

# Iterative Robust Graph for Unsupervised Change Detection of Heterogeneous Remote Sensing Images

Yuli Sun<sup>id</sup>, Lin Lei<sup>id</sup>, Dongdong Guan<sup>id</sup>, and Gangyao Kuang, *Senior Member, IEEE*

**Abstract**—This work presents a robust graph mapping approach for the unsupervised heterogeneous change detection problem in remote sensing imagery. To address the challenge that heterogeneous images cannot be directly compared due to different imaging mechanisms, we take advantage of the fact that the heterogeneous images share the same structure information for the same ground object, which is imaging modality-invariant. The proposed method first constructs a robust  $K$ -nearest neighbor graph to represent the structure of each image, and then compares the graphs within the same image domain by means of graph mapping to calculate the forward and backward difference images, which can avoid the confusion of heterogeneous data. Finally, it detects the changes through a Markovian co-segmentation model that can fuse the forward and backward difference images in the segmentation process, which can be solved by the co-graph cut. Once the changed areas are detected by the Markovian co-segmentation, they will be propagated back into the graph construction process to reduce the influence of changed neighbors. This iterative framework makes the graph more robust and thus improves the final detection performance. Experimental results on different data sets confirm the effectiveness of the proposed method. Source code of the proposed method is made available at <https://github.com/yulisun/IRG-McS>.

**Index Terms**—Unsupervised change detection, heterogeneous data, self-similarity,  $K$ -nearest neighbor graph, co-segmentation.

## I. INTRODUCTION

### A. Background

THE change detection (CD) of remote sensing (RS) images is a technique to identify the changes of an object or phenomenon on the earth's surface by analyzing the difference between images acquired at different times [1]. It is one of the important topics in earth observation, which has been found a wide range of applications in urban studies [2], environmental monitoring [3], and nature disaster assessment [4].

Currently, most of the CD techniques are based on homogeneous RS images, i.e., images are from the same sensor,

such as CD of SAR images [5], optical images [6], [7], and hyperspectral images [8]. In recent years, with the rapid development of RS-related technologies, more and more image data representing the real information of the earth's surface can be obtained from different sensors at the same time, which puts forward new requirements on CD methods, i.e., they should not only be applied to the homogeneous data, but should also consider the broader issue of CD on heterogeneous data obtained from different sensors [9]. More importantly, heterogeneous CD has a very urgent need in rescue and assessment of emergency disasters. In such sudden scenarios (earthquakes, floods, etc.), it is important to exploit the first available acquisition after the event, independently of its modality. This post-event image is often likely to be acquired by a SAR sensor, due to weather, cloud, smoke, and daylight constraints. On the other hand, the pre-event image should be selected from the archived data of remote sensing platforms as up to date as possible in order to reduce the time delay between the two acquisitions and to better focus on the ground changes caused by the event.

Despite its undeniable importance, relatively little research has been devoted to heterogeneous CD compared to homogeneous CD. Due to the different imaging mechanisms of different sensors, heterogeneous images provide different descriptions of the same object and exhibit quite different characteristics. Unlike in the homogeneous CD, the heterogeneous images cannot be directly compared with each other. Therefore, it is not feasible to directly use arithmetical operators such as the image differencing (usually in homogeneous optical CD) or image ratio/log-ratio (usually in homogeneous SAR CD) to calculate the difference image (DI) of multi-temporal heterogeneous images.

Generally, the existing heterogeneous CD methods can be classified into supervised and unsupervised based on whether or not a labeled training sample is required. In the supervised method, the labeled samples are used to establish the local joint statistical model between heterogeneous data [10], [11], or to train the image regression model [12], or to train the classifier [13]. However, labeling samples is labor-intensive and accurate labeling of heterogeneous data requires extensive expert knowledge as a guide and sometimes even a field trip, so unsupervised methods are more popular in practice.

According to the basic analysis unit used in image comparison, the heterogeneous CD method can also be divided into pixel-based and object-based. In the former, the individual pixel is treated as the basic unit to calculate the pixel-wise

Manuscript received December 28, 2020; revised May 5, 2021 and June 7, 2021; accepted June 26, 2021. Date of publication July 7, 2021; date of current version July 13, 2021. This work was supported by the National Natural Science Foundation of China under Grant 61701508. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Xiaolin Hu. (*Corresponding author: Lin Lei.*)

Yuli Sun, Lin Lei, and Gangyao Kuang are with the College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China (e-mail: sunyuli@mail.ustc.edu.cn; alaleilin@163.com; kuangyeats@hotmail.com).

Dongdong Guan is with the High-Tech Institute of Xi'an, Xi'an 710025, China (e-mail: guandd@hotmail.com).

Digital Object Identifier 10.1109/TIP.2021.3093766

difference between heterogeneous images and is then divided into changed or unchanged classes. The pixel-based methods are more suitable for low- and medium-resolution images [12]–[15], and may bring the salt-and-pepper noise on CD results of very-high-resolution (VHR) images due to the neglect of spatial dependence between pixels [16]. To address this challenge, the object-based methods are proposed to fuse spectral, texture, and geometry information to obtain meaningful change information and improve the detection performance [17], [18].

Since the heterogeneous images reflect different physical quantities of the object and show quite different statistical behaviors [5], [9], [19], the goal of the heterogeneous CD method is to transform the “incomparable” heterogeneous images to a space where they are “comparable”. According to the transformed space, the existing heterogeneous CD method can also be roughly divided into three categories [20]: firstly, the classification-based methods transform the multitemporal images to a common category space (such as water, buildings, roads, etc.) by classifying the images separately, and then compare the image classification results in the category space to detect the changes, such as the post-classification comparison method (PCC) [21], the multitemporal segmentation and compound classification method (MS-CC) [18], [22]. Although classification-based method is intuitive, feasible and easy to implement, it has the following drawbacks: the CD accuracy is affected by the image classification performance, especially the accurate classification is difficult for SAR image due to the inherent speckle; the CD granularity is determined by the fineness of image classification, i.e., it can only detect changes between classes, for example, when the category is small (e.g., only water, farmland, building, tree), it is difficult to detect changes within the same category and outside the all categories (e.g., vehicles, roads); in addition, such post-classification comparisons often suffers from the error accumulation or error propagation [23].

Secondly, the traditional similarity-based methods transform the multitemporal images to a constructed feature space, in which the constructed imaging-modality-invariant-features are used to distinguish the changed and unchanged regions, such as sorted histogram distance (SHD) [24], pixel pair method (PP) [25], homogeneous pixel transformation method (HPT) [12], affinity matrix-based image regression (AM-IR) [26]. For this kind of methods, it is very important to mine the modal invariant features between heterogeneous images and make full use of these features to design the CD operator. As a result, such methods often face the challenge that when the detection scene is complex (e.g., diversity of ground objects, the difference of imaging conditions) or when the noise level is very high (especially the speckle noise in SAR images), these manually constructed connections may not exist or may no longer fully characterize the relationship between heterogeneous images, which may result in a dramatic degradation of CD performance.

Thirdly, the deep learning-based methods transform the multitemporal images to a common latent feature space, which is learned from the unchanged samples of heterogeneous images by using the deep neural networks (DNN), such as the

symmetrical convolutional coupling network (SCCN) [14], approximately symmetrical DNN [27], logarithmic transformation feature learning network (LT-FL) [28]. Although these methods perform relatively well in terms of detection accuracy, they usually either require the construction of large training set under the supervision mode, or need a pre-constructed pseudo-training set/change prior to guide the training process, or involve a complex and time-consuming iterative coarse-to-fine filtering process to construct the pseudo-training set to learn the latent feature space.

## B. Motivation

In this paper, we propose a structure consistency based heterogeneous CD method, which belongs to the above-mentioned unsupervised, object-based and traditional similarity-based method. The proposed method takes advantage of the self-similarity (or self-redundancy) property of images, which has been widely used in the image denoising with the so-called “nonlocal-based” methods [29], [30]. At the same time, it has recently been exploited in the heterogeneous CD [20], [31], [32].

In the fractal projection and Markovian segmentation-based method (FPMS) [32], the self-similarity property is used to map the pre-event image to the domain of the post-event image by fractal projection, and then pixel-wise DI is obtained by comparing the projected image and the post-event image. In [20], the self-similarity is further used as image self-expression, which learns a patch similarity graph matrix (PSGM) of pre-event (post-event) image, then obtain the regression image by multiplying the post-event (pre-event) image with this learned PSGM, and measures the change level by comparing the regression image and the post-event (pre-event) image. In this paper, the self-similarity is used to construct graphs representing the structures for each image and establish the relationships between heterogeneous images, which is similar to the nonlocal patch similarity graph-based method (NPSG) in [31]. However, the proposed method is different from the previous self-similarity based methods in the following aspects.

- 1) The proposed method is object-based, while the others are pixel-based. FPMS [32], PSGM [20] and NPSG [31] all calculate the pixel-wise difference by using the square patch centered on this pixel. Although the context information is considered in the CD process, the square image patch may include pixels from different object, and it cannot maintain the shape and structure of the object. The proposed method uses superpixel as the basic unit, which has two main advantages: one is that it can maintain the structure and edge of the object as the interior of each superpixel is homogeneous (representing the same kind of object); the other is that it can reduce the computational complexity, especially for the large-scale high-resolution images.

- 2) Different from the other methods which calculate the DI and binary change map (CM) relatively independently, the proposed method combines the DI generation and the CM calculation by using an iterative framework. This can bring two benefits: first, it reduces the influence of changed regions during the graph construction to obtain a more robust graph,

making the DI better distinguish the changed parts; second, it can reduce the false detection and missed detection in the CM with the generated high-quality DI.

3) Different from FPMS and PSGM, the proposed method neither reconstructs the image nor translates the image. It only focuses on the structure difference between images similar as NPSG, which is represented by the superpixel-based  $K$ -nearest neighbor (KNN) graph. However, unlike NPSG which select nearest neighbors (NNs) for each vertex with fixed  $K$  and without considering the properties of neighbors (changed or unchanged), the graph constructed by the proposed method is more robust by adaptively selecting NNs for each superpixel with different  $K$  and unchanged neighbors, which eliminates the changed neighbors that cause confusion.

4) Different from the other methods which compute the binary CM directly on the fused DI, the proposed method compute the CM by using the Markov random field (MRF) with a co-graph cut segmentation model, which combines the DI fusion and segmentation in the same model. Then, the forward and backward DIs can be fully fused to improve the CM accuracy.

### C. Contribution

In particular, our contribution refers to an unsupervised, object-based, iterative robust graph and Markovian co-segmentation method (IRG-McS) for heterogeneous CD. The method exploits the inherent structure consistency between heterogeneous images, and constructs robust graph to represent the structure of images, and then compares the graphs to measure the change level by graph mapping. The main contributions of the proposed method are as follows.

- An object-based heterogeneous CD method based on structure consistency is proposed, which measures the consistency between image structures by graph mapping.
- A robust adaptive KNN graph of each image is constructed by adaptively selecting unchanged NNs with appropriate  $K$  for each superpixel though an iterative framework combining the DI generation and CM calculation processes.
- A superpixel-based MRF co-segmentation model is designed to fuse the forward and backward DIs in the segmentation process to improve the CD accuracy, which is solved by the co-graph cut.

The rest of this paper is organized as follows: Section II describes the details of the proposed IRG-McS. Section III presents the experimental results and comparisons with some existing state-of-the-art methods. Finally, we conclude this paper in Section IV.

## II. ITERATIVE ROBUST GRAPH AND MARKOVIAN CO-SEGMENTATION (IRG-MCS)

We consider two co-registered heterogeneous RS images acquired at the same geographical area by different sensors at different times (before and after an event), which are denoted as  $\mathbf{X} \in \mathbb{R}^{M \times N \times C_X}$  in domain  $\mathcal{X}$  and  $\mathbf{Y} \in \mathbb{R}^{M \times N \times C_Y}$  in domain  $\mathcal{Y}$ , and their pixels are denoted as  $x(m, n, c)$  and  $y(m, n, c)$ , respectively. Here,  $M$ ,  $N$ ,  $C_X$  ( $C_Y$ ) represent the height, width and channel number of two images, respectively.

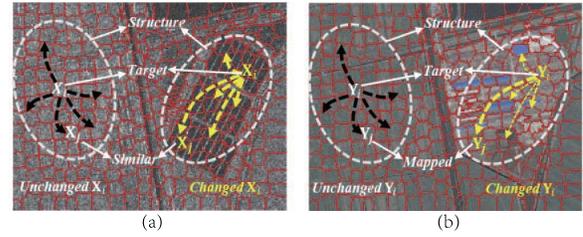


Fig. 1. Illustration of the structure consistency in heterogeneous images: (a) SAR image; (b) Optical image. The similarity between image parts ( $X_i$  and  $X_j$ ,  $Y_i$  and  $Y_j$ ) is reflected by the thickness of connecting lines. The structure of the unchanged target part  $X_i$  in the SAR image can be preserved by the part  $Y_i$  in the optical image. However, for the changed target part  $X_j$ , the structure in the SAR image is no longer conformed by the optical image.

