Unsupervised change detection between multi-sensor high resolution satellite images

Gang Liu*, Julie Delon†, Yann Gousseau*, and Florence Tupin*
*LTCI, CNRS, Telecom-ParisTech, Université Paris Saclay, 75013, Paris, France
{gang.liu, yann.gousseau, florence.tupin}@telecom-paristech.fr
†MAP 5, CNRS, Université Paris Descartes, Sorbonne Paris Cité, 75006, Paris, France
julie.delon@parisdescartes.fr

***Abstract*—In this paper, we present a novel unsupervised framework for change detection between two high resolution remote sensing images. Thanks to the use of local descriptors, the method does not need any image co-registration and is able to identify changes even with images acquired from different incidence angles and by different sensors. Local descriptors are used to both locally align images and identify changes. The setting of thresholds as well as the final grouping of isolated changes are performed thanks to *a contrario* statistical procedures. This provides a complete and automatic pipeline, whose efficiency is shown through several challenging pairs of high resolution urban images, acquired through different satellites.**

I. Introduction

Being able to detect changes between remote sensing images, with as little human intervention as possible, is crucial for many applications such as urban planning, disaster evaluation or land use monitoring. For this reason many methods have been developed to perform unsupervised change detection, see e.g. [1].

Over the years, the resolution of satellite acquisitions has increased significantly, yielding very complex scenes [2], especially over urban areas. Advanced sensors, such as Worldview-2 or Geoeye-2, can reach submetric resolution and produce images with abundant geometric details. In such cases, traditional pixel to pixel methods are prone to fail [2], and one should prefer methods relying on local structures instead of pixels.

Another tendency of modern remote sensing imaging is that scenes are acquired by a large variety of sensors. Comparing heterogeneous images is difficult because of different resolutions, incidence angles, shadows, or cloud cover. In particular, different incidence angles produce strong parallax effects over urban zones.

Now, the traditional and most widely used approaches to change detection rely on a first step where images are co-registered [3]–[5], 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 [6], MRFs [7], *a contrario* methods [8]–[10], morphological attribute profiles [11] or neural network (NN) models [12].

Such pixel-based approaches fail when the registration is not accurate enough, in particular on urban areas, for which more sophisticated methods have been proposed. These methods usually rely on an explicit modeling of buildings and the use of a Digital Surface Model (DSM), as in [13]–[15].

These approaches yield accurate results and are not sensitive to the parallax issue previously mentioned. However, they necessitate the knowledge of DSM models 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 rely on local descriptors having a prescribed degree of invariance, such as the descriptors developed for computer vision applications, the most well known of which is the SIFT descriptor [16]. Indeed, these descriptors make it possible to identify common geometric structures between images even in the presence of strong geometrical or radiometrical distortions. Because such distortions will also be encountered in remote sensing imaging, especially in the aforementioned cases (different resolutions, parallax distortions, cloud coverage, shadows, etc.), such descriptors may be helpful to identify changes. However, these have been used mostly for the pre-registration of images [17], and hardly for the change detection step itself, with the exception of [18], where some of the authors of the present paper have proposed to detect changes based on the assumption that few correspondences between descriptors will be found in changed areas. In this paper, we build from similar ideas and develop a full change detection pipeline, taking into account the local geometry of the scenes and explicitly detecting changed descriptors, yielding a more robust and precise identification of changed regions. The approach does not assume any pre-registration between images and also yield satisfactory results when heterogeneous sensors are used. Because the method relies on geometric descriptors and explicitly model local deformations, it is especially adapted to high resolution urban scenes. To the best of our knowledge, this is the first operational method of this kind that does not explicitly rely on building or surface modeling.

