

# Locality Preservation for Unsupervised Multimodal Change Detection in Remote Sensing Imagery

Yuli Sun<sup>1</sup>, Lin Lei<sup>1</sup>, Dongdong Guan, Gangyao Kuang, *Senior Member, IEEE*,  
Zhang Li, and Li Liu<sup>1</sup>, *Senior Member, IEEE*

**Abstract**—Multimodal change detection (MCD) is a topic of increasing interest in remote sensing. Due to different imaging mechanisms, the multimodal images cannot be directly compared to detect the changes. In this article, we explore the topological structure of multimodal images and construct the links between class relationships (same/different) and change labels (changed/unchanged) of pairwise superpixels, which are imaging modality-invariant. With these links, we formulate the MCD problem within a mathematical framework termed the locality-preserving energy model (LPEM), which is used to maintain the local consistency constraints embedded in the links; the structure consistency based on feature similarity and the label consistency based on spatial continuity. Because the foundation of LPEM, i.e., the links, is intuitively explainable and universal, the proposed method is very robust across different MCD situations. Noteworthy, LPEM is built directly on the label of each superpixel, so it is a paradigm that outputs the change map (CM) directly without the need to generate intermediate difference image (DI) as most previous algorithms have done. Experiments on different real datasets demonstrate the effectiveness of the proposed method. Source code of the proposed method is made available at <https://github.com/yulisun/LPEM>.

**Index Terms**—Energy, heterogeneous, locality preservation, multimodal change detection (MCD), topological structure.

## I. INTRODUCTION

### A. Background

A fundamental and challenging problem in the fields of computer vision and remote sensing, change detection aims at identifying changes of a particular area on the Earth surface over time by comparing multitemporal remote sensing imagery of the same geographical area taken at different times [1], [2]. It has received significant attention in recent

Manuscript received 23 April 2023; revised 29 November 2023 and 29 February 2024; accepted 7 May 2024. This work was supported in part by the Postdoctoral Fellowship Program of CPSF under Grant GZC20233545, in part by the Natural Science Foundation of Hunan Province of China under Grant 2024JJ6466, in part by the National Key Research and Development Program of China under Grant 2021YFB3100800, and in part by the National Natural Science Foundation of China under Grant 62376283. (*Corresponding author: Lin Lei*)

Yuli Sun and Zhang Li are with the College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China, and also with Hunan Provincial Key Laboratory of Image Measurement and Vision Navigation, Changsha 410073, China.

Lin Lei, Gangyao Kuang, and Li Liu are with the College of Electronic Science, National University of Defense Technology, Changsha 410073, China (e-mail: alaleilin@163.com).

Dongdong Guan is with the High-Tech Institute of Xi'an, Xi'an 710025, China.

Digital Object Identifier 10.1109/TNNLS.2024.3401696

years, due to the vital role it plays in a wide variety of applications including damage assessment (e.g., natural disasters such as floods, earthquakes, forest fires.), land management, urban development, and environment monitoring [3], [4], [5]. Especially with the rapid development of imaging techniques, high-resolution multitemporal remote sensing images at a large geographical scale can be acquired more easily and conveniently than before, enabling even more research opportunities for detecting and monitoring subtle changes in the Earth surface at a finer scale [6], [7], [8].

Homogeneous change detection, performing conventional change detection with images that come from the same sensor, has been widely investigated [9], [10], [11]. In other words, traditional change detection heavily relies on homogeneous data acquired by the same sensor, which has serious limitation for many realistic applications. In recent years, as an emerging and challenging problem, multimodal change detection (MCD) has received increasing attention. MCD deals with images that come from different sources, such as different types of sensors (e.g., synthetic aperture radar (SAR) and optical) and same sensor type but with different modalities (e.g., different spectral channels for optical sensors and different microwave frequencies or polarizations for radar instruments) [12], [13]. It is highly demanding to perform MCD that combines various Earth observation data, including multispectral and radar with different frequency bands and polarisations, etc. [14], [15]. One important application is in case of sudden events (such as floods and earthquakes), where MCD allows to use the first available images to assess the damages instead of waiting for collection of a homogeneous image pair.<sup>1</sup> This shortens the response time of the change analysis [16], [17]. Another use of MCD is to enable comparisons with old data acquired by outdated sensors [18], [19].

Despite its practical value, MCD poses challenges because multimodal images come from different data domains and have significantly different characteristics [20], [21], [22], such as SAR and optical images. This diversity makes direct comparisons infeasible, e.g., simple pixel differences commonly used in traditional monomodal change detection. Therefore, it becomes critical to reveal the connections between the multimodal images and transform them to a common domain in which they can be compared [23], [24]. To meet this challenge, several kinds of methodologies have been proposed,

<sup>1</sup>In such cases, homogeneous images may not be available due to the sensor's repeat cycle and adverse light and weather conditions.

which are related to the topics of data transformation and domain adaptation [25], [26].

### B. Related Work

Comparing with the monomodal change detection, there are relatively few works that have been devoted to MCD. To make the multitemporal images of  $X$  and  $Y$  comparable, the existing MCD methods can basically be generalized in terms of the sequential operations  $\text{DI} = \mathcal{M}_1(X) \ominus \mathcal{M}_2(Y)$  and  $\text{CM} = \mathcal{B}(\text{DI})$  [14], [27], which first calculate the difference image (DI) with the transform functions  $\mathcal{M}_1$  and  $\mathcal{M}_2$  and the pixelwise difference operator  $\ominus$ , and then compute the change map (CM) with the binary segmentation operator  $\mathcal{B}$ . Thus, previous approaches have been devoted to the design of  $\mathcal{M}_1$ ,  $\mathcal{M}_2$ ,  $\ominus$ , and  $\mathcal{B}$ .

According to whether the labels of training samples are required in the transformation process or not, the MCD methods can be classified as supervised and unsupervised. The supervised methods perform the transformation with the help of labeled samples [15], [20], [28], [29], [30], [31], [32], [33], while the unsupervised methods usually construct or learn a transformation that is independent of the imaging modality based on certain assumptions [26], [34], [35], [36], [37], [38], [39]. Currently, the unsupervised MCD methods are more attractive than supervised ones, and more applicable for real scenarios since the ground truth in Earth observation is very difficult to collect.

According to the nature of the transformed common domain, the MCD methods are based on postclassification comparison, feature transformation, or image regression. The classification comparison methods first transform the images into a common category space by taking  $\mathcal{M}_1$  and  $\mathcal{M}_2$  as classifiers, and then compare the classification results to detect changes [40], [41]. The feature transformation methods transform the images into a common constructed feature space [15], [23], [31] or latent learning feature space [42], [43], [44] as  $\mathcal{M}_1 : X \rightarrow Z$  and  $\mathcal{M}_2 : Y \rightarrow Z'$ . The image regression methods transform one image to the domain of the other image as  $\mathcal{M}_1 : X \rightarrow Y'$  or  $\mathcal{M}_2 : Y \rightarrow X'$ , such as the traditional image translation methods [24], [28] and the deep translation methods [26], [45], [46] with generative adversarial networks (GANs) [47] or cycle GAN [48].

According to the methods used for  $\mathcal{M}_1$  and  $\mathcal{M}_2$ , the MCD methods can be classified as: 1) traditional signal-processing-based methods, such as dictionary learning [49], [50], copular theory [20], multidimensional scaling [51], fractal projection [37], and graph processing [23]; and 2) deep-learning-based methods, some of which use convolutional neural networks (CNNs) to learn new representation [14], [44], [52], [53], and some use the GAN to accomplish domain adaptation [26], [45], [46]. However, the datasets for MCD are relatively limited until now, which is because constructing a ground-truth map that reflects real change information requires a high cost of manual operation and great expert knowledge in practice [54]. Although there are large datasets available for homogeneous change detection that can support pretrained models [55], [56], [57], there is currently no such dataset for MCD, and these deep-learning-based methods are mostly based on the pre-event and post-event images

themselves to detect the changes. Therefore, the traditional unsupervised MCD methods (e.g., the proposed LPEM) are still very appealing: first, they can quickly and automatically extract change information and second, they can provide assistance to deep-learning-based methods, such as constructing high-confidence pseudotraining sets [24], or supporting the training process [26].

Once the DI is obtained by comparing the transformed images of  $\text{DI} = \mathcal{M}_1(X) \ominus \mathcal{M}_2(Y)$ , the final CM solution can be treated as an image segmentation problem with the paradigm  $\text{CM} = \mathcal{B}(\text{DI})$ , which divides the DI into changed and unchanged classes, e.g., the thresholding methods such as Otsu threshold [58], or clustering methods such as the k-means clustering [59] and fuzzy c-means (FCM) clustering [60], or Markov random field (MRF)-based methods [23], [35], [37].

We can find that there are two important issues to cope with the problem of unsupervised MCD.

- 1) How to learn the robust transformations of  $\mathcal{M}_1$  and  $\mathcal{M}_2$ ? This requires that these transformations can bring the data to a common latent space where they can be efficiently compared, and that they are available for different MCD situations [37]. The transformations between multimodal images established by previous methods are generally based on certain assumptions (e.g., some imaging modality-invariant assumptions in traditional methods) or trained on pseudolabeled samples (e.g., some networks in deep-learning-based methods). These transformations may not be sufficiently stable nor universal, such as when the MCD scene is very complex (e.g., diversity of ground objects, difference in imaging conditions), the noise in image is severe (especially the speckle noise in SAR image), or the training samples are not sufficient or mixed with wrong samples.
- 2) How to suppress the influence of unknown change samples on the transformation? It includes the influence on the process of learning the transformation functions  $\mathcal{M}_1$  and  $\mathcal{M}_2$  and completing the transformations  $\mathcal{M}_1(X)$  and  $\mathcal{M}_2(Y)$  when they are not based on individual pixels [38]. This point is rarely mentioned by other studies for two reasons. First, it is partly due to the fact that this challenge is unique to MCD, i.e., it is not a problem in monomodal change detection that directly compares the images without the design of transformations [6]. Second, previous research usually treats the MCD as a two-step process, with the first step transforming images to the same domain to obtain the  $\text{DI} = \mathcal{M}_1(X) \ominus \mathcal{M}_2(Y)$ , and the second step segmenting the DI to obtain the final  $\text{CM} = \mathcal{B}(\text{DI})$  [14], [43]. Although this two-step strategy is intuitive, it carries a hidden risk of tending to ignore the influence of changes on the transformation process ( $\mathcal{M}_1$ ,  $\mathcal{M}_2$ ), i.e., the DI calculation will be affected by the change samples in the CM. Even though a few studies have focused on this point, they still require the use of a complex, coarse-to-fine iterative process to alleviate the influence [23].

### C. Motivation

Recently, some image-structure-based unsupervised MCD methods have been proposed [12], [23], [24], [26], [38], [61],

[62], which are based on the assumption that the topological structure of the multimodal images is the same in the unchanged regions and different in the changed regions. Luppino et al. [12], [24], [26] divide the images into overlapped patches and compute two affinity matrices to capture the structure within the patch. They then directly compare the affinity matrices to obtain the change prior, which can be further used to select the unchanged pseudosamples for traditional image regression [24], or assist the training process for deep image regression [12], [26]. Touati and Mignotte [51] calculate the DI using a nonlocal pairwise energy-based model (EBM), which encodes the pairwise pixel relationships between each pixel and all other pixels within the whole image, and then segments the DI by five different automatic thresholding algorithms to obtain a fused CM. In [62],  $k$ -nearest neighbor (KNN) graphs that connect each patch with its KNNs within the whole image are constructed to represent the structure of an image. Then, it compares the graphs within the same domain by means of graph mapping to calculate the DI, and then segments the DI to obtain the final CM. Although such structure comparison methods based on pairwise pixel relationships are intuitive and robust, they still suffer from a drawback: the influence of change samples on the structure comparison is ignored. Although an attempt to solve this problem has been made in our previous work [23], which propagates the CM obtained in the previous iteration to the next round of DI calculation to alleviate the change influence, it is still a complex, redundant iterative process. This prompts us to think about using a paradigm that outputs the CM directly without computing the intermediate variable of DI.

