



Iterative structure transformation and conditional random field based method for unsupervised multimodal change detection

Yuli Sun^a, Lin Lei^{a,*}, Dongdong Guan^b, Junzheng Wu^a, Gangyao Kuang^a

^aCollege of Electronic science, National University of Defense Technology, Changsha 410073, China
^bHigh-Tech Institute of Xi'an, Xi'an 710025, China



ARTICLE INFO

Article history:

Received 20 November 2021

Revised 5 June 2022

Accepted 9 June 2022

Available online 13 June 2022

Keywords:

Unsupervised change detection
 KNN graph
 Image transformation
 Multimodal
 Conditional random field

ABSTRACT

Change detection between heterogeneous images has become an increasingly interesting research topic in remote sensing. The different appearances and statistics of heterogeneous images bring great challenges to this task. In this paper, we propose an unsupervised iterative structure transformation and conditional random field (IST-CRF) based multimodal change detection (MCD) method, combining an imaging modality-invariant based structure transformation method with a random field framework specifically designed for MCD, to acquire an optimal change map within a global probabilistic model. IST-CRF first constructs graphs to represent the structures of the images, and transforms the heterogeneous images to the same differential domain by using graph based forward and backward structure transformations. Then, the change vectors are calculated to distinguish the changed and unchanged areas. Finally, in order to classify the change vectors and compute the binary change map, a CRF model is designed to fully explore the spectral-spatial information, which incorporates the change information, local spatially-adjacent neighbor information, and global spectrally-similar neighbor information with a random field framework. As the changed samples will influence the structure transformation and reduce the quality of change vectors, we use an iterative framework to propagate the CRF segmentation results back to the structure transformation process that removes the changed samples, and thus improve the accuracy of change detection. Experiments conducted on different real data sets show the effectiveness of IST-CRF. Source code of the proposed method will be made available at <https://github.com/yulisun/IST-CRF>.

© 2022 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Background

Change detection (CD) is a process to detect changed regions by analyzing multitemporal remote sensing images acquired over the same area, which plays an important role in the military (such as missile early-warning, battlefield dynamic monitoring) and civil applications (such as environmental monitoring, damage assessment) [1–3].

Until now, many CD techniques have been proposed for solving the monomodal CD (or named homogeneous CD) problem, which assumes that the multitemporal images are obtained from the same imaging modality, i.e., the images are acquired by the same sensor (e.g., optical sensor and synthetic aperture radar (SAR) sensor) with similar imaging conditions. In this monomodal case, the multitemporal images can be directly compared to generate a

difference image (DI) to measure the change level by using arithmetical operators, such as the image differencing of optical images [4] and image rationing/log-rationing of SAR images [5]. However, the homogeneity assumption does not always hold in many practical applications, especially when different sensors are involved.

With the rapid development of remote sensing related technologies, more and more image data representing the information about the Earth's surface can be acquired from different sensors, which brings multimodal CD (or named heterogeneous CD) into the spotlight. The multimodal CD (MCD) is a procedure for identifying changes based on heterogeneous sources of data, which roughly contains two types: images acquired by different sensor types (multisource images, such as a pair of optical and SAR images); and images acquired by the same sensor type but with different sensors (cross-sensor or multisensor images, such as two multispectral images obtained from Landsat-8 and Sentinel-2, respectively). Thus, MCD can be regarded as a generalization of the traditional monomodal or homogeneous CD problem.

MCD is particularly attractive for the following two reasons: first, it can increase the temporal resolution or extend the time

* Corresponding author.

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

frame of long-term trend monitoring by inserting heterogeneous images along the timeline [6,7]; second, it can shorten the response time of CD analysis in the case of emergency natural disasters (such as flood and earthquake) and rescue operations [8]. MCD allows to detect changes using the first image of opportunity instead of waiting to obtain a comparable homogeneous image and, more importantly, the obtained homogeneous image may not be usable due to the adverse light and weather conditions accompanying the disasters.