In short, the proposed method works as follows. Starting with a pair of unregistered images, keypoints and their corresponding SIFT-like descriptors are first extracted and matched between images. From these matches, a single global transform is inferred between images, as well as several locally evaluated transforms. For each keypoint in one of the image, we test the computed transforms and retain the best location in the other image. We call this location the *mapping* of the

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

keypoint. Thanks to the local transforms, this step allows to deal with different incidence angles and disparate heights. Then we proceed to the change detection itself. Thanks to an *a contrario* approach [19], we identify keypoints that are significantly different from their mapping as *changed*. Then, these *changed* keypoints are grouped, again using an *a contrario* methodology, resulting in the final detected regions.

The remainder of the paper is organized as follows: in Section II, we define keypoints, descriptors and their mapping. In Section III, we define the changed keypoints and the corresponding grouping procedure. We test the resulting algorithm in Section IV on several challenging pairs of high resolution images.

II. Keypoints mapping

We assume that we have two unregistered images at hand. In this section, we briefly explain how the local features are extracted and matched, and then how their mapping (best corresponding location) is found in the other image.

A. SIFT extraction and matching

We use a robust variant of the SIFT [16] method, as described in [20], but virtually any local descriptor could be used. Keypoints (and their scale) are detected as space-time extrema of the Laplacian and then filtered by a multi-scale Harris test. One or two dominant directions are then associated to each keypoint. Descriptors are then made of the concatenation of histograms of the gradient orientation on S non-overlapping pie-shaped regions. As in the original SIFT, these histograms are weighted by the gradient magnitude. We refer to [20] for the full description of these descriptors. In what follows, we write $\{p_i^a, \boldsymbol{d}_i^a\}_{i=1,2,\cdots,N^a}$ 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.

As a first step, we match keypoints between the two images following the procedure from [20]. To compute the similarity between descriptors, we use the circular earth mover distance (CEMD) [20].

B. Key points mapping

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. The mapping system aims at solving two problems: the registration of the global geometric position and the correction of parallax effects that occur when images have been 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. We therefore proceed as follows : first a global mapping is computed using feature correspondences, and then local transforms are estimated on sliding windows. In this work, we consider affine transforms (six parameters), both for the global mapping and the local ones.

This appears to be a reasonable choice for urban scenes, where the structures are locally plane. In order to infer these transforms from the initial keypoints correspondences, we rely on a robust multiple RANSAC procedure as presented in [21], the so-called Multiple *a contrario* RANSAC (MAC-RANSAC).

The global mapping f_0 is computed from all matched keypoints. In this case, we keep only the best transform computed by the MAC-RANSAC algorithm, which is then similar to a standard RANSAC procedure.

Then, we split the images into local windows with size $2L \times 2L$ using a sliding window with 50 percent overlapping. For each window, we consider all matched keypoints (from the first step of the algorithm) falling into the window. From these correspondences, we obtain a set of transformations using the MAC-RANSAC algorithm. Notice that this is made possible by the ability of the MAC-RANSAC algorithm to find several transforms and to automatically decide which are meaningful, a task that would be far from trivial using the plain RANSAC algorithm.

Fig.1 illustrates this step by giving a simple example of computation of local mapping functions in two windows (red and yellow).

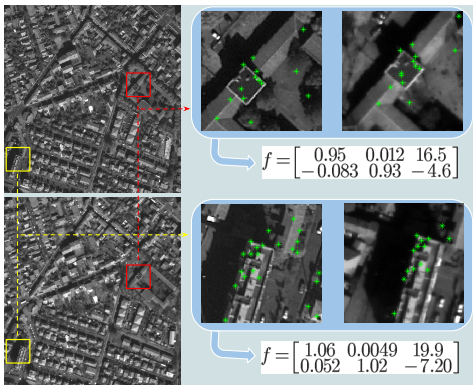

Fig. 1. Illustration of the necessity of local transforms. The buildings in the red window are lower than the buildings in the yellow window, so that the displacement of pixels in the yellow window is much larger than the displacement of pixels in the red window, which gives rise to two different mapping functions.

