Robust Change Detection in High Resolution Satellite Images with Geometric Distortions

Dongkwon Jin, Kyungsun Lim, and Chang-Su Kim
School of Electrical Engineering, Korea University, Seoul, Korea
E-mail: dongkwonjin@mcl.korea.ac.kr, kslim@mcl.korea.ac.kr, changsukim@korea.ac.kr

***Abstract*—A robust change detection algorithm for high resolution satellite images, which are not perfectly registered, is proposed in this work. To achieve this goal, a change detection technique for registered images and an image registration technique are employed in a cooperative way. Specifically, we use not only hand-crafted features but also change detection results to match keypoints extracted from two images. We then align the images using the matching pairs of keypoints. Finally, we obtain a change map from the aligned images. These steps of image registration and change detection are alternately iterated until the convergence. Experimental results demonstrate that proposed algorithm outperforms the conventional change detection technique significantly, when there are geometric distortions between temporal satellite images.**

I. Introduction

Remote sensing using satellite imagery, which captures meaningful changes on the surface [1], [2], has been applied to a variety of applications, including environmental monitoring, deforestation assessment, and agricultural expansion. With the advance of satellite imaging technology, it is possible to analyze variations in objects, such as moving vehicles and building construction sites, by comparing two images of the same area but taken at different time instances. However, there are various types of distortions when acquiring high resolution satellite images, which interfere with the detection of actual changes. In such satellite images, tall objects, such as high-rise buildings, look different depending on the photographic angles of cameras. Also, there are significant color variations due to seasonal and illumination conditions.

To distinguish actual changes from temporal distortions between satellite images, many change detection techniques have been proposed. An approach is to categorize regions in two images into predefined classes and then determine changed regions by comparing the classification results [3], [4]. Alternatively, the difference between satellite images can be analyzed to obtain a feature map for change detection [5]. Recently, convolutional neural networks (CNNs) have been used in a wide range of vision applications, including change detection [6]. CNN-based change detection techniques can identify actual changes reliably in spite of the presence of noise between images [7]–[11]. Especially, Lim *et al.* [7] proposed three CNNs in the encoder-decoder structure to yield change maps. Their algorithm does not require any pre-processing, such as ortho-rectification and object classification.

Meanwhile, registration errors (or geometric distortions) between satellite images may cause wrong results in change detection. Since most change detection algorithms require correctly registered images, image registration is an important pre-processing step for change detection [7]. It should geometrically align two images obtained in different imaging conditions, such as different times, various camera angles, and different sensors. There are two types of registration: area-based or feature-based methods [12]. Area-based methods, such as cross-correlation (CC) and mutual information (MI), exploit image intensities without any structural analysis [13], [14]. Feature-based methods utilize distinctive information, provided by local shapes and structure, and are more appropriate for satellite image registration [15]. Most feature-based methods focus on the design of robust feature descriptors invariant to imaging conditions.

Scale-invariant feature transform (SIFT) [16], [17] is a popular feature-based registration technique for extracting distinctive features, which are invariant to affine distortions and viewpoint changes. However, it may yield inaccurate matching results in case of large geometric variations or repeating structures in satellite images [18], [19]. To alleviate this problem, many attempts have been made to modify SIFT, such as SURF [20], BRIEF [21], and ORB [22]. But, several studies [19], [23], [24] demonstrated that the original SIFT is more robust than its variants. On the other hand, high-level features from CNN also can be used for image registration [25]–[27]. Also, He *et al.* [28] focused on registering satellite images with background variations, but their algorithm requires a large, annotated training dataset.

In this paper, we propose a robust algorithm to detect actual changes in temporal satellite images with geometric distortions. To reduce the impacts of geometric distortions, we perform image registration and change detection collaboratively. More specifically, we first extract keypoint descriptors from two images using SIFT. We then divide the images into blocks and find matching pairs of keypoints in corresponding blocks. To overcome matching errors of SIFT, we select reliable matching pairs only, by employing a change detection result. We then align the images using an affine transform, estimated from the matching pairs. Finally, we obtain a change detection map from the aligned images. These two steps of registration and change detection are repeatedly performed until the convergence.

This work has three main contributions:

- We propose a robust change detection algorithm for satellite images, which are not perfectly registered.

