



# Nonlocal patch similarity based heterogeneous remote sensing change detection

Yuli Sun, Lin Lei\*, Xiao Li, Hao Sun, Gangyao Kuang

*College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China*



## ARTICLE INFO

### Article history:

Received 28 April 2020

Revised 17 July 2020

Accepted 16 August 2020

Available online 17 August 2020

### Keywords:

Unsupervised change detection

Heterogeneous data

Nonlocal similarity

Graph

## ABSTRACT

Change detection of heterogeneous remote sensing images is an important and challenging topic, which has found a wide range of applications in many fields, especially in the emergency situation resulting from nature disaster. However, the difference in imaging mechanism of heterogeneous sensors makes it difficult to carry out a direct comparison of images. In this paper, we propose a new change detection method based on similarity measurement between heterogeneous images. The method constructs a graph for each patch based on the nonlocal patch similarity to establish a connection between heterogeneous data, and then measures the change level by measuring how much the graph structure of one image still conforms to that of the other image. The graph structures are compared in the same domain, so it can avoid the leakage of heterogeneous data and bring more robust change detection results. Experiments demonstrate the effective performance of the proposed nonlocal patch similarity based heterogeneous change detection method.

© 2020 Elsevier Ltd. All rights reserved.

## 1. Introduction

### 1.1. Background

The change detection (CD) of remote sensing images is a process of identifying the changes of objects or phenomenon in the same geographical area at different times by analyzing the differences between images [1]. CD has been widely used in many real-world applications, such as land-use and land-cover evaluation [2], urban growth monitoring [3,4], nature disaster assessment [5,6], etc.

CD with homogeneous images, i.e. image collected by the same kind of sensors, e.g., radar or optical sensors, has been of interest for a long time [7]. Researchers have proposed many relatively mature algorithms for homogeneous CD, such as the change vector analysis (CVA) [8], compressed change vector analysis ( $C^2VA$ ) [9], multivariate alteration detection (MAD) [10], iteratively reweighted MAD (IR-MAD) [11], the generalized Kittler and Illingworth thresholding (GKIT) algorithm [12], reformulated fuzzy local information C-means clustering algorithm (RFLICM) [13], principal component analysis with K-means clustering (PCA-KM) [14], Gabor wavelets with two level fuzzy c-means clustering (GaborTLC) [15]. However, the wide range of different sensors found in remote sensing makes

the detection of changes in images acquired by heterogeneous sensors a growing interest topic. In particular, the heterogeneous CD has great practical significance for the immediate evaluation and emergency disasters. In such scenarios (e.g., earthquake or flood), the rapid mapping of damages is needed. The pre-event SAR image is sometimes unavailable and the pre-event optical image can be obtained from the archived data of remote sensing platforms, whereas maybe only the post-event SAR image can be available due to adverse atmospheric conditions.

However, heterogeneous CD is very challenging because of the distinct feature representations of ground object in images acquired by different sensors, especially for these images obtained from optical and SAR sensors respectively. The optical sensor is passive, which can measure the intensity of reflected light in visible and near-infrared spectral bands. Therefore, the optical image reflects the surface reflection and illumination information of the object. SAR sensor is active, which can measure the backscatter characteristics of objects by transmitting radar waves. Therefore, the SAR image reflects the geometric and dielectric characteristics of the target. As the images acquired by heterogeneous sensors show different physical quantities and different statistical behaviors, it is difficult to calculate pixelwise difference between heterogeneous images. This is different from that in homogeneous images, which only requires a simple arithmetical operation such as image differencing (usually for optical images) and image ratio/log-ratio (usually for SAR images).

\* Corresponding author.

E-mail address: [alaleilin@163.com](mailto:alaleilin@163.com) (L. Lei).

Generally, according to whether the label information is used, the existing heterogeneous CD methods can be divided into supervised and unsupervised. Meanwhile, it can also be divided into three categories according to the different methods for generating the binary change map. The first is the method based on classification. These methods first classify the images separately and transform the heterogeneous images into the same category space, then compare the classification results to detect the changes. Among them, post-classification comparison (PCC) is the most widely used method [16,17]. The accuracy of PCC strongly depends on the performance of classification, and the PCC may suffer from the error propagation or error accumulation [18]. Recently, the PCC based on multitemporal segmentation and compound classification (MS-CC) [19] and its extended version based on cooperative multitemporal segmentation and hierarchical compound classification (CMS-HCC) [20] have been proposed to overcome the error propagation. However, as shown in these papers, image segmentation will affect the accuracy of CD, in particular, it is well known that the SAR image segmentation is very difficult, which brings a big challenge to its accurate detection.

The second is the method based on deep learning. Some of them use the Convolutional Neural-Networks (CNNs) and stacked denoising autoencoders (SDAEs) to extract the high-level feature representation and explore the inner relationships of heterogeneous images, such as the symmetric convolutional coupling network (SCCN) [21], and logarithmic transformation feature learning (LTFL) with SDAE [22]. Some of them use the translation network to translate the two heterogeneous images into homogeneous ones, between which the difference image can be obtained in a common observation space, such as the conditional generative adversarial network (cGAN) [23], the X-Net with two fully convolutional networks and the Adversarial Cyclic Encoders Network (ACE-Net) with two autoencoders whose code spaces are aligned by adversarial training [24]. Despite their excellent performance on the detection accuracy, deep learning-based method still has two major flaws: the time-consuming training process and the construction of large training set, which requires a high cost of manual operation in practice under the supervision mode or requires a complicated screening process to select the training samples under the unsupervision mode.

The third is the method based on similarity measure. Such methods usually define a function to measure the difference between images by using a sliding or analysis window. Mercier et al. use the copula theory to model the dependence between unchanged areas and then employ the Kullback-Leibler (KL) distance on local statistical measures to calculate the changes [25]. Prendes et al. propose a multivariate statistical approach, which models the objects contained in analysis window by local joint distributions and then uses the manifold to measure the change indices [26]. In [27], they further introduce a Bayesian nonparametric framework to deal with the unknown number of objects in the analysis window. This kind of method based on parameter estimation requires an explicit data distribution, a complex parameter estimation and a large amount of training data. In [28], a series of similarity measures are employed for automatic CD of optical and SAR images, such as the distance to independence, mutual information, cluster reward algorithm [29], Woods criterion [30], robust Woods criterion [31], and correlation ratio [32]. By assuming that the heterogeneous images with absence of change have similar local internal layouts, the distance of sorted histogram (SH) is employed to estimate the dissimilarity between the images [33]. The pixels pair (PP) method assumes that the mapping between the pixel values of images in the image pair are monotonic, then it computes differences between pixels in each image separately, and the difference scores are then compared between images in the pair to generate the change map [34,35]. Luppino et al. propose the affinity ma-

trices distance (AMD) method to calculate the change possibility of each pixel, which can be used to generate the CD map directly by thresholding operation, and can be further used to construct the pseudo-training data for the unsupervised traditional image regression based CD [36] or can be treated as the change prior to guide the deep image translation based heterogeneous CD [24]. The main advantages of these imaging modality-invariant operator based methods (such as SH, PP and AMD) are intuitive, unsupervised and easy to implement. However, they do not take into account the different statistical characteristics of the heterogeneous images when the scene is complex or the noise in the image is very high (especially the speckle noise of SAR image), these designed operators can not fully represent the similarity (or difference) between heterogeneous images, resulting in a sharp deterioration in CD performance.

From the above discussion, we can see that the goal of these heterologous CD methods is to transform the heterogeneous images to a space where they can be directly compared, such as the category space of classification based methods, the learned high-dimensional feature space of deep learning based methods, and the constructed feature space of similarity measure based methods. When direct comparison is possible, methods for homogeneous CD can be used to extract the changes.

## 1.2. Contribution

In this paper, we propose a nonlocal self-similarity based method for heterogeneous CD, which belongs to the family of similarity-based methods. Although the heterogeneous images refer to completely different imaging mechanisms, and hence, there is no direct relationship between their pixel values, they share important structure information, which can be used to detect the changes between images. To exploit this structure information, we use the nonlocal patch similarity based graph (NPSG) to measure the structure consistency between heterogeneous images. The proposed CD method mainly consists of two parts. First, the image is divided into a number of overlapped squared patches. In the forward detection, for each target patch in the pre-event image, its  $K$ -nearest NPSG within the pre-event image is calculated by using the statistical based similarity measurement, then we map this  $K$ -nearest NPSG to the post-event image, and compare its own  $K$ -nearest NPSG of the post-event image with this mapped  $K$ -nearest NPSG from the pre-event image by calculating the similarity difference. If changes occur within this patch area, then the graph structure of this patch will change. Intuitively, the more changes occur in the patch, the more the two  $K$ -nearest NPSGs will diverge. After that, we also similarly do the backward detection by mapping the  $K$ -nearest NPSG of post-event image to the pre-event image. Then the forward and backward detections are combined to further improve the robustness of CD with respect to the noise and modality difference of images. Second, with combined difference image, the final binary change map (CM) can be treated as an image binary segmentation problem, which can be solved by the thresholding methods such as Otsu threshold [37], or clustering methods such as the PCAKM [14] and GaborTLC [15] for the homogeneous CD.

The core idea of the proposed method is that we use the NPSG to represent the image structure, then the change level depends on how much one image still conforms to the NPSG from the other image. Meanwhile, there are two points should be noted. First, to measure the structure differences of the post-event (pre-event) target patch, we only use the location information of the mapped  $K$ -nearest NPSG from the pre-event (post-event) images, then the structure difference calculation is only carried out within this post-event (pre-event) image by comparing the similarity difference between the target patch and patches in the two  $K$ -nearest NPSG. It is worth underlining that this operation avoids the leakage of hetero-

geneous data, which exists in different degrees in other similarity based heterogeneous CD methods such as SH, PP, and AMD. Second, to make the fusion process of forward detection and backward detection more reasonable and accurate, we calculate the patch similarity based on the image statistical distribution and analyze the characteristics of these similarities, and make the expectations of forward and backward similarity difference balanced. The main contributions of this work are summarized as follows:

1) We propose a novel CD framework with constructed graph based on nonlocal patch similarity, which makes the heterogeneous data comparable. This method measures the change level between the pre- and post-event images by measuring how much the graph structure of one image still conforms to that of the other image.

2) The proposed method avoids the leakage of heterogeneous data and reasonably combines the forward and backward detection results based on the statistical distribution, which can efficiently reduce the false alarms in CD of heterogeneous images.

3) The proposed heterogeneous CD framework is completely unsupervised, its parameters are easy to determine, and it exhibits a good level of usability as shown in the experiments.

### 1.3. Outline

The overall structure of this study takes the form of four sections, including this introductory section. [Section 2](#) introduces the proposed method. [Section 3](#) presents the numerical results. In the end, we provide our conclusion in [Section 4](#).

## 2. Nonlocal patch similarity based heterogeneous CD

We consider two co-registered images that represent the same region acquired by heterogeneous sensors before (time  $t_1$ ) and after an event (time  $t_2$ ), which are denoted as  $\mathbf{X} = \{x(m, n, c) | 1 \leq m \leq M, 1 \leq n \leq N, 1 \leq c \leq C_X\}$  lining in the domain  $\mathcal{X}$  and  $\mathbf{Y} = \{y(m, n, c) | 1 \leq m \leq M, 1 \leq n \leq N, 1 \leq c \leq C_Y\}$  lining in the domain  $\mathcal{Y}$ , respectively. Here  $M$  and  $N$  are the height and the width of two images,  $C_X$  and  $C_Y$  are the number of channels of two images. As mentioned in [Section 1](#), the heterogeneous sensors take completely different imaging mechanisms, it is meaningless to directly compare their signal values for detecting changes. An intuitive method is to find a relationship between the heterogeneous data and make them comparable, which makes the unchanged areas and changed areas show completely different forms.

By exploiting image self-similarity, a small patch in the image can always find some very similar patches in an extended search window (or the whole image) based on the patch-wise similarity. This self-similarity is also the basis of the widely used nonlocal SAR image despeckling algorithms, such as probabilistic patch-based (PPB) algorithm [38], SAR block-matching 3-D (SAR-BM3D) [39], fast adaptive nonlocal SAR despeckling (FANS) [40], and it has also been applied to SAR image change detection [41]. On the other hand, this kind of nonlocal patch similarity also represents the structural information of the image, which can establish the relationship between heterogeneous images. As shown in [Fig. 1](#), in the SAR image  $\mathbf{X}$ , a small target patch  $T^{\mathbf{X}}$  is connected with several similar patches  $S^{\mathbf{X}}$ . If the objects (areas) represented by these patches do not change between  $t_1$  and  $t_2$ , this nonlocal similarity structure can be followed by the optical image  $\mathbf{Y}$ . This means that the patch  $T^{\mathbf{Y}}$  is also very similar to patches  $S^{\mathbf{Y}}$  in the optical image  $\mathbf{Y}$ , where the positions of  $T^{\mathbf{Y}}$  and  $S^{\mathbf{Y}}$  in  $\mathbf{Y}$  are the same as that of the  $T^{\mathbf{X}}$  and  $S^{\mathbf{X}}$  in  $\mathbf{X}$ . If there is a change in the area represented by  $T^{\mathbf{X}}$ , we can find that this nonlocal similarity structure is no longer conformed and the patch  $T^{\mathbf{Y}}$  is very different from the patches  $S^{\mathbf{Y}}$  in the image  $\mathbf{Y}$ , as shown in target patch 2 in [Fig. 1](#). From this illustration, we can find that the change level in the heterogeneous

images can be measured by how dissimilar between the nonlocal patch similarity structures of two images.

The proposed method consists of three steps: 1) construct NPSG for each target patch; 2) compare the similarity structure and generate the difference image; 3) generate the binary change map with thresholding or clustering method. The framework is illustrated in [Fig. 2](#).

### 2.1. Nonlocal patch similarity based graph

Let us consider a square target patch,  $\mathbf{X}_{(m,n)} = \{x(m + \vartheta_m, n + \vartheta_n, c); (\vartheta_m, \vartheta_n) \in \mathbb{P}, 1 \leq c \leq C_X\}$  with the size  $p_s \times p_s \times C_X$ , where  $(m, n)$  is the position of the anchor pixel (for example, the patch center),  $\mathbb{P}$  indicates the set of square spatial offsets with respect to the anchor pixel and its cardinality is  $|\mathbb{P}| = p_s^2$ .

As the graph model can efficiently capture crucial information and local structure of an image, it is an effective tool for image representation and analysis [42,43], which has also been used in SAR image change detection using a pointwise approach [44]. Here, we propose to construct a weighted graph  $G$  to represent the geometric structure for each target patch. Thus, denoting the graph  $G_{\mathbf{X}_{(m,n)}} = \{V_{\mathbf{X}_{(m,n)}}, E_{\mathbf{X}_{(m,n)}}, w_{\mathbf{X}}\}$  for each target patch  $\mathbf{X}_{(m,n)}$ , we construct the NPSG within a  $\xi_s \times \xi_s$  search window  $\mathbb{W}$  centered on this  $\mathbf{X}_{(m,n)}$  as

$$\begin{aligned} V_{\mathbf{X}_{(m,n)}} &= \{\mathbf{X}_{(i,j)}; (i + \vartheta_i, j + \vartheta_j) \in \mathbb{W}, (\vartheta_i, \vartheta_j) \in \mathbb{P}\}, |V_{\mathbf{X}_{(m,n)}}| = (\xi_s - p_s)^2 \\ E_{\mathbf{X}_{(m,n)}} &= \{(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}); \mathbf{X}_{(i,j)} \in V_{\mathbf{X}_{(m,n)}}\} \\ w(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}) &= \exp(-\lambda d(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)})), (\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}) \in E_{\mathbf{X}_{(m,n)}} \end{aligned} \quad (1)$$

where the term  $d(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)})$  represents a distance measure of two vertices  $\mathbf{X}_{(m,n)}$  and  $\mathbf{X}_{(i,j)}$ , and  $\lambda > 0$  is a parameter controlling the bandwidth of the exponential kernel. Within this graph  $G_{\mathbf{X}_{(m,n)}} = \{V_{\mathbf{X}_{(m,n)}}, E_{\mathbf{X}_{(m,n)}}, w_{\mathbf{X}}\}$ , each patch in the search window  $\mathbb{W}$  becomes a vertex, and each vertex  $\mathbf{X}_{(i,j)}$  is connected with the target vertex  $\mathbf{X}_{(m,n)}$  by a set of edges  $E_{\mathbf{X}_{(m,n)}}$ , and their associate weights  $w$  involving a measure of similarity between each vertex and the target  $\mathbf{X}_{(m,n)}$ . In this case, the structure information of the patch  $\mathbf{X}_{(m,n)}$  can be characterized by this graph. To exploring this nonlocal patch similarity based graph, we need to calculate the distance  $d(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)})$  between the target vertex  $\mathbf{X}_{(m,n)}$  and its all neighbors  $\mathbf{X}_{(i,j)}$  within the search window  $\mathbb{W}$ .