1.2. Related work

MCD encounters greater challenges since it must be capable of comparing images of any existing different modalities to detect changes. Let \mathbf{X} and \mathbf{Y} be the co-registered images recorded at time t_1 and t_2 respectively, and let x and y be two data samples drawn on the same spatial location from \mathbf{X} and \mathbf{Y} , respectively. According to the basic analysis unit of CD method, x and y can be individual pixels, squared patches, or irregular superpixels. Let H_0 and H_1 indicate the “unchanged” and “changed” hypotheses, respectively. Then in the homogeneous CD, one can directly compare x and y to obtain the DI under the assumption that x and y exhibit the same (or similar) statistical properties when conditioned to H_0 . However, this assumption is violated in MCD because the multimodal images provide different descriptions of the same object and exhibit quite different characteristics.

The main purpose of MCD method is to make the “incomparable” heterogeneous images “comparable”, which are related to the topic of image transformation [9–11]. Generally, most MCD methods are designed to accomplish the following two types of transformations. The first can be regarded as image translation or image regression and expressed as two mappings of $\xi : x \rightarrow y'$ and $\zeta : y \rightarrow x'$, such that y' and y , x' and x have the same distribution when conditioned to H_0 . That is, the y' (or x') and y (or x) can be directly compared to obtain the DI as in homogeneous CD by these two mappings of (ξ, ζ) . In order to construct these mappings, Liu et al. [11] propose a homogeneous pixel transformation (HPT) method by using kernel regression trained by the known unchanged pixels. To eliminate the dependence on labeled data (unchanged samples), Luppino et al. [12] propose an unsupervised image regression method that first picks out the probably unchanged pixels as the pseudo-training data by affinity matrix difference (AMD), and then performs the image translation. With the self-expression property, another unsupervised regression method is proposed in [13], which learns a patch similarity graph matrix (PSGM) to represent the structure of one image and then transforms the PSGM to the other image domain to obtain the regression image. Some deep learning based translation methods have also been proposed, such as the AMD change prior based X-Net and adversarial cyclic encoder network (ACE-Net) [14], the conditional generative adversarial network (cGAN) [15], the coupling translation networks (CPTN) [16], and the image style transfer based method [17].

The second type of transformation can be expressed as two mappings of $\mu : x \rightarrow z$ and $\vartheta : y \rightarrow z'$, such that z and z' have the same (or quite similar) distribution when conditioned to H_0 , which can be regarded as transforming the heterogeneous images to a new common domain. The classification-based MCD methods transform the multitemporal images into a common category space, such as the post-classification comparison method (PCC) [18], the multitemporal segmentation and compound classification method (MS-CC) [19], and the classified adversarial network based method (CAN) [20]. The feature-based MCD methods transform the images to a common constructed or learned feature space, such as the manifold learning based method [21], the kernel canonical correlation analysis based method (kCCA) [22], the logarithmic transformation feature learning network (LT-FL) [23], the probabilistic model based on bipartite convolutional neural network (BCCN) [24], the deep sparse residual model based on anomaly feature learning (AFL-DSR) [25], and the commonality autoencoder based common feature learning method (CACFL) [26].

According to the construction process of the mappings of (ξ, ζ) and (μ, ϑ) , the transformation based MCD methods can be divided into supervised, semi-supervised and unsupervised. In the supervised and semi-supervised methods, the labeled unchanged samples are used to train the image regression model [8,11], the classifier [18,27], or the common feature learning [21,28]. Since labeling samples requires a high cost of manual operation and needs extensive expert knowledge in practice, the unsupervised MCD methods are more remarkably interesting for applications. Since there is no ground truth to guide the transformation process, unsupervised methods usually need a pre-constructed pseudo-training set [12,14,29], or involve a self-supervised framework [30], or use a coarse-to-fine filtering process [31,32].

Once the image transformations of (ξ, ζ) or (μ, ϑ) are completed, the DI can be calculated by comparing the transformed y' and y (x' and x) or z' and z . Then, the final change map (CM) solution can be treated as an image segmentation problem as in homogeneous CD, which can be accomplished by using the thresholding methods such as Otsu threshold [33], or clustering methods such as the K-means clustering [34] and fuzzy c-means (FCM) clustering [35], or the random field models such as Markov random field (MRF) and conditional random field (CRF) models. In the random field models based CD (either monomodal or multimodal), the local interactive information of pixels (or patches, superpixels) can be passed to the global information with the help of a global probabilistic framework.

