

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/348640790>

# Structure Consistency-Based Graph for Unsupervised Change Detection With Homogeneous and Heterogeneous Remote Sensing Images

Article in IEEE Transactions on Geoscience and Remote Sensing · January 2021

DOI: 10.1109/TGRS.2021.3053571

CITATIONS

7

READS

426

5 authors, including:



**Yuli Sun**

University of Science and Technology of China

35 PUBLICATIONS 159 CITATIONS

[SEE PROFILE](#)



**Xiao Li**

National University of Defense Technology

31 PUBLICATIONS 100 CITATIONS

[SEE PROFILE](#)



**Xiang Tan**

National University of Defense Technology

7 PUBLICATIONS 25 CITATIONS

[SEE PROFILE](#)



**Gangyao Kuang**

National University of Defense Technology

273 PUBLICATIONS 3,142 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



SAR/PoSAR Image Processing [View project](#)

# Structure Consistency based Graph for Unsupervised Change Detection with Homogeneous and Heterogeneous Remote Sensing Images

Yuli Sun, Lin Lei, Xiao Li, Xiang Tan, and Gangyao Kuang, *Senior Member, IEEE*

**Abstract**—Change detection of remote sensing images is one of the important problems in earth observation, which has been extensively studied in recent years. However, with the development of remote sensing technology, the specific characteristics of remotely sensed images including sensor characteristics, resolutions, noises and distortions in imagery make the change detection more complex. In this paper, we propose a structure consistency based method for change detection, which detects changes by comparing the structures of two images, rather than comparing the pixel values of images. Because the image structure is imaging modality-invariant and not sensitive to noise, illumination and other interference factors, the proposed method can be applied to a variety of change detection scenarios and has strong robustness. Structural comparison is realized by constructing and mapping an improved nonlocal patch based graph (NLPG) to avoid the data leakage of two images. First, we demonstrate the effectiveness of the method in homogeneous and heterogeneous change detection, which shows that the proposed method can be used as a unified change detection framework. Second, we extend the method to the heterogeneous change detection with multichannel SAR image, which can provide a reference for future research as the heterogeneous change detection with multichannel SAR is rarely studied. Third, through the decomposition and in-depth analysis of NLPG, we modify the graph construction process, structure difference calculation, and the difference image fusion to make it more robust and accurate. Experiments on 6 scenarios 12 data sets demonstrate the effectiveness of the proposed method.

**Index Terms**—Unsupervised change detection, Heterogeneous data, Structure consistency, Nonlocal similarity, Graph.

## I. INTRODUCTION

### A. Background

The change detection (CD) of remote sensing (RS) images is a process of identifying changes of objects or phenomena that have occurred in the same geographical area at different times [1]. CD has been widely used in many practical applications, such as urban studies [2], human activity monitoring [3], resource and environment management [4], and nature disaster assessment [5].

Generally, CD algorithms can be divided into three categories according to whether the label information is used: unsupervised [6], semi-supervised [7], and supervised [8]. Although the semi-supervised and supervised approaches can

Manuscript received September 20, 2020; revised December 24, 2020; accepted January 16, 2021. (*Corresponding author: Lin Lei*)

Y. Sun, L. Lei, X. Li, X. Tan and G. Kuang are with College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China (e-mail: sunyuli@mail.ustc.edu.cn; alaleilin@163.com; lxcherishm@126.com; tanxiang@nudt.edu.cn; kuangyeats@hotmail.com).

theoretically provide better performance, they require the ground truth to provide the labeled samples. The traditional unsupervised CD methods can be divided into three processes: image preprocessing, difference image (DI) generation and analysis of DI [9]. In the preprocessing, geometric correction is usually accomplished by image-to-image registration to ensure that corresponding pixels in the multitemporal images refer to the same geographic location. In the second process, the two registered images are compared to generate the DI, which aims to increase the contrast between changed and unchanged areas. In the last process, the DI can be divided into the changed class and unchanged class to obtain the binary change map (CM). Among these processes, the DI generation plays an important role in the whole process of CD. First, a high-quality DI can clearly highlight the changed part, then the final CM can be obtained just by some threshold segmentation or clustering operations on the DI, which means that DI directly affects the performance of the final CM. Second, for some difficult problems or problems that must be treated with caution, the CD results are only used as a reference for experts, which means that in such case, the DI representing the probability of change may be more useful to experts than the binary CM. Third, for some unsupervised CD methods (such as the deep learning-based methods), a high-quality DI can be used to construct the pseudo training set or assist the training process. Therefore, in this paper, we will focus on how to generate a high-quality DI, and as for the CM, we just obtain it through some conventional threshold segmentation or clustering methods.

CD using multitemporal RS images is a complicated process, and it can be affected by many factors. The challenges for CD are summarized as follows.

1) *Inherent noise*: The noise inherent in the imaging process makes it difficult to obtain a high-quality DI, which may lead to many false and missed alarms in the CM [10]. This problem is especially serious in the CD task with synthetic aperture radar (SAR) images, where the coherent processing of SAR data makes images susceptible to speckles and the salt-and-pepper appearances because of the existence of many scatterers within the resolution cell [11]. Speckle in SAR images complicates the CD task by reducing the quality of DI and causing confusion between changes and unchanges. Although the denoising operation can be implemented in the preprocessing step, it may bring some irreparable errors. For example, denoising may over smooth the image and cause the loss of details, or it may also bring some undesired

artifacts such as the so-called staircasing artifact. Therefore, the CD method should be carefully designed to achieve good performance in terms of resisting noise and preserving details.

2) *High-resolution*: The high-resolution (HR) images significantly increase the image size, which causes an increase in processing time. More importantly, unlike the moderate- and coarse-resolution data, the HR images show some different characteristics, such as the increase of intensity variations within the same land-cover class in the HR SAR images [12], and the salt-and-pepper noise in the CM of very-high-resolution (VHR) optical images [13], [14]. In moderate- and coarse-resolution cases, the pixel-based methods are usually used, which treat the individual pixel as the basic unit of image analysis. However, in the HR images, each pixel is closely related to its spatial neighboring pixels. Therefore, the context information should be considered in the CD method.

3) *Heterogeneous CD*: Recently, with the rapid development of RS-related technologies, more and more image data representing the real information of the earth's surface can be obtained from different sensors at the same time, and the related heterogeneous RS image processing problem has been studied [15]-[18]. Heterogeneous CD [19], which is defined as a CD procedure with multitemporal images acquired from different satellite sensors, has also attracted a growing interest due to the great practical significance for the immediate evaluation and emergency disasters. In such scenarios (e.g., earthquake or flood), the pre-event SAR image is sometimes unavailable, whereas maybe only the post-event SAR image can be available due to the adverse atmospheric conditions [20]. However, since the heterogeneous images reflect different physical quantities of the object and show quite different statistical behaviors, it is difficult to calculate the difference between the heterogeneous images as the directly comparing is infeasible.

4) *Limited data sets*: The data sets for CD are relatively limited. This is because that constructing a ground-truth map that reflects real change information requires a high cost of manual operation in practice.

Different types of remotely sensed data require sensor-specific considerations, such as the SAR (single polarization, interferometric, polarimetric)/spectral sensor (multispectral, hyperspectral) with moderate/coarse/high resolution, and different applications call for different approaches, such as the land cover, flood mapping, building changing. Therefore, due to its complexity, there is no single method can deal with all types of change detection problems [10]. In this paper, we try to propose a CD framework that can be applied to a wider range, although we do not expect it to be applicable to all the CD problems.

### B. Motivations and contributions

Aiming at the aforementioned challenges, the proposed CD framework is expected to meet the following requirements: robust to noise, adapt to different resolution requirements, applicable to different data sets (homogeneous and heterogeneous CD), and without requiring any ground reference.

In our previous work, we have proposed a nonlocal patch-based graph (NLPG) based heterogeneous CD method [21],

which is based on the nonlocal self-similarity. It assumes that the heterogeneous images share the same structure information, and then it uses graph to measure the structure consistency between heterogeneous images. This NLPG brings three benefits: 1) since it uses the similarity/distance of patch to measure the changes, it can reduce the impact of noise; 2) it takes advantage of context information because it uses patches as processing units rather than individual pixels; 3) it constructs a comparable relationship between heterogeneous images and calculates the structure difference within the same image domain by mapping the graph of one domain to the other domain, which avoids heterogeneous data leakage. However, NLPG still has the following problems: 1) NLPG has only been tested in the heterogeneous CD task, but its effect in the homogeneous CD has not been reported and theoretically proven; 2) NLPG only considers the single-polarization SAR images but not polarimetric SAR (PolSAR) image, which limits its application; 3) NLPG individually constructs a graph for each patch, which is not efficient due to a lot of redundant computation; 4) the forward and backward DIs of NLPG are fused based on statistical distribution, which needs to estimate the noise level of images.

In this paper, we propose a structure consistency based method by the improved NLPG (INLPG) to solve the above problems. The main contributions are summarized as follows.

- A unified CD framework by using the structure consistency is proposed for both homogeneous and heterogeneous RS images. In particular, we show that structure consistency can establish connections between images for heterogeneous CD, and we also demonstrate that structure consistency is more robust compared with traditional patch difference/ratio operators in homogeneous CD. Therefore, the application scope of INLPG is greatly extended.
- INLPG is extended to the multichannel SAR (such as the PolSAR) data sets with two strategies: channel-based and covariance-matrix-based. Although there are studies on the CD of homogeneous PolSAR images, the heterogeneous CD with PolSAR images is rarely studied. This work can provide a reference for future researches.
- The graph construction process of NLPG is simplified and accelerated, which makes the INLPG to represent image structure more accurately and efficiently. We also modify the structure difference calculation to increase the divergence between the changed and unchanged classes. At the same time, the wavelet fusion technique is used to avoid the noise estimation. Therefore, based on these improvements, INLPG is more robust and has higher detection accuracy.
- INLPG is tested with experiments on 6 different scenarios 12 data sets, and the experimental results demonstrate the effectiveness of the INLPG (source code is made available at <https://github.com/yulisun/INLPG>).

### C. Outline

The overall structure of this study takes the form of six sections, including this introductory section. Section II reviews the related work. Section III describes the details of the proposed method. Section IV extends the application scope

of the method. Section V shows the experimental results. In the end, we provide our conclusion in Section VI.

## II. RELATED WORK

As the quality of DI directly affects the accuracy of CD, we will review some methods of generating DI for different types of sensors (domains) in this section.

### A. DI for multitemporal optical images

The mathematical operators used to compare multitemporal optical images mainly include image differencing, image regression, change vector analysis (CVA) and multivariate alteration detection (MAD). CVA [22] is an extension of the concept of image differencing, and it computes spectral change vectors to produce two types of change information: change magnitude representing the intensity of change, and change direction which provides information about the spectral behavior of the change vector. Bovolo et. al proposed a compressed change vector analysis (C2VA) [23], which compresses the information presented in spectral change vectors by computing the direction as the angular distance between the multispectral difference vector and a reference vector. Different from CVA, which is mainly applied to the original image feature space, some methods are performed in the transformed feature space, where the changed and unchanged areas show significantly different values. For example, the MAD [24] and iteratively reweighted (IR)-MAD [25] have been widely investigated, which are based on the established technique of canonical correlation analysis (CCA). In addition, to quantitatively measure the changes in the VHR optical images, a novel adaptive histogram trend (AHT) similarity approach has been proposed recently [13], [14]. However, these methods except AHT rarely take into account the effects of very different imaging conditions, such as seasons, illuminations, and phenological phenomena, which causes the same object to exhibit quite different spectral properties on multitemporal images. Therefore, direct comparison of spectral values (or simple transformations) will lead to many pseudo-changes.

### B. DI for multitemporal SAR images

In this case, the image difference operator becomes poorly effective due to the multiplicative noise model of the SAR image. Therefore, the ratio operator [26], log-ratio operator [27] and mean-ratio detector [28] are often used for its robustness with respect to the speckle noise. Furthermore, there are some other works proposed recently to generate a better DI based on the fusion of different methods, such as the wavelet fusion technique on different operators (on log-ratio and mean-ratio [29], on Gauss-log-ratio and log-ratio [30]), the saliency extraction guided log-ratio images [31], the shearlet fusion technique on saliency extraction and Gauss-log-ratio images [32]. In [33], different despeckling methods are performed on the input multitemporal images to test how can despeckling benefit to CD performance. Instead of separately applying the despeckling method to the multitemporal images or to the ratio/log-ratio DI directly, a nonlocal low-rank (NLR) model

that jointly uses the statistical characteristics of the logarithm transformed multitemporal images is proposed [34], which can avoid the information loss in the subtraction process and lead to a better DI. Note that a complete removal of speckle without destroying the fine structural details in the SAR images is not feasible. In addition, strong intensity variations associated with HR images is exacerbated by SAR speckle, which could also negatively affect the detection accuracy. Therefore, these methods usually faces such difficulties: how to generate a high-quality DI to keep tradeoff between robustness to speckle noise and strong intensity variation, and effectiveness of preserving the high-resolution geometrical information.

### C. DI for heterogeneous RS images

In order to obtain the DI for heterogeneous CD, some researchers use the similarity measures to detect the changes between heterogeneous images. The copula theory is used to model the dependence between unchanged areas [19], then the Kullback-Leibler (KL) distance is employed to measure the changes. Prendes et al. model the objects with a sliding window by local joint distributions, then use the manifold to measure the change indices [35], [36]. Five similarity measures are tested for CD of SAR and optical images in [37], including measures using probability: distance to independence, mutual information, cluster reward algorithm [38], and measures combining probability and radiometric value: Woods criterion [39], correlation ratio [40]. By assuming that the heterogeneous images with absence of change have some similar features, some special detection operators are designed, such as the sorted histogram distance (SHD) [41], the pixels pair (PP) method [42], [43], and the affinity matrices distance (AMD) [44], [45]. In addition, some regression-based methods are also employed to map the first image to the domain of the other image, such as the homogeneous pixel transformation (HPT) method [46] based on kernel regression and  $K$ -nearest neighbors technique. In [44], the AMD is first used to pick out the identified unchanged pixels as pseudo-training data, then four different regression methods are used to measure the changes. In [47], the image regression is implemented by using a learned patch similarity graph matrix (PSGM) with the self-similarity property. For these similarity- and regression-based methods, two points are very important: first, how to find the connections between heterogeneous images accurately; and second, how to design the change detection operator to fully utilize these connections. Therefore, this approaches usually face the difficulty that when the scene is complex or the noise in the image is very severe (especially the speckle noise of SAR images), these connections are no longer applicable, or the designed operators no longer fully represent the connections between heterogeneous images, resulting in a sharp decrease in CD performance.

### D. Deep learning based methods

Meanwhile, benefit from the achievement of deep neural network in the field of computer vision, the deep learning based CD methods have also been proposed to extract the high-level feature representation and explore the inner relationships

of multitemporal images. For example, the deep learning based CD for multitemporal spectral images: general end-to-end 2-D convolutional neural network (GETNET) [48], deep change vector analysis (DCVA) [49], semisupervised CD using graph convolutional network (GCNCD) [50]; for the multitemporal SAR images: deep nonsmooth nonnegative matrix factorization network (nsNMF) [51], PCANet [52], convolutional-wavelet neural network (CWNN) [53]; for the multitemporal heterogeneous images: the symmetric convolutional coupling network (SCCN) [54], conditional generative adversarial network (cGAN) [55], anomaly feature learning based deep sparse residual model (AFL-DSR) [56], the X-Net with two fully convolutional networks [45] and the adversarial cyclic encoders network (ACE-Net) [45]. Although the CD method based on deep learning performs well in detection accuracy, it still has two flaws: one is the time-consuming training process; the other is the construction of large size training set under the supervision mode or the complicated screening process for selecting the pseudo-training set under the unsupervision mode.

Since constructing a ground-truth map requires a high cost of manual operation in practice, we focus on the unsupervised CD method to reflect the change information. As reviewed above, although many algorithms have been proposed for CD, few methods can be well applied to both homogeneous and heterogeneous CD problems. In this paper, we will propose a structure-consistency-based CD method by using similarity relationship of image patches. Although image patch based CD methods have been proposed by researchers, such as the mean-ratio detector [28] and robust CVA [57], most of these methods directly compare the difference between the patch in one image and the corresponding patch in the other image, so they can not be used in heterogeneous CD due to the distinct image characteristics. Different from these methods, the proposed method characterizes each patch by its similarity relationship with other patches within the image, and then calculates the difference between patches of different images by using these relationships to measure the changes. It is not a comparison of pixel values of two image patches, but a comparison between structures (characterized by patch similarity relationships), which are represented by nonlocal patch based graphs. Therefore, it can establish the connection between images for heterogeneous CD. At the same time, because it compares the structures which are not sensitive to the interference factors such as illumination, season and noise, it can also be used in the homogeneous CD of both SAR and optical images.

### III. METHODOLOGY

We consider two coregistered images acquired by sensors (homogeneous or heterogeneous) before and after an event, which are denoted as  $\mathbf{X} = \{x(m, n, c); 1 \leq m \leq M, 1 \leq n \leq N, 1 \leq c \leq C_X\}$  in  $\mathcal{X}$  domain and  $\mathbf{Y} = \{y(m, n, c); 1 \leq m \leq M, 1 \leq n \leq N, 1 \leq c \leq C_Y\}$  in  $\mathcal{Y}$  domain, respectively. Here,  $M$  and  $N$  represent the length and width of the image, respectively;  $C_X$  and  $C_Y$

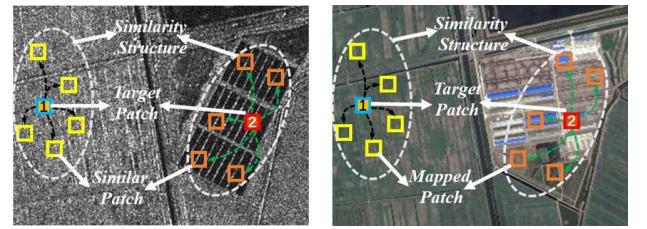


Fig. 1. Illustration of the structure consistency in heterogeneous images. The similarity structure of target patch 1 in SAR image and optical image are consistent, while the similarity structure of target patch 2 are not.

represent the number of channels/bands of two images, respectively. Because we expect that the CD method can be applied to different data sets, especially heterogeneous multitemporal images, we can only abandon those traditional mathematical operators in the homogeneous CD task, such as difference operator, ratio/log-ratio operator. Since the heterogeneous sensors take different imaging mechanisms, which makes it meaningless to directly compare pixel values of heterogeneous images, we need to find a relationship between the heterogeneous data and make them comparable, and then extract the changes.

The proposed method is based on the image self-similarity, that is each small patch in the image can always find some similar patches within the image. The relationship between this target patch and its similar patches can be regarded as the structure of this target patch, which is quite well preserved across the different types of imaging modality, then it can be used to build the relationship between heterogeneous images [21]. As shown in Fig. 1, for the unchanged targeted patch 1, its similarity structure in the SAR image, which is represented by the connections with its similar patches, is conformed by the optical image, and showing that the target patch is also similar to the mapped patches in the optical image. On the contrary, if the area represented by the target patch has changed in the event, this similarity structure is no longer conformed, showing that the target patch 2 are very different from the mapped patches in the optical image. Therefore, the change level can be measured by calculating the difference between the similarity structures of multitemporal images. Next, we briefly review the NLPG method [21], and then introduce how to improve and extend it.