As illustrated in Fig. 1, in the heterogeneous CD, directly comparing the pixel values of  $x(m, n, c)$  and  $y(m, n, c)$ , such as  $|x(m, n, c) - y(m, n, c)|$  and  $\left| \log \frac{x(m, n, c)}{y(m, n, c)} \right|$  in the homogeneous CD, is not feasible. However, with the self-similarity property, each small part of the image ( $X_i$  in the SAR image) can always find some similar parts ( $X_j$  in the SAR image) within the same image. Then, the structure of  $X_i$ , which is represented by this similarity relationships between  $X_i$  and its similar parts  $X_j$ , can be well preserved by the unchanged part  $Y_i$  in the optical image, showing that  $Y_i$  and  $Y_j$  are also very similar, as illustrated by the unchanged part in Fig. 1. On the contrary, if the area represented by image  $X_i$  has changed in the event, the structure of  $X_i$  is no longer preserved by  $Y_i$ , showing that  $Y_i$  and  $Y_j$  are very different, as illustrated by the changed part in Fig. 1. Therefore, we can find that the structure consistency is quite imaging modality-invariant, and the change level can be measured by how dissimilar between the structure of pre-event image and post-event image. Then, three main problems need to be solved: how to represent the structure, how to calculate the structural difference, and how to detect the changes.

The proposed heterogeneous CD method consists of four steps: 1) superpixel segmentation and feature extraction; 2) structure representation by constructing KNN graph; 3) structural difference calculation and DI generation; 4) binary CM calculation by MRF model. The framework is illustrated in Fig. 2. In the general CD process, the calculation of binary CM is after the construction of DI, which is a one-way process. However, we find that the quality of generated DI is affected by the CM in the proposed method, and in light of this, we use an iterative framework to combine the DI construction process and binary CM calculation process. This framework is also applicable to some other unsupervised algorithms, such as those that need to pick pseudo-training sets [26], or those where the DI is affected by changed regions [20], [31], [32].

### A. Superpixel Segmentation and Feature Extraction

Rather than focusing on the individual pixel, image block is chosen as the basic analysis unit in the proposed method, which is obtained by superpixel segmentation. In this paper, the simple liner iterative clustering (SLIC) method [33] is

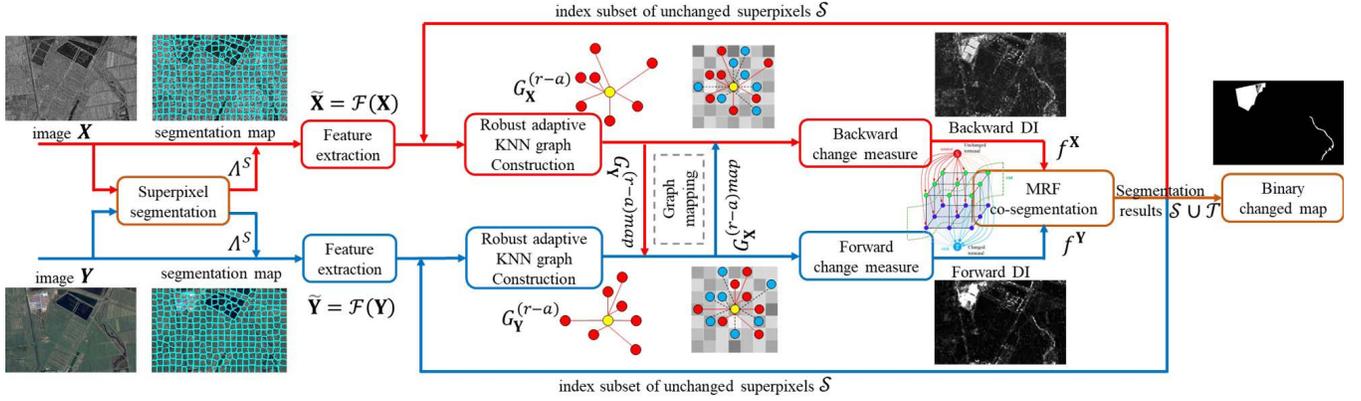


Fig. 2. Framework of the proposed heterogeneous change detection method.

applied on each image to generate the superpixels for its superior in both efficiency and boundary preservation. The original SLIC method is designed for optical image with R, G, B bands, which transforms the RGB space to the CIELAB space and then computes the color distance  $d$  as

$$d = \sqrt{(l_i - l_j)^2 + (a_i - a_j)^2 + (b_i - b_j)^2} \quad (1)$$

where  $l$ ,  $a$ , and  $b$  represent the pixel values in CIELAB color space. However, the images involved in heterogeneous CD are not only RGB optical images, but also multispectral images and SAR images. For the multispectral images  $\mathbf{X}$  with  $C_X > 3$ , the principle component analysis (PCA) method is firstly used to reduce its dimension to obtain the first three principle components  $\mathbf{X}' \in \mathbb{R}^{M \times N \times 3}$ , and then SLIC is used to complete the superpixel segmentation by directly using (1) without the CIELAB space transformation. For the SAR image, it is usually assumed to be contaminated by the well-accepted multiplicative speckle noise modeled by a Gamma distribution. It is obviously inappropriate to use (1) to calculate the distance for SAR image. Inspired by the generalized likelihood ratio (GLR) based similarity criterion of patches in SAR image proposed in [34], the following distance  $d_{GLR}$  is used to produce the superpixels instead of (1)

$$d_{GLR} = \log \left( \frac{x_i + x_j}{2\sqrt{x_i x_j}} \right) \quad (2)$$

where  $x_i$  and  $x_j$  are intensity values of two pixels.

We first segment  $\mathbf{X}$  and  $\mathbf{Y}$  independently into superpixels, and denote the corresponding segmentation maps of  $\Lambda^{\mathbf{X}}$  and  $\Lambda^{\mathbf{Y}}$  as the region sets of each superpixel, i.e.,  $\Lambda^{\mathbf{X}} = \{\Lambda_i^{\mathbf{X}} | i = 1, \dots, N_1\}$  and  $\Lambda^{\mathbf{Y}} = \{\Lambda_j^{\mathbf{Y}} | j = 1, \dots, N_2\}$  respectively, where  $\Lambda_i^{\mathbf{X}}$  is the region of the  $i$ -th superpixel of  $\mathbf{X}$  and  $\Lambda_j^{\mathbf{Y}}$  is the region of the  $j$ -th superpixel of  $\mathbf{Y}$ . Denote  $\Lambda_{(i,j)}^S = \Lambda_i^{\mathbf{X}} \cap \Lambda_j^{\mathbf{Y}}$ , we have a new individual region  $\Lambda_{(i,j)}^S$  in the segmentation map  $\Lambda^S$  as  $\Lambda_{(i,j)}^S = \{(m, n) | (m, n) \in \Lambda_i^{\mathbf{X}}, (m, n) \in \Lambda_j^{\mathbf{Y}}\}$ . By eliminating the empty sets in  $\Lambda^S$  and merging the small regions (i.e., regions that are smaller than the  $N_S$ -th smallest region in  $\Lambda^S$ ) into the nearest larger regions, we can get the co-segmentation

map  $\Lambda^S$  with the new subscript index as

$$\Lambda^S = \{\Lambda_i^S | i = 1, \dots, N_S\}; \quad \Lambda_i^S \cap \Lambda_j^S = \emptyset, \text{ if } i \neq j \quad (3)$$

$$\bigcup_{i=1}^{N_S} \Lambda_i^S = \{(m, n) | m = 1, \dots, M; n = 1, \dots, N\};$$

The  $i$ -th superpixels of  $\mathbf{X}$  and  $\mathbf{Y}$  are defined as  $\mathbf{X}_i = \{x(m, n, c) | (m, n) \in \Lambda_i^S, c = 1, \dots, C_X\}$  and  $\mathbf{Y}_i = \{y(m, n, c) | (m, n) \in \Lambda_i^S, c = 1, \dots, C_Y\}$ , respectively.

Therefore, each superpixel in the co-segmentation map  $\Lambda^S$  exhibits homogeneous structural property in both pre-event image  $\mathbf{X}$  and post-event image  $\mathbf{Y}$ . Compared with the square patch based methods (such as FPMS, PSGM and NPSG), the proposed superpixel based method can group more adjacent pixels, adhere to edges, and preserve the object structures of both images. Hence, it can explore the contextual information more accurately and improve the CD performance in accuracy and efficiency.

Once the co-segmentation map  $\Lambda^S$  is obtained, different kinds of feature information can be extracted from the superpixel, such as the spectral (intensity), textural, and spatial information. Denote the feature extraction operator as  $\mathcal{F}$ , we have the feature vectors of  $\mathbf{X}_i$  and  $\mathbf{Y}_i$  as  $\tilde{\mathbf{X}}_i = \mathcal{F}(\mathbf{X}_i)$  and  $\tilde{\mathbf{Y}}_i = \mathcal{F}(\mathbf{Y}_i)$ , respectively. In the proposed method, the mean, median, and variance values of each band are taken as the superpixel's features. Then we can obtain the feature matrices of  $\mathbf{X}$  and  $\mathbf{Y}$  as  $\tilde{\mathbf{X}} \in \mathbb{R}^{3C_X \times N_S}$  and  $\tilde{\mathbf{Y}} \in \mathbb{R}^{3C_Y \times N_S}$  respectively, each column of which represents a feature vector.

### B. Robust Adaptive KNN Graph Construction and Change Level Measurement

1) *KNN Graph Construction*: As the graph model can efficiently capture crucial information of an image, we propose to construct a weighted graph to represent the structure of each image.

*Definition 1 (KNN Graph)*: Given a set of data points  $\mathbf{z} = \{z_1, z_2, \dots, z_n\}$  with  $z_i \in \mathbb{R}^d$  in domain  $\mathcal{Z}$ ,  $G = \{V, E, w\}$  is a weighted directed graph, where  $V = \mathbf{z}$ , and  $(z_i, z_j) \in E$  if and only if  $dist_{i,j}^{\mathcal{Z}}$  is among the  $k$  smallest elements of the set

$\left\{ \text{dist}_{i,l}^Z | l = 1, \dots, n; l \neq i \right\}$ , and  $w(i, j) = \exp(-\eta \text{dist}_{i,j}^Z)$  for  $(z_i, z_j) \in E$ , where  $\text{dist}_{i,j}^Z$  represents the distance between  $z_i$  and  $z_j$  with defined distance metric.

Given the pre-event image  $\mathbf{X}$  with the co-segmentation map  $\Lambda^S$ , we set each superpixel as a vertex and construct the KNN graph  $G_{\mathbf{X}} = \{V_{\mathbf{X}}, E_{\mathbf{X}}, w_{\mathbf{X}}\}$  as

$$\begin{aligned} V_{\mathbf{X}} &= \{\mathbf{X}_i | i = 1, 2, \dots, N_S\} \\ E_{\mathbf{X}} &= \left\{ (\mathbf{X}_i, \mathbf{X}_j) | i = 1, 2, \dots, N_S; j \in \mathcal{N}_{\mathbf{X}_i}^K \right\} \\ w_{\mathbf{X}}(i, j) &= \exp(-\eta \text{dist}_{i,j}^{\mathcal{X}}), \forall (\mathbf{X}_i, \mathbf{X}_j) \in E_{\mathbf{X}} \end{aligned} \quad (4)$$

where  $\eta > 0$  is a bandwidth parameter of exponential kernel,  $\text{dist}_{i,j}^{\mathcal{X}} = \|\tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_j\|_2^2$  represents the distance between superpixels  $\mathbf{X}_i$  and  $\mathbf{X}_j$  by using the feature vectors,  $\text{dist}_i^{\mathcal{X}} \in \mathbb{R}^{N_S}$  denotes the distance vector composed of  $\text{dist}_{i,i}^{\mathcal{X}}$ ,  $\mathcal{N}_{\mathbf{X}_i}^K$  represents the position set of the  $K$  NNs of  $\mathbf{X}_i$  by sorting all elements except  $\text{dist}_{i,i}^{\mathcal{X}}$  in the distance vector  $\text{dist}_i^{\mathcal{X}}$  and taking out the  $K$  smallest elements. For the post-event image  $\mathbf{Y}$  with the co-segmentation map  $\Lambda^S$ , we can construct KNN graph  $G_{\mathbf{Y}} = \{V_{\mathbf{Y}}, E_{\mathbf{Y}}, w_{\mathbf{Y}}\}$  in domain  $\mathcal{Y}$  similar as  $G_{\mathbf{X}}$  in (4).

2) *Change Level Measurement*: In the constructed directed graph  $G_{\mathbf{X}}$ , each superpixel of  $\mathbf{X}$  is treated as a vertex, and each vertex is connected with its neighbors by  $K$  directed edges with associate weights  $w_{\mathbf{X}}$  representing the similarities between the vertex and its NNs. In this way, the structure information of the image  $\mathbf{X}$  can be characterized by the constructed KNN graph. Therefore, we can compare the difference between graphs  $G_{\mathbf{X}}$  and  $G_{\mathbf{Y}}$  to measure the changes between  $\mathbf{X}$  and  $\mathbf{Y}$ .

