



**HAL**  
open science

# A contrario comparison of local descriptors for change detection in Very High spatial Resolution (VHR) satellite images of urban areas

Florence Tupin, Gang Liu, Yann Gousseau

## ► To cite this version:

Florence Tupin, Gang Liu, Yann Gousseau. A contrario comparison of local descriptors for change detection in Very High spatial Resolution (VHR) satellite images of urban areas. *IEEE Transactions on Geoscience and Remote Sensing*, In press, 10.1109/TGRS.2018.2888985 . hal-01739459v2

**HAL Id: hal-01739459**

**<https://hal.science/hal-01739459v2>**

Submitted on 19 Dec 2018

**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.

# A *contrario* comparison of local descriptors for change detection in Very High spatial Resolution (VHR) satellite images of urban areas

Gang Liu, Yann Gousseau, Florence Tupin, *Senior Member, IEEE*,

**Abstract**—Change detection is a key problem for many remote sensing applications. In this paper, we present a novel unsupervised method for change detection between two high resolution remote sensing images possibly acquired by two different sensors. This method is based on keypoints matching, evaluation and grouping, and does not require any image co-registration. It consists of two main steps. First, global and local mapping functions are estimated through keypoints extraction and matching. Secondly, based on these mappings, keypoint matchings are used to detect changes and then grouped to extract regions of changes. Both steps are defined through an *a contrario* framework, simplifying the parameter setting and providing a robust pipeline. The proposed approach is evaluated on synthetic and real data from different optic sensors with different resolutions, incidence angles and illumination conditions.

**Index Terms**—Remote Sensing, Change Detection, a *contrario*

## I. INTRODUCTION

Change detection, aiming at detecting changes between remote sensing images, has been widely studied in the literature [1]–[20]. Many applications have been investigated based on change detection, such as disaster evaluation [4], [21]–[24] or land use monitoring [9], [25]–[28]. The unsupervised detection of changes, which requires less human intervention, has attracted more and more attention [3], [7], [18], [29]–[38].

In the past few years, two main trends have led to drastic evolutions in the change detection approaches. First, numerous sensors have been launched both by national and international space agencies and by private actors. This wealth of sensors (either optic or SAR) and acquisition conditions (sensor agility, acquisition mode diversity) makes it necessary to develop robust approaches overcoming changes of resolutions, of lighting conditions and incidence angles. The challenges are specially difficult when dealing with urban areas where object elevation impairs the fine registration, particularly in the absence of a Digital Surface Model (DSM). Secondly, the spatial resolution improvement leads to scenes with higher complexity [39], the abundance of small and potentially moving objects giving rise to many irrelevant changes.

Traditional change detection approaches usually rely on a first step of co-registration [40]–[43], followed by a pixel-based comparison. For this second step, many methods have

been proposed to efficiently discriminate between changed and unchanged pixels, relying on SVM [30], MRFs [19], [44], a *contrario* methods [28], [45], [46], morphological attribute profiles [23], graph model [47], physically-based quantities [20] or neural network (NN) models [48].

One difficulty with pixel-based methods is that image data captured from different sensors, at different time or in different weather conditions follow different distributions, which gives rise to an inaccurate comparison between pixels. To solve these problems, Pacifici *et.al.* [20] demonstrate that the physically-based quantities and angular information are robust to the changes of illumination, viewing geometries, and atmospheric conditions, and that they are both useful to analyze high resolution remote sensing images, especially for change detection. Furthermore, Solano-Correa, *et.al.* [38] indicate that high level physical features, such as radiometric index, are robust enough to process images from multi-sensors. In our approach, thanks to the use of invariant descriptors, there is no need to compute these physical features.

Recently, and following the general trend in computer vision, significant improvements have been obtained for change detection tasks by using convolutional neural networks, either for optical or SAR images [49]. Different architectures have been explored, either to extract discriminative features [50], or to directly perform end-to-end learning of the change map [51]. Fully convolutional architecture have also been explored [52], in order to get accurate pixel-level predictions. In [53] recurrent networks are used to detect changes and encode temporal dependency in multitemporal images. Similarly to pixel-based methods, these approaches necessitate an accurate registration step before both the learning and the detection steps. As of today, such approaches are not able to deal with parallax effects encountered in high resolution urban area.

In order to overcome the need to finely register data, common to all aforementioned methods, more sophisticated methods using 3D information have been developed. They usually rely on an explicit modeling of buildings [54]–[57] and provide robust results. However, they necessitate the knowledge of digital surface models (DSM) and/or the use of sophisticated acquisition procedures such as LIDAR, which may be impracticable in cases such as disaster evaluations. An alternative is to build DSM with stereo images and detect changes by the joint use of height changes and image comparisons, as in [58], where the use of a Kullback Leibler divergence similarity measure is proposed. However, such approaches are of course very sensitive to matching errors

Gang Liu is with Idiap Research Institute, Switzerland and this work was finished when he was pursuing Ph.D in Telecom ParisTech. Yann Gousseau, Florence Tupin are with LTCI, Telecom ParisTech, Université Paris Saclay, 75013, Paris, France. E-mail: gangliu.spl@gmail.com, {yann.gousseau, florence.tupin}@telecom-paristech.fr

between stereo views.

Without such DSM information and to overcome the drawbacks of pixel-based approaches, feature-based approaches or object-based approaches have been developed (see [59] for a review). Local descriptors such as SIFT [60] have the great advantage of incorporating radiometric and geometric invariances. Therefore they are well adapted to deal with challenging situations such as multi-sensors change detection where strong geometric and radiometric distortions may occur. Although widely used in computer vision, these descriptors have been mostly used for the pre-registration step of remote sensing images [61], and hardly for the change detection step itself, with the exception of [10].

In this paper, which is an extended version of the conference paper [62], we go further into some ideas of [10] to propose a complete processing pipeline for change detection between optic images acquired by heterogeneous sensors. Our method does not require a pre-registration step, except for some simple spatial sampling to balance the resolution, and is able to cope with radiometric and geometric distortions. The proposed approach relies on local descriptors and is robust to parallax effects thanks to the modeling of local deformations. The difficult setting of parameters discriminating change and "no change" situations relies on the *a contrario* framework [63] giving an intuitive understanding of their behaviour.

The method is divided into two main steps. First, using SIFT-like descriptors extraction and matching, we compute a global mapping and several local mappings between the two images. Indeed, the *a contrario* framework permits the reliable search of multiple local mappings. This step allows to take into account local geometric deformations due to ground or building elevation. The output of this first step is a new set of optimally matched local descriptors. The second step of the method consists in detecting changes between descriptors. The *a contrario* model yields thresholds to detect changes, defined through an expected number of false detections. These changes are then grouped to detect changed areas by evaluating the number of changed keypoint under a no change hypothesis, again in an *a contrario* framework. The proposed method is then evaluated on synthetic and real images through quantitative and qualitative experiments.

Recently two works relying on keypoints for change detection have been proposed [47], [64]. In [64], SIFT descriptors are used to select candidate regions, on which traditional pixel-based methods are applied. In [47], local maximal points are extracted to construct graphs in pair of images and change maps are detected by comparing the graphs using local patches around each keypoint. Our work, in contrast, directly compares the keypoints for matching and change detection between two un-registered images.

The paper is organized as follows (see the flowchart Fig. 1). The first step computing the global and local mappings through keypoints and descriptors matching is presented in Section II. The second step of change detection and grouping in an *a contrario* framework is described in Section III. Then Section IV presents and discusses experiments of the proposed approach on synthetic and real data.

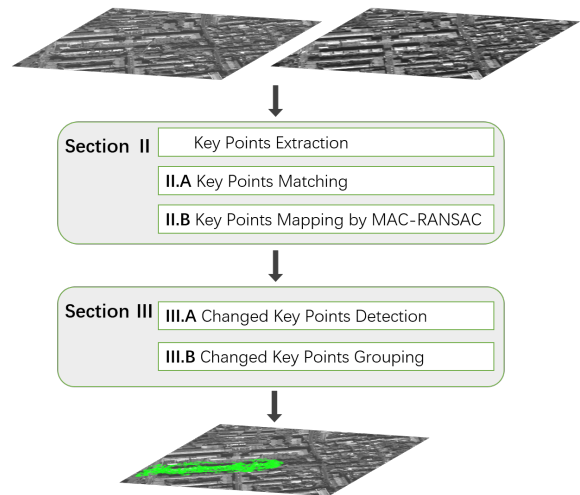


Fig. 1: The flowchart of the proposed approach. The keypoint matching and mapping computations are presented in section II.A and II.B. The keypoints evaluation for change detection is described in section III.A and the grouping in section III.B.

## II. FINE SPARSE REGISTRATION

The first step of the algorithm is, starting from two unregistered images, to establish accurate and reliable correspondences between spatial positions. First, keypoints matching permits the estimation of one global transform and then the same matching permit the estimation of locally refined transforms. Then, for each keypoint in the first image, its best corresponding spatial position is found in the other image, using the various refined transforms to generate candidate positions. The same is performed between the second and first image. In short, we perform an accurate registration of the positions given by local keypoints.

### A. Keypoints matching

We choose to rely on a SIFT-like procedure to extract and characterize local keypoints since SIFT keypoints and descriptors are robust to scale and illumination changes, and to a certain extent to spatial affine transforms. Precisely, we use a robust variant of the SIFT [60] method, as described in [65]. Note that this choice, using SIFT for encoding the local geometry, is not unique and that other descriptors can be used, in particular in order to take advantage of different invariance properties.

The first step of the procedure is to identify interesting spatial positions. This is classically performed by detecting scale-space extreme of the Laplacian operator. Similarly as in the original SIFT, we use 13 scales with a scale ratio of 1.2. Then, the detected maxima are filtered by applying a multi-scale Harris test. For a point  $x$  detected in an image  $I_\sigma$  at scale  $\sigma$ , this test is classically defined [66] as:

$$\text{Det}(M_\sigma) - k\text{Tr}(C_\sigma)^2 > t,$$

where  $C_\sigma$  is a smoothed structure tensor computed on the image  $I_\sigma$ , and  $k, t$  are two parameters. The outcome of this step is a set of keypoints, each of which is associated to a detection scale.