#### A. NLPG

For each target patch,  $\mathbf{X}_{(m,n)} = \{x(m + \vartheta_m, n + \vartheta_n, c); \vartheta_m, \vartheta_n \in [p, -p], 1 \leq c \leq C_X\}$  centered on  $(m, n)$ , NLPG first constructs a weighted  $K$ -nearest graph  $G_{\mathbf{X}_{(m,n)}}^K = \{V_{\mathbf{X}_{(m,n)}}^K, E_{\mathbf{X}_{(m,n)}}^K, w_{\mathbf{X}}\}$  within a  $\omega_s \times \omega_s$  search window  $\mathbb{W}$  centered on this  $\mathbf{X}_{(m,n)}$  as

$$\begin{aligned} V_{\mathbf{X}_{(m,n)}}^K &= \{\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}; (i, j) \in \mathcal{N}_{\mathbf{X}_{(m,n)}}^K\}, |V_{\mathbf{X}_{(m,n)}}^K| = K + 1 \\ E_{\mathbf{X}_{(m,n)}}^K &= \{(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}) ; \mathbf{X}_{(i,j)} \in V_{\mathbf{X}_{(m,n)}}^K, (i, j) \neq (m, n)\} \\ w(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}) &= \exp \{-\lambda d(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)})\}, \\ &\forall (\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}) \in E_{\mathbf{X}_{(m,n)}}^K \end{aligned} \quad (1)$$

where  $\mathbf{X}_{(i,j)}$  represents the square patch within  $\mathbb{W}$  with search step size  $\Delta_s$ ,  $d(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)})$  represents a distance measure of two patches  $\mathbf{X}_{(m,n)}$  and  $\mathbf{X}_{(i,j)}$ ,  $\mathcal{N}_{\mathbf{X}_{(m,n)}}^K$  represents the center position set of the  $K$ -nearest neighbors ( $K$ -NN) of  $\mathbf{X}_{(m,n)}$  within the search window  $\mathbb{W}$  by sorting the distances  $d(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)})$ , and  $\lambda > 0$  is a bandwidth controlling parameter. Therefore, the graph  $G_{\mathbf{X}_{(m,n)}}^K$  can characterize the structure information of patch  $\mathbf{X}_{(m,n)}$ .

Then, NLPG maps the  $G_{\mathbf{X}_{(m,n)}}^K$  to the other image domain  $\mathcal{Y}$  to construct the mapped graph  $G_{\mathbf{Y}_{(m,n)}}^{map} = \{V_{\mathbf{Y}_{(m,n)}}, E_{\mathbf{Y}_{(m,n)}}, w_{\mathbf{Y}}\}$  as

$$\begin{aligned} V_{\mathbf{Y}_{(m,n)}}^{map} &= \left\{ \mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)}; (i, j) \in \mathcal{N}_{\mathbf{X}_{(m,n)}}^K \right\}, |V_{\mathbf{Y}_{(m,n)}}^{map}| = K + 1 \\ E_{\mathbf{Y}_{(m,n)}}^{map} &= \left\{ (\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)}) ; \mathbf{Y}_{(i,j)} \in V_{\mathbf{Y}_{(m,n)}}^{map}, (i, j) \neq (m, n) \right\} \\ w(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)}) &= \exp \left\{ -\lambda d(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)}) \right\}, \\ &\forall (\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)}) \in E_{\mathbf{Y}_{(m,n)}}^{map} \end{aligned} \quad (2)$$

NLPG calculates the forward structure difference  $f_{(m,n)}^{\mathbf{Y}}$  by comparing this mapped  $G_{\mathbf{Y}_{(m,n)}}^{map}$  with its own  $K$ -nearest graph  $G_{\mathbf{Y}_{(m,n)}}^K$  (similar to the construction process of  $G_{\mathbf{X}_{(m,n)}}^K$  in (1)) as

$$f_{(m,n)}^{\mathbf{Y}} = \frac{1}{K} \sum_{k=1}^K \left| d(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)^k}) - d(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i',j')^k}) \right| \quad (3)$$

where  $(i, j)^k \in \mathcal{N}_{\mathbf{Y}_{(m,n)}}^K$  represents the center position of the  $k$ -th nearest patch to  $\mathbf{Y}_{(m,n)}$  in  $V_{\mathbf{Y}_{(m,n)}}^K$ , and  $(i', j')^k \in \mathcal{N}_{\mathbf{X}_{(m,n)}}^K$  represents the center position of the  $k$ -th nearest patch to  $\mathbf{X}_{(m,n)}$  in  $V_{\mathbf{X}_{(m,n)}}^K$ .

Then,  $f_{(m,n)}^{\mathbf{Y}}$  is assigned to all the pixels in the patch  $\mathbf{Y}_{(m,n)}$ . After the structure difference is calculated on all the overlapping patches, the forward DI can be obtained as

$$\text{DI}^{fw}(i, j) = \frac{1}{|F_{(i,j)}^{\mathbf{Y}}|} \sum_{f_{(m,n)}^{\mathbf{Y}} \in F_{(i,j)}^{\mathbf{Y}}} f_{(m,n)}^{\mathbf{Y}} \quad (4)$$

where  $F_{(i,j)}^{\mathbf{Y}}$  is the set of structure differences that cover the pixel  $(i, j)$ . The backward DI<sup>bw</sup> can be obtained by using similar operation of mapping  $G_{\mathbf{Y}_{(m,n)}}^K$  to the image domain  $\mathcal{X}$ . Then, the final DI can be fused as

$$\text{DI}^{final} = (\text{DI}^{fw} + \text{DI}^{bw}) / 2 \quad (5)$$

The framework work of NLPG based DI generation is summarized in Table I, where the patch step size  $\Delta_p$  and search step size  $\Delta_s$  are used to accelerate the algorithm and improve the performance. From this framework, we can find that the calculation process of the structure difference is time-consuming. It operates with each target patch as a unit and individually creates a  $K$ -nearest graph for each patch, therefore, there are a lot of repeated distance calculations in this process.

### B. Improved NLPG

We still choose to exploit the structure information of the image in terms of patches, mainly for three reasons: 1) the patch contains the context, texture information; 2) the

TABLE I  
THE GENERATION STEPS OF NLPG BASED DI.

Algorithm 1. NLPG based DI	
<b>Input:</b>	Images of $\mathbf{X}$ and $\mathbf{Y}$ , parameters of $p$ , $\Delta_p$ , $\omega_s$ , $\Delta_s$ and $K$ .
<b>1. Calculation of the similarity structure difference</b>	$\text{for all the target patches } \mathbf{X}_{(m,n)}, \mathbf{Y}_{(m,n)} \text{ with step size } \Delta_p \text{ do}$ $\text{Compute } d(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}) \text{ and } d(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)}), \forall (i, j) \in \mathbb{W} \text{ with step size } \Delta_s.$ $\text{Compute } K\text{-nearest graph } G_{\mathbf{X}_{(m,n)}}^K \text{ and } G_{\mathbf{Y}_{(m,n)}}^K.$ $\text{Compute } f_{(m,n)}^{\mathbf{X}} \text{ and } f_{(m,n)}^{\mathbf{Y}} \text{ by graph mapping.}$ $\text{Add } f_{(m,n)}^{\mathbf{X}} \text{ to } F_{(i,j)}^{\mathbf{X}}, \text{ and add } f_{(m,n)}^{\mathbf{Y}} \text{ to } F_{(i,j)}^{\mathbf{Y}}.$ <b>end for</b>
<b>2. Computation of the forward and backward difference images</b>	$\text{for all the pixels } (i, j), 1 \leq i \leq M, 1 \leq j \leq N \text{ do}$ $\text{Compute } \text{DI}^{fw}(i, j) \text{ and } \text{DI}^{bw}(i, j).$ <b>end for</b>
<b>3. Fusion of the forward and backward difference images</b>	$\text{DI}^{final} = (\text{DI}^{fw} + \text{DI}^{bw}) / 2$

patchwise distance/similarity is more robust than the pixelwise distance/similarity under the noisy condition; 3) it can reduce the number of vertices when constructing the graph, so that the algorithm can be applied to large-scale data sets.

*1) Construct the K-NN Graph:* In the INLPG, we first divide the images  $\mathbf{X}$  into a number of overlapped square patches  $\mathbf{X}_{(m,n)}$  with the step size  $\Delta_p \in [1, 2p+1]$ , then we stack these patches into patch group matrix (PGM)  $\tilde{\mathbf{X}}$  with the size  $(2p+1) \times (2p+1) \times C_{\mathbf{X}} \times N_{\Delta_p}$ , where  $N_{\Delta_p} = \lceil M/\Delta_p \rceil \cdot \lceil N/\Delta_p \rceil$  with  $\lceil \cdot \rceil$  denoting the up rounding operation. Here, we denote the  $i$ -th patch as  $\tilde{\mathbf{X}}_i = \tilde{\mathbf{X}}(:, :, :, i) \in \mathbb{R}^{(2p+1) \times (2p+1) \times C_{\mathbf{X}}}$ ,  $i = 1, 2, \dots, N_{\Delta_p}$ . For the image  $\mathbf{Y}$ , we follow the same operation to obtain the PGM  $\tilde{\mathbf{Y}}$  with the size  $(2p+1) \times (2p+1) \times C_{\mathbf{Y}} \times N_{\Delta_p}$ . In this way, each patch pair in the original images  $\mathbf{X}$  and  $\mathbf{Y}$  still have the same position relationship in the new PGM  $\tilde{\mathbf{X}}$  and  $\tilde{\mathbf{Y}}$ .

As the graph model can efficiently capture crucial information and structure of an image [58], [59], we propose to construct a weighted graph  $\mathcal{G}$  to represent the geometric structure for the multitemporal images (in order to avoid the symbol confusion with the previous  $K$ -nearest graph  $G$  of NLPG, we use  $\mathcal{G}$  to represent the  $K$ -NN graph constructed in INLPG).

**Definition 1. (K-NN Graph).** Given a set of data points  $\mathbf{z} = \{z_1, z_2, \dots, z_n\}$  with  $z_i \in \mathbb{R}^d$ ,  $\mathcal{G} = \{V, E, w\}$  is a weighted directed graph, where  $V = Z$ , and  $(z_i, z_j) \in E$  if and only if  $\text{dist}(z_i, z_j)$  is among the  $K$  smallest elements of the set  $\{\text{dist}(z_i, z_l) | l = 1, \dots, i-1, i+1, \dots, n\}$ , and  $w(i, j) = \exp(-\lambda \text{dist}(z_i, z_j))$  for  $(z_i, z_j) \in E$ , where  $\text{dist}(\cdot, \cdot)$  is a defined distance metric such as Euclidean distance and Minkowski distance.

Given the PGM  $\tilde{\mathbf{X}}$ , we set each patch as a vertex and construct the  $K$ -NN graph  $\mathcal{G}_{\tilde{\mathbf{X}}} = \{V_{\tilde{\mathbf{X}}}, E_{\tilde{\mathbf{X}}}, w_{\tilde{\mathbf{X}}}\}$  as

$$\begin{aligned} V_{\tilde{\mathbf{X}}} &= \left\{ \tilde{\mathbf{X}}_i; i = 1, 2, \dots, N_{\Delta_p} \right\}, |V_{\tilde{\mathbf{X}}}| = N_{\Delta_p} \\ E_{\tilde{\mathbf{X}}} &= \left\{ (\tilde{\mathbf{X}}_i, \tilde{\mathbf{X}}_j); i = 1, 2, \dots, N_{\Delta_p}, j \in \mathcal{N}_{\tilde{\mathbf{X}}_i}^K \right\} \\ w_{\tilde{\mathbf{X}}}(i, j) &= \exp \left\{ -\lambda \text{dist}(\tilde{\mathbf{X}}_i, \tilde{\mathbf{X}}_j) \right\}, \forall (\tilde{\mathbf{X}}_i, \tilde{\mathbf{X}}_j) \in E_{\tilde{\mathbf{X}}} \end{aligned} \quad (6)$$

where  $\mathcal{N}_{\tilde{\mathbf{X}}_i}^K$  represents the position set for the  $K$ -NN of  $\tilde{\mathbf{X}}_i$  by sorting the distance  $dist(\tilde{\mathbf{X}}_i, \tilde{\mathbf{X}}_j)$ ,  $j = 1, \dots, i-1, i+1, \dots, N_{\Delta_p}$  and taking our the  $K$  smallest elements.

By comparing the graph construction process of NLPG and INLPG, we can find that their differences are: 1) INLPG removes the limitation of local search window  $\mathbb{W}$  in NLPG and searches for similar patches in the whole image, which can establish a wider connection for each patch. Thus, it is able to find more similar nearest neighbors (NNs) for each vertex and also to avoid over-concentration of NNs. Therefore, the graph  $\mathcal{G}$  of INLPG is more robust. 2) NLPG individually creates a graph  $G$  for each target patch as illustrated in (1), which contains  $K+1$  vertices and  $K$  edges. Then, NLPG constructs a total of  $N_{\Delta_p}$  graphs for each image. However, INLPG is not to construct the graph one by one for each target patch, but to construct the graph  $\mathcal{G}_{\tilde{\mathbf{X}}}$  for the whole image  $\mathbf{X}$  as illustrated in (6), which contains  $N_{\Delta_p}$  vertices and  $K \times N_{\Delta_p}$  directed edges. In this way, INLPG can avoid the repeated distance calculation in NLPG and use some efficient graph construction methods.

For each individual graph of NLPG in (1), the time complexity of calculating distance vector requires  $\mathcal{O}(C_{\mathbf{X}}(2p+1)^2 N_{\Delta_s})$ , where  $N_{\Delta_s}$  is the number of candidate NNs in the search window, and that of sorting the distance vector requires  $\mathcal{O}(N_{\Delta_s} \log N_{\Delta_s})$  by using some accelerated sorting algorithms, such as the Block sort or Tree sort. Therefore, the time complexity of constructing graphs in NLPG is  $\mathcal{O}((C_{\mathbf{X}}(2p+1)^2 + \log N_{\Delta_s}) N_{\Delta_s} N_{\Delta_p})$ . For the time complexity of INLPG, calculating the complete distance matrix in (6) requires  $\mathcal{O}(C_{\mathbf{X}}(2p+1)^2 N_{\Delta_p}^2 / 2)$  and sorting the distance matrix by column requires  $\mathcal{O}(N_{\Delta_p}^2 \log N_{\Delta_p})$ . Therefore, the time complexity of constructing  $K$ -NN graph in INLPG is  $\mathcal{O}((C_{\mathbf{X}}(2p+1)^2 / 2 + \log N_{\Delta_p}) N_{\Delta_p}^2)$ , which is larger than NLPG as  $N_{\Delta_s}$  is usually a fraction of  $N_{\Delta_p}$ . However, due to the large amount of redundancy in the distance matrix, this  $K$ -NN graph construction process can be accelerated by using some efficient graph construction methods. For example, the  $K$ -NN graph with NN-Descent uses the principle: “a neighbor of a neighbor is also likely to be a neighbor” [60], [61]; Locality Sensitive Hashing (LSH) uses families of functions that hash signatures of similar objects to the same bucket with high probability [62]; and L2Knng solves the exact  $K$ -NN graph construction problem by pruning much of the similarity search space [63]. Specifically, for the kgraph [61], its empirical cost is around  $\mathcal{O}(n^{1.14})$  for constructing the  $K$ -NN graph with a set of  $n$  objects [60]. By using these fast graph construction methods, INLPG can be greatly accelerated.

2) *Measure the change level:* In the directed graph  $\mathcal{G}_{\tilde{\mathbf{X}}} = \{V_{\tilde{\mathbf{X}}}, E_{\tilde{\mathbf{X}}}, w_{\tilde{\mathbf{X}}}\}$ , each patch in  $\tilde{\mathbf{X}}$  becomes a vertex, and each vertex is connected with  $K$  NNs by a set of directed edges, and the associate weights  $w$  involving the similarities between each vertex and its NNs. In this way, the structure information of the image can be characterized by this  $K$ -NN graph. Directly comparing the graph  $\mathcal{G}_{\tilde{\mathbf{X}}} = \{V_{\tilde{\mathbf{X}}}, E_{\tilde{\mathbf{X}}}, w_{\tilde{\mathbf{X}}}\}$  and  $\mathcal{G}_{\tilde{\mathbf{Y}}} = \{V_{\tilde{\mathbf{Y}}}, E_{\tilde{\mathbf{Y}}}, w_{\tilde{\mathbf{Y}}}\}$  (similar to the construction process of  $\mathcal{G}_{\tilde{\mathbf{X}}}$  in (6)) is difficult, which is mainly because they come

from different domains. To avoid the leakage of heterogeneous data, we first map the  $\mathcal{G}_{\tilde{\mathbf{X}}}$  to the other image domain  $\mathcal{Y}$ , then we have the mapped graph  $\mathcal{G}_{\tilde{\mathbf{Y}}}^{map} = \{V_{\tilde{\mathbf{Y}}}^{map}, E_{\tilde{\mathbf{Y}}}^{map}, w_{\tilde{\mathbf{Y}}}^{map}\}$  as

$$\begin{aligned} V_{\tilde{\mathbf{Y}}}^{map} &= \left\{ \tilde{\mathbf{Y}}_i; i = 1, 2, \dots, N_{\Delta_p} \right\}, |V_{\tilde{\mathbf{Y}}}^{map}| = N_{\Delta_p} \\ E_{\tilde{\mathbf{Y}}}^{map} &= \left\{ (\tilde{\mathbf{Y}}_i, \tilde{\mathbf{Y}}_j); i = 1, 2, \dots, N_{\Delta_p}, j \in \mathcal{N}_{\tilde{\mathbf{X}}_i}^K \right\} \\ w_{\tilde{\mathbf{Y}}}(i, j) &= \exp \left\{ -\lambda dist(\tilde{\mathbf{Y}}_i, \tilde{\mathbf{Y}}_j) \right\}, \forall (\tilde{\mathbf{Y}}_i, \tilde{\mathbf{Y}}_j) \in E_{\tilde{\mathbf{Y}}}^{map} \end{aligned} \quad (7)$$

We can see that the mapped graph  $\mathcal{G}_{\tilde{\mathbf{Y}}}^{map}$  is constructed by using the edges  $E_{\tilde{\mathbf{X}}}$  of  $\mathcal{G}_{\tilde{\mathbf{X}}}$ , and calculating the weights in the  $\mathcal{Y}$  domain. As the  $\mathcal{G}_{\tilde{\mathbf{Y}}}^{map}$  and  $\mathcal{G}_{\tilde{\mathbf{Y}}}$  are in the same domain, then we can calculate the structure difference by using the distance criterion or similarity criterion as

$$f_i^{\tilde{\mathbf{Y}}} = \frac{1}{K} \left( \sum_{j' \in \mathcal{N}_{\tilde{\mathbf{X}}_i}^K} dist(\tilde{\mathbf{Y}}_i, \tilde{\mathbf{Y}}_{j'}) - \sum_{j \in \mathcal{N}_{\tilde{\mathbf{Y}}_i}^K} dist(\tilde{\mathbf{Y}}_i, \tilde{\mathbf{Y}}_j) \right) \quad (8)$$

$$\begin{aligned} f_i^{\tilde{\mathbf{Y}}} &= \frac{1}{K} \left( \sum_{j \in \mathcal{N}_{\tilde{\mathbf{Y}}_i}^K} \exp \left\{ -\lambda dist(\tilde{\mathbf{Y}}_i, \tilde{\mathbf{Y}}_j) \right\} \right. \\ &\quad \left. - \sum_{j' \in \mathcal{N}_{\tilde{\mathbf{X}}_i}^K} \exp \left\{ -\lambda dist(\tilde{\mathbf{Y}}_i, \tilde{\mathbf{Y}}_{j'}) \right\} \right) \end{aligned} \quad (9)$$

Intuitively, the structure difference is measured by how much is the difference between the  $K$ -NN position sets of  $\mathcal{N}_{\tilde{\mathbf{X}}_i}^K$  and  $\mathcal{N}_{\tilde{\mathbf{Y}}_i}^K$ . If the region represented by the patch  $\tilde{\mathbf{Y}}_i$  does not change, the patch  $\tilde{\mathbf{Y}}_{j'}, j' \in \mathcal{N}_{\tilde{\mathbf{X}}_i}^K$  will be similar as  $\tilde{\mathbf{Y}}_i$  with a high probability, so the difference value  $f_i^{\tilde{\mathbf{Y}}}$  will be very small; on the contrary, if this region changes, the patch  $\tilde{\mathbf{Y}}_{j'}, j' \in \mathcal{N}_{\tilde{\mathbf{X}}_i}^K$  will be different from  $\tilde{\mathbf{Y}}_i$  with a high probability, and then resulting in a large  $f_i^{\tilde{\mathbf{Y}}}$ .