Because the heterogeneous images  $\mathbf{X}$  and  $\mathbf{Y}$  represent different physical quantities, the graphs  $G_{\mathbf{X}}$  and  $G_{\mathbf{Y}}$  are constructed in different domains of  $\mathcal{X}$  and  $\mathcal{Y}$  respectively, directly comparing  $G_{\mathbf{X}}$  and  $G_{\mathbf{Y}}$  (such as  $|w_{\mathbf{X}} - w_{\mathbf{Y}}|$ ) will cause the leakage of heterogeneous data. In order to make  $G_{\mathbf{X}}$  and  $G_{\mathbf{Y}}$  comparable, we first map the  $G_{\mathbf{X}}$  to the other image domain  $\mathcal{Y}$  to obtain the mapped graph  $G_{\mathbf{Y}}^{\text{map}} = \{V_{\mathbf{Y}}^{\text{map}}, E_{\mathbf{Y}}^{\text{map}}, w_{\mathbf{Y}}^{\text{map}}\}$  as

$$\begin{aligned} V_{\mathbf{Y}}^{\text{map}} &= \{\mathbf{Y}_i | i = 1, 2, \dots, N_S\} \\ E_{\mathbf{Y}}^{\text{map}} &= \left\{ (\mathbf{Y}_i, \mathbf{Y}_j) | i = 1, 2, \dots, N_S, j \in \mathcal{N}_{\mathbf{X}_i}^K \right\} \\ w_{\mathbf{Y}}(i, j) &= \exp(-\eta \text{dist}_{i,j}^{\mathcal{Y}}), \forall (\mathbf{Y}_i, \mathbf{Y}_j) \in E_{\mathbf{Y}}^{\text{map}} \end{aligned} \quad (5)$$

From this mapping process, we can find that the  $G_{\mathbf{Y}}^{\text{map}}$  is constructed in domain  $\mathcal{Y}$  by using the edges information of  $G_{\mathbf{X}}$  but computing the corresponding weights by using  $\text{dist}_{i,j}^{\mathcal{Y}}$ , which is also a  $K$  connected graph. In this way, since  $G_{\mathbf{Y}}^{\text{map}}$  and  $G_{\mathbf{Y}}$  are constructed in the same domain  $\mathcal{Y}$ , they are comparable. Then, the forward change level  $f_i^{\mathbf{Y}}$  can be measured by the structure difference between  $G_{\mathbf{Y}}^{\text{map}}$  and  $G_{\mathbf{Y}}$  with the distance criterion or similarity criterion as

$$\begin{aligned} f_i^{\mathbf{Y}} &= \frac{1}{K} \left( \sum_{j' \in \mathcal{N}_{\mathbf{X}_i}^K} \text{dist}_{i,j'}^{\mathcal{Y}} - \sum_{j \in \mathcal{N}_{\mathbf{Y}_i}^K} \text{dist}_{i,j}^{\mathcal{Y}} \right) \\ f_i^{\mathbf{Y}} &= \frac{1}{K} \left( \sum_{j \in \mathcal{N}_{\mathbf{Y}_i}^K} w_{\mathbf{Y}}(i, j) - \sum_{j' \in \mathcal{N}_{\mathbf{X}_i}^K} w_{\mathbf{Y}}(i, j') \right) \end{aligned} \quad (6)$$

Intuitively, the structure difference is calculated by how different the two KNN position sets of  $\mathcal{N}_{\mathbf{X}_i}^K$  and  $\mathcal{N}_{\mathbf{Y}_i}^K$  are in the image  $\mathbf{Y}$ . If the area represented by superpixel  $\mathbf{X}_i$  does not change in the event, the superpixel  $\mathbf{Y}_{j'}$ ,  $j' \in \mathcal{N}_{\mathbf{X}_i}^K$  will be similar as  $\mathbf{Y}_i$  with a high probability, then the change level  $f_i^{\mathbf{Y}}$  of (6) will be very small; on the contrary, if the area represented by superpixel  $\mathbf{X}_i$  changes in the event, the superpixel  $\mathbf{Y}_{j'}$ ,  $j' \in \mathcal{N}_{\mathbf{X}_i}^K$  is no longer similar to  $\mathbf{Y}_i$  with a high probability, and then it leads to a large  $f_i^{\mathbf{Y}}$ . In this way, the changed and unchanged superpixels can be distinguished.

3) *Adaptive KNN*: In the KNN construction and structure difference calculation (6), we can find that the number “ $K$ ” of the NNs has an important role in the performance of change level measurement. Obviously, a very small  $K$  is not appropriate, which will make the measurement (6) not robust enough. For example, for the unchanged superpixel  $\mathbf{Y}_i$ , when a mapped neighbor  $\mathbf{Y}_{j'}$ ,  $j' \in \mathcal{N}_{\mathbf{X}_i}^K$  in (5) is polluted by noise,  $\text{dist}_{i,j'}^{\mathcal{Y}}$  will be a large value and it brings errors in (6). In this case, a larger  $K$  will reduce the influence of this polluted  $\mathbf{Y}_{j'}$  because (6) takes the mean value of all  $K$  neighbors. However, a particularly large  $K$  is also not appropriate, which will make the measurement (6) less discriminative. For example, for the changed superpixel  $\mathbf{Y}_i$ , when the value of  $K$  exceeds the actual number of its real similar neighbors, some large  $\text{dist}_{i,j}^{\mathcal{Y}}$  will be introduced into (6). In the extreme case of  $K = N_S$ , the change level will always equal to 0. Therefore, the choice of  $K$  is a challenge and it is not feasible to set the same  $K$  for all vertexes. Instead, we need to set a suitable  $K$  for each vertex according to the principle: “each vertex is connected to as many truly similar vertices as possible”. However, it is difficult to achieve this goal completely. Here, we propose a strategy to adaptively select  $K$  for each vertex.

*Step 1*: Choose a large  $K_{\max}$  and a small  $K_{\min}$ , and construct the KNN graphs  $G_{\mathbf{X}}$  and  $G_{\mathbf{Y}}$  with  $K = K_{\max}$ ;

*Step 2*: Calculate the in-degree  $di(\mathbf{X}_i)$  and  $di(\mathbf{Y}_i)$  of each vertex in  $G_{\mathbf{X}}$  and  $G_{\mathbf{Y}}$ , respectively;

*Step 3*: Calculate  $K_i^{\mathbf{X}}$  and  $K_i^{\mathbf{Y}}$  as  $K_i^{\mathbf{X}} = \min\{K_{\max}, \max\{di(\mathbf{X}_i), K_{\min}\}\}$  and  $K_i^{\mathbf{Y}} = \min\{K_{\max}, \max\{di(\mathbf{Y}_i), K_{\min}\}\}$ , respectively;

*Step 4*: Set  $K_i = \min\{K_i^{\mathbf{X}}, K_i^{\mathbf{Y}}\}$  for each vertex  $\mathbf{X}_i$  and  $\mathbf{Y}_i$ , and construct the adaptive KNN graph  $G_{\mathbf{X}}^{(a)}$  by replacing the  $\mathcal{N}_{\mathbf{X}_i}^K$  in (4) with  $\mathcal{N}_{\mathbf{X}_i}^{K_i}$ , and construct  $G_{\mathbf{Y}}^{(a)}$  with a similar operations.

Here, the  $K_{\max}$  and  $K_{\min}$  should satisfy the following desirable properties: the large  $K_{\max}$  should be small enough to represent the real structure of objects with the most superpixels, while the small  $K_{\min}$  should be large enough to preserve the structure of objects with the least superpixels;  $K_{\max}$  and  $K_{\min}$  should be data-dependent. In the proposed method, we set  $K_{\max} = \lceil \sqrt{N_S} \rceil$  and  $K_{\min} = \lceil \sqrt{N_S/10} \rceil$  to satisfy these conditions, where  $\lceil \cdot \rceil$  represents the rounding up operation. The in-degree  $di(\mathbf{X}_i)$  in  $G_{\mathbf{X}}$  is the number of times  $\mathbf{X}_i$  occurs among the  $K$  NNs of all other vertices in  $G_{\mathbf{X}}$ , which can be used to measure the “popularity” of  $\mathbf{X}_i$ . A larger  $di(\mathbf{X}_i)$  means that  $\mathbf{X}_i$  is “popular” in the KNN graph, while a very small  $di(\mathbf{X}_i)$  means that  $\mathbf{X}_i$  is “alone” in the graph. With this strategy, we can choose an appropriate  $K$  for each

vertex, whether it has many truly similar vertices (belonging to high density) or few truly similar vertices (belonging to low density).

4) *Robust KNN Graph*: In the KNN graph construction, we find the NNs of each  $\mathbf{X}_i$  in the whole  $\mathbf{X}$  image without considering the stability of neighbors, that is, whether the neighbors will change in the event. Next, we will show that the unstable (changed) neighbors degrade the performance of structure difference measurement (6).

For the unchanged superpixel  $\mathbf{X}_i$ , when its  $K$  NNs contain the changed superpixel  $\mathbf{X}_{j'}$ ,  $j' \in \mathcal{N}_{\mathbf{X}_i}^K$ , although  $\mathbf{X}_i$  and  $\mathbf{X}_{j'}$  are very similar in  $G_{\mathbf{X}}$ ,  $\mathbf{Y}_i$  and  $\mathbf{Y}_{j'}$  in the mapped  $G_{\mathbf{Y}}^{map}$  are quite different (resulting a large  $dist_{i,j'}^{\mathcal{Y}}$ ). Then, it increases the value of unchanged  $f_i^{\mathbf{Y}}$ . On the contrary, for the changed superpixel  $\mathbf{X}_i$ , when its  $K$  NNs contain the changed superpixel  $\mathbf{X}_{j'}$ ,  $j' \in \mathcal{N}_{\mathbf{X}_i}^K$ , if  $\mathbf{X}_i$  and  $\mathbf{X}_{j'}$  changed to the same category,  $\mathbf{Y}_i$  and  $\mathbf{Y}_{j'}$  in the mapped  $G_{\mathbf{Y}}^{map}$  are also quite similar (resulting a small  $dist_{i,j'}^{\mathcal{Y}}$ ). Then, it reduces the value of changed  $f_i^{\mathbf{Y}}$ . From the above analysis, we can find that whether for the changed or unchanged superpixel  $\mathbf{X}_i$ , its unstable  $K$  NNs will make the structure difference measurement less discriminative (as illustrated by Fig. 6 in the subsection III-C).

In order to reduce the impact of unstable neighbors, we need to eliminate them in the KNN graph construction. However, we can't know which superpixels are changed in advance, so we use an iterative framework to achieve this elimination. That is, we use the binary CM of previous round (in Section II-C) to guide the adaptive KNN graph construction of next round, to enhance the quality of the DI and improve the accuracy of the final CM, which is an iterative coarse-to-fine process.

Suppose the previous CD results with MRF segmentation in Section II-C is the index subset  $\mathcal{S}$  of unchanged superpixels and the index subset  $\mathcal{T}$  of changed superpixels, we can construct the robust KNN graph  $G_{\mathbf{X}}^{(r)} = \{V_{\mathbf{X}}^{(r)}, E_{\mathbf{X}}^{(r)}, w_{\mathbf{X}}\}$  by replacing the  $E_{\mathbf{X}}$  in (4) with  $E_{\mathbf{X}}^{(r)} = \{(\mathbf{X}_i, \mathbf{X}_j) | i = 1, 2, \dots, N_S, j \in \mathcal{N}_{\mathbf{X}_i}^K\}$ , where  $\mathcal{N}_{\mathbf{X}_i}^K$  represents the position set of the  $K$  NNs of the  $\mathbf{X}_i$  in the unchanged subset  $\mathcal{S}$ . This means that it finds the  $K$  NNs of each superpixel in the set of unchanged superpixels rather than the whole  $\mathbf{X}$ .

With the adaptive  $K$  selection strategy and unchanged index subset  $\mathcal{S}$ , we can construct the robust adaptive KNN graph  $G_{\mathbf{X}}^{(r-a)} = \{V_{\mathbf{X}}^{(r-a)}, E_{\mathbf{X}}^{(r-a)}, w_{\mathbf{X}}\}$  as

$$\begin{aligned} V_{\mathbf{X}}^{(r-a)} &= \{\mathbf{X}_i | i = 1, 2, \dots, N_S\} \\ E_{\mathbf{X}}^{(r-a)} &= \{(\mathbf{X}_i, \mathbf{X}_j) | i = 1, 2, \dots, N_S, j \in \mathcal{N}_{\mathbf{X}_i}^K\} \\ w_{\mathbf{X}}(i, j) &= \exp(-\eta dist_{i,j}^{\mathcal{X}}), \forall (\mathbf{X}_i, \mathbf{X}_j) \in E_{\mathbf{X}}^{(r-a)} \end{aligned} \quad (7)$$

where the adaptive  $K$  is calculated with steps 1 to 4 in subsection II-B.3 by setting  $K_{\max} = \lceil \sqrt{|\mathcal{S}|} \rceil$ ,  $K_{\min} = \lceil \sqrt{|\mathcal{S}|}/10 \rceil$  and replacing KNN graphs ( $G_{\mathbf{X}}$  and  $G_{\mathbf{Y}}$ ) with robust KNN graphs ( $G_{\mathbf{X}}^{(r)}$  and  $G_{\mathbf{Y}}^{(r)}$ ), respectively.