The next step is to compute SIFT-like descriptors at each keypoint. Again, we follow a robust statistical procedure as explained in [65]. Dominant directions are first associated to each keypoint, using the robust statistical procedure from [67] (a maximum of two directions is retained). Then, for each point having scale  $\sigma$  and each dominant direction, a circular region with radius  $12\sigma$  is associated to it, divided into  $S$  sub-regions [65] (in all experiments in this paper, we use  $S = 9$ ). For each sub-region, a histogram of the orientation of the gradient with respect to the dominant direction is built over 12 bins. As in the original SIFT, orientations are weighted by the magnitude of the gradient. The final descriptor is made of the concatenation of the histograms. Observe that the building of these descriptors is given for completeness and is in no case a contribution of the present work. Again the approach to be described next is generic and could be applied with different local descriptors. In what follows, we write  $\{p_i^a, \mathbf{d}_i^a\}_{i=1,2,\dots,N^a}$  for the  $N^a$  keypoints and descriptors extracted from the first image  $I_a$ , and similarly for those extracted from the second image  $I_b$ .

The last step of this classical preliminary matching of keypoints is the matching step. We compare descriptors using the Circular Earth Mover Distance (CEMD) (noted as  $D_{CEMD}$  in what follows), obtained by summing circular transportation distances between orientation histograms, see [68]. For setting matching thresholds, we use the automatic *a contrario* procedure from [65]. An alternative would be to use the criterion on the ratio of distances proposed in [60]. The outcome of this step is a set of matched keypoints pairs between images  $I_a$  and  $I_b$ .

The previous procedure is able to deal with unregistered pairs of images having arbitrary displacement between them. However, the procedure is straightforwardly speeded up by assuming a maximum displacement amplitude between the two images (e.g. by using a rough pre-registration). In our experiments, we restrict the search for similar keypoints to a window of size  $W$ , a parameter to be specified in the experimental section.

## B. Keypoints mapping

The next step of our procedure is to refine the spatial correspondences found in the previous section by keypoint matching. There are two reasons for this. First, the detection of keypoints may be unstable from one image to the other, in particular when images are acquired with different sensors. Second, many keypoints in one image are not matched with any keypoint in the other image, or are wrongly matched. In order to associate to each keypoint a reliable corresponding position in the other image, we rely on the robust estimation of local transforms.

1) *Global and local mappings extraction*: From the matched pairs of keypoints, we seek a set of transforms that will be complete enough so that, for each keypoint, one of the transform will map the keypoint to its corresponding position in the other image. These transforms mostly account for two sources of disparity between images: one global rigid disparity and several local disparities due to parallax effects between

images acquired with different incidence angles. These effects are particularly strong in high resolution urban images, where tall buildings can shift over a large number of pixels.

The global transform will first be estimated thanks to all keypoints correspondences, and then parallax effects will be accounted for by the estimations of local transforms over sliding windows. In all the paper, we consider affine transforms (six parameters), both for the global mapping and the local ones, a reasonable choice in urban scenes for which structures are locally flat.

Both global and local transforms are classically obtained using a RANSAC procedure. Once the global transform has been estimated, the difficulty is to obtain locally the remaining local transforms needed to explain the disparity. For this, we rely on a *multiple* RANSAC procedure, the so-called *Multiple a contrario* RANSAC (MAC-RANSAC) introduced in [69].

The global mapping, that we denote  $f_0$ , is computed from all pairs of matched keypoints. It is obtained by a classical RANSAC procedure (or equivalently by retaining the most meaningful transform returned by the MAC-RANSAC algorithm). Then, we consider sliding windows in the first image, with size  $2L \times 2L$  and 50 percent overlap. For each window, the corresponding set of pairs of matched keypoints are used to feed the MAC-RANSAC algorithm, resulting in a set of local transforms. It is worth emphasizing that estimating several transforms using a RANSAC procedure is a tricky task, here permitted by the use of MAC-RANSAC. In particular, this algorithm automatically set the number of transforms that should be retained for each window.

Next, we map each point in the first image to its corresponding position in the second one, a position that we call the *mapping* of the point. For each keypoint  $p_i^a$  in  $I_a$ , we consider the global transform  $f_0$  and the  $n$  local transforms  $f_1, \dots, f_n$  corresponding to the local windows the point belongs to. The mapping of  $p_i^a$  is defined as the position  $f_k(p_i^a)$  in  $I_b$  most similar to  $p_i^a$ . To assess the similarity and characterize the mapping, each position  $f_k(p_i^a)$  is associated a new local descriptor, the scale and orientation of which are those of  $\mathbf{d}_i^a$ , the descriptor associated to the keypoint  $p_i^a$ . For the experiments considered in this paper, images roughly have the same scales and orientations. If this is not the case, the scales and orientations of the new descriptors have to be computed (in case the two images differ by a known rigid transform) or estimated as in the original descriptor definition. We write  $\mathbf{d}_{i,k}^b$  for this "projected" descriptor. The mapping of  $p_i^a$  is then defined as  $f_{\hat{k}}(p_i^a)$ , where the index  $\hat{k}$  is defined as

$$\hat{k} = \arg \min_k (D_{CEMD}(\mathbf{d}_i^a, \mathbf{d}_{i,k}^b)), \quad k = 0, 1, \dots, n, \quad (1)$$

$D$  being the Circular Earth Mover's Distance (CEMD) between descriptors and the set of transforms  $f_0, f_1, \dots, f_n$  being obtained as described above from all local windows containing the point  $p_i^a$ .

Similarly, each keypoint in  $I_b$  is mapped to a position in  $I_a$ .

Observe that the global transform  $f_0$  is added to the set of candidate transforms in Equation (1). This is useful if no local transform is found by the MAC-RANSAC procedure.

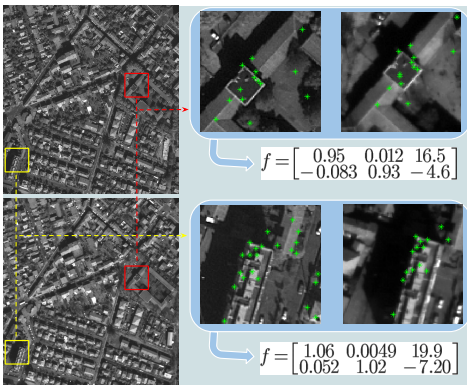


Fig. 2: A simple situations where two local transforms correspond to two groups of buildings with different heights, yielding different transform by parallax effect. Buildings in the red window are lower than in the yellow window and correspond to much smaller displacements.

Following the mapping of keypoints just described (from  $I_a$  to  $I_b$  and conversely) we obtain a set of pairs of descriptors that should correspond to the same physical positions on the scene. We write

$$\{x_i^a, x_i^b, \mathbf{d}_i^a, \mathbf{d}_i^b\}_{i=1, \dots, N}$$

for these positions and for the associated descriptors.

2) *Experiments on local mapping*: A simple situation illustrating the computation of local mappings is given in Fig. 2, where the two colored (red and yellow) windows correspond to group of buildings with different heights. Thanks to the MAC-RANSAC algorithm, different local mappings are found for each window.

To further illustrate the utility of the local mappings, we perform two experiments. In the first one, a synthetic image is obtained by juxtaposing two buildings extracted from a larger image. A second synthetic image is obtained by placing the buildings in different relative positions. We can see on Fig. 3(a) that the matched keypoints are detected as two different groups (corresponding to two different local transforms) by the MAC-RANSAC procedure. In Fig. 3(b), we show a more realistic experiment. We compare two small images containing buildings with different heights, extracted respectively from Geoeye-1 and Wordview-2. The MAC-RANSAC algorithm detect a different group (local transforms) for each building.

### III. CHANGE DETECTION

The main idea to detect changes is quite simple. First, we identify descriptors corresponding to the same physical location (that is, pairs  $(\mathbf{d}_i^a, \mathbf{d}_i^b)$ ) that have changed significantly. Second, we perform spatial grouping of these changed descriptors. For both steps, we draw on an *a contrario* methodology [70], that yields robust decision criteria.

#### A. Detection of changed keypoints

For each pair of positions  $(x_i^a, x_i^b)$  (ideally corresponding to the same physical location if the previous step has succeeded) we want to check if a change has occurred between images  $I_a$  and  $I_b$ . This is achieved by computing a distance between the

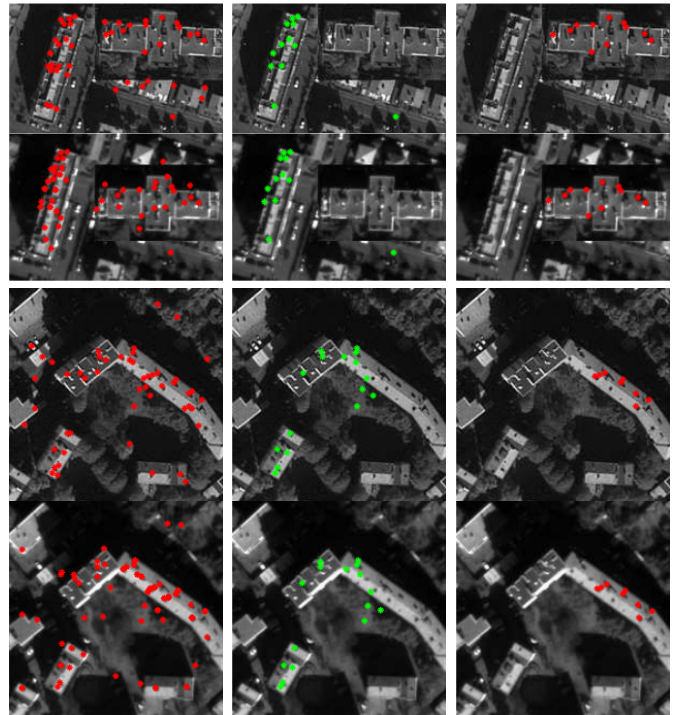


Fig. 3: Top two rows: Synthetic images (1-st and 2-nd row) made of the juxtaposition of two sub-images of buildings in different relative positions. Bottom two rows: Two image patches cut from Geoeye-1 (3-rd row) and Worldview-2 (4-th row).

descriptors and then setting a threshold thanks to a statistical procedure. For this, we rely on an *a contrario* approach [63], [70], both for robustness and genericity. The approach is similar to the one introduced in [65] for the comparison of local descriptors, except that in our case we want to assert the dissimilarity of local descriptors and not their resemblance.

The principle of *a contrario* methods is to set a random model, called *background model* or null hypothesis, for which it is assumed that no event should be detected. This automatically yield detection thresholds [63], [70]. In our case "no detection" means "no change". We will set the detection threshold so that, under the background model, most pairs of descriptors (the meaning of "most" is specified by parameter  $\epsilon_1$ , to be discussed in Section IV-B) will be at a distance smaller than this threshold.

In order to define the background model, we assume that the distance  $D$  between descriptors can be written as  $D(\mathbf{d}_i^a, \mathbf{d}_i^b) = \sum_{s=1}^S d(\mathbf{s}_{i,s}^a, \mathbf{s}_{i,s}^b)$ , where  $S$  is the number of sectors composing the local descriptors (see Section II-A),  $\mathbf{s}_{i,s}^a$  is the  $s$ th sector of descriptor  $\mathbf{d}_i^a$ , and  $d$  is some distance between sectors. This is in particular the case for the distance  $D_{CEMD}$  that we use, initially introduced in [65], for which  $d$  is the circular earth mover distance between orientation histograms.