Meanwhile, by comparing the distance criterions (8) of INLPG and (3) of NLPG, we can find that the INLPG calculates the difference between the mean values, while NLPG calculates each cumulative difference. As a result,  $f_i^{\tilde{\mathbf{Y}}} \leq f_{(m,n)}^{\mathbf{Y}}$ , with  $\tilde{\mathbf{Y}}_i$  and  $\mathbf{Y}_{(m,n)}$  representing the same patch. The mean difference of (8) is more robust than the cumulative difference of (3) in the noisy conditions. For the changed patch  $\tilde{\mathbf{Y}}_i$ , the  $dist(\tilde{\mathbf{Y}}_i, \tilde{\mathbf{Y}}_{j'})$ ,  $j' \in \mathcal{N}_{\tilde{\mathbf{X}}_i}^K$  may be very large. Therefore, we basically have  $dist(\tilde{\mathbf{Y}}_i, \tilde{\mathbf{Y}}_j) \leq dist(\tilde{\mathbf{Y}}_i, \tilde{\mathbf{Y}}_{j'})$  for all  $j \in \mathcal{N}_{\tilde{\mathbf{Y}}_i}^K$ ,  $j' \in \mathcal{N}_{\tilde{\mathbf{X}}_i}^K$ , and then we have  $f_i^{\tilde{\mathbf{Y}}} = f_{(m,n)}^{\mathbf{Y}}$ . On the contrary, for the unchanged patch  $\tilde{\mathbf{Y}}_i$ , although the  $dist(\tilde{\mathbf{Y}}_i, \tilde{\mathbf{Y}}_{j'})$ ,  $j' \in \mathcal{N}_{\tilde{\mathbf{X}}_i}^K$  may be very small,  $f_i^{\tilde{\mathbf{Y}}}$  always be smaller than or equal to  $f_{(m,n)}^{\mathbf{Y}}$ . Specifically, we have  $f_i^{\tilde{\mathbf{Y}}} = 0$  if  $\mathcal{N}_{\tilde{\mathbf{Y}}_i}^K = \mathcal{N}_{\tilde{\mathbf{X}}_i}^K$ , and  $f_{(m,n)}^{\mathbf{Y}} = 0$  if and only if  $\mathcal{N}_{\tilde{\mathbf{Y}}_i}^K = \mathcal{N}_{\tilde{\mathbf{X}}_i}^K$  and the order of the elements in them should be the same (that is  $(i, j)^k = (i', j')^k$  for  $k = 1, 2, \dots, K$  in (3)). However, due to the presence of noise, this condition of  $f_i^{\tilde{\mathbf{Y}}} = f_{(m,n)}^{\mathbf{Y}}$  is too strict when  $\tilde{\mathbf{Y}}_i$  has not changed. From this, we can find that the mean difference is more able to increase the divergence between the changed and the unchanged classes.

After the structure difference  $f_i^{\tilde{Y}}$  is calculated for all the patches  $i = 1, 2, \dots, N_{\Delta_p}$ , then for each pixel  $(s, t)$ ,  $1 \leq s \leq M$ ,  $1 \leq t \leq N$  in the image, we denote  $F_{(s,t)}^{\tilde{Y}}$  as the set of structure difference covering the pixel  $(s, t)$ . The forward DI can be calculated as

$$\text{DI}^{fw}(s, t) = \frac{1}{|F_{(s,t)}^{\tilde{Y}}|} \sum_{f_i^{\tilde{Y}} \in F_{(s,t)}^{\tilde{Y}}} f_i^{\tilde{Y}} \quad (10)$$

Similarly, the backward  $\text{DI}^{bw}$  can be calculated by mapping  $\mathcal{G}_{\tilde{Y}}$  to the image domain  $\mathcal{X}$ , and then comparing the mapped  $\mathcal{G}_{\tilde{X}}^{\text{map}}$  with  $\mathcal{G}_{\tilde{X}}$ .

**3) Fuse the DI:** In NLPG, the forward DI and backward DI are fused by arithmetic average as (5). Since the values of  $\text{DI}^{fw}$  and  $\text{DI}^{bw}$  may not be at the same level due to different noise levels and different distance criteria of the heterogeneous data, it needs to balance the distance measurements  $\text{dist}(\tilde{Y}_i, \tilde{Y}_{j'})$  and  $\text{dist}(\tilde{X}_i, \tilde{X}_{j'})$  to keep the  $\text{DI}^{fw}$  and  $\text{DI}^{bw}$  at the same level. NLPG adds normalized parameters on the distance criteria to make their expectations equal in the unchanged area. Although this method is feasible and proved to be effective, it has two disadvantages: 1) it is derived based on a specific distribution model, such as additive white Gaussian noise (AWGN) model and multiplicative speckle noise model of Gamma distribution. When the noise in the image does not belong to these two types, it may cause performance degradation; 2) it needs to estimate the noise level of the image, such as the variance of the AWGN and the equivalent number of looks (ENL) of speckle noise model, which is not easy and introduces extra computation. In order to fuse the image and avoid the above problems, we use a simple wavelet fusion method in INLPG.

Due to its low computational complexity and the ability to preserve image details, discrete wavelet transform (DWT) has been widely used in pixel level image fusion [64], [65], and it has also been extended to the CD task [29], [30]. The main step of DWT fusion method in INLPG can be described as follows.

**Step 1.** Compute the DWT of forward and backward DIs, and obtain the multiresolution decomposition of each DI as:  $D_{LL}^{fw}$  and  $D_{LL}^{bw}$  representing the low-frequency wavelet coefficients of  $\text{DI}^{fw}$  and  $\text{DI}^{bw}$ , respectively;  $D_{\varepsilon}^{fw}$  and  $D_{\varepsilon}^{bw}$  ( $\varepsilon \in \{LH, HL, HH\}$ ) standing for three high-frequency wavelet coefficients of  $\text{DI}^{fw}$  and  $\text{DI}^{bw}$ , respectively.

**Step 2.** Fuse corresponding coefficients as follows:

$$D_{LL}^{fuse} = (D_{LL}^{fw} + D_{LL}^{bw}) / 2$$

$$D_{\varepsilon}^{fuse}(i, j) = \begin{cases} D_{\varepsilon}^{fw}(i, j), & E_{\varepsilon}^{fw}(i, j) \leq E_{\varepsilon}^{bw}(i, j) \\ D_{\varepsilon}^{bw}(i, j), & E_{\varepsilon}^{fw}(i, j) > E_{\varepsilon}^{bw}(i, j) \end{cases} \quad (11)$$

where  $E_{\varepsilon}(i, j)$  is the Gaussian weighted local area energy coefficient defined as

$$E_{\varepsilon}(i, j) = \sum_{h=-p}^p \sum_{t=-p}^p g_{h,t} [D_{\varepsilon}(i+h, j+t)]^2 \quad (12)$$

where  $g_{h,t}$  is the element of the rotationally symmetric Gaussian low-pass filter  $g$  of size  $(2p+1) \times (2p+1)$  with standard deviation  $\sigma = 1$ .

TABLE II  
THE GENERATION STEPS OF INLPG BASED DI.

Algorithm 1. INLPG based DI

**Input:** Images of  $\mathbf{X}$  and  $\mathbf{Y}$ , parameters of  $p$ ,  $\Delta_p$ , and  $K$ .

**1.** Construction of the  $K$ -NN graph

Construct the PGM  $\tilde{\mathbf{X}}$  and  $\tilde{\mathbf{Y}}$  with step size  $\Delta_p$ .

Construct the  $K$ -NN graphs  $\mathcal{G}_{\tilde{\mathbf{X}}}$  and  $\mathcal{G}_{\tilde{\mathbf{Y}}}$ .

**2.** Calculation of the nonlocal patch similarity structure difference

for all the target patches  $\tilde{\mathbf{X}}_i$ ,  $\tilde{\mathbf{Y}}_i$ ,  $i = 1, 2, \dots, N_{\Delta_p}$  do

Compute  $f_i^{\tilde{\mathbf{X}}}$  and  $f_i^{\tilde{\mathbf{Y}}}$  by graph mapping with the mean difference.

Add  $f_i^{\tilde{\mathbf{X}}}$  to  $F_{(s,t)}^{\tilde{\mathbf{X}}}$ , and add  $f_i^{\tilde{\mathbf{Y}}}$  to  $F_{(s,t)}^{\tilde{\mathbf{Y}}}$ .

end for

**3.** Computation of the forward and backward difference images

for all the pixels  $(s, t)$ ,  $1 \leq s \leq M$ ,  $1 \leq t \leq N$  do

Compute  $\text{DI}^{fw}(i, j)$  and  $\text{DI}^{bw}(i, j)$ .

end for

**4.** Fusion of the forward and backward difference images

Compute the DWT of  $\text{DI}^{fw}$  and  $\text{DI}^{bw}$ , respectively.

Fuse corresponding coefficients using the fusion rule (11).

Compute the fused image  $\text{DI}^{final}$  by applying the inverse DWT.

**Step 3.** The fused image  $\text{DI}^{final}$  can be obtained by inverse DWT of low-frequency  $D_{LL}^{fuse}$  and three high-frequency  $D_{\varepsilon}^{fuse}$  ( $\varepsilon \in \{LH, HL, HH\}$ ).

In this DWT fusion process, the low-frequency and high-frequency wavelet coefficients are fused separately, which can make full use of the forward and backward DIs.

**4) Summarize the algorithm:** The framework of the INLPG based DI generation is summarized in Table II, which reduces the parameters needed in the NLPG, such as the search windows size, the search step size, the estimated noise level parameters (the variance and ENL). As mentioned above, although the basic ideas of the NLPG (Table I) and INLPG (Table II) are the similar, their implementation process is different.

First, their graph construction processes are different. INLPG searches the NNs in the whole image instead of a local search window, which can establish a broader and more accurate connection for each patch to obtain a more robust graph; INLPG constructs the  $K$ -NN graph for the image as a whole, rather than a separate graph for each patch individually, which can avoid the repeated distance calculation and can be accelerated by some efficient graph construction methods, such as the kgraph [60], [61]. Second, their calculations of structure difference are different. INLPG uses the mean difference instead of the cumulative difference, which is more robust and can increase the divergence between the changed and the unchanged classes. Third, their methods of DIs fusion are different. INLPG uses the DWT fusion method instead of the noise estimation based arithmetic average, which avoids the complex noise estimation process, simplifies the fusion process, and improves the fusion effect.

#### IV. EXTENSION OF INLPG

From the framework of the INLPG based DI generation, we can find that the proposed method exploits the inherent structure property of images and appeals quite imaging modality-invariant. So, it can be used for heterogeneous CD. At the same time, as INLPG compares the structures which are not sensitive to the illumination, season and noise, it can also

be used for homogeneous CD. Moreover, it can be used for images acquired by optical, single channel SAR, multichannel SAR and other sensors, as long as we can construct the corresponding nonlocal patch-based graphs representing the structure of these images. In this section, we first propose some patch distance calculations for INLPG and show how to extend it to the CD problem with multichannel SAR images, and then we demonstrate that structure difference is more robust compared with traditional patch difference/ratio operators in homogeneous CD.

#### A. Patch distance calculation

It can be found from (6) that the construction of  $K$ -NN graph largely depends on the selection of patch distance. For two patches, e.g.  $\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2 \in \mathbb{R}^{(2p+1) \times (2p+1) \times C_{\mathbf{X}}}$ , we vectorize them and denote each element as  $x_1(i)$  and  $x_2(i)$  with  $i = 1, \dots, (2p+1)^2 C_{\mathbf{X}}$ , respectively. In the NLPG, we have given three distance criteria as

$$dist_{spe}^{(1)}(\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2) = \frac{\sum_{i=1}^{(2p+1)^2 C_{\mathbf{X}}} (x_1(i) - x_2(i))^2}{(2p+1)^2 C_{\mathbf{X}}} \quad (13)$$

$$dist_{SAR}^{(1)}(\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2) = \frac{\sum_{i=1}^{(2p+1)^2 C_{\mathbf{X}}} \log\left(\frac{(x_1(i)+x_2(i))^2}{4x_1(i)x_2(i)}\right)}{(2p+1)^2 C_{\mathbf{X}}} \quad (14)$$

$$dist_{SAR}^{(2)}(\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2) = \frac{\sum_{i=1}^{(2p+1)^2 C_{\mathbf{X}}} (\log x_1(i) - \log x_2(i))^2}{(2p+1)^2 C_{\mathbf{X}}} \quad (15)$$

The  $dist_{spe}^{(1)}$  can be applied to the optical image with the AWGN, the  $dist_{SAR}^{(1)}$  and  $dist_{SAR}^{(2)}$  are suitable for the SAR image with the multiplicative Gamma distribution model. These distance formulas are derived based on different criteria [66], such as Bayesian joint likelihood, generalized likelihood ratio and variance stabilization criterion. In [21], these patch distances have been studied in depth. Here, we expand the patch distance in two aspects.

1) *Robust distance criterion for spectral images:* For the real-world spectral images, there usually exists a combination of several different types of noise, e.g., Gaussian noise and impulse noise. In some circumstances, the impulse noise is dominant or there are many bad points in the image (such as missing data and stripe corruption in the hyperspectral images [67]), so a robust distance criterion is needed. In this case, the following criterion is more suitable than the squared Euclidean distance  $dist_{spe}^{(1)}$

$$dist_{spe}^{(2)}(\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2) = \frac{\sum_{i=1}^{(2p+1)^2 C_{\mathbf{X}}} |x_1(i) - x_2(i)|}{(2p+1)^2 C_{\mathbf{X}}} \quad (16)$$

2) *Extending to the multichannel SAR:* For the  $D$ -channel SAR image  $\mathbf{X}$  (e.g.,  $D = 2$  for dual PolSAR or interferometric SAR (InSAR), and  $D = 3$  for fully PolSAR), we have two strategies for calculating the patch distance: channel-based and covariance matrix-based. The channel-based strategy is to

treat the  $D$ -channel SAR image as the ordinary multi-channel images with  $C_{\mathbf{X}} = D$ . Therefore, the  $dist_{SAR}^{(1)}$  and  $dist_{SAR}^{(2)}$  can be directly used for the multichannel SAR.

For the covariance-matrix-based strategy, these channels can be gathered together at each pixel by forming a complex covariance matrix. Denote the  $L$  vectors of  $D$  complex values recorded for a pixel  $i$  as  $\vec{k}_i^{(1)}, \dots, \vec{k}_i^{(L)}$ , then the  $D \times D$  empirical covariance matrix  $\mathbf{C}_i$  is given by

$$\mathbf{C}_i = \frac{1}{L} \sum_{t=1}^L \vec{k}_i^{(t)} \vec{k}_i^{(t)\dagger} \quad (17)$$

where  $\dagger$  indicates the hermitian transpose. With Goodman's model [68], the covariance matrix  $\mathbf{C}_i$  is circular complex Wishart distributed, for  $L \geq D$ :

$$p_{\mathbf{C}_i}(\mathbf{C}_i | \Sigma_i) = \frac{L^{LD} \det(\mathbf{C}_i)^{L-D}}{\Gamma_D(L) \det(\Sigma_i)^L} \exp(-L \text{tr}(\Sigma_i^{-1} \mathbf{C}_i)) \quad (18)$$

where  $\Sigma_i = \mathbb{E}[\vec{k}_i \vec{k}_i^\dagger]$  is the underlying  $D \times D$  complex covariance matrix that codes for reflexivity and complex correlation,  $\det(\cdot)$  denotes the determinant of a matrix,  $\text{tr}(\cdot)$  denotes the trace of a matrix,  $\Gamma(\cdot)$  denotes the multivariate Gamma function. When  $L < D$ , the  $\mathbf{C}_i$  is singular ( $\det(\mathbf{C}_i) = 0$ ). To ensure that covariance matrix is full rank, a rescaling operation [69] is performed on the off-diagonal elements of  $\mathbf{C}_i$  to obtain matrix  $\mathbf{C}'_i$  as

$$\begin{aligned} \forall d, \quad & C'_i(d, d) = C_i(d, d), \\ \forall d \neq e, \quad & C'_i(d, e) = \gamma C_i(d, e), \quad d, e = 1, \dots, D \end{aligned} \quad (19)$$

where  $\gamma = \min\{L/D, 1\}$ . With the generalized likelihood ratio criterion, a distance criterion between two patches  $\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2$  is given in [69] as follows:

$$\begin{aligned} dist_{M-SAR}^{(1)}(\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2) &= \frac{1}{(2p+1)^2} \sum_{i=1}^{(2p+1)^2} \log\left(\frac{\det\left(\frac{1}{2}(\mathbf{C}'_{1,i} + \mathbf{C}'_{2,i})\right)^2}{\det(\mathbf{C}'_{1,i}) \det(\mathbf{C}'_{2,i})}\right) \end{aligned} \quad (20)$$

where  $\mathbf{C}'_{1,i}$  and  $\mathbf{C}'_{2,i}$  denote the rescaled covariance matrix for the  $i$ -th pixel in patch  $\tilde{\mathbf{X}}_1$  and  $\tilde{\mathbf{X}}_2$ , respectively.