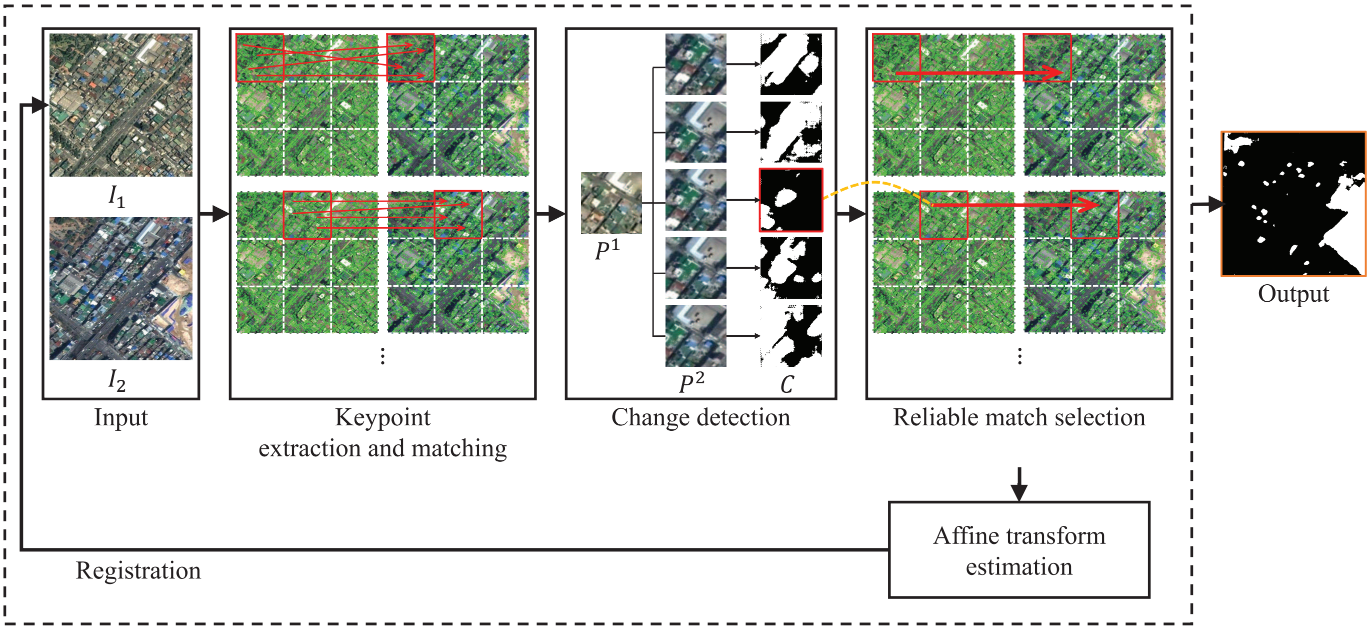


Fig. 1: An overview of the proposed algorithm.

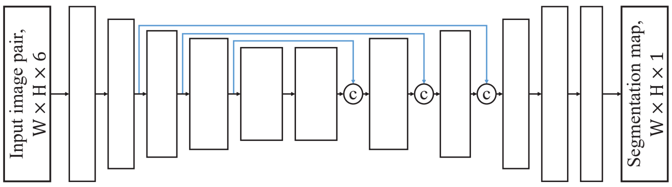


Fig. 2: The structure of the single long network (SLN) in [7].

- We develop a novel technique to overcome the limitation of SIFT-based registration using change detection results.
- The proposed algorithm yields better results than the existing change detection algorithm [7] for temporal satellite images with geometric distortions.

The rest of the paper is organized as follows. Section II describes the proposed algorithm, and Section III discusses experimental results. Finally, Section IV concludes this work.

II. Proposed Algorithm

This section describes the proposed change detection algorithm. The input is a pair of high resolution temporal satellite images, which are not perfectly registered, and the output is a binary map representing actual changes. Fig. 1 is an overview of the proposed algorithm. We first extract keypoint descriptors from the pair of images and find matching pairs of keypoints. Second, we select reliable matching pairs, by employing a change detection map. Third, we estimate the affine transform from the matching pairs and align the images. Finally, we update the change detection map using the aligned images. The image registration and change detection are alternately iterated until the convergence.