As it is done in most *a contrario* approaches, the random model is defined through an independence assumption. In our case, this results in the following definition for the background model (or null hypothesis).

**Hypothesis 1 (Background model  $\mathcal{H}_0$ )** for any  $i$ , the random variables  $\{d(\mathbf{d}_{i,s}^a, \mathbf{d}_{i,s}^b)\}_{s=1, \dots, S}$  are mutually independent.

We denote  $z$  the random variable corresponding to the distance between two descriptors under  $\mathcal{H}_0$  hypothesis (background model for "no change" case). Using the previous independence assumption, we get that the probability that  $z$  is larger than a threshold is:

$$\Theta(\delta) := \mathbb{P}(z \geq \delta | \mathcal{H}_0) = \int_{\delta}^{+\infty} \mathbb{G}_s(z) dz. \quad (2)$$

where the probability density function (PDF) of the random variable  $d(\mathbf{d}_{i,s}^a, \mathbf{d}_{i,s}^b)$  is denoted as  $\mathbb{G}_s$  and  $*$  is the convolution.

The next step is then to set a threshold  $\delta$  on distances, so that false alarms are controlled. One of the key idea of *a contrario* methods is that it is much easier to compute the expected number of false alarms than the probability that a certain number of false alarms occurs [63]. Observing that the total number of comparisons between descriptors is  $N = N_a + N_b$ , we define as *changed* any pair  $(x_i^a, x_i^b)$ ,  $i = 1, 2, \dots, N$ , satisfying:

$$\Theta(D_{CEMD}(\mathbf{d}_i^a, \mathbf{d}_i^b)) := \mathbb{P}(z \geq D_{CEMD}(\mathbf{d}_i^a, \mathbf{d}_i^b) | \mathcal{H}_0) \leq \frac{\varepsilon_1}{N}, \quad (3)$$

where  $\varepsilon_1$  sets the bound on the expected number of false detections. With such a definition of "changed keypoints", we easily get that if we perform  $N$  comparisons between descriptors under  $\mathcal{H}_0$ , the expected number of changed keypoints is smaller than  $\varepsilon_1$ .

In order to be able to compute the probability given by Equation (2), we must estimate the density  $\mathbb{G}_s$ . In this work, in a way similar to [71], the densities are empirically estimated through the histograms of observed distances  $d(\mathbf{d}_{i,s}^a, \mathbf{d}_{i,s}^b)$ , for  $i = 1, \dots, N$ .

### B. Changed keypoints grouping

In the previous section, we have set a threshold on the distance between descriptors. After this step, we retain all pairs of descriptors at distance larger than the threshold as "changed".

Although the descriptors and matching method used are robust, they only use local information (at the scale of the descriptors) and some isolated changes, such as cars that have moved, may be detected. In most circumstances, such as the monitoring of natural disasters or for urban planning, such changes are not relevant. Therefore, we now introduce an automatic way to group the descriptors detected as changed, yielding larger "changed" areas that will be the final output of our algorithm. Again, this grouping is performed using an *a contrario* method. More precisely, we will first define a set of candidate regions by multi-scale circulars centered at each keypoints, and then for each region we will compare the number of detected keypoints and the number of changed descriptors. When these two numbers are suspiciously different (according to a statistical criterion to be defined) we will detect the region as "changed". Eventually, a maximality principle permits to avoid the detection of multiple regions.

We arbitrarily choose to perform the grouping in image  $I_a$ , but performing it on  $I_b$  would yield very similar results. Because keypoints are mapped symmetrically, so image  $I_a$  and image  $I_b$  have same distribution of keypoints and changed keypoints.

In order to estimate when the number of changed descriptors is high in a given region, we must first compute the probability that a descriptor is changed between the two images. We assume that this probability can be estimated globally over the image, although using a local estimation could also provide interesting results. If  $N_c$  is the total number of changed descriptors (according to the procedure of the previous section), then the probability  $\varrho$  that a descriptor is changed is estimated as  $\varrho = N_c/N$ .

Following the *a contrario* methodology, we detect groups of changed keypoints that are very unlikely under some background model. Again, this background model relies on an independence assumption between the individual elements to be grouped, in our cases changed descriptors. This hypothesis asserts that each descriptor may be changed with some probability, independently of the other descriptors. That is, we define the background model  $\mathcal{H}'_0$  as:

**Hypothesis 2 ( $\mathcal{H}'_0$ )** The number of changed keypoints in a local region containing  $n$  keypoints follows a binomial distribution  $B(n, \varrho)$ .

Therefore, the probability that this region contains more than  $m$  changed keypoints is

$$\Psi(m, n) := \mathbb{P}(k > m | \mathcal{H}'_0) = \sum_{k=m}^n B(k, n, \varrho), \quad (4)$$

where  $B(k, n, \varrho) = \binom{n}{k} \cdot \varrho^k \cdot (1 - \varrho)^{n-k}$ .

The meaningfulness of a given region is then defined through a Number of False Alarm (NFA) [70]. If the region has  $m$  changed descriptors for a total number of  $n$  keypoints, then the NFA is defined as  $\Psi(m, n)$  times the total number of considered regions [63]. In order to be able to detect changed regions with various scales, we consider, for each keypoint, a set of disks with different radiuses centered on this point. The set of radiuses we consider is obtained by uniformly sampling the interval  $[r_{min}, r_{max}]$ , with a step of  $n_r$ . Therefore the total number of tested regions is  $\Lambda = n_r \cdot N$ . Therefore, the NFA of a region containing  $n$  keypoints among which  $m$  descriptors have changed is:

$$NFA = \Lambda \cdot \Psi(m, n). \quad (5)$$

When working with large images, numerical values of the NFAs may become intractably small. In such cases, one can rely on the Hoeffding bound [72] on the tail of the binomial distribution to get an approximation of the NFA.

Then, a region will be detected if its NFA is smaller than a detection parameter  $\varepsilon_2$ . As before, this detection parameter has an intuitive meaning since it is an upper bound of the number of (false) groups that should be detected if hypothesis  $\mathcal{H}'_0$  is true.

The last step in order to decide which regions are changed is a maximality principle. Indeed, the set of all meaningful

changed regions (regions for which the NFA is smaller than  $\varepsilon_2$ ) is highly redundant and in particular contains many nested regions. Now, a strong asset of the NFA is that it may be compared between entities of different sizes (see e.g. [63], [73]). More precisely, We use a multi-scale strategy here, which means that at each change keypoint, we select several candidate regions with different sizes. Therefore, among all meaningful regions (disks) centered on a given keypoint, we only retain the most meaningful one, that is, the one having the smallest number of false alarms.

### C. Algorithm pipeline

The complete approach is summarized in Algorithm 1. The choice of parameters will be discussed in the next section, where we gather experimental results.

---

#### Algorithm 1 Change detection between two digital images

---

**Input:** Two images  $I^a$  and  $I^b$ , searching scope  $W$ , threshold for NFA  $\varepsilon_1$  and  $\varepsilon_2$ .

**Output:** Change detection result (a binary mask)

- 1: Extract SIFT-like keypoints and descriptors
  - 2: Match keypoints using an *a contrario* approach;
  - 3: Mappings. Compute a global transform; compute local transforms for each local region (overlapping sliding windows)
  - 4: Best match computation. Using the estimated transforms, associate to each keypoint a *mapping* (best corresponding position) in the other image, yielding pairs of keypoints and descriptors through Eq. (1)
  - 5: Change detection. Calculate the CEMD distance between each pair of descriptors and evaluate whether the pair is changed or not, using an *a contrario* approach with threshold  $\varepsilon_1$  through Eq. (2) and Eq. (3).
  - 6: Grouping. Consider multi-scale concentric circular regions centered at each key point and decide whether a region is changed (containing a large number of changed keypoints), by using an *a contrario* approach with threshold  $\varepsilon_2$  through Eq. (4) and Eq. (5).
- 

## IV. EXPERIMENTS AND ANALYSIS

We now proceed to the experimental validation of the proposed approach for change detection. We first detail the datasets we use. We then present the experimental setup, including a presentation of the parameters of the algorithm. Next, we provide a numerical evaluation.

In order to quantify the results, we build pairs of synthetic images on which changes are artificially added. Next, we show the utility of the local transforms involved in our algorithm. Eventually, we provide results on real and challenging high resolution image pairs, originating from different satellites or even directly acquired from Google Earth.

### A. Datasets

Our experimental datasets are taken from the following scenes.

- Scene-1: GeoEYE-1 and WorldView-2 data. The GeoEye-1 satellite sensor was launched in 2008 and provides a sub-metric resolution (between 0.4 and 0.6m depending on the viewing angle). WorldView-2 was launched in 2009 and also provides approximately a 0.5m resolution. The images are obtained on the city of Toulouse, France. The image from GeoEye-1 was taken in 2010, while the image from Worldview-2 was taken in 2012
- Scene-2: Google maps data. Both images are taken from the same place but at different times. More importantly, they also appear to be taken by satellites with different incident angles. Thus, the buildings in the images shift heavily from the first image to the second image. It is also worth noticing that these images have unknown origin, and in particular have undergone unknown processing stages (enhancement, compressions, etc.). This is both a challenging and realistic application case. We have roughly resized the spatial resolution of the images to 0.48m. The ground truth is obtained manually. The images were acquired near Avenue du Général Leclerc, Paris, France, in 2008 and 2014 respectively, with sizes  $850 \times 1000$  pixels.
- Scene-3: GeoEye-1 and QuickBird data. QuickBird II was launched in October 18, 2001 and provides a sub-metric resolution of 0.65m. Considering that the GeoEye-1 data has a different resolution, we pre-process them by ENVI to ensure the same spatial resolution of 0.65m and rotation by bicubic interpolation. The images are obtained on the city of West Ujimqin Banner, China. The images from QuickBird and GeoEye-1 were taken in 2007 and 2013 respectively.

### B. Experimental setup

Four parameters need to be set by users in our method. This section describes them in detail.