Touati et al. [35] have proposed a Markov random field model for MCD (named M3CD for short), which, to our knowledge, is the first energy model that directly outputs the CM for MCD. Based on three constraints on pixel pairs, M3CD constructs a visual cue to distinguish two pixels that belong to two different class labels (different pairwise labels) from two pixels who share the same label (identical pairwise label) and assumes that the visual cue obeys a Gaussian distribution or an exponential distribution in these respective cases. The attractiveness of M3CD lies in its use of pairwise constraints to construct an energy model for directly outputting the pixel label (change/unchanged), i.e., the CM.

Inspired by the pairwise relationship-based methods [24], [62] and the energy model [35], [51], and to address the challenges of unsupervised MCD, we propose a simple yet highly effective approach, named locality-preserving energy model (LPEM). LPEM extends the concept of detecting changes by measuring structural differences between images, which is achieved by comparing distances or affinities [23], [24]. Similar to methods proposed in [35] and [51], we convert MCD into an energy minimization problem based on the links between class relationships and change labels, but with quite different terms of constraint construction, energy building, metrics, structure representation, basic unit, model solution, etc., as discussed in the Appendix. For a superpixel pair at the same location in multimodal images of the same scene (without changes), the superpixels vary significantly and their statistical features cannot be directly compared due to different appearances, but the neighborhood similarity relationship among superpixels representing the topological structure of one image can be well-preserved by the other image due to

physical constraints. Based on this observation, we establish links between class relationships (same/different) and change labels (changed/unchanged) of pairwise superpixels,<sup>2</sup> which is more adequate and more accurate than in M3CD [35]. Since these links are intuitively explainable and universal, it ensures that our proposed method is highly robust and adaptable to a variety of different MCD situations. Furthermore, to use these links for detecting changes, LPEM implements three types of constraints: 1) a feature-similarity-based structure consistency that encodes the interactions between each superpixel and its spectral neighborhoods; 2) a spatial-continuity-based label consistency that encodes the interactions between each superpixel and its spatial neighborhoods; and 3) a typical prior constrain that the changes happen in small regions. The LPEM is built directly on the label (changed/unchanged) of each superpixel, so minimizing it will directly output the CM, without calculating the intermediate DI while avoiding the negative influence of changes.

#### D. Contribution

The main contributions are summarized as follows.

- 1) We construct robust, imaging modality-invariant connections between multimodal images by exploiting all the links between class relationships (same/different) and change labels (changed/unchanged) of pairwise superpixels. From these links, we extract six mappings as the criteria to identify superpixel labels for the MCD problem, which is widely available across different MCD situations.
- 2) We use an EBM to release the ability of detecting changes in the extracted criteria, which can directly compute the CM without calculating the intermediate DI, unlike the two-step process of most previous approaches (calculating DI and segmenting DI).
- 3) The proposed model is used for locality preservation defined as structure consistency in the feature space and label consistency in the geographic space between multimodal images. Moreover, by considering the label of each sample in the structure consistency-based constraint, the proposed model reduces the negative impact of change samples. The experimental results on five real datasets show that the proposed method obtains better performance than state-of-the-art (SOTA) methods.

## II. LINKS BETWEEN RELATIONSHIPS AND LABELS

Let us consider the bitemporal MCD problem and assume that two coregistered images of the same Earth region  $\mathbf{X} \in \mathbb{R}^{M \times N \times C_x}$  and  $\mathbf{Y} \in \mathbb{R}^{M \times N \times C_y}$  are given, where  $M \times N$  denotes the spatial size of the image, and  $C_x$  and  $C_y$  represent their number of image channels. The objective of MCD is to compute a binary CM  $\mathbf{B} \in \mathbb{R}^{M \times N}$  where each pixel label indicates changed or unchanged.

The core issue in tackling MCD lies in the difficulty in comparing two different physical quantities measured by the two sensors. In other words, direct comparison of two multimodal images in terms of simple pixelwise differencing,

<sup>2</sup>Here, the “pairwise” term comprises two aspects: first, pairs of superpixels in different regions on the same image and second, pairs of superpixels in the same region of different images, as illustrated by Fig. 1 and Table I.

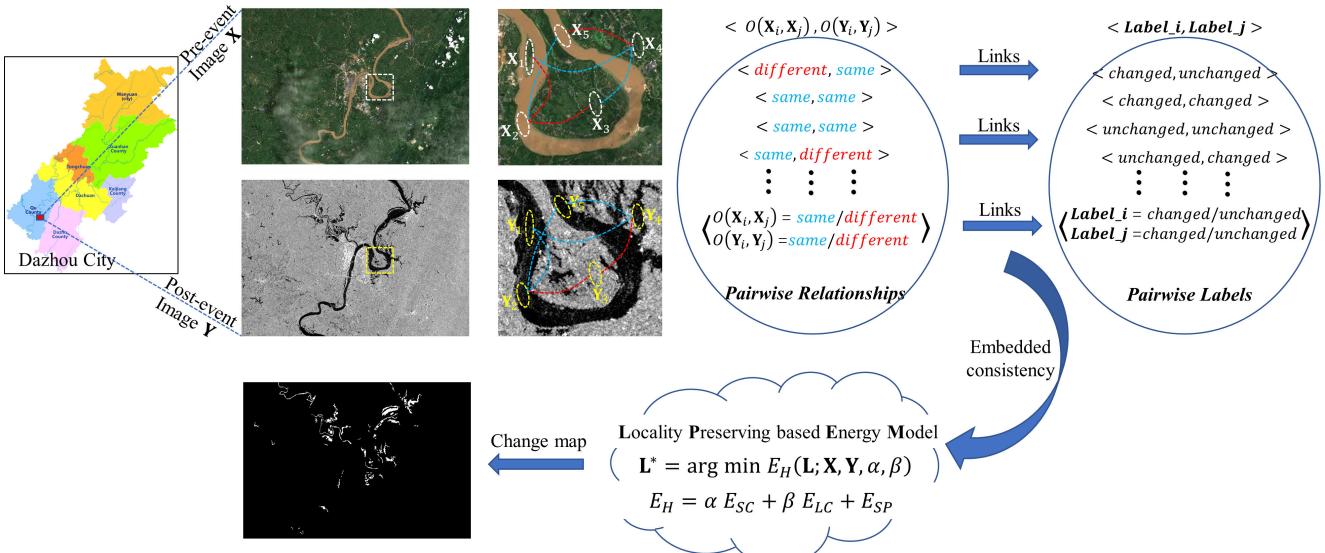


Fig. 1. Illustration of the LPEM. The two images show the same area in Qu, Dazhou, China. The pre-event optical image is collected by Sentinel-2 on 22 June, and the post-event SAR image is collected by Sentinel-1 on 11 July (data processed by ESA, <http://www.copernicus.eu>). These two images show the flooding caused by the dramatic rise in the level of Qujiang River as a result of heavy rainfall on 11 July 2021. In the pairwise relationships, “ $O(\mathbf{X}_i, \mathbf{X}_j) = \text{same}$ ” and “ $O(\mathbf{X}_i, \mathbf{X}_j) = \text{different}$ ” means that superpixels of  $\mathbf{X}_i$  and  $\mathbf{X}_j$  belong to the same kind and different kinds of objects, respectively. In the pairwise labels,  $\text{Label}_i$  and  $\text{Label}_j$  represent the labels of the  $i$ th and  $j$ th superpixels, respectively.

TABLE I  
LINKS BETWEEN CLASS RELATIONSHIPS AND CHANGE  
LABELS OF PAIRWISE SUPERPIXELS

Pairwise Relationships	Examples	Pairwise Labels
$<O(\mathbf{X}_i, \mathbf{X}_j), O(\mathbf{Y}_i, \mathbf{Y}_j)>$	$X_i, X_j \rightarrow Y_i, Y_j$	$<\text{Label}_i, \text{Label}_j>$
$<\text{same}, \text{same}>$	$\Delta \triangle \rightarrow \Delta \Delta$ $\Delta \triangle \rightarrow \circ \circ$	$<\text{unchanged}, \text{unchanged}>$
$<\text{same}, \text{different}>$	$\Delta \triangle \rightarrow \Delta \circ$ $\Delta \triangle \rightarrow \circ \Delta$ $\Delta \triangle \rightarrow \square \square$	$<\text{unchanged}, \text{changed}>$
$<\text{different}, \text{same}>$	$\square \circ \rightarrow \Delta \Delta$ $\square \circ \rightarrow \Delta \circ$ $\square \circ \rightarrow \square \Delta$	$<\text{changed}, \text{unchanged}>$
$<\text{different}, \text{different}>$	$\square \circ \rightarrow \square \circ$ $\square \circ \rightarrow \square \square$ $\square \circ \rightarrow \square \square$	$<\text{changed}, \text{changed}>$

which is commonly used for traditional monomodal change detection, is meaningless, as illustrated in Fig. 1. Therefore, we need to find a robust, transferable connection between the multimodal images. The structure consistency [23], [24], [62] is based on the self-similarity property of images, that is, for the multimodal images of the same region without any changes, although direct pixelwise difference cannot be applied here, the topological structure of the two images is the same. Specifically, if the pixels (or patches/superpixels) of  $x_1$  and  $x_2$  located at positions of  $(m_1, n_1)$  and  $(m_2, n_2)$ , respectively, in image  $\mathbf{X}$  belong to the same kind of object,<sup>3</sup> showing a very small difference of  $x_1 - x_2$ , then the pixels (or patches/superpixels) of  $y_1$  and  $y_2$  located at the same positions of  $(m_1, n_1)$  and  $(m_2, n_2)$ , respectively, in image  $\mathbf{Y}$  also represent the same kind of object, and showing that the difference of  $y_1 - y_2$  is also very small.

In this article, we define the similarity relationships between each superpixel and its KNNs within the whole image as the topological structure of this image. Then the other image with a different modality from the same scene should also conform to this topological structure, unless changes have

<sup>3</sup>In the object-based change detection approaches, the term “object” is usually defined as a group of pixels with homogeneous spectra/intensity and spatial continuity [6], [7], [40], [63].

occurred. The more straightforward explanation would be that the nonlocal similarity within the image itself could eliminate the discrepancy between images across different modalities.

#### A. Preprocessing

As aforementioned, we need to consider the pairwise relationships among all the pixels in an image. Obviously, we have quadratic time complexity that is proportional to the square of the number of pixels, i.e.,  $\mathcal{O}(M^2 N^2)$ . For computational efficiency, we choose superpixels as the basic analysis unit, instead of using individual pixels or rectangular image patches. As a result of perceptual grouping of pixels, superpixels encapsulate more information than pixels and align better with image edges than rectangular image patches. In addition, using superpixels can lead to significant speed-up of subsequent processing, as the number of superpixels of an image is substantially smaller than the number of pixels.

To obtain the cosegmented superpixels, each image is independently oversegmented with the simple linear iterative clustering (SLIC) method [64]. For different types of images, such as optical images with RGB bands, multispectral images, and SAR images, the SLIC algorithm can be modified to match the image statistics, such as in [38] and [65]. Then, the superpixel segmentation maps from  $\mathbf{X}$  and  $\mathbf{Y}$  are combined through the intersection operator to obtain the cosegmentation map  $\Lambda = \{\Lambda_i | i = 1, \dots, N_S\}$  which consists of  $N_S$  cosegmented superpixels, denoted as  $\mathbf{X}_i = \{x(m, n, c) | (m, n) \in \Lambda_i, c = 1, \dots, C_x\}$  and  $\mathbf{Y}_i = \{y(m, n, c) | (m, n) \in \Lambda_i, c = 1, \dots, C_y\}$ , respectively. Thus, the superpixels of  $\mathbf{X}_i$  or  $\mathbf{Y}_i$  represent the same geographical area, and the set of pixels inside each superpixel ( $\mathbf{X}_i$  or  $\mathbf{Y}_i$ ) in  $\Lambda$  have the property to be internally homogeneous simultaneously in  $\mathbf{X}$  and  $\mathbf{Y}$ .

#### B. Connections Between Multimodal Images

To investigate the topological structure of the images, we consider class relationships (same/different) between pairs

of superpixels  $\langle i, j \rangle$  in each of two images  $\mathbf{X}$  and  $\mathbf{Y}$  and use the relationships to identify the change labels (changed/unchanged) of superpixels  $\langle i, j \rangle$ .