For a given point p_i^a in I_a, we assume that we have computed n local transforms $f_1, \ldots, f_n$ in the corresponding local window. We then find its *mapping* in I_b as the position $f_k(p_i^a)$ which is the most similar in I_b to p_i^a in I_a. For this, we compute, for each position $f_k(p_i^a)$, a new local descriptor inheriting its scale and orientation from $\boldsymbol{d}_i^a$, the descriptor at point p_i^a. We call this new descriptor $\boldsymbol{d}_{i,k}^b$. The mapping of p_i^a is then defined as $f_{\hat{k}}(p_i^a)$, where the index $\hat{k}$ is defined as

$$\hat{k} = \arg\min_k \left(D_{CEMD}(\boldsymbol{d}_i^a, \boldsymbol{d}_{i,k}^b) \right), \tag{1}$$

where D denotes the Circular Earth Mover's Distance(CEMD) between descriptors, as introduced above, and the set of transforms $f_1, \ldots, f_n$ corresponds to the local window the keypoint p_i^a belongs to. The same process is then repeated with keypoints from I_b that are mapped to positions in I_a. In

short, we choose for each keypoint the best position in the other image according to a set of possible transforms.

Following this mapping step, we are given a set of pair of positions, together with associated local descriptors, that ideally correspond to the same physical positions of the imaged scene, in both images I_a and I_b. From now on, we write

$$\left\{x_i^a, x_i^b, \boldsymbol{d}_i^a, \boldsymbol{d}_i^b\right\}_{i=1,\dots,N}, \quad N = N_a + N_b$$

for these positions and associated descriptors.

III. Change detection

We can now proceed with the change detection, by comparing the descriptors corresponding to the same position, that is $\boldsymbol{d}_i^a$ and $\boldsymbol{d}_i^b$, for $i = 1, \dots, N$. We first detect positions at which the descriptors are significantly different, by relying on an *a contrario* methodology to set the detection thresholds. Then, we group the detected positions, again using an *a contrario* methodology. The choice of this statistical framework for the detection of changed regions is motivated by the robustness of such approaches, as well as by their genericity.

A. Changed keypoints evaluation

Given a pair of corresponding positions (mapped keypoints) (x_i^a, x_i^b), we want to decide whether they correspond to a change between image I_a and I_b or not. For this, we compute their similarity as $D(\boldsymbol{d}_i^a, \boldsymbol{d}_i^b)$ in order to label as *changed* the positions at which descriptors differ significantly. In order to set a threshold on this similarity, we rely on an *a contrario* approach [19] similar to the one in [20]. The principle of this approach is to set some random model, called *background model* or null hypothesis, in which we assume there should be no detections (changed keypoints in our case). For this, we assume that the distance D between descriptors can be written as $D(\boldsymbol{d}_i^a, \boldsymbol{d}_i^b) = \sum_{s=1}^{S} d(\boldsymbol{d}_{i,s}^a, \boldsymbol{d}_{i,s}^b)$, where S is the number of sectors the local descriptors are made of (see Section II-A). This is in particular the case for the CEMD [20].

As it is common when setting matching thresholds using the *a contrario* framework [20], we define the null hypothesis $\mathcal{H}_0$ as

Hypothesis 1 ($\boldsymbol{\mathcal{H}_0}$) *for any index i, the random variables $\{d(\boldsymbol{d}_{i,s}^a, \boldsymbol{d}_{i,s}^b)\}_{s=1,\cdots,S}$ are independent.*