- Searching scope  $W$ :  
This parameter controls the searching scope used in the matching step. For each keypoint, only keypoints at a distance smaller than  $W$  in the other image are sought for. This restriction relies on the hypothesis that images are roughly registered and is mostly aimed at speeding up the process (although it may also eliminate some outliers). As such, it is not critical and can easily be adapted to a given situation. In our case, the spatial resolution of the considered image pairs varies from 0.46m to 0.65m. These images are roughly registered by their latitude and longitude. We have set  $W = 50$  (roughly 23m to 33m) once and for all, which appeared as a good compromise between matching accuracy and speed. The value  $W = 100$  has also been tested, yielding comparable results at the expense of speed.
- Local patch size  $w$ :  
This parameter controls the size of the sliding windows on which local transforms are computed. It should be big enough to contain the larger building of the analyzed scene and can be fixed depending on the image resolution. Again, for the considered images, the spatial resolution

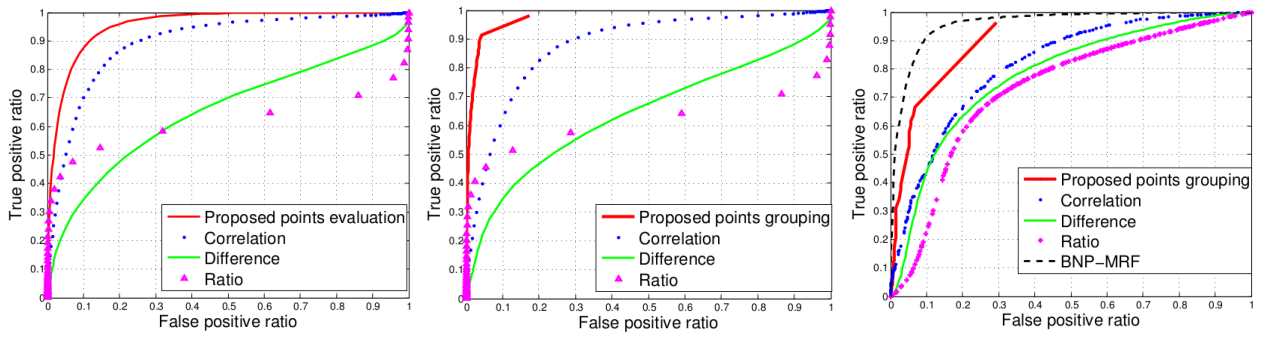


Fig. 4: Left: evaluation of the detection of *changed* descriptors. Middle: evaluation of the detection of *changed* regions. Right: comparison with the approach from BNP-MRF [74].

varies from 0.46m to 0.65m, so that a local patch with  $w=50$  covers a region with size 23m to 33m, covering most single buildings. Some large buildings cannot be covered by a single patch, but the chosen size appears large enough for a reliable estimation. Besides, the value of  $w$  should not be too small because in this case there may not be enough keypoints to estimate the local mapping functions.

- Threshold for the detection of changed descriptors  $\epsilon_1$ : Descriptors corresponding to the same spatial position in the two images are evaluated as changed or not by an *a contrario* method. This requires to set a bound  $\epsilon_1$  on the number of false detections. Since this parameter has a statistical meaning, its setting is easier and more robust than the setting of a threshold on the spatial distance  $D(p_i^a, p_i^b)$ . More information is given in Sec. IV E.
- Threshold for the grouping of changed keypoints  $\epsilon_2$ : The changed key points are grouped into changed regions using an *a contrario* multiscale strategy. Again this requires the setting of a parameter  $\epsilon_2$ , for which the same remarks as for  $\epsilon_1$  apply. More information is given in Sec. IV E.

We provide the setting of these parameters and discuss their effects for each foregoing experiment.

### C. Numerical evaluation on synthetic realistic images

We first consider synthetic images, on which the changes are artificially added, permitting a numerical evaluation. We start from twenty pairs of images acquired from Scene-1.

For each pair, we create changes by inserting excerpts from other images from the same satellites. These excerpts locations then provide the reference data for the detection of changes. Images are available on this web site<sup>1</sup>. The performances of our algorithm are then compared to methods relying on pixel-based comparisons. These methods are applied after a global registration using an homography estimated by SIFT and RANSAC-based methods. We use the same protocol as the one recently proposed in [74]. Performances are compared through ROC curves, displayed in Fig. 4. The first curve, on the left, evaluates the detection of changed descriptors. The Probability of False Alarms and the Probability of Detection

are computed using each detected keypoint. Keypoints outside the artificially added excerpt are considered as not changed, and the other ones as changed. We can see on this curve that the proposed method outperforms pixel-based ones, even though the correlation method gives good results in this case.

The second curve, in the middle, evaluates the grouping of keypoints. Contrarily to the previous ROC curve, this curve is obtained by considering all pixels of the image. Indeed the ultimate goal is to correctly classified all pixels as "changed" or "not changed". Here we can see that the proposed approach is clearly better than pixel-based approaches, which in particular are not able to deal with parallax effects.

The third curve in Fig. 4 includes also a comparison with the state of art method [74]. For this comparison, we use the candidate images from [74]. The used reference data not only includes the removed buildings, but also includes the changed plain ground, which is not taken into account by our method. As a consequence, the comparison is not balanced but we presented it for the sake of completeness.

### D. Usefulness of the local transforms

This section shows the necessity of estimating local transforms for the detection of changes between descriptors. In order to do so, we compare the results obtained with and without introducing the local transforms in addition to the global transform on two real images of the same sensor.

The corresponding results are shown in Fig. 5. The first column shows the input images  $I^a$  and  $I^b$ , the second column shows the evaluated changed keypoints when only a global mapping function is used. The third column shows the evaluated changed keypoints by using the proposed algorithm (combination of global and local transforms). Framed in blue and pink are details of the results. In particular, one sees that when no local transform is used, points on the circular elevated structure are mapped to wrong positions, which results in false detections of changed keypoints (red points in the blue rectangles for image  $I_a$  and  $I_b$ ). This is corrected by the estimation of local transforms (pink rectangles).

### E. Experimental results on heterogeneous images

In this section, we perform change detection on non-synthetic images, acquired by different satellites on the city of

<sup>1</sup><http://idiap.ch/~gliu/eusipco2016/changedet.html>



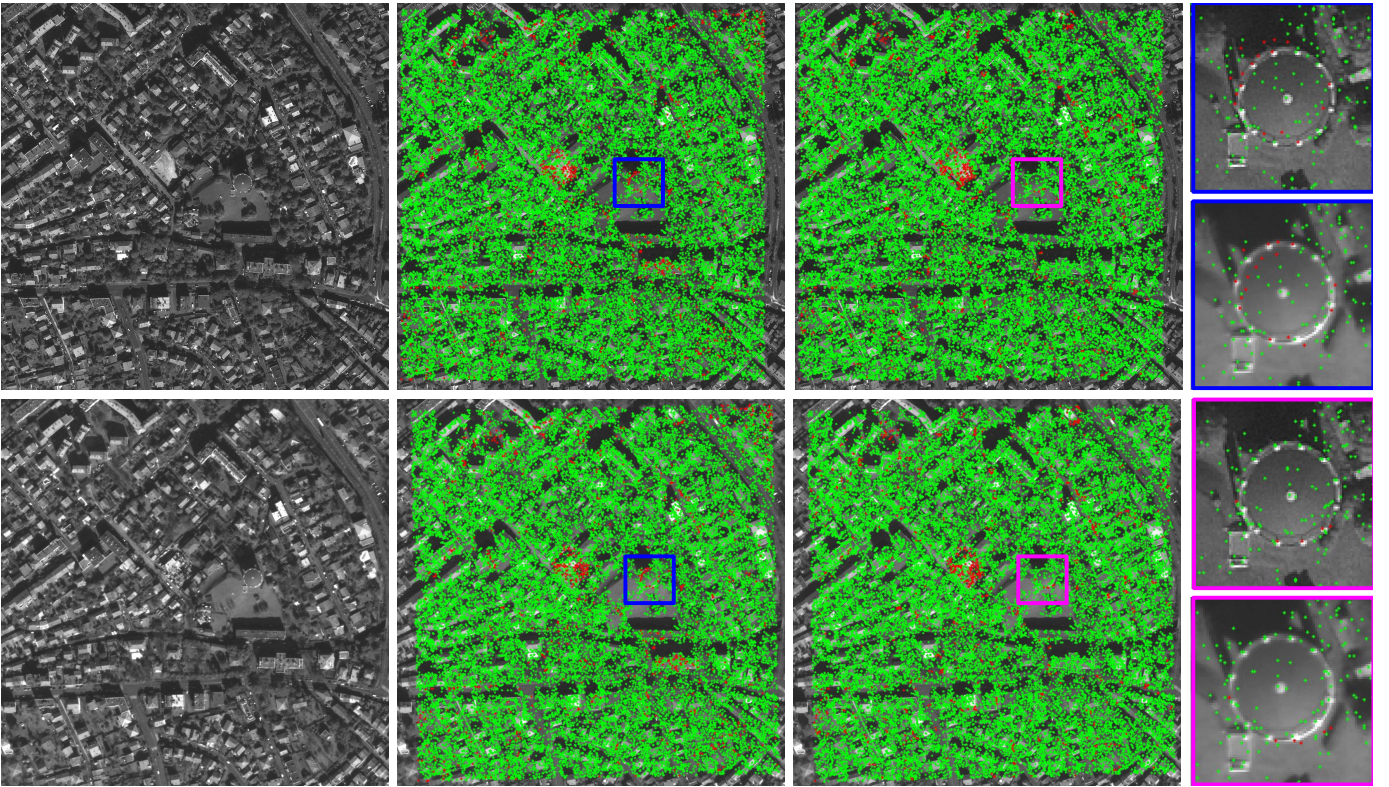


Fig. 5: This figure illustrates the utility of estimating local transforms. On the first column are the two original images taken by different sensors (Geoeye-1 and Worldview-2). On the second column, changed keypoints (in red) using the global approach (a single transform is estimated to map the keypoints) are shown for image  $I_a$  on the top and image  $I_b$  on the bottom. On the third column are shown changed keypoints obtained when estimating local transforms for both images. The fourth column displays details on the top without local transforms and on the bottom with local transforms.

Toulouse. Then, we apply our algorithm on images extracted from Google Earth, for which no calibration information is available.

1) *Experiment 1:* In this first experiment, we apply our algorithm to high resolution remote sensing images from Scene-1.

We choose three pairs of parts of the images and detect the changes between them, which have size of  $700 \times 700$ ,  $850 \times 700$  and  $1000 \times 1000$  pixels.

In this experiment, the parameter  $W$ , which used to restrict the searching scope for each key point, is set to  $W = 50$ . The two parameters  $\varepsilon_1$  and  $\varepsilon_2$ , which are used to control the number of false alarm (NFA) are set to be  $\varepsilon_1 = 1$  and  $\varepsilon_2 = 10^{-5}$  respectively.