In the homogeneous CD, the DI is usually used as the observation field, and the spectral and spatial contextual information of the image are incorporated in the segmentation process by the MRF-based methods [36,37] and CRF-based methods [38–40]. For the MCD, Mignotte [41] proposes a fractal projection and Markovian segmentation based method (FPMS) to binarize the DI, which contains a parameter estimation process with a mixture of two Gaussian likelihood distributions (for the two class labels: “changed” and “unchanged”) by using expectation maximization (EM) or stochastic EM algorithms and a segmentation process based on the estimated parameters. Touati et al. [42] propose a Markov model for MCD (named M3CD for short). M3CD constructs a visual cue to distinguish the two pixels that belong to two different class labels (different pixel-pairwise labels) or share the same class label (identical pixel-pairwise label), and assumes that the visual cue obeys a Gaussian distribution under the different pixel-pairwise label and obeys an exponential distribution under the identical pixel-pairwise label. Recently, an iterative robust graph and MRF co-segmentation model (IRG-McS) is proposed for MCD [43]. IRG-McS calculates the DI by graph mapping and fuses the DI in the MRF co-segmentation process, which assumes that the DI obeys a uniform distribution.

1.3. Motivations and contributions

1) A main challenge in unsupervised MCD is that the image transformation must be learnt from a dataset that includes noise and changes, which means that the noise and changed pixels will contaminate the mappings of (ξ, ζ) and (μ, ϑ) . To alleviate the influence of noise, the proposed method calculates the change vector by comparing the structure difference between heterogeneous images with a self-similarity based structure transformation, which is more robust to the noise than the common transformation based methods that aim to learn a luminance transformation function [11,12]. More importantly, we pay particular attention to the in-

fluence of changed pixels on image transformation, which is rarely mentioned by other studies. Similar to the iterative framework in IRG-McS [43], the proposed method combines the image transformation process and image segmentation process to reduce the negative impact of changed pixels, which propagates the detection result calculated by CRF segmentation model back to the structure transformation process to calculate change vectors. This can bring two benefits: first, it makes the transformation more robust and makes the change vectors highlight more the changed regions; second, it improves the detection accuracy of final CM with the generated high-quality change vectors.

2) The self-similarity property has been widely used in the image denoising and image super-resolution fields with the so-called “nonlocal-based” methods [44], which is a widespread property across different types of satellite images (also common to natural images) that can be expressed as “any small part of the image has many similar parts within the same image”. Recently, it has also been exploited in the MCD to establish a connection between heterogeneous images. For example, the self-similarity is used to complete the image regression in PSGM [13] and FPMS [41] as $\xi : x \xrightarrow{w} y'$ by image reconstruction with a self-expression matrix w and fractal projection with fractal code w , respectively. In [45,46], the self-similarity is used to complete the graph mapping $\mu : x \rightarrow w$, $\vartheta : y \rightarrow w'$ with w and w' denoting the K-nearest neighbor (KNN) graphs, which is similar to the approach adopted in the proposed method. However, different from the above literature, we apply the self-similarity to structure transformation and analyze the transformation process in detail in this paper. We explicitly give the transformation expression, that is, how to transform the heterogeneous images to the same differential domain and how to compare the transformed images. Moreover, unlike the KNN graphs in [45] and [46] that select nearest neighbors (NNs) for each vertex with fixed K and without considering the status of the neighbors (changed or unchanged), the graph constructed in the proposed method is more robust by adaptively selecting NNs in the unchanged class for each vertex with different K . At the same time, different from IRG-McS [43] that calculates a change probability value for each pixel, the proposed method computes two change vectors (change features) for each superpixel by comparing the structure difference between heterogeneous images, which contains more change information than the univariate probability value.