By constructing and mapping these robust adaptive KNN graphs, we can re-calculate the structure difference (6) as

$$\begin{aligned} f_i^{\mathbf{Y}} &= \frac{1}{K_i} \left( \sum_{j' \in \mathcal{N}_{\mathbf{X}_i}^{K_i}} dist_{i,j'}^{\mathcal{Y}} - \sum_{j \in \mathcal{N}_{\mathbf{X}_i}^{K_i}} dist_{i,j}^{\mathcal{Y}} \right) \\ f_i^{\mathbf{Y}} &= \frac{1}{K_i} \left( \sum_{j \in \mathcal{N}_{\mathbf{X}_i}^{K_i}} w_{\mathbf{Y}}(i, j) - \sum_{j' \in \mathcal{N}_{\mathbf{X}_i}^{K_i}} w_{\mathbf{Y}}(i, j') \right) \end{aligned} \quad (8)$$

5) *DI With Robust Adaptive KNN Graph*: In addition, by defining  $\mathbf{D}^{\mathbf{X}} \in \mathbb{R}^{N_S \times N_S}$  as the distance matrix of image  $\mathbf{X}$  with element  $[\mathbf{D}^{\mathbf{X}}]_{i,j} = dist_{i,j}^{\mathcal{X}}$ , and  $\mathbf{Z}^{\mathbf{X}} \in \mathbb{R}^{N_S \times N_S}$  as the  $K$ -connection matrix of image  $\mathbf{X}$  with element  $[\mathbf{Z}^{\mathbf{X}}]_{i,j} = 1$  if  $j \in \mathcal{N}_{\mathbf{X}_i}^{K_i}$  and  $[\mathbf{Z}^{\mathbf{X}}]_{i,j} = 0$  if  $j \notin \mathcal{N}_{\mathbf{X}_i}^{K_i}$ , then the robust adaptive KNN graph  $G_{\mathbf{X}}^{(r-a)}$  can be represented by the weighting matrix  $\mathbf{W}^{\mathbf{X}} = \exp(-\eta \mathbf{D}^{\mathbf{X}}) \odot (\mathbf{Z}^{\mathbf{X}})$ , where  $\odot$  represents the Hadamard product, and  $[\mathbf{W}^{\mathbf{X}}]_{i,j} > 0$  represents a directed edges from  $\mathbf{X}_i$  to  $\mathbf{X}_j$  with weights  $[\mathbf{W}^{\mathbf{X}}]_{i,j}$ , and  $[\mathbf{W}^{\mathbf{X}}]_{i,j} = 0$  means that there is no directed edges from  $\mathbf{X}_i$  to  $\mathbf{X}_j$ . Through similar definition, we can define  $\mathbf{D}^{\mathbf{Y}}$ ,  $\mathbf{Z}^{\mathbf{Y}}$ , and  $\mathbf{W}^{\mathbf{Y}}$  for image  $\mathbf{Y}$ . Using these defined matrices, we can rewrite (8) into its equivalent form as

$$\begin{aligned} f_i^{\mathbf{Y}} &= \frac{1}{K_i} \sum_{j=1}^{N_S} [\mathbf{D}^{\mathbf{Y}} \odot (\mathbf{Z}^{\mathbf{X}} - \mathbf{Z}^{\mathbf{Y}})]_{i,j} \\ f_i^{\mathbf{Y}} &= \frac{1}{K_i} \sum_{j=1}^{N_S} [\exp(-\eta \mathbf{D}^{\mathbf{Y}}) \odot (\mathbf{Z}^{\mathbf{Y}} - \mathbf{Z}^{\mathbf{X}})]_{i,j} \end{aligned} \quad (9)$$

From (9), we can find that  $f^{\mathbf{Y}}$  is computed in the  $\mathcal{Y}$  domain by calculating the difference between the mapped edges  $\exp(-\eta \mathbf{D}^{\mathbf{Y}}) \odot \mathbf{Z}^{\mathbf{X}}$  and its own edges  $\mathbf{W}^{\mathbf{Y}}$ , which can be seen as the  $\exp(-\eta \mathbf{D}^{\mathbf{Y}})$  weighted difference between two  $K$ -connection matrices  $\mathbf{Z}^{\mathbf{X}}$  and  $\mathbf{Z}^{\mathbf{Y}}$ .

Similarly, the backward change level  $f^{\mathbf{X}}$  can be measured by mapping the  $G_{\mathbf{Y}}^{(r-a)}$  to the  $\mathcal{X}$  domain, and comparing the structure difference between the mapped  $G_{\mathbf{X}}^{(r-a)map}$  and  $G_{\mathbf{X}}^{(r-a)}$  as

$$\begin{aligned} f_i^{\mathbf{X}} &= \frac{1}{K_i} \sum_{j=1}^{N_S} [\mathbf{D}^{\mathbf{X}} \odot (\mathbf{Z}^{\mathbf{Y}} - \mathbf{Z}^{\mathbf{X}})]_{i,j} \\ f_i^{\mathbf{X}} &= \frac{1}{K_i} \sum_{j=1}^{N_S} [\exp(-\eta \mathbf{D}^{\mathbf{X}}) \odot (\mathbf{Z}^{\mathbf{X}} - \mathbf{Z}^{\mathbf{Y}})]_{i,j} \end{aligned} \quad (10)$$

Before fusing the forward  $f^{\mathbf{Y}}$  and backward  $f^{\mathbf{X}}$ , it is reasonable to clip the change levels  $f^{\mathbf{Y}}$  and  $f^{\mathbf{X}}$  beyond some standard deviations of the mean value (e.g.,  $f^{\mathbf{X}} > \text{mean}(f^{\mathbf{X}}) + 3 \times \text{std}(f^{\mathbf{X}})$ ), so that outliers do not compromise the fusion step. Then, we can obtain the final change level as

$$f^{final} = f^{\mathbf{X}} / \text{mean}(f^{\mathbf{X}}) + f^{\mathbf{Y}} / \text{mean}(f^{\mathbf{Y}}) \quad (11)$$

By assigning the forward, backward and final change levels to the  $i$ -th superpixel  $\Lambda_i^S$ , we can obtain the forward ( $DI^{fw}$ ),

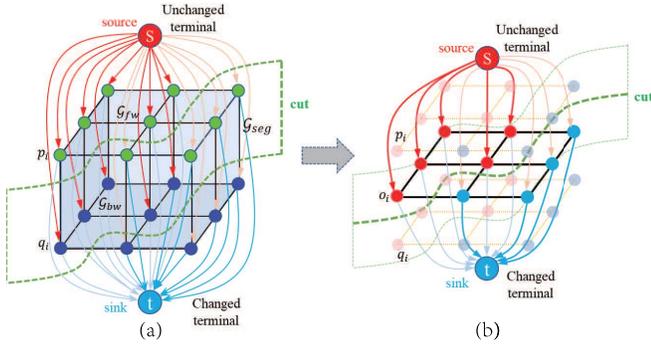


Fig. 3. Example of graph cut co-segmentation: (a) graph  $\mathcal{G}_{seg}$  connects two subgraphs  $\mathcal{G}_{fw}$  and  $\mathcal{G}_{bw}$ ; (b) transformed single graph segmentation.

backward ( $DI^{bw}$ ) and final ( $DI^{final}$ ) DIs, respectively.

$$DI_{m,n}^{fw} = f_i^Y, \quad DI_{m,n}^{fw} = f_i^X, \quad DI_{m,n}^{final} = f_i^{final}; \quad (m, n) \in \Lambda_i^S \quad (12)$$

### C. Binary CM Calculation by MRF Model

Once the DIs are obtained, the CD problem can be converted into a binary segmentation problem. We define the binary CM calculation as a superpixel-labeling problem, which assigns a label  $L_i$  for each superpixel  $\Lambda_i^S$  with  $L_i = 1$  representing the changed class and  $L_i = 0$  representing the unchanged class. This superpixel-labeling problem can be realized by minimizing the energy function  $\mathcal{E}$ , which is the log likelihood of the posterior distribution of an MRF [35], [36], composing a data energy  $\mathcal{E}_d$  and a smoothness energy  $\mathcal{E}_s$

$$\min_{L \in \mathbb{R}^{N_S}} \mathcal{E}(L) = \mathcal{E}_d(L) + \mathcal{E}_s(L) \quad (13)$$

where  $\mathcal{E}_d$  mainly captures the change level information and the  $\mathcal{E}_s$  mainly captures the spatial contextual information in the MRF framework.

In order to intuitively understand the construction of energy function, we design a graph  $\mathcal{G}_{seg} = \{V_{seg}, E_{seg}, w_{seg}\}$  for the MRF model (in order to avoid the symbol confusion with the previous KNN graph  $G$ , we use  $\mathcal{G}$  to represent the graph constructed in MRF model). Each edge  $E_{seg}$  in the graph  $\mathcal{G}_{seg}$  is assigned a nonnegative weight to form the energy function  $\mathcal{E}(L)$ , and the minimum cost of the graph cut gives an optimal labeling  $L$  in (13).

The designed graph  $\mathcal{G}_{seg}$  connects two subgraphs  $\mathcal{G}_{fw} = \{V_{fw}, E_{fw}, w_{fw}\}$  and  $\mathcal{G}_{bw} = \{V_{bw}, E_{bw}, w_{bw}\}$ , which represent forward and backward DIs respectively, as shown in Fig. 3. In the subgraphs, the nodes  $p_i \in V_{fw}$  and  $q_i \in V_{bw}$  in  $\mathcal{G}_{fw}$  and  $\mathcal{G}_{bw}$  represents the superpixels  $\Lambda_i^S$  in the forward and backward DIs, respectively. Meanwhile, there are two additional nodes: the unchanged terminal (source  $s$ ) and the changed terminal (sink  $t$ ). Therefore, we have

$$V_{seg} = V_{fw} \cup V_{bw} \cup \{s, t\} \quad (14)$$

The set of edges consists of three types of edges:  $t$ -links (terminal links),  $g$ -links (subgraph links), and  $n$ -links (neighborhood links). Each superpixel in  $DI^{fw}$  and  $DI^{bw}$  is connected to two terminals (unchanged and change) through two  $t$ -links, and each superpixel at the same location in

$DI^{fw}$  and  $DI^{bw}$  is connected together through a  $g$ -link. The  $R$ -adjacency neighborhood system is used for  $n$ -links, that is, as long as the boundaries of two superpixels ( $\Lambda_i^S$  and  $\Lambda_j^S$ ) intersect or the distance between the centers of two superpixels is less than  $R$  (we set  $R = 2\sqrt{MN}/N_S$  in this paper for simplicity), the two superpixels are marked to be  $R$ -adjacency neighbors of each other as  $i \in \mathcal{N}_j^{R-adj}$  (or  $j \in \mathcal{N}_i^{R-adj}$ ). Then, each pair of neighboring superpixels ( $p_i, p_j$ ) or ( $q_i, q_j$ ) is connected by an  $n$ -link. We have

$$\begin{aligned} E_{seg} &= E_t \cup E_g \cup E_n \\ E_t &= \{(s, p_i), (p_i, t) \mid \forall p_i \in V_{fw}\} \cup \{(s, q_i), (q_i, t) \mid \forall q_i \in V_{bw}\} \\ E_g &= \{(p_i, q_i) \mid p_i \in V_{fw}, q_i \in V_{bw}, i = 1, \dots, N_S\} \\ E_n &= \{(p_i, p_j), (q_i, q_j) \mid \forall p_i \in V_{fw}, \forall q_i \in V_{bw}, j \in \mathcal{N}_i^{R-adj}\} \end{aligned} \quad (15)$$

All edges in  $E_{seg}$  are assigned some weights or costs. The cost of  $t$ -links corresponds to a penalty for assigning the corresponding label to the superpixel

$$\begin{aligned} w_t(s, p_i) &= \max\{-\lambda \log(f_i^Y/2T_1), 0\} \\ w_t(s, q_i) &= \max\{-\lambda \log(f_i^X/2T_2), 0\} \\ w_t(p_i, t) &= \begin{cases} -\lambda \log(1 - f_i^Y/2T_1), & \text{if } f_i^Y \leq 2T_1 \\ W_1, & \text{otherwise} \end{cases} \\ w_t(q_i, t) &= \begin{cases} -\lambda \log(1 - f_i^X/2T_2), & \text{if } f_i^X \leq 2T_2 \\ W_2, & \text{otherwise} \end{cases} \end{aligned} \quad (16)$$

and it can be used to form the data energy term  $\mathcal{E}_d$  in (13) as

$$\begin{aligned} \mathcal{E}_d &= \sum_{i=1}^{N_S} D_{p_i}(L_{p_i}) + \sum_{i=1}^{N_S} D_{q_i}(L_{q_i}) \\ D_{p_i}(L_{p_i}) &= \begin{cases} w_t(s, p_i), & \text{if } L_{p_i} = 1 \\ w_t(p_i, t), & \text{if } L_{p_i} = 0 \end{cases} \\ D_{q_i}(L_{q_i}) &= \begin{cases} w_t(s, q_i), & \text{if } L_{q_i} = 1 \\ w_t(q_i, t), & \text{if } L_{q_i} = 0 \end{cases} \end{aligned} \quad (17)$$

where  $\lambda \in (0, 1)$  is a balance parameter,  $T_1$  and  $T_2$  are two threshold parameters, which are determined by the maximum inter class variance criterion with the Otsu method [37] on  $f^Y$  and  $f^X$ , respectively. Specially, when  $f_i^Y > 2T_1$  (or  $f_i^X > 2T_2$ ), the cost of edge ( $s, p_i$ ) (or ( $s, q_i$ )) is assigned to zero because of the hard constraint condition, and the cost of edge ( $p_i, t$ ) (or ( $q_i, t$ )) is assigned to  $W_1$  (or  $W_2$ ), which is set to be the maximum sum of all  $n$ -links weights for each superpixel. In addition, this type of  $\mathcal{E}_d$  can be regarded as derived from the assumption that  $f_i^Y$  (or  $f_i^X$ ) obeys a uniform distribution between 0 and  $2T_1$  (or  $2T_2$ ).