The corresponding results are shown in Fig. 6, Fig. 7 and Fig. 8 respectively. In this experiment, images (a) and (b) are not registered and not aligned at the pixel level. In particular, it can be noticed in the left bottom corner of Fig. 7 (a) and (b), that the pixels corresponding to a tall building shift widely from the first image to the second one, over roughly 7 pixels. However, applying the MAC-RANSAC method over each local region with the matched keypoints, we can obtain multiple mapping functions. Ideally, the matched keypoints belonging to any single building can be grouped into a single set and be mapped to the other image. Thus, in any local region, each keypoint in the first image can be mapped to several locations in the second image. By choosing the most

similar ones as the mapped keypoints, we are able to ensure that most keypoints are mapped to the correct places in the second image.

Fig. (f) in Fig. 6, Fig. 7 and Fig. 8 illustrate the grouping results. Green keypoints are the points eventually detected as "changed". Note that our results are not so accurate at the border of the changed region. This is due to the use of sparse SIFT-like keypoints rather than a dense patching. Nevertheless, the final region covers most of the (manually estimated) ground truth.

Observe also that due to the use of SIFT descriptors, the proposed change detection procedure is not sensitive to very small changes, such as the moving cars in the street. This may be seen as an advantage, since change detections in high resolution remote sensing images are usually more concerned by changes on buildings than on moving cars. Of course, this aspect may be finely tuned by selecting the scale of the considered SIFT descriptors, depending on the targeted application.

2) *Experiment 2:* In this experiment, we process one pair of optical remote sensing images extracted from scene-2 (Google Earth), as shown in (a),(b) in Fig. 9. In particular, a large bus deposit was demolished between these dates and at this place (bottom right of the image).

In this experiment, the parameter  $W$ , which used to restrict the searching scope of each keypoint, is set to  $W = 50$ . The parameter  $\varepsilon_1$  which is used to control the number of false

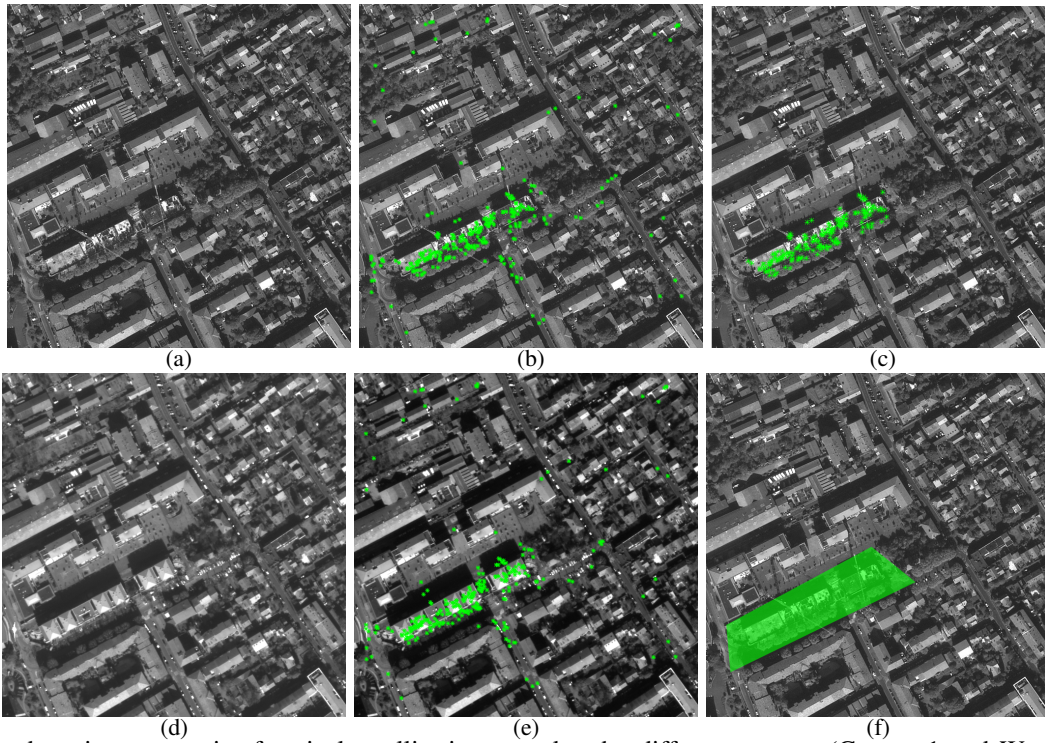


Fig. 6: Change detection on a pair of optical satellite images taken by different sensors (Geoeye-1 and Worldview-2). Both images have a resolution of  $0.46m$ . Figure (a) and figure (b) show the original images without co-registration. The middle two images, (c) and (d), show the changed key points (in green) for each image. Figure (e) shows the final result of change detection (green) overlaid on image (a) and figure (f) is the manual annotation.

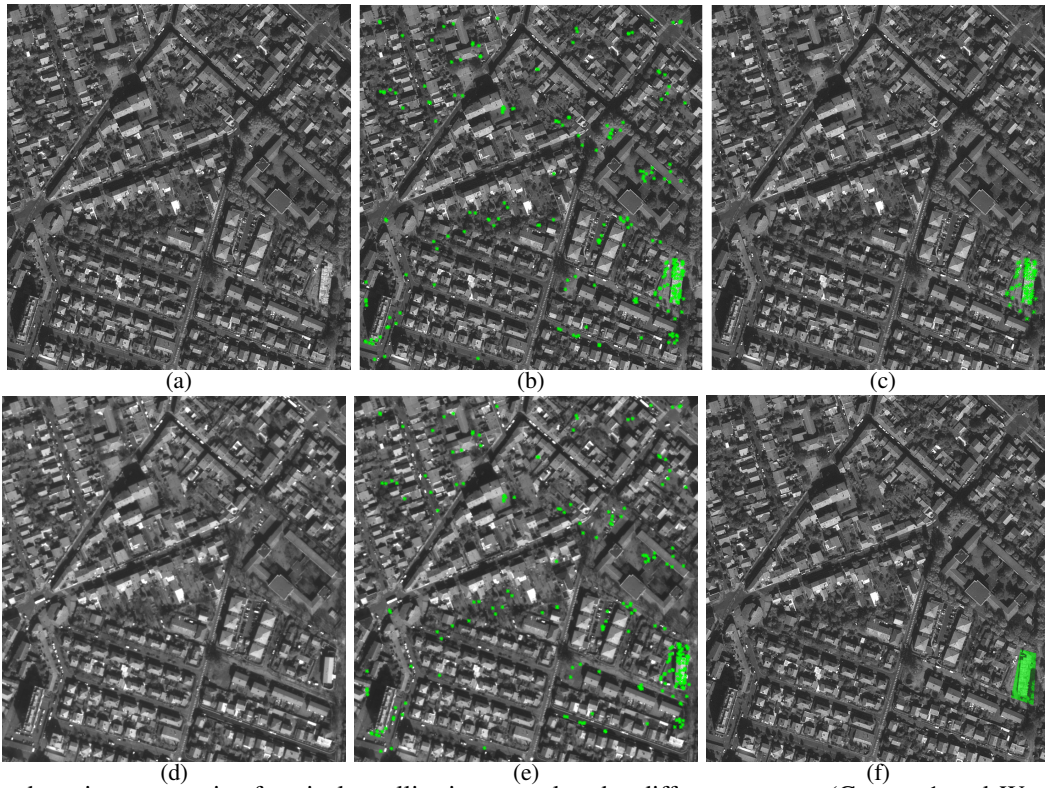


Fig. 7: Change detection on a pair of optical satellite images taken by different sensors (Geoeye-1 and Worldview-2). Both images have a pixel sampling of approximately  $0.5m$ . Figure (a) and figure (b) show the original images without co-registration. The middle two images, (c) and (d), show the changed key points (in green). Figure (e) shows the final result of change detection (green) and figure (f) is the manually annotation.

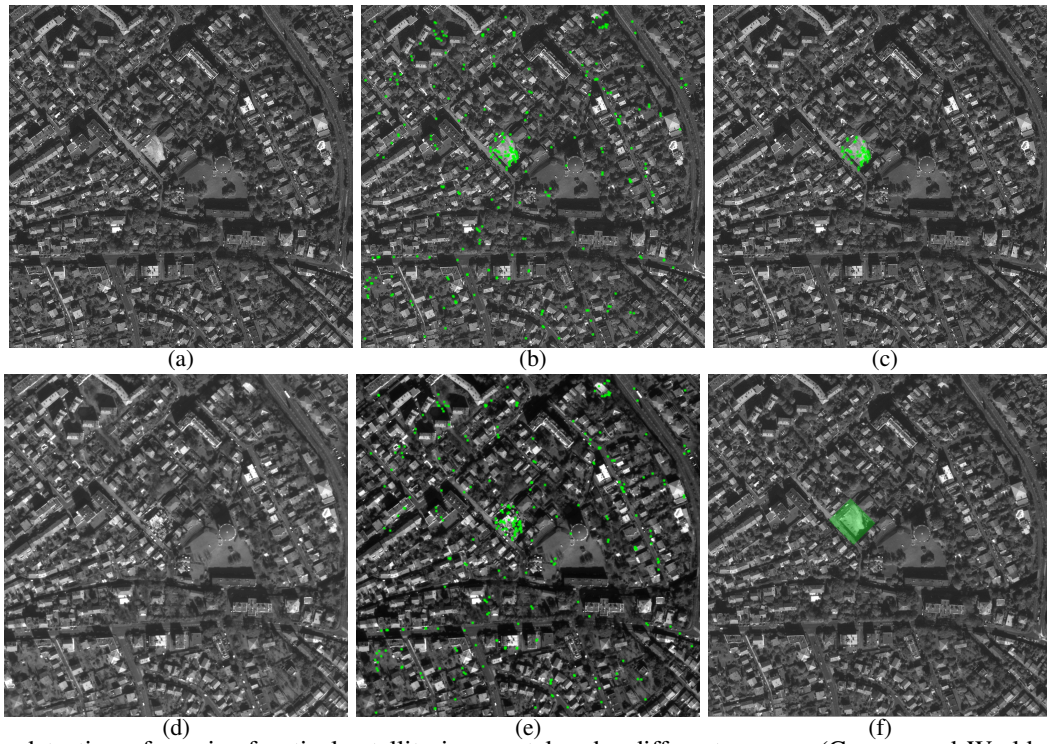


Fig. 8: Change detection of a pair of optical satellite images taken by different sensors (Geoeye and Worldview). Figure (a) and figure (b) show the original images without co-registration. The middle two images, (c) and (d), show the changed key points (in green) for both images. Figure (e) shows the final result of change detection (green) overlaid on figure (a) and figure (f) is the manual annotation.