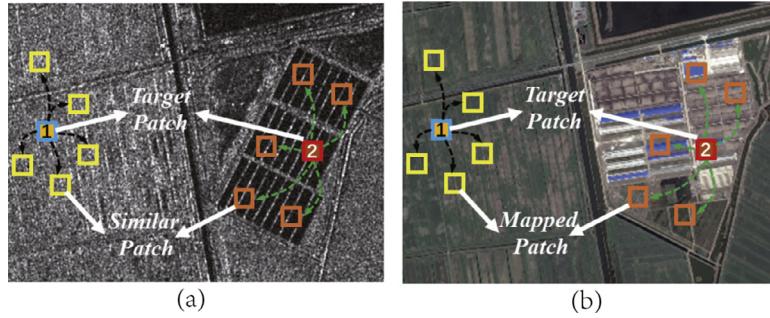
For the optical image, assuming the additive white Gaussian noise (AWGN) model,  $x = z + u$ , with  $z$  being the true signal value and  $u$  being a zero mean Gaussian random variable with constant variance  $\sigma^2$ , i.e.,  $u \sim \mathcal{N}(0, \sigma^2)$ . Under this AWGN model, the traditional Euclidean distance is usually used, which has been proved to be optimal under most criteria of interest [45]. Here we add a normalized parameter  $\gamma_\sigma$  (described in more detail in [Section 2.4](#) and in the Appendix) to the Euclidean distance as

$$d_{opt}(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}) = \frac{\gamma_\sigma}{C_X p_s^2} \|\mathbf{X}_{(m,n)} - \mathbf{X}_{(i,j)}\|_F^2 \quad (2)$$

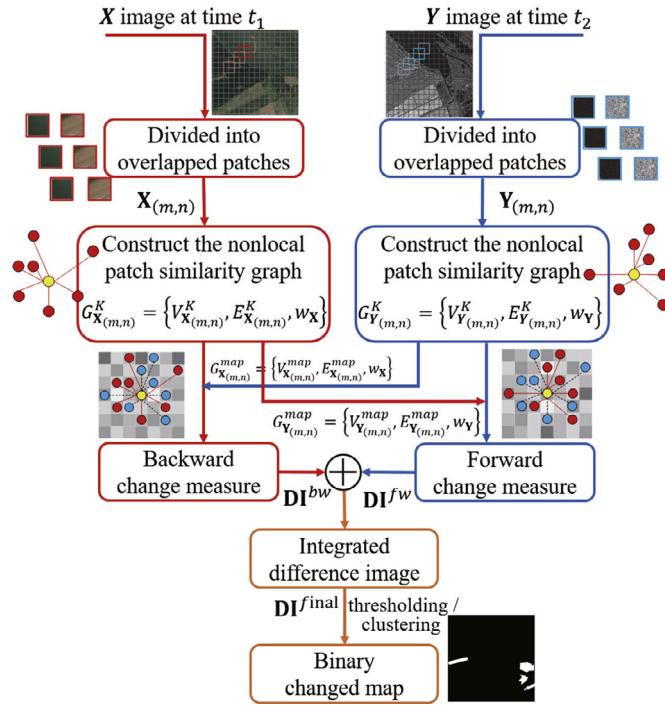
The normalized parameter  $\gamma_\sigma$  can make the expectation of distance  $E[d_{opt}(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)})] = 1$  when  $\mathbf{X}_{(m,n)}$  and  $\mathbf{X}_{(i,j)}$  are two realizations of the same true value (noise-free patch).

For the SAR image, assuming the usual multiplicative model,  $x = zu$ , with  $z$  being the true intensity value and  $u$  being a unit-mean Gamma random variable modeling the speckle. The probability density function (pdf) of  $u$  is given by

$$p_u(u) = \frac{L^L}{\Gamma(L)} u^{L-1} e^{-Lu}, u \geq 0 \quad (3)$$



**Fig. 1.** Illustration of nonlocal patch similarity in heterogeneous images. The target patches 1 and 2 are connected with several similar patches in the SAR image (a). Target patch 1 is also similar to the mapped patches in optical image (b), which means that the similarity structure in SAR image is followed by the unchanged target patch 1 and its mapped patches in optical image. However, the similarity structure is no longer conformed by the changed target patch 2 and its mapped patches, where they show a lot of difference.



**Fig. 2.** Framework of the NPSG based heterogeneous change detection method.

where  $\Gamma(\cdot)$  denotes the Gamma function, and  $L$  is the equivalent number of looks (ENL) that describes the degree of averaging applied to the SAR measurements during data formation and postprocessing. Here we use the following two types of distances

$$\begin{aligned} d_{\text{SAR}}^{(1)}(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}) \\ = \frac{\gamma_L^{(1)}}{C_x p_s^2} \sum_{c=1}^{C_x} \sum_{(\vartheta_m, \vartheta_n) \in \mathbb{P}} \\ \times \log \left[ \frac{x(m + \vartheta_m, n + \vartheta_n, c) + x(i + \vartheta_m, j + \vartheta_n, c)}{2\sqrt{x(m + \vartheta_m, n + \vartheta_n, c)x(i + \vartheta_m, j + \vartheta_n, c)}} \right] \end{aligned} \quad (4)$$

$$d_{\text{SAR}}^{(2)}(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}) = \frac{\gamma_L^{(2)}}{C_x p_s^2} \sum_{c=1}^{C_x} \sum_{(\vartheta_m, \vartheta_n) \in \mathbb{P}} \left[ \log \left( \frac{x(i + \vartheta_m, j + \vartheta_n, c)}{x(m + \vartheta_m, n + \vartheta_n, c)} \right) \right]^2 \quad (5)$$

where  $\gamma_L^{(1)}$  and  $\gamma_L^{(2)}$  are normalized parameters similar to  $\gamma_\sigma$  (described in more detail in Section 2.4 and in the Appendix). The first

type distance (4) has also been used in [46]. Intuitively,  $d_{\text{SAR}}^{(1)}$  can be regarded as the normalized sum of logarithmic ratio between arithmetic and geometric means of the signal intensities and  $d_{\text{SAR}}^{(2)}$  can be regarded as the normalized sum squares of logarithmic ratio between signal intensities.

With these defined patch distances, the edge weight between  $\mathbf{X}_{(m,n)}$  and  $\mathbf{X}_{(i,j)}$  is generated by using the Gaussian kernel type similarity criterion

$$w(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}) = \exp(-\lambda d(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)})) \quad (6)$$

where  $d$  corresponds  $d_{\text{opt}}$ ,  $d_{\text{SAR}}^{(1)}$  and  $d_{\text{SAR}}^{(2)}$  according to different situations, respectively. Then for each patch in the pre-event image  $\mathbf{X}$  and the post-event image  $\mathbf{Y}$ , we can construct its NPSG by this operation.

## 2.2. Calculating the difference image

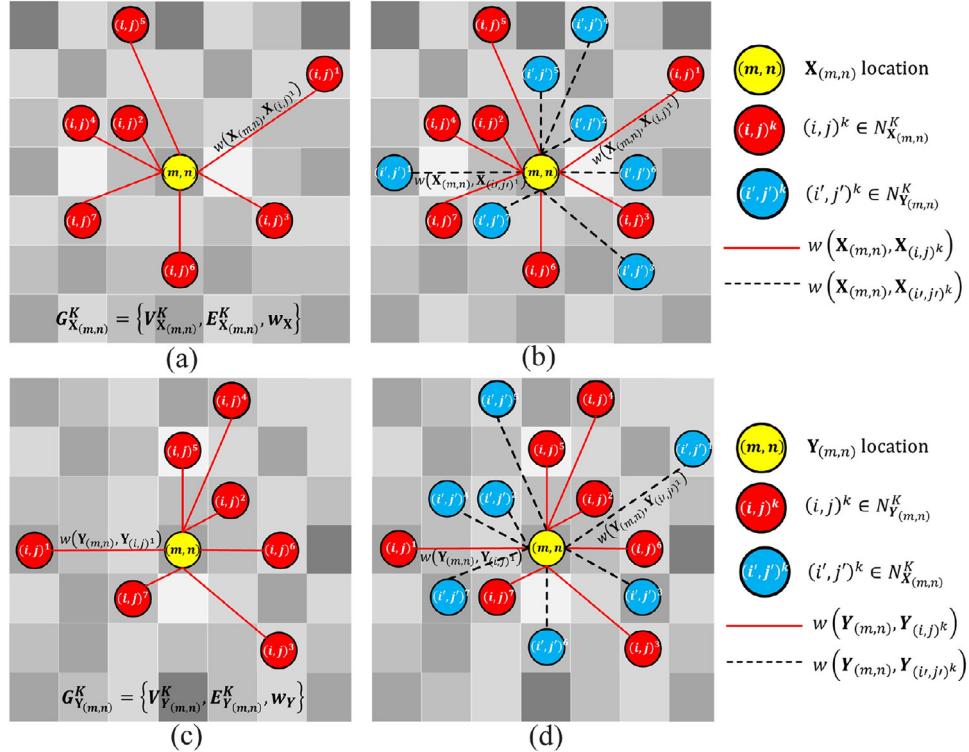
As the graph  $G$  contains rich robust structure information, we can use this to detect the changes between heterogeneous images, which means that we need to compare the difference between the graph  $G_{\mathbf{X}_{(m,n)}}$  of the pre-event  $\mathbf{X}_{(m,n)}$  and the graph  $G_{\mathbf{Y}_{(m,n)}}$  of the post-event  $\mathbf{Y}_{(m,n)}$ . A simple and intuitive idea is to directly compare the weighted graph as follows

$$\begin{aligned} f_{(m,n)} = \sum_{(i,j)} |w(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)})h(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}) \\ - w(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)})h(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)})| \end{aligned} \quad (7)$$

where  $h(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)})$  is the function on graph vertices  $\mathbf{X}_{(m,n)}$  and  $\mathbf{X}_{(i,j)}$ , such as the simplest case  $h(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}) = 1$ .

Although this difference criterion is simple and easy to understand, just as it directly compares the patch similarities between all the neighborhoods, there is a serious risk. The operation of direct similarity (or weighted similarity) subtraction in (7) will cause confusion of similarities and leakage of heterogeneous data. The “confusion” means that all the neighbors are used in computation of structure difference of (7), which may make the measure less discriminative. The “leakage” means that the two terms directly subtracted in (7) are generated based on different images, that is, they come from different domains  $\mathcal{X}$  and  $\mathcal{Y}$ , which should not be directly compared.

Through further observation, we can find that the structural information of each patch is concentrated on its  $K$ -NN ( $K$  most similar neighbors). Then we construct the  $K$ -nearest NPSG  $G_{\mathbf{X}_{(m,n)}}^K =$



**Fig. 3.** The process of constructing and mapping  $K$ -nearest NPSG. (a)  $K$ -nearest NPSG  $G_{X_{(m,n)}}^K$  for the target patch  $\mathbf{X}_{(m,n)}$ ; (b)  $G_{X_{(m,n)}}^K$  and the mapped  $G_{X_{(m,n)}}^{map}$ ; (c)  $K$ -nearest NPSG  $G_{Y_{(m,n)}}^K$  for the target patch  $\mathbf{Y}_{(m,n)}$ ; (d)  $G_{Y_{(m,n)}}^K$  and the mapped  $G_{Y_{(m,n)}}^{map}$ .  $G_{X_{(m,n)}}^{map}$  is obtained by mapping the  $G_{Y_{(m,n)}}^K$  to the image  $\mathbf{X}$ , the vertex positions of  $G_{X_{(m,n)}}^{map}$  and  $G_{Y_{(m,n)}}^K$  are the same, but these two graphs are constructed in different domains. The same is true for  $G_{Y_{(m,n)}}^{map}$  and  $G_{X_{(m,n)}}^K$ .

$\left\{ V_{\mathbf{X}_{(m,n)}}^K, E_{\mathbf{X}_{(m,n)}}^K, w_{\mathbf{X}} \right\}$  of each target patch  $\mathbf{X}_{(m,n)}$  as

$$\begin{aligned} V_{\mathbf{X}_{(m,n)}}^K &= \left\{ \mathbf{X}_{(i,j)} ; (i + \vartheta_i, j + \vartheta_j) \in \mathbb{W}, (\vartheta_i, \vartheta_j) \in \mathbb{P}, (i, j) \in \mathcal{N}_{\mathbf{X}_{(m,n)}}^K \right\}, \\ \left| V_{\mathbf{X}_{(m,n)}}^K \right| &= K \\ E_{\mathbf{X}_{(m,n)}}^K &= \left\{ (\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}) ; \mathbf{X}_{(i,j)} \in V_{\mathbf{X}_{(m,n)}}^K \right\} \\ w(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}) &= \exp(-\lambda d(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)})), \forall (\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)}) \in E_{\mathbf{X}_{(m,n)}}^K \end{aligned} \quad (8)$$

where  $\mathcal{N}_{\mathbf{X}_{(m,n)}}^K$  represents the anchor pixel position set of the  $K$ -NN of  $\mathbf{X}_{(m,n)}$  by sorting the distances  $d(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)})$  and taking out the  $K$  smallest  $\mathbf{X}_{(i,j)}$ .

Then, instead of directly comparing the similarity (distance) vectors between the heterogeneous images, we first map the  $G_{\mathbf{X}_{(m,n)}}^K$  to the post-event image  $\mathbf{Y}$  to obtain the mapped graph  $G_{\mathbf{Y}_{(m,n)}}^{\text{map}} = \left\{ V_{\mathbf{Y}_{(m,n)}}^{\text{map}}, E_{\mathbf{Y}_{(m,n)}}^{\text{map}}, w_{\mathbf{Y}} \right\}$  as

$$\begin{aligned} V_{\mathbf{Y}_{(m,n)}}^{map} &= \left\{ \mathbf{Y}_{(i,j)} ; (i + \vartheta_i, j + \vartheta_j) \in \mathbb{W}, (\vartheta_i, \vartheta_j) \in \mathbb{P}, (i, j) \in \mathcal{N}_{\mathbf{X}_{(m,n)}}^K \right\}, \\ |V_{\mathbf{X}_{(m,n)}}^{map}| &= K \\ E_{\mathbf{Y}_{(m,n)}}^{map} &= \left\{ (\mathbf{Y}_{(m,n)}, \mathbf{X}_{(i,j)}) ; \mathbf{Y}_{(i,j)} \in V_{\mathbf{Y}_{(m,n)}}^{map} \right\} \\ w(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)}) &= \exp(-\lambda d(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)})), \forall (\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)}) \in E_{\mathbf{Y}_{(m,n)}}^{map} \end{aligned} \quad (9)$$

We can easily find that the mapped graph  $G_{Y_{(m,n)}}^{map}$  is a subgraph of  $G_{Y_{(m,n)}}$ . The process of constructing and mapping  $K$ -nearest NPSG is shown Fig. 3.

Then we compare its own  $K$ -nearest graph  $G_{\mathbf{Y}_{(m,n)}}^K$  (similar to (8)) with this mapped  $G_{\mathbf{Y}_{(m,n)}}^{map}$  within the post-event image  $\mathbf{Y}$  for the

forward detection. We calculate the structure difference by using the similarity or distance criteria as

$$f_{(m,n)}^{\mathbf{Y}} = \frac{1}{K} \sum_{k=1}^K \left| w\left(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)^k}\right) h\left(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)^k}\right) - w\left(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i',j')^k}\right) h\left(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i',j')^k}\right) \right| \quad (10)$$

$$f_{(m,n)}^Y = \frac{1}{K} \sum_{k=1}^K \left| d\left(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)^k}\right) h\left(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)^k}\right) - d\left(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i',j')^k}\right) h\left(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i',j')^k}\right) \right| \quad (11)$$

where  $(i, j)^k \in \mathcal{N}_{\mathbf{Y}_{(m,n)}}^K$  denotes the position of  $k$ th nearest patch to  $\mathbf{Y}_{(m,n)}$  in  $V_{\mathbf{Y}_{(m,n)}}^K$ , and  $(i', j')^k \in \mathcal{N}_{\mathbf{X}_{(m,n)}}^K$  denotes the position of  $k$ th nearest patch to  $\mathbf{X}_{(m,n)}$  in  $V_{\mathbf{X}_{(m,n)}}^K$ . In this paper, for the sake of simplicity and to be able to derive the normalized parameters based on statistical distribution to fuse the difference images, we uniformly set the function  $h(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)})$  to be constant 1, which means that we only use the difference of edge weights. Then (10) and (11) can be reduced to

$$f_{(m,n)}^{\mathbf{Y}} = \frac{1}{K} \sum_{k=1}^K \left| w\left(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)^k}\right) - w\left(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i',j')^k}\right) \right| \quad (12)$$

$$f_{(m,n)}^{\mathbf{Y}} = \frac{1}{K} \sum_{k=1}^K \left| d\left(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)^k}\right) - d\left(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i',j')^k}\right) \right| \quad (13)$$

From this, we can find that the structure difference criteria (12), (13) use the position information of the mapped  $G_{Y_{(m,n)}}^{map}$ , and they calculate the structure difference within the same image domain, which avoids direct comparison between heterogeneous data such

as the similarity difference in (7), and other types of difference criteria in SH [33], PP [34,35], and AMD [36].

After the structure difference  $f_{(m,n)}^Y$  is calculated, we assign it to all the pixels in the patch  $\mathbf{Y}_{(m,n)}$ . Then we turn to another target path  $\mathbf{Y}_{(\tilde{m},\tilde{n})}$  and repeat the same operation to obtain the structure difference  $f_{(\tilde{m},\tilde{n})}^Y$ . After the operations described above are performed on all the overlapping patches, then for each pixel  $(i, j)$ ,  $1 \leq i \leq M$ ,  $1 \leq j \leq N$  in the forward change image, there is a set  $F_{(i,j)}^Y$  of structure difference corresponding to the patches covering pixel  $(i, j)$ . Then the final change level of this pixel can be obtained by averaging this set as

$$DI^{fw}(i, j) = \frac{1}{|F_{(i,j)}^Y|} \sum_{f_{(m,n)}^Y \in F_{(i,j)}^Y} f_{(m,n)}^Y \quad (14)$$

Then we can obtain the forward difference image  $DI^{fw}$ . The backward difference image  $DI^{bw}$  to associate the pre-event image  $\mathbf{X}$  with the post-event image's NPSG  $G_{Y_{(m,n)}}^K$  can be similarly done as the generation of the forward  $DI^{fw}$ .

For the integrated difference image, we calculate the  $DI^{fw}$  and  $DI^{bw}$  by using distance criterion (13), and then fuse them to obtain the difference image (DI) as

$$DI^{final} = (DI^{fw} + DI^{bw})/2 \quad (15)$$

To improve the performance of the proposed method, we introduce two step sizes in the DI generation procedure. First, we use a search step  $\Delta_s > 1$  in the construction of  $K$ -nearest NPSG, which makes the distance between adjacent candidate patches in the search window  $\mathbb{W}$  greater than the step size. Therefore, the search space of the candidate patches  $\mathbb{W}$  can be reduced to  $\Omega$  by this  $\Delta_s$ , that is,  $\Omega$  can be regarded as the  $\Delta_s$  times down-sampling of  $\mathbb{W}$  and its cardinality is  $|\Omega| = \left\lfloor \frac{\xi_s - p_s}{\Delta_s} \right\rfloor^2$  with  $\lfloor \cdot \rfloor$  representing the rounding down operation. Thus, it can not only reduce the computation of constructing the  $K$ -nearest NPSG, but also forces these patches of  $K$ -nearest NPSG away from each other to avoid local aggregation, so that the corresponding  $K$ -nearest NPSG can contain richer and more stable structure information. Second, we use a target patch step  $\Delta_p \in [1, p_s]$  in the division of the target patches, which means that it can approximately reduce the amount of target patches by a factor  $\Delta_p^2$ , thereby speeding up the DI generation process. The framework of the NPSG based DI generation is summarized in Table 1, where  $\Lambda$  is the reduced position set of the anchor pixels of the target patches, whose cardinality is  $|\Lambda| = \left\lceil \frac{M}{\Delta_p} \right\rceil \left\lceil \frac{N}{\Delta_p} \right\rceil$  with  $\lceil \cdot \rceil$  representing the rounding up operation.