Under this hypothesis, and writing z for the random variable corresponding to the total distance between two descriptors, we get, for any threshold δ,

$$\Theta(\delta) := \mathbb{P}(z \geq \delta | \mathcal{H}_0) = \int_{\delta}^{+\infty} \underset{s=1}{\overset{S}{*}} \mathbb{G}_s(z) dz. \qquad (2)$$

where the probability density function (PDF) of the random variable $d(\boldsymbol{d}_{i,s}^a, \boldsymbol{d}_{i,s}^b)$ is denoted as $\mathbb{G}_s$. The principle of *a contrario* method is then to set a threshold on this probability so that the expectation of the total number of false detections is bounded [19]. Recalling that the total number of pairs on which the detection is tested is equal to $N = N_a + N_b$, we define as *changed* any pair $(x_i^a, x_i^b), i = 1, 2, \cdots, N$, satisfying

$$\Theta\left(D_{CEMD}(\boldsymbol{d}_i^a, \boldsymbol{d}_i^b)\right) := \mathbb{P}(z \geq D_{CEMD}(\boldsymbol{d}_i^a, \boldsymbol{d}_i^b) | \mathcal{H}_0) \leq \frac{\varepsilon}{N}, \qquad (3)$$

where ε sets the bound on the number of false detections, see [19]. The last step, in order to compute the above probability, is to estimate the density $\mathbb{G}_s$. These are simply estimated through the histograms of observed distances $d(\boldsymbol{d}_{i,s}^a, \boldsymbol{d}_{i,s}^b)$, when $i = 1, \dots, N$.

B. Changed keypoints grouping

Once the *changed* keypoints have been detected as explained in the previous section, we aim at detecting local regions by grouping these keypoints. Again, we rely on an *a contrario* approach to do this in an automatic way . More precisely, local regions are evaluated by considering, in a given region, the total number of keypoints and the number of changed points.

As the keypoints $\{x_i^a\}$ in image I^a and $\{x_i^b\}$ in image I^b share a one to one mapping, we can consider only points in I_a to perform the grouping. We first assume that the probability for a given keypoint to be changed can be estimated globally, so that, writing N_c for the total number of changed keypoints (points detected by the procedure of the previous section),

Hypothesis 2 *The probability ϱ that any keypoint is a changed keypoint is $\varrho = N_c/N$.*

Following the *a contrario* methodology, we detect groups of changed keypoints that are unlikely under the hypothesis that each keypoint is labelled as "changed" independently of the other keypoints. That is, we define the null hypothesis $\mathcal{H}_0'$ as

Hypothesis 3 ($\boldsymbol{\mathcal{H}_0'}$) *The number of changed keypoints in a local region containing n keypoints follows a binomial distribution $B(n, \varrho)$.*

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

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

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

Following the *a contrario* method, we calculate the meaningfulness of each local region by its Number of False Alarm (NFA) and set a threshold on it. This NFA is defined as the above probability times the total number of tested regions [19]. In order to cover a wide enough range of possible scales for the regions to be detected, without testing every possible regions, our candidate local regions are chosen as multi-scale circulars regions with radius $r \in [r_{min}, r_{max}]$ centered at each keypoint. Suppose there are n_r scales for each keypoint, then the total number of candidate local regions can be computed as $\Lambda = n_r \cdot N$. Therefore, the NFA of a region containing n keypoints among which m are changed keypoints is

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

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

For each series of regions from r_{min} to r_{max} centered at the single keypoint, we calculate an NFA for each region and choose the minimal one for this series (therefore following the same exclusion principle as in [22]). Eventually, we threshold the NFAs by ε_2 for the detection of the final regions,

$$NFA_i < \varepsilon_2, i = 1, \cdots, N, \tag{6}$$

IV. Experiments and analysis

In this section, we validate the proposed algorithm though artificial and real experiments. For all experiments, we use the following parameters. Local descriptors are computed with the same parameters as in [20] and are matched using a matching parameter value of 1. Sliding windows used to compute the local transforms is $L = 50$, the MAC-RANSAC algorithm [21] is run with a detection parameter of 1.