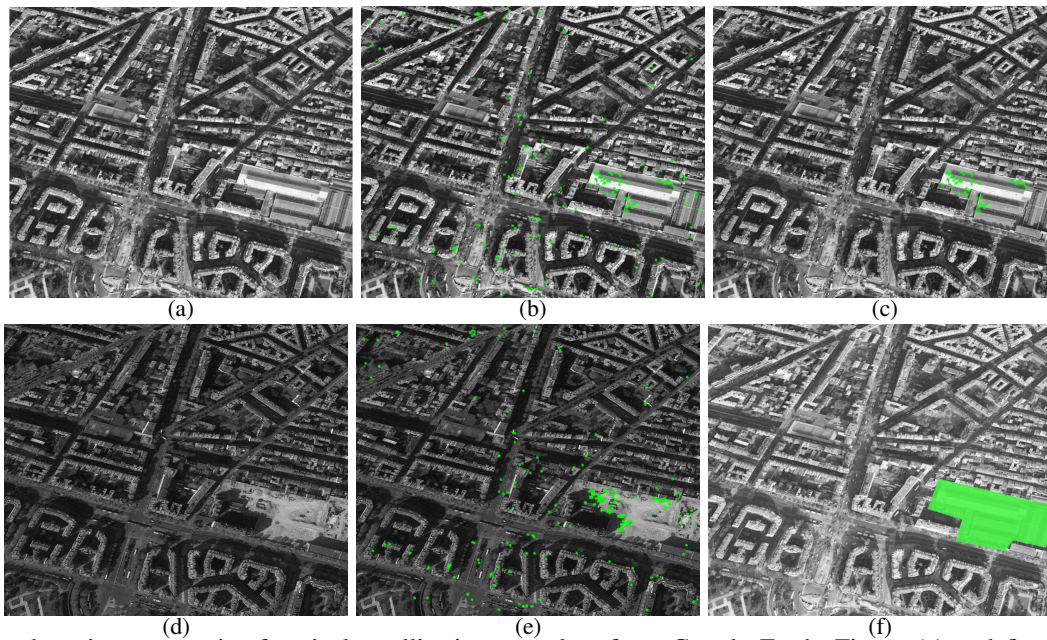


Fig. 9: Change detection on a pair of optical satellite images taken from Google Earth. Figure (a) and figure (b) show the original images without co-registration. The middle two images, (c) and (d), show the changed key points (in green). Figure (e) shows the final result of change detection (green) and figure (f) is the manual annotation.

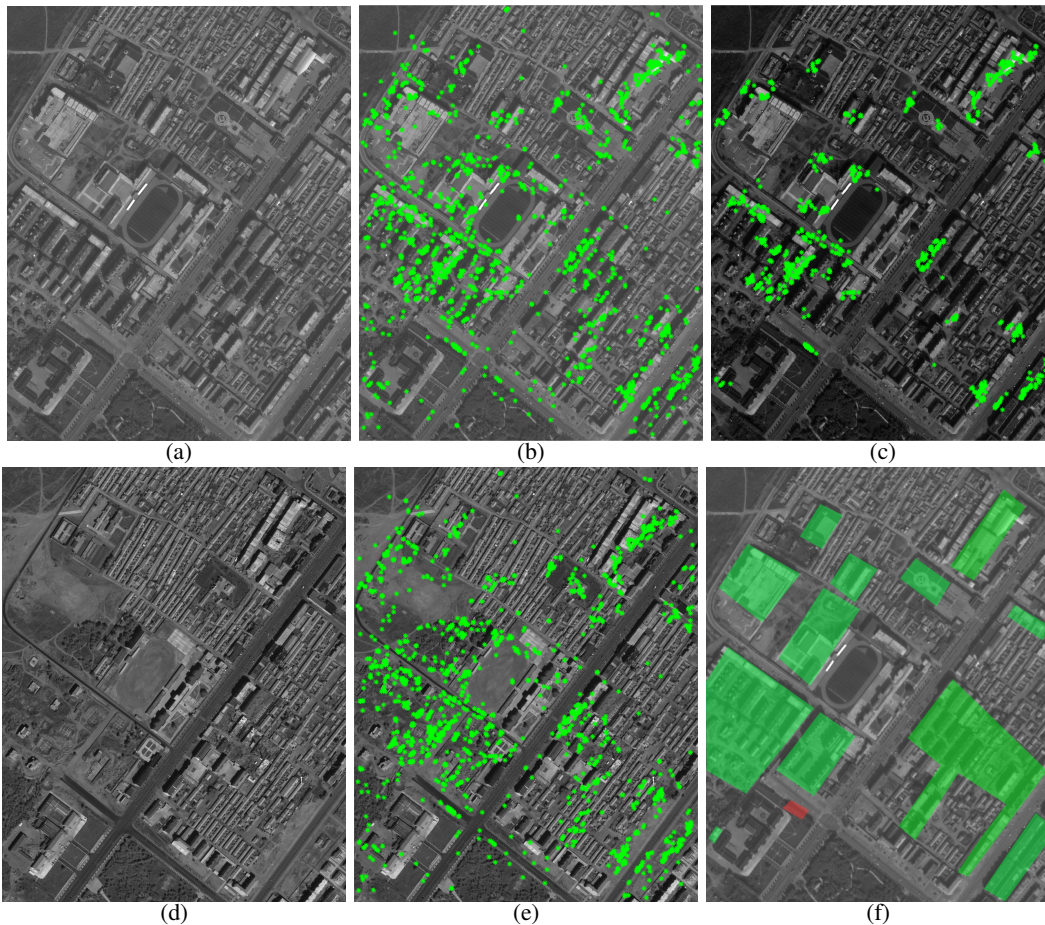


Fig. 10: Change detection of a pair of optical satellite images taken by different sensors (Geoeeye and Worldview). Figure (a) and figure (b) show the original images without co-registration. The middle two images, (c) and (d), show the changed key points (in green) for both images. Figure (e) shows the final result of change detection (green) overlaid on figure (a) and figure (f) is the manual annotation.

alarm (NFA) when detecting changed key points is set to be  $\varepsilon_1 = 1$ . However, as these high resolution remote sensing images from Google Earth are acquired in different situations, the other parameter  $\varepsilon_2$ , used to control the number of false alarm (NFA) when grouping of key points is set to  $\varepsilon_2 = 10^{-5}$ .

In Fig. 9 (e), our result indicates the exact location of the (manually evaluated) changed region. However, compared to the ground truth in (f), our mask does not cover some flat parts of the region. This is because there are few SIFT keypoints in these flat regions. This is one of the limitations of our algorithm, which should be adapted to dense points comparisons.

3) *Experiments 3*: In this experiment, regions of size  $2000 \times 2000$  have been extracted from Scene-3, as shown in Fig. 10(a) and (b), which represents a difficult case. The parameters for this experiment are  $W = 50$ ,  $\varepsilon_1 = 1$ ,  $\varepsilon_2 = -1$ .

In the scene, there are buildings, sport court, trees *etc.* and we annotated many changes in Fig. 10(g). The resolutions of the images are not similar and we first re-scaled and rotated the data by bicubic Interpolation using the ENVI software. It would have been possible to either automatically compute the scaling and rotation using methods in [75] or to rely on the scale-invariant SIFT to detect the scale difference between the two images, but these steps are beyond the scope of this paper.

Compared to the annotated groundtruth of Fig. 10(g), our result indicates most of the changed regions with a few key-

points. However, the key-points do not completely cover the changed regions. Besides, our result also has a false alarm indicated by the red rectangle in Fig. 10(f), which is caused by changing trees.

#### F. Discussion on the choice of parameters

In contrast with other methods, and in particular the pixel-based methods, the proposed method is largely based on an *a contrario* strategy, which alleviates the choice of parameters to some extent. Nevertheless, and as it was illustrated on Google images, this parameter choice still has its importance.

The first parameter  $\varepsilon_1$  is the threshold for detecting changed descriptors. The bigger this value, the more descriptors will be classified as changed. Fig. 11 illustrates the relationship between the number of changed descriptors and  $\varepsilon_1$ , from which we can see that the total number of changed descriptors increases slowly as  $\varepsilon_1$  becomes larger. Fig. 13 shows the changed descriptors for four different values of  $\varepsilon_1 = 2^{-3}, 2^0, 2^3, 2^6$ . If  $\varepsilon_1$  is very small, we detect less keypoints in changed and unchanged regions. On the contrary, if  $\varepsilon_1$  becomes larger, we detect more keypoints in both changed and unchanged regions. Under the *a contrario* hypothesis, the final result is robust if  $\varepsilon_1$  varies, which can be seen from Fig. 11 and Fig. 13. It can be seen that the parameter  $\varepsilon_1$  is a robust parameter. In our experiment, we set  $\varepsilon_1 = 1$ .

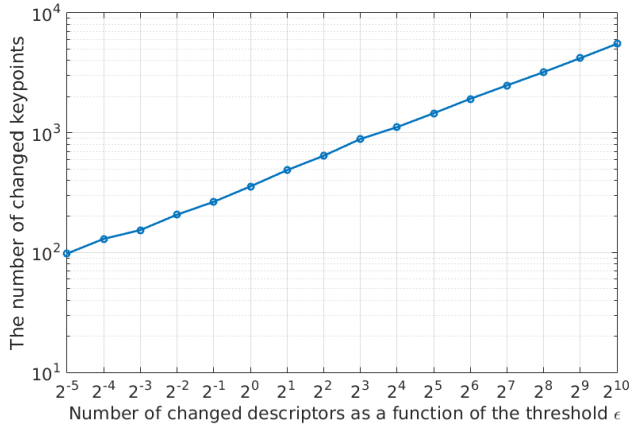


Fig. 11: Number of changed descriptors as a function of the threshold  $\varepsilon_1$ .

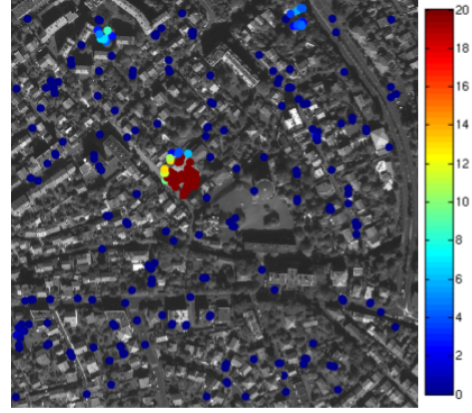


Fig. 12: Illustration of the  $-\log(NFA)$  for each keypoint. For better visualization.