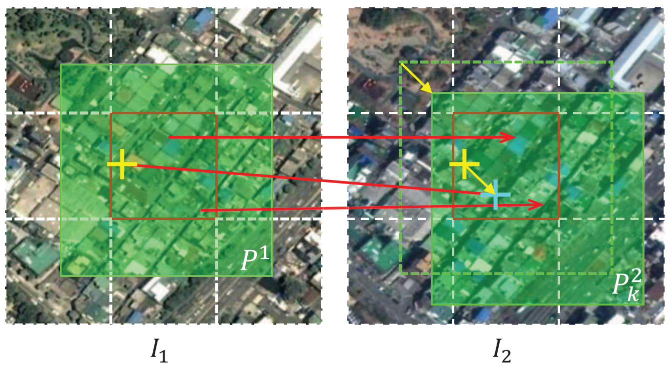


Fig. 3: Keypoint matching between corresponding blocks. Dashed white lines are block boundaries. Yellow and blue crosses denote SIFT keypoints. Red arrows depict matching results. The change detection between green patches is performed to select reliable matching pairs.

A. Preliminary

For the sake of completeness, let we briefly review the baseline method, Lim *et al.* [7], which is a CNN-based change detection algorithm for high resolution temporal satellite images. In [7], three CNNs in the encoder-decoder architecture were developed. In the encoder-decoder architecture, the encoder extracts a feature map from an image, and then the decoder converts the feature map into a segmentation map. Each pixel in the segmentation map represents the likelihood that any change occurs at the pixel location. By connecting the encoder and the decoder, the single short network (SSN), the single long network (SLN), and the double long network (DLN) were constructed. The baseline method yields segmentation maps

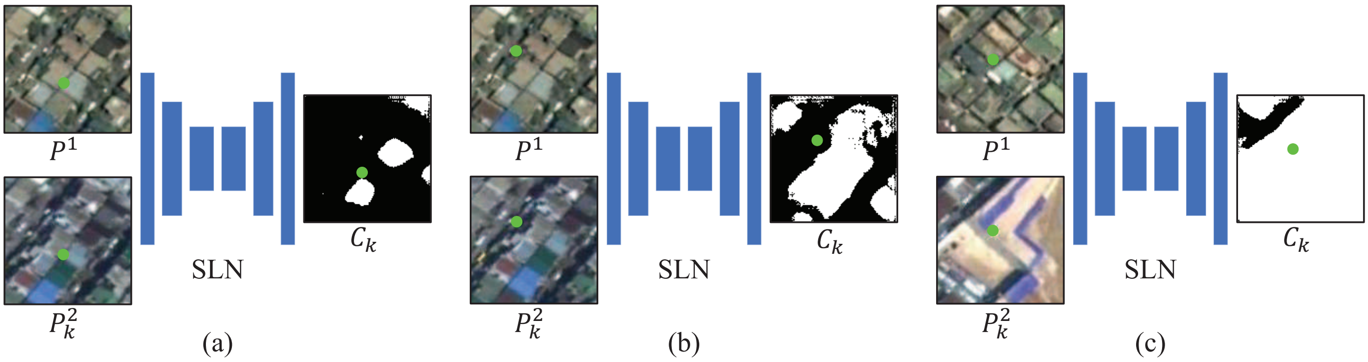


Fig. 4: Three cases of change detection results between image patches: (a) correct match between the descriptor pair (f_k^1, f_k^2), (b) incorrect match between (f_k^1, f_k^2), (c) useless match belonging to an actually changed region. In C_k, white and black pixels mean changed and unchanged pixels, respectively. Also, green points represent the locations of keypoints.

from the three CNNs, respectively, and obtains a final map using the average of the three maps. The ensemble of these networks provides excellent performance. In this work, we use SLN, whose network structure is illustrated in Fig. 2. Note that we modify the backbone from VGG16 [29] to ResNet50 [30].