**Table 1**  
Implementation steps of generating the NPSG based DI.

Algorithm 1. NPSG based DI generation

**Input:** Heterogeneous images  $\mathbf{X}$ ,  $\mathbf{Y}$ , parameters  $p_s$ ,  $\xi_s$ ,  $\Delta_s$ ,  $\Delta_p$  and  $K$ .

1. Calculate the nonlocal patch similarity structure difference
  - for all target patches  $\mathbf{X}_{(m,n)}$ ,  $\mathbf{Y}_{(m,n)}$ ,  $(m, n) \in \Lambda$  do
    - Compute  $d(X_{(m,n)}, X_{(i,j)})$  and  $d(Y_{(m,n)}, Y_{(i,j)})$ ,  $\forall (i, j) \in \Omega$ .
    - Compute  $K$ -nearest NPSG  $G_{X_{(m,n)}}^K$  and  $G_{Y_{(m,n)}}^K$ .
    - Compute  $f_{(m,n)}^X$  by mapping  $G_{X_{(m,n)}}^K$  to  $\mathbf{X}$ .
    - Compute  $f_{(m,n)}^Y$  by mapping  $G_{Y_{(m,n)}}^K$  to  $\mathbf{Y}$ .
    - Add  $f_{(m,n)}^X$  and  $f_{(m,n)}^Y$  to the sets  $F_{(i,j)}^X$  and  $F_{(i,j)}^Y$ .
    - $\forall (i, j) \in \{(m + \vartheta_m, n + \vartheta_n); (\vartheta_m, \vartheta_n) \in \mathbb{P}\}$ , respectively.
  - end for
2. Compute the forward and backward DIs
  - for all pixels  $(i, j)$ ,  $1 \leq i \leq M$ ,  $1 \leq j \leq N$  do
    - Compute the forward  $DI^{fw}(i, j)$  and backward  $DI^{bw}(i, j)$ .
  - end for
3. Fuse the forward and backward DIs
 
$$DI^{final} = (DI^{fw} + DI^{bw})/2$$

### 2.3. Binary change map generation

Once the DI is obtained, the CD task can be treated as an image binary segmentation problem, which is similar as that in the homogeneous CD. The simplest way to obtain the binary change map  $CM = \{cm(i, j) | 1 \leq i \leq M, 1 \leq j \leq N\}$  is using thresholding method as

$$cm(i, j) = \begin{cases} 1, & DI^{final}(i, j) \geq T \\ 0, & DI^{final}(i, j) < T \end{cases} \quad (16)$$

where  $cm(i, j) = 1$  indicates a change and vice versa, and  $T$  is the threshold such as the Otsu threshold [37]. However, how to choose the optimal threshold is a very difficult problem. The clustering method can also be used in this binary segmentation problem, which is a process of grouping a set of features into meaningful clusters. Here, we select the PCAKM algorithm [14] to obtain the final binary CM for its simplicity and effectiveness, which uses the component analysis (PCA) method to extract the features and employs the  $k$ -means clustering to feature vectors with  $k = 2$ .

### 2.4. Patch distance calculation

The proposed CD method finds heavily on the patch similarity. Therefore, it is worth investigating in more depth the patch distance. This mainly includes two aspects. First, how to compare noisy patches? As there are many ways to measure distance, which one to choose in different situations needs to be determined, especially for the SAR image. Second, how to balance these distances of heterogeneous images? Because we need to fuse forward and backward DI as (15), which are based on their respective distance measurements on the heterogeneous images, therefore, these distance representation methods on heterogeneous images should be comparable, that is, they should be kept at the same level, otherwise the effect of this fusion will be greatly reduced.

We consider uncorrelated noise. First, for the patchwise optical-domain distance, where the AWGN model  $\mathbf{x} = \mathbf{z} + \mathbf{u}$  is usually used, with  $\mathbf{x} \in \mathbb{R}^{p_s^2}$  being the vectorized patch, and each element of  $\mathbf{z}$  being the true signal value and  $\mathbf{u}$  being a Gaussian random variable, in this case, the mean squared Euclidean distance is usually used as

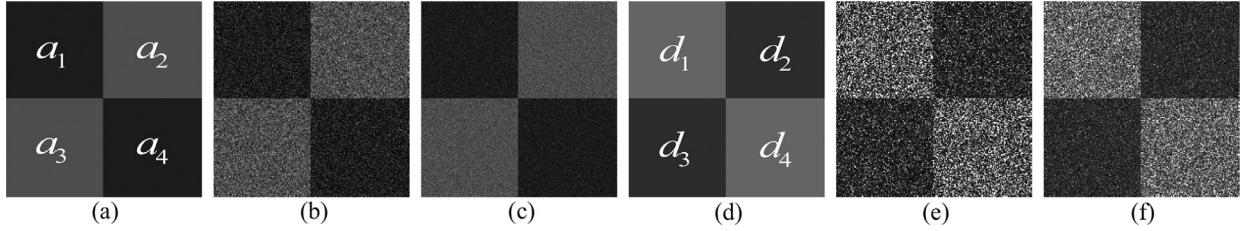
$$d_1(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2/p_s^2 \quad (17)$$

and its induced similarity criterion  $g(\mathbf{x}_1, \mathbf{x}_2) = \exp(-\lambda \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2)$  is optimal under most criteria of interest [45].

Second, for the SAR-domain distance, where the multiplicative model  $\mathbf{x} = \mathbf{z} \odot \mathbf{u}$  is usually used, with each element of  $\mathbf{z}$  being the true intensity value and  $\mathbf{u}$  being a Gamma random variable. Deledalle et. al [45] have derived three different similarity criteria by using different evaluation methods:

$$\begin{aligned} g(\mathbf{x}_1, \mathbf{x}_2) &= \prod_{i=1}^{p_s^2} \alpha_L \frac{1}{x_1(i)x_2(i)} \left( \frac{x_1(i)x_2(i)}{(x_1(i) + x_2(i))^2} \right)^L \\ g(\mathbf{x}_1, \mathbf{x}_2) &= \prod_{i=1}^{p_s^2} 2^{2L} \left( \frac{x_1(i)x_2(i)}{(x_1(i) + x_2(i))^2} \right)^L \\ g(\mathbf{x}_1, \mathbf{x}_2) &= \exp \left( - \sum_{i=1}^{p_s^2} (\log x_1(i) - \log x_2(i))^2 \right) \end{aligned} \quad (18)$$

where  $\alpha_L$  is a function of  $L$ , the first similarity criterion is derived based on the Bayesian joint likelihood with  $\alpha_L = \frac{\sqrt{L}\Gamma(2L)}{\Gamma(L)^2}$  (or the maximum joint likelihood with  $\alpha_L = \frac{(2L)^{2L}}{\Gamma(L)^2 e^{2L}}$ ), the second criterion is based on the generalized likelihood ratio (or the mutual



**Fig. 4.** A simple example of the differences between different distance criteria. (a) represents the true optical image with size  $200 \times 200$ , which is divided into four  $100 \times 100$  patches with  $a_1 = a_4 = 0.25$  and  $a_2 = a_3 = 0.75$ ; (b) and (c) are images acquired by different optical sensors with different noisy levels,  $\sigma_1 = 0.25$  and  $\sigma_2 = 0.1$ ; (d) represents the true radar image with the transformation  $d(i, j) = 1 - a(i, j)^2$  for each pixel; (e) and (f) are images acquired by different SAR sensors with different number of looks  $L_1 = 1$  and  $L_2 = 4$ . Then we can calculate the patch distances by using (17), (19) and (20) as

$$\begin{bmatrix} d_1(b_1, b_4) & d_1(b_1, b_2) & d_1(c_1, c_4) & d_1(c_1, c_2) \\ d_2(e_1, e_4) & d_2(e_1, e_2) & d_2(f_1, f_4) & d_2(f_1, f_2) \\ d_3(e_1, e_4) & d_3(e_1, e_2) & d_3(f_1, f_4) & d_3(f_1, f_2) \end{bmatrix} = \begin{bmatrix} 0.1248 & 0.3750 & 0.0200 & 0.2700 \\ 0.6140 & 0.7101 & 0.5311 & 1.0389 \\ 3.2921 & 3.8740 & 0.5677 & 1.1490 \end{bmatrix}.$$

From this, we can find that the distance contrast is larger in the image with low noise, which means that it is easier to distinguish the image patches in the low noise image, for example the distance ratio  $d_1(c_1, c_2)/d_1(c_1, c_4)$  is about 4 times of  $d_1(b_1, b_2)/d_1(b_1, b_4)$ . Meanwhile, we can also find that the distances obtained by using these standard criteria are unbalanced, for example,  $d_3(e_1, e_4)$  is more than 25 times of  $d_1(b_1, b_4)$ . Therefore, the fusion contrast between the changed and unchanged regions is not obvious, for example,  $r_1 = [d_3(e_1, e_2) + d_1(b_1, b_2)]/[d_3(e_1, e_4) + d_1(b_1, b_4)] = 1.244$  and  $r_2 = [d_2(f_1, f_2) + d_1(c_1, c_2)]/[d_2(f_1, f_4) + d_1(c_1, c_4)] = 2.375$ . After normalization with (24), the normalized patch distances are

$$\begin{bmatrix} d_1(b_1, b_4) & d_1(b_1, b_2) & d_1(c_1, c_4) & d_1(c_1, c_2) \\ d_2(e_1, e_4) & d_2(e_1, e_2) & d_2(f_1, f_4) & d_2(f_1, f_2) \\ d_3(e_1, e_4) & d_3(e_1, e_2) & d_3(f_1, f_4) & d_3(f_1, f_2) \end{bmatrix} = \begin{bmatrix} 0.9987 & 2.9999 & 1.0000 & 13.5008 \\ 1.0004 & 1.1570 & 1.0002 & 1.9564 \\ 1.0007 & 1.1776 & 1.0002 & 2.0241 \end{bmatrix},$$

and the fusion contrasts between the changed and unchanged regions become  $r_1 = 2.089$  and  $r_2 = 7.728$ , respectively.

information kernel), and the last criterion is based on the variance stabilizing transform. However, as the first similarity criterion lacks some important properties such as the Maximal self-similarity, which means that there exists a distinct pair  $(\mathbf{x}_1, \mathbf{x}_2)$  makes  $g(\mathbf{x}_1, \mathbf{x}_2) > g(\mathbf{x}_1, \mathbf{x}_1)$ , the performance of this similarity criterion is not satisfactory compared with the other two criteria (for more details, please refer to [45]). Therefore, we only refer to the last two similarity criteria, and the corresponding distance expressions are

$$d_2(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{p_s^2} \sum_{i=1}^{p_s^2} 2L \log \left( \frac{x_1(i) + x_2(i)}{2\sqrt{x_1(i)x_2(i)}} \right) \quad (19)$$

$$d_3(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{p_s^2} \sum_{i=1}^{p_s^2} (\log x_1(i) - \log x_2(i))^2 \quad (20)$$

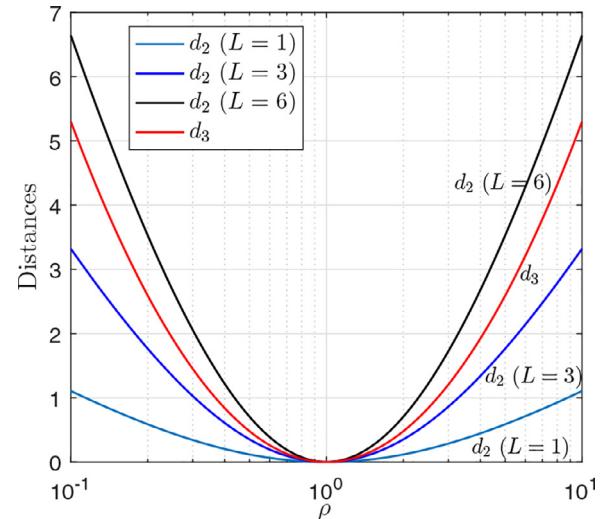
Next, we consider how to make these distances for heterogeneous data at the same level. We give a simple example to illustrate that the distances obtained by using these standard criteria are unbalanced as shown in Fig. 4, where the ratios of different criteria are even dozens of times. Therefore, we can not directly fuse the forward DI and backward DI by using these standard distance formulations, and we need to normalize them.

We further analyze these patch distance criteria. Suppose pixels  $x_1$  and  $x_2$  in the image are independent identically distributed (i.i.d.), we want to identify pixels with similar signal components to the target based on the different measured distances, that is,  $x_1 \approx x_2$ . Since the observed pixel value is polluted by noise,  $x_i = z_i + u_i$  or  $x_i = z_i u_i$ , the distance also depends on the noise. Let us consider three limiting cases: 1) noiseless signal, 2) uniform noisy signal, and 3) non-uniform noisy signal.

In the first case,  $u_i = 0$  for the optical signal or  $u_i = 1$  for the SAR signal, define the signal difference in optical image as  $\eta = x_1 - x_2$  and the signal ratio in SAR image as  $\rho = x_1/x_2$ , and substitute this into (17), (19) and (20), we have

$$\begin{aligned} d_1(x_1, x_2) &= \eta^2; & d_2(x_1, x_2) &= 2L \log \left( \frac{\sqrt{\rho}}{2} + \frac{1}{2\sqrt{\rho}} \right); \\ d_3(x_1, x_2) &= [\log(\rho)]^2 \end{aligned} \quad (21)$$

In Fig. 5, we plot the distances  $d_2$  and  $d_3$  as the functions of signal ratio. We can find that their shapes are symmet-



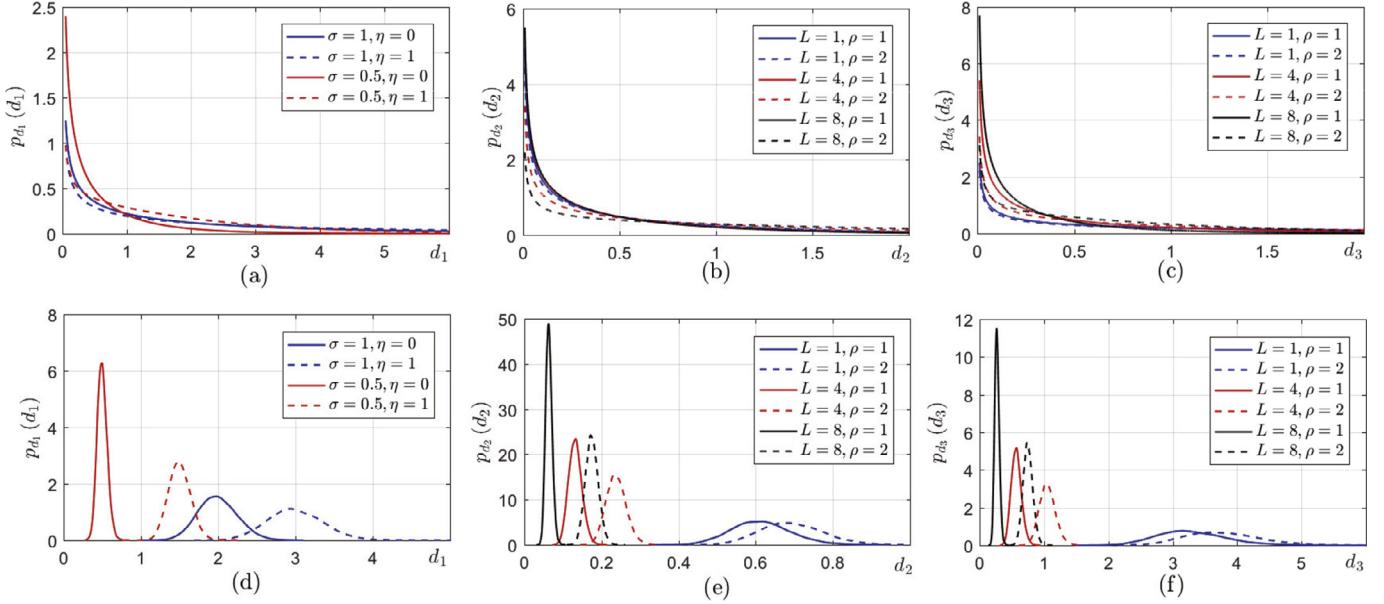
**Fig. 5.** Plot of the speckle-free distances of  $d_2$  and  $d_3$ .

rical. Define  $\rho = e^t$ , we have  $d_2(x_1, x_2) = 2L \log \left( \frac{e^{t/2} + e^{-t/2}}{2} \right)$  and  $d_3(x_1, x_2) = t^2$ . Therefore, in semilog axes,  $d_3$  is quadratic,  $d_2$  is almost quadratic around the minimum, zero, and begins growing linearly for larger/smaller values. Generally, when the difference between the object and the background in the SAR image is not high, sharp curve is more suitable, such as the  $d_3$ , which can accurately find the similar target patches rather than the background patch for the target patch. Meanwhile, when the noise level is high, we should choose a gentle curve, such as the  $d_2$  with  $L \leq 4$ . This is because that it is not much discriminative for the samples with relatively close intensity caused by the noisy, thus it is more robust.