The cost of  $g$ -links corresponds to a penalty for discontinuity between the  $i$ -th superpixel in  $DI^{fw}$  and the  $i$ -th superpixel in  $DI^{bw}$ , which can be used to form the consistency energy term  $\mathcal{E}_c$ . Since the pre-event and post-event images are completely registered, the position of the changed area in these

two images is the same, that is,  $L_p$  and  $L_q$  should be the same. Then, we have

$$w_g(p_i, q_i) = \infty$$

$$\mathcal{E}_c = \sum_{i=1}^{N_S} -\log(1 - |L_{p_i} - L_{q_i}|) \quad (18)$$

The cost of  $n$ -links corresponds to a penalty for discontinuity between the superpixels and their  $R$ -adjacency neighbors, which can be used to form the adjacency energy term  $\mathcal{E}_a$ . Here, we propose a novel costs of  $n$ -links, which takes into account not only the spatial continuity, but also the relationship between two subgraphs in the heterogeneous CD task.

$$w_n(p_i, p_j)_{j \in \mathcal{N}_i^{R-adj}}$$

$$= w_n(q_i, q_j)_{j \in \mathcal{N}_i^{R-adj}} = \frac{1 - \lambda}{2d(\Lambda_i^S, \Lambda_j^S)}$$

$$\times \begin{cases} \exp\left(-\frac{dist_{i,j}^Y}{2\sigma_1^2} - \frac{dist_{i,j}^X}{2\sigma_2^2}\right), \\ \quad \text{if } dist_{i,j}^Y \leq \sigma_1^2, dist_{i,j}^X \leq \sigma_2^2 \\ \exp\left(\frac{dist_{i,j}^Y}{2\sigma_1^2} - \frac{dist_{i,j}^X}{2\sigma_2^2} - 1\right), \\ \quad \text{if } dist_{i,j}^Y \leq \sigma_1^2, dist_{i,j}^X > \sigma_2^2 \\ \exp\left(-\frac{dist_{i,j}^Y}{2\sigma_1^2} + \frac{dist_{i,j}^X}{2\sigma_2^2} - 1\right), \\ \quad \text{if } dist_{i,j}^Y > \sigma_1^2, dist_{i,j}^X \leq \sigma_2^2 \\ \exp(-1), \\ \quad \text{if } dist_{i,j}^Y > \sigma_1^2, dist_{i,j}^X > \sigma_2^2 \end{cases} \quad (19)$$

where  $d(\Lambda_i^S, \Lambda_j^S)$  represents the Euclidean spatial distance between superpixels,  $\sigma_1^2 = \langle dist_{i,j}^Y \rangle$  and  $\sigma_2^2 = \langle dist_{i,j}^X \rangle$  are two normalization parameters with  $\langle \cdot \rangle$  representing the average feature distance over the whole image. Then we have the corresponding adjacency energy term  $\mathcal{E}_a$  as

$$\mathcal{E}_a = \sum_{i=1}^{N_S} \sum_{j \in \mathcal{N}_i^{R-adj}} \{w_n(p_i, p_j) \delta(L_{p_i} - L_{p_j})$$

$$+ w_n(q_i, q_j) \delta(L_{q_i} - L_{q_j})\} \quad (20)$$

where function  $\delta(\cdot)$  is defined as  $\delta(x) = 0$  if  $x = 0$  and  $\delta(x) = 1$  if  $x \neq 0$ . From (19) and (20), we can find that it gives a penalty for the discontinuity of  $L_{p_i} \neq L_{p_j}$  and  $L_{q_i} \neq L_{q_j}$  in four cases: a large penalty when  $\mathbf{Y}_i$  and  $\mathbf{Y}_j$ ,  $\mathbf{X}_i$  and  $\mathbf{X}_j$  are similar; a small penalty when  $\mathbf{Y}_i$  and  $\mathbf{Y}_j$  are similar and  $\mathbf{X}_i$  and  $\mathbf{X}_j$  are dissimilar; a small penalty when  $\mathbf{Y}_i$  and  $\mathbf{Y}_j$  are dissimilar and  $\mathbf{X}_i$  and  $\mathbf{X}_j$  are similar; and a median penalty when  $\mathbf{Y}_i$  and  $\mathbf{Y}_j$ ,  $\mathbf{X}_i$  and  $\mathbf{X}_j$  are dissimilar.

With these constructed costs of different links (16), (18), (19) and the corresponding energy terms (17), (18), (20), we can obtain the total energy  $\mathcal{E}$

TABLE I

THE OVERALL FRAMEWORK OF IRG-MCS

Algorithm 1. IRG-McS

**Input:** Images of  $\mathbf{X}$  and  $\mathbf{Y}$ , parameters of  $N_S$ ,  $Iter$ , and  $\lambda$ .**Preprocessing:** Superpixel segmentation and feature extractionImplement the (modified) SLIC on  $\mathbf{X}$  and  $\mathbf{Y}$  to obtain  $\Lambda^{\mathbf{X}}$  and  $\Lambda^{\mathbf{Y}}$ .Intersect  $\Lambda^{\mathbf{X}}$  and  $\Lambda^{\mathbf{Y}}$  to obtain  $\Lambda^{\mathbf{S}}$  with  $N_S$  superpixels.Extract the features to obtain  $\tilde{\mathbf{X}}$  and  $\tilde{\mathbf{Y}}$ .**Main iteration loop of IRG-McS:**Set initial index subset as  $S^0 = \{1, 2, \dots, N_S\}$ .for  $i = 1, 2, \dots, Iter$  do**1. DI generation:**Construct the KNN graphs of  $G_{\tilde{\mathbf{X}}}^{(r-a)}$  and  $G_{\tilde{\mathbf{Y}}}^{(r-a)}$  with  $S^{i-1}$ .Calculate the change levels of  $f_{\tilde{\mathbf{X}}}^{(r-a)}$  and  $f_{\tilde{\mathbf{Y}}}^{(r-a)}$  by graph mapping.Compute the DIs of  $DI^{fw}$  and  $DI^{bw}$  with (12).**2. Binary CM calculation:**Calculate the edge weights of  $w_t$  (16) and  $w_n$  (19).Solve the graph cut problem (22) to obtain  $S^i$  and  $\mathcal{T}^i$ .Exit criterion: calculate  $\xi^i = 1 - |S^{i-1} \cap S^i| / |S^i|$ if  $\xi^i < \xi^0$ 

exit

End if

End for

**Output:** Compute the final change map with (23).

as

$$\mathcal{E} = \mathcal{E}_d + \mathcal{E}_c + \mathcal{E}_a$$

$$= \sum_{i=1}^{N_S} \{D_{p_i}(L_{p_i}) + D_{q_i}(L_{q_i}) - \log(1 - |L_{p_i} - L_{q_i}|)\}$$

$$+ \sum_{j \in \mathcal{N}_i^{R-adj}} w_n(p_i, p_j) \delta(L_{p_i} - L_{p_j})$$

$$+ w_n(q_i, q_j) \delta(L_{q_i} - L_{q_j}) \quad (21)$$

This energy minimization problem is equivalent to the minimum cut problem on the graph  $\mathcal{G}_{seg} = \{V_{seg}, E_{seg}, w_{seg}\}$ , which is to find a cut that has the minimum cut cost among all cuts to partition the nodes in the graph into two disjoint subsets: the unchanged nodes connecting to source  $s$  and the changed nodes connecting to sink  $t$ .

Due to the hard constraint of  $\mathcal{E}_c$ , we have an equality constraint  $L_{p_i} = L_{q_i}$ ,  $i = 1, \dots, N_S$  and the optimal cut take the form of plane cuts as shown in Fig. 3(a). Then, the minimization problem of (21) can be rewritten as

$$\min_{L_o \in \mathbb{R}^{N_S}} \mathcal{E} = \sum_{i=1}^{N_S} \{D_{o_i}(L_{o_i})$$

$$+ \sum_{j \in \mathcal{N}_i^{R-adj}} w_n(o_i, o_j) \delta(L_{o_i} - L_{o_j})\} \quad (22)$$

where  $D_{o_i}(L_{o_i}) = D_{p_i}(L_{p_i}) + D_{q_i}(L_{q_i})$  and  $w_n(o_i, o_j) = w_n(p_i, p_j) + w_n(q_i, q_j)$ . Therefore, the co-graph segmentation problem is transformed to a single graph segmentation as shown in Fig. 3(b), which can be solved by the min-cut/max-flow algorithm in [38]. Once the optimal  $L_o$  is assigned, we can obtain the unchanged index subset of nodes connecting to source  $\mathcal{S} = \{i | L_{o_i} = 0; i = 1, \dots, N_S\}$  and the changed index subset of nodes connecting to sink  $\mathcal{T} = \{i | L_{o_i} = 1; i = 1, \dots, N_S\}$ .

The final changed map can be obtained as

$$cm(m, n) = L_{o_i}; \text{ if } (m, n) \in \Lambda_i^S \quad (23)$$

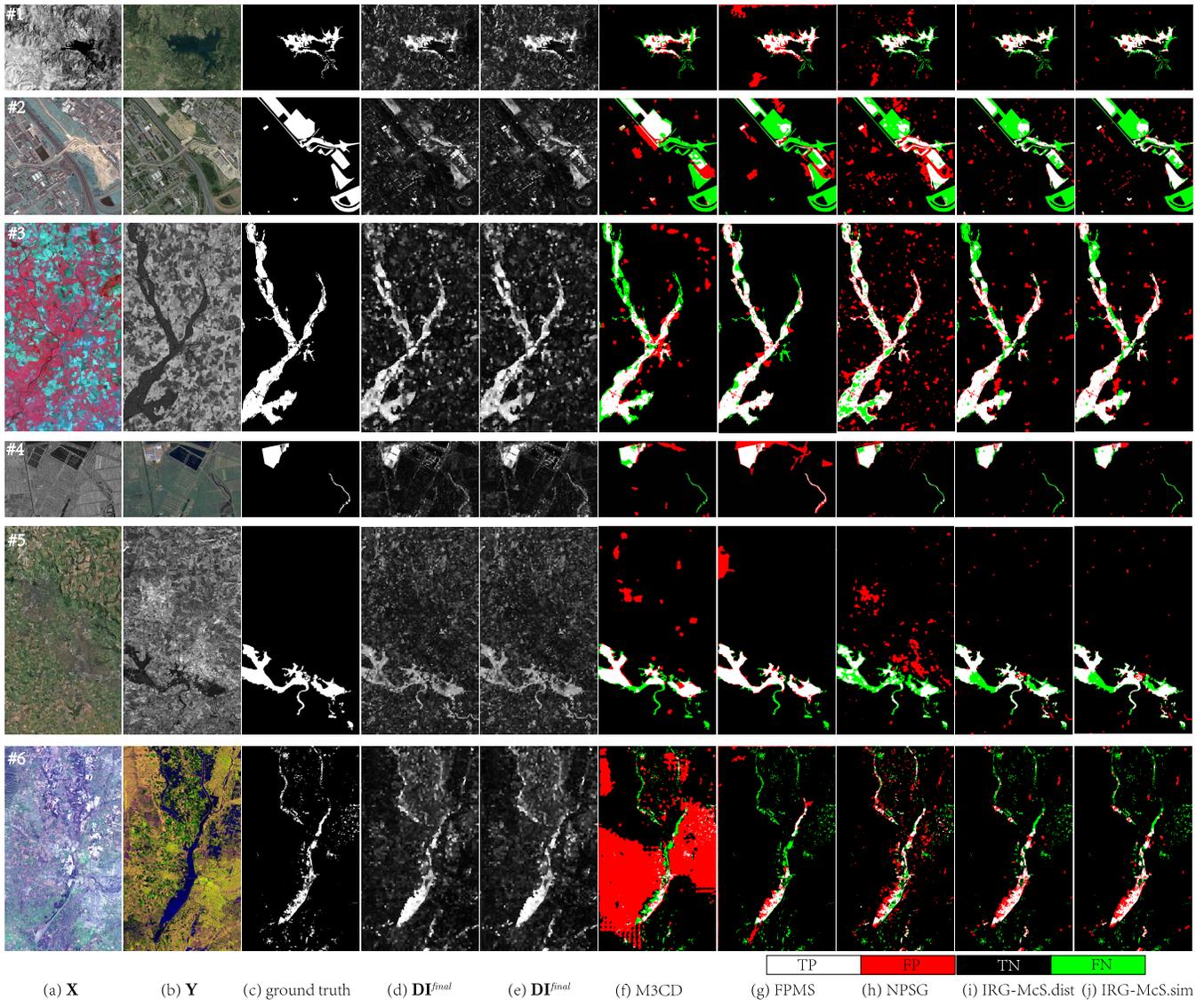


Fig. 4. Fused DI of IRG-McS and binary CM of different methods on heterogeneous data set. From top to bottom, they correspond to datasets #1 to #6, respectively. From left to right are: (a) pre-event image; (b) post-event image; (c) the ground truth; (d) fused DI of IRG-McS with distance criterion; (e) fused DI of IRG-McS with similarity criterion; (f) binary CM of M3CD; (g) binary CM of FPMS; (h) binary CM of NPSG; (i) binary CM of IRG-McS with distance criterion; (j) binary CM of IRG-McS with similarity criterion. In the binary CM, White: true positives (TP); Red: false positives (FP); Black: true negatives (TN); Green: false negatives (FN).