Next, similar to  $dist_{SAR}^{(2)}$ , we derive another distance calculation for the covariance-matrix-based strategy with the variance stabilization criterion. For the Wishart distributed  $\mathbf{C}_i$  in (18), denoted by  $\mathcal{W}(\Sigma_i; L)$ , according to [70], we have

$$\begin{aligned} \mathbb{E}[\mathbf{C}_i] &= \Sigma_i; \\ \text{var}[\mathbf{C}_i(d, e)] &= \frac{1}{L} \Sigma_i(d, d) \Sigma_i(e, e); \quad d, e = 1, \dots, D \end{aligned} \quad (21)$$

Since the variance for off-diagonal elements does not depend on  $\Sigma_i(d, e)$  but on  $\Sigma_i(d, d)$  and  $\Sigma_i(e, e)$ , it is difficult to find an invertible application  $\mathcal{S}$  which stabilizes the variance of  $\mathbf{C}_i$  (makes  $\text{var}[\mathcal{S}(\mathbf{C}_i)]$  to be a constant). With the matrix logarithm transform  $\hat{\mathbf{C}}_i = \log \mathbf{C}_i$  and  $\hat{\Sigma}_i = \log \hat{\Sigma}_i$ , the covariance matrices  $\hat{\Sigma}_i$  are distributed according to a Wishart-Fisher-Tippett distribution [70]. However, the expectation and variance of  $\hat{\mathbf{C}}_i$  do not seem to be known in closed form in the literature. Nevertheless, as the determinant  $\det(\mathbf{C}_i)$  also

represents information of  $\mathbf{C}_i$ , we can use  $\det(\mathbf{C}_i)$  to measure the similarity between different patches. According to [71], we have

$$\frac{\det^D(\mathbf{C}_i)}{\det(\Sigma_i)} \sim \prod_{i=0}^{D-1} \chi_{L-i}^2 \quad (22)$$

The distribution of  $\frac{\det^D(\mathbf{C}_i)}{\det(\Sigma_i)}$  is similar to the product of chi-square variables with different degrees of freedom. The first two moments of Log-Determinant  $\log(\det(\mathbf{C}_i))$  (which coincides with the trace of  $\hat{\mathbf{C}}_i$ :  $\log(\det(\mathbf{C}_i)) = \text{tr}\hat{\mathbf{C}}_i$ ) are known in closed form [71] as

$$\begin{aligned} \mathbb{E}[\log(\det(\mathbf{C}_i))] &= \log[\det(\Sigma_i)] \\ &\quad + \sum_{i=1}^D \psi(0, L-i+1) - D \log L \\ \text{var}[\log(\det(\mathbf{C}_i))] &= \sum_{i=1}^D \psi(1, L-i+1) \end{aligned} \quad (23)$$

where  $\psi(m, \cdot)$  is the  $m$ -order Polygamma function. Therefore, variance stabilization of  $\det(\mathbf{C}_i)$  can be performed by using a logarithm transform, and then the resulting distance criterion is given by:

$$\begin{aligned} \text{dist}_{M-SAR}^{(2)}(\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2) \\ = \frac{1}{(2p+1)^2} \sum_{i=1}^{(2p+1)^2} [\log(\det(\mathbf{C}'_{1,i})) - \log(\det(\mathbf{C}'_{2,i}))]^2 \end{aligned} \quad (24)$$

The channel-based strategy is simple and easy to calculated, however, it does not use important inter-channel cross-correlations, resulting in less accuracy than the covariance-matrix-based strategy, as shown in the experimental section.

### B. Robustness in homogeneous CD

Next, we will show that INLPG also has advantages in homogeneous CD. As mentioned above, because INLPG compares the structures of images, it is not sensitive to the interference factors, such as illumination, season and noise in homogeneous CD. Here, we demonstrate the effectiveness of INLPG from another perspective by proving that the comparison of patch relationships in INLPG is more robust than the traditional patch difference/ratio operators.

We simply consider two noisy homogeneous images  $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{M \times N}$  divided into  $\sqrt{Q} \times \sqrt{Q}$  square patches. Let  $\mathbf{X}^*$  and  $\mathbf{Y}^*$  being the corresponding true images, we make the following assumptions.

**Assumption 1.** For the spectral images  $\mathbf{X}$  and  $\mathbf{Y}$ , their pixels are independently distributed according to the AWGN model as  $x \sim N(x^*, \sigma^2)$  and  $y \sim N(y^*, \sigma^2)$ , where  $\sigma^2$  is the variance of the images.

**Assumption 2.** For the SAR images  $\mathbf{X}$  and  $\mathbf{Y}$ , their pixels are independent distributed according to the usual multiplicative unit-mean Gamma model as  $x/x^* \sim \text{Gamma}(L, L)$  and  $y/y^* \sim \text{Gamma}(L, L)$ , where  $L$  is the ENL of the images.

For the vectorized target patches  $\mathbf{X}_1$  and  $\mathbf{Y}_1$  with their elements defined as  $x_1(q), y_1(q), q = 1, \dots, Q$ , and their

$k$ -th NN are defined as  $\mathbf{X}_{ik}, i^k \in \mathcal{N}_{\mathbf{X}_1}^K$  and  $\mathbf{Y}_{jk}, j^k \in \mathcal{N}_{\mathbf{Y}_1}^K$  with  $\mathbf{X}_{ik}^* = \mathbf{X}_1^*$  and  $\mathbf{Y}_{jk}^* = \mathbf{Y}_1^*$ . Next, we compare different operators of measuring the change level between the target patches  $\mathbf{X}_1$  and  $\mathbf{Y}_1$ .

The traditional patch-wise difference operator and log-ratio operator are respectively defined as

$$\begin{aligned} dt_{spe} &= \|\mathbf{X}_1 - \mathbf{Y}_1\|_2^2 \\ dt_{SAR} &= \|\log \mathbf{X}_1 - \log \mathbf{Y}_1\|_2^2 \end{aligned} \quad (25)$$

The proposed patch-wise relationship difference operators are defined as (corresponding to (8))

$$\begin{aligned} dr_{spe} &= \frac{1}{K} \sum_{k=1}^K \left( \|\mathbf{Y}_{ik} - \mathbf{Y}_1\|_2^2 - \|\mathbf{Y}_{jk} - \mathbf{Y}_1\|_2^2 \right) \\ dr_{SAR} &= \frac{1}{K} \sum_{k=1}^K \left( \|\log \mathbf{Y}_{ik} - \log \mathbf{Y}_1\|_2^2 - \|\log \mathbf{Y}_{jk} - \log \mathbf{Y}_1\|_2^2 \right) \end{aligned} \quad (26)$$

where  $\mathbf{Y}_{ik}, i^k \in \mathcal{N}_{\mathbf{X}_1}^K$  is the mapped patch of  $\mathbf{X}_{ik}$ , and  $\mathbf{Y}_{jk}, j^k \in \mathcal{N}_{\mathbf{Y}_1}^K$  is the  $k$ -th NN of  $\mathbf{Y}_1$ . To compare these operators, we define the change level contrast ratio (CCR) as

$$ccr(d) = \frac{\mathbb{E}[d|_{\mathbf{X}_1 \text{ changed}}]}{\mathbb{E}[d|_{\mathbf{X}_1 \text{ unchanged}}]} = \frac{\mathbb{E}[d|_{\mathbf{X}_1^* \neq \mathbf{Y}_1^*}]}{\mathbb{E}[d|_{\mathbf{X}_1^* = \mathbf{Y}_1^*}]}, \quad d \in \{dt_{spe}, dt_{SAR}, dr_{spe}, dr_{SAR}\} \quad (27)$$

which can measure the contrast of change levels on the changed and unchanged patches obtained by different operators.

**Theorem 1.** Suppose that the  $K$  NNs of the target patch are not changed in the event. For the multitemporal spectral images, if Assumption 1 is satisfied, then  $ccr(dr_{spe}) > ccr(dt_{spe})$ ; for the multitemporal SAR images, if Assumption 2 is satisfied, then  $ccr(dr_{SAR}) > ccr(dt_{SAR})$ .

Proof of Theorem 1 can be found in the Appendix, where we also derive the degree of CCR improvement. From Theorem 1, we can find that the proposed relationship difference operator is more robust to noise, which plays an important role in the CD of SAR images as shown in the experimental results of Fig. 3 in Section V-B. Therefore, the performance of INLPG in homogeneous CD is also guaranteed.

## V. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, experiments are performed to evaluate the performance of the INLPG<sup>1</sup>, which are conducted on real images of different CD tasks (homogeneous and heterogeneous) with six scenarios. The quantitative measures, experimental results and parameter analysis are as follows.

### A. Figures-of-merit

The quality of DIs generated by different methods can be evaluated by the empirical receiver operating characteristics (ROC) curves, which takes the probability of false alarm (PFA) as abscissa and probability of detection (PD) as ordinate. Moreover, the area under the curve (AUC) is used as the

<sup>1</sup>INLPG is available at <https://github.com/yulisun/INLPG>.

TABLE III  
DATA SET DESCRIPTION OF SIX SCENARIOS.

Scenario (Dataset)	Sensor	Size (pixels)	Date	Location	Event (& Spatial resolution)
1 #1 #2	Radarsat-2	289 × 257 × 1(1)	June 2008 - June 2009	Yellow River Estuary, China Shangqiu City, China	Farmland irrigation (8m.)
	Sentinel-1A	468 × 669 × 1(1)	Jan. 2016 - May 2016		Lake and building changes (20m.)
2 #3 #4	UAVSAR	2360 × 600 × 3(3)	Apr. 2009 - May 2015	Los Angeles, California, USA Los Angeles, California, USA	New buildings (Rg: 1.67m, Az: 0.6m.)
	UAVSAR	2300 × 600 × 3(3)	Apr. 2009 - May 2015		New buildings (Rg: 1.67m, Az: 0.6m.)
3 #5 #6	Google Earth	500 × 500 × 3(3)	Sept. 2012 - Mar. 2013	Beijing, China Beijing, China	Construction (1m.)
	Google Earth	500 × 500 × 3(3)	Sept. 2012 - Mar. 2013		Construction (1m.)
4 #7 #8	TerraSAR-X/Pleiades	2604 × 4404 × 1(1)	Feb. 2009 - July 2013	Toulouse, France Sutter County, California, USA	Construction (2m.)
	Landsat-8/Sentinel-1A	875 × 500 × 11(3)	Jan. 2017 - Feb. 2017		Flooding (≈15m.)
5 #9 #10	Landsat-5/Google Earth	300 × 412 × 1(3)	Sept. 1995 - July 1996	Sardinia, Italy Bastrop County, Texas, USA	Lake expansion (30m.)
	Landsat-5/EO-1 ALI	1534 × 808 × 7(10)	Aug. 2011 - Sept. 2011		Forest fire (30m.)
6 #11 #12	Landsat-8/Sentinel-1A	468 × 669 × 7(2)	Dec. 2015 - May 2016	Shangqiu City, China Los Angeles, California, USA	Lake and building changes (20m.)
	UAVSAR/Google Earth	2300 × 600 × 3(3)	Apr. 2009 - June 2017		New buildings (Rg: 1.67m, Az: 0.6m.)

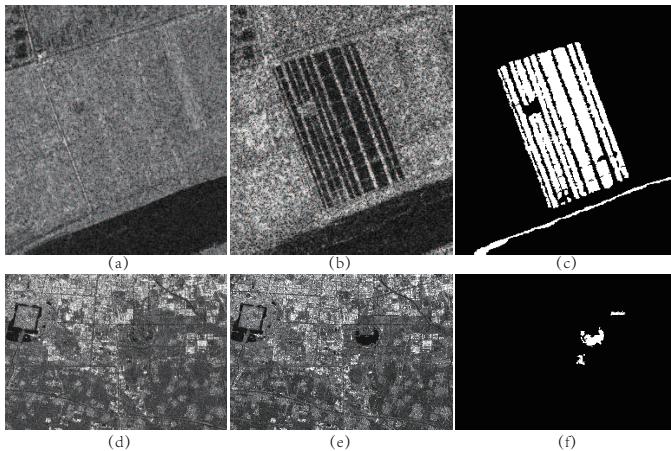


Fig. 2. Dataset #1 and Dataset #2 of Scenario 1. (a)-(c) the pre-event image, post-event image, and the ground truth image of Dataset #1, respectively. (d)-(f) the pre-event image, post-event image, and the ground truth image of Dataset #2, respectively.

quantitative criterion for the ROC curve, which is the integral of the ROC curve. The higher the AUC, the better the quality of DI.

### B. Experimental results and discussion

We construct six distinct scenarios to illustrate the performance of the proposed CD framework, as listed in Table III. Homogeneous CD: Scenario 1 with two single-polarization SAR data sets; Scenario 2 with two PolSAR data sets; Scenario 3 with two optical image data sets. Heterogeneous CD: Scenario 4 with two SAR/optical (multispectral) data sets; Scenario 5 with two multispectral data sets of different bands acquired from different sensors; Scenario 6 with two PolSAR/optical (multispectral) data sets. All considered images have been manually geographically aligned to fulfill the requirements of the CD framework.

The main parameters of the INLPG are the patch size  $p$ , the step size  $\Delta_p$ , and the number of the most similar neighbors  $K$ . For all the experimental results, we set  $\Delta_p = p$ ,  $K = \lceil 0.01N_{\Delta_p} \rceil$ , and adjust the  $p$  for different data sets. The impact of these parameters will be analyzed in detail in subsection V-C.

TABLE IV  
AUC COMPARISON OF DIFFERENT METHODS ON DATASET #1 AND DATASET #2.

Methods	Dataset #1	Dataset #2
Diff	0.657	0.742
LR	0.764	0.881
MR	0.902	0.968
NR	0.905	0.978
SDCD	0.901	0.950
NLR-ADMM	0.939	0.972
INLPG	<b>0.978</b>	<b>0.992</b>

1) *Homogeneous CD of single-polarization SAR data sets:* Two pairs of homogeneous SAR images are used in this scenario. Dataset #1 is acquired by Radarsat-2 SAR sensor over Yellow River Estuary, China, as shown in Figs. 2(a)-(c). The noise levels of multitemporal images are quite different, the pre-event image is a four-look image, while the post-event image is a single-look image. This different speckle noise level between two images makes the CD more complicated. Dataset #2 is acquired by Sentinel-1A SAR sensor with VV polarization over Shangqiu City, China, as shown in Figs. 2(d)-(f). Both SAR images in Dataset #2 are four-look images.

To compare the DIs on the homogeneous SAR images, we choose the difference operator (Diff), log-ratio operator (LR), mean-ratio method (MR) [28], neighborhood-ratio method (NR) [72], sparsity-driven change detection (SDCD) [73] and nonlocal low-rank model with the alternating direction methods of multipliers (NLR-ADMM) [34] as the comparison methods. For the MR and NR, we set the neighborhood size as  $3 \times 3$ ; for the SDCD, we set the regularization parameter  $\lambda$  from  $10^{-4}$  to  $10^{-1}$  with 20 logarithmic equally spaced, and then select the best one as the detection result; for the NLR-ADMM, we set the patch size as  $5 \times 5$ ; for the INLPG, we set  $p = 2$  for both two data sets.

Fig. 3 shows the DIs generated by different methods of Dataset #1 and Dataset #2, where all DIs are displayed in ‘jet’ colormap. Fig. 4 plots the ROC curves, and the corresponding AUC are listed in Table IV. As can be seen from Fig. 3, the INLPG generated DI can reduce the influence of speckle noise as it compares the patchwise similarity, so as to obtain a cleaner and more accurate DI. Therefore, the DI of INLPG shows the best performance on ROC curves in Fig. 4 and get the highest AUC in Table IV.

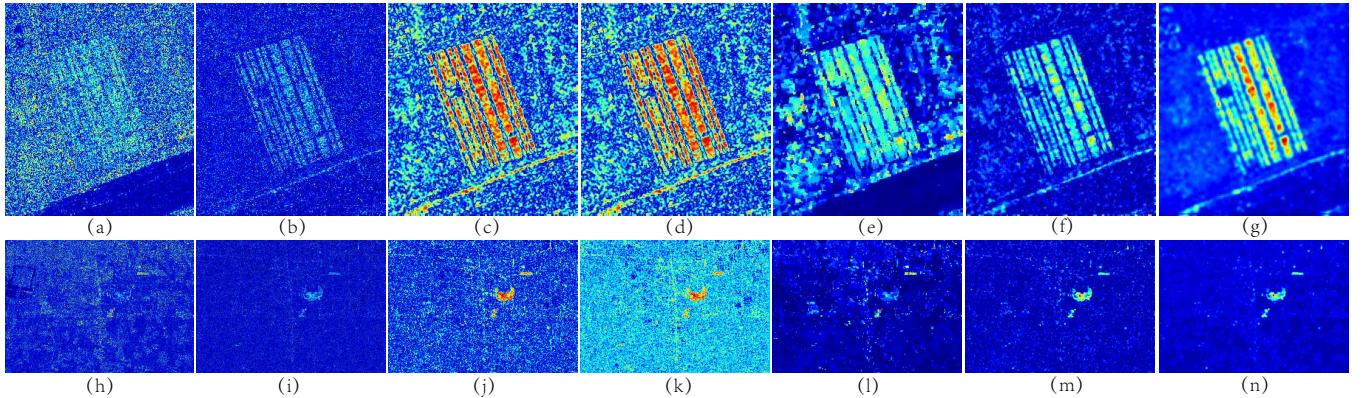


Fig. 3. DIs generated by different methods of Dataset #1 and Dataset #2. The top row is the DIs of Dataset #1: (a) Diff; (b) LR; (c) MR; (d) NR; (e) SDCD; (f) NLR-ADMM; (g) INLPG. The bottom row is the DIs of Dataset #2: (h) Diff; (i) LR; (j) MR; (k) NR; (l) SDCD; (m) NLR-ADMM; (n) INLPG.

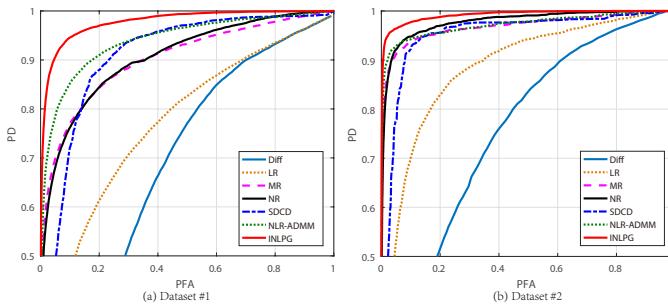


Fig. 4. ROC curves on Dataset #1 and Dataset #2: (a) Dataset #1; (b) Dataset #2.

TABLE V  
QUANTITATIVE MEASURES OF BINARY CMs GENERATED BY DIFFERENT METHODS ON DATASET #1.

Methods	FP(%)	FN(%)	KC
PCANet	2.46	2.14	0.846
INLPG-PCANet	1.86	1.90	0.873
CWNN	0.96	2.28	0.887
INLPG-CWNN	1.00	1.62	0.910

In addition, in order to show the advantages of DIs generated by the proposed INLPG, we also apply it to the unsupervised deep learning networks to construct the pseudo training set, which are the patches with high probability to be changed or unchanged. We test the application of INLPG in PCANet<sup>2</sup> [52] and CWNN<sup>3</sup> [53] to replace the original filtering based log-ratio operator. Fig. 5 shows the binary CMs generated by the original PCANet, CWNN and the INLPG based PCANet and CWNN (denoted as INLPG-PCANet and INLPG-CWNN for short). Table V reports three quantitative measures of these CMs: the false-positive (FP) rate, the false-negative (FN) rate, and the Kappa coefficient (KC). From Fig. 5 and Table V, we can find that the INLPG induced pseudo training set can obtain better detection performance with higher accuracy.

2) *Homogeneous CD of PolSAR data sets:* In Scenario 2, two pairs of images from L-band UAVSAR acquired over the city of Los Angeles, California, on April 23, 2009 and May

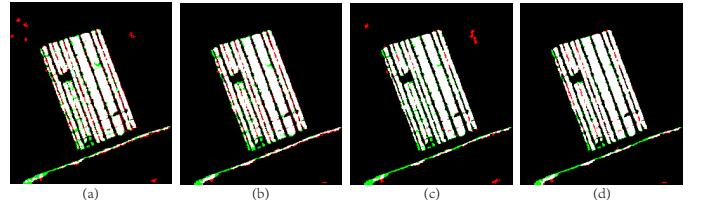


Fig. 5. Binary CMs generated by different methods on Dataset #1: (a) PCANet; (b) INLPG based PCANet; (c) CWNN; (d) INLPG based CWNN. In the binary CM, White: true positives (TP); Red: false positives (FP); Black: true negatives (TN); Green: false negatives (FN).