B. Keypoint Extraction and Matching

To find the correspondence between reference and target images, we first extract keypoint descriptors from the two images using the SIFT algorithm [16], [17]. SIFT consists of a detector and a descriptor, which detects keypoints and describes local feature for each keypoint as a 128-dimensional vector, respectively. Let I_1 denote the reference image, and I_2 the target image. We divide each image into blocks of size $r \times r$. Then, we match keypoints between corresponding blocks in I_1 and I_2. For each descriptor in a block of I_1, we find the best matching descriptor in the corresponding block of I_2 with the smallest descriptor distance. From all such matching pairs, we keep only K matching pairs with the smallest distances, $\mathcal{M} = \{m_k \mid k = 1, \dots, K\}$. Here, m_k is the SIFT descriptor pair (f_k^1, f_k^2), where f_k^1 is the descriptor in I_1 and f_k^2 is that in I_2. Fig. 3 illustrates matching results, which are depicted by red arrows.

C. Reliable Match Selection Using Change Detection

Due to similar top appearance of buildings in satellite images, the descriptors located on buildings may be not distinctive and the descriptor matching may be unreliable [18], [19]. Moreover, if matching results are obtained within changed regions, they are useless. However, it is not easy to distinguish false matching results from true ones. Hence, we refine matching results using the change detection algorithm in [7].

Given an image pair, we obtain a binary change map C by thresholding the output of the change detection algorithm. Note that the change detection algorithm assumes that the two images are geometrically registered. If they are not registered, semantically unchanged pixels can be falsely detected as changed pixels. Thus, accurate registration tends to reduce the number of changed pixels in C.

Let P^1 be a 128×128 patch, which is centered at the block including f_{k}^1 in I_1, as shown in Fig. 3. We obtain the corresponding patch P_k^2 based on the keypoint position of the matching descriptor f_k^2 in I_2. From P^1 and P_k^2, we get the change map C_k. If $m_k = (f_k^1, f_k^2)$ is a correct match, only actual changes are detected as in Fig. 4(a). On the other hand, if m_k is not correct, many false positives are detected in Fig. 4(b). Therefore, we select only the most reliable match among $\mathcal{M}$, which minimizes the changed region. To compute the number of changed pixels, let s denote a score function that adds all pixel values in the change map. For each matching pair m_k in $\mathcal{M}$, we compute the score $s(C_k)$. Then, we select the pair that has the smallest score. Meanwhile, in some cases, keypoints are included in actually changed regions, as in Fig. 4(c). Since matching is impossible in a changed region, we exclude such matches.