3) The distribution assumption in MRF, such as the Gaussian distribution assumption in FPMS [41], the exponential distribution assumption in M3CD [42], and the uniform distribution in IRG-McS [43], is not always suitable for MCD with different kinds of heterogeneous images. Different from MRF which is a Bayesian generative model, CRF is a discriminative probability model which directly builds the posterior distribution of the CM conditioned on the DI, so it is more flexible and robust. In this paper, a novel CRF model is designed to calculate the CM by incorporating the change information, local spatially-adjacent-neighbor information (LSAN), and global spectrally-similar-neighbor information (GSSN) with a random field framework. Specifically, the unary potential that represents the probability of a superpixel being labeled as changed/unchanged is obtained by adopting the FCM algorithm on the change vectors, without prior assumption of the distribution of changed/unchanged class. Considering the specificity of the MCD problem, we design two pairwise potentials for the CRF model: LSAN based pairwise potential and GSSN based pairwise potential. The LSAN takes into account not only the spatial continuity of the pairwise superpixels, but also their similarity relationships in the original multitemporal images. The interaction of spectrally similar superpixels is considered by introducing the GSSN into the CRF model, which can make full use of the structure information of the heterogeneous images. By using the ro-

bust unary potential and combining the local and global information based pairwise potentials, the CD results of the proposed method can be greatly improved with less false detection and miss detection.

The main contributions of the proposed iterative structure transformation and conditional random field (IST-CRF) based method are summarized as follows:

- A structure transformation is proposed to transform the heterogeneous images to the same differential domain.
- A CRF model is designed for MCD by incorporating the change information based unary potential, LSAN and GSSN based pairwise potentials.
- An iterative framework is used to combine the structure transformation and CRF segmentation to improve the detection accuracy.

1.4. Outline

The rest of this paper is organized as follows. Section II describes the structure transformation process of the proposed MCD method. Section III describes the details of the proposed CRF model. Section IV presents the experimental results by comparing them with some existing state-of-the-art (SOTA) methods and gives some discussions. Finally, we conclude this paper in Section V.

2. Structure transformation

We consider a pair of co-registered heterogeneous images obtained at time t_1 (pre-event) and t_2 (post-event) denoted as $\mathbf{X} \in \mathbb{R}^{H \times W \times C_X}$ and $\mathbf{Y} \in \mathbb{R}^{H \times W \times C_Y}$, which are from different domains of \mathcal{X} and \mathcal{Y} , respectively. We define the pixel vectors in \mathbf{X} and \mathbf{Y} as $\mathbf{x}(h, w) \in \mathbb{R}^{C_X}$ and $\mathbf{y}(h, w) \in \mathbb{R}^{C_Y}$, $1 \leq h \leq H$, $1 \leq w \leq W$. Here, H , W and C_X (or C_Y) represent the height, width, and number of channels of the image \mathbf{X} (or \mathbf{Y}), respectively.

As mentioned in the introduction, the heterogeneous images give different descriptions of the same object and exhibit different characteristics, so it is meaningless to directly compare the pixel values of $\mathbf{x}(h, w)$ and $\mathbf{y}(h, w)$ in MCD. Therefore, we first need to transform the “incomparable” heterogeneous images into a common domain to make them “comparable”, and then segment the comparison result into binary CM to detect changes.

The proposed MCD method consists of three steps: 1) pre-processing, which contains superpixel segmentation and feature extraction; 2) image transformation, which contains structure representation, structure transformation, and change vector calculation; 3) image segmentation, which contains CRF model construction and solution. The framework of IST-CRF is illustrated in Fig. 1.

2.1. Pre-processing

In the pre-processing, the main task is to generate the co-segmentation superpixels and extract the features of superpixels. Rather than focusing on the individual pixel or squared patch (such as methods in [41,45]), the image block (superpixel) is chosen as the basic analysis unit in IST-CRF, which can bring two benefits: first, the superpixel can maintain the structure, edge and context information of object as the interior of each superpixel itself is homogeneous (it internally belongs to the same kind of object); second, it can reduce the computational complexities of the subsequent image transformation and segmentation, which is very useful for large-scale very-high-resolution images.