We use  $O(\cdot, \cdot)$  to define the class relationships of pairwise superpixels: if the superpixels of  $\mathbf{X}_i$  and  $\mathbf{X}_j$  belong to the same kind of object, we define

$$O(\mathbf{X}_i, \mathbf{X}_j) = \text{same} \quad (1)$$

and if the superpixels of  $\mathbf{X}_i$  and  $\mathbf{X}_j$  belong to the different kinds of objects, we define

$$O(\mathbf{X}_i, \mathbf{X}_j) = \text{different}. \quad (2)$$

The  $O(\cdot, \cdot)$  will be determined by radiometric distances that can be computed with sensor specific distance measures adapted to the statistics of the sensor, under the assumption that superpixels from same/different objects or classes will have low/high distance that can be distinguished sufficiently well. Then, the pairwise relationships of  $\langle O(\mathbf{X}_i, \mathbf{X}_j), O(\mathbf{Y}_i, \mathbf{Y}_j) \rangle$  have a total of four combinations, as listed in the first column of Table I.

We let  $\text{Label}_i$  and  $\text{Label}_j$  represent the change labels of superpixels in the CM located at  $\Lambda_i$  and  $\Lambda_j$ , respectively, being either changed or unchanged. Then, the pairwise labels of  $\langle \text{Label}_i, \text{Label}_j \rangle$  also have a total of four combinations, as listed in the last column of Table I.

Although there are 16 links between  $\langle O(\mathbf{X}_i, \mathbf{X}_j), O(\mathbf{Y}_i, \mathbf{Y}_j) \rangle$  and  $\langle \text{Label}_i, \text{Label}_j \rangle$  when using exhaustive enumeration, four links are unrealistic in MCD problem. For example, if we have known that  $O(\mathbf{X}_i, \mathbf{X}_j) = \text{same}$  and  $O(\mathbf{Y}_i, \mathbf{Y}_j) = \text{same}$ , then the superpixels  $\langle i, j \rangle$  can only be labeled as both changed and unchanged. In Table I, we list all the possible links between the relationships of pairwise superpixels  $\langle \mathbf{X}_i, \mathbf{X}_j \rangle$ ,  $\langle \mathbf{Y}_i, \mathbf{Y}_j \rangle$  and the labels of  $\langle i, j \rangle$ . The symbols of  $\Delta$ ,  $\bigcirc$ ,  $\square$ ,  $\dagger$  in the second column of Table I are used as examples to illustrate different kinds of objects. It should be noted that Touati et al. [35], [51] have also established the connections between the relationship and labels of pairwise pixels. However, these are not as accurate and complete as Table I in this article. For example, they ignore the 5th, 8th, 9th, and 12th links of Table I and directly assume that the pairwise relationships of  $\langle \text{same}, \text{different} \rangle$ ,  $\langle \text{different}, \text{same} \rangle$ , and  $\langle \text{different}, \text{different} \rangle$  all correspond to the different pairwise labels of  $\langle \text{unchanged}, \text{changed} \rangle$ , or  $\langle \text{changed}, \text{unchanged} \rangle$ . Fig. 2 shows some examples of the links between the relationships and labels.

Because the task of MCD is to obtain the change label of superpixels, the links listed in Table I can provide us guidance. On one hand, we can find that these links are universal and constant, and no other assumptions are used, so it can embrace change detection in both homogeneous and different kinds of heterogeneous data. On the other hand, how to use these links to detect changes needs to be solved, i.e., which links are useful and how to turn these links in mathematical decision rules. Next, we use the EBM to address this problem, which can encode the pairwise superpixel interactions embedded in the links.

### III. LOCALITY-PRESERVING ENERGY MODEL

An EBM can capture dependencies between variables by associating a scalar energy to each configuration of the variables [66]. The attractiveness of EBM is that one is free to

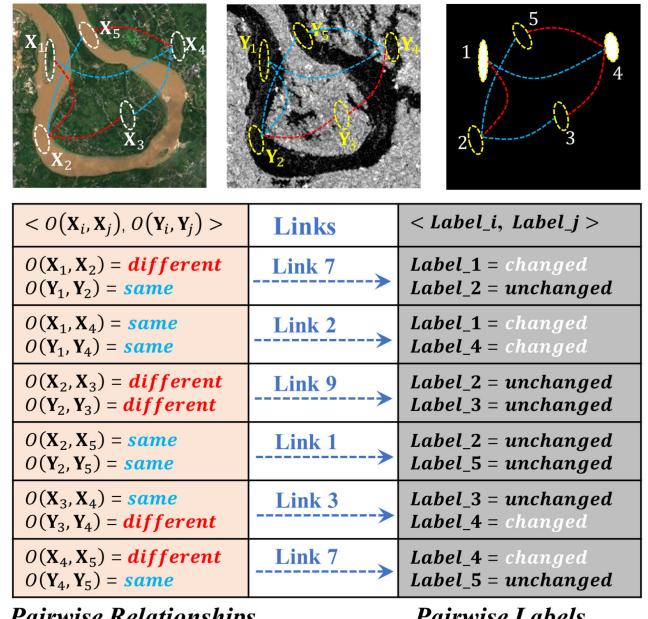


Fig. 2. Some examples of the links between the relationships and labels.

construct the energy in any reasonable way, thus giving it a great deal of flexibility and expressiveness [67], [68], [69], which has also been used for MCD [35], [51]. Here, we use the EBM to realize the potential value of links listed in Table I for MCD.

We define the index set as  $\mathcal{I} = \{1, 2, \dots, N_s\}$ , and the putative label set as an binary vector  $\mathbf{L} \in \mathbb{R}^{N_s}$ , where  $L_i \in \{0, 1\}$  represents the label of the  $i$ th superpixel. Specifically, we divide  $\mathcal{I}$  into the unchanged subset  $\mathcal{U} = \{i | L_i = 0, i \in \mathcal{I}\}$  and the changed subset  $\mathcal{C} = \{i | L_i = 1, i \in \mathcal{I}\}$ . Then, we convert the MCD task into an EBM optimization problem

$$\mathbf{L}^* = \arg \min_{\mathbf{L} \in \{0, 1\}^{\mathcal{I}}} E_H(\mathbf{L}; \mathbf{X}, \mathbf{Y}) \quad (3)$$

where the energy function  $E_H$  can measure the “goodness” of each possible configuration of  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{L}$  with the criteria extracted from the links.

#### A. From Links to Constraints

Although there are 12 different links in total in Table I, none of them is a one-to-one mapping from class relationships to change labels, i.e., each category of  $O(\mathbf{X}_i, \mathbf{X}_j)$ ,  $O(\mathbf{Y}_i, \mathbf{Y}_j)$  corresponds to more than one label of  $\langle \text{Label}_i, \text{Label}_j \rangle$ . For example, when  $O(\mathbf{X}_i, \mathbf{X}_j) = \text{same}$  and  $O(\mathbf{Y}_i, \mathbf{Y}_j) = \text{different}$ , there are three possible label pairs for the superpixels  $\langle i, j \rangle$ , namely,  $\langle \text{unchanged}, \text{changed} \rangle$ ,  $\langle \text{changed}, \text{unchanged} \rangle$ , and  $\langle \text{changed}, \text{changed} \rangle$ , as illustrated by Links 3–5 in Table I. Therefore, we need to add some restrictions to extract some one-to-one mappings that can help define an EBM to identify the superpixel labels.

For each superpixel in the image, we mainly consider the relationships between itself and its neighbors in feature space or geographical space, i.e., superpixels with feature similarity or spatial proximity. Accordingly, two types of pairwise relationship are studied: #1, each superpixel and its feature space neighbors and #2, each superpixel and its spatial neighbors, which can help extract some one-to-one mappings

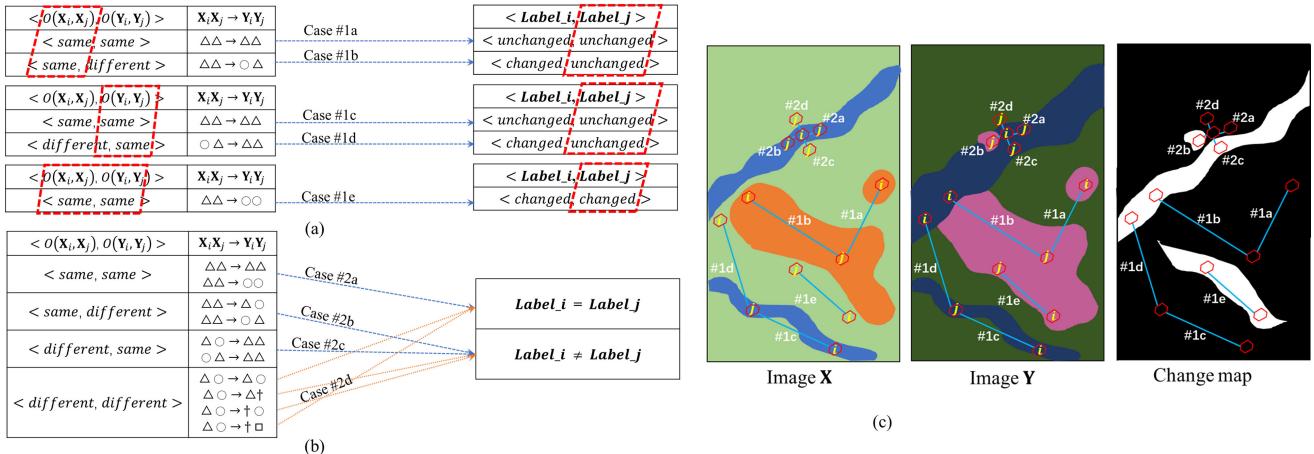


Fig. 3. Illustrations of the mappings in LPEM. The red dashed lines in (a) are the restrictions of each case. (a) Mappings of case #1. (b) Mappings of case #2. (c) Illustrations of case #1 and case #2.

to be used in formulating the EBM, as shown in cases #1 and #2 of Fig. 3.

Specifically, with these mappings, we can construct  $E_H$  with three types of constrains, i.e., the feature-similarity-based structure consistency, spatial-continuity-based label consistency, and change-prior-based sparse penalty, as follows:

$$E_H = \alpha E_{SC} + \beta E_{LC} + E_{SP} \quad (4)$$

where  $E_{SC}$ ,  $E_{LC}$ , and  $E_{SP}$  are the corresponding energy functions, respectively, and  $\alpha, \beta > 0$  are two balancing parameters. Next, we will present these criteria and  $E_{SC}$ ,  $E_{LC}$ , and  $E_{SP}$  in detail.

### B. Feature-Similarity-Based Structure Consistency

**Cases #1a and #1b.** In Table I, if we restrict that  $\mathbf{X}_i$  and  $\mathbf{X}_j$  belong to the same kind of objects (e.g.,  $\Delta$ ) and the  $j$ th superpixel is unchanged, we can obtain two one-to-one mappings: case #1a, the  $i$ th superpixel belongs to the unchanged class if  $\mathbf{Y}_i$  and  $\mathbf{Y}_j$  also belong to the same kind of objects (e.g.,  $\Delta$ ); case #1b, the  $i$ th superpixel belongs to the changed class if  $\mathbf{Y}_i$  and  $\mathbf{Y}_j$  belong to the different kinds of objects (e.g.,  $\bigcirc$  for  $\mathbf{Y}_i$  and  $\Delta$  for  $\mathbf{Y}_j$ ), as shown in Fig. 3.

To further enable these two mappings to help detect changes (i.e., to identify the label of the  $i$ th superpixel), we construct a function as

$$f_{i,j}^y = d_{i,j}^y - \hat{d}_i^y, \quad j \in \mathcal{N}_i^x \cap \mathcal{U} \quad (5)$$

where  $d_{i,j}^y$  represents the feature distance between superpixels  $\mathbf{Y}_i$  and  $\mathbf{Y}_j$  (such as  $d_{i,j}^y = \|\tilde{\mathbf{Y}}_i - \tilde{\mathbf{Y}}_j\|_2^2$  with  $\tilde{\mathbf{Y}}_i$  and  $\tilde{\mathbf{Y}}_j$  being the feature vectors). We define the index set of the KNN of  $\mathbf{Y}_i$  in  $\mathbf{Y}$  as  $\mathcal{N}_i^y$  by sorting the distance vector  $\{d_{i,j}^y | j \neq i, j \in \mathcal{I}\}$  and define  $\hat{d}_i^y = \max_{j' \in \mathcal{N}_i^y} (d_{i,j'}^y)$  as the maximum distance between  $\mathbf{Y}_i$  and its KNNs. Similarly, we can define  $d_{i,j}^x$ ,  $\hat{d}_i^x$  and  $\mathcal{N}_i^x$  for image  $\mathbf{X}$ .