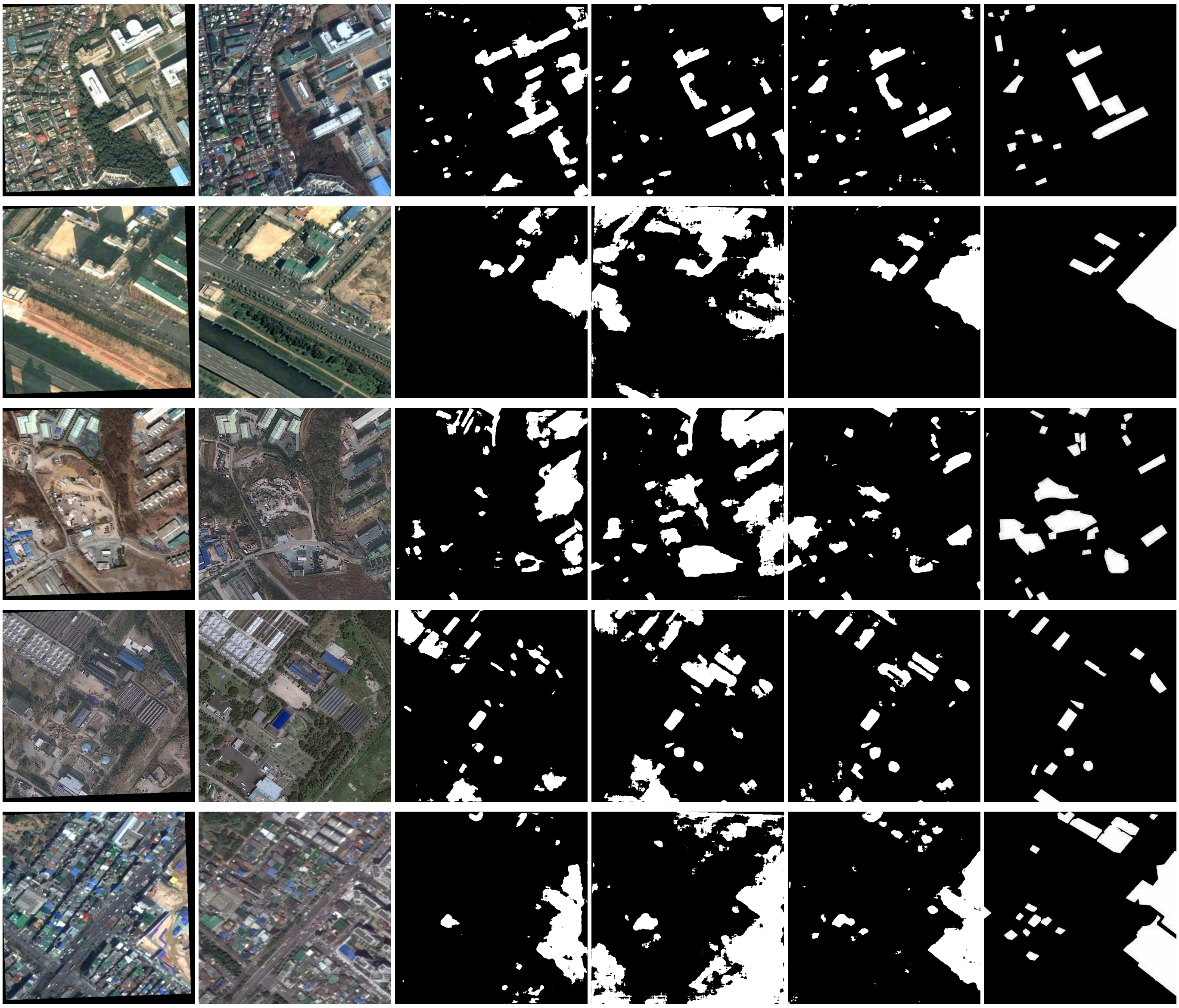
D. Affine Transform Estimation and Alignment

After selecting the most reliable matching pair of keypoints from matching pairs of corresponding blocks, we estimate an affine transform that maps the selected keypoints from I_1 to their matching keypoints in I_2. To this end, we employ the random sample consensus (RANSAC) [31]. RANSAC divides data into inliers that conform to the estimated model and outliers that do not. The pruned matching pairs, obtained by removing outliers, are used to estimate the affine transform based on the least squares method. Then, we align the reference image I_1 with the target image I_2, and find the change map between I_1 and I_2 using the algorithm in [7]. The change map, in turn, is used to register I_1 with I_2, as described in Section II-C. These image registration and change detection are iteratively performed until the convergence. After the convergence, we obtain the final change map from the registered images I_1 and I_2.

III. Experimental Results

A. Dataset

We use the change detection dataset in [7] to evaluate the performance of proposed algorithm. The dataset consists of



(a) Reference (b) Target (c) Baseline (d) SIFT + ratio test (e) Proposed (f) Ground-truth

Fig. 5: Examples of change detection results: Given a pair of temporal images with geometric distortions in (a) and (b), binary change maps are detected by the baseline [7] in (c), SIFT + ratio test in (d), and the proposed algorithm in (e), respectively. The ground-truth change map is in (f).

1000 bi-temporal satellite image pairs, which are geometrically registered, and the corresponding ground-truth binary maps, which represent changed regions. We divide the 1,000 image pairs into 900 training and 100 test pairs. To evaluate the robustness of the proposed algorithm, we generate image pairs with geometric distortions from the 100 test pairs. We warp each reference image using affine transforms, which mimic scale, translation, and rotation distortions. Specifically, in case A, the scale factor, translation, and rotation angle are set to 0.02, 0.02, $1°$, respectively. In case B, they are set to 0.04, 0.04, and $2°$.

B. Evaluation Metrics

For the performance assessment, we classify pixels in a change detection map using the corresponding ground-truth. True positive (TP) and true negative (TN) denote the numbers of pixels correctly predicted as changed and unchanged pixels, respectively. False alarm (FA) is the number of pixels predicted as changed but actually unchanged in the ground-truth, and miss alarm (MA) is the number of inverse cases. Then, precision and recall are defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FA}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{MA}}. \tag{1}$$

TABLE I: Performance comparison in terms of F1- and F2-scores. The highest score in each test is boldfaced.