Meanwhile, it should be noted that this fused graph segmentation is different from the segmentation on the fused DI, that is, instead of first fusing DI and then segmenting it, the propose method combines fusion and segmentation in the same framework. The benefit of this combination is that it can further explore the information in the forward and backward DIs (as shown in (16), (18), and (19)) and better use them to improve the CD accuracy. The overall framework of the proposed IRG-McS is summarized in Table I.

### III. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, experiments are performed to evaluate the performance of IRG-McS, which are conducted on six pairs of heterogeneous RS images.

#### A. Heterogeneous Data Sets and Quantitative Measures

Six heterogeneous data sets are used to verify the effectiveness of the proposed approach, as listed in Table II. These data sets include two different types of heterogeneity: multisensor optical image pairs (images are acquired by different optical sensors, e.g., #1, #2, and #3 in the first three rows of Figs. 4(a)-(c)) and multisource image pairs (images are acquired by different types of sensors, e.g., #4, #5, and #6<sup>1</sup> [26] in the last three rows of Figs. 4(a)-(c)). These data sets cover different image sizes (varying from 300 to 4135 pixels in length or width), different resolution levels (varying from 0.52m to 30m) and different types of changes (lake overflow, flooding and building construction), which can evaluate the robustness of the algorithm in different scenarios.

<sup>1</sup>Dataset #6 is kindly available at <https://sites.google.com/view/luppino>

TABLE II  
DESCRIPTION OF THE SIX HETEROGENEOUS DATA SETS USED IN THE MANUSCRIPT

Dataset	Sensor	Size (pixels)	Date	Location	Event (& Spatial resolution)
#1	Landsat-5/Google Earth	300 × 412 × 1(3)	Sept. 1995 - July 1996	Sardinia, Italy	Lake expansion (30m.)
#2	Pleiades/WorldView2	2000 × 2000 × 3(3)	May 2012 - July 2013	Toulouse, France	Construction (0.52m.)
#3	Spot/NDVI	990 × 554 × 3(1)	1999 - 2000	Gloucester, England	Flooding (≈ 25m.)
#4	Radarsat-2/Google Earth	593 × 921 × 1(3)	June 2008 - Sept. 2012	Shuguang Village, China	Building construction (8m.)
#5	Quick Bird 2/TerraSAR-X	4135 × 2325 × 3(1)	July 2006 - July 2007	Gloucester, England	Flooding (0.65m)
#6	Landsat-8/Sentinel-1A	875 × 500 × 11(3)	Jan. 2017 - Feb. 2017	Sutter County, USA	Flooding (≈ 15m.)

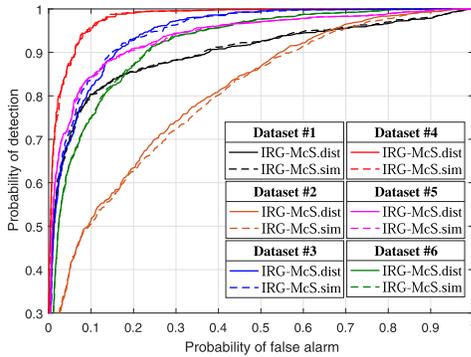


Fig. 5. ROC curves of IRG-McS generated DIs.

In order to discuss and compare CD results of different methods, two types of quantitative measures are selected. The quality of DI can be evaluated by the empirical receiver operating characteristics (ROC) curves, which take the probability of detection (PD) as a function of the probability of false alarm (PFA). Moreover, the area under the curve (AUC) is used as the quantitative criterion for the ROC curve. The quality of binary CM can be evaluated by the widely used measures: the percentage of correct classification (PCC), the Kappa coefficient (KC) and the F-measure:  $F_m = (2TP)/(2TP + FP + FN)$ , where TP, TN, FN, FP represent the true positives, negatives, false negatives and positives, respectively.

### B. Experimental Results

The main parameters of IRG-McS are the  $N_S$  of the superpixels numbers, the maximum number of iterations  $Iter$ , and the balance parameter  $\lambda$ . For all the experiments, we set  $N_S = 5000$ ,  $Iter = 6$ ,  $\lambda = 0.05$  (except  $\lambda = 0.1$  for dataset #2). The impact of these parameters will be analyzed in detail in subsection III-C.

Figs. 4(d)-(e) show the fused DIs  $DI^{final}$  (12) of different data sets generated by IRG-McS with distance criterion in (8) (denoted as IRG-McS.dist for short) or similarity criterion in (8) (denoted as IRG-McS.sim for short). The ROC curves of these DIs are plotted in Fig. 5. From these results, we can find that the proposed IRG-McS can well build the relationship between heterogeneous images and highlight the changes in the DIs, which can obtain high quality ROC curves and gain large AUC (as listed in the last row of Table V).

In order to evaluate the binary CM generated by IRG-McS<sup>2</sup>, we choose the recently proposed Markov Model for Multimodal CD method (M3CD)<sup>3</sup> [39], FPMS<sup>4</sup>, and NPSG<sup>5</sup> for comparison, and use the default parameters in their codes, which are also consistent with their papers. Figs. 4(f)-(j) show the final CM of different methods on all the evaluated heterogeneous data sets. Table III reports the PCC, KC and Fm of these different methods. From the comparison of CMs in Fig. 4, we can see that the changed and unchanged areas are well detected by IRG-McS with relatively small FN and FP. From the quantitative measures of CMs in Table III, we can find that the IRG-McS can obtain the best or second-best result whether using distance or similarity criterion, which demonstrates the effectiveness of the proposed heterogeneous CD method.

In order to further evaluate the performance of IRG-McS, we also select some other representative and state-of-the-art (SOTA) methods for comparison as summarized in Table IV, including HPT [12], SCCN [14], RMN (reliable mixed-norm based method) [17], PSGM [20], AM-IR [26], LT-FL [28], DFR-MT (deep feature representation and mapping transformation based method) [40], MDS (multidimensional scaling based method) [41], ALSC (adaptive local structure consistency based method) [42], AFL-DSR (anomaly feature learning based deep sparse residual model) [43], MDER (multidimensional evidential reasoning based method) [44], and NLPem (nonlocal pairwise energy-based model) [45], where SCCN, LT-FL, DFR-MT and AFL-DSR are deep learning based methods. For the sake of fairness, we directly quote the results of the corresponding datasets in their original published papers. From Table IV, we can see that the IRG-McS can achieve better or quite competitive accuracy rate by comparing with these SOTA methods, and show the ability to gain consistent good results on different data sets. The average PCC obtained on the six heterogeneous data sets of IRG-McS is about 0.95.

### C. Discussion

1) *The Effectiveness of Robust Adaptive KNN Graph:* We have proposed a strategy to adaptively select  $K$  for each vertex of the KNN graph. To verify the effectiveness of the proposed adaptive  $K$ -selection strategy, we will compare the

<sup>2</sup>IRG-McS will be available at <https://github.com/yulisun/IRG-McS>

<sup>3</sup>M3CD is kindly available at <http://www-labs.iro.umontreal.ca/~mignotte>

<sup>4</sup>FPMS is kindly available at <http://www-labs.iro.umontreal.ca/~mignotte>

<sup>5</sup>NPSG is available at <https://github.com/yulisun/NPSG>

TABLE III  
QUANTITATIVE MEASURES OF BINARY CMS ON THE HETEROGENEOUS DATA SETS

Methods	Dataset #1			Dataset #2			Dataset #3			Dataset #4			Dataset #5			Dataset #6			Average		
	PCC	KC	Fm																		
M3CD	0.963	0.669	0.689	0.863	0.405	<b>0.481</b>	0.915	0.588	0.636	0.962	0.602	0.622	0.955	0.641	0.665	0.575	0.021	0.077	0.872	0.487	0.528
FPMS	0.925	0.552	0.588	0.838	0.215	0.296	<b>0.962</b>	<b>0.816</b>	<b>0.837</b>	0.938	0.569	0.597	0.970	<b>0.771</b>	<b>0.787</b>	0.947	0.329	0.356	0.930	0.542	0.577
NPSG	0.947	0.559	0.587	0.830	0.346	0.446	0.902	0.608	0.663	0.975	0.729	0.742	0.929	0.413	0.451	0.941	0.419	0.449	0.921	0.512	0.556
IRG-McS.dist	<b>0.971</b>	<b>0.739</b>	<b>0.754</b>	<b>0.882</b>	0.420	0.478	0.939	0.714	0.749	<b>0.983</b>	<b>0.794</b>	<b>0.804</b>	<b>0.972</b>	0.737	0.751	<b>0.959</b>	<b>0.490</b>	<b>0.512</b>	<b>0.951</b>	<b>0.649</b>	<b>0.675</b>
IRG-McS.sim	<b>0.971</b>	0.733	0.749	0.881	<b>0.421</b>	<b>0.481</b>	0.942	0.735	0.768	0.978	0.755	0.767	<b>0.972</b>	0.740	0.755	0.958	0.487	0.508	0.950	0.645	0.671

TABLE IV  
ACCURACY RATE OF CM GENERATED BY DIFFERENT METHODS ON DIFFERENT DATA SETS. THE RESULTS OF THESE COMPARISON METHODS ARE REPORTED BY THEIR ORIGINAL PUBLISHED PAPERS

Dataset #1	OA	Dataset #2	OA	Dataset #3	OA	Dataset #4	OA	Dataset #5	OA	Dataset #6	OA
DFR-MT <sup>[40]</sup>	0.975	<b>IRG-McS</b>	<b>0.882</b>	HPT <sup>[12]</sup>	0.957-0.964	<b>IRG-McS</b>	<b>0.983</b>	<b>IRG-McS</b>	<b>0.972</b>	<b>IRG-McS</b>	<b>0.959</b>
<b>IRG-McS</b>	<b>0.971</b>	AFL-DSR <sup>[43]</sup>	0.880	<b>IRG-McS</b>	<b>0.942</b>	AFL-DSR <sup>[43]</sup>	0.980	NLPEM <sup>[45]</sup>	0.949	ALSC <sup>[42]</sup>	0.944
ALSC <sup>[42]</sup>	0.965	RMN <sup>[17]</sup>	0.877	AFL-DSR <sup>[43]</sup>	0.836	PSGM <sup>[20]</sup>	0.977	RMN <sup>[17]</sup>	0.943	AM-IR <sup>[26]</sup>	0.933
PSGM <sup>[20]</sup>	0.961	NLPEM <sup>[45]</sup>	0.853	MDER <sup>[44]</sup>	0.818	SCCN <sup>[14]</sup>	0.976	AFL-DSR <sup>[43]</sup>	0.892		
MDS <sup>[41]</sup>	0.942					MDS <sup>[41]</sup>	0.967				
AFL-DSR <sup>[43]</sup>	0.929					LT-FL <sup>[28]</sup>	0.964				
RMN <sup>[17]</sup>	0.847					ALSC <sup>[42]</sup>	0.963				
						RMN <sup>[17]</sup>	0.884				

TABLE V  
AUC OF DIS GENERATED BY IRG-McS.SIM WITH DIFFERENT  $K$  ON DIFFERENT DATA SETS

Methods	Datasets					
	#1	#2	#3	#4	#5	#6
$K = K_{\min}$	0.882	0.768	0.820	0.961	0.896	0.880
$K = K_{\max}$	0.896	0.792	0.941	0.974	0.929	0.911
$K = K_{\text{mean}}$	0.894	0.794	0.933	0.969	0.931	0.909
adaptive $K_i$	0.900	0.798	0.948	0.980	0.936	0.918

adaptive KNN graph with  $K_i$  and the fixed KNN graphs with fixed  $K = K_{\min}, K_{\max}, K_{\text{mean}}$  in the DI generation, where  $K_{\text{mean}} = \frac{1}{N_S} \sum_{i=1}^{N_S} K_i$  is the mean value of the adaptive  $K_i$ . Table V lists the AUC of DIs generated by IRG-McS with different  $K$  on datasets #1 to #6. We can find that the adaptive  $K_i$  produces a higher AUC than the fixed  $K$ , whether  $K$  takes the maximum value  $K_{\max}$ , the minimum value  $K_{\min}$ , or the mean value  $K_{\text{mean}}$ . This demonstrates the superiority of the adaptive KNN graph.