In the function  $f_{i,j}^y$  of (5), the constraint of  $j \in \mathcal{N}_i^x \cap \mathcal{U}$  is used to satisfy the restrictions of cases #1a and #1b, that is,  $j \in \mathcal{N}_i^x$  is used to constrain that  $\mathbf{X}_i$  and  $\mathbf{X}_j$  belong to the same kind of objects, and  $j \in \mathcal{U}$  means that the  $j$ th superpixel is unchanged. The function  $f_{i,j}^y$  can be used to measure the likelihood that  $\mathbf{Y}_i$  and  $\mathbf{Y}_j$  belong to the same kind of objects:

if  $O(\mathbf{Y}_i, \mathbf{Y}_j) = \text{same}$ , then  $d_{i,j}^y$  will be small and we have a small  $f_{i,j}^y$ ; on the contrary, if  $O(\mathbf{Y}_i, \mathbf{Y}_j) = \text{different}$ , then  $d_{i,j}^y$  will be larger and we have a larger  $f_{i,j}^y$ . Therefore,  $f_{i,j}^y | j \in \mathcal{N}_i^x \cap \mathcal{U}$  of (5) can be used to measure the constraint power of cases #1a and #1b. It can also be found that case #1a makes use of Link 1 and case #1b makes use of Link 4 and Link 3 (just swapping  $i$  and  $j$ ) of Table I.

**Cases #1c and #1d.** On the other hand, if we restrict that  $\mathbf{Y}_i$  and  $\mathbf{Y}_j$  belong to the same kind of objects (e.g.,  $\Delta$ ) and the  $j$ th superpixel is unchanged in Table I, we can also obtain two one-to-one mappings: cases #1c and #1d, similar to cases #1a and #1b, respectively, as shown in Fig. 3.

Similarly, we can construct the function  $f_{i,j}^x$  to measure the constraint power of cases #1c and #1d as

$$f_{i,j}^x = d_{i,j}^x - \hat{d}_i^x, \quad j \in \mathcal{N}_i^y \cap \mathcal{U}. \quad (6)$$

It can also be found that case #1c makes use of Link 1 and case #1d makes use of Link 7 and Link 6 (swapping  $i$  and  $j$ ) of Table I.

**Case #1e.** Meanwhile, if we restrict that  $\mathbf{X}_i$  and  $\mathbf{X}_j$ ,  $\mathbf{Y}_i$  and  $\mathbf{Y}_j$  represent the same kind of objects, respectively (e.g.,  $\Delta$  for  $\mathbf{X}_i$  and  $\mathbf{X}_j$ ,  $\Delta$  for  $\mathbf{Y}_i$  and  $\mathbf{Y}_j$ ), and the  $j$ th superpixel is changed in Table I, we can obtain a one-to-one mapping: **case #1e**, the  $i$ th superpixel belongs to the changed class, as shown in Fig. 3.

We construct the function for case #1e as follows:

$$g_{i,j} = d_{i,j}^y - \hat{d}_i^y + d_{i,j}^x - \hat{d}_i^x, \quad j \in \mathcal{N}_i^x \cap \mathcal{N}_i^y \cap \mathcal{C} \quad (7)$$

where the constraint  $j \in \mathcal{N}_i^x \cap \mathcal{N}_i^y \cap \mathcal{C}$  is used to satisfy the restrictions of case #1e. The  $g_{i,j}$  of (7) can be used to maintain the mapping of case #1e, which makes use of Link 2 of Table I.

**1) Structure-Consistency-Based Energy:** By combining  $f_{i,j}^y$  of (5),  $f_{i,j}^x$  of (6), and  $g_{i,j}$  of (7), we have the structure-consistency-based energy function as

$$E_{SC} = \sum_{i \in \mathcal{U}} \left( \sum_{j \in \mathcal{N}_i^x \cap \mathcal{U}} f_{i,j}^y + \sum_{j \in \mathcal{N}_i^y \cap \mathcal{U}} f_{i,j}^x \right) + \sum_{i \in \mathcal{C}} \sum_{j \in \mathcal{N}_i^x \cap \mathcal{N}_i^y \cap \mathcal{C}} g_{i,j}. \quad (8)$$

With  $\mathcal{U} = \{i | L_i = 0, i \in \mathcal{I}\}$  and  $\mathcal{C} = \{i | L_i = 1, i \in \mathcal{I}\}$ ,  $E_{SC}$  can be rewritten as

$$E_{SC} = \sum_{i \in \mathcal{I}} (1 - L_i) \left( \sum_{j \in \mathcal{N}_i^x} (1 - L_j) f_{i,j}^y + \sum_{j \in \mathcal{N}_i^y} (1 - L_j) f_{i,j}^x \right) + \sum_{i \in \mathcal{I}} L_i \sum_{j \in \mathcal{N}_i^x \cap \mathcal{N}_i^y} L_j g_{i,j}. \quad (9)$$

### C. Spatial-Continuity-Based Label Consistency

Based on the reasonable assumption that spatially adjacent nodes may be very similar and thus should share the same label with high probability, the label-consistency-based energy is constructed by taking the contextual information into consideration, which can reduce the salt-and-pepper noise in the CM.

First, a spatially adjacent neighbor is defined for each superpixel: if two superpixels  $\langle i, j \rangle$  (located at  $\Lambda_i$  and  $\Lambda_j$ ) intersect or the spatial distance between their center points is less than  $R$ , superpixels  $\langle i, j \rangle$  are marked as spatially adjacent neighbors of each other, denoted as  $i \in \mathcal{N}_j^R$  or  $j \in \mathcal{N}_i^R$ . Because the average size of a superpixel generated by the SLIC segmentation averages  $MN/N_S$ , here we set  $R = 2(MN/N_S)^{1/2}$  for simplicity.

Second, for superpixels  $\langle i, j \rangle$  that are spatially adjacent, we drop two mappings of  $O(\mathbf{X}_i, \mathbf{X}_j) = \text{same}$  and  $O(\mathbf{Y}_i, \mathbf{Y}_j) = \text{different} \rightarrow \text{Label}_i = \text{changed}, \text{Label}_j = \text{changed}$  and  $O(\mathbf{X}_i, \mathbf{X}_j) = \text{different}, O(\mathbf{Y}_i, \mathbf{Y}_j) = \text{same} \rightarrow \text{Label}_i = \text{changed}, \text{Label}_j = \text{changed}$  listed in Table I (the fifth and eighth links), which has a small probability of appearing in practice. That is, it is not probable that adjacent superpixels belonging to the same type of objects at time  $t_1$  (e.g.,  $\Delta$ ) will change into two different objects at time  $t_2$ , respectively (e.g.,  $\dagger$  and  $\square$ ). With the spatial continuity, there are four possible cases (#2a–#2d) as shown in Fig. 3.

We construct a novel spatially adjacent-based energy using cases #2a–#2d, which not only takes into account the traditional spatial continuity but also the specificity of MCD task by considering the similarity relationships of original multi-temporal images. The label-consistency-based energy function  $E_{LC}$  is defined as follows:

$$E_{LC} = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{N}_i^R} \frac{\phi_{i,j}}{d(\Lambda_i, \Lambda_j)} \delta(L_i \neq L_j) \quad (10)$$

with  $d(\Lambda_i, \Lambda_j)$  being the Euclidean spatial distance between two superpixels,  $\delta(\cdot)$  being 1 if the specified condition inside parenthesis holds, and 0 otherwise, and  $\phi_{i,j}$  is defined as

$$\phi_{i,j} = \begin{cases} 1/2, & \text{if } d_{i,j}^x > \rho_1^2, d_{i,j}^y > \rho_2^2 \\ \sigma \left( \frac{2(d_{i,j}^x - \rho_1^2)(d_{i,j}^y - \rho_2^2)}{\rho_1^2 \rho_2^2} \right), & \text{else} \end{cases} \quad (11)$$

where the sigmoid function is defined as  $\sigma(z) = 1/(1 + e^{-z})$ , and the normalization parameters of  $\rho_1^2$  and  $\rho_2^2$  are set to be  $\rho_1^2 = (\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{N}_i^R} d_{i,j}^x / \sum_{i \in \mathcal{I}} |\mathcal{N}_i^R|)$  and  $\rho_2^2 = (\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{N}_i^R} d_{i,j}^y / \sum_{i \in \mathcal{I}} |\mathcal{N}_i^R|)$ , respectively, representing the average neighborhood feature distance over the whole image.

The function  $\phi_{i,j}$  gives a penalty for the discontinuity of  $L_i \neq L_j$  in four situations, corresponding to the four cases in Fig. 3.

*Case #2a:*  $O(\mathbf{X}_i, \mathbf{X}_j) = \text{same}$  and  $O(\mathbf{Y}_i, \mathbf{Y}_j) = \text{same}$ . In this case,  $\phi_{i,j}$  gives a large penalty for the discontinuity of  $L_i \neq L_j$ , and as  $d_{i,j}^x$  and  $d_{i,j}^y$  decreases, the discontinuity penalty is larger.

*Case #2b:*  $O(\mathbf{X}_i, \mathbf{X}_j) = \text{same}$  and  $O(\mathbf{Y}_i, \mathbf{Y}_j) = \text{different}$ . In this case,  $\phi_{i,j}$  gives a small penalty for the discontinuity of  $L_i \neq L_j$ , and as  $d_{i,j}^x$  decreases and  $d_{i,j}^y$  increases, the discontinuity penalty is smaller.

*Case #2c:*  $O(\mathbf{X}_i, \mathbf{X}_j) = \text{different}$  and  $O(\mathbf{Y}_i, \mathbf{Y}_j) = \text{same}$ . Similar to case #2b,  $\phi_{i,j}$  gives a small penalty for the discontinuity of  $L_i \neq L_j$  in this case, and as  $d_{i,j}^x$  increases and  $d_{i,j}^y$  decreases, the discontinuity penalty is smaller.

*Case #2d:*  $O(\mathbf{X}_i, \mathbf{X}_j) = \text{different}$  and  $O(\mathbf{Y}_i, \mathbf{Y}_j) = \text{different}$ . In this case, the  $i$ th and  $j$ th superpixels are likely to be located at the junction of two objects of different kinds. This means that the  $i$ th superpixel and the  $j$ th superpixel are not closely related to each other, i.e., the relationship between their labels is also ambiguous. Then,  $\phi_{i,j}$  gives a median discontinuity penalty for this case.

### D. Change-Prior-Based Sparse Penalty

Based on the fact that only a small part of the area changes and most of the area remains unchanged during the event in the practical change detection problem, we have the following sparse penalty  $E_{SP}$  as

$$E_{SP} = \|\mathbf{L}\|_0 = \sum_{i \in \mathcal{I}} L_i \quad (12)$$

which discourages changes and avoids trivial solution of  $\mathbf{L}^* = \mathbf{1}$ . Using the penalty  $E_{SP}$ , a sparse CM can be obtained which reduces the false alarms in the change detection results.

### E. Minimization of LPEM

By combining the energies of  $E_{SC}$ ,  $E_{LC}$ , and  $E_{SP}$ , we can find the unknown label set by solving

$$\mathbf{L}^* = \arg \min_{\mathbf{L} \in \{0, 1\}^{\mathcal{I}}} \{E_H(\mathbf{L}; \mathbf{X}, \mathbf{Y}, \alpha, \beta) := \alpha E_{SC} + \beta E_{LC} + E_{SP}\} \quad (13)$$

with the hybrid energy function  $E_H$  defined in (4). We further set

$$\alpha = \frac{\alpha^* N_S}{\sum_{i \in \mathcal{I}} \left( \sum_{j \in \mathcal{N}_i^x} f_{i,j}^y + \sum_{j \in \mathcal{N}_i^y} f_{i,j}^x \right)} \quad (14)$$

$$\beta = \frac{\beta^* N_S}{\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{N}_i^R} \phi_{i,j} / d(\Lambda_i, \Lambda_j)}$$

as  $\alpha^*$  and  $\beta^*$  are relatively easier to tune.