In the second case, with the uniform signal,  $z_1 = z_2$ , by substituting this in (17), (19) and (20), we have

$$\begin{aligned} d_1(x_1, x_2) &= (u_1 - u_2)^2; & d_2(x_1, x_2) &= 2L \log \left( \frac{u_1 + u_2}{2\sqrt{u_1 u_2}} \right); \\ d_3(x_1, x_2) &= (\log u_1 - \log u_2)^2 \end{aligned} \quad (22)$$

where  $u_i \sim \mathcal{N}(0, \sigma^2)$  is the independent Gaussian distributed random variable with zero mean and  $\sigma^2$  variance for the optical im-



**Fig. 6.** The theoretical pdfs of the pixel distances and empirical pdfs of the patch distances. (a) pdfs of  $d_1$  under different noise levels  $\sigma = 1, 0.5$ , and with different signal intensity differences  $\eta = 0, 1$ ; (b) pdfs of  $d_2$  under different noise levels  $L = 1, 4, 8$ , and with different signal intensity ratios  $\rho = 1, 2$ ; (c) pdfs of  $d_3$  under distance noise level  $L = 1, 4, 8$ , and with different signal intensity ratios  $\rho = 1, 2$ ; (d) empirical pdfs of the patch distances using  $d_1$  with different  $\sigma$  and  $\eta$ ; (e) empirical pdfs of the patch distances using  $d_2$  with different  $L$  and  $\rho$ ; (f) empirical pdfs of the patch distances using  $d_3$  with different  $L$  and  $\rho$ .

age in  $d_1$ , and  $u_i \sim \text{Gamma}(L, L)$  is the independent Gamma distributed random variable with shape parameter  $L$  and inverse scale parameter  $L$  (equal to the number of looks) for the SAR image in  $d_2$  and  $d_3$ . In appendix A, we calculate the pdfs and expected values of these distances as

$$\begin{aligned} E[d_1(x_1, x_2)] &= 2\sigma^2; \quad E[d_2(x_1, x_2)] = -L[\varphi(0, L) - \varphi(0, L+1/2)]; \\ E[d_3(x_1, x_2)] &= 2\varphi(1, L) \end{aligned} \quad (23)$$

In the proposed heterogeneous CD method, for each target patch, take  $\mathbf{Y}_{(m,n)}$  (or  $\mathbf{X}_{(m,n)}$ ) in the forward (or backward) detection as an example, if no change occurs in this patch, each patch  $\mathbf{Y}_{(i,j)}$  (or  $\mathbf{X}_{(i,j)}$ ) in its own  $K$ -nearest NPSG and mapped  $K$ -nearest NPSG is very similar to the target patch  $\mathbf{Y}_{(m,n)}$  (or  $\mathbf{X}_{(m,n)}$ ), ideally, their true values are equal  $\mathbf{Z}_{(i,j)}^{\mathbf{Y}} = \mathbf{Z}_{(m,n)}^{\mathbf{Y}}$  (or  $\mathbf{Z}_{(i,j)}^{\mathbf{X}} = \mathbf{Z}_{(m,n)}^{\mathbf{X}}$ ),  $(i, j) \in \mathcal{N}_{\mathbf{Y}_{(m,n)}}^K \cup \mathcal{N}_{\mathbf{X}_{(m,n)}}^K$ . As the patch distance level between heterogeneous images is very different as (23), we need to normalize these distances to make the pixel value in the forward and backward DIs balanced for a better fusion performance. An intuitive and reasonable approach is to keep their values at the same level in the unchanged areas. Then, suppose  $\mathbf{X}_{(m,n)}$ ,  $\mathbf{X}_{(i,j)}$  are optical patches with the same true values, and  $\mathbf{Y}_{(m,n)}$ ,  $\mathbf{Y}_{(i,j)}$  are SAR patches with the same true values, we need to have  $E[d_{opt}(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)})] = E[d_{SAR}^{(1)}(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)})] = E[d_{SAR}^{(2)}(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)})]$ , where  $d_{opt}$ ,  $d_{SAR}^{(1)}$  and  $d_{SAR}^{(2)}$  correspond to (2), (4) and (5), respectively. Therefore, we add normalized parameters in the distance criteria in  $d_{opt}$ ,  $d_{SAR}^{(1)}$  and  $d_{SAR}^{(2)}$  with

$$\begin{aligned} \gamma_{\sigma} &= 1/[2\sigma^2]; \quad \gamma_L^{(1)} = 1/[L(\varphi(0, L+1/2) - \varphi(0, L))]; \\ \gamma_L^{(2)} &= 1/[2\varphi(1, L)] \end{aligned} \quad (24)$$

In the third case, we consider a more common model that  $\eta = z_1 - z_2$ ,  $u_1, u_2 \sim \mathcal{N}(0, \sigma^2)$  in the optical image and  $\rho = z_1/z_2$ ,  $u_1, u_2 \sim \text{Gamma}(L, L)$  in the SAR image. We calculate the pdfs of the distances (17), (19) and (20) in the Appendix (refer to equations (A.5), (A.16) and (A.24)). For a more precise analysis, we plot the pdfs of the distances for several different signal intensity differences/ratios ( $\eta, \rho$ ) as shown in Fig. 6. For the uniform pixels  $z_1 = z_2$ , if we set the detection rate to 80% for detecting the uni-

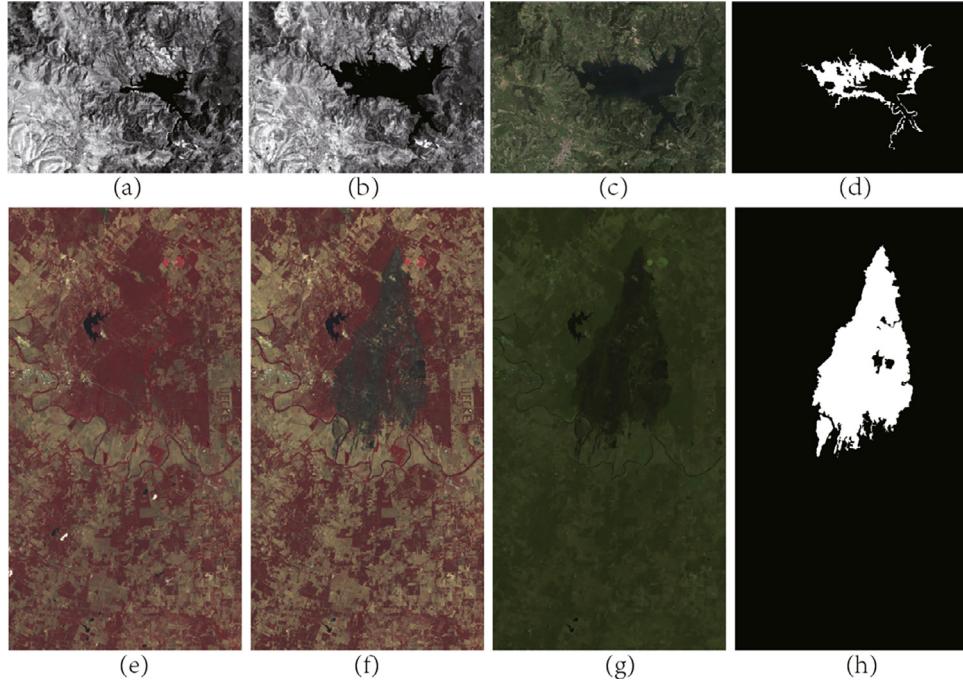
form  $x_1$  and  $x_2$ , then the detection thresholds are  $1/9 < x_1/x_2 < 9$  under the signal-look SAR image, which is obtained by computing the probability  $P(d_2(x_1, x_2) < 2 \log \frac{5}{3} | L = 1) \approx 0.8$  or  $P(d_3(x_1, x_2) < [\log(9)]^2 | L = 1) \approx 0.8$ . However, this detection thresholds are  $1/2.59 < x_1/x_2 < 2.59$  and  $1/1.93 < x_1/x_2 < 1.93$  for 4-look and 8-look SAR images, respectively. This intuitively shows the influence of different noise levels on the detection effect. Meanwhile, from Fig. 6, we can also find that the pdfs largely overlap for different signal intensity differences or ratios ( $\eta$  or  $\rho$ ) under the high noise level conditions ( $\sigma = 1$  or  $L = 1$ ), which means that it is difficult to carry out reliable discrimination. However, in this paper, we use the patchwise distance instead of the pixelwise distance. For the sake of simplicity, we assume two patches with constant signal intensity ratio as

$$z(m + \vartheta_m, m + \vartheta_n)/z(i + \vartheta_m, j + \vartheta_n) = \rho, \forall (\vartheta_m, \vartheta_n) \in \mathbb{P} \quad (25)$$

Then, the patchwise distance becomes the mean of  $|\mathbb{P}|$  i.i.d. random variables, well approximate by a Gaussian distribution with the Central Limit Theorem. Fig. 6(d)-(f) show the estimated pdfs of different patch distances when  $11 \times 11$ -pixel patches are considered with the Monte Carlo simulation. By comparing Fig. 6(a)-(c) with (d)-(f), we can find that the patchwise distance can provide a much better discrimination than the pixelwise distance, which is one of the reasons why we choose the robust patchwise based NPSG for the heterogeneous CD. At the same time, our in-depth discussion on these distance criteria can also provide a good reference for other applications, such as the nonlocal based speckle reduction algorithms.

### 3. Experimental results and discussion

In this Section, simulations are performed to demonstrate the proposed conclusions and evaluate the performance of proposed nonlocal patch similarity based heterogeneous CD method. The experiments are performed on five pairs of heterogeneous remote sensing images. Detailed descriptions of datasets, quantitative measures, parameter analysis, and experimental results are provided in the following.



**Fig. 7.** Sardinia and Texas datasets. The top row is the Sardinia dataset: (a) the NIR band image acquired in September 1995; (b) the NIR band image acquired in July 1996; (c) the optical image acquired in July 1996; (d) the ground truth image representing the changes between (a) and (c). The bottom row is the Texas dataset: (e) the image acquired by Landsat 5 TM in August 2011; (f) the image acquired by Landsat 5 TM in September 2011; (g) the image acquired by Advanced Land Imager from the Earth Observing mission in September 2011; (h) the ground truth image representing the changes between (e) and (g).

### 3.1. Experimental datasets and quantitative measures

Five pairs of heterogeneous datasets are used to assess the effectiveness of the proposed CD method.

(1) Dataset A: The *Sardinia* dataset consists of two near-infrared (NIR) band images and one optical image (Sardinia, Italy), as shown in Fig. 7(a)-(c). The two NIR band images, with  $300 \times 412 \times 1$  pixels, are acquired by Landsat-5 (L5t1 and L5t2), in September 1995 and July 1996, respectively, whereas the optical image, with  $300 \times 412 \times 3$  pixels, is obtained from Google Earth (GEt2) with red, green, and blue bands in July 1996. The dataset represents the expansion of Lake Mulargia as shown in Fig. 7(d), which is built by the on-the-spot investigation.

(2) Dataset B: The *Texas* dataset consists of three multispectral images: a pair of pre-event and first post-event images with  $1534 \times 808 \times 7$  pixels from the same sensor, the Landsat 5 TM (L5t1 and L5t2), in August 2011 and September 2011, as shown in Fig. 7(e) and (f),<sup>1</sup> respectively; and the second post-event image with  $1534 \times 808 \times 10$  pixels acquired by the Advanced Land Imager (ALIt2) from the Earth Observing mission in September 2011, as shown in Fig. 7(g)<sup>2</sup>. The dataset represents a forest fire in Bastrop County, Texas, and the ground truth is provided by Volpi et al. [47]. As the L5t2 and ALIt2 are acquired within 1 day interval, there is no apparent change between them. The change between L5t1 and L5t2 (or ALIt2) is shown in Fig. 7(h), which represents the fire damage.

(3) Dataset C: The *Shuguang* dataset consists of one SAR image and one optical satellite image, as shown in Fig. 8(a) and (b), respectively. The SAR image, with size  $593 \times 921 \times 1$ , is taken by the Radarsat-2 with C-band on June 2008, whereas the optical image, with size  $593 \times 921 \times 3$ , is acquired from Google Earth with red, green, and blue bands in September 2012. The dataset represents the changes of land use in the farmland as shown in

Fig. 8(c), which is generated by manual annotation that combines expert knowledge and prior information.

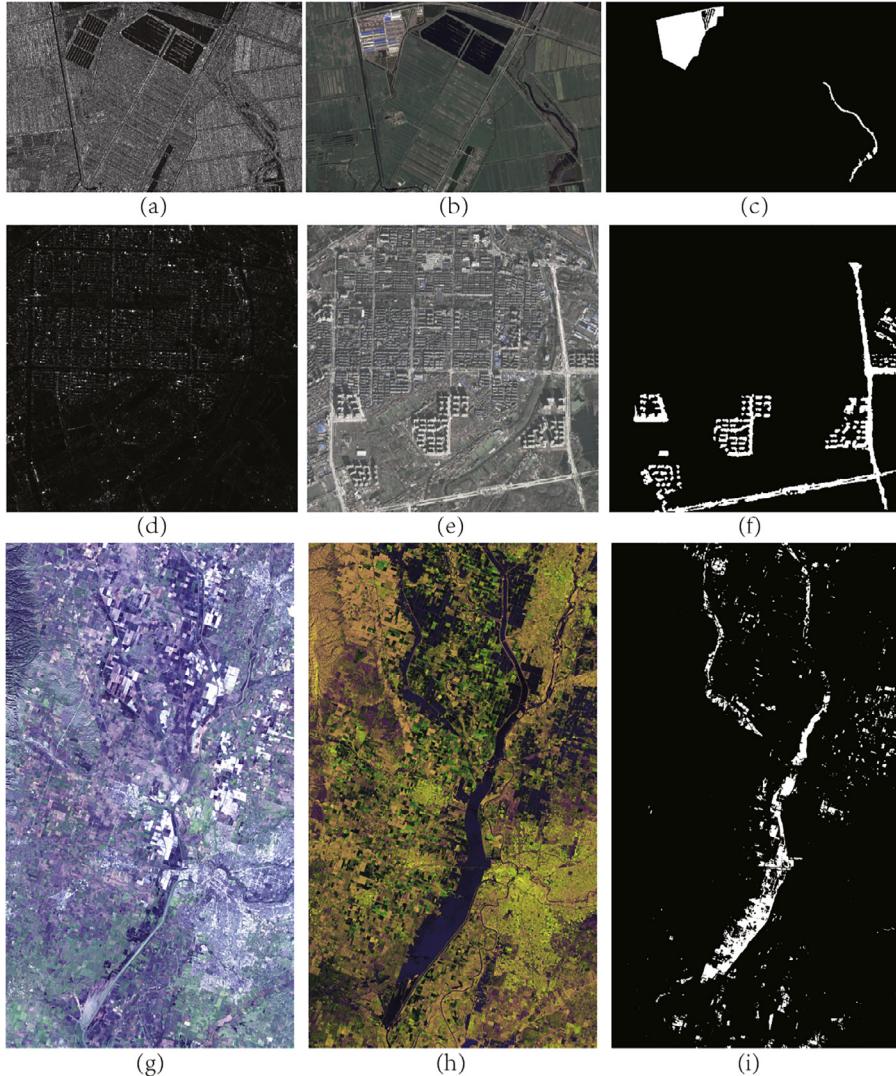
(4) Dataset D: The *Wuhan* dataset is a pair of SAR/optical satellite images (Wuhan City, China), as shown in Fig. 8(d) and (e). The SAR image, with  $495 \times 503 \times 1$  pixels, is taken by the Radarsat-2 with C-band in June 2008, whereas the optical image, with  $495 \times 503 \times 3$  pixels, is acquired from Google Earth with red, green, and blue bands in November 2011. The dataset represents the changes of new buildings and roads as shown in Fig. 8(f), which is generated through the on-the-spot investigation.

(5) Dataset E: the *California* dataset is a pair of multispectral/SAR images, as shown in Fig. 8(g)<sup>2</sup> and (h)<sup>2</sup>. The multispectral image, with size  $875 \times 500 \times 11$ , is taken by Landsat 8 on January 5, 2017, with nine channels covering the spectrum from deep blue to shortwave infrared and two long-wave infrared channels; whereas the SAR image, with size  $875 \times 500 \times 3$ , is acquired by Sentinel-1A recorded in polarisations VV and VH on 18 February 2017 and augmented with the ratio between the two intensities as the third channel (all these channels are log-transformed). The ground truth in Fig. 8(i) is provided by Luppino et al. [36], which represents a flood in Sacramento County, Yuba County, and Sutter County, California. These images are re-sampled from the original  $3500 \times 2000$  images to reduce the computation time.

The performance of DIs generated by the comparing methods can be assessed by the empirical receiver operating characteristics (ROC) curves, which represents the estimated pixel-wise probability of detection (PD) as a function of the probability of false alarm (PFA) by varying of the binary segmentation threshold  $T$  in (16). Moreover, two quantitative criteria derived from the ROC curve can be computed: (1) the area under the curve (AUC), corresponding to the integral of the ROC curve; (2) the diagonal distance (Ddist) between the no detection point ( $PFA=1$ ,  $PD=0$ ) and the point at the interception of the ROC curve with the diagonal line defined

<sup>1</sup> Distributed by LP DAAC, <http://lpdaac.usgs.gov>.

<sup>2</sup> Data processed by ESA, <http://www.copernicus.eu/>.



**Fig. 8.** *Shuguang*, *Wuhan* and *California* datasets. The top row is the *Shuguang* dataset: (a) the SAR image acquired in June 2008; (b) the optical image acquired in September 2012; (c) the ground truth image representing the changes between (a) and (b). The middle row is the *Wuhan* dataset: (d) the SAR image acquired in June 2008; (e) the optical image acquired in November 2011; (f) the ground truth image representing the changes between (d) and (e). The bottom row is the *California* dataset: (g) the multispectral image acquired in January 2017; (h) the SAR image acquired in February 2017; (i) the ground truth image representing the changes between (g) and (h).

by PFA=1-PD. For these two metrics, the greater the criterion, the better the detection. In order to measure the binary CM, we select the widely used quantitative parameters: false positives rate (FPR), false negatives rate (FNR), the percentage correct classification (PCC) and the Kappa coefficient (Kappa).