a) Synthetic experiments: We first detect changes that are artificially added to pairs of high resolution images of urban scenes, acquired from different satellites, Geoeye-1 and Worldview-2, at two different times, 2009 and 2010. Both satellites provide a resolution of 0.46 meters. We use three pairs of images on which changes are created by inserting excerpts from other urban images from the same satellites. Images are available at this address[1]. We use such synthetic image pairs because we are not aware of databases of annotated high resolution heterogeneous images. On these images, we evaluate performances through ROC curves. Our algorithm is compared with classical pixel-based methods (difference, ratio and correlation) after a global co-registration performed by SIFT matching. This evaluation protocol is actually the same one as recently proposed in [23]. We provide two ROC curves. The first one, Figure 2, is aimed at evaluating the detection of *changed* keypoints. Probability of False Alarms and Probability of Detection are computed only on detected keypoints for all methods. The second curve, Figure 3 is aimed at evaluating the grouping stage of the method. This time, we evaluate results on all pixels (for our approach, a pixel is classified as changed if it belongs to one of the disk detected as explained in Section III-B).

The first curve shows the proposed method is more efficient than the three proposed alternatives, although the correlation method after global registration gives good results. On the second curve, we can see the proposed method yields much better final regions than pixel-based methods after a global registration. In particular these methods are not able to deal with parallax effects that are very common in high resolution urban images. Future comparison should be made with more sophisticated methods such as those from [7] or [23].

b) Real experiments: Next, we provide detection results between challenging pairs of images that do contain changes. For these two experiments, the detection parameters are set

[1]http://perso.telecom-paristech.fr/∼gliu/eusipco2016/changedet.html

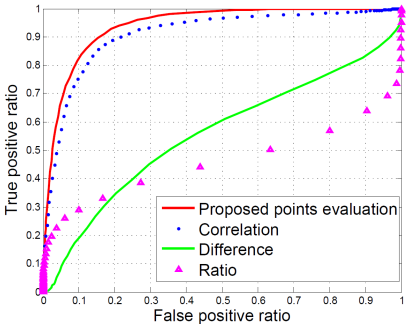

Fig. 2. Evaluation of the detection of *changed* keypoints.

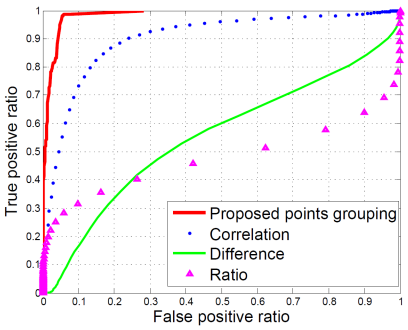

Fig. 3. Evaluation of the grouping of *changed* keypoints.

to $\varepsilon = 1$ and $\varepsilon_2 = 10^{-5}$. The scales of the grouping region are set to $r_{min} = 20$ and $r_{min} = 50$. The first example is again made of two images from Geoeye-1 and Worldview-2, on the city of Toulouse. The second example is made of images extracted from Google Earth, acquired near Avenue du Général Leclerc, Paris, in 2008 and 2014 respectively. In particular, a large bus deposit was demolished between these dates and at this place (bottom right of the image). On these examples, one can visually check that the detected changes correspond to large changed areas. On these examples, it also appears that the detected changed regions are slightly bigger than the real changed areas, so that a further spatial refinement step could be developed. Observe also that they are many local changes between images from the pairs, in particular because of tall buildings, as well as radiometric changes, and that these are well handled by the method.

V. Conclusion

In this work, we have proposed a change detection method enabling one to compare images that are not registered. Thanks to the use of invariant local descriptors, the method is robust to radiometric changes and local geometric distortions. 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, 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