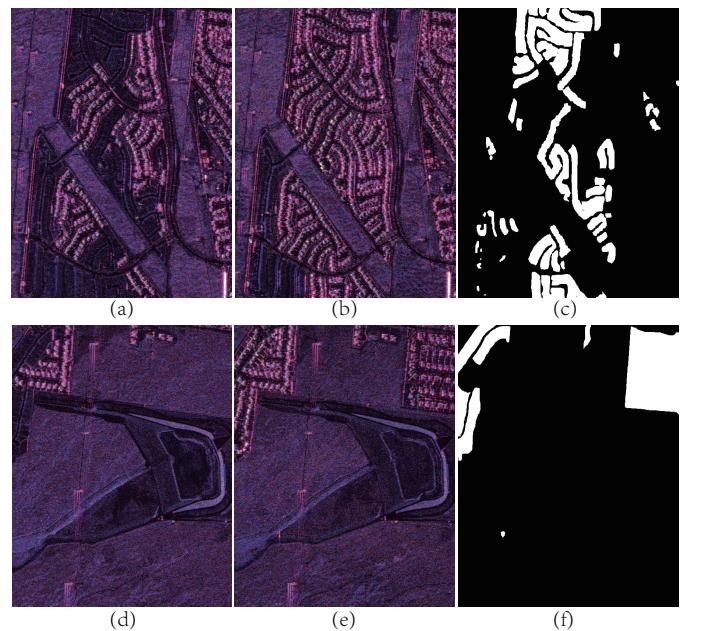


Fig. 6. Dataset #3 and Dataset #4 of Scenario 2. (a)-(c) the pre-event image, post-event image, and the ground truth image of Dataset #3, respectively. (d)-(f) the pre-event image, post-event image, and the ground truth image of Dataset #4, respectively.

11, 2015 with three channels of HH, HV and VV, and with a spatial resolution of 0.6m in azimuth and 1.67m in range are used. The size of the first pair and the second pair of images are  $2360 \times 600 \times 3$  and  $2300 \times 600 \times 3$ , respectively. The ground

<sup>2</sup>PCANet is kindly available at <https://github.com/summitgao>.

<sup>3</sup>CWNN is kindly available at <https://github.com/summitgao>.

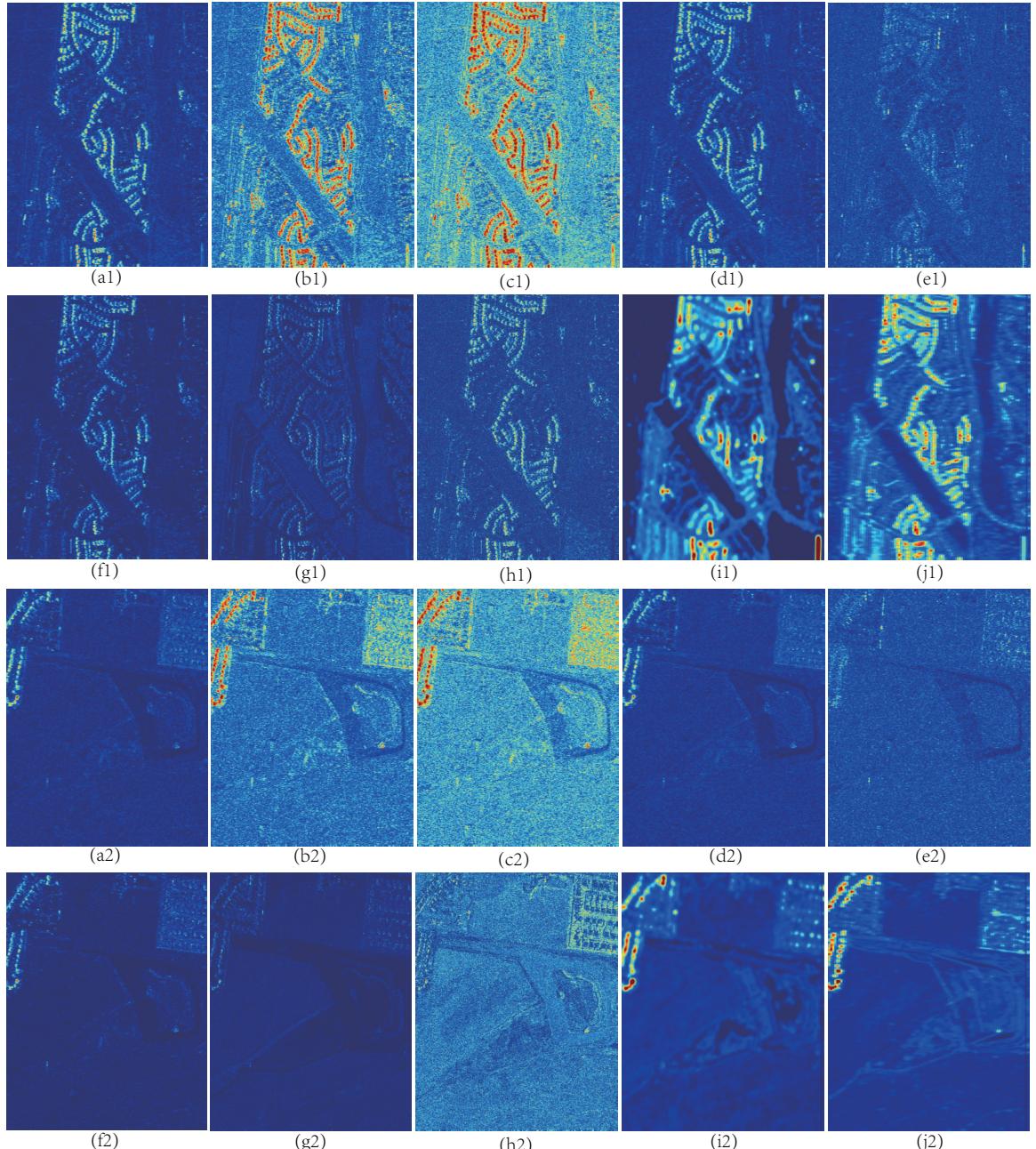


Fig. 7. DIs generated by different methods of Dataset #3 and Dataset #4. The top row is the DIs of Dataset #3: (a1)  $\hat{\Lambda}_G$ ; (b1)  $\hat{\Lambda}_{t1}$ ; (c1)  $\hat{\Lambda}_{Wald}$ ; (d1)  $\hat{\Lambda}_{MT}$ ; (e1)  $\hat{\Lambda}_{Mat}$ ; (f1)  $\hat{\Lambda}_{Tex}$ ; (g1) log-ratio (channel-based); (h1) log-ratio (covariance-matrix-based); (i1) INLPG ( $dist_{SAR}^{(2)}$ ); (j1) INLPG ( $dist_{M-SAR}^{(2)}$ ). The bottom row is the DIs of Dataset #4: (a2)  $\hat{\Lambda}_G$ ; (b2)  $\hat{\Lambda}_{t1}$ ; (c2)  $\hat{\Lambda}_{Wald}$ ; (d2)  $\hat{\Lambda}_{MT}$ ; (e2)  $\hat{\Lambda}_{Mat}$ ; (f2)  $\hat{\Lambda}_{Tex}$ ; (g2) log-ratio (channel); (h2) log-ratio (covariance matrix); (i2) INLPG ( $dist_{SAR}^{(2)}$ ); (j2) INLPG ( $dist_{M-SAR}^{(2)}$ )

truth is provided by Mian et al. [71], which represents the effects of urbanization. Fig. 6 shows these two data sets, where the images are compressed vertically by 3 times to display conveniently.

To compare the DIs on these homogeneous PolSAR images, we choose some statistics based CD methods as comparison: 1) Gaussian model based [75]: the generalised likelihood ratio test (GLRT) statistic  $\hat{\Lambda}_G$ , the t1 statistic  $\hat{\Lambda}_{t1}$  and the Wald statistic  $\hat{\Lambda}_{Wald}$ ; 2) non-Gaussian model based [74], the GLRT

statistic<sup>4</sup> for three different problems denoted as  $\hat{\Lambda}_{MT}$ ,  $\hat{\Lambda}_{Mat}$ , and  $\hat{\Lambda}_{Tex}$ , respectively (more details in [74]). For these statistical methods, we use the  $11 \times 11$  observation window as default; for the INLPG, we use both the channel-based (using  $dist_{SAR}^{(2)}$ ) and covariance-matrix-based (using  $dist_{M-SAR}^{(2)}$ ) strategies, and set  $p = 6$ . In addition, we also add the channel-based log-ratio DI (DI  $(i, j) = \sum_{d=1}^3 |\log \frac{x(i,j,d)}{y(i,j,d)}|$ ) and covariance-matrix-

<sup>4</sup>code is kindly available at <https://github.com/AmmarMian>.

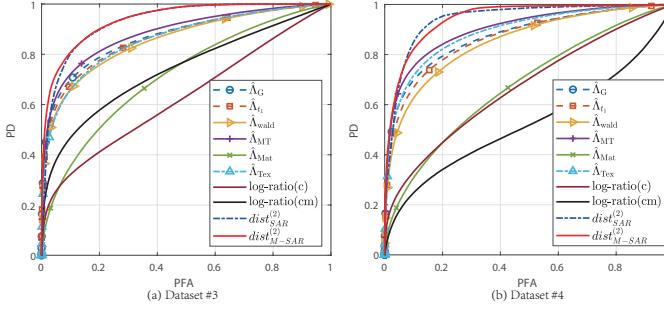


Fig. 8. ROC curves on Dataset #3 and Dataset #4: (a) Dataset #3; (b) Dataset #4. In the legend, the Log-ratio (c) and log-ratio (cm) represent channel-based log-ratio operator and covariance-matrix-based log-ratio operator, respectively; the  $dist_{SAR}^{(2)}$  and  $dist_{M-SAR}^{(2)}$  represent the INLPG using different patch distance criteria.

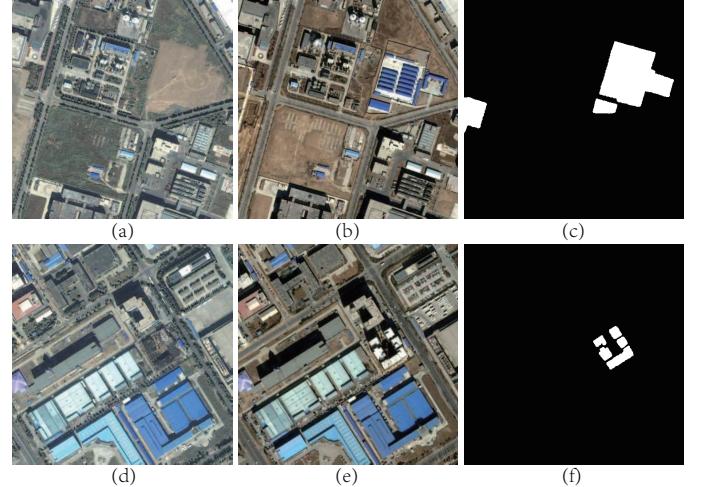


TABLE VI  
AUC COMPARISON OF DIFFERENT METHODS ON DATASET #3 AND DATASET #4.

Methods	Dataset #3	Dataset #4
$\hat{\Lambda}_G$	0.872	0.875
$\hat{\Lambda}_{t1}$	0.866	0.867
$\hat{\Lambda}_{Wald}$	0.857	0.852
$\hat{\Lambda}_{MT}$	0.886	0.903
$\hat{\Lambda}_{Mat}$	0.716	0.678
$\hat{\Lambda}_{Tex}$	0.861	0.890
log-ratio (channel)	0.628	0.671
log-ratio (covariance matrix)	0.738	0.528
INLPG ( $dist_{SAR}^{(2)}$ )	<b>0.932</b>	<b>0.942</b>
INLPG ( $dist_{M-SAR}^{(2)}$ )	<b>0.940</b>	<b>0.944</b>

based log-ratio DI ( $DI(i, j) = \left| \log \frac{\det(\mathbf{C}'_{x(i, j)})}{\det(\mathbf{C}'_{y(i, j)})} \right|$  with  $\mathbf{C}'_{x(i, j)}$  and  $\mathbf{C}'_{y(i, j)}$  representing the rescaled covariance matrix of pixel  $x(i, j)$  and  $y(i, j)$ , respectively).

Fig. 7 shows the DIs of Dataset #3 and Dataset #4 generated by different methods. Fig. 8 plots the ROC curves, and Table VI reports the corresponding AUC. From these results on Scenario 2, we can find that the statistical methods (such as  $\hat{\Lambda}_G$ ,  $\hat{\Lambda}_{t1}$ , and  $\hat{\Lambda}_{Tex}$ ) generally perform better than direct comparison methods (such as channel-based log-ratio operator and covariance-matrix-based log-ratio operator). At the same time, the  $\hat{\Lambda}_{MT}$  based on the non-Gaussian model is the best among these statistical methods. By comparing the INLPG based DIs with other DIs, we can observe that the INLPG based method can detect the change more easily, which is also confirmed by Fig. 8 and Table VI.

3) *Homogeneous CD of optical image data sets*: In this scenario, two pairs of optical images obtained from Google Earth over the city of Beijing on September 30, 2012 and March 4, 2013 are used, which are R, G and B channels, with a spatial resolution of 1m. The size of both pairs of images is  $500 \times 500 \times 3$ . The ground truth of Dataset #5 and Dataset #6 represents the changes of buildings, as shown in Fig. 9.

We choose the CVA [22], MAD<sup>5</sup> [24], IR-MAD<sup>6</sup> [25], the deep slow feature analysis network (DSFANet)<sup>7</sup> [76] and

Fig. 9. Dataset #5 and Dataset #6 of Scenario 3. (a)-(c) the pre-event image, post-event image, and the ground truth image of Dataset #5, respectively. (d)-(f) the pre-event image, post-event image, and the ground truth image of Dataset #6, respectively.

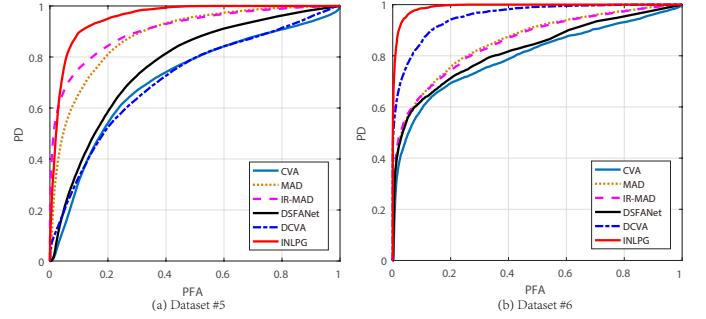


Fig. 11. ROC curves on Dataset #5 and Dataset #6: (a) Dataset #5; (b) Dataset #6.

DCVA<sup>8</sup> [49] as the DI comparison methods. In these two data sets, we set  $p = 4$  for INLPG. The DIs generated by different methods on this scenario are shown in Fig. 10, and the corresponding ROC curves are plotted in Fig. 11. Table VII reports the AUC comparison of different methods on Scenario 3. The analysis of these results shows that the INLPG based method can suppress the interference caused by light intensity, seasonal changes in the homogeneous CD of optical images, especially for the Dataset #6.

It should be noted that the proposed INLPG is not sensitive to illumination, season and noise because it measures the similarity structure changes of image patches, rather than directly comparing the differences between pixels. Therefore, it can achieve the best performance in the above scenarios of homogeneous CD (Scenarios 1, 2, and 3). Accordingly, if there are no such interference factors in the pre-event and post-event images, and satisfactory results can be obtained just by using intensity ratio or spectral difference operators, the performance advantage of the INLPG will be discounted.

4) *Heterogeneous CD of SAR/optical (multispectral) data sets*: In this scenario, Dataset #7 and Dataset #8 are used to test the performance of INLPG, as shown in Fig. 12. Dataset

<sup>5</sup>MAD is kindly available at <https://people.compute.dtu.dk/alan>.

<sup>6</sup>IR-MAD is kindly available at <https://people.compute.dtu.dk/alan>.

<sup>7</sup>DSFANet is kindly available at <https://github.com/rulixiang/DSFANet>.

<sup>8</sup>DCVA is kindly available at <https://github.com/sudipansaha>.

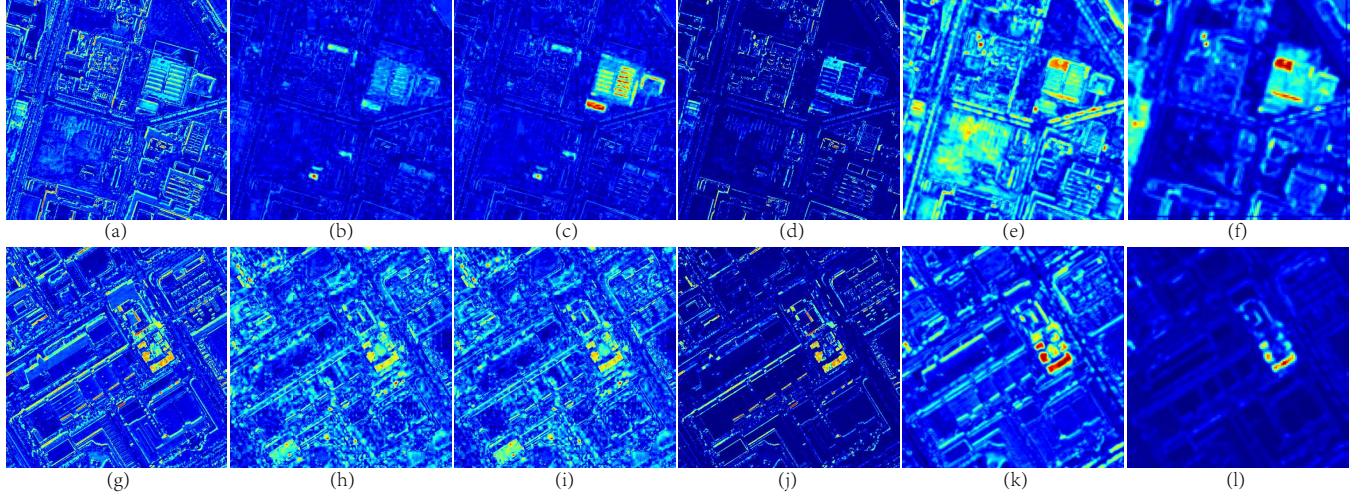


Fig. 10. DIs generated by different methods of Dataset #5 and Dataset #6. The top row is the DIs of Dataset #5: (a) CVA; (b) MAD; (c) IR-MAD; (d) DSFANet; (e) DCVA; (f) INLPG. The bottom row is the DIs of Dataset #6: (g) CVA; (h) MAD; (i) IR-MAD; (j) DSFANet; (k) DCVA; (l) INLPG.

TABLE VII  
AUC COMPARISON OF DIFFERENT METHODS ON DATASET #5 AND  
DATASET #6.