In IST-CRF, the simple linear iterative clustering (SLIC) method [47] is applied on each image to generate the superpixels. Since

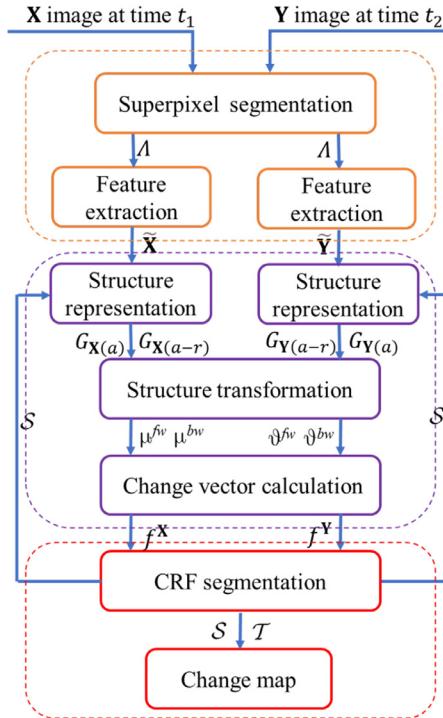


Fig. 1. Framework of the proposed IST-CRF.

the original SLIC is designed for the optical image with RGB bands, we need to adjust it to fit the remote sensing images involved in MCD. For the optical image \mathbf{X} with RGB bands, the original SLIC is directly employed. For the multispectral image \mathbf{X} with $C_{\mathbf{X}} > 3$, the principal component analysis (PCA) method is firstly adopted to reduce its dimension so that each pixel contains three principal components, and then SLIC is used to segment the image. For the SAR image, inspired by the generalized likelihood ratio (GLR) based similarity criterion for multiplicative speckle noise model with Gamma distribution proposed in [48], we use the GLR based distance, $d_{GLR} = \log \left(\frac{x_i + x_j}{2\sqrt{x_i x_j}} \right)$, to replace the Euclidean distance in original SLIC and generate the superpixels for SAR image, with x_i and x_j being the intensity values of two pixels.

Once the SLIC (or adjusted SLIC) is performed on each image independently, we can obtain the segmentation map of each image, denoted as $\Lambda^{\mathbf{X}}$ and $\Lambda^{\mathbf{Y}}$. By taking the intersection of segmentation maps $\Lambda^{\mathbf{X}}$ and $\Lambda^{\mathbf{Y}}$, eliminating the empty sets, and merging the very small regions (i.e., regions that are smaller than the N_S -th smallest region in the co-segmentation map) into the nearest larger regions whose centers are the closest to the centers of these small regions in spatial Euclidean distance, we can obtain the co-segmentation map $\Lambda = \{\Lambda_i | i = 1, \dots, N_S\}$ and the segmented superpixels of \mathbf{X} and \mathbf{Y} , denoted as $\mathbf{X}_i = \{\mathbf{x}(h, w) | (h, w) \in \Lambda_i\}$ and $\mathbf{Y}_i = \{\mathbf{y}(h, w) | (h, w) \in \Lambda_i\}$ respectively with $i = 1, \dots, N_S$. Since the co-segmentation map Λ is an intersection of $\Lambda^{\mathbf{X}}$ and $\Lambda^{\mathbf{Y}}$, the set of pixels inside each superpixel in Λ exhibits homogeneous structure in both multitemporal images of \mathbf{X} and \mathbf{Y} .

After the superpixel co-segmentation is completed, features representing different information can be extracted from the superpixel, such as the spectral (intensity), textural, and spatial information. In this paper, the mean, median, and variance values of each band are selected as the features. By stacking the feature vectors of $\tilde{\mathbf{X}}_i \in \mathbb{R}^{3C_{\mathbf{X}}}$ and $\tilde{\mathbf{Y}}_i \in \mathbb{R}^{3C_{\mathbf{Y}}}$ of superpixels \mathbf{X}_i and \mathbf{Y}_i , we can obtain the feature matrices of $\tilde{\mathbf{X}} \in \mathbb{R}^{3C_{\mathbf{X}} \times N_S}$ and $\tilde{\mathbf{Y}} \in \mathbb{R}^{3C_{\mathbf{Y}} \times N_S}$, respectively.