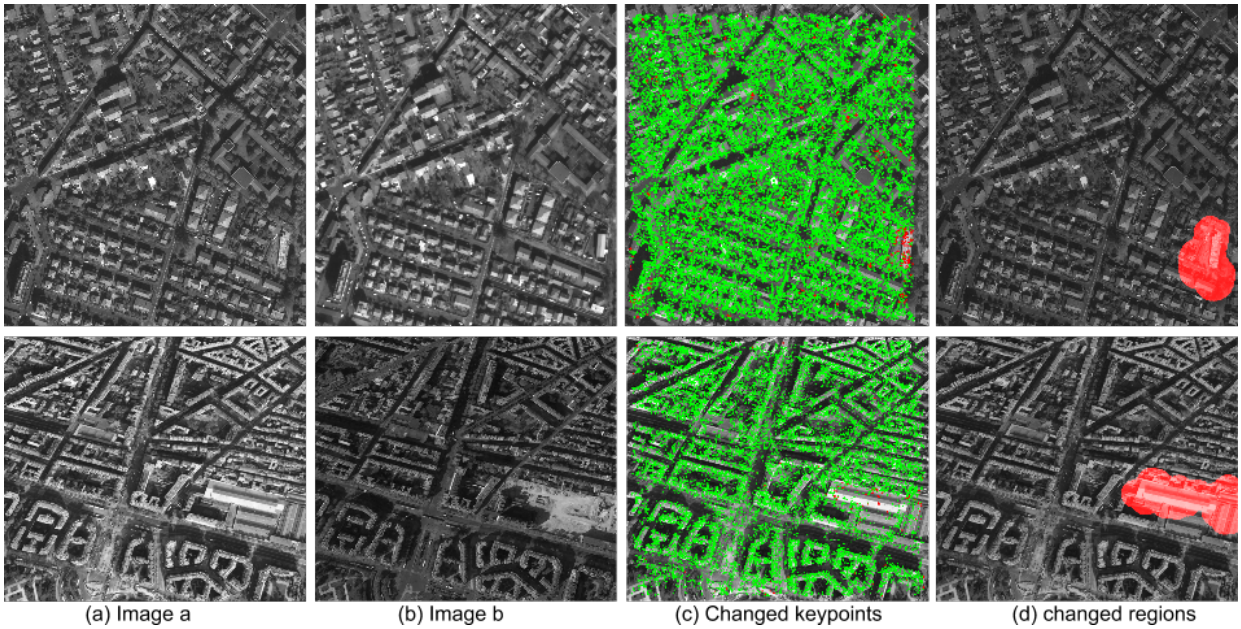

Fig. 4. Examples of change detection. First line : Images are from Geoeye-1 (leftmost image) and Wordview-2. Second line : Images are extracted from Google earth and correspond to two different times (2008 and 2010). From left to right on each line: two original images ((a) and (b)), changed keypoints (c), changed regions (d).

low contrast scenes, where few keypoints are detected. In such situations, a dense keypoint extraction could be considered.

REFERENCES

[1] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, 2000.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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 Journal of Photogrammetry and Remote Sensing*, vol. 107, pp. 22–37, 2015.

[8] J. L. Lisani and J.-M. Morel, "Detection of major changes in satellite images," in *International Conference on Image Processing*, vol. 1, pp. I–941, 2003.

[9] A. Robin, L. Moisan, and S. Le Hégarat-Mascle, "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.

[10] A. Flenner and G. Hewer, "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.

[11] N. Falco, M. D. Mura, F. Bovolo, *et al.*, "Change detection in VHR images based on morphological attribute profiles," *EEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 636–640, 2013.

[12] 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.

[13] N. Champion, D. Boldo, M. Pierrot-Deseilligny, and G. Stamon, "2D building change detection from high resolution satelliteimagery: A two-step hierarchical method based on 3D invariant primitives," *Pattern Recognition Letters*, vol. 31, no. 10, pp. 1138–1147, 2010.

[14] L. Matikainen, J. Hyyppä, 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.

[15] 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.

[16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[17] 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.

[18] 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 International Geoscience and Remote Sensing Symposium*, pp. 1281–1284, IEEE, 2014.

[19] A. Desolneux, L. Moisan, and J.-M. Morel, "Meaningful alignments," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 7–23, 2000.

[20] 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.

[21] 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.

[22] G.-S. Xia, J. Delon, and Y. Gousseau, "Accurate junction detection and characterization in natural images," *International journal of computer vision*, vol. 106, no. 1, pp. 31–56, 2014.

[23] 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.