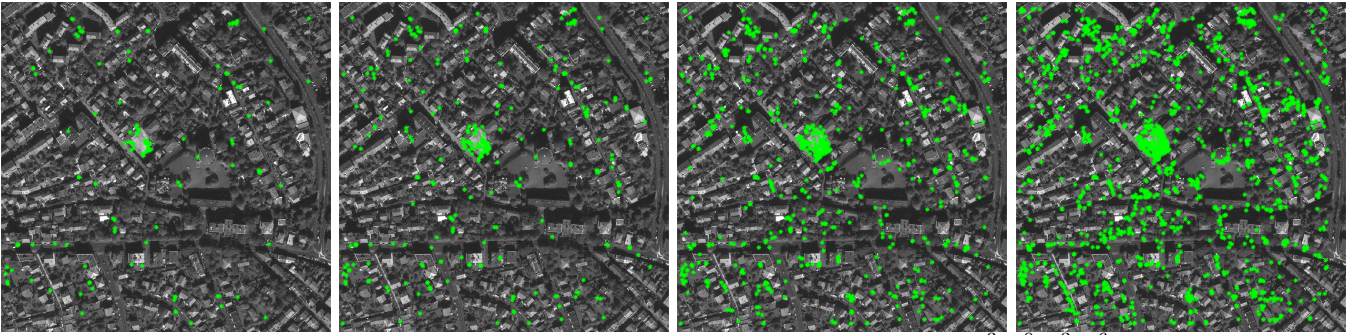


Fig. 13: Detected changed descriptors for various values of  $\varepsilon_1 = 2^{-3}, 2^0, 2^3, 2^6$ .



Fig. 14: Changed regions for different threshold  $\varepsilon_2 = 2^0, 2^{-3}, 2^{-6}, 2^{-9}$ , all experiments being performed with  $\varepsilon_1 = 1$ .

The second parameter  $\varepsilon_2$  is the threshold for grouping the keypoints. In Fig. 12 we show the regions whose number of false alarm (NFA) are less than 1. Here for better visualizing, the NFA value is transformed by  $-\log(NFA)$ . We can observe that the NFA is different in different regions. As expected, the regions with a larger number of changed descriptors have a larger value  $-\log(NFA)$ . In Fig. 12, the region where a construction site has completely disappeared has the largest value  $-\log(NFA)$ , while the regions with small changes like moving cars have a smaller value for  $-\log(NFA)$ . This indicates that the NFA is a reasonable indicator of the magnitude of changes. Fig. 14 illustrates the detection results with  $\varepsilon_2 = 2^0, 2^{-3}, 2^{-6}, 2^{-9}$ .

### G. Computing time

In this section we discuss the computing time of the algorithm. Recall that the algorithm is made of five steps:

keypoints extraction, matching, mapping, changed keypoints detection and grouping.

In the first step, the keypoints are first detected and the corresponding descriptor is classical. It accounts for about a quarter of the overall process. In the second step, we detect matched keypoints for each one from a large set of candidates. This is a costly step which is kept reasonable thanks to the pre-defined searching scope  $W$ . In the third step, keypoints from one image are mapped into the other image by using the MAC-RANSAC algorithm, which is processed in a sliding window with width  $w$  and 50% overlapping. In each window, affine mapping functions are estimated from 4 pairs of matched keypoints, using 1000 trials. The fourth step, estimating the threshold automatically by an *a contrario* procedure from the mapped keypoints, has marginal computational cost. The last step is more costly since it necessitates the grouping of the changed keypoints using Eq. (4) and Eq. (5) with a multiscale

TABLE I: Average time cost (in s) of each step for detecting changes on 33 pairs of images with size  $2000 \times 2000$ .

steps	Extraction	Matching	Mapping	Changed	Grouping
cost	$60.2 \times 2$	225.1	86.6	2.8	114.8

strategy. However, evaluation of Formula (4) is greatly reduced using Hoeffding inequality [72].

Table I shows the average cost of each step for 33 pairs of images with size  $2000 \times 2000$ . We set  $W = 50$  and  $w = 50$  in this experiment. The CPU is an Intel(R) Core(TM) i7-5930K with 3.50GHz. The program is written in C++ with MATLAB interface. The total cost of the change detection procedure is about 10 minutes, which is practicable in an offline scenario.

## V. CONCLUSION

In this work, we have proposed a change detection method enabling one to compare images having the same spatial resolution but that are not necessarily pre-registered. Thanks to the use of invariant local descriptors, the method is robust to radiometric changes and local geometric distortions, for instance corresponding to parallax effects. Because of this geometric robustness and because these descriptors are adapted to geometric structures, the method is especially adapted to high resolution urban scenes. There are several ways this work could be continued. First, the method could be extended to images with different spatial resolutions by taking advantage of the scale feature of the local descriptors. Second, we wish to produce a larger scale evaluation involving ground truth on high resolution urban images. Next, one should take into account shadows that in practice are responsible for frequent false detections. Last, the method will fail in case of low contrast scenes, where few keypoints are detected. In such situations, a dense keypoint extraction could be considered.

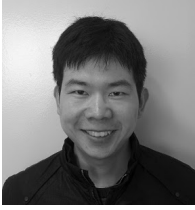
## ACKNOWLEDGMENT

This work has been supported by a Ph.D grant from the “Futur et Ruptures” program from IMT. The authors thank Jorge Prendes from University of Toulouse and Professor Wen Yang from Wuhan University for sharing some datasets.

## REFERENCES

- [1] Tung Fung, “An assessment of TM imagery for land-cover change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 28, pp. 681–684, jul 1990.
- [2] E. Rignot and J. van Zyl, “Change detection techniques for ERS-1 SAR data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 31, pp. 896–906, jul 1993.
- [3] L. Bruzzone and D. Prieto, “Automatic analysis of the difference image for unsupervised change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 38, pp. 1171–1182, may 2000.
- [4] P. Gamba, F. Dell’Acqua, and G. Lisini, “Change detection of multi-temporal SAR data in urban areas combining feature-based and pixel-based techniques,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2820–2827, 2006.
- [5] F. Bovolo, “A multilevel parcel-based approach to change detection in very high resolution multitemporal images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 6, pp. 33–37, Jan 2009.
- [6] F. Bovolo, G. Camps-Valls, and L. Bruzzone, “A support vector domain method for change detection in multitemporal images,” *Pattern Recognition Letters*, vol. 31, no. 10, pp. 1148–1154, 2010.
- [7] Y. Bazi, F. Melgani, and H. D. Al-Sharari, “Unsupervised change detection in multispectral remotely sensed imagery with level set methods,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 8, pp. 3178–3187, 2010.
- [8] F. Pacifici, “Unsupervised change detection frameworks for very high spatial resolution images,” *IEEE Int. Geosci. Remote Sens. Symp.*, pp. 2567–2570, 2010.
- [9] D. Lu, E. Moran, S. Hetrick, and G. Li, “Land-use and land-cover change detection,” *Advances in Environmental Remote Sensing Sensors, Algorithms, and Applications. CRC Press Taylor & Francis Group, New York*, pp. 273–290, 2011.
- [10] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, “Change detection for high resolution satellite images, based on SIFT descriptors and an a contrario approach,” in *IEEE Int. Geosci. Remote Sens. Symp.*, pp. 1281–1284, IEEE, jul 2014.
- [11] Y. Chen and Z. Cao, “An improved mrf-based change detection approach for multitemporal remote sensing imagery,” *Signal processing*, vol. 93, no. 1, pp. 163–175, 2013.
- [12] X. Su, C.-A. Deledalle, F. Tupin, and H. Sun, “NORCAMA: Change analysis in SAR time series by likelihood ratio change matrix clustering,” *ISPRS J. Photogramm. Remote Sens.*, vol. 101, pp. 247–261, mar 2015.
- [13] C. Marin, F. Bovolo, and L. Bruzzone, “Building change detection in multitemporal very high resolution SAR images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, pp. 2664–2682, may 2015.
- [14] P. Xiao, X. Zhang, D. Wang, M. Yuan, X. Feng, and M. Kelly, “Change detection of built-up land: A framework of combining pixel-based detection and object-based recognition,” *ISPRS J. Photogramm. Remote Sens.*, vol. 119, pp. 402–414, sep 2016.
- [15] A. P. Tewkesbury, A. J. Comber, N. J. Tate, A. Lamb, and P. F. Fisher, “A critical synthesis of remotely sensed optical image change detection techniques,” *Remote Sens. Environ.*, vol. 160, pp. 1–14, 2015.
- [16] Y. Zhong, W. Liu, J. Zhao, and L. Zhang, “Change detection based on pulse-coupled neural networks and the nmi feature for high spatial resolution remote sensing imagery,” *IEEE Geosci. Remote Sens. Lett.*, vol. 12, pp. 537–541, March 2015.
- [17] D. Peng and Y. Zhang, “Building change detection by combining LiDAR data and ortho image,” *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLI-B3, pp. 669–676, jun 2016.
- [18] G. Cao, “Unsupervised change detection in high spatial resolution remote sensing images based on a conditional random field mode,” *Eur. J. Remote Sens.*, vol. 49, pp. 225–237, jun 2016.
- [19] H. Yu, W. Yang, G. Hua, H. Ru, and P. Huang, “Change detection using high resolution remote sensing images based on active learning and markov random fields,” *Remote Sensing*, vol. 9, no. 12, p. 1233, 2017.
- [20] F. Pacifici, N. Longbotham, and W. J. Emery, “The importance of physical quantities for the analysis of multitemporal and multiangular optical very high spatial resolution images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6241–6256, 2014.
- [21] R. Dianat and S. Kasaei, “Change detection in optical remote sensing images using difference-based methods and spatial information,” *Int. Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 215–219, 2010.
- [22] N. Longbotham, F. Pacifici, T. Glenn, A. Zare, M. Volpi, D. Tuia, E. Christophe, J. Michel, J. Inglada, J. Chanussot, *et al.*, “Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009–2010 data fusion contest,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 5, no. 1, pp. 331–342, 2012.
- [23] N. Falco, M. D. Mura, F. Bovolo, *et al.*, “Change detection in VHR images based on morphological attribute profiles,” *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 636–640, 2013.
- [24] L. Giustarini, R. Hostache, P. Matgen, G. J.-P. Schumann, P. D. Bates, and D. C. Mason, “A change detection approach to flood mapping in urban areas using TerraSAR-X,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, pp. 2417–2430, apr 2013.
- [25] J. Chen, M. Lu, X. Chen, J. Chen, and L. Chen, “A spectral gradient difference based approach for land cover change detection,” *ISPRS J. Photogramm. Remote Sens.*, vol. 85, no. November, pp. 1–12, 2013.
- [26] D. Jin-Song, W. Ke, L. Jun, and D. Yan-Hua, “Urban land use change detection using multisensor satellite images,” *Pedosphere*, vol. 19, no. 1, pp. 96–103, 2009.
- [27] S. Pang, X. Hu, Z. Wang, and Y. Lu, “Object-based analysis of airborne LiDAR data for building change detection,” *Remote Sens.*, vol. 6, pp. 10733–10749, nov 2014.
- [28] A. Robin, L. Moisan, and S. Le Hégarat-Masclé, “An a contrario approach for subpixel change detection in satellite imagery,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1977–1993, 2010.
- [29] T. Celik and K.-K. Ma, “Unsupervised change detection for satellite images using dual-tree complex wavelet transform,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, pp. 1199–1210, mar 2010.
- [30] F. Bovolo, L. Bruzzone, and M. Marconcini, “A novel approach to unsupervised change detection based on a semisupervised SVM and a