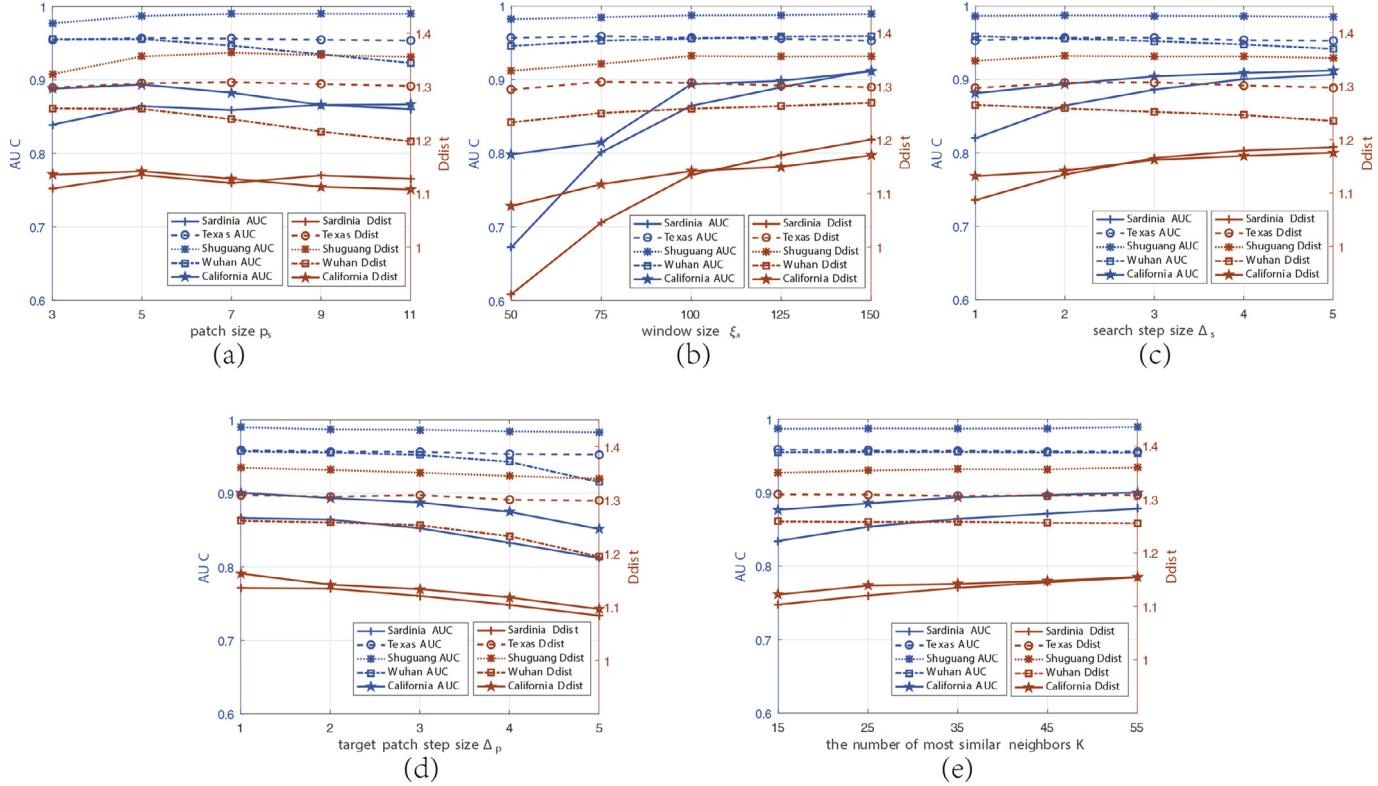
### 3.2. Parameter analysis

The main parameters in the proposed NPSG are the patch size  $p_s$ , the search window size  $\xi_s$ , the search step size  $\Delta_s$ , the target patch step size  $\Delta_p$ , and the number of the selected most similar neighbors  $K$ . To measure the impact of these parameters, we analyze the influences of them using the above datasets. The AUC and Ddist of the ROC curve are adopted to evaluate the generated DI, respectively, which can describe the global performance of the NPSG.

(1) Influence of the patch size  $p_s$ . In Fig. 9(a), we vary  $p_s$  from 3 to 11 with step two and keep other parameters fixed as  $\xi_s = 100$ ,  $\Delta_s = \Delta_p = 2$  and  $K = 35$ . It can be found that the patch size has an important impact on the CD performance. As the  $p_s$  increase, the

AUC and Ddist gradually increase first and then decrease, which is mainly because that a too small  $p_s$  cannot fully reflect structural information and is not robust to the noise, whereas a too large  $p_s$  makes it very hard to find enough similar patches. Generally, for high-resolution images, a larger  $p_s$  can be selected; on the contrary, for low-resolution images,  $p_s$  should be appropriately smaller. From Fig. 9(a), we can see that setting  $p_s = 5$  is appropriate for our experiments.

(2) Influence of the search window size  $\xi_s$ . We vary the  $\xi_s$  from 50 to 150 with step 25 and keep other parameters fixed as  $p_s = 5$ ,  $\Delta_s = \Delta_p = 2$  and  $K = 35$ . From Fig. 9(b), we can find that the datasets of *Shuguang*, *Wuhan* and *Texas* are not as sensitive to  $\xi_s$  as the datasets of *Sardinia* and *California*. For the *Sardinia* and *California* datasets, a larger search window can bring better results, which is mainly due to that a larger  $\xi_s$  can bring more similar patches. However, a very large  $\xi_s$  will increase the search space  $\Omega$  of the candidate patches ( $|\Omega| = \left\lfloor \frac{\xi_s - p_s}{\Delta_s} \right\rfloor^2$ ), and bring high computing cost. In addition, the value of  $\xi_s$  is also affected by  $p_s$ . When  $p_s$  increases,  $\xi_s$  should also increase. According to our experience, it is suggested to set the window size to  $\xi_s = 20p_s$ .



**Fig. 9.** Influences of parameters on the NPSG performance: (a) influence of the patch size  $p_s$ ; (b) influence of the search window size  $\xi_s$ ; (c) influence of the search step size  $\Delta_s$ ; (d) influence of the target patch step size  $\Delta_p$ ; (e) influence of the number of the most similar neighbors  $K$ .

(3) Influence of the search step size  $\Delta_s$ . In Fig. 9(c), we vary  $\Delta_s$  from 1 to 5 with step one and keep other parameters fixed as  $p_s = 5$ ,  $\xi_s = 100$ ,  $\Delta_p = 2$  and  $K = 35$ . We can see that the Shuguang and Texas datasets are more robust to  $\Delta_s$  than other datasets. As the large step size can reduce the search space to accelerate the algorithm, and avoid local aggregation of similar patches, we suggest setting the search step size to  $\lfloor p_s/2 \rfloor \leq \Delta_s \leq p_s$ . In our experiments, we fix this parameter to  $\Delta_s = 2$  in a comprehensive consideration.

(4) Influence of the target patch step size  $\Delta_p$ . We vary  $\Delta_p$  from 1 to 5 with step one and keep other parameters fixed as  $p_s = 5$ ,  $\xi_s = 100$ ,  $\Delta_s = 2$  and  $K = 35$ . From Fig. 9(d), we can find that as the value of  $\Delta_p$  increases, the performance of NPSG will gradually decrease. It starts to slow down, and then it drops quickly. Although the  $\Delta_p$  can speed up the DI generation process by nearly  $\Delta_p^2$  times, at the same time it will smooth the DI as can be seen from (14). The value range of  $\Delta_p$  is  $2 \leq \Delta_p \leq p_s$ , and we suggest to set  $\Delta_p$  to 2 or 3 as a compromise choice.

(5) Influence of the number of the selected most similar neighbors  $K$ . In Fig. 9(e), we vary  $K$  from 15 to 55 with step 10 and keep other parameters fixed as  $p_s = 5$ ,  $\xi_s = 100$  and  $\Delta_s = \Delta_p = 2$ . It can be found that the detection performance is not very sensitive to the value of  $K$ . Of course, an extra small  $K$  is not recommended, such as  $K < 10$ . This is because that a very small  $K$ -nearest NPSG is not robust enough. At the same time, a very big  $K$  is also not recommended, such as  $K \geq 0.2 * |\Omega| = 0.2 * \left\lceil \frac{\xi_s - p_s}{\Delta_s} \right\rceil^2$ . There are two main reasons: first, a large  $K$  will lead to confusion, that is, there may be no  $K$  really similar patches in the search window, and some patches that do not really represent the same object as the target patch will also be introduced into the  $K$ -nearest NPSG; second, a large  $K$  will increase the computational complexity. Therefore, we empirically suggest to set this parameter  $20 \leq K \leq \min \left\{ 75, 0.05 * \left\lceil \frac{\xi_s - p_s}{\Delta_s} \right\rceil^2 \right\}$  by considering the CD perfor-

mance and the computational cost. In our experiments, we fix this parameter to  $K = 35$ .

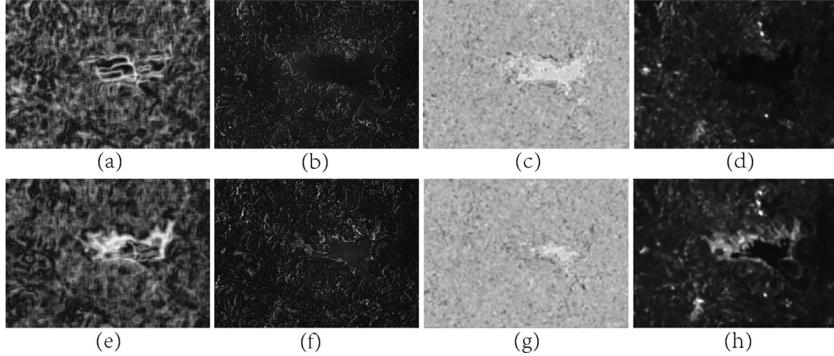
(6) Meanwhile, the performance of the final CD is also influenced by clustering process when we select the PCAKM algorithm to obtain the binary CM. The main parameters of PCAKM are the feature vector (principal components) size  $s$  and the image block size  $h$ . In [14], these two parameters are analyzed in detail. Generally, setting  $s \in \{3, 5, 7\}$  can meet most of the requirements. The parameter  $h$  defines the local neighborhood size, which affects the contribution of spatial contextual information on the feature extraction. A large  $h$  will smoothen the effect of noise and reduce the false detections, but it will bring more miss detections, and vice versa. In this paper, we set  $s$  equal to  $h$  for convenience in PCA feature extraction. We try different sizes of  $h$  ( $3 \leq h \leq 7$ ), and then choose the best result for each method.

### 3.3. Experimental results

We apply four methods for comparison: SH [33], PP [34,35], AMD [36] and Markov Model for Multimodal Change Detection method (M3CD) [48], which is relying on an observation field built up from a pixel pairwise modeling on heterogeneous image pair. Among them, M3CD does not generate DI, but directly generates binary CM. For the SH, PP, AMD and the proposed NPSG based CD method (called NPSG for short), we adopt the Otsu thresholding method [37] and PCAKM [14] to generate the final CM after obtaining the DI. For each DI, we try different feature vector sizes and image block sizes for the PCAKM, and then choose the best result for each method.

#### 3.3.1. Experiments on Sardinia and Texas datasets

In the first experiment, we show that the NPSG can establish a connection between heterogeneous images, which means that NPSG can be consistent in the unchanged area, but no longer consistent in the changed area. Then, we test each similarity-based



**Fig. 10.** DIs and binary CMs on *Sardinia* dataset. The top row is the DI<sub>1</sub> between L5t2 and GEt2: (a) SH; (b) PP; (c) AMD; (d) NPSG. The bottom row is the DI<sub>2</sub> between L5t1 and GEt2: (e) SH; (f) PP; (g) AMD; (h) NPSG.

**Table 2**  
Quantitative measures (AUC and Ddist) on *Sardinia* and *Texas* datasets.

| Methods | <i>Sardinia</i> dataset |               | <i>Texas</i> dataset |               |
|---------|-------------------------|---------------|----------------------|---------------|
|         | AUC                     | Ddist         | AUC                  | Ddist         |
| SH      | 0.8840                  | 1.1390        | 0.5578               | 0.7655        |
| PP      | 0.7445                  | 0.9489        | 0.3718               | 0.5633        |
| AMD     | 0.3645                  | 0.5549        | 0.8810               | 1.1403        |
| NPSG    | <b>0.9129</b>           | <b>1.2006</b> | <b>0.9528</b>        | <b>1.2983</b> |

method by comparing the performances on the heterogeneous images acquired at the same time and different times, respectively.

Using the *Sardinia* dataset, we assume that the noise on this dataset is the AWGN. Then, by selecting some uniform regions in the images, we estimate the noise standard deviations of NIR and optical images to be 1.08e-3 and 9.71e-3, respectively. Fig. 10 shows the difference maps between heterogeneous image pairs of L5t2 and GEt2, L5t1 and GEt2. Obviously, since L5t2 and GEt2 are obtained at similar times, there is no obvious changes between them. Ideally, there should be no obvious structural features in the DI<sub>1</sub> between L5t2 and GEt2, however, the changing parts should be highlighted in the DI<sub>2</sub> between L5t1 and GEt2. From Fig. 10, we can find that the SH can highlight the changes in DI<sub>2</sub>, but it also shows some undesirable structures in DI<sub>1</sub>; the PP and AMD fail to detect the changes in DI<sub>2</sub> as shown in Figs. 10(f) and 10(g); the NPSG can not only find the changed parts in DI<sub>2</sub>, but also show little structural features in DI<sub>1</sub>.

For the *Texas* dataset, we assume that the dataset is contaminated by AWGN, and estimate the standard deviations to be 1.63e-2 and 5.65e-3 for L5t1 (L5t2) and ALIt2, respectively. Fig. 11 shows the difference maps between heterogeneous image pairs of L5t2 and ALIt2, L5t1 and ALIt2. It can be found that the SH and PP cannot find the changes between the heterogeneous L5t1 and ALIt2, as shown in Fig. 11(e) and (f). Both AMD and NPSG perform well in *Texas* dataset, which show the consistency between L5t2 and ALIt2 and the changed part between L5t1 and ALIt2. In order to further compare the ability of these methods to detect changes, we also plot ROC curves in Fig. 12 and list quantitative comparisons of AUC and Ddist in Table 2. The analysis of these results shows that the NPSG outperforms these comparing methods in these two datasets.

### 3.3.2. Experiments on Shuguang, Wuhan and California datasets

In the second experiment, we test the NPSG on three pairs of SAR/optical (multispectral) images. For the intensity SAR images of these datasets, the ENL parameter  $L$  can be estimated by selecting some uniform regions in the images as  $L = (\frac{\mu_{ur}}{\sigma_{ur}})^2$  [49], where  $\mu_{ur}$  is the mean value of the selected uniform region and  $\sigma_{ur}$  is the corresponding standard deviation. In order to fully compare the

**Table 3**  
Quantitative measures of binary CMs on *Shuguang* dataset.

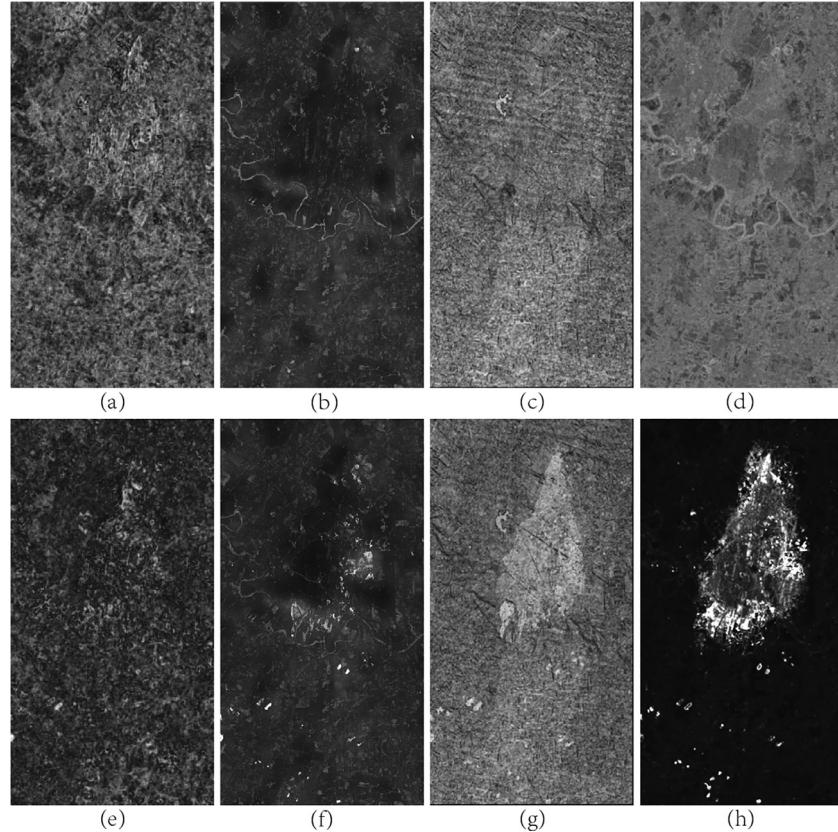
| Methods    | FPR(%) | FNR(%) | PCC(%) | Kappa  |
|------------|--------|--------|--------|--------|
| M3CD       | 2.33   | 1.47   | 96.20  | 0.6017 |
| SH-Otsu    | 54.55  | 0.40   | 45.05  | 0.0517 |
| SH-PCAKM   | 53.76  | 0.35   | 45.89  | 0.0553 |
| PP-Otsu    | 14.33  | 1.95   | 83.71  | 0.1862 |
| PP-PCAKM   | 17.01  | 0.05   | 82.94  | 0.2942 |
| AMD-Otsu   | 51.50  | 1.38   | 47.12  | 0.0258 |
| AMD-PCAKM  | 48.55  | 1.32   | 50.13  | 0.0348 |
| NPSG-Otsu  | 2.66   | 0.53   | 96.81  | 0.7024 |
| NPSG-PCAKM | 1.55   | 0.97   | 97.48  | 0.7288 |

performances of different methods, we not only compare the DI, but also compare the binary CM. To generate the binary CM, we employ Otsu thresholding method and PCAKM on the DIs of each similarity based heterogeneous CD method, respectively.