In the IRG-McS, a robust KNN graph is constructed by adaptively selecting unchanged NNs though an iterative framework combining the DI generation and CM calculation processes, which can reduce the impact of unstable neighbors in the change level calculation. To demonstrate the effectiveness of this iterative framework, we will compare the DI generated by each iteration of IRG-McS with similarity criterion in detail. Fig. 6(c) shows the proportion of the changed superpixels in the  $K$  NNs of each superpixel in the pre-event image  $\mathbf{X}$  (or in the graph  $G_{\mathbf{X}}^{(a)}$ ), that is,  $p_{m,n} = \frac{|\mathcal{NC}_{\mathbf{X}_i}^{K_i}|}{K_i}$ ,  $(m, n) \in \Lambda_i^S$ , where  $\mathcal{NC}_{\mathbf{X}_i}^{K_i} = \{j | j \in \mathcal{N}_{\mathbf{X}_i}^{K_i}, \Lambda_j^S \text{ is changed}\}$  and  $|\mathcal{NC}_{\mathbf{X}_i}^{K_i}|$  represents the number of the changed superpixels in  $\mathcal{N}_{\mathbf{X}_i}^{K_i}$ . Fig. 6(d) shows the proportion of the changed superpixels in

the  $K$  NNs of the post-event event image  $\mathbf{Y}$  (or in the graph  $G_{\mathbf{Y}}^{(a)}$ ). From Figs. 6(a)-(d), we can easily find that the NNs of some superpixels in the graph  $G_{\mathbf{Y}}^{(a)}$  contain a large number of changed superpixels, as shown in the middle of Fig. 6(d) (representing the lake). As analyzed in subsection II-B.4, these changed NNs will make the structure difference measurement less discriminative. Therefore, we can find that the quality of backward DI in Fig. 6(e) ( $DI^{bw}$  of the initial iteration) is much worse than that of forward DI in Fig. 6(f) ( $DI^{fw}$  of the initial iteration), and the backward DI is hard to distinguish between the changed and the unchanged areas, which is consistent with our theoretical analysis. Figs. 6(g) and (h) show the backward and forward DIs of the final iteration respectively, which are generated by calculating the structure difference of (8) with the robust adaptive KNN graphs by removing the detected changed superpixels of previous iteration in the NNs. By comparing Figs. 6(e) and (g), we can clearly find that the backward DI of the final iteration is much better than that of the initial iteration, which demonstrates the effectiveness of the proposed robust KNN graph with the iterative framework. In addition, we also plot the ROC curves of these DIs of initial and final iterations in Fig. 7 for comparison.

2) *Parameter Analysis*: Another issue to be discussed is the sensitivity of parameters used in IRG-McS. For the superpixels number  $N_S$ , it has two impacts on the algorithm: first, it affects the detection granularity of the algorithm; second, it determines the complexity of the algorithm. A larger  $N_S$  will make the segmented superpixel smaller, which improves the detection granularity. Fig. 8 plots the fused DIs and CMs generated by IRG-McS.dist on dataset #3 with  $N_S = 2500, 5000, 10000$  and  $20000$ . In order to fully compare these detection results, we mark some details with the white regions in the DIs of Fig. 8. We can find that when  $N_S$  is smaller, the size of the generated superpixels is larger, the block effect of the DI is more obvious, and then some minor changes in the CM are easier to be ignored as shown in Fig. 8. On the other

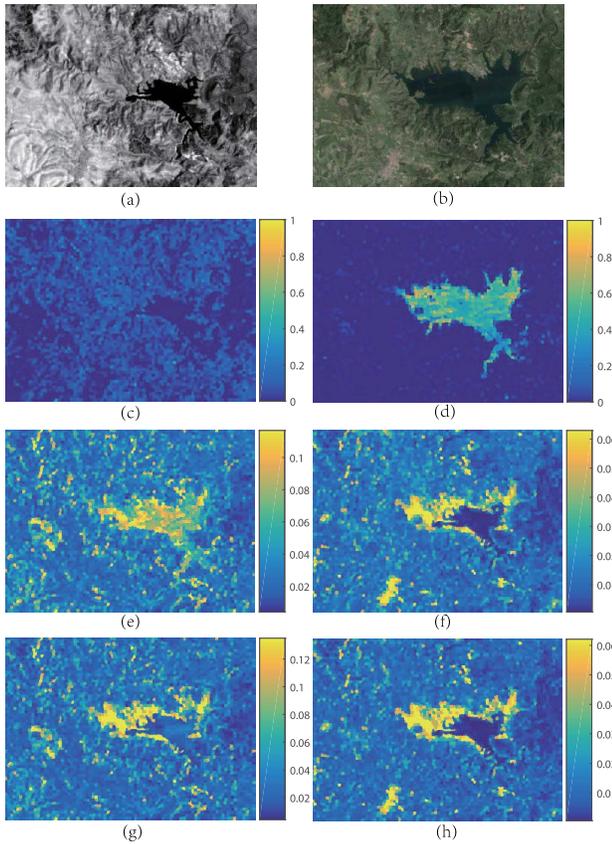


Fig. 6. DIs generated by the initial and final iterations of IRG-McS on dataset #1. (a)-(b) the pre-event and post-event images of dataset #1, respectively. (c)-(d) the proportion of the changed superpixels in the graphs  $G_X^{(a)}$  and  $G_Y^{(a)}$ , respectively. (e)-(f) the backward and forward DIs generated by the initial iteration of IRG-McS, respectively. (g)-(h) the backward and forward DIs generated by the final iteration of IRG-McS, respectively.

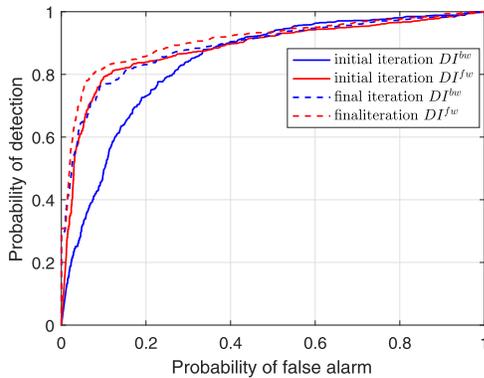


Fig. 7. ROC curves of forward and backward DIs generated by the initial iteration and final iteration of IRG-McS.

hand, a large  $N_S$  also increases the computational complexity as analyzed in the following Subsection III-C.3. In this paper, we simply set  $N_S = 5000$  as a compromise choice, which can be adjusted according to the granularity requirement and the computing environment in practical applications.

For the maximum number of iterations  $Iter$ , according to our experiments, the most obvious benefit of iterative framework occurs after the initial iteration, and then gradually

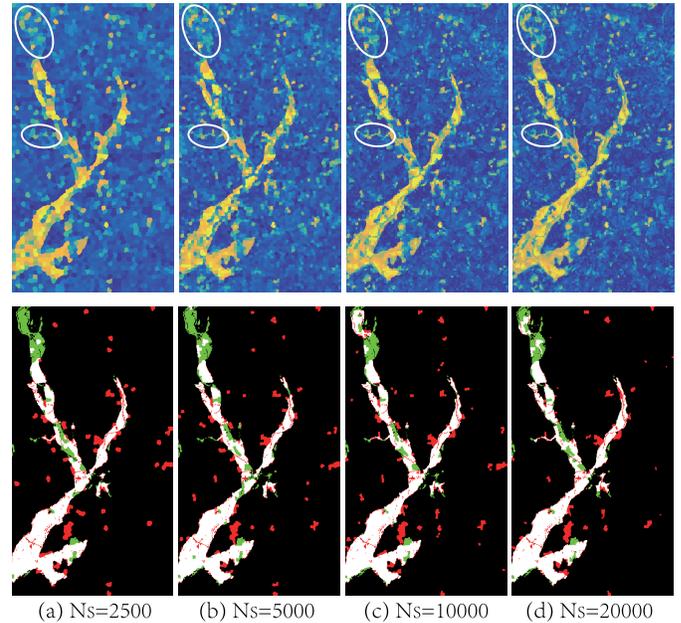


Fig. 8. The fused DIs (top row) and CMs (bottom row) generated by IRG-McS-dist on dataset #3 with different  $N_S$ . From left to right are: (a)  $N_S = 2500$ ; (b)  $N_S = 5000$ ; (c)  $N_S = 10000$ ; (d)  $N_S = 20000$ .

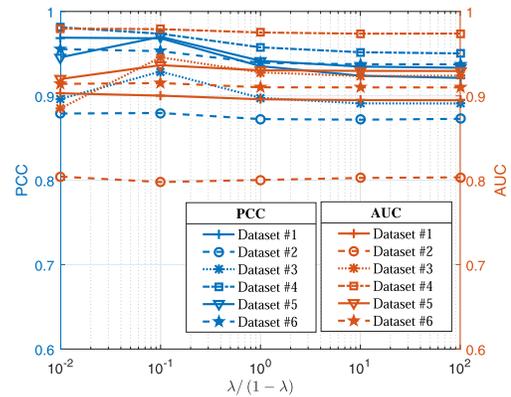


Fig. 9. Influences of parameter  $\lambda$  on the IRG-McS performance.

tends to be stable. At the same time, a larger  $Iter$  will also increase the computational time of IRG-McS. Therefore, we fix  $Iter = 6$  as a compromise choice in our experiments.

For the parameter  $\lambda \in (0, 1)$ , which is used to balance the data energy term and the smoothness energy term in the MRF co-segmentation model. In order to investigate the effect of  $\lambda$  on the IRG-McS, we test different  $\lambda$  on the abovementioned data sets and select the AUC of the fused DI and PCC of the final CM to evaluate the CD results obtained by IRG-McS with similarity criterion, which can describe the comprehensive performance of the method. In Fig. 9, we set  $\lambda$  to make  $\lambda/(1-\lambda)$  vary from 0.01 to 100 with the ratio of 10. It can be found that the IRG-McS is very robust for this balance parameter. There are two main reasons for this: first, for the high-quality DI, using threshold segmentation method alone can obtain good CD results. With the increase of  $\lambda$ , the data energy term becomes more important. In particular,

TABLE VI  
COMPUTATIONAL TIME (SECONDS) OF M3CD, FPMS, NPSG AND IRG-McS WITH  $N_S = 5000$  AND  $N_S = 10000$

Data sets	Image size	M3CD	FPMS	NPSG	IRG-McS ( $N_S = 5000$ )				IRG-McS ( $N_S = 10000$ )			
					$t_{sf}$	$t_{DI}$	$t_{cs}$	$t_{total}$	$t_{sf}$	$t_{DI}$	$t_{cs}$	$t_{total}$
Dataset #1	$300 \times 412 \times 1(3)$	2259.33	23.26	79.71	0.50	0.94	1.34	8.03	1.14	1.99	5.03	22.07
Dataset #2	$2000 \times 2000 \times 3(3)$	2611.58	92.37	176.52	4.18	1.20	1.76	20.53	4.88	1.81	4.93	39.13

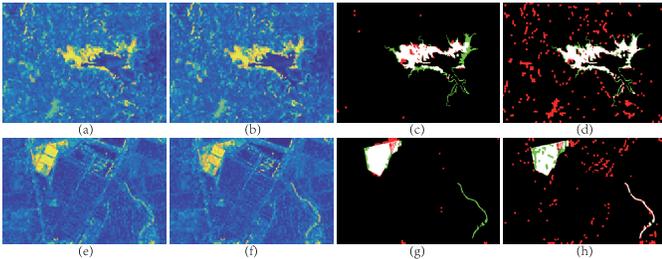


Fig. 10. Fused DIs and final CMs generated by IRG-McS with different  $\lambda$ . The top row is the result on dataset #1: (a) fused DI with  $\lambda = 0.01$ ; (b) fused DI with  $\lambda = 0.99$ ; (c) final CM with  $\lambda = 0.01$ ; (d) final CM with  $\lambda = 0.99$ . The bottom row is the result on dataset #4: (e) fused DI with  $\lambda = 0.01$ ; (f) fused DI with  $\lambda = 0.99$ ; (g) final CM with  $\lambda = 0.01$ ; (h) final CM with  $\lambda = 0.99$ .

when we set  $\lambda = 1$ , the MRF co-segmentation degenerates to the threshold (Otsu) co-segmentation. Second, different from the general smoothness energy term which only consider the spatial constraints, the proposed adjacency energy term  $\mathcal{E}_a$  (20) also contains the relationship between two subgraphs by using (19), which can be used to assist segmentation. That is, even if we set  $\lambda = 0.01$  to reduce the weight of data energy terms  $\mathcal{E}_d$ , we can obtain satisfactory results by using the adjacency energy term  $\mathcal{E}_a$ . Fig. 10 shows the fused DIs and final CMs generated by IRG-McS with  $\lambda = 0.01$  and  $\lambda = 0.99$  on datasets #1 and #4. We can find that IRG-McS is very robust with this balance parameter  $\lambda$ .