Methods	Dataset #5	Dataset #6
CVA	0.712	0.798
MAD	0.885	0.859
IR-MAD	0.908	0.856
DSFANet	0.769	0.824
DCVA	0.715	0.955
INLPG	<b>0.955</b>	<b>0.992</b>

#7 consists of a pair of SAR/optical images acquired over Toulouse, France. The SAR image is sensed by TerraSAR-X satellite, and the optical image is obtained by Pleiades (High-Resolution Optical Imaging Constellation of CNES, French National Centre for Space) satellite. The SAR image was co-registered and re-sampled with a spatial resolution of 2m to match the optical image. The ground truth of Dataset #7 shows the change of construction. Dataset #8 consists of a pair of multispectral/SAR images acquired over California. The multispectral image is sensed by Landsat-8 with 11 bands, and the SAR image is sensed by Sentinel-1A with three channels (two channels are VV and VH polarization data, and the third channel is the ratio between them). The ground truth of Dataset #8 shows a flood in California, which is constructed by Luppino et al. [66]. Although the SAR image in Dataset #8 is multichannel, it does not use the cross-correlation between different polarization data, but treats them as color images with three channels as shown in Fig. 12(e).

For the heterogeneous CD, we apply five methods for comparing the DI: SHD [41], PP [43], AMD<sup>9</sup> [44], NLPG<sup>10</sup> [21], and the recently proposed fractal projection and Markovian segmentation-based method (FPMS)<sup>11</sup> [77], which contains a fractal encoding step and a fractal projection/decoding step. We set  $p = 6$  for Dataset #7, and  $p = 2$  for Dataset #8.

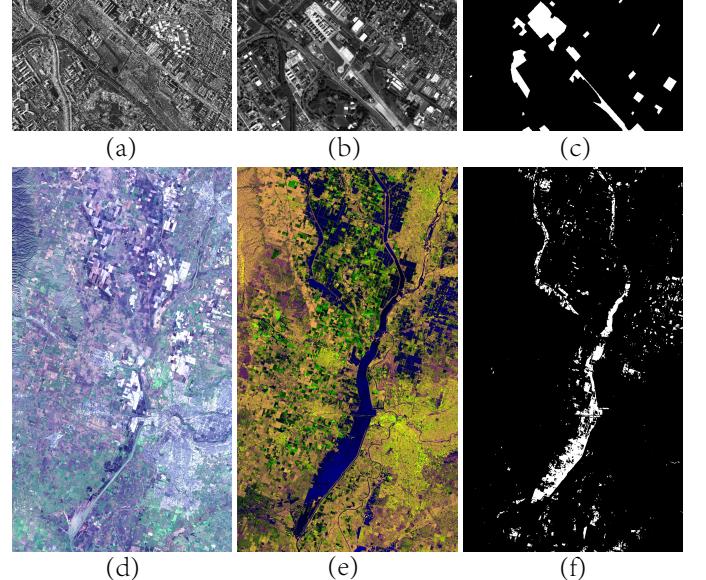


Fig. 12. Dataset #7 and Dataset #8 of Scenario 4. (a)-(c) the pre-event image, post-event image, and the ground truth image of Dataset #7, respectively. (d)-(f) the pre-event image, post-event image, and the ground truth image of Dataset #8, respectively.

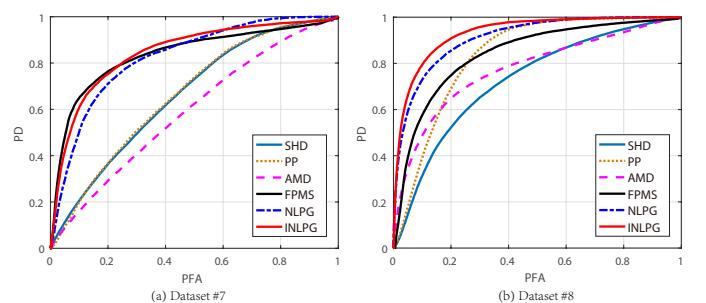


Fig. 14. ROC curves on Dataset #7 and Dataset #8: (a) Dataset #7; (b) Dataset #8.

Fig. 13 shows the DIs of Dataset #7 and Dataset #8

<sup>9</sup>AMD is kindly available at <https://sites.google.com/view/luppino>.

<sup>10</sup>NLPG is available at <https://github.com/yulisun/NPSG>.

<sup>11</sup>FPMS is kindly available at <http://www-labs.iro.umontreal.ca/~mignotte>.

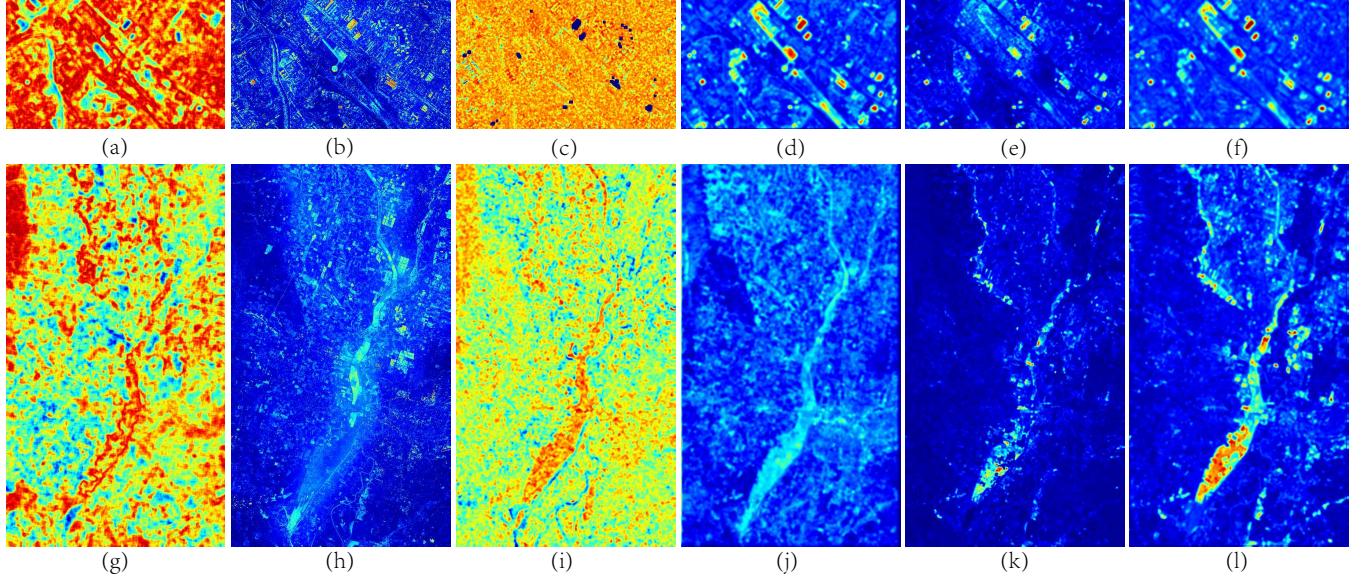


Fig. 13. DIs generated by different methods of Dataset #7 and Dataset #8. The top row is the DIs of Dataset #7: (a) SHD; (b) PP; (c) AMD; (d) FPMS; (e) NLPG; (f) INLPG. The bottom row is the DIs of Dataset #8: (g) SHD; (h) PP; (i) AMD; (j) FPMS; (k) NLPG; (l) INLPG.

TABLE VIII  
AUC COMPARISON OF DIFFERENT METHODS ON DATASET #7, DATASET #8, DATASET #9 AND DATASET #10.

Methods	Dataset #7	Dataset #8	Dataset #9	Dataset #10
SHD	0.660	0.726	0.884	0.558
PP	0.663	0.836	0.744	0.372
AMD	0.590	0.775	0.410	0.890
FPMS	0.836	0.844	0.925	0.704
NLPG	0.824	0.911	0.913	0.953
INLPG	<b>0.845</b>	<b>0.932</b>	<b>0.949</b>	<b>0.971</b>

generated by different methods. Fig. 14 plots the ROC curves, and the corresponding AUC are listed in Table VIII. As shown in Fig. 13, INLPG performs better than other methods, followed by NLPG and FPMS, which can clearly distinguish the changed part from the unchanged part in the DIs. At the same time, we can also find that INLPG is obviously better than NLPG in terms of visual effect in Fig. 13 and AUC value in Table VIII. This is mainly because INLPG improves the graph construction process, structure difference calculation and DI fusion process, which makes the detection results more robust and accurate.

5) *Heterogeneous CD of multispectral data sets:* In Scenario 5, we use two pairs of multispectral images with different bands, as shown in Fig. 15. Dataset #9 consists of one near-infrared (NIR) band image and one optical image acquired over Sardinia, Italy. The NIR band image is sensed by Landsat-5, and the optical image is obtained from Google Earth with R, G, B bands. The ground truth of Dataset #9 shows the Lake expansion on Sardinia, Italy. Dataset #10 consists of two multispectral images acquired by different sensors. The pre-event image is sensed by Landsat-5 with 7 bands, and post-event image is sensed by Advanced Land Imager from the Earth Observing mission (EO-1 ALI) with 10 bands. The ground truth of Dataset #10 is provided by Volpi et al. [78], which represents a forest fire in Bastrop County, Texas. Both

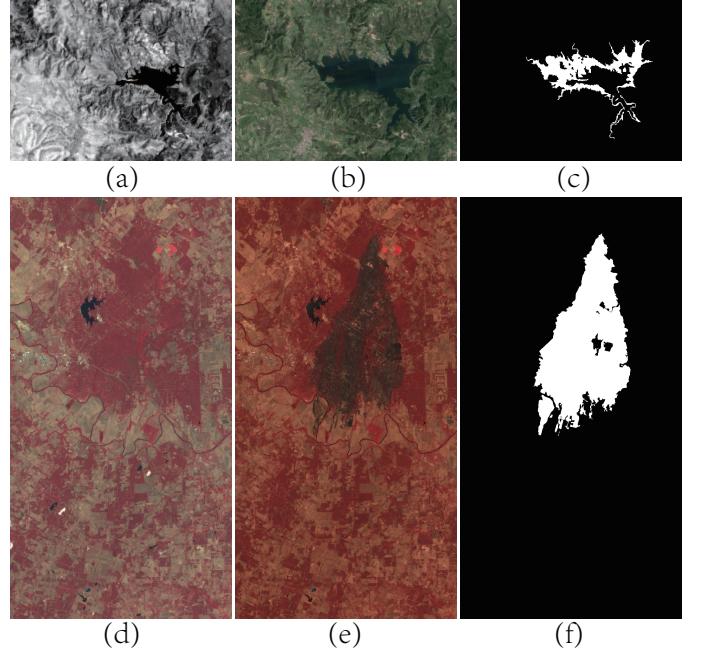


Fig. 15. Dataset #9 and Dataset #10 of Scenario 5. (a)-(c) the pre-event image, post-event image, and the ground truth image of Dataset #9, respectively. (d)-(f) the pre-event image, post-event image, and the ground truth image of Dataset #10, respectively.

Dataset #9 and Dataset #10 have a spatial resolution of 30m. In these two data sets, we set  $p = 3$  for INLPG.

Fig. 16 shows the DIs of Dataset #9 and Dataset #10 generated by SHD, PP, AMD, FPMS, NLPG and INLPG. Fig. 17 plots the ROC curves of these DIs, and the corresponding AUC are listed in Table VIII. As expected, INLPG shows better performance than NLPG in terms of DI visual effect and ROC curves on this scenario. The AUC of NLPG on Dataset #9 and Dataset #10 are increased from 0.913, 0.953 to 0.949

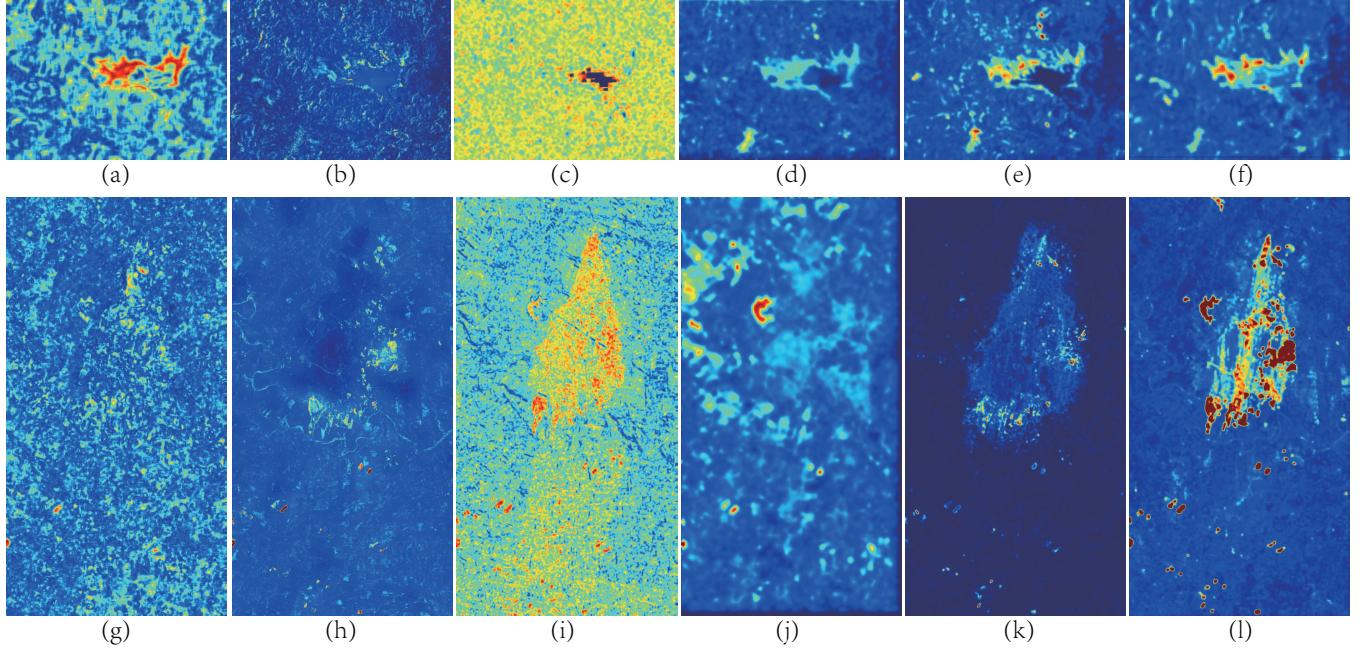


Fig. 16. DIs generated by different methods of Dataset #9 and Dataset #10. The top row is the DIs of Dataset #9: (a) SHD; (b) PP; (c) AMD; (d) FPMS; (e) NLPG; (f) INLPG. The bottom row is the DIs of Dataset #10: (g) SHD; (h) PP; (i) AMD; (j) FPMS; (k) NLPG; (l) INLPG.

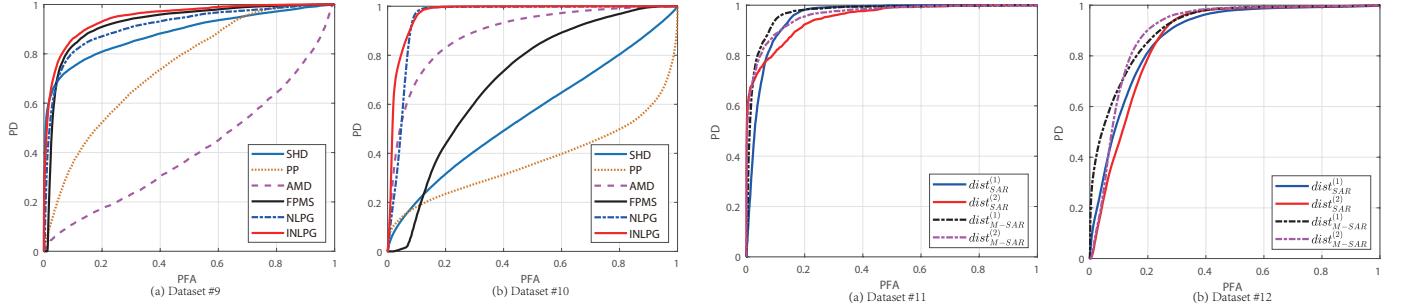


Fig. 17. ROC curves on Dataset #9 and Dataset #10: (a) Dataset #9; (b) Dataset #10.

and 0.971 by INLPG, respectively. The experimental results on Scenario 5 and Scenario 6 demonstrate the improvement effect of the INLPG.

*6) Heterogeneous CD of PolSAR/multispectral (optical) data sets:* In this scenario, we test the heterogeneous CD including PolSAR images, as shown in Fig. 18. Dataset #11 and Dataset #2 are acquired in the same region, which represent the same event. Unlike Dataset #2, the pre-event image of Dataset #11 is acquired by Landsat-8 without the panchromatic and “cirrus” bands, and the post-event image of Dataset #11 is acquired by Sentinel-1A with VV and VH polarizations. Similarly, Dataset #12 and Dataset #4 are also acquired in the same region and represent the same event. They share the same pre-event PolSAR image, but the post-event image of Dataset #12 is obtained from Google Earth with R, G, B bands.

Since heterogeneous CD with PolSAR images is rarely studied, we only compare the DIs generated by the proposed INLPG with different patch distance criteria, i.e., channel-based  $dist_{SAR}^{(1)}$  and  $dist_{SAR}^{(2)}$ , covariance-matrix-based

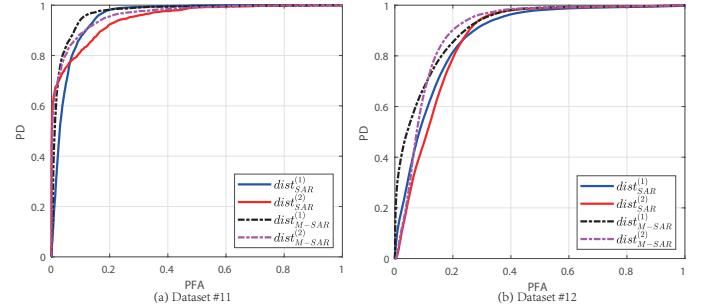


Fig. 19. ROC curves on Dataset #11 and Dataset #12: (a) Dataset #11; (b) Dataset #12.

$dist_{M-SAR}^{(1)}$  and  $dist_{M-SAR}^{(2)}$ . Fig. 18 shows the DIs of Dataset #11 and Dataset #12 generated by INLPG. Fig. 19 plots the ROC curves, and Table IX reports the corresponding AUC. By comparing the INLPG based DIs of Dataset #11 and Dataset #12 with the homogeneous CD of Dataset #2 and Dataset #4, it is obvious that heterogeneous CD is much more difficult than homogeneous CD. The distinct imaging mode of the pre-event and post-event images cause many confusions of changed and unchanged in the DI, as shown in Fig. 18. At the same time, we can also find that the covariance-matrix-based strategy performs better than the channel-based strategy on the ROC curves and AUC, which is mainly because that the covariance-matrix-based strategy uses the cross-correlations of the channels.