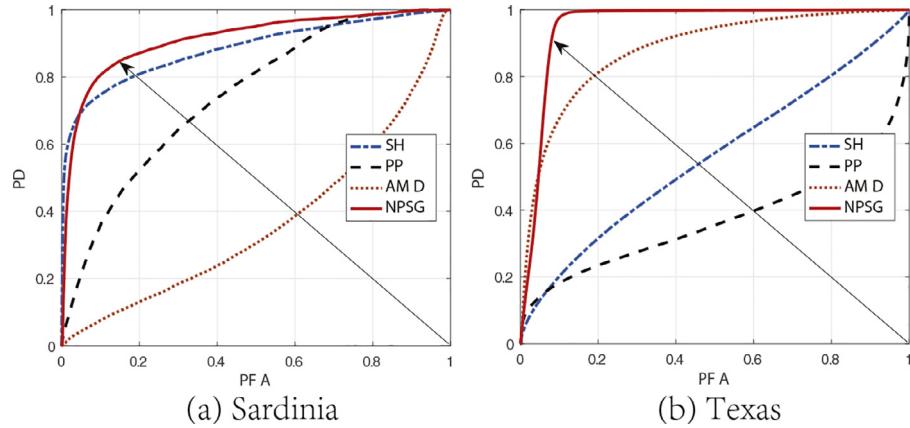
Figs. 13, 14 and 15 show the DIs, CMs of different methods on *Shuguang*, *Wuhan* and *California* datasets, respectively. Fig. 16 plots the ROC curves of SH, PP, AMD and NPSG on these three datasets. For the *Shuguang* dataset, we estimate that the ENL of SAR image is 11 and the noise standard deviation of optical image is 1.12e-2. Therefore, we choose the second type of distance criterion  $d_{SAR}^{(2)}$  in (5) for the SAR image. Form the detection performance, we can find that the SH and AMD fail to detect the changes as shown in Fig. 13(a) and (c), thus results in smaller AUC and Ddist, as shown in Fig. 16(a) and listed in Table 6. From the comparison of CMs, we can see that the NPSG can obtain the best result whether using the Otsu thresholding or PCAKM, so it gets Kappa coefficients of 0.7024 and 0.7288 for NPSG-Otsu and NPSG-PCAKM, respectively, followed by M3CD with Kappa coefficient 0.6017, which outperforms the PP, SH, and AMD, as listed in Table 3.

For the *Wuhan* dataset, the estimated ENL of SAR image is 4, and the noise standard deviation of optical image is 3.59e-2. Therefore, we choose the first type of distance criterion  $d_{SAR}^{(1)}$  in (4) for the SAR image. From Fig. 14(c), we can see that the AMD cannot highlight the changes in the generated DI, which leads to a poor performance on the corresponding ROC curve as shown in Fig. 16(b). The PP performs better than SH on *Wuhan* dataset with a higher AUC and Ddist as listed in Table 6. Visually, the DI generated by NPSG is cleaner than other DIs as shown in Fig. 14(a)-(d), that is, it is more sensitive to real changes and more robust to the false changes caused by noise, which is further confirmed in Fig. 16(b) and Table 6. Meanwhile, by comparing the CMs of different methods in Fig. 14(e)-(g) and Table 4, the NPSG obtain the best CM with highest Kappa coefficient and PCC, followed by PP-PCAKM, M3CD and PP-Otsu, which outperform other methods.

For the last *California* dataset, as the SAR image has been de-speckled, the estimated ENL is 23.47, thus we choose the  $d_{SAR}^{(2)}$  in



**Fig. 11.** DIs and binary CMs on Texas dataset. The top row is the  $\text{DI}_1$  between L5t2 and ALlt2: (a) SH; (b) PP; (c) AMD; (d) NPSG. The bottom row is the  $\text{DI}_2$  between L5t1 and ALlt2: (e) SH; (f) PP; (g) AMD; (h) NPSG.

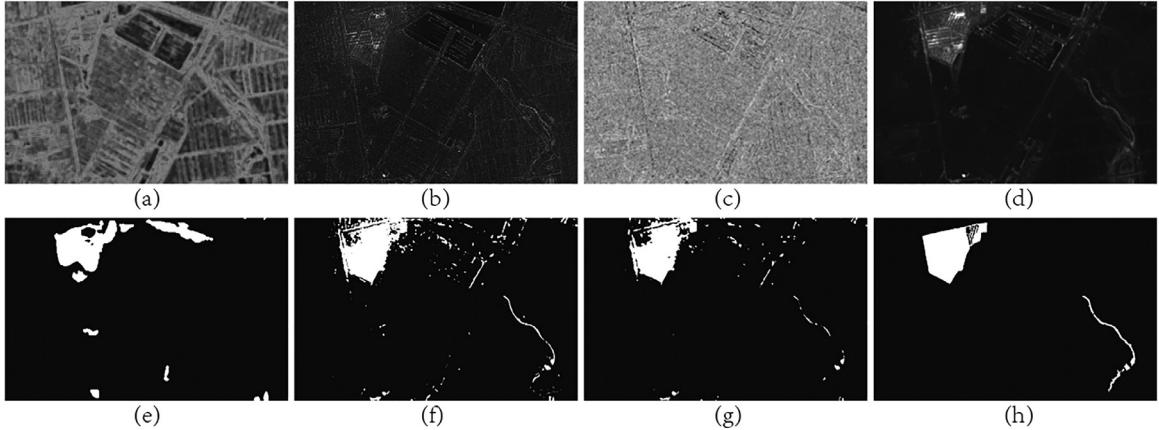


**Fig. 12.** ROC curves derived by varying of the threshold  $T$  in (16) on DIs generated by different methods on the heterogeneous datasets: (a) Sardinia; (b) Texas.

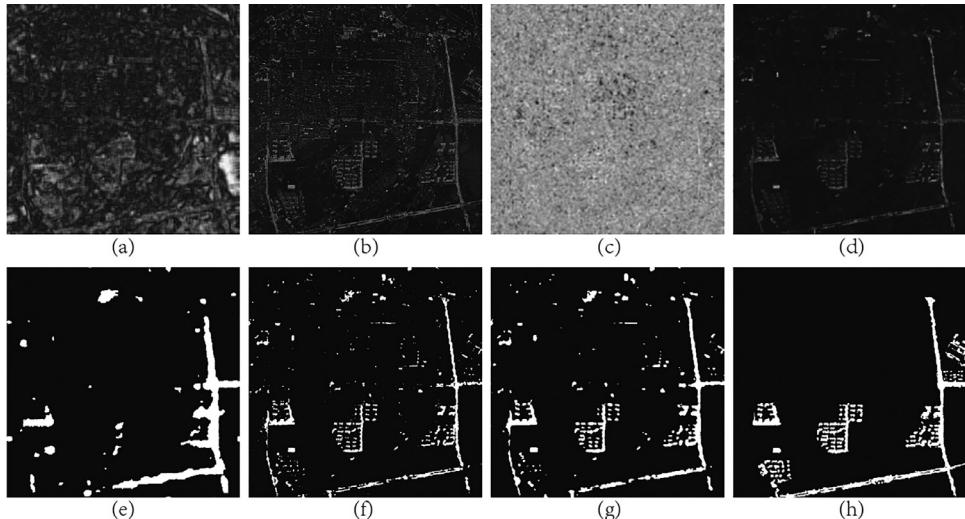
(5) as the distance criterion, at the same time, the estimated noise standard deviation of multispectral image is  $3.02\text{e-}2$ . Because of the low speckle level in SAR image, the DIs of these comparing methods can detect the changes between heterogeneous SAR and multispectral images with different degrees, as shown in Fig. 15(a)-(d). The corresponding ROC curves are shown in Fig. 16(c), and the quantitative criteria of AUC and Ddist are listed in Table 6. Due to poor performance of M3CD with low PCC and Kappa coefficient on this dataset as listed in Table 5, we show the CM of PP-Otsu with higher value instead of M3CD in Fig. 15(e). By comparing the CMs of different methods in Fig. 15(e)-(g) and Table 5, we can find that the proposed NPSG gain the highest PCC (94.07% of NPSG-Otsu) and Kappa coefficient (0.4185 of NPSG-Otsu), which is a lot ahead of the second PP method on this dataset with PCC 73.47%

(PP-Otsu) and Kappa coefficient 0.1486 (PP-Otsu). The analysis of these results shows that the NPSG generated DI and CMs outperform other DIs and CMs both in terms of qualitative vision and quantitative measurement, which means that the proposed non-local patches similarity based graph can establish a more robust connection between the heterogeneous images.

In addition, we can find that the similarity-based SH, PP and AMD don't always detect changes between heterogeneous images. For example, SH fails in the Texas and Shuguang datasets, PP fails in the Sardinia and Texas datasets, and AMD fails in the Sardinia, Shuguang and Wuhan datasets. However, only the NPSG method can succeed on all datasets. This may be due to two reasons: first, the imaging modality-invariant assumptions they use are no longer true in some complicated cases; second, these methods all have



**Fig. 13.** DIIs and binary CMs on *Shuguang* dataset. (a) DI of SH; (b) DI of PP; (c) DI of AMD; (d) DI of NPSG; (e) binary CM of M3CD; (f) binary CM of NPSG-Otsu; (g) binary CM of NPSG-PCAkm; (h) the reference image.



**Fig. 14.** DIIs and binary CMs on *Wuhan* dataset. (a) DI of SH; (b) DI of PP; (c) DI of AMD; (d) DI of NPSG; (e) binary CM of M3CD; (f) binary CM of NPSG-Otsu; (g) binary CM of NPSG-PCAkm; (h) the reference image.

**Table 4**  
Quantitative measures of binary CMs on *Wuhan* dataset.

| Methods    | FPR(%) | FNR(%) | PCC(%) | Kappa  |
|------------|--------|--------|--------|--------|
| M3CD       | 3.53   | 3.09   | 93.38  | 0.4758 |
| SH-Otsu    | 18.15  | 2.29   | 79.55  | 0.2147 |
| SH-PCAkm   | 18.07  | 1.92   | 80.01  | 0.2397 |
| PP-Otsu    | 8.28   | 2.25   | 89.47  | 0.3985 |
| PP-PCAkm   | 6.98   | 1.26   | 91.76  | 0.5220 |
| AMD-Otsu   | 59.64  | 1.68   | 38.68  | 0.0206 |
| AMD-PCAkm  | 56.19  | 1.65   | 42.15  | 0.0301 |
| NPSG-Otsu  | 1.49   | 3.19   | 95.32  | 0.5662 |
| NPSG-PCAkm | 2.00   | 2.20   | 95.80  | 0.6525 |

**Table 6**  
Quantitative measures (AUC and Ddist) on *Shuguang*, *Wuhan* and *California* datasets.

| Methods | <i>Shuguang</i> dataset |               | <i>Wuhan</i> dataset |               | <i>California</i> dataset |               |
|---------|-------------------------|---------------|----------------------|---------------|---------------------------|---------------|
|         | AUC                     | Ddist         | AUC                  | Ddist         | AUC                       | Ddist         |
| SH      | 0.7496                  | 0.9719        | 0.8049               | 1.0607        | 0.7263                    | 0.9548        |
| PP      | 0.8549                  | 1.0835        | 0.8662               | 1.1316        | 0.8362                    | 1.0807        |
| AMD     | 0.6195                  | 0.8322        | 0.5780               | 0.7871        | 0.7572                    | 0.9983        |
| NPSG    | <b>0.9890</b>           | <b>1.3562</b> | <b>0.9591</b>        | <b>1.2694</b> | <b>0.9111</b>             | <b>1.1709</b> |

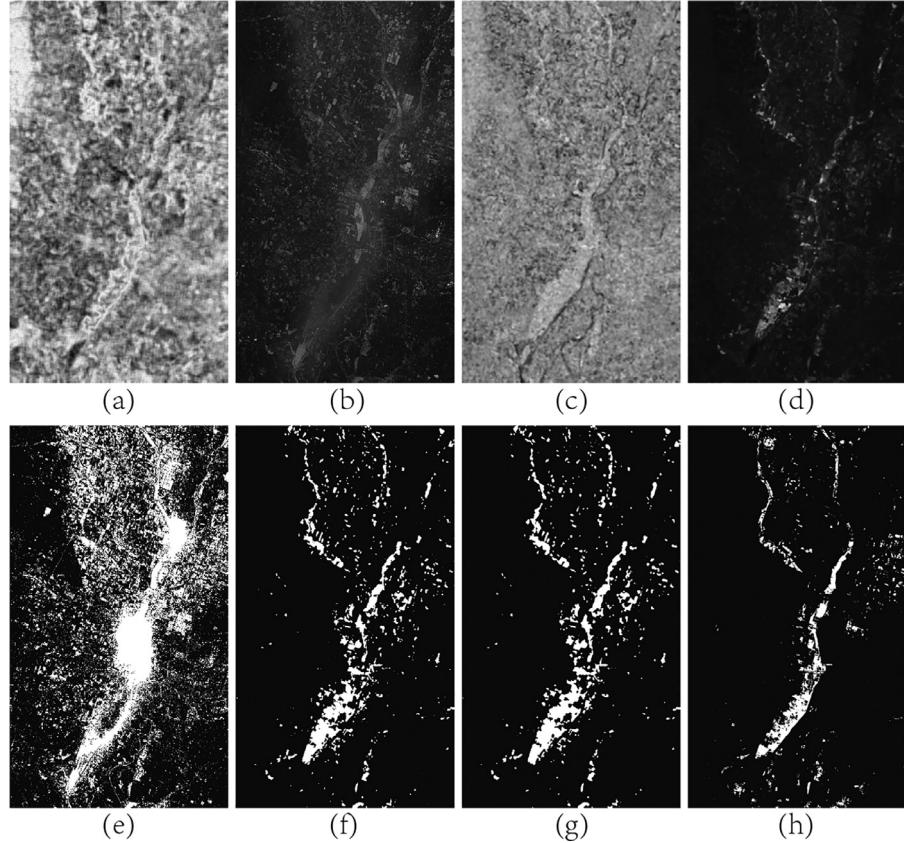
some problems of mixing heterogeneous data when they generate the difference images. However, the proposed NPSG based heterogeneous CD method can overcome these shortcomings, so it can achieve more robust and better results.

**Table 5**  
Quantitative measures of binary CMs on *California* dataset.

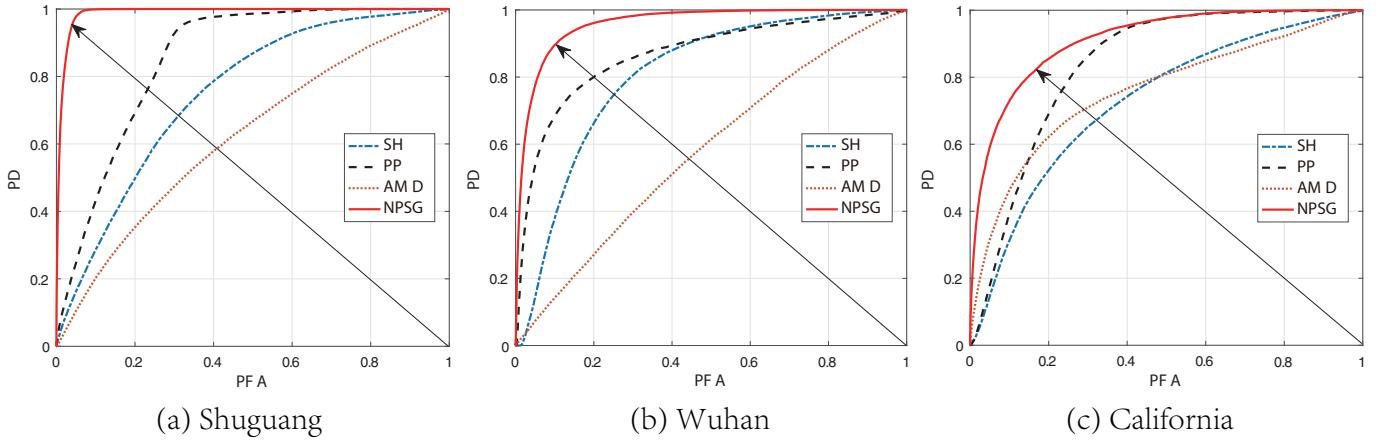
| Methods    | FPR(%) | FNR(%) | PCC(%) | Kappa  |
|------------|--------|--------|--------|--------|
| M3CD       | 40.41  | 2.05   | 57.53  | 0.0208 |
| SH-Otsu    | 44.53  | 0.91   | 54.56  | 0.0565 |
| SH-PCAkm   | 44.57  | 0.83   | 54.60  | 0.0595 |
| PP-Otsu    | 25.75  | 0.78   | 73.47  | 0.1486 |
| PP-PCAkm   | 34.62  | 0.24   | 65.16  | 0.1229 |
| AMD-Otsu   | 56.28  | 0.68   | 43.04  | 0.0363 |
| AMD-PCAkm  | 53.78  | 0.72   | 45.51  | 0.0403 |
| NPSG-Otsu  | 3.97   | 1.96   | 94.07  | 0.4185 |
| NPSG-PCAkm | 4.08   | 1.93   | 93.99  | 0.4178 |

### 3.4. Comparison of computational time

The main space and time complexity of the NPSG based CD method is concentrating on the calculation of the nonlocal patch similarity structure difference (step 1 of Algorithm 1 listed in Table 1). As can be seen from Table 1, the structure difference  $f_{(m,n)}^X$  (or  $f_{(m,n)}^Y$ ) is calculated on a patch by patch basis.



**Fig. 15.** Dis and binary CMs on *California* dataset. (a) DI of SH; (b) DI of PP; (c) DI of AMD; (d) DI of NPSG; (e) binary CM of PP-Otsu; (f) binary CM of NPSG-Otsu; (g) binary CM of NPSG-PCAkm; (h) the reference image.



**Fig. 16.** ROC curves derived by varying of the threshold  $T$  in (16) on Dis generated by different methods on the heterogeneous datasets: (a) *Shuguang*; (b) *Wuhan*; (c) *California*.

1) For the space complexity, storing the distance set of  $d(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)})$  (or  $d(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)})$ ) for all  $(i, j) \in \Omega$  requires  $\mathcal{O}\left(\frac{\xi_s^2}{\Delta_s^2}\right)$ , and storing the structure difference set of  $F^X$  (or  $F^Y$ ) for all pixels requires  $\mathcal{O}\left(\frac{p_s^2 MN}{\Delta_p^2}\right)$ . Therefore, the space complexity of the proposed algorithm is very low.