Models	Registered		Case A		Case B	
	F1-score	F2-score	F1-score	F2-score	F1-score	F2-score
Lim *et al.* [7]	**67.41**	**72.30**	57.57	63.64	45.32	53.32
SIFT + ratio test	61.32	68.87	58.40	67.73	61.49	68.43
Proposed	65.45	71.46	**65.82**	**71.24**	**66.05**	**71.52**

Also, we compute two types of F-measure. The F-measure is defined as

$$F_\beta = (1+\beta^2) \cdot \frac{\text{Precision} \cdot \text{Recall}}{(\beta^2 \cdot \text{Precision}) + \text{Recall}} \tag{2}$$

where β determines the ratio of the influence of precision and recall. We use F1-score, which is the traditional F-measure with $\beta = 1$. It is the harmonic mean of precision and recall, so it is influenced by precision and recall with equal strength. We also use F2-score, which weighs recall more importantly than precision with $\beta = 2$. Therefore, F2-score gets a larger penalty by MAs than by FAs. In a surveillance system, FAs can be double-checked by personnel, whereas MAs do not have such opportunities. Thus, F2-score is more suitable as an assessment tool for change detection techniques.

C. Comparison Results

Table I compares the performance of the proposed algorithm with SLN in [7]. Note that [7] cannot be expected to provide high performance in cases A and B. On the contrary, the proposed algorithm yields robust performances in both cases. In the case of registered image pairs, the proposed algorithm provides comparable scores to [7].

Fig. 5 compares change detection results. Lim *et al.*'s algorithm [7] yields lots of false positives, since the image pair contain geometric distortions. In contrast, the proposed algorithm provides more reliable results with fewer errors.

D. Ablation Study

We carry out an ablation study to analyze the effectiveness of the reliable match selection using change detection. As a baseline, in the keypoint matching phase, for each keypoint in I_1, the two closest matching keypoints are found, whose distances are d_1 and d_2, where $d_1 \leq d_2$. Then, the ratio test [17] is performed, and the closest match is accepted if $d_1 < 0.75 \cdot d_2$. Then, for each pair of corresponding blocks, among the accepted pairs, the best pair with the smallest descriptor distance is selected as a reliable match pair. Table I also includes the performance of this baseline (SIFT + ratio test). We see that, as compared with the baseline, the proposed algorithm provides significantly higher F-scores. Fig. 5(d) shows change detection results of the baseline as well. Many pixels are falsely detected due to land cover changes and similarly looking buildings. Especially, in the second and fifth rows, there are significant land cover changes, which cause registration failures and yield lots of false positives. In contrast, Fig. 5(e) shows more robust change detection results.

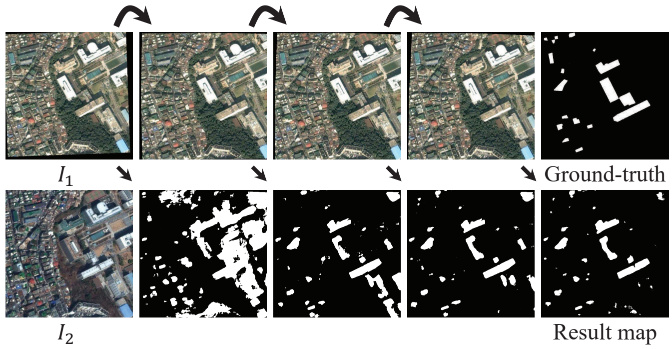


Fig. 6: Update of change detection maps according to the iterative registration of satellite images.

Fig. 6 shows how the image registration and the change detection are collaboratively performed in the proposed algorithm. We see that, as the iteration goes on, the reference image I_1 is more accurately aligned with the target image I_2 and a more accurate change detection map is obtained.