3) *Computational Time*: In addition, the computational efficiency of the proposed IRG-McS is also evaluated. In Table VI, we report the computational times of M3CD, FPMS, NPSG, and each process of IRG-McS with  $N_S = 5000, 10000$  on different data sets. The C++ codes of M3CD and FPMS are executed in a Linux computer with Intel(R) Xeon(R) Silver 4110 CPU and 31 GB of RAM. The NPSG and IRG-McS are performed in MATLAB 2016a running on a Windows desktop with Intel(R) Core(TM) i7-8700K CPU and 32GB of RAM. For the large scale dataset #2, M3CD, FPMS and NPSG all reduce the size of the images to be  $500 \times 500$  with a bilinear interpolation for reducing the computation. In Table VI,  $t_{sf}$  represents the computational time spent on the preprocessing (superpixel segmentation and feature extraction),  $t_{DI}$  and  $t_{cs}$  represent the time spent on the DI generation and MRF co-segmentation in each iteration respectively, and  $t_{total}$  represents the total computational time of IRG-McS. From Table VI, we can find that IRG-McS is very efficient compared with other method even with the dataset #2 without any downsampling, which is mainly because IRG-McS uses the superpixel as the basic analysis unit. Meanwhile, the two most time-consuming processes of IRG-McS are the DI generation and MRF co-segmentation.

For the former, calculating the distance matrices of  $\mathbf{D}^X$  and  $\mathbf{D}^Y$  requires  $\mathcal{O}((3C_X + 3C_Y)N_S^2/2)$ , and sorting the distance matrix by column to construct the robust adaptive KNN graph requires  $\mathcal{O}(N_S^2 \log N_S)$ . For the latter, calculating  $\mathcal{E}_d$  and  $\mathcal{E}_a$  requires  $\mathcal{O}(N_S)$  and  $\mathcal{O}(N_R)$  respectively, where  $N_R$  is the number of edges in the  $R$ -adjacency neighbor system. The complexity of min-cut/maxflow algorithm has been studied in [38], that is, the theoretical complexity of the worst-case is  $\mathcal{O}(2N_R N_S^2)$  and the empirical complexity is relatively low on typical problem instances in vision, which can also be seen in Table VI. In addition, for the large scale  $N_S$ , the DI generation process can be accelerated by using some efficient graph construction methods, such as the Kgraph<sup>6</sup> [42], whose empirical cost is around  $\mathcal{O}(n^{1.14})$  for constructing the KNN graph with a set of  $n$  nodes.

#### IV. CONCLUSION

In this paper, we mainly address the problem of unsupervised CD of heterogeneous RS images. The proposed IRG-McS method uses an iterative framework combining the DI generation and CM calculation processes, which can improve the quality of DI and detection accuracy of CM. In the DI generation process, IRG-McS exploits the inherent self-similarity property and detects the changes in heterogeneous images by using the structure consistency, which is represented by the constructed superpixel-based robust adaptive KNN graph. To avoid the leakage of heterogeneous data, IRG-McS calculates the structure differences between heterogeneous images by graph mapping, which is a comparison of similarity relationships. In the CM calculation process, a MRF co-segmentation model is designed to fuse the forward and backward DIs in the segmentation process, which can make full use of the change information and spatial information in DIs and the relationship between DIs. Once the changed areas are detected by the MRF co-segmentation, they are propagated back into the DI generation process to construct the robust KNN graph, which can reduce the influence of changed neighbors in the KNN graph. Experimental results clearly show that IRG-McS can effectively detect the changes in different heterogeneous data sets.

In this paper, we construct robust adaptive KNN graphs to represent the image structure, focusing on the choice of  $K$  and the nearest-neighbors, while we only use the Euclidean distance for the similarity metric of KNN graph and do not take into account the noise condition, such as the speckle noise [47] or stripe noise in the images [48]. Our future work is to design an appropriate similarity metric for the KNN graph under different noise environments and study the distribution

<sup>6</sup>kgraph: <http://www.kgraph.org/>

model of DI generated by KNN graph mapping on different heterogeneous image pairs, so as to design a more accurate segmentation model.

#### ACKNOWLEDGMENT

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

#### REFERENCES

- [1] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, Jun. 1989.
- [2] Y. Ban and O. A. Yousif, "Multitemporal spaceborne SAR data for urban change detection in China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1087–1094, Aug. 2012.
- [3] R. E. Kennedy *et al.*, "Remote sensing change detection tools for natural resource managers: Understanding concepts and tradeoffs in the design of landscape monitoring projects," *Remote Sens. Environ.*, vol. 113, no. 7, pp. 1382–1396, Jul. 2009.
- [4] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using VHR optical and SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2403–2420, May 2010.
- [5] Y. Sun, L. Lei, D. Guan, X. Li, and G. Kuang, "SAR image change detection based on nonlocal low-rank model and two-level clustering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 293–306, 2020.
- [6] Z. Y. Lv, T. F. Liu, P. Zhang, J. A. Benediktsson, T. Lei, and X. Zhang, "Novel adaptive histogram trend similarity approach for land cover change detection by using bitemporal very-high-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9554–9574, Dec. 2019.
- [7] C. Wu, B. Du, and L. Zhang, "Slow feature analysis for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2858–2874, May 2014.
- [8] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2019.
- [9] H. Chen, C. Wu, B. Du, L. Zhang, and L. Wang, "Change detection in multisource VHR images via deep siamese convolutional multiple-layers recurrent neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2848–2864, Apr. 2020.
- [10] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 799–812, Mar. 2015.
- [11] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "A Bayesian nonparametric model coupled with a Markov random field for change detection in heterogeneous remote sensing images," *SIAM J. Imag. Sci.*, vol. 9, no. 4, pp. 1889–1921, Jan. 2016.
- [12] Z. Liu, G. Li, G. Mercier, Y. He, and Q. Pan, "Change detection in heterogenous remote sensing images via homogeneous pixel transformation," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1822–1834, Apr. 2018.
- [13] Y. Qin, Z. Niu, F. Chen, B. Li, and Y. Ban, "Object-based land cover change detection for cross-sensor images," *Int. J. Remote Sens.*, vol. 34, no. 19, pp. 6723–6737, Jun. 2013.
- [14] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 545–559, Mar. 2018.
- [15] X. Niu, M. Gong, T. Zhan, and Y. Yang, "A conditional adversarial network for change detection in heterogeneous images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 45–49, Jan. 2019.
- [16] C. Wu, H. Chen, B. Do, and L. Zhang, "Unsupervised change detection in multi-temporal VHR images based on deep kernel PCA convolutional mapping network," 2019, *arXiv:1912.08628*. [Online]. Available: <http://arxiv.org/abs/1912.08628>
- [17] R. Touati, M. Mignotte, and M. Dahmane, "A reliable mixed-norm-based multiresolution change detector in heterogeneous remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3588–3601, Sep. 2019.
- [18] L. Wan, Y. Xiang, and H. You, "A post-classification comparison method for SAR and optical images change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1026–1030, Jul. 2019.
- [19] Y. Chang, M. Chen, L. Yan, X.-L. Zhao, Y. Li, and S. Zhong, "Toward universal stripe removal via wavelet-based deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2880–2897, Apr. 2020.
- [20] Y. Sun, L. Lei, X. Li, X. Tan, and G. Kuang, "Patch similarity graph matrix-based unsupervised remote sensing change detection with homogeneous and heterogeneous sensors," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 4841–4861, Jun. 2021.
- [21] W. Zhou, A. Troy, and M. Grove, "Object-based land cover classification and change analysis in the baltimore metropolitan area using multi-temporal high resolution remote sensing data," *Sensors*, vol. 8, no. 3, pp. 1613–1636, Mar. 2008.
- [22] L. Wan, Y. Xiang, and H. You, "An object-based hierarchical compound classification method for change detection in heterogeneous optical and SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9941–9959, Dec. 2019.
- [23] M. Volpi, D. Tuia, F. Bovolo, M. Kanevski, and L. Bruzzone, "Supervised change detection in VHR images using contextual information and support vector machines," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 20, pp. 77–85, Feb. 2013.
- [24] L. Wan, T. Zhang, and H. J. You, "Multi-sensor remote sensing image change detection based on sorted histograms," *Int. J. Remote Sens.*, vol. 39, no. 11, pp. 3753–3775, Jun. 2018.
- [25] C. Kwan, B. Ayhan, J. Larkin, L. Kwan, S. Bernabé, and A. Plaza, "Performance of change detection algorithms using heterogeneous images and extended multi-attribute profiles (EMAPs)," *Remote Sens.*, vol. 11, no. 20, p. 2377, Oct. 2019.
- [26] L. T. Luppino, F. M. Bianchi, G. Moser, and S. N. Anfinsen, "Unsupervised image regression for heterogeneous change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9960–9975, Dec. 2019.
- [27] W. Zhao, Z. Wang, M. Gong, and J. Liu, "Discriminative feature learning for unsupervised change detection in heterogeneous images based on a coupled neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7066–7080, Dec. 2017.
- [28] T. Zhan, M. Gong, X. Jiang, and S. Li, "Log-based transformation feature learning for change detection in heterogeneous images," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 9, pp. 1352–1356, Sep. 2018.
- [29] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
- [30] D. Guan, D. Xiang, X. Tang, and G. Kuang, "SAR image despeckling based on nonlocal low-rank regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3472–3489, Jun. 2019.
- [31] Y. Sun, L. Lei, X. Li, H. Sun, and G. Kuang, "Nonlocal patch similarity based heterogeneous remote sensing change detection," *Pattern Recognit.*, vol. 109, Jan. 2021, Art. no. 107598.
- [32] M. Mignotte, "A fractal projection and Markovian segmentation-based approach for multimodal change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8046–8058, Nov. 2020.
- [33] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2281, Nov. 2012.
- [34] C.-A. Deledalle, L. Denis, and F. Tupin, "How to compare noisy patches? Patch similarity beyond Gaussian noise," *Int. J. Comput. Vis.*, vol. 99, no. 1, pp. 86–102, Mar. 2012.
- [35] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.
- [36] R. Szeliski *et al.*, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1068–1080, Jun. 2008.
- [37] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
- [38] Y. Boykov and V. Kolmogorov, "An experimental comparison of mincut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.

- [39] R. Touati, M. Mignotte, and M. Dahmane, "Multimodal change detection in remote sensing images using an unsupervised pixel pairwise-based Markov random field model," *IEEE Trans. Image Process.*, vol. 29, pp. 757–767, Aug. 2020.
- [40] P. Zhang, M. Gong, L. Su, J. Liu, and Z. Li, "Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 24–41, Jun. 2016.
- [41] R. Touati, M. Mignotte, and M. Dahmane, "Change detection in heterogeneous remote sensing images based on an imaging modality-invariant MDS representation," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Athens, Greece, Oct. 2018, pp. 3998–4002.
- [42] L. Lei, Y. Sun, and G. Kuang, "Adaptive local structure consistency-based heterogeneous remote sensing change detection," *IEEE Geosci. Remote Sens. Lett.*, early access, Nov. 26, 2020, doi: [10.1109/LGRS.2020.3037930](https://doi.org/10.1109/LGRS.2020.3037930).
- [43] R. Touati, M. Mignotte, and M. Dahmane, "Anomaly feature learning for unsupervised change detection in heterogeneous images: A deep sparse residual model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 588–600, 2020.
- [44] Z.-G. Liu, G. Mercier, J. Dezert, and Q. Pan, "Change detection in heterogeneous remote sensing images based on multidimensional evidential reasoning," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 168–172, Jan. 2014.
- [45] R. Touati and M. Mignotte, "An energy-based model encoding nonlocal pairwise pixel interactions for multisensor change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 1046–1058, Feb. 2018.
- [46] W. Dong, C. Moses, and K. Li, "Efficient k-nearest neighbor graph construction for generic similarity measures," in *Proc. 20th Int. Conf. World Wide Web*, 2011, pp. 577–586.
- [47] Y. Chang, L. Yan, M. Chen, H. Fang, and S. Zhong, "Two-stage convolutional neural network for medical noise removal via image decomposition," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 2707–2721, Jun. 2020.
- [48] Y. Chang, L. Yan, H. Fang, S. Zhong, and W. Liao, "HSI-DeNet: Hyperspectral image restoration via convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 667–682, Feb. 2019.



**Yuli Sun** received the B.S. degree from Xiangtan University, Xiangtan, China, in 2009, and the M.S. degree from the University of Science and Technology of China, Hefei, China, in 2014. He is currently pursuing the Ph.D. degree in information and communication engineering with the College of Electronic Science, National University of Defense Technology. His research interests include machine learning, remote sensing image processing, and multitemporal image analysis.



**Lin Lei** received the Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2008.

She is currently a 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.



**Dongdong Guan** received the B.S. degree in surveying and mapping engineering from Wuhan University, Wuhan, China, in 2013, and the M.S. degree in photogrammetry and remote sensing and the Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2015 and 2019, respectively.

He currently works with the High-Tech Institute of Xi'an. His research interests include synthetic aperture radar image processing, machine learning, and computation vision in remote sensing applications.



**Gangyao Kuang** (Senior Member, IEEE) received the B.S. and M.S. degrees in geophysics from the Central South University of Technology, Changsha, China, in 1988 and 1991, respectively, and the Ph.D. degree in communication and information from the National University of Defense Technology, Changsha, in 1995.

He is currently a Professor with the 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.