From the formulation of  $E_H$ , we can see that  $E_H$  encodes the interactions between each superpixel and its KNNs ( $j \in \mathcal{N}_i^x$  or  $j \in \mathcal{N}_i^y$ ) using  $E_{SC}$  of (9) and encodes the interactions between each superpixel and its spatial neighbors ( $j \in \mathcal{N}_i^R$ ) using  $E_{LC}$  of (10). This means that we have considered locality preservation of multimodal images in both feature space (structure consistency) and geographic space (label consistency), where the “locality” means the “neighborhood relationships

in feature space and geographic space.” At the same time, the influence of changed superpixels is also taken into account in the locality preservation-based constraint with  $j \in \mathcal{U}$  and  $j \in \mathcal{C}$  in the function of  $E_{SC}$ .

The energy minimization of (13) is an integer quadratic optimization problem. However, because  $E_H$  is a nonsubmodular energy function [70], [71], [72], the minimization problem (13) cannot be solved by the traditional graph cuts (such as min-cut/max-flow algorithm [73]), which are designed for minimizing submodular energy functions. To solve (13), the quadratic pseudo-Boolean optimization (QPBO) [72], [74] and the local submodular approximation (LSA)-based method [70] can be used, which iteratively uses nonlinear submodular approximations and optimizes them without leaving the domain of integer solutions.

Once the label set  $L^*$  is obtained by solving the model of (13) with LSA, we can compute the binary CM  $\mathbf{B}$  as

$$B(m, n) = L_i^*, \quad (m, n) \in \Lambda_i \quad (15)$$

where the unchanged areas are shown in black and changed areas are shown in white. The overall framework of the LPEM-based MCD method is shown in Fig. 4.

#### IV. EXPERIMENTS AND ANALYSIS

In this section, experiments are performed to evaluate the performance of LPEM. We first provide a brief description of the experimental setting, including the datasets, implementation details, and evaluation metrics. Following that, the experimental results are presented and analyzed. Finally, some discussions are given.

##### A. Experimental Setting

1) *Dataset*: Five multimodal datasets<sup>4</sup> are used to evaluate the performance of MCD methods as listed in Table II and shown in Fig. 5, which contain different types of heterogeneity: multisensor image pairs (same sensor type but with different sensors), e.g., images acquired from Landsat-5 and Advanced Land Imager (ALI) from the Earth Observing (EO-1) in dataset #1, and images acquired from Pleiades and WorldView2 in dataset #2; multsource image pairs (different sensor types), e.g., images obtained from Radarsat-2 and Google Earth in datasets #3 and #4, and images obtained from Landsat-8 and Sentinel-1A in dataset #5. These datasets reflect quite different MCD conditions: different resolution levels, different image sizes, and different change events, which can evaluate the generalizability and robustness of the proposed method.

2) *Implementation Detail*: The proposed LPEM is implemented with MATLAB 2016a. We set  $N_S = 5000$  for superpixel segmentation and choose the mean and median values of each band as the features to calculate the squared distances of  $d_{i,j}^x$  and  $d_{i,j}^y$  for simplicity (this is not exclusive, i.e., other features are also available). The neighborhood size  $K$  for each superpixel is set using an in-degree based  $k$ -selection strategy proposed in [38]. We set the balancing

<sup>4</sup>Dataset #1 is provided by Volpi et al. [15] and made available at <https://sites.google.com/site/michelevolpiresearch/cross-sensor>. Dataset #6 is provided by Luppino et al. [24] and made available at <https://sites.google.com/view/luppino>.

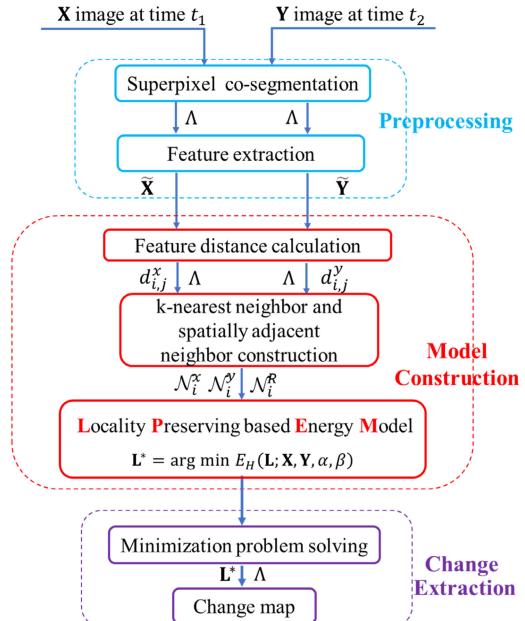


Fig. 4. Framework of the proposed LPEM-based MCD method.

parameters  $\beta^* = 5$  for all the datasets and adjust  $\alpha^*$  (set  $\alpha^* = 0.3$  for #3 and #5,  $\alpha^* = 0.5$  for #4,  $\alpha^* = 0.8$  for #1 and #2) in the hybrid energy function  $E_H$  of (13). These parameters will be analyzed in Section IV-C2.

3) *Metrics*: The binary change detection belongs to the binary classification task, so the overall accuracy (OA), F1 score, and Kappa coefficient (Kc) are used to evaluate the detection result, which are computed by:  $OA = (TP + TN)/(TP + TN + FP + FN)$ ,  $F1 = (2TP)/(2TP + FP + FN)$ , and  $Kc = (OA - PRE)/(1 - PRE)$  with

$$PRE = \frac{(TP + FN)(TP + FP) + (TN + FP)(TN + FN)}{(TP + TN + FP + FN)^2} \quad (16)$$

where TP, FP, TN, and FN represent the true positives, false positives, true negatives, and false negatives, respectively. Besides, we also provide the detection maps to evaluate their qualities by visual inspection.

##### B. Results

We compare the proposed LPEM-based method with ten SOTA methods,<sup>5</sup> including M3CD [35], FPMS [37], NPSG [61], IRG-McS [23], AGSCC [39], SCCN [14], ACE-Net [26], X-Net [26], CAAE [12], and ITSA [33], where the latter five methods are deep-learning-based and ITSA is a supervised method. For these comparison methods, the publicly available codes with default parameters were run to obtain the results. Quantitative results are presented in Table III, and visualization results are shown in Fig. 5.

1) *Qualitative Results*: Fig. 5 shows the CMs of different methods on all the evaluated datasets. We can find that some methods do not perform robustly enough and their performance degrades considerably on some datasets, such

<sup>5</sup>M3CD and FPMS are available at <http://www-labs.iro.umontreal.ca/~mignotte>. NPSG, IRG-McS, and AGSCC are available at <https://github.com/yulisun>. ACE-Net, X-Net, and CAAE are available at <https://github.com/llu025>. ITSA is available at <https://github.com/ImgSciGroup/ITSA>.

TABLE II  
DESCRIPTION OF THE FIVE MULTIMODAL DATASETS

Dataset	Sensor	Size (pixels)	Date	Location	Event (& Spatial resolution)
#1	Landsat-5/EO-1 ALI	1534 × 808 × 7(10)	Aug. 2011 - Sept. 2011	Texas, USA	Forest fire (30m.)
#2	Pleiades/WorldView2	2000 × 2000 × 3(3)	May 2012 - July 2013	Toulouse, France	Construction (0.52m.)
#3	RadarSat-2/Google Earth	593 × 921 × 1(3)	June 2008 - Sept. 2012	Shuguang Village, China	Building construction (8m.)
#4	RadarSat-2/Google Earth	495 × 503 × 1(3)	June 2008 - Nov. 2011	Wuhan, China	Building construction (8m.)
#5	Landsat-8/Sentinel-1A	875 × 500 × 11(3)	Jan. 2017 - Feb. 2017	Sutter County, USA	Flooding ( $\approx 15\text{m.}$ )

Fig. 5. Binary CMs of different methods on multimodal datasets. From top to bottom, they correspond to datasets #1–#5, respectively. From left to right are (a) pre-event image, (b) post-event image, (c) M3CD, (d) FPMS, (e) NPSG, (f) IRG-McS, (g) AGSCC, (h) SCCN, (i) ACE-Net, (j) X-Net, (k) CAAE, (l) ITSA, (m) LPEM, and (n) ground truth. In the binary CM, white: true positives (TPs), red: false positives (FP), black: true negatives (TN), and green: false negatives (FN).

TABLE III

QUANTITATIVE MEASURES OF BINARY CMs ON THE MULTIMODAL DATASETS. THE HIGHEST SCORE ARE HIGHLIGHTED IN BOLD

Methods	Dataset #1			Dataset #2			Dataset #3			Dataset #4			Dataset #5			Average		
	OA	Kc	F1	OA	Kc	F1	OA	Kc	F1									
M3CD [35]	0.883	0.044	0.075	0.863	0.405	0.481	0.962	0.602	0.622	0.934	0.476	0.511	0.575	0.021	0.077	0.843	0.310	0.353
FPMS [37]	0.830	0.019	0.112	0.838	0.215	0.296	0.938	0.569	0.597	0.944	0.610	0.640	0.947	0.329	0.356	0.899	0.348	0.400
NPSG [61]	0.913	0.657	0.703	0.830	0.346	0.446	0.975	0.729	0.742	0.953	0.566	0.590	0.941	0.419	0.449	0.922	0.543	0.586
IRG-McS [23]	0.912	0.448	0.493	0.882	0.420	0.478	0.983	0.794	0.804	0.948	0.553	0.581	0.959	0.490	0.512	0.937	0.541	0.574
AGSCC [39]	0.971	0.876	0.890	0.897	0.490	0.540	0.983	0.773	0.782	0.943	0.447	0.476	0.960	0.486	0.506	0.951	0.614	0.639
SCCN [14]	0.918	0.551	0.596	0.865	0.387	0.461	0.972	0.652	0.667	0.941	0.610	<b>0.641</b>	0.932	0.414	0.447	0.926	0.523	0.562
ACE-Net [26]	0.951	0.747	0.775	0.867	0.322	0.383	0.969	0.684	0.701	0.947	0.498	0.525	0.927	0.432	0.466	0.932	0.537	0.570
X-Net [26]	0.960	0.781	0.803	0.859	0.376	0.454	0.973	0.711	0.725	0.942	0.520	0.551	0.927	0.419	0.453	0.932	0.561	0.597
CAAЕ [12]	0.975	0.861	0.875	0.868	0.463	0.538	0.964	0.651	0.670	0.893	0.376	0.423	0.929	0.503	0.540	0.926	0.571	0.609
ITSA [33]	<b>0.990</b>	<b>0.951</b>	<b>0.956</b>	0.903	0.543	0.594	0.977	0.752	0.764	0.934	0.515	0.550	0.951	0.459	0.484	0.951	0.644	0.670
LPEM (proposed)	0.984	0.916	0.925	<b>0.917</b>	<b>0.702</b>	<b>0.752</b>	<b>0.986</b>	<b>0.835</b>	<b>0.842</b>	<b>0.957</b>	<b>0.616</b>	0.639	0.957	<b>0.503</b>	<b>0.526</b>	<b>0.960</b>	<b>0.714</b>	<b>0.737</b>

as M3CD, FPMS, and IRG-McS on dataset #1, FPMS and ACE-Net on dataset #2, and M3CD and SCCN on dataset #5. Conversely, the proposed LPEM can consistently achieve better results across different CD conditions. On the whole, the proposed LPEM can suppress the false positives and reduce the false negatives simultaneously, and it outperforms most comparison methods. These performances can be attributed to two main factors: 1) the imaging modality-invariant link-based locality-preserving constraint is robust to different MCD conditions, such as scenes, noises, and sensors; and 2) the negative influence of changes are taken into account in the structure-consistency-based energy  $E_{SC}$  in LPEM.

2) *Quantitative Results*: To further illustrate the superiority of LPEM, the quantitative measures of CMs are reported in

Table III. It can be seen that the proposed LPEM outperforms the SOTA methods on most datasets. For example, the LPEM achieves the highest OA, Kc, and F1 in datasets #2, and #3 (even comparing with the deep-learning-based ACE-Net, X-Net, CAAE, SCCN, and ITSA). The average OA, Kc, and F1 obtained by LPEM on all the evaluated datasets are about 0.960, 0.714, and 0.737, respectively. These scores are significantly higher than other methods, e.g., the average Kc and F1 are improved by 0.070 and 0.067, respectively, compared with the second-ranked supervised ITSA.