### C. Parameter analysis and computational time

The main parameters of the INLPG are the patch size  $p$ , the step size  $\Delta_p$ , and the number of the most similar neighbors  $K$ . In the NLPG [21], the influence of these parameters are deeply

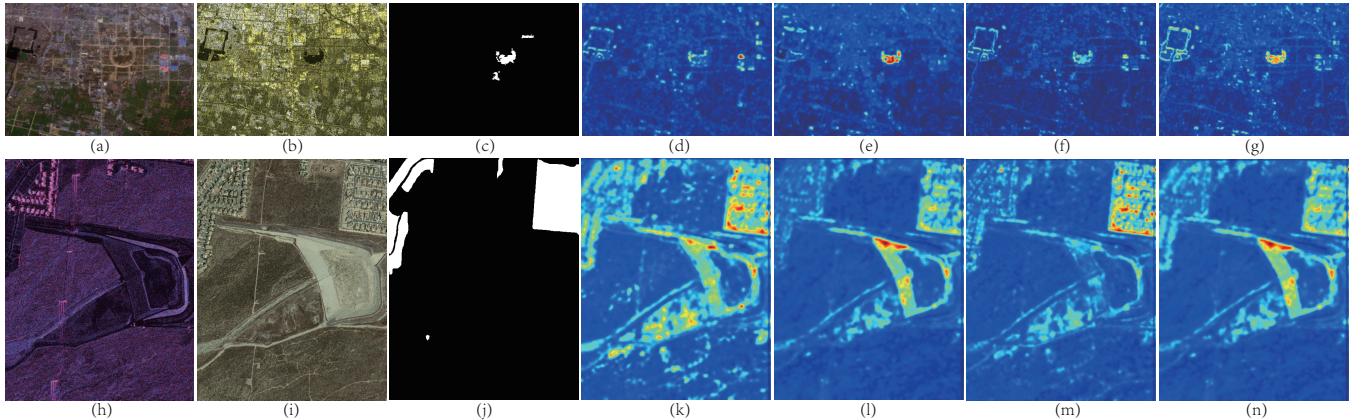


Fig. 18. DIs generated by INLPG with different patch distance criteria of Dataset #11 and Dataset #12. (a)-(c) the pre-event image, post-event image, and the ground truth image of Dataset #11, respectively. (d)-(g) DIs generated by INLPG with  $dist_{SAR}^{(1)}$ ,  $dist_{SAR}^{(2)}$ ,  $dist_{M-SAR}^{(1)}$ , and  $dist_{M-SAR}^{(2)}$  of Dataset #11, respectively. (h)-(j) the pre-event image, post-event image, and the ground truth image of Dataset #12, respectively. (k)-(n) DIs generated by INLPG with  $dist_{SAR}^{(1)}$ ,  $dist_{SAR}^{(2)}$ ,  $dist_{M-SAR}^{(1)}$ , and  $dist_{M-SAR}^{(2)}$  of Dataset #12, respectively.

TABLE IX  
AUC COMPARISON OF INLPG WITH DIFFERENT PATCH DISTANCE CRITERIA ON DATASET #11 AND DATASET #12.

Distance criteria	Dataset #11	Dataset #12
$dist_{SAR}^{(1)}$	0.953	0.877
$dist_{SAR}^{(2)}$	0.949	0.868
$dist_{M-SAR}^{(1)}$	0.970	0.911
$dist_{M-SAR}^{(2)}$	0.962	0.899

analyzed. Generally, a larger step size  $\Delta_p$  can speed up INLPG by reducing the number of target patch  $N_{\Delta_p}$ , but it will also make the DI more fuzzy as illustrated by (10). Considering the computational complexity and detection accuracy of the algorithm, we suggest to set  $\Delta_p = p$  as a compromise choice based on our experiments. For the parameter  $K$ , it is obvious that both extremely large and very small values are not appropriate. An extremely large  $K$  will increase the computational complexity and cause confusion, which means that some image patches that are not really similar to the target patch are also introduced into the  $K$ -NN graph. At the same time, a very small  $K$  will cause that the  $K$ -NN graph is not robust enough. Therefore, we empirically set  $K = \lceil 0.01N_{\Delta_p} \rceil$ . The parameter that has the greatest impact on performance of INLPG is the patch size  $p$ . In order to measure the impact of different  $p$ , we vary  $p$  from 1 to 6 and fix  $\Delta_p = 3$ ,  $K = \lceil 0.01N_{\Delta_p} \rceil$  to observe the change in AUC on different data sets. From Fig. 20, we can see that for the low-resolution data sets, a small patch size is more appropriate, such as the  $p = 2$  for Dataset #1 and Dataset #8,  $p = 3$  for Dataset #9. On the contrary, for the high-resolution data sets, a larger patch size is more appropriate, such as the  $p = 6$  for Dataset #4, Dataset #6 and Dataset #11. This is mainly due to the fact that in high-resolution images, the relationship between adjacent pixels is closer and the context information is richer.

To compare the computational time of NLPG and INLPG, we report the CPU times of these two methods on Dataset #8 and Dataset #9 in Table X. The patch size of NLPG and

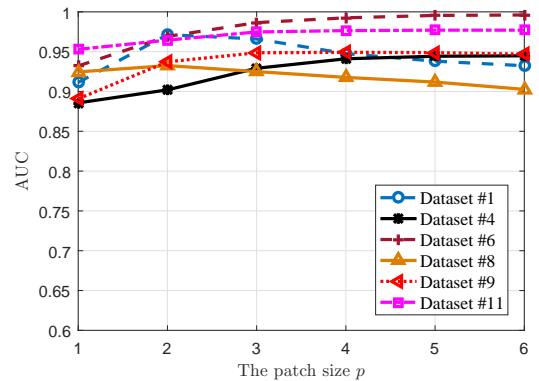


Fig. 20. INLPG performance with different patch size  $p$

INLPG are set to  $5 \times 5$ , the search window size of NLPG is set to  $100 \times 100$ , the search step size of NLPG is set to  $\Delta_s = \Delta_p$ , the  $K$  of NLPG is set to 35, while the  $K$  of INLPG is set to  $K = \lceil 0.01N_{\Delta_p} \rceil$ . The algorithms are performed in MATLAB 2016a running on a Windows desktop with Intel(R) Core(TM) i7-8700K CPU and 32GB of RAM. As can be seen from Table X, with the increase of  $\Delta_p$ , the time required for the algorithms is greatly reduced. In general, when  $N_{\Delta_p}$  is large, both INLPG and NLPG are time-consuming. However, the INLPG can be accelerated by using some efficient graph construction methods, such as the kgraph [61]. As shown in Table X, with the increase of  $N_{\Delta_p}$ , the acceleration effect of INLPG using kgraph is more obvious.

## VI. CONCLUSION

In this paper, we mainly focus on the unsupervised CD of both homogeneous and heterogeneous RS images. The proposed INLPG method is an extension of NLPG, which exploits the inherent nonlocal self-similarity of images and assumes that the unchanged pre-event and post-event images shares the imaging modality-invariant structure information. In order to expand the application range of NLPG and improve its detection performance, we have made the following improvements

TABLE X  
COMPUTATIONAL TIME (SECONDS) OF NLPG, INLPG AND INLPG USING KGRAPH UNDER DIFFERENT  $\Delta_p$ .

Data sets	Image size	NLPG			INLPG			INLPG using kgraph		
		$\Delta_p = 2$	$\Delta_p = 3$	$\Delta_p = 4$	$\Delta_p = 2$	$\Delta_p = 3$	$\Delta_p = 4$	$\Delta_p = 2$	$\Delta_p = 3$	$\Delta_p = 4$
Dataset #9	$300 \times 412 \times 1(3)$	387.9	79.7	26.8	27.1	5.9	3.3	20.0	5.3	3.1
Dataset #8	$875 \times 500 \times 11(3)$	1434.3	368.9	120.7	1487.7	262.5	77.0	525.9	154.1	49.6

in INLPG. First, we demonstrate that the structure consistency also has advantages in homogeneous CD by comparing it with traditional detection operators. Therefore, INLPG can be applied to both homogeneous and heterogeneous CD. Second, we extend it to the CD task with multichannel SAR image by using two strategies with different patch distance criteria. This work is very enlightening as the heterogeneous CD with PolSAR images is rarely studied. Third, we improve the graph construction process, structure difference calculation, and DI fusion process of NLPG, which makes INLPG more robust and the detection result more accurate. At last, we test INLPG on 6 scenarios with 12 data sets, which demonstrates that INLPG can be applied to a wide range of unsupervised CD tasks, including homogeneous/heterogeneous, SAR (single polarization, polarimetric)/spectral (optical, multispectral), moderate/coarse/high resolution CD tasks. However, the square image patch is used as the basic unit in INLPG, although it can simplify the preprocessing and distance calculation, it can not maintain the shape and structure of objects. Our future work is to use the superpixel as the basic unit, which can ensure that each unit represents the same kind of object and reduce the computational complexity by reducing the number of vertices. Therefore, we need to solve the problems of accurate superpixel segmentation and superpixel comparison in the future.

## APPENDIX A PROOF OF THEOREM 1

In this appendix, we compute the CCR of different operators. Here,  $i^k$  and  $j^k$  represent the position indices of the  $k$ -th NN of  $\mathbf{X}_1$  and  $\mathbf{Y}_1$ , respectively. Suppose that the target patch  $\mathbf{Y}_1$  have  $R$  ( $R \geq K$ ) really similar patches  $\mathbf{Y}_j$ ,  $j \in \mathcal{N}_{\mathbf{Y}_1}^R$  in the image with  $\mathbf{Y}_j^* = \mathbf{Y}_1^*$ .

### A. Case 1: the spectral images

1. Compute  $ccr(dt_{spe})$ . Rewrite the  $dt_{spe}$  as follows

$$dt_{spe} = \|\mathbf{X}_1 - \mathbf{Y}_1\|_2^2 = \sum_{q=1}^Q (x_1(q) - y_1(q))^2 \quad (\text{A.1})$$

Let  $\rho(q) = \frac{x_1(q) - y_1(q)}{\sqrt{2}\sigma}$ , by using Assumption 1, we have  $\rho(q) \sim N\left(\frac{x_1^*(q) - y_1^*(q)}{\sqrt{2}\sigma}, 1\right)$ . Let  $A = \sum_{q=1}^Q (\rho(q))^2$ , we have  $A$  is distributed according to the noncentral chi-square distribution with  $Q$  degree-of-freedom and

$$\mathbb{E}[A] = Q + \frac{1}{2\sigma^2} \sum_{q=1}^Q (x_1^*(q) - y_1^*(q))^2 \quad (\text{A.2})$$

Because  $dt_{spe} = 2\sigma^2 A$ , we have

$$\mathbb{E}[dt_{spe}] = 2Q\sigma^2 + \|\mathbf{X}_1^* - \mathbf{Y}_1^*\|_2^2 \quad (\text{A.3})$$

Then, we can obtain

$$ccr(dt_{spe}) = \frac{\mathbb{E}[dt_{spe} | \mathbf{X}_1^* \neq \mathbf{Y}_1^*]}{\mathbb{E}[dt_{spe} | \mathbf{X}_1^* = \mathbf{Y}_1^*]} = 1 + \frac{\|\mathbf{X}_1^* - \mathbf{Y}_1^*\|_2^2}{2Q\sigma^2} \quad (\text{A.4})$$

2. Compute  $ccr(dr_{spe})$ . Rewrite the  $dr_{spe}$  as follows

$$\begin{aligned} dr_{spe} &= \frac{1}{K} \sum_{k=1}^K \left( \|\mathbf{Y}_{i^k} - \mathbf{Y}_1\|_2^2 - \|\mathbf{Y}_{j^k} - \mathbf{Y}_1\|_2^2 \right) \\ &= \frac{1}{K} \sum_{k=1}^K \sum_{q=1}^Q \left\{ (y_{i^k}(q) - y_1(q))^2 - (y_{j^k}(q) - y_1(q))^2 \right\} \end{aligned} \quad (\text{A.5})$$

As the  $K$  NNs of target patch are not changed, by using Assumption 1, we have  $y_{i^k}(q) \sim N(x_1^*(q), \sigma^2)$  and  $y_1(q) \sim N(y_1^*(q), \sigma^2)$ . Similar as the calculation of  $\mathbb{E}[dt_{spe}]$ , we can obtain

$$\mathbb{E}[\|\mathbf{Y}_{i^k} - \mathbf{Y}_1\|_2^2] = 2Q\sigma^2 + \|\mathbf{X}_1^* - \mathbf{Y}_1^*\|_2^2 \quad (\text{A.6})$$

Define the distance between the target patch  $\mathbf{Y}_1$  and its really similar patches  $\mathbf{Y}_j$ ,  $j \in \mathcal{N}_{\mathbf{Y}_1}^R$  as  $d_j = \|\mathbf{Y}_j - \mathbf{Y}_1\|_2^2$ . Then, the distance of  $d_{(k)} = \|\mathbf{Y}_{j^k} - \mathbf{Y}_1\|_2^2$ ,  $j^k \in \mathcal{N}_{\mathbf{Y}_1}^R$  is the  $k$ -th order statistic of the sample  $\{d_j; j \in \mathcal{N}_{\mathbf{Y}_1}^R\}$ .

Let  $\rho_j = \frac{d_j}{2\sigma^2}$ . As  $y_j(q), y_1(q) \sim N(y_1^*(q), \sigma^2)$ , we have  $\rho_j$  is distributed according to the chi-square distribution with  $Q$  degrees of freedom as  $\rho_j \sim \chi^2(Q)$ .

Let  $\rho_{(k)} = \frac{d_{(k)}}{2\sigma^2}$ , we have

$$\sum_{k=1}^K \rho_{(k)} \leq \frac{K}{R} \sum_{k=1}^R \rho_{(k)} \quad (\text{A.7})$$

By using  $\sum_{k=1}^R \rho_{(k)} = \sum_{j \in \mathcal{N}_{\mathbf{Y}_1}^R} \rho_j$ ,  $\mathbb{E}[\rho_j] = Q$  and (A.7), we can obtain  $\mathbb{E}\left[\sum_{k=1}^K \rho_{(k)}\right] \leq KQ$  and  $\mathbb{E}\left[\sum_{k=1}^R \rho_{(k)}\right] = RQ$ . The cumulative distribution function (CDF) of  $\rho_j$  is

$$F_{\rho_j}(\rho) = \frac{\gamma(Q/2, \rho/2)}{\Gamma(Q/2)} \quad (\text{A.8})$$

where  $\gamma(s, t)$  is the lower incomplete gamma function. Because  $\rho_{(1)}$  is the minimum of  $R$  variables  $\rho_j$ , we have

$$\mathbb{E}[\rho_{(1)}] = \int_0^{+\infty} (1 - F_{\rho_j}(\rho))^R d\rho \quad (\text{A.9})$$

By using  $\sum_{k=1}^K \rho_{(k)} \geq K\rho_{(1)}$ , we can obtain

$$2K\sigma^2 \mathbb{E} [\rho_{(1)}] \leq \mathbb{E} \left[ \sum_{k=1}^K \|\mathbf{Y}_{j^k} - \mathbf{Y}_1\|_2^2 \right] \leq 2KQ\sigma^2 \quad (\text{A.10})$$

Substitute (A.6) into (A.5), we have

$$\begin{aligned} ccr(dr_{spe}) &= \frac{\mathbb{E} [dr_{spe} | \mathbf{X}_1^* \neq \mathbf{Y}_1^*]}{\mathbb{E} [dr_{spe} | \mathbf{X}_1^* = \mathbf{Y}_1^*]} \\ &= 1 + \frac{\|\mathbf{X}_1^* - \mathbf{Y}_1^*\|_2^2}{2Q\sigma^2 - \frac{1}{K} \mathbb{E} \left[ \sum_{k=1}^K \|\mathbf{Y}_{j^k} - \mathbf{Y}_1\|_2^2 \right]} \end{aligned} \quad (\text{A.11})$$

3. Compare  $ccr(dr_{spe})$  and  $ccr(dt_{spe})$ . Substitute (A.10) into (A.11) and use (A.4), we can obtain

$$ccr(dr_{spe}) - ccr(dt_{spe}) \geq \frac{\|\mathbf{X}_1^* - \mathbf{Y}_1^*\|_2^2 \mathbb{E} [\rho_{(1)}]}{2Q\sigma^2 (Q - \mathbb{E} [\rho_{(1)}])} \quad (\text{A.12})$$

Then, we can find that  $ccr(dr_{spe})$  is always larger than  $ccr(dt_{spe})$  from (A.12), and the difference between them will be greater with the increase of  $K$ . In particular, if we choose  $K = R$ , then  $\mathbb{E} \left[ \sum_{k=1}^K \|\mathbf{Y}_{j^k} - \mathbf{Y}_1\|_2^2 \right] = 2KQ\sigma^2$ , and the difference of  $ccr(dr_{spe}) - ccr(dt_{spe})$  is infinite.

### B. Case 2: the SAR images

1. Compute  $ccr(dt_{SAR})$ . Rewrite the  $dt_{SAR}$  as follows

$$dt_{SAR} = \|\log \mathbf{X}_1 - \log \mathbf{Y}_1\|_2^2 = \sum_{q=1}^Q (\log x_1(q) - \log y_1(q))^2 \quad (\text{A.13})$$

By using Assumption 2, we have  $x_1(q)/x_1^*(q)$ ,  $y_1(q)/y_1^*(q) \sim \text{Gamma}(L, L)$ . The expected value and variance value of Logarithmically transformed speckle is given in [79] as

$$\begin{aligned} \mathbb{E} \left[ \log \frac{x_1(q)}{x_1^*(q)} \right] &= \mathbb{E} \left[ \log \frac{y_1(q)}{y_1^*(q)} \right] = \psi(0, L) - \log L \\ \text{var} \left[ \log \frac{x_1(q)}{x_1^*(q)} \right] &= \text{var} \left[ \log \frac{y_1(q)}{y_1^*(q)} \right] = \psi(1, L) \end{aligned} \quad (\text{A.14})$$

where  $\psi(m, \cdot)$  is the  $m$ -order Poygamma function. Then, we have

$$\mathbb{E} \left[ \left( \log \frac{x_1(q)}{y_1(q)} \right)^2 \right] = 2\psi(1, L) + \left( \log \frac{x_1^*(q)}{y_1^*(q)} \right)^2 \quad (\text{A.15})$$

Substitute (A.15) into (A.13), we can obtain

$$\mathbb{E} [dt_{SAR}] = 2Q\psi(1, L) + \|\log \mathbf{X}_1^* - \log \mathbf{Y}_1^*\|_2^2 \quad (\text{A.16})$$

Substitute (A.16) into (27), we have

$$ccr(dt_{SAR}) = \frac{\mathbb{E} [dt_{SAR} | \mathbf{X}_1^* \neq \mathbf{Y}_1^*]}{\mathbb{E} [dt_{SAR} | \mathbf{X}_1^* = \mathbf{Y}_1^*]} = 1 + \frac{\|\log \mathbf{X}_1^* - \log \mathbf{Y}_1^*\|_2^2}{2Q\psi(1, L)} \quad (\text{A.17})$$

2. Compute  $ccr(dr_{SAR})$ . Rewrite the  $dr_{SAR}$  as follows

$$\begin{aligned} dr_{SAR} &= \frac{1}{K} \sum_{k=1}^K \|\log \mathbf{Y}_{j^k} - \log \mathbf{Y}_1\|_2^2 - \|\log \mathbf{Y}_{j^k} - \log \mathbf{Y}_1\|_2^2 \\ &= \frac{1}{K} \sum_{k=1}^K \sum_{q=1}^Q \left\{ \left( \log \frac{y_{j^k}(q)}{y_1(q)} \right)^2 - \left( \log \frac{y_{j^k}(q)}{y_1(q)} \right)^2 \right\} \end{aligned} \quad (\text{A.18})$$

As the  $K$  NNs of target patch are not changed, by using Assumption 2, we have  $y_{j^k}(q)/x_1^*(q)$ ,  $y_1(q)/y_1^*(q) \sim \text{Gamma}(L, L)$ . Similar as the calculation of  $\mathbb{E}[dt_{SAR}]$ , we can obtain

$$\mathbb{E} [\|\log \mathbf{Y}_{j^k} - \log \mathbf{Y}_1\|_2^2] = 2Q\psi(1, L) + \|\log \mathbf{X}_1^* - \log \mathbf{Y}_1^*\|_2^2 \quad (\text{A.19})$$