IV. Conclusions

We proposed a robust change detection algorithm for temporal satellite images with geometric distortions. We extracted keypoint descriptors and obtained matching pairs of descriptors between two images. By employing existing change detection and image registration techniques collaboratively, we selected reliable matching pairs to reduce the impacts of inaccurate matching, caused by land cover changes and similarly looking buildings. We then estimated an affine transform and aligned the two images. Finally, we obtained a change map from the aligned images. The image registration and the change detection were iteratively performed until the convergence. Experimental results showed that the proposed algorithm detects actual change regions robustly even if image pairs are not correctly registered.

V. Acknowledgment

This work was supported in part by the MSIT, Korea, under the ITRC support program (IITP-2019-2016-0-00464) supervised by the IITP, and in part by the National Research Foundation of Korea (NRF) through the Korea Government (MSIP) under Grant NRF2018R1A2B3003896.

References

[1] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989.

[2] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Review articledigital change detection methods in ecosystem monitoring: a review," *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1565–1596, 2004.

[3] J. Im, J. Jensen, and J. Tullis, "Object-based change detection using correlation image analysis and image segmentation," *Int. J. Remote Sens.*, vol. 29, no. 2, pp. 399–423, 2008.

[4] G. Conchedda, L. Durieux, and P. Mayaux, "An object-based method for mapping and change analysis in mangrove ecosystems," *ISPRI J. Photogramm. Remote Sens.*, vol. 63, no. 5, pp. 578–589, 2008.

[5] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and k-means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, 2009.

[6] K. Lim, W.-D. Jang, and C.-S. Kim, "Background subtraction using encoder-decoder structured convolutional neural network," in *Proc. IEEE AVSS*, 2017.

[7] K. Lim, D. Jin, and C.-S. Kim, "Change detection in high resolution satellite images using an ensemble of convolutional neural networks," in *Proc. IEEE APSIPA*, 2018.

[8] 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. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1845–1849, 2017.

[9] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional Siamese networks for change detection," in *Proc. IEEE ICIP*, 2018.

[10] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Urban change detection for multispectral earth observation using convolutional neural networks," in *Proc. IEEE IGARSS*, 2018.

[11] Y. Chu, G. Cao, and H. Hayat, "Change detection of remote sensing image based on deep neural networks," in *Proc. AIIE*, 2016.

[12] X. Dai and S. Khorram, "A feature-based image registration algorithm using improved chain-code representation combined with invariant moments," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2351–2362, 1999.

[13] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," in *Proc. MICCAI*, 1998.

[14] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imaging*, vol. 16, no. 2, pp. 187–198, 1997.

[15] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, 2003.

[16] D. G. Lowe *et al.*, "Object recognition from local scale-invariant features.," in *Proc. IEEE ICCV*, 1999.

[17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[18] M. Hasan, X. Jia, A. Robles-Kelly, J. Zhou, and M. R. Pickering, "Multi-spectral remote sensing image registration via spatial relationship analysis on SIFT keypoints," in *Proc. IEEE IGARSS*, 2010.

[19] X. Zhuo, T. Koch, F. Kurz, F. Fraundorfer, and P. Reinartz, "Automatic UAV image geo-registration by matching UAV images to georeferenced image data," *Remote Sens.*, vol. 9, no. 4, p. 376, 2017.

[20] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[21] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in *Proc. ECCV*, 2010.

[22] E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. IEEE ICCV*, 2011.

[23] L. Juan and L. Gwon, "A comparison of SIFT, PCA-SIFT and SURF," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 8, no. 3, pp. 169–176, 2007.

[24] J. Heinly, E. Dunn, and J.-M. Frahm, "Comparative evaluation of binary features," in *Proc. ECCV*, 2012.

[25] H. Altwaijry, E. Trulls, J. Hays, P. Fua, and S. Belongie, "Learning to match aerial images with deep attentive architectures," in *Proc. IEEE CVPR*, 2016.

[26] I. Melekhov, J. Kannala, and E. Rahtu, "Siamese network features for image matching," in *Proc. IEEE ICPR*, 2016.

[27] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proc. IEEE CVPR*, 2015.

[28] H. He, M. Chen, T. Chen, and D. Li, "Matching of remote sensing images with complex background variations via Siamese convolutional neural network," *Remote Sens.*, vol. 10, no. 2, p. 355, 2018.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, 2016.

[31] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.