Finally, to further compare the performance of the proposed LPEM, the results obtained by some representative and SOTA methods [15], [17], [24], [44], [51], [75], [76], [77], [78], [79], [80], [81], [82], [83], [84], [85] are summarized in

TABLE IV

ACCURACY RATE OF CMs GENERATED BY DIFFERENT METHODS ON DIFFERENT DATASETS. THE RESULTS OF THESE COMPARISON METHODS ARE REPORTED BY THEIR ORIGINAL PUBLISHED ARTICLES, EXCEPT RESULTS INDICATED WITH  $\dagger$  ARE REPORTED BY [75] (THEY ARE CONSISTENT WITH THEIR OPEN SOURCE CODES IN [15]). ITALICIZED AND UNDERLINED MARKS ARE USED FOR DEEP-LEARNING-BASED METHODS

Dataset #1	OA	Dataset #3	OA	Dataset #5	OA
<i>DPFL</i> [76]	0.989	<i>DPFL</i> [76]	0.987	<i>GIR-MRF</i> [78]	0.959
<b>LPEM</b>	<b>0.984</b>	<b>LPEM</b>	<b>0.986</b>	<b>LPEM</b>	<b>0.957</b>
<i>AMD-IR</i> [24]	0.975	<i>AFL-DSR</i> [79]	0.980	<i>ALSC</i> [84]	0.945
<i>DCCAE</i> [77]	0.957	<i>CACFL</i> [44]	0.979	<i>PSGM</i> [81]	0.944
<i>DCCA</i> [75]	0.947	<i>RMN</i> [80]	0.877	<i>LT-FL</i> [83]	0.933
$\dagger$ <i>KCCA</i> [15]	0.915	<i>PUSM</i> [17]	0.865	<i>SSL</i> [85]	0.902
$\dagger$ <i>CCA</i> [15]	0.812	<i>NPPEM</i> [51]	0.853	<i>ALSC</i> [84]	
				<i>RMN</i> [80]	0.884

Table IV, except for the methods that have been compared in detail in Fig. 5 and Table III. Among these comparison approaches, *DCCAE* [77], *DCCA* [75], *AFL-DSR* [79], *DPFL* [76], *CACFL* [44], *LT-FL* [83], and *SSL* [85] are deep-learning-based methods. For the sake of fairness, we directly quote the results of the corresponding datasets in their original published articles in Table IV (because the datasets used in each article are not identical, Table IV is not aligned). As can be seen in Table IV, the LPEM consistently yields better or very competitive accuracy across different datasets by comparing with these SOTA approaches, which again demonstrates the effectiveness of the proposed locality preserving model. In addition, the reason why the proposed LPEM does not achieve the highest OA (but ranked second) on datasets #1, #3, and #5 may lie in the following aspects: first, the chosen weighting parameters in the hybrid energy function do not balance the energy components well and second, the superpixel feature distance metric is too simple to characterize the similarity relationships within the image in complex scenes.

### C. Discussions

1) *Difference Images*: As the proposed LPEM is built on the label of each superpixel, it can directly output the CM by solving the minimization problem of (13), without the need of calculating the intermediate DI as other MCD methods. However, it does not mean that the LPEM has no capability to output the DI. Sometimes the DI is also required, for example, a high-quality DI can be used to construct the pseudotraining set or assist the training process for some unsupervised deep-learning-based methods [26], [86].

Once the label set  $L^*$  is obtained by solving the model of (13), we can calculate the change level of each superpixel as

$$p_i = \sum_{j \in \mathcal{N}_i^x} (1 - L_j^*) f_{i,j}^y + \sum_{j \in \mathcal{N}_i^y} (1 - L_j^*) f_{i,j}^x. \quad (17)$$

The reverse process from calculating CM to calculating DI makes the DI more accurate due to the removal of the negative influence of changes. Fig. 6 shows the DIs calculated by LPEM on datasets #3 and #4. It can be found that the DIs are able to highlight the changes very well, which achieve the

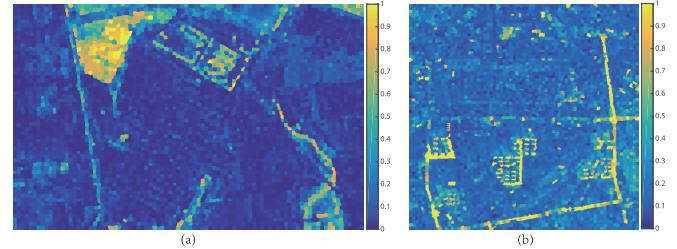


Fig. 6. DIs calculated by LPEM with (17) on (a) dataset #3 and (b) dataset #4.

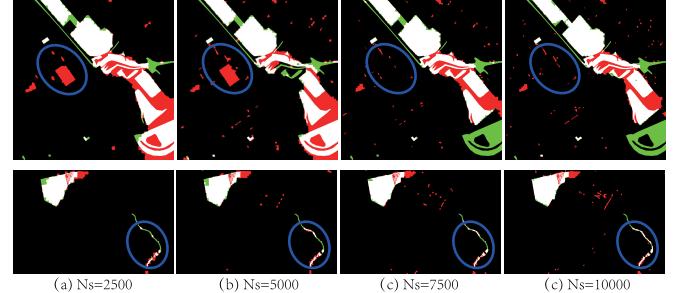


Fig. 7. CMs calculated by LPEM on datasets #2 (top row) and #3 (bottom row) with different  $N_S$ . The F1 scores of CMs on dataset #2 from (a) to (d) are 0.737, 0.741, 0.703, and 0.711, respectively; the F1 scores of CMs on dataset #3 from (a) to (d) are 0.811, 0.817, 0.837, and 0.832, respectively.

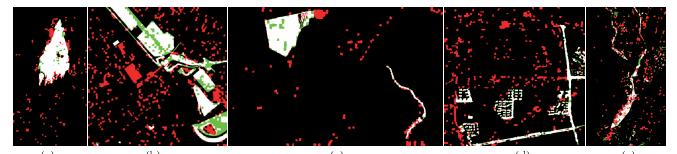


Fig. 8. CMs generated by LPEM without  $E_{LC}$ . From (a) to (e) are the results on datasets #1–#5, respectively.

areas under the receiver operating characteristic (ROC) curves equal to 0.968 and 0.916, respectively.

2) *Parameter Analysis*: The main parameters in LPEM are the number of superpixels  $N_S$  and the balancing parameters of  $\alpha^*$  and  $\beta^*$ .

Generally,  $N_S$  should be selected according to image resolution and granularity requirement of MCD task. A larger  $N_S$  will improve the detection granularity as the segmented superpixels are smaller. For example, when the detection task focuses on changes in land cover (such as crops, rivers, forests), a smaller  $N_S$  can be chosen; when the detection

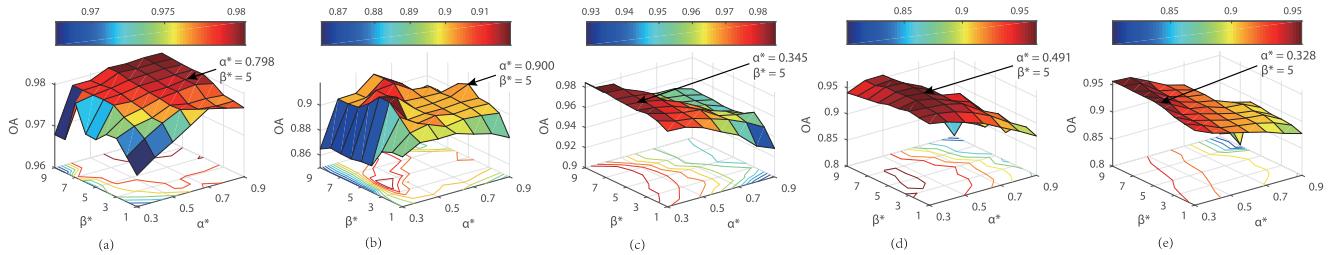


Fig. 9. Sensitivity analysis of parameters  $\alpha^*$  and  $\beta^*$  in LPEM on different datasets. The value identified by the arrow is  $\alpha^* = \max\{7.5 \times cr, 0.9\}$  and  $\beta^* = 5$ , where  $cr$  is the ratio of changed areas in ground truth. (a) Dataset #1. (b) Dataset #2. (c) Dataset #3. (d) Dataset #4. (d) Dataset #5.

task focuses on changes in targets with small size (such as vehicles), a larger  $N_S$  should be chosen. Fig. 7 shows the CMs generated by LPEM on datasets #2 and #3 with  $N_S = 2500, 5000, 7500$ , and  $10\,000$ , where some details are marked with the blue ellipse to fully compare these detection results. We can see that more details are detected as  $N_S$  increases. The average F1 scores of CMs generated by LPEM with different  $N_S$  (from  $2500$  to  $10\,000$ ) on the five evaluated datasets are  $0.707, 0.732, 0.722$ , and  $0.726$ , respectively. On the other hand, a larger  $N_S$  also increases the computational time as analyzed later. Therefore, we simply fix  $N_S = 5000$  as a compromise choice, which guarantees detection accuracy while allowing LPEM to be faster.

The parameters of  $\alpha^*$  and  $\beta^*$  are used to balance the weights of  $E_{SC}$ ,  $E_{LC}$  and  $E_{SP}$  in the LPEM. First, we perform the ablation analysis.

- 1) If we only focus on  $E_{SC}$  (i.e., set  $\alpha^* \rightarrow \infty$ ), the solution of (13) will be  $L^* = \mathbf{1}$ ; and if we remove  $E_{SC}$  (i.e., set  $\alpha^* \rightarrow 0$ ), it will be  $L^* = \mathbf{0}$  with  $E_H = 0$ .
- 2) If we only focus on  $E_{LC}$  (i.e.,  $\beta^* \rightarrow \infty$ ), the solution of (11) is  $L^* = \mathbf{0}$  or  $L^* = \mathbf{1}$ , and a larger  $\beta^*$  indicates a smoother CM. In Fig. 8, we show the CMs generated by LPEM without  $E_{LC}$ . By comparing Figs. 5(k) and 8, a lot of salt-and-pepper noise and false alarms appear in the CM when  $E_{LC}$  is removed, making the detection performance degrade.
- 3) If we only focus on  $E_{SP}$  (i.e., set  $\alpha^* \rightarrow 0$  and  $\beta^* \rightarrow 0$ ), we have the trivial solution of  $L^* = \mathbf{0}$ ; and if we remove  $E_{SP}$ , we have the trivial solution of  $L^* = \mathbf{1}$ .

Second, we show the performance of LPEM with varying  $\alpha^*$  and  $\beta^*$  in Fig. 9. We can find that LPEM can achieve good results for a certain range of both  $\alpha^*$  and  $\beta^*$ , which indicates that the algorithm is certain robust to the balance parameters. To further observe the effects of  $\alpha^*$  on the algorithm, we show the CMs of LPEM on dataset #4 by varying  $\alpha^*$  from  $0.3$  to  $0.8$  in Fig. 10. It can be found that  $\alpha^*$  acts like a threshold that controls the number of changes detected. Based on the compositional form of  $E_H$  and the results in Fig. 9, we can see that the value of  $\alpha^*$  should be related to the degree of real change. Here, we recommend setting the balancing parameters as  $\alpha^* = \max\{7.5 \times ecr, 0.9\}$  and  $\beta^* = 5$  in the hybrid energy function  $E_H$  of (13), where  $ecr$  represents the estimated ratio of changed areas, i.e., the estimated sparsity of the ground truth, as shown in Fig. 9. However, how to obtain the optimal  $\alpha^*$  and  $\beta^*$  without using  $ecr$  still needs to be studied, which is also our future work.

