. We tested several methods such as dense neural network, SVM and gradient boosting [8], which is an iterative optimization technique using decision trees. We chose the gradient boosting as it gave the best results and seemed to be best suited to this problem: it can handle input data with very different behaviors, is more resilient to overfitting and requires less data than neural networks, since it has less parameters to optimize.

4. RESULTS

We considered the set of 35 primary detectors described in Section 3, trained the fusion model on a dataset of 18 image pairs (800x800 pixels) extracted from images of different scenes, taken from different satellites, and tested it on 4 new pairs, including the pair in Fig. 1. The ground truth binary images were generated by hand-tagging interesting changes at pixel scale. We tagged as "change" small object appearances or disappearances, such as vehicles, trees or small buildings. Shadows and global changes due to vegetation or illumination variations were tagged as "non change".

The results obtained on the pair of images of Fig. 1 are shown in Fig. 2. Given the difficulty of the generic change detection problem, the results are encouraging: the algorithm detects every object appearances (in black and red) and as concerns false detections, three of them are due to shadow shifts, none is caused by variations of contrast or ground track positions, and the others are caused by dark spots whose irrelevance as changes may be questionable.

To assess the contribution of all primary detectors to the whole algorithm, we computed their weights in the fusion model and displayed the associated raw detection masks in Fig. 3. In Fig. 4 are shown the precision-recall curves (with detections counted object-wise) achieved on the 4 test image
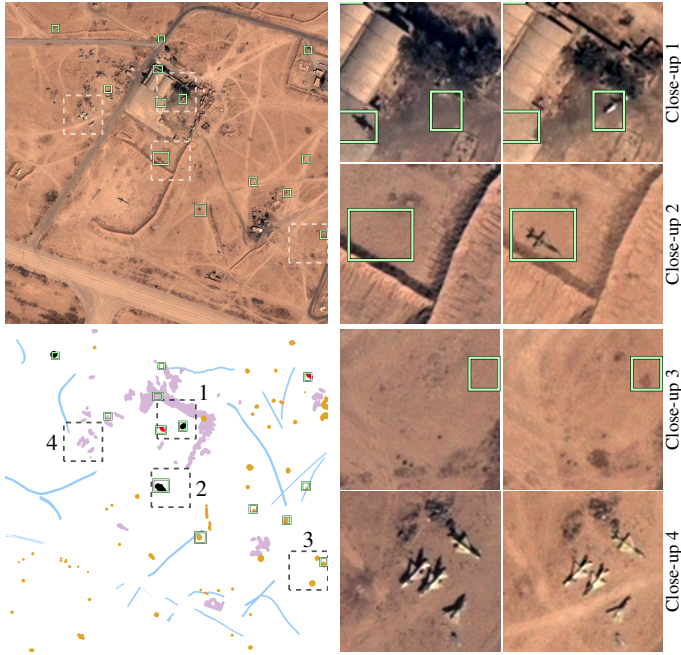
Fig. 2. Results obtained on the image pair of Fig. 1. Detections are represented by green boxes in all images. **Top-left:** Second image of the pair. **Bottom-left:** Color representation of the different types of changes presented in Fig. 1. **Right:** Close-ups of the four areas squared with dashes. We observe that the algorithm detects the changes related to vehicle appearances and disappearances (Close-Ups 1 and 2) and some small black spots (Close-Up 3), while being robust to most of the perturbations such as intense shadow variations and appearances of linear ground tracks (Close-Ups 1, 2 and 4).

pairs for several versions of the algorithm: the whole algorithm (obtained with the 35 primary detectors issued by the 4 models) and four incomplete algorithms obtained by selecting primary detectors from all but one model. As we can see, the removal of any of the 4 primary detector models induces a substantial loss in terms of precision-recall, which again illustrates the complementarity of the 4 models.

5. REFERENCES

[1] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS JPRS*, vol. 80, pp. 91–106, 2013.

[2] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional siamese networks for change detection," in *2018 25th IEEE ICIP*. 2018, pp. 4063–4067.

[3] A. Desolneux, L. Moisan, and J.-M. Morel, *From gestalt theory to image analysis: a probabilistic approach*, vol. 34, Springer, 2007.
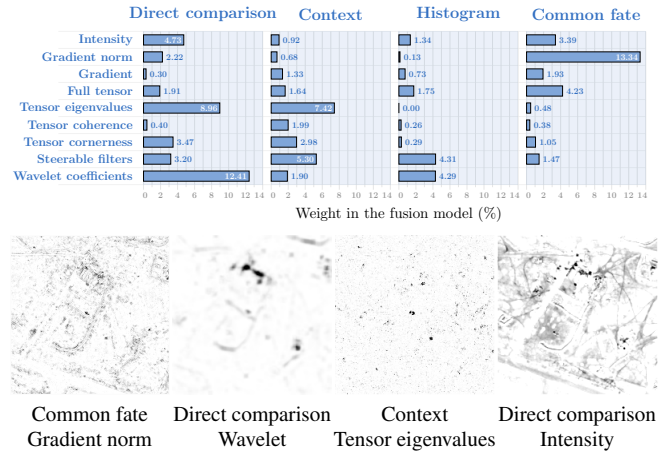
Common fate   Direct comparison   Context   Direct comparison
Gradient norm   Wavelet   Tensor eigenvalues   Intensity

Fig. 3. **First row:** Weight of each of the 35 primary detectors in the fusion model. The maximum weight is below 14% and 12 detectors contribute more than 3%, which highlights their complementarity and the relevance of their fusion. **Second row:** Raw gray level masks of four of the most contributing primary detectors in the fusion model, illustrating their spatial complementarity.
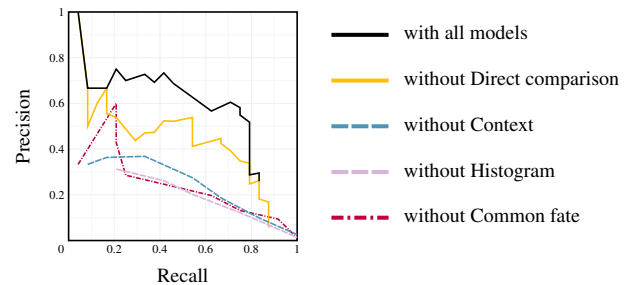

Fig. 4. Precision-recall curves obtained on 4 pairs of images with the proposed algorithm (all models) and with incomplete versions (all models but one).

[4] U. Köthe, "Edge and junction detection with an improved structure tensor," in *JPRS*. Springer, 2003, pp. 25–32.

[5] R. Abergel and L. Moisan, "Fast and accurate evaluation of a generalized incomplete gamma function," preprint MAP5 2016, to appear in *TOMS*.

[6] J. Zhou, C. Kwan, B. Ayhan, and M. T Eismann, "A novel cluster kernel RX algorithm for anomaly and change detection using hyperspectral images," *IEEE TGRS*, vol. 54, no. 11, pp. 6497–6504, 2016.

[7] G. Kanizsa, *Grammatica del vedere: saggi su percezione e gestalt*, Il mulino, 1980.

[8] J. H Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.