2) For the time complexity, calculating the distance between patches  $d(\mathbf{X}_{(m,n)}, \mathbf{X}_{(i,j)})$  (or  $d(\mathbf{Y}_{(m,n)}, \mathbf{Y}_{(i,j)})$ ) for all  $(m, n) \in \Lambda$  and  $(i, j) \in \Omega$  requires  $\mathcal{O}\left(p_s^2 C_X \frac{MN}{\Delta_p^2} \frac{\xi_s^2}{\Delta_s^2}\right)$ , and constructing the  $K$ -nearest NPSG NPSG  $G_{\mathbf{X}_{(m,n)}}^K$  (or  $G_{\mathbf{Y}_{(m,n)}}^K$ ) for all  $(m, n) \in \Lambda$  requires

$\mathcal{O}\left(\frac{MN}{\Delta_p^2} \frac{\xi_s^2}{\Delta_s^2} \log\left(\frac{\xi_s^2}{\Delta_s^2}\right)\right)$  by using some accelerated sorting algorithms, such as the Block sort or Tree sort.

The time complexity of the proposed algorithm is relatively high in the abovementioned theoretical analysis, which requires  $\mathcal{O}\left(\left(p_s^2 C_X + \log\left(\frac{\xi_s^2}{\Delta_s^2}\right)\right) \frac{MN}{\Delta_p^2} \frac{\xi_s^2}{\Delta_s^2}\right)$ . Table 7 reports the CPU times of different methods on *Sardinia*, *Wuhan* and *Shuguang* datasets. The parameters of NPSG are set as  $p_s = 5$ ,  $\xi_s = 100$  and  $K = 35$ . The algorithms, excluding M3CD, were performed in MATLAB 2016a running on a Windows laptop with Intel(R) Core(TM) i7-8550U CPU and 8 GB of RAM. The C++ code of M3CD algorithm was executed

**Table 7**  
Computational time (seconds) of different methods.

| Datasets | Image size                   | M3CD    | SH      | PP     | AMD    | NPSG with different $\Delta_s$ and $\Delta_p$ |                           |                           |
|----------|------------------------------|---------|---------|--------|--------|---|---------------------------|---------------------------|
|          |                              |         |         |        |        | $\Delta_s = \Delta_p = 2$                     | $\Delta_s = \Delta_p = 3$ | $\Delta_s = \Delta_p = 4$ |
| Sardinia | $300 \times 412 \times 1(3)$ | 2206.64 | 1083.31 | 205.72 | 76.97  | 652.89  | 130.23                    | 42.73                     |
| Wuhan    | $495 \times 503 \times 1(3)$ | 2616.22 | 2286.94 | 429.48 | 169.97 | 1360.54                                       | 287.27                    | 98.64                     |
| Shuguang | $593 \times 921 \times 1(3)$ | 4691.82 | 7957.84 | 979.46 | 304.50 | 2948.06                                       | 602.38                    | 196.62                    |

in a Linux computer with Intel(R) Xeon(R) Silver 4110 CPU and 31 GB of RAM.

From Table 7, we can find that the two most time-consuming methods are M3CD and SH, followed by NPSG, PP and AMD. In addition, we can also find that the speed of the NPSG can be significantly improved by using the search step  $\Delta_s$  and target patch step  $\Delta_p$  as analyzed in Section 2.2 and Section 3.2. Meanwhile, from the above complexity analysis and Table 7, it can be found that several strategies can be considered in NPSG to avoid a large amount of computational time:

1) Using the patch-wisely parallel solution. Due to the target patch-wise independence property of the structure difference calculation (step 1 in Algorithm 1), NPSG can be easily accelerated by the parallel solution.

2) Using large search step and target patch step to reduce the search space of similar patches and the amount of target patches, seeking the balance between computation time and change detection accuracy.

3) Using the superpixel as the basic unit instead of the square patch. Compared with the square image patch, the superpixel has two main advantages: it can maintain the structure and edge of the object, and the interior of each superpixel is homogeneous (representing the same kind of object); it can greatly reduce the computational complexity, especially for large-scale high-resolution images. However, we have two problems to solve when using the superpixel as the basic unit: one is the accurate superpixel segmentation (especially for SAR images), and the other is to find a suitable criterion to measure the similarity (or distance) of superpixels, which is also our future work.

#### 4. Conclusion

This paper mainly focuses on the change detection for heterogeneous remote sensing images. Since the heterogeneous images refer to distinct feature representations of ground object by different imaging mechanisms, it is difficult to measure the changes between heterogeneous images by direct comparison. We present an unsupervised change detection method to make the heterogeneous data comparable, which is based on the structure consistency between images. To exploit the structure information, we construct the NPSG for each image based on the nonlocal self-similarity. Therefore, the change level can be measured by how much the graph structure of one image still conforms to that of the other image. Then, the NPSG based CD method can be implemented in two steps. It first constructs the K-nearest NPSG for each input image, then map the K-nearest NPSG of one image to the other image, and compare the difference between the graph and mapped graph to obtain the DI. Secondly, it uses the thresholding or clustering method to obtain the binary CM with the combined DI. In the process of obtaining the combined DI, the proposed method can avoid the leakage of heterogeneous data by comparing the graph difference on the same domain, and reasonably fuse the forward and backward detection results based on the statistical distribution. In this way, it can achieve robust and effective performance in the difficult heterogeneous CD task.

In this paper, we only consider two commonly used noise distribution models (AWGN and multiplicative Gamma distribution noise models) to measure the distances for the optical image and SAR image, but there are some other noise models not considered, such as the complex Wishart distribution based polarimetric SAR (PolSAR) data. Therefore, how to apply this model to PolSAR is our next research work. Moreover, as the NPSG represents the structure information of patches, it is also applicable to the land cover classification of remote sensing images, especially to the classification based on fusion of heterogeneous data. In the future, we will also evaluate the NPSG model on other complex heterogeneous data and plan to extend the NPSG model to the classification problem.

#### Declaration of Competing Interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled "Nonlocal patch similarity based heterogeneous remote sensing change detection".

#### Acknowledgment

This work is partially supported by the National Natural Science Foundation of China (61971426). The authors would like to thank the editors and the anonymous reviewers for their careful reading of an earlier version of this article and constructive suggestions that improved the presentation of this article.

#### Appendix A. Statistical characteristics of the distance criteria

In this appendix, we compute distribution, mean, and variance of the random variables (17), (19) and (20). We denote two observed signals  $x, y$  and their corresponding true constant signals  $z_x, z_y$ , and their correlation noise  $u_x, u_y$ , respectively. Next, we analyze the statistical characteristics of the distance criteria (17), (19) and (20) under different models.

##### A1. The additive noise model of optical image

We have  $x = z_x + u_x$  and  $y = z_y + u_y$ , where  $u_x$  and  $u_y$  are i.i.d. random variables that  $u_x, u_y \sim \mathcal{N}(0, \sigma^2)$ . We analyze the random variable of distance criterion  $d = (x - y)^2$ .

1. Fist case, with the uniform signal,  $z_x = z_y$ , then we have  $d = (u_x - u_y)^2$ . Easily, we have the expected value of  $d$  as

$$E(d) = 2\sigma^2 \quad (\text{A.1})$$

Let  $A = u_x - u_y$ , we have  $A$  obeys the Gaussian distribution  $A \sim \mathcal{N}(0, 2\sigma^2)$ . Let  $B = \left(\frac{A}{\sqrt{2}\sigma}\right)^2$ , we have that  $B$  is distributed according to the chi-square distribution with 1 degrees of freedom, denoted as  $B \sim \chi^2(1)$ . The pdf of the chi-square distribution is

$$p_B(b; 1) = \frac{b^{-1/2} e^{-b/2}}{2^{1/2} \Gamma(1/2)} \mathbf{1}(b) \quad (\text{A.2})$$

with  $\mathbf{1}(b)$  the unit-step function. As  $d = 2\sigma^2 B$ , then we have the pdf of random variable  $d$  is

$$p_d(d) = p_B\left(\frac{d}{2\sigma^2}, 1\right) \frac{1}{2\sigma^2} = \frac{d^{-1/2}}{2\sigma\Gamma(1/2)} e^{-\frac{d}{4\sigma^2}} \mathbf{1}(d) \quad (\text{A.3})$$

2. Second case, the true signals with different values  $z_x = v_1$  and  $z_y = v_2$ , then the  $x$  and  $y$  are i.i.d. random variables with different means that  $x \sim \mathcal{N}(v_1, \sigma^2)$  and  $y \sim \mathcal{N}(v_2, \sigma^2)$ . Let  $\eta = v_1 - v_2$  and  $A = \frac{x-y-\eta}{\sqrt{2}\sigma}$ , then we have  $A \sim \mathcal{N}(0, 1)$ . As  $d = 2\sigma^2\left(A + \frac{\eta}{\sqrt{2}\sigma}\right)^2$ , then the cumulative distribution function (cdf) of  $d$  can be written as

$$P(d) = \int_{-\frac{\sqrt{d}-\eta}{\sqrt{2}\sigma}}^{\frac{\sqrt{d}-\eta}{\sqrt{2}\sigma}} \phi(x) dx \quad (\text{A.4})$$

where  $\phi(\cdot)$  is the standard normal density. Then we have the pdf of random variable  $d$  as

$$\begin{aligned} p_d(d) &= \frac{1}{2\sqrt{2}\sigma} \left[ \phi\left(\frac{\sqrt{d}-\eta}{\sqrt{2}\sigma}\right) + \phi\left(\frac{\sqrt{d}+\eta}{\sqrt{2}\sigma}\right) \right] \mathbf{1}(d) \\ &= \frac{1}{4\sqrt{\pi d}\sigma} \left[ e^{-\frac{(\sqrt{d}-\eta)^2}{4\sigma^2}} + e^{-\frac{(\sqrt{d}+\eta)^2}{4\sigma^2}} \right] \mathbf{1}(d) \end{aligned} \quad (\text{A.5})$$

## 2.2. The multivariate noise model of SAR image

We have  $x = z_x u_x$  and  $y = z_y u_y$ , where  $u_x$  and  $u_y$  are i.i.d. Gamma distributed random variables with shape parameter  $L$  and inverse scale parameter  $L$  as  $u_x, u_y \sim \text{Gamma}(L, L)$ .

### 2.2.1. The distance criterion of $d = 2L \log\left(\frac{x+y}{2\sqrt{xy}}\right)$

1. First case, with the uniform signal,  $z_x = z_y$ , we have

$$d = 2L \log\left(\frac{u_x+u_y}{2\sqrt{u_x u_y}}\right) \quad (\text{A.6})$$

Let  $A = \log\left(\frac{u_x+u_y}{2\sqrt{u_x u_y}}\right)$ , then we calculate the expected value and pdf of  $A$  by directly using the conclusion in [46]

$$E(A) = -\frac{1}{2}[\varphi(0, L) - \varphi(0, L + 1/2)] \quad (\text{A.7})$$

$$p_A(A) = \frac{1}{2^{2L-2}B(L,L)} \frac{e^{-2LA}}{\sqrt{1-e^{-2A}}} \mathbf{1}(A) \quad (\text{A.8})$$

where  $\varphi(m, x)$  is the  $m$ -order Polygamma function defined as the  $(m+1)$ th derivative of the Digamma function  $\varphi(m, x) = \frac{d^{m+1} \log(\Gamma(x))}{dx^{m+1}}$ , and  $B(\cdot, \cdot)$  is a Beta function with  $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ . Therefore, we have the expected value and pdf of  $d = 2LA$  as

$$E(d) = -L[\varphi(0, L) - \varphi(0, L + 1/2)] \quad (\text{A.9})$$

$$p_d(d) = p_A(d/2L) \frac{1}{2L} = \frac{1}{2^{2L-1}LB(L,L)} \frac{e^{-d}}{\sqrt{1-e^{-d/L}}} \mathbf{1}(d) \quad (\text{A.10})$$

2. Second case, the true signals with different values  $z_x = v_1$  and  $z_y = v_2$ . Let  $\rho = z_x/z_y$  and  $A = u_x/u_y$ , then we have that the distance criterion becomes

$$d = 2L \log\left(\frac{x+y}{2\sqrt{xy}}\right) = 2L \log\left(\frac{1+\rho A}{2\sqrt{\rho A}}\right) \quad (\text{A.11})$$

As  $A = u_x/u_y$  and  $u_x, u_y \sim \text{Gamma}(L, L)$ , then  $A$  is beta-distributed as  $A \sim \text{Beta}(L, L)$ , and the pdf of  $A$  is

$$p_A(A) = \frac{A^{L-1}(1+A)^{-2L}}{B(L,L)} \mathbf{1}(A) \quad (\text{A.12})$$

Define  $w = \frac{1+\rho A}{2\sqrt{\rho A}}$ , then we have  $A = \frac{1}{\rho} \left( \frac{w+\sqrt{w^2-1}}{w-\sqrt{w^2-1}} \right)$  or  $A = \frac{1}{\rho} \left( \frac{w-\sqrt{w^2-1}}{w+\sqrt{w^2-1}} \right)$ . The cdf of  $w$  can be written as

$$P(w) = \int_{-\frac{1}{\rho}}^{\frac{\alpha}{\rho}} p_A(A) dA \quad (\text{A.13})$$

where  $\alpha = \frac{w+\sqrt{w^2-1}}{w-\sqrt{w^2-1}}$ . Then the pdf of  $w$  can be computed as

$$\begin{aligned} p_w(w) &= p_A\left(\frac{1}{\rho} \left( \frac{w+\sqrt{w^2-1}}{w-\sqrt{w^2-1}} \right)\right) \frac{2}{\rho\sqrt{w^2-1}} \left( \frac{w+\sqrt{w^2-1}}{w-\sqrt{w^2-1}} \right) \\ &\quad + p_A\left(\frac{1}{\rho} \left( \frac{w-\sqrt{w^2-1}}{w+\sqrt{w^2-1}} \right)\right) \frac{2}{\rho\sqrt{w^2-1}} \left( \frac{w-\sqrt{w^2-1}}{w+\sqrt{w^2-1}} \right) \\ &= \frac{2}{B(L,L)\sqrt{w^2-1}} \left( \left( \frac{\alpha}{\rho} \right)^L \left( 1 + \frac{\alpha}{\rho} \right)^{-2L} \right. \\ &\quad \left. + \left( \frac{1}{\rho\alpha} \right)^L \left( 1 + \frac{1}{\rho\alpha} \right)^{-2L} \right) \mathbf{1}(w) \\ &= \frac{2}{B(L,L)\sqrt{w^2-1}} \left( \left( \frac{\rho\alpha}{(\rho+\alpha)^2} \right)^L + \left( \frac{\rho\alpha}{(\rho\alpha+1)^2} \right)^L \right) \mathbf{1}(w) \end{aligned} \quad (\text{A.14})$$

By substituting  $\frac{\rho\alpha}{(\rho+\alpha)^2} = \frac{\rho}{[(\rho+1)w+(1-\rho)\sqrt{w^2-1}]^2}$  and  $\frac{\rho\alpha}{(\rho\alpha+1)^2} = \frac{\rho}{[(\rho+1)w+(\rho-1)\sqrt{w^2-1}]^2}$ , (A.14) can be written as

$$\begin{aligned} p_w(w) &= \frac{2\rho^L}{B(L,L)\sqrt{w^2-1}} \left( \left( (\rho+1)w + (1-\rho)\sqrt{w^2-1} \right)^{-2L} \right. \\ &\quad \left. + \left( (\rho+1)w + (\rho-1)\sqrt{w^2-1} \right)^{-2L} \right) \mathbf{1}(w) \end{aligned} \quad (\text{A.15})$$

As  $d = 2L \log w$ , the pdf of  $d$  is

$$\begin{aligned} p_d(d) &= p_w(e^{d/2L}) \frac{e^{d/2L}}{2L} \\ &= \frac{\rho^L}{LB(L,L)\sqrt{1-e^{-d/L}}} \\ &\quad \times \left( \left( (\rho+1)e^{d/2L} + (1-\rho)\sqrt{e^{d/L}-1} \right)^{-2L} \right. \\ &\quad \left. + \left( (\rho+1)e^{d/2L} + (\rho-1)\sqrt{e^{d/L}-1} \right)^{-2L} \right) \mathbf{1}(d) \end{aligned} \quad (\text{A.16})$$

We can easily find that (A.10) is the special case of (A.16) with  $\rho = 1$ .