3) *Computational Time:* The main computational complexity of LPEM is concentrating on the processes of preprocessing, energy model construction, and energy model

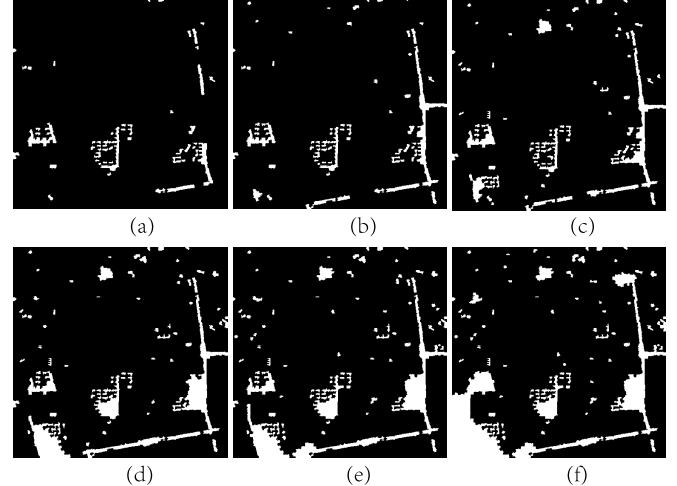


Fig. 10. CMs calculated by LPEM on dataset #4 with (a)  $\alpha^* = 0.3$ , (b)  $\alpha^* = 0.4$ , (c)  $\alpha^* = 0.5$ , (d)  $\alpha^* = 0.6$ , (e)  $\alpha^* = 0.7$ , and (f)  $\alpha^* = 0.8$ .

minimization. In the preprocessing, the complexity of SLIC method is linear in the number of pixels in the image, i.e.,  $\mathcal{O}(MN)$ , as reported in [64]. In the construction of energy function  $E_H$  of (6), computing the feature distances between all the superpixels requires  $\mathcal{O}((C_x + C_y)N_S^2)$ , and sorting the feature distances to find the KNN of all the superpixels requires  $\mathcal{O}(N_S^2 \log N_S)$ , and computing the spatial distance  $d(\Lambda_i, \Lambda_j)$  and  $\phi_{i,j}$  between adjacent superpixels in the  $E_{LC}$  of (12) requires  $\mathcal{O}(N_R)$  with  $N_R = \sum_{i \in \mathcal{I}} |\mathcal{N}_i^R|$ . In the LPEM minimization of (13), the LSA method for the optimization of binary nonsubmodular energy is used, whose running time for problems of different scales is presented in detail in [70]. Generally, the LPEM is relatively efficient, which is faster than the comparison traditional methods listed in Table III. For example, the computational times (seconds) on dataset #2 are<sup>6</sup>: M3CD: 2611.6; FPMS: 92.4; NPSG: 176.5; IRG-McS: 20.5; AGSCC: 46.5; and LPEM: 19.7, where the preprocessing takes 2.5 s, the energy model construction takes 3.1 s, and the LSA takes 14.1 s in LPEM.

## V. CONCLUSION

In this article, we propose an unsupervised MCD method by exploring the links between class relationships of pairwise superpixels and their change labels, which is intuitively explainable and universal across different MCD scenarios.

<sup>6</sup>LPEM, NPSG, IRG-McS, and AGSCC are performed in MATLAB 2016a running on a Windows desktop with Intel Core i7-8700K CPU, the C++ codes of M3CD and FPMS are executed in a Linux computer with Intel Xeon Silver 4110 CPU. Besides, M3CD, FPMS, and NPSG all downsample the images of dataset #2 by a factor of 4, as suggested by the authors, to ensure that the procedures are not interrupted by memory overflow errors.

Moreover, we extract the locality preservation property from the links and build an EBM by constructing three types of constraints. It is a paradigm for MCD that can output the CM directly by solving an energy minimization problem, without the need to generate intermediate DI as most previous algorithms have done, while overcoming the disadvantage that other algorithms tend to ignore the influence of changes on MCD. The experimental results demonstrate that our method achieves excellent performance for MCD with various scenarios.

In this article, we use the fixed KNN graph to characterize the topological structure of image, which may have limitations in some scenarios where the image structure is very complex. We manually select the weighting parameters to balance the energies, which is not robust enough. A future work would be to use the adaptive graph and incorporate geometrical structure information to make the EBM more accurate and use adaptive balancing parameters to improve the robustness of the proposed method. We hope that the proposed simple approach will inspire a rethinking of MCD, especially as deep learning methods are to be used systematically. Moreover, since the links described in this article are widely available between multimodal images, we believe that the proposed method may also be useful for other fields, such as detecting lesions using multimodal medical images (e.g., CT and MRI).

## APPENDIX

### DISCUSSIONS WITH OTHER ENERGY-BASED MCD METHODS

In the appendix, we will discuss in detail the connections and differences between the energy model proposed in this article and previous energy models, to facilitate further understanding of these energy-based MCD methods by interested readers.

Touati et al. [35], [51] have exploited energy models for the MCD task, which construct three constraints (#1–#3) for two distinct pixels  $i, j$  using the following connections between pairwise pixel relationships and label:

- #1,  $\langle \text{same}, \text{same} \rangle \rightarrow \text{Label}_i = \text{Label}_j$ ;
- #2,  $\langle \text{same}, \text{different} \rangle \rightarrow \text{Label}_i \neq \text{Label}_j$ ;  
 $\langle \text{different}, \text{same} \rangle \rightarrow \text{Label}_i \neq \text{Label}_j$ ;
- #3,  $\langle \text{different}, \text{different} \rangle \rightarrow \text{Label}_i \neq \text{Label}_j$ .

Then, a visual cue  $\eta_{(i,j)}$  is proposed for each pixel pairs  $(i, j)$  based on these connections as

$$\eta_{(i,j)} = \left| d_{i,j}^x - d_{i,j}^y \right| \quad (18)$$

where  $d_{i,j}^x$  and  $d_{i,j}^y$  are distance measures that are defined as  $d_{i,j}^x = \max\{|x_i - x_j|/x_i|, |x_i - x_j|/x_j|\}$  in [51] and  $d_{i,j}^x = \|\mathbf{x}_i - \mathbf{x}_j\|_1$  in [35] with  $x_i$  denoting the gray level and  $\mathbf{x}_i$  denoting a local statistics vector at pixel  $i$  in the pre-event image.

With this visual cue, a nonlocal pixel pairwise EBM (NPPEM) is proposed in [51] to calculate the similarity map  $\hat{S}$  (can be regarded as DI)

$$\hat{S} = \arg \min_{\mathbf{S}} \sum_{(i,j)_{i \neq j}} (\eta_{(i,j)} - \|S_i - S_j\|_2)^2. \quad (19)$$

Finally,  $\hat{S}$  is segmented using different thresholding algorithms to obtain a fused CM.

In M3CD [35], the visual cue  $\eta_{(i,j)}$  (18) is assumed to obey a Gaussian distribution in the case of identical pixel-pairwise label of  $L_i = L_j$ , and an exponential distribution in the case of different pixel-pairwise label of  $L_i \neq L_j$ . Then, M3CD builds an energy model as

$$E(\mathbf{L}, \eta) = \sum_{i=1}^{MN} \sum_{j \in \mathcal{G}_i} -\ln P(\eta_{(i,j)} | L_{(i,j)}) + \sum_{j \in \mathcal{E}_i} \beta \delta(L_i \neq L_j) \quad (20)$$

and obtain the pixel label by  $\hat{L} = \operatorname{argmin}_{\mathbf{L}} E(\mathbf{L}, \eta)$ , where  $L_{(i,j)}$  represents the above two cases of  $L_i = L_j$  and  $L_i \neq L_j$ .

By comparing NPPEM [51], M3CD [35], and the proposed LPEM, it can be found that there are two common aspects between them: they all use the pairwise relationship and pairwise labels of the multitemporal images and build energy models for the MCD problem. However, by deeply analyzing these methods, we can find that LPEM differs from the NPPEM and M3CD in the following.

1) Constraints construction: Both NPPEM and M3CD are based on constraints #1–#3, which is a subset of the 12 links in Table I. First, they ignore the Links 5, 8, 9, and 12 of Table I and make constraints #2–#3 not accurate enough. Second, NPPEM and M3CD only consider the pairwise labels of “identical” ( $L_i = L_j$ ) and “different” ( $L_i \neq L_j$ ) in the constraints, so the models of (19) and (20) have two opposite optimal solutions. However, LPEM not only considers the identical/different pairwise labels in case #2 of Table I using the spatial continuity but also considers the individual labels of changed/unchanged in case #1 of Table I using the structure consistency.

2) Energy model: NPPEM builds energy model (19) with a set of  $MN(MN-1)/2$  equality constraints that require that the distance between pixels  $i, j$  in  $\hat{S}$  is close to the visual cue  $\eta_{(i,j)}$ , which is similar to multidimensional scaling [87]. M3CD builds the MRF model (20) by assuming the distributions of the visual cue  $\eta_{(i,j)}$ , whose solution process consists of a parameter estimation step for the likelihood distributions and a segmentation step for the maximum a posteriori (MAP) solution of  $\hat{L}$ . However, the LPEM directly attributes energy to each constraint in a reasonable way as (9), (10), and (12), which takes full advantage of the EBM model’s flexibility and expressiveness. Moreover, the energy model built from the links can release the potential meaning they embed: locality preservation between multimodal images.

3) Metrics: Both NPPEM and M3CD calculate the visual cue  $\eta_{(i,j)}$  that measures the difference in pairwise relationships (i.e., structure difference) using (18), which directly compares the distances calculated in different domains as  $d_{i,j}^x - d_{i,j}^y$ . To make the distances comparable and to reduce the confusion, NPPEM and M3CD first use the gray level (or a local statistics vector), and then NPPEM uses the normalized distance metric and M3CD uses a double histogram matching for the gray images, respectively. However, LPEM measures the difference in pairwise relationships using  $f_{i,j}^y$  (5) and  $f_{i,j}^x$  (6), which compares the distances in the same domain. For example, for the  $(i, j)$ th pixels or superpixels known to belong to the same kind of objects in the pre-event image  $\mathbf{X}$ , to determine whether they also belong to the same kind of objects in the post-event image  $\mathbf{Y}$ , NPPEM and M3CD calculate the

difference between  $d_{i,j}^x$  and  $d_{i,j}^y$ , while LPEM calculates  $f_{i,j}^y$  by comparing  $d_{i,j}^y$  with  $\max_{j' \in \mathcal{N}_i^y}(d_{i,j'}^y)$  in (5). Intuitively, LPEM measures the structure difference by how different the two KNN position sets of  $\mathcal{N}_i^x$  and  $\mathcal{N}_i^y$  are in the same domain. In our previous article [61], [62], we have compared the direct distance comparison with the mapped KNN comparison.

4) Structure representation: NPPEM considers the pixel pairwise relationships between each pixel and all other pixels with (19), and M3CD uses the pairwise relationships between the  $i$ th pixel and a subsample of 8 pixels regularly distributed around a squared window of  $\mathcal{G}_i$  ( $41 \times 41$  in [35]) with (20). However, LPEM constructs the KNN graph that uses the relationships between each superpixel and its KNNs within the whole image as the structure, which brings two benefits: first, it serves as a constraint in case #1 of Fig. 3, and second, it allows to measure structural difference in the same domain.

5) There are other differences, such as the basic processing unit and the solution method of the energy model. For example, NPPEM and M3CD are pixel-pairwise-based, while LPEM is superpixel-pairwise-based, which greatly reduces the computational complexity.