Define the distance between the target  $\mathbf{Y}_1$  and its really similar patches  $\mathbf{Y}_j$ ,  $j \in \mathcal{N}_{\mathbf{Y}_1}^R$  as  $d_j = \|\log \mathbf{Y}_j - \log \mathbf{Y}_1\|_2^2$ . Then the distance of  $d_{(k)} = \|\log \mathbf{Y}_{j^k} - \log \mathbf{Y}_1\|_2^2$ ,  $j^k \in \mathcal{N}_{\mathbf{Y}_1}^R$  is the  $k$ -th order statistic of the sample  $\{d_j; j \in \mathcal{N}_{\mathbf{Y}_1}^R\}$ . Let  $\rho_{j,q} = \left( \log \frac{y_j(q)}{y_1(q)} \right)^2$ . Sort  $\rho_{j,i}$  in ascending order to obtain the sequence  $\rho_{(1)}, \rho_{(2)}, \dots, \rho_{(RQ)}$ . Then, we have

$$\begin{aligned} \sum_{k=1}^K d_{(k)} &\leq \frac{K}{R} \sum_{k=1}^R d_{(k)} = \frac{K}{R} \sum_{j \in \mathcal{N}_{\mathbf{Y}_1}^R} d_j \\ \sum_{k=1}^K d_{(k)} &\geq \sum_{r=1}^{KQ} \rho_{(r)} \geq KQ\rho_{(1)} \end{aligned} \quad (\text{A.20})$$

As  $y_j(q)/y_1^*(q)$ ,  $y_1(q)/y_1^*(q) \sim \text{Gamma}(L, L)$ , we have  $y_j(q)/y_1(q)$  is distributed according to the beta prime distribution as  $y_j(q)/y_1(q) \sim \beta'(L, L)$ . Using algebraic transformations, we can obtain the CDF of  $\rho_{j,q}$  as

$$\begin{aligned} F_{\rho_{j,i}}(\rho) &= \int_{-\sqrt{\rho}}^{\sqrt{\rho}} \frac{\left( e^{-\frac{t}{2}} + e^{\frac{t}{2}} \right)^{-2L}}{B(L, L)} dt \\ &= \int_{e^{-\sqrt{\rho}}}^{e^{\sqrt{\rho}}} \frac{t^{L-1}(1+t)^{-2L}}{B(L, L)} dt \\ &= \frac{(-1)^{L-1}}{B(L, L)} \left\{ B(-e^{-\sqrt{\rho}}; L, 1-2L) \right. \\ &\quad \left. - B(-e^{\sqrt{\rho}}; L, 1-2L) \right\} \end{aligned} \quad (\text{A.21})$$

where  $B(a, b)$  represents the Beta function and  $B(x; a, b)$  represents the incomplete beta function. In particular, when  $L = 1$ , we have

$$F_{\rho_{j,i}}(\rho) = \tanh \left( \frac{\sqrt{\rho}}{2} \right); \quad \text{if } L = 1 \quad (\text{A.22})$$

Because  $\rho_{(1)}$  is the minimum of  $RQ$  variables  $\rho_{j,i}$ , we have

$$\mathbb{E} [\rho_{(1)}] = \int_0^{+\infty} (1 - F_{\rho_{j,i}}(\rho))^{RQ} d\rho \quad (\text{A.23})$$

By using  $E[\rho_{j,q}] = 2\psi(1, L)$  and (A.20), we can obtain

$$KQ\mathbb{E} [\rho_{(1)}] \leq \mathbb{E} \left[ \sum_{k=1}^K \|\log \mathbf{Y}_{j^k} - \log \mathbf{Y}_1\|_2^2 \right] \leq 2KQ\psi(1, L) \quad (\text{A.24})$$

Substitute (A.19) into (A.18), we have

$$\begin{aligned} ccr(dr_{SAR}) &= \frac{\mathbb{E}[dr_{SAR} | \mathbf{x}_1^* \neq \mathbf{y}_1^*]}{\mathbb{E}[dr_{SAR} | \mathbf{x}_1^* = \mathbf{y}_1^*]} \\ &= 1 + \frac{\|\log \mathbf{x}_1^* - \log \mathbf{y}_1^*\|_2^2}{2Q\psi(1, L) - \frac{1}{K}\mathbb{E}\left[\sum_{k=1}^K \|\log \mathbf{y}_{j^k} - \log \mathbf{y}_1\|_2^2\right]} \end{aligned} \quad (\text{A.25})$$

3. Compare  $ccr(dr_{SAR})$  and  $ccr(dt_{SAR})$ . Substitute (A.24) into (A.25) and use (A.17), we can obtain

$$ccr(dr_{SAR}) - ccr(dt_{SAR}) \geq \frac{\|\log \mathbf{x}_1^* - \log \mathbf{y}_1^*\|_2^2 \mathbb{E}[\rho_{(1)}]}{2Q\psi(1, L)(2\psi(1, L) - \mathbb{E}[\rho_{(1)}])} \quad (\text{A.26})$$

Then, we can also find that  $ccr(dr_{SAR})$  is always larger than  $ccr(dt_{SAR})$  from (A.26), and the difference between them will be greater with the increase of  $K$ . In particular, if we choose  $K = R$ , then  $\mathbb{E}\left[\sum_{k=1}^K \|\log \mathbf{y}_{j^k} - \log \mathbf{y}_1\|_2^2\right] = 2KQ\psi(1, L)$  from (A.20), and the difference of  $ccr(dr_{SAR}) - ccr(dt_{SAR})$  is infinite.

#### ACKNOWLEDGMENT

The author would like to thank the researchers for their friendly sharing of their change detection codes and data sets, which provide a wealth of resources for this study. The authors would also like to thank the editors and anonymous reviewers for their careful reading of an earlier version of this article and constructive suggestions that improved the presentation of this work.

#### 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, Jun. 1989.
- [2] Y. Ban and O. A. Yousif, "Multitemporal spaceborne SAR data for urban change detection in China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1087-1094, Aug. 2012.
- [3] F. Bovolo, C. Marin, and L. Bruzzone, "A hierarchical approach to change detection in very high resolution SAR images for surveillance applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2042-2054, Apr. 2013.
- [4] R. E. Kennedy et al., "Remote sensing change detection tools for natural resource managers: understanding concepts and tradeoffs in the design of landscape monitoring projects," *Remote Sens. Environ.*, vol. 113, no. 7, pp. 1382-1396, Jul. 2009.
- [5] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using VHR optical and SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2403-2420, May 2010.
- [6] G. Moser and S. B. Serpico, "Unsupervised change detection from multichannel SAR data by Markovian data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2114-2128, Jul. 2009.
- [7] Y. Yuan, H. Lv, and X. Lu, "Semi-supervised change detection method for multi-temporal hyperspectral images," *Neurocomputing*, vol. 148, pp. 363-375, Jan. 2015.
- [8] Y. Li, M. Gong, L. Jiao, L. Li, and R. Stolk, "Change-detection map learning using matching pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4712-4723, Aug. 2015.
- [9] L. Bruzzone and D. F. Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 452-466, Apr. 2002.
- [10] Y. F. Ban and O. Yousif, "Change detection techniques: a review," In *Multitemporal Remote Sensing*, Springer, Cham, 2016.
- [11] A. Ferretti, A. Monti-Guarnieri, C. Prati, F. Rocca, and D. Massonet, *InSAR Principles-Guidelines for SAR Interferometry Processing and Interpretation*, vol. 19. Noordwijk, The Netherlands: ESA, 2007.
- [12] O. Yousif and Y. F. Ban, "A novel approach for object-based change image generation using multitemporal high-resolution SAR images," *Int. J. Remote Sens.*, vol. 38, no. 7, pp. 1765-1787, Aug. 2017.
- [13] Z. Y. Lv, T. F. Liu, P. Zhang, J. A. Benediktsson, T. Lei, and X. Zhang, "Novel adaptive histogram trend similarity approach for land cover change detection by using bitemporal very-high-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9554-9574, Aug. 2019.
- [14] Z. Y. Lv, T. F. Liu, C. Shi, and J. A. Benediktsson, "Local histogram-based analysis for detecting land cover change using VHR remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, Early Access, 2020, doi: 10.1109/LGRS.2020.2998684.
- [15] Y.L. Chang, C.C. Han, H. Ren, C.T. Chen, K.S. Chen, and K.C. Fan, "Data fusion of hyperspectral and SAR images," *Opt. Eng.*, vol. 43, no. 8, pp. 1787-1797, Aug. 2004.
- [16] D. Lahat, T. Adali and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449-1477, Sept. 2015.
- [17] D. Hong et al., "More diverse means better: multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, Early Access, 2020, doi: 10.1109/TGRS.2020.3016820.
- [18] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, Early Access, 2020, doi: 10.1109/TGRS.2020.3015157.
- [19] G. Mercier, G. Moser, and S. B. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1428-1441, May 2008.
- [20] J. Prendes, "New statistical modeling of multi-sensor images with application to change detection," Ph.D. dissertation, Univ. Toulouse, Toulouse, France, 2015.
- [21] Y. Sun, L. Lei, X. Li, H. Sun, and G. Kuang, "Nonlocal patch similarity based heterogeneous remote sensing change detection," *Pattern Recognit.*, vol. 109, pp. 107598, Jan. 2021.
- [22] W. A. Malila, "Change vector analysis: an approach for detecting forest changes with Landsat," in *Proc. LARS symposia*, Jan. 1980, pp. 385.
- [23] F. Bovolo, S. Marchesi, and L. Bruzzone, "A framework for automatic and unsupervised detection of multiple changes in multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2196-2212, Jun. 2012.
- [24] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sens. Environ.*, vol. 64, no. 1, pp. 1-19, 1998, Apr. 1998.
- [25] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 463-478, Feb. 2007.
- [26] G. Moser and S. B. Serpico, "Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2972-2982, Oct. 2006.
- [27] F. Bovolo and L. Bruzzone, "A detail-preserving scale-driven approach to change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 12, pp. 2963-2972, Dec. 2005.
- [28] J. Ingla and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1432-1445, May 2007.
- [29] J. Ma, M. Gong, and Z. Zhou, "Wavelet fusion on ratio images for change detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 6, pp. 1122-1126, Nov. 2012.
- [30] B. Hou, Q. Wei, Y. Zheng, and S. Wang, "Unsupervised change detection in SAR image based on Gauss-log ratio image fusion and compressed projection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3297-3317, Aug. 2014.
- [31] Y. Zheng, L. Jiao, H. Liu, B. Hou, and S. Wang, "Unsupervised saliency-guided SAR image change detection," *Pattern Recognit.*, vol. 61, pp. 309-326, Jan. 2017.
- [32] Y. Zhang, S. Wang, C. Wang, J. Li, and H. Zhang, "SAR image change detection using saliency extraction and shearlet transform," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4701-4710, Dec. 2018.
- [33] R. Wang, J. W. Chen, L. Jiao, and M. Wang, "How can despeckling and structural features benefit to change detection on bitemporal SAR images?" *Remote Sens.*, vol. 11, no. 4, pp. 421, 2019.
- [34] Y. Sun, L. Lei, D. Guan, X. Li, and G. Kuang, "SAR image change detection based on nonlocal low-rank model and two-level clustering,"

- IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 293-306, 2020.
- [35] 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. Image Process.*, vol. 24, no. 3, pp. 799-812, Mar. 2015.
- [36] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "A Bayesian nonparametric model coupled with a Markov random field for change detection in heterogeneous remote sensing images," *SIAM J. Imaging Sci.*, vol. 9, no. 4, pp. 1889-1921, Jan. 2016.
- [37] V. Alberga, "Similarity measures of remotely sensed multi-sensor images for change detection applications," *Remote Sens.*, vol. 1, no. 3, pp. 122-143, Jul. 2009.
- [38] J. Ingla and A. Giros, "On the possibility of automatic multisensor image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2104-2120, Oct. 2004.
- [39] R.P. Woods, J.C. Mazziotta, and S. R. Cherry, "MRI-PET registration with automated algorithm," *J. Comput. Assist. Tomogr.*, vol. 17, no. 4, pp. 536-546, Jul. 1993.
- [40] G.J. Szekely, M.L. Rizzo, and N.K. Bakirov, "Measuring and testing dependence by correlation of distances," *Ann. Stat.*, vol. 35, no. 6, pp. 2769-2794, Dec. 2007.
- [41] L. Wan, T. Zhang, and H. J. You, "Multi-sensor remote sensing image change detection based on sorted histograms," *Int. J. Remote Sens.*, vol. 39, no. 11, pp. 3753-3775, Jun. 2018.
- [42] B. Ayhan and C. Kwan, "A new approach to change detection using heterogeneous images," in *Proc. IEEE UEMCON*, Oct. 2019, pp. 0192-0197.
- [43] C. Kwan, B. Ayhan, J. Larkin, L. Kwan, S. Bernab, and A. Plaza, "Performance of change detection algorithms using heterogeneous images and extended multi-attribute profiles (EMAPs)," *Remote Sens.*, vol. 11, no. 20, pp. 2377, Oct. 2019.
- [44] L. T. Luppino, F. M. Bianchi, G. Moser, and S. N. Anfinsen, "Unsupervised image regression for heterogeneous change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9960-9975, Dec. 2019.
- [45] L. T. Luppino et al., "Deep image translation with an affinity-based change prior for unsupervised multimodal change detection," *arXiv preprint*, arXiv: 2001.04271, 2020.
- [46] Z. Liu, G. Li, G. Mercier, Y. He, and Q. Pan, "Change detection in heterogenous remote sensing images via homogeneous pixel transformation," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1822-1834, Apr. 2018.
- [47] Y. Sun, L. Lei, X. Li, X. Tan, and G. Kuang, "Patch similarity graph matrix based unsupervised remote sensing change detection with homogeneous and heterogeneous sensors," *IEEE Trans. Geosci. Remote Sens.*, Early Access, 2020, doi: 10.1109/TGRS.2020.3013673.
- [48] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: a general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3-13, Jan. 2019.
- [49] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised deep change vector analysis for multiple-change detection in VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3677-3693, Jun. 2019.
- [50] S. Saha, L. Mou, X. X. Zhu, F. Bovolo, and L. Bruzzone, "Semisupervised change detection using graph convolutional network," *IEEE Geosci. Remote Sens. Lett.*, Early Access, 2020, doi: 10.1109/LGRS.2020.2985340.
- [51] H. C. Li, G. Yang, W. Yang, Q. Du, and W. J. Emery, "Deep nonsmooth nonnegative matrix factorization network with semi-supervised learning for SAR image change detection," *ISPRS-J. Photogramm. Remote Sens.*, vol. 160, pp. 167-179, Feb. 2020.
- [52] F. Gao, J. Dong, B. Li, and Q. Xu, "Automatic change detection in synthetic aperture radar images based on PCANet," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1792-1796, Dec. 2016.
- [53] F. Gao, X. Wang, Y. Gao, J. Dong, and S. Wang, "Sea ice change detection in SAR images based on convolutional-wavelet neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1240-1244, Aug. 2019.
- [54] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 545-559, Mar. 2018.
- [55] X. Niu, M. Gong, T. Zhan, and Y. Yang, "A conditional adversarial network for change detection in heterogeneous images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 45-49, Jan. 2019.
- [56] R. Touati, M. Mignotte, and M. Dahmane, "Anomaly feature learning for unsupervised change detection in heterogeneous images: a deep sparse residual model," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 588-600, 2020.
- [57] F. Thonfeld, H. Feilhauer, M. Braun and G. Menz, "Robust change vector analysis (RCVA) for multi-sensor very high resolution optical satellite data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 50, pp. 131-140, Aug. 2016.
- [58] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. Zhu, "Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS-J. Photogramm. Remote Sens.*, vol. 147, pp. 193-205, Jan. 2019.
- [59] L. Lei, Y. Sun, and G. Kuang, "Adaptive local structure consistency based heterogeneous remote sensing change detection," *IEEE Geosci. Remote Sens. Lett.*, Early Access, 2020, doi: 10.1109/LGRS.2020.3037930.
- [60] W. Dong, C. Moses and K. Li, "Efficient k-nearest neighbor graph construction for generic similarity measures," in *Proc. 20th Int. Conf. World Wide Web*, Mar. 2011, pp. 577-586.
- [61] kggraph: <http://www.kgraph.org/>.
- [62] A. Gionis, P. Indyk, and R. Motwani, "Similarity search in high dimensions via hashing," in *Proc. 30th Int. Conf. Very Large Data Bases*, 1999, pp. 518-529.
- [63] D. C. Anastasiu and G. Karypis, "L2Knng: Fast exact k-nearest neighbor graph construction with L2-norm pruning," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, Oct. 2015, pp. 791-800.
- [64] G. Piella, "A general framework for multiresolution image fusion from pixels to regions," *Inf. Fusion*, vol. 4, no. 4, pp. 259-280, Dec. 2003.
- [65] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, "Contrast-based fusion of noisy images using discrete wavelet transform," *IET Image Process.*, vol. 4, no. 5, pp. 374-384, 2010.
- [66] C.-A. Deledalle, L. Denis, and F. Tupin, "How to compare noisy patches? patch similarity beyond gaussian noise," *Int. J. Comput. Vis.*, vol. 99, no. 1, pp. 86-102, Mar. 2012.
- [67] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729-4743, Aug. 2014.
- [68] J. W. Goodman, "Some fundamental properties of speckle," *J. Opt. Soc. Amer.*, vol. 66, no. 11, pp. 1145-1150, Nov. 1976.
- [69] C. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jager, "NL-SAR: a unified nonlocal framework for resolution-preserving (Pol)(In)SAR denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2021-2038, Apr. 2015.
- [70] C.-A. Deledalle, L. Denis, S. Tabti, and F. Tupin, "MuLoG, or how to apply Gaussian denoisers to multi-channel SAR speckle reduction?" *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4389-4403, Sept. 2017.
- [71] S. N. Anfinsen, A. P. Doulgeris, and T. Eltoft, "Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3795-3809, Nov. 2009.
- [72] M. Gong, Y. Cao, and Q. Wu, "A neighborhood-based ratio approach for change detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 307-311, Mar. 2012.
- [73] F. Nar, A. Ozgur, and A. N. Saran, "Sparsity-driven change detection in multitemporal SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 7, pp. 1032-1036, Jul. 2016.
- [74] A. Mian, G. Ginolhac, J. Ovarlez, and A. M. Atto, "New robust statistics for change detection in time series of multivariate SAR images," *IEEE Trans. Signal Process.*, vol. 67, no. 2, pp. 520-534, Jan. 2019.
- [75] D. Ciunzo, V. Carotenuto, and A. De Maio, "On multiple covariance equality testing with application to SAR change detection," *IEEE Trans. Signal Process.*, vol. 65, no. 19, pp. 5078-5091, Oct. 2017.
- [76] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9976-9992, Dec. 2019.
- [77] M. Mignotte, "A fractal projection and Markovian segmentation-based approach for multimodal change detection," *IEEE Trans. Geosci. Remote Sens.*, Early Access, 2020, doi: 10.1109/TGRS.2020.2986239.
- [78] M. Volpi, G. Camps-Valls, and D. Tuia, "Spectral alignment of multitemporal cross-sensor images with automated kernel canonical correlation analysis," *ISPRS-J. Photogramm. Remote Sens.*, vol. 107, pp. 50-63, Sept. 2015.
- [79] H. Xie, L.E. Pierce, and F.T. Ulaby, "Statistical properties of logarithmically transformed speckle," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 3, pp. 721-727, Mar. 2002.