### 2.2.2. The distance criterion of $d = [\log\left(\frac{x}{y}\right)]^2$

1. First case, the uniform signal,  $z_x = z_y$ . Let  $A = x/y$ , then  $A \sim \text{Beta}(L, L)$ , and the pdf of  $A$  is given in (A.12). Define  $w = \log A$ , the pdf of  $w$  is

$$p_w(w) = p_A(e^w) e^w = \frac{e^{wL}(1+e^w)^{-2L}}{B(L,L)} \mathbf{1}(w) = \frac{(e^{-w/2}+e^{w/2})^{-2L}}{B(L,L)} \mathbf{1}(w) \quad (\text{A.17})$$

As  $d = w^2$ , the cdf of  $d$  is

$$P(d) = \int_{-\sqrt{d}}^{\sqrt{d}} p_w(w) dw \quad (\text{A.18})$$

Then, the pdf of  $d$  is

$$p_d(d) = p_w(\sqrt{d}) \frac{1}{2\sqrt{d}} + p_w(-\sqrt{d}) \frac{1}{2\sqrt{d}} = \frac{(e^{-\sqrt{d}/2}+e^{\sqrt{d}/2})^{-2L}}{\sqrt{dB(L,L)}} \mathbf{1}(d) \quad (\text{A.19})$$

Meanwhile, the expected value and variance value of Logarithmically transformed speckle is given in [50] as

$$E(\log u_x) = E(\log u_y) = \varphi(0, L) - \log L \quad (\text{A.20})$$

$$\text{var}(\log u_x) = \text{var}(\log u_y) = \varphi(1, L) \quad (\text{A.21})$$

Then the expected value of  $d$  is

$$E(d) = E[(\log u_x - \log u_y)^2] = 2\varphi(1, L) \quad (\text{A.22})$$

2. Second case, the true signals with different values  $z_x = v_1$  and  $z_y = v_2$ . Let  $\rho = z_x/z_y$  and  $w = \log \frac{u_x}{u_y}$ , then we have that the distance criterion becomes  $d = (w + \ln \rho)^2$ . Then, the cdf of  $d$  is

$$P(d) = \int_{-\sqrt{d}-\ln \rho}^{\sqrt{d}-\ln \rho} p_w(w) dw \quad (\text{A.23})$$

where  $p_w(w)$  is given in (A.17). Then the pdf of  $d$  is

$$\begin{aligned} p_d(d) &= p_w\left(\sqrt{d} - \ln \rho\right) \frac{1}{2\sqrt{d}} + p_w\left(-\sqrt{d} - \ln \rho\right) \frac{1}{2\sqrt{d}} \\ &= \frac{\left(e^{\frac{-\sqrt{d}+\ln \rho}{2}} + e^{\frac{\sqrt{d}-\ln \rho}{2}}\right)^{-2L} + \left(e^{\frac{\sqrt{d}+\ln \rho}{2}} + e^{\frac{-\sqrt{d}-\ln \rho}{2}}\right)^{-2L}}{2\sqrt{d}B(L, L)} \mathbf{1}(d) \end{aligned} \quad (\text{A.24})$$

## References

- [1] A. SINGH, Review article digital change detection techniques using remotely-sensed data, *Int. J. Remote Sens.* 10 (1989) 989–1003, doi:[10.1080/01431168908903939](https://doi.org/10.1080/01431168908903939).
- [2] J.L. Gil-Yepes, L.A. Ruiz, J.A. Recio, A. Balaguer-Beser, T. Hermosilla, Description and validation of a new set of object-based temporal geostatistical features for land-use/land-cover change detection, *ISPRS J. Photogramm. Remote Sens.* 121 (2016) 77–91, doi:[10.1016/j.isprsjprs.2016.08.010](https://doi.org/10.1016/j.isprsjprs.2016.08.010).
- [3] H. Taubenbock, T. Esch, A. Felbier, M. Wiesner, A. Roth, S. Dech, Monitoring urbanization in mega cities from space, *Remote Sens. Environ.* 117 (2012) 162–176, doi:[10.1016/j.rse.2011.09.015](https://doi.org/10.1016/j.rse.2011.09.015).
- [4] Y. Ban, O.A. Yousif, Multitemporal spaceborne SAR data for urban change detection in China, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 5 (2012) 1087–1094, doi:[10.1109/JSTARS.2012.2201135](https://doi.org/10.1109/JSTARS.2012.2201135).
- [5] D. Brunner, G. Lemoine, L. Bruzzone, Earthquake damage assessment of buildings using VHR optical and SAR imagery, *IEEE Trans. Geosci. Remote Sens.* 48 (2010) 2403–2420, doi:[10.1109/TGRS.2009.2038274](https://doi.org/10.1109/TGRS.2009.2038274).
- [6] M. Zhao, Q. Ling, F. Li, An iterative feedback-based change detection algorithm for flood mapping in SAR images, *IEEE Geosci. Remote Sens. Lett.* 16 (2019) 231–235, doi:[10.1109/LGRS.2018.2871849](https://doi.org/10.1109/LGRS.2018.2871849).
- [7] Y. Ban, O. Yousif, Change detection techniques: a review, *Multitemporal Remote Sensing*, 2016, doi:[10.1007/978-3-319-47037-5\\_2](https://doi.org/10.1007/978-3-319-47037-5_2), 19–43.
- [8] E.F. Lambin, A.H. Strahler, Change-vector analysis in multitemporal space: a tool to detect and categorize land-cover change processes using high temporal-resolution satellite data, *Remote Sens. Environ.* 48 (1994) 231–244, doi:[10.1016/0034-4257\(94\)90144-9](https://doi.org/10.1016/0034-4257(94)90144-9).
- [9] F. Bovolo, S. Marchesi, L. Bruzzone, A framework for automatic and unsupervised detection of multiple changes in multitemporal images, *IEEE Trans. Geosci. Remote Sens.* 50 (2012) 2196–2212, doi:[10.1109/TGRS.2011.2171493](https://doi.org/10.1109/TGRS.2011.2171493).
- [10] A.A. Nielsen, K. Conradsen, J.J. Simpson, Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: new approaches to change detection studies, *Remote Sens. Environ.* 64 (1998) 1–19, doi:[10.1016/S0034-4257\(97\)00162-4](https://doi.org/10.1016/S0034-4257(97)00162-4).
- [11] A.A. Nielsen, The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data, *IEEE Trans. Image Process.* 16 (2007) 463–478, doi:[10.1109/TIP.2006.888195](https://doi.org/10.1109/TIP.2006.888195).
- [12] G. Moser, S.B. Serpico, Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery, *IEEE Trans. Geosci. Remote Sens.* 44 (2006) 2972–2982, doi:[10.1109/TGRS.2006.876288](https://doi.org/10.1109/TGRS.2006.876288).
- [13] M. Gong, Z. Zhou, J. Ma, Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering, *IEEE Trans. Image Process.* 21 (2012) 2141–2151, doi:[10.1109/TIP.2011.2170702](https://doi.org/10.1109/TIP.2011.2170702).
- [14] T. Celik, Unsupervised change detection in satellite images using principal component analysis and k-means clustering, *IEEE Geosci. Remote Sens. Lett.* 6 (2009) 772–776, doi:[10.1109/LGRS.2009.2025059](https://doi.org/10.1109/LGRS.2009.2025059).
- [15] H.-C. Li, T. Celik, N. Longbotham, W.J. Emery, Gabor feature based unsupervised change detection of multitemporal SAR images based on two-level clustering, *IEEE Geosci. Remote Sens. Lett.* 12 (2015) 2458–2462, doi:[10.1109/LGRS.2015.2484220](https://doi.org/10.1109/LGRS.2015.2484220).
- [16] K. Mubea, G. Menz, Monitoring land-use change in Nakuru (Kenya) using multi-sensor satellite data, *Adv. Remote Sens.* 01 (2012) 74–84, doi:[10.4236/ars.2012.13008](https://doi.org/10.4236/ars.2012.13008).
- [17] W. Zhou, A. Troy, M. Grove, Object-based land cover classification and change analysis in the baltimore metropolitan area using multitemporal high resolution remote sensing data, *Sensors* 8 (2008) 1613–1636, doi:[10.3390/s8031613](https://doi.org/10.3390/s8031613).
- [18] M. Volpi, D. Tuia, F. Bovolo, M. Kaniewski, L. Bruzzone, Supervised change detection in VHR images using contextual information and support vector machines, *Int. J. Appl. Earth Obs. Geoinf.* 20 (2013) 77–85, doi:[10.1016/j.jag.2011.10.013](https://doi.org/10.1016/j.jag.2011.10.013).
- [19] L. Wan, Y. Xiang, H. You, A post-classification comparison method for SAR and optical images change detection, *IEEE Geosci. Remote Sens. Lett.* 16 (2019) 1026–1030, doi:[10.1109/LGRS.2019.2892432](https://doi.org/10.1109/LGRS.2019.2892432).
- [20] L. Wan, Y. Xiang, H. You, An object-based hierarchical compound classification method for change detection in heterogeneous optical and SAR images, *IEEE Trans. Geosci. Remote Sens.* 57 (2019) 9941–9959, doi:[10.1109/TGRS.2019.2930322](https://doi.org/10.1109/TGRS.2019.2930322).
- [21] J. Liu, M. Gong, K. Qin, P. Zhang, A deep convolutional coupling network for change detection based on heterogeneous optical and radar images, *IEEE Trans. Neural Netw. Learn. Syst.* 29 (2018) 545–559, doi:[10.1109/TNNLS.2016.2636227](https://doi.org/10.1109/TNNLS.2016.2636227).
- [22] T. Zhan, M. Gong, X. Jiang, S. Li, Log-based transformation feature learning for change detection in heterogeneous images, *IEEE Geosci. Remote Sens. Lett.* 15 (2018) 1352–1356, doi:[10.1109/LGRS.2018.2843385](https://doi.org/10.1109/LGRS.2018.2843385).
- [23] X. Niu, M. Gong, T. Zhan, Y. Yang, A conditional adversarial network for change detection in heterogeneous images, *IEEE Geosci. Remote Sens. Lett.* 16 (2019) 45–49, doi:[10.1109/LGRS.2018.2868704](https://doi.org/10.1109/LGRS.2018.2868704).
- [24] L.T. Luppino, M. Kampffmeyer, F.M. Bianchi, G. Moser, S.B. Serpico, R. Jenssen, S.N. Anfinsen, Deep image translation with an affinity-based change prior for unsupervised multimodal change detection, *ArXiv Preprint arXiv:2001.04271*.
- [25] G. Mercier, G. Moser, S.B. Serpico, Conditional copulas for change detection in heterogeneous remote sensing images, *IEEE Trans. Geosci. Remote Sens.* 46 (2008) 1428–1441, doi:[10.1109/TGRS.2008.916476](https://doi.org/10.1109/TGRS.2008.916476).
- [26] J. Prendes, M. Chabert, F. Pascal, A. Giros, J.Y. Tourneret, A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors, *IEEE Trans. Image Process.* 24 (2015) 799–812, doi:[10.1109/TIP.2014.2387013](https://doi.org/10.1109/TIP.2014.2387013).
- [27] J. Prendes, M. Chabert, F. Pascal, A. Giros, 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.* 9 (2016) 1889–1921, doi:[10.1137/15M1047908](https://doi.org/10.1137/15M1047908).
- [28] V. Alberga, Similarity measures of remotely sensed multi-sensor images for change detection applications, *Remote Sens.* 1 (2009) 122–143.
- [29] J. Inglaña, A. Giro, On the possibility of automatic multisensor image registration, *IEEE Trans. Geosci. Remote Sens.* 42 (2004) 2104–2120, doi:[10.1109/TGRS.2004.835294](https://doi.org/10.1109/TGRS.2004.835294).
- [30] R.P. Woods, J.C. Mazziotta, S.R. Cherry, MRI-PET registration with automated algorithm, *J. Comput. Assist. Tomogr.* 17 (1993) 536–546, doi:[10.1007/BF0004728-199307000-00004](https://doi.org/10.1007/BF0004728-199307000-00004).
- [31] C. Nikou, F. Heitz, J.-P. Armsbach, I.J. Namer, Mesures de similarité robustes pour le recalage d'images médicales volumiques multimodales, *Traitem. Du Signal* 16 (1999) 255–272.
- [32] G.J. Szekely, M.L. Rizzo, N.K. Bakirov, Measuring and testing dependence by correlation of distances, *Ann. Stat.* 35 (2007) 2769–2794, doi:[10.1214/09053607000000505](https://doi.org/10.1214/09053607000000505).
- [33] L. Wan, T. Zhang, H.J. You, Multi-sensor remote sensing image change detection based on sorted histograms, *Int. J. Remote Sens.* 39 (2018) 3753–3775, doi:[10.1080/01431161.2018.1448481](https://doi.org/10.1080/01431161.2018.1448481).
- [34] B. Ayhan, C. Kwan, A new approach to change detection using heterogeneous images, in: 2019 IEEE 10th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON), 2019, pp. 0192–0197, doi:[10.1109/UEMCON47517.2019.8993038](https://doi.org/10.1109/UEMCON47517.2019.8993038).
- [35] C. Kwan, B. Ayhan, J. Larkin, L. Kwan, S. Bernabe, A. Plaza, Performance of change detection algorithms using heterogeneous images and extended multi-attribute profiles (EMAPs), *Remote Sens.* 11 (2019) 2377, doi:[10.3390/rs11202377](https://doi.org/10.3390/rs11202377).
- [36] L.T. Luppino, F.M. Bianchi, G. Moser, S.N. Anfinsen, Unsupervised image regression for heterogeneous change detection, *IEEE Trans. Geosci. Remote Sens.* 57 (2019) 9960–9975, doi:[10.1109/TGRS.2019.2930348](https://doi.org/10.1109/TGRS.2019.2930348).
- [37] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Trans. Syst. Man. Cybern.* 9 (1979) 62–66, doi:[10.1109/TSMC.1979.4310076](https://doi.org/10.1109/TSMC.1979.4310076).
- [38] C.-A. Deledalle, L. Denis, F. Tupin, Iterative weighted maximum likelihood denoising with probabilistic patch-based weights, *IEEE Trans. Image Process.* 18 (2009) 2661–2672, doi:[10.1109/TIP.2009.2029593](https://doi.org/10.1109/TIP.2009.2029593).
- [39] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3-d transform-domain collaborative filtering, *IEEE Trans. Image Process.* 16 (2007) 2080–2095, doi:[10.1109/TIP.2007.901238](https://doi.org/10.1109/TIP.2007.901238).
- [40] D. Cozzolino, S. Parrilli, G. Scarpa, G. Poggi, L. Verdoliva, Fast adaptive nonlocal SAR despeckling, *IEEE Geosci. Remote Sens. Lett.* 11 (2014) 524–528, doi:[10.1109/LGRS.2013.2271650](https://doi.org/10.1109/LGRS.2013.2271650).
- [41] Y. Sun, L. Lei, D. Guan, X. Li, G. Kuang, SAR image change detection based on nonlocal low-rank model and two-level clustering, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 13 (2020) 293–306, doi:[10.1109/JSTARS.2019.2960518](https://doi.org/10.1109/JSTARS.2019.2960518).
- [42] A. Sanfelix, R. Alquezar, J. Andrade, J. Climent, F. Serratosa, J. Verges, Graph-based representations and techniques for image processing and image analysis, *Pattern Recognit.* 35 (2002) 639–650, doi:[10.1016/S0031-3203\(01\)00066-8](https://doi.org/10.1016/S0031-3203(01)00066-8).
- [43] C. de Mauro, M. Diligenti, M. Gori, M. Maggini, Similarity learning for graph-based image representations, *Pattern Recognit. Lett.* 24 (2003) 1115–1122, doi:[10.1016/S0167-8655\(02\)00258-1](https://doi.org/10.1016/S0167-8655(02)00258-1).
- [44] M.-T. Pham, G. Mercier, J. Michel, Change detection between SAR images using a pointwise approach and graph theory, *IEEE Trans. Geosci. Remote Sens.* 54 (2016) 2020–2032, doi:[10.1109/TGRS.2015.2493730](https://doi.org/10.1109/TGRS.2015.2493730).
- [45] C.-A. Deledalle, L. Denis, F. Tupin, How to compare noisy patches? patch similarity beyond Gaussian noise, *Int. J. Comput. Vis.* 99 (2012) 86–102, doi:[10.1007/s11263-012-0519-6](https://doi.org/10.1007/s11263-012-0519-6).
- [46] S. Vitalé, D. Cozzolino, G. Scarpa, L. Verdoliva, G. Poggi, Guided patchwise non-local SAR despeckling, *IEEE Trans. Geosci. Remote Sens.* 57 (2019) 6484–6498, doi:[10.1109/TGRS.2019.2906412](https://doi.org/10.1109/TGRS.2019.2906412).

- [47] M. Volpi, G. Camps-Valls, D. Tuia, Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis, *ISPRS J. Photogramm. Remote Sens.* 107 (2015) 50–63, doi:[10.1016/j.isprsjprs.2015.02.005](https://doi.org/10.1016/j.isprsjprs.2015.02.005).
- [48] R. Touati, M. Mignotte, M. Dahmane, Multimodal change detection in remote sensing images using an unsupervised pixel pairwise based Markov random field model, *IEEE Trans. Image Process.* (2019) 1, doi:[10.1109/TIP.2019.2933747](https://doi.org/10.1109/TIP.2019.2933747).
- [49] H. Feng, B. Hou, M. Gong, SAR Image despeckling based on local homogeneous-region segmentation by using pixel-relativity measurement, *IEEE Trans. Geosci. Remote Sens.* 49 (2011) 2724–2737, doi:[10.1109/TGRS.2011.2107915](https://doi.org/10.1109/TGRS.2011.2107915).
- [50] H. Xie, L.E. Pierce, F.T. Ulaby, Statistical properties of logarithmically transformed speckle, *IEEE Trans. Geosci. Remote Sens.* 40 (2002) 721–727, doi:[10.1109/TGRS.2002.1000333](https://doi.org/10.1109/TGRS.2002.1000333).



**Yuli Sun** received the B.S. degree from Xiangtan University, Xiangtan, China, in 2009, and the M.S. degree from University of Science and Technology of China, Hefei, China, in 2014. Currently, He is working toward the Ph.D. degree in information and communication engineering in the College of Electronic Science, National University of Defense Technology, since 2019. His research interests cover machine learning and remote sensing image processing.



**Lin Lei** received the Ph.D. degree in Information and Communication Engineering from National University of Defense Technology, Changsha, China, in 2008. She is currently an Associate Professor with the school of Electronic Science, National University of Defense Technology. Her research interests include computer vision, remote sensing image interpretation and data fusion.



**Xiao Li** received the B.S. degree in the electrical engineering and automation from the University of Jinan, Jinan, China, in 2015, and the M.S. degrees in control science and engineering at Xiangtan University, Xiangtan, China, in 2018. He is currently pursuing the Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China. His research interests include image processing and pattern recognition, representation and dictionary learning, computational pathology applications.



**Hao Sun** received the M.S. degree in information and communication engineering and the Ph.D. degree in electronic science and technology from National University of Defense Technology, Changsha, China, in 2009 and 2011, respectively. He is currently an Associate Professor in the College of Electrical Science and Engineering, National University of Defense Technology. His research interests include multimodal remote sensing image processing and machine learning.



**Gangyao Kuang** received B.S. and M.S. degrees in geophysics from the Central South University of Technology, Changsha, China, in 1988 and 1991, and the Ph.D. degree in communication and information from the National University of Defense Technology, Changsha, China, in 1995. He is currently a Professor at School of Electronic Science, National University of Defense Technology. His research interests include remote sensing, SAR image processing, change detection, SAR ground moving target indication, and classification with polarimetric SAR images.