#### ACKNOWLEDGMENT

The authors would 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] L. Bruzzone and F. Bovolo, "A novel framework for the design of change-detection systems for very-high-resolution remote sensing images," *Proc. IEEE*, vol. 101, no. 3, pp. 609–630, Mar. 2013.
- [3] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogramm. Remote Sens.*, vol. 80, pp. 91–106, Jun. 2013.
- [4] M. Yang, L. Jiao, B. Hou, F. Liu, and S. Yang, "Selective adversarial adaptation-based cross-scene change detection framework in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2188–2203, Mar. 2021.
- [5] M. Zhang, T. Gao, M. Gong, S. Zhu, Y. Wu, and H. Li, "Semisupervised change detection based on bipartitional feature aggregation and extraction network," *IEEE Trans. Neural Netw. Learn. Syst.*, 2023, doi: 10.1109/TNNLS.2023.3242075.
- [6] Z. Lv, T. Liu, J. A. Benediktsson, and N. Falco, "Land cover change detection techniques: Very-high-resolution optical images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 44–63, Mar. 2022.
- [7] D. Wen et al., "Change detection from very-high-spatial-resolution optical remote sensing images: Methods, applications, and future directions," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 68–101, Dec. 2021.
- [8] H. Zhang, M. Lin, G. Yang, and L. Zhang, "ESCNet: An end-to-end superpixel-enhanced change detection network for very-high-resolution remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 1, pp. 28–42, Jan. 2023.
- [9] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 125–138, Jan. 2016.
- [10] F. Liu, L. Jiao, X. Tang, S. Yang, W. Ma, and B. Hou, "Local restricted convolutional neural network for change detection in polarimetric SAR images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 818–833, Mar. 2019.
- [11] M. Yang, L. Jiao, F. Liu, B. Hou, and S. Yang, "Transferred deep learning-based change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6960–6973, Sep. 2019.
- [12] L. T. Luppino et al., "Code-aligned autoencoders for unsupervised change detection in multimodal remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 1, pp. 60–72, Jan. 2024, doi: 10.1109/TNNLS.2022.3172183.
- [13] Z. Lv et al., "Land cover change detection with heterogeneous remote sensing images: Review, progress, and perspective," *Proc. IEEE*, vol. 110, no. 12, pp. 1976–1991, Dec. 2022.
- [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] M. Volpi, G. Camps-Valls, and D. Tuia, "Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis," *ISPRS J. Photogramm. Remote Sens.*, vol. 107, pp. 50–63, Sep. 2015.
- [16] 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.
- [17] R. Touati, "Détection de changement en imagerie satellitaire multimodale," Ph.D. dissertation, Département d'informatique et de recherche opérationnelle, Montreal, QC, Canada, 2019.
- [18] P. Ghamisi et al., "Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 1, pp. 6–39, Mar. 2019.
- [19] F. Bovolo and L. Bruzzone, "The time variable in data fusion: A change detection perspective," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 8–26, Sep. 2015.
- [20] G. Mercier, G. Moser, and S. B. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1428–1441, May 2008.
- [21] T. G. Rudner et al., "Multi3Net: Segmenting flooded buildings via fusion of multiscale, multisensor, and multitemporal satellite imagery," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 702–709.
- [22] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, 2019.
- [23] Y. Sun, L. Lei, D. Guan, and G. Kuang, "Iterative robust graph for unsupervised change detection of heterogeneous remote sensing images," *IEEE Trans. Image Process.*, vol. 30, pp. 6277–6291, 2021.
- [24] 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.
- [25] G. Moser, S. N. Anfinsen, L. T. Luppino, and S. B. Serpico, "Change detection with heterogeneous remote sensing data: From semi-parametric regression to deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2020, pp. 3892–3895.
- [26] L. T. Luppino et al., "Deep image translation with an affinity-based change prior for unsupervised multimodal change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4700422.
- [27] J. Liu, M. Gong, A. K. Qin, and K. C. Tan, "Bipartite differential neural network for unsupervised image change detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 3, pp. 876–890, Mar. 2020.
- [28] Z. Liu, G. Li, G. Mercier, Y. He, and Q. Pan, "Change detection in heterogeneous remote sensing images via homogeneous pixel transformation," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1822–1834, Apr. 2018.
- [29] K. Jiang, J. Liu, W. Zhang, F. Liu, and L. Xiao, "MANet: An efficient multidimensional attention-aggregated network for remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4706118.
- [30] K. Jiang, W. Zhang, J. Liu, F. Liu, and L. Xiao, "Joint variation learning of fusion and difference features for change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4709918.
- [31] 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.
- [32] 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, 2016.

- [33] Z. Lv, H. Huang, W. Sun, M. Jia, J. A. Benediktsson, and F. Chen, "Iterative training sample augmentation for enhancing land cover change detection performance with deep learning neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 21, 2023, doi: 10.1109/TNNLS.2023.3282935.
- [34] V. Ferraris, N. Dobigeon, and M. Chabert, "Robust fusion algorithms for unsupervised change detection between multi-band optical images—A comprehensive case study," *Inf. Fusion*, vol. 64, pp. 293–317, Dec. 2020.
- [35] 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, 2020.
- [36] D. Marcos, R. Hamid, and D. Tuia, "Geospatial correspondences for multimodal registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5091–5100.
- [37] 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.
- [38] Y. Sun, L. Lei, D. Guan, M. Li, and G. Kuang, "Sparse-constrained adaptive structure consistency-based unsupervised image regression for heterogeneous remote-sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4405814.
- [39] Y. Sun, L. Lei, D. Guan, J. Wu, G. Kuang, and L. Liu, "Image regression with structure cycle consistency for heterogeneous change detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 1613–1627, Feb. 2024, doi: 10.1109/TNNLS.2022.3184414.
- [40] 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.
- [41] Y. Wu, Z. Bai, Q. Miao, W. Ma, Y. Yang, and M. Gong, "A classified adversarial network for multi-spectral remote sensing image change detection," *Remote Sens.*, vol. 12, no. 13, p. 2098, Jun. 2020.
- [42] 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.
- [43] J. Liu, W. Zhang, F. Liu, and L. Xiao, "A probabilistic model based on bipartite convolutional neural network for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4701514.
- [44] Y. Wu, J. Li, Y. Yuan, A. K. Qin, Q.-G. Miao, and M.-G. Gong, "Commonality autoencoder: Learning common features for change detection from heterogeneous images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4257–4270, Sep. 2022.
- [45] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 179, pp. 14–34, Sep. 2021.
- [46] Z.-G. Liu, Z.-W. Zhang, Q. Pan, and L.-B. Ning, "Unsupervised change detection from heterogeneous data based on image translation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4403413.
- [47] I. J. Goodfellow et al., "Generative adversarial nets," in *Proc. NeurIPS*, 2014, pp. 2672–2680.
- [48] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2223–2232.
- [49] M. Gong, P. Zhang, L. Su, and J. Liu, "Coupled dictionary learning for change detection from multisource data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7077–7091, Dec. 2016.
- [50] V. Ferraris, N. Dobigeon, Y. Cavalcanti, T. Oberlin, and M. Chabert, "Coupled dictionary learning for unsupervised change detection between multimodal remote sensing images," *Comput. Vis. Image Understand.*, vol. 189, Dec. 2019, Art. no. 102817.
- [51] 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.
- [52] T. Zhan, M. Gong, J. Liu, and P. Zhang, "Iterative feature mapping network for detecting multiple changes in multi-source remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 146, pp. 38–51, Dec. 2018.
- [53] 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.
- [54] Z. Zheng, A. Ma, L. Zhang, and Y. Zhong, "Change is everywhere: Single-temporal supervised object change detection in remote sensing imagery," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 15193–15202.
- [55] R. Gupta et al., "Creating xBD: A dataset for assessing building damage from satellite imagery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPR)*, Jun. 2019, pp. 10–17.
- [56] L. Shen et al., "S2Looking: A satellite side-looking dataset for building change detection," 2021, *arXiv:2107.09244*.
- [57] D. Peng, L. Bruzzone, Y. Zhang, H. Guan, H. Ding, and X. Huang, "SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5891–5906, Jul. 2021.
- [58] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jun. 1979.
- [59] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-means clustering algorithm," *J. Roy. Stat. Soc. Ser. C*, vol. 28, no. 1, pp. 100–108, May 1979.
- [60] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy C-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, Jan. 1984.
- [61] 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.
- [62] Y. Sun, L. Lei, X. Li, X. Tan, and G. Kuang, "Structure consistency-based graph for unsupervised change detection with homogeneous and heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4700221.
- [63] Y. Ban and O. Yousif, *Change Detection Techniques: A Review*. Cham, Switzerland: Springer, 2016, pp. 19–43.
- [64] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [65] 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, Aug. 2012.
- [66] Y. LeCun, S. Chopra, R. Hadsell, M. A. Ranzato, and F. J. Huang, "A tutorial on energy-based learning," in *Predicting Structured Data*. MA, USA: MIT Press, 2006.
- [67] H. Isack and Y. Boykov, "Energy-based geometric multi-model fitting," *Int. J. Comput. Vis.*, vol. 97, no. 2, pp. 123–147, Apr. 2012.
- [68] J. Zhao, M. Mathieu, and Y. LeCun, "Energy-based generative adversarial networks," in *Proc. ICLR*, 2017, pp. 1–17. [Online]. Available: <https://openreview.net/forum?id=ryh9pmce>
- [69] Y. Wang, B. Li, T. Che, K. Zhou, Z. Liu, and D. Li, "Energy-based open-world uncertainty modeling for confidence calibration," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9302–9311.
- [70] L. Gorelick, Y. Boykov, O. Veksler, I. B. Ayed, and A. Delong, "Submodularization for binary pairwise energies," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1154–1161.
- [71] S. Jegelka and J. Bilmes, "Submodularity beyond submodular energies: Coupling edges in graph cuts," in *Proc. CVPR*, Jun. 2011, pp. 1897–1904.
- [72] V. Kolmogorov and C. Rother, "Minimizing nonsubmodular functions with graph cuts—A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 7, pp. 1274–1279, Jul. 2007.
- [73] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.
- [74] P. L. Hammer, P. Hansen, and B. Simeone, "Roof duality, complementation and persistency in quadratic 0–1 optimization," *Math. Program.*, vol. 28, no. 2, pp. 121–155, Feb. 1984.
- [75] J. Yang, Y. Zhou, Y. Cao, and L. Feng, "Heterogeneous image change detection using deep canonical correlation analysis," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 2917–2922.
- [76] M. Yang, L. Jiao, F. Liu, B. Hou, S. Yang, and M. Jian, "DPFL-Nets: Deep pyramid feature learning networks for multiscale change detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6402–6416, Nov. 2022.
- [77] Y. Zhou, H. Liu, D. Li, H. Cao, J. Yang, and Z. Li, "Cross-sensor image change detection based on deep canonically correlated autoencoders," in *Proc. Int. Conf. Artif. Intell. Commun. Netw.* Harbin, China: Springer, 2019, pp. 251–257.

- [78] Y. Sun, L. Lei, X. Tan, D. Guan, J. Wu, and G. Kuang, "Structured graph based image regression for unsupervised multimodal change detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 185, pp. 16–31, Mar. 2022.
- [79] 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.
- [80] 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.
- [81] 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.
- [82] 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)*, Oct. 2018, pp. 3998–4002.
- [83] 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.
- [84] L. Lei, Y. Sun, and G. Kuang, "Adaptive local structure consistency-based heterogeneous remote sensing change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [85] Y. Chen and L. Bruzzone, "Self-supervised change detection in multiview remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5402812.
- [86] C. Wu, B. Du, and L. Zhang, "Fully convolutional change detection framework with generative adversarial network for unsupervised, weakly supervised and regional supervised change detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 9774–9788, Aug. 2023, doi: [10.1109/TPAMI.2023.3237896](https://doi.org/10.1109/TPAMI.2023.3237896).
- [87] W. S. Torgerson, "Multidimensional scaling: I. Theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, Dec. 1952.



**Yuli Sun** received the Ph.D. degree in information and communication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2023.

He is currently a Post-Doctoral Researcher with the College of Aerospace Science and Engineering, NUDT. He has published articles in *IEEE TRANSACTIONS ON IMAGE PROCESSING*, *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, *ISPRS Journal of Photogrammetry and Remote Sensing*, and *Pattern Recognition*. His research interests include 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 was with the High-Tech Institute of Xi'an, Xi'an, China. 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, 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.



**Zhang Li** received the Ph.D. degree from Delft University of Technology, Delft, The Netherlands, in 2015.

He is an Associate Professor with the College of Aerospace Science and Engineering, National University of Defense Technology, Changsha, China. He authored more than 20 papers in high-ranking journals and conferences, such as *IEEE TRANSACTIONS ON MEDICAL IMAGING*, *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, *IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS*, and *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*. His research interests include biomedical image analysis, particularly image registration and vision navigation.



**Li Liu** (Senior Member, IEEE) received the Ph.D. degree in information and communication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2012.

During the Ph.D. study, she spent more than two years as a Visiting Student with the University of Waterloo, Waterloo, Canada, from 2008 to 2010. From 2015 to 2016, she spent ten months visiting the Multimedia Laboratory, The Chinese University of Hong Kong, Hong Kong. From 2016 to 2018, she was a Senior Researcher with the Machine Vision Group, University of Oulu, Oulu, Finland. She is currently a Full Professor with the College of Electronic Science and Technology, NUDT. Her articles have currently more than 13 900 citations according to Google Scholar. Her current research interests include computer vision, pattern recognition, and machine learning.

Dr. Liu was the Co-Chair of nine international workshops with CVPR, ICCV, and ECCV. She served as the leading Guest Editor for special issues on *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE* and *International Journal of Computer Vision*. She currently serves as an Associate Editor for *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, and *Pattern Recognition*.