- similarity measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2070–2082, 2008.
- [31] C. Carincotte, S. Derode, and S. Bourennane, "Unsupervised change detection on SAR images using fuzzy hidden Markov chains," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, pp. 432–441, feb 2006.
- [32] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, pp. 874–887, apr 2005.
- [33] Z. Yetgin, "Unsupervised change detection of satellite images using local gradual descent," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, pp. 1919–1929, may 2012.
- [34] M. Volpi, D. Tuia, G. Camps-Valls, and M. Kanevski, "Unsupervised change detection with kernels," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, pp. 1026–1030, nov 2012.
- [35] 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. Geosci. Remote Sens.*, vol. 45, pp. 778–789, mar 2007.
- [36] A. Ghosh, N. S. Mishra, and S. Ghosh, "Fuzzy clustering algorithms for unsupervised change detection in remote sensing images," *Inf. Sci. (Ny.)*, vol. 181, no. 4, pp. 699–715, 2011.
- [37] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Hierarchical unsupervised change detection in multitemporal hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, pp. 244–260, jan 2015.
- [38] Y. T. Solano-Correa, F. Bovolo, and L. Bruzzone, "An approach for unsupervised change detection in multitemporal vhr images acquired by different multispectral sensors," *Remote Sensing*, vol. 10, no. 4, p. 533, 2018.
- [39] L. Bruzzone and F. Bovolo, "A novel framework for the design of change-detection systems for very-high-resolution remote sensing images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 609–630, 2013.
- [40] X. Dai and S. Khorram, "The effects of image misregistration on the accuracy of remotely sensed change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 5, pp. 1566–1577, 1998.
- [41] Y. Bentoutou, N. Taleb, K. Kpalma, and J. Ronsin, "An automatic image registration for applications in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2127–2137, 2005.
- [42] M. N. Klaric, B. C. Claywell, G. J. Scott, N. J. Hudson, O. Sjahputera, Y. Li, S. T. Barratt, J. M. Keller, and C. H. Davis, "Geocdx: An automated change detection and exploitation system for high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2067–2086, 2013.
- [43] Y. Han, F. Bovolo, and L. Bruzzone, "An approach to fine coregistration between very high resolution multispectral images based on registration noise distribution," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6650–6662, 2015.
- [44] C. Benedek, M. Shadaydeh, Z. Kato, T. Szirányi, and J. Zerubia, "Multilayer markov random field models for change detection in optical remote sensing images," *ISPRS J. of Photogrammetry and Remote Sens.*, vol. 107, pp. 22–37, 2015.
- [45] J. L. Lisani and J.-M. Morel, "Detection of major changes in satellite images," in *International Conference on Image Processing*, vol. 1, pp. 1–941, 2003.
- [46] A. Fleener and G. Hoyer, "A Helmholtz principle approach to parameter free change detection and coherent motion using exchangeable random variables," *SIAM J. Imaging Sci.*, vol. 4, no. 1, pp. 243–276, 2011.
- [47] M.-T. Pham, G. Mercier, and J. Michel, "Change detection between sar images using a pointwise approach and graph theory," *IEEE Trans. on Geosci. and Remote Sens.*, vol. 54, no. 4, pp. 2020–2032, 2016.
- [48] F. Pacifici, F. Del Frate, C. Solimini, and W. J. Emery, "An innovative neural-net method to detect temporal changes in high-resolution optical satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 9, pp. 2940–2952, 2007.
- [49] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE trans. on neural networks and learning systems*, vol. 27, no. 1, pp. 125–138, 2016.
- [50] Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu, "Change detection based on deep siamese convolutional network for optical aerial images," *IEEE Geosci. and Remote Sens. Lett.*, vol. 14, no. 10, pp. 1845–1849, 2017.
- [51] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Urban change detection for multispectral earth observation using convolutional neural networks," in *Int. Geosci. and Remote Sens. Symp.*, IEEE, 2018.
- [52] R. C. Daudt, B. Le Saux, and A. Boulch, "fully convolutional siamese networks change detection," in *Int. conference on Image Processing*, vol. 2018, 2018.
- [53] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *arXiv preprint arXiv:1803.02642*, 2018.
- [54] N. Champion, D. Boldo, M. Pierrot-Descilligny, and G. Stamon, "2D building change detection from high resolution satellite imagery: A two-step hierarchical method based on 3D invariant primitives," *Pattern Recognition Letters*, vol. 31, no. 10, pp. 1138–1147, 2010.
- [55] L. Matikainen, J. Hyypä, E. Ahokas, L. Markelin, and H. Kaartinen, "Automatic detection of buildings and changes in buildings for updating of maps," *Remote Sens.*, vol. 2, no. 5, pp. 1217–1248, 2010.
- [56] D. Crispell, J. Mundy, and G. Taubin, "A variable-resolution probabilistic three-dimensional model for change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 489–500, 2012.
- [57] C. Guerin, R. binet, and M.-P. Deseilligny, "Automatic Detection of Elevation Changes by Differential DSM Analysis: Application to Urban Areas," *IEEE Trans. on Geosci. and Remote Sens.*, pp. 4020–4038, 2014.
- [58] J. Tian, S. Cui, and P. Reinartz, "Building change detection based on satellite stereo imagery and digital surface models," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 406–417, 2014.
- [59] 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 J. of Photogrammetry and Remote Sensing*, pp. 91–106, 1998.
- [60] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [61] H. Goncalves, L. Corte-Real, and J. A. Goncalves, "Automatic image registration through image segmentation and SIFT," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2589–2600, 2011.
- [62] G. Liu, J. Delon, Y. Gousseau, and F. Tupin, "Unsupervised change detection between multi-sensor high resolution satellite images," in *Eur. Signal Process. Conf.*, pp. 2435–2439, IEEE, aug 2016.
- [63] A. Desolneux, L. Moisan, and J.-M. Morel, "Meaningful alignments," *Int. J. of Computer Vision*, vol. 40, no. 1, pp. 7–23, 2000.
- [64] Y. Wang, L. Du, and H. Dai, "Unsupervised sar image change detection based on sift keypoints and region information," *IEEE Geosci. and Remote Sens. Lett.*, vol. 13, no. 7, pp. 931–935, 2016.
- [65] J. Rabin, J. Delon, and Y. Gousseau, "A statistical approach to the matching of local features," *SIAM Journal on Imaging Sciences*, vol. 2, no. 3, pp. 931–958, 2009.
- [66] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," in *IEEE International Conference on Computer Vision*, vol. 1, pp. 525–531, IEEE, 2001.
- [67] J. Delon, A. Desolneux, J.-L. Lisani, and A. B. Petro, "A nonparametric approach for histogram segmentation," *IEEE Trans. on Image Processing*, vol. 16, no. 1, pp. 253–261, 2007.
- [68] J. Rabin, J. Delon, and Y. Gousseau, "Circular earth movers distance for the comparison of local features," in *Int. Conference on Pattern Recognition*, pp. 1–4, IEEE, 2008.
- [69] J. Rabin, J. Delon, Y. Gousseau, L. Moisan, et al., "Mac-ransac: a robust algorithm for the recognition of multiple objects," *Proceedings of 3DPTV 2010*, 2010.
- [70] A. Desolneux, L. Moisan, and J.-M. Morel, *From gestalt theory to image analysis: a probabilistic approach*, vol. 34. Springer Science & Business Media, 2007.
- [71] J. Rabin, J. Delon, and Y. Gousseau, "A statistical approach to the matching of local features," *SIAM J. Imaging Sci.*, vol. 2, pp. 931–958, jan 2009.
- [72] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American statistical association*, vol. 58, no. 301, pp. 13–30, 1963.
- [73] G.-S. Xia, J. Delon, and Y. Gousseau, "Accurate junction detection and characterization in natural images," *Int. J. of computer vision*, vol. 106, no. 1, pp. 31–56, 2014.
- [74] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors," *IEEE Trans. on Image Processing*, vol. 24, no. 3, pp. 799–812, 2015.
- [75] B. Luo, J.-F. Aujol, Y. Gousseau, S. Ladjal, and H. Maître, "Resolution-independent characteristic scale dedicated to satellite images," *IEEE Trans. on Image Processing*, vol. 16, no. 10, pp. 2503–2514, 2007.



**Gang Liu** received the B.S. degree in communication engineering and the M.S. degree in electrical engineering from Wuhan University, China, in 2011 and 2013, respectively. Then he obtained the Ph.D. degree in computer vision from Telecom ParisTech, France. Currently he is working as a postdoc in Idiap Research Institute, Switzerland. His research involves texture analysis, including texture retrieving, and classification, remote sensing image processing and gaze estimation.



**Yann Gousseau** received the Engineering degree from the École Centrale de Paris, France, in 1995, and the Ph.D. degree in applied mathematics from the University of Paris-Dauphine, in 2000. He is currently a Professor at Telecom ParisTech. His research interests include the mathematical modeling of natural images and textures, stochastic geometry, computational photography, computer vision and image processing.



**Florence Tupin** (SM'07) is currently Professor at Telecom ParisTech, France. She received the Engineer degree in 1994 and the Ph.D. degree in signal and image processing from Ecole Nationale Supérieure des Telecommunications (ENST), Paris, France in 1997, and the Habilitation Diriger des Recherches degree from the University of Rennes in 2007. From 1997 to 1998, she worked with SAGEM on fingerprint recognition. Since november 1998, she has been Associate Professor and then Professor of image and signal processing at Telecom ParisTech in the Image, Data, Signal Department. Since 2014, she is head of the IMAGES (Image, Modeling, Analysis, GEometry, Synthesis) group. Her research interests are image processing and analysis, specially for remote sensing and SAR (Synthetic Aperture radar) imagery applications, and earth observation. In 2007, she was chair of the Urban Remote Sensing Joint Event held in Paris. Since 2003, she is also member of international and national technical committees. From 2005 to 2007, she was an Associate Editor of the French journal "Signal Processing" and from 2007 to 2016 she served as Associate Editor of IEEE Transactions on Geoscience and Remote Sensing. She is a coauthor of over 200 papers. She has received several awards, among them the IEEE TGRS Transactions Prize Paper Award in 2016 for works on non-local speckle filtering.