2.2. Structure transformation

The structure transformation of IST-CRF is based on the inherent self-similarity property of images, that is, each small part of the image can always find some similar parts within the same image. At the same time, this self-similarity can be preserved across different modalities. As the multitemporal images involved in MCD are acquired at the same geographical area, if \mathbf{X}_i and \mathbf{X}_j represent the same kind of object (showing that they are very similar in the pre-event image) and neither of them changed during the event, then \mathbf{Y}_i and \mathbf{Y}_j also represent the same kind of object (showing that they are also very similar in the post-event image). We use the similarity relationships between the target part (\mathbf{X}_i) and its similar parts (\mathbf{X}_j) to represent the structure of this target part (\mathbf{X}_i). This nonlocal similarity within the image itself would eliminate the discrepancy between the two modalities. Therefore, these similarity relationships based on structure consistency are quite imaging modality invariant.

2.2.1. Structure representation

We construct a graph to capture the structure information of each image, which is an effective tool for image representation and analysis [49,50]. Given the pre-event image \mathbf{X} with the superpixel segmentation map Λ , we set each superpixel as a vertex and construct the directed KNN graph $G_{\mathbf{X}} = \{V_{\mathbf{X}}, B_{\mathbf{X}}\}$, with the set of vertices $V_{\mathbf{X}}$ and the set of edges $B_{\mathbf{X}}$ such that

$$\begin{aligned} V_{\mathbf{X}} &= \{\mathbf{X}_i | i = 1, \dots, N_S\}, \\ B_{\mathbf{X}} &= \{(\mathbf{X}_i, \mathbf{X}_j) | i = 1, \dots, N_S; j \in \mathcal{N}_{\mathbf{X}_i}^K\}, \end{aligned} \quad (1)$$

where $\mathcal{N}_{\mathbf{X}_i}^K$ represents the position set of the K NNs of \mathbf{X}_i . Here, we define $\mathbf{D}_{\mathbf{X}}^{\mathbf{X}} \in \mathbb{R}^{N_S \times N_S}$ as the feature distance matrix of image \mathbf{X} with the element $D_{i,j}^{\mathbf{X}} = \|\tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_j\|_2^2$ being the distance between superpixels \mathbf{X}_i and \mathbf{X}_j . Then, $j \in \mathcal{N}_{\mathbf{X}_i}^K$ if and only if $D_{i,j}^{\mathbf{X}}$ is among the K -smallest elements except $D_{i,i}^{\mathbf{X}}$ in the distance vector $\mathbf{D}_{\mathbf{X}}^{\mathbf{X}}$. The graph $G_{\mathbf{X}}$ can be formally represented by its adjacent matrix $\mathbf{A}_{\mathbf{X}} \in \mathbb{R}^{N_S \times N_S}$, with each element defined as

$$A_{i,j}^{\mathbf{X}} \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } (\mathbf{X}_i, \mathbf{X}_j) \in B_{\mathbf{X}} \\ 0, & \text{if } (\mathbf{X}_i, \mathbf{X}_j) \notin B_{\mathbf{X}} \end{cases} \quad (2)$$

For the post-event image \mathbf{Y} , we can construct the directed KNN graph $G_{\mathbf{Y}} = \{V_{\mathbf{Y}}, B_{\mathbf{Y}}\}$ and obtain the corresponding adjacent matrix $\mathbf{A}_{\mathbf{Y}} \in \mathbb{R}^{N_S \times N_S}$ in a similar way.

2.2.2. Structure transformation and change vector calculation

Since the KNN graphs of $G_{\mathbf{X}}$ and $G_{\mathbf{Y}}$ are constructed in different domains, it is not appropriate to compare them directly (such as $A^{\mathbf{X}} - A^{\mathbf{Y}}$), which will cause the leakage of heterogeneous data. Next, we transform the heterogeneous images of \mathbf{X} and \mathbf{Y} into a common domain by using the structure based graphs of $G_{\mathbf{X}}$ and $G_{\mathbf{Y}}$, respectively.

For the forward transformation, the pre-event image \mathbf{X} and post-event image \mathbf{Y} are transformed to their respective differential domains to calculate the mean difference between each superpixel and its own K NNs as

$$\begin{aligned} \mu^{fw}(\mathbf{X}_i) &= \frac{1}{K} \sum_{j \in \mathcal{N}_{\mathbf{X}_i}^K} |\tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_j| = \frac{1}{K} \sum_{j \in \mathcal{N}_{\mathbf{X}_i}^K} |\Delta_{i,j}^{\mathbf{X}}|, \\ \vartheta^{fw}(\mathbf{Y}_i) &= \frac{1}{K} \sum_{j \in \mathcal{N}_{\mathbf{Y}_i}^K} |\tilde{\mathbf{Y}}_i - \tilde{\mathbf{Y}}_j| = \frac{1}{K} \sum_{j \in \mathcal{N}_{\mathbf{Y}_i}^K} |\Delta_{i,j}^{\mathbf{Y}}|, \end{aligned} \quad (3)$$

where $\Delta_{i,j}^{\mathbf{X}} = \tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_j$ and $\Delta_{i,j}^{\mathbf{Y}} = \tilde{\mathbf{Y}}_i - \tilde{\mathbf{Y}}_j$ are the feature difference vectors between superpixels, and $|\cdot|$ denotes the absolute operation.

For the backward transformation, the pre-event image \mathbf{X} and post-event image \mathbf{Y} are transformed into each other's differential domain to calculate the mean difference between each superpixel and its mapped K NNs as

$$\begin{aligned}\mu^{bw}(\mathbf{X}_i) &= \frac{1}{K} \sum_{j \in \mathcal{N}_{\mathbf{X}_i}^K} |\tilde{\mathbf{Y}}_i - \tilde{\mathbf{Y}}_j| = \frac{1}{K} \sum_{j \in \mathcal{N}_{\mathbf{X}_i}^K} |\Delta_{i,j}^{\mathbf{Y}}|, \\ \vartheta^{bw}(\mathbf{Y}_i) &= \frac{1}{K} \sum_{j \in \mathcal{N}_{\mathbf{Y}_i}^K} |\tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_j| = \frac{1}{K} \sum_{j \in \mathcal{N}_{\mathbf{Y}_i}^K} |\Delta_{i,j}^{\mathbf{X}}|.\end{aligned}\quad (4)$$

Then, with these structure transformations, we can compare the structure of \mathbf{X}_i and \mathbf{Y}_i in the same differential domains to calculate the change vectors of $f_i^{\mathbf{X}}$ and $f_i^{\mathbf{Y}}$ as

$$\begin{aligned}f_i^{\mathbf{X}} &= |\mu^{fw}(\mathbf{X}_i) - \vartheta^{bw}(\mathbf{Y}_i)|, \\ f_i^{\mathbf{Y}} &= |\mu^{bw}(\mathbf{X}_i) - \vartheta^{fw}(\mathbf{Y}_i)|.\end{aligned}\quad (5)$$

By using the adjacent matrices of $\mathbf{A}^{\mathbf{X}}$ and $\mathbf{A}^{\mathbf{Y}}$, the change vectors can be rewritten as

$$\begin{aligned}f_i^{\mathbf{X}} &= \frac{1}{K} \left| \sum_{j=1}^{N_s} (\mathbf{A}_{i,j}^{\mathbf{X}} - \mathbf{A}_{i,j}^{\mathbf{Y}}) |\Delta_{i,j}^{\mathbf{X}}| \right|, \\ f_i^{\mathbf{Y}} &= \frac{1}{K} \left| \sum_{j=1}^{N_s} (\mathbf{A}_{i,j}^{\mathbf{X}} - \mathbf{A}_{i,j}^{\mathbf{Y}}) |\Delta_{i,j}^{\mathbf{Y}}| \right|.\end{aligned}\quad (6)$$

Intuitively, the change vectors are calculated by how different the adjacent matrices of $\mathbf{A}^{\mathbf{X}}$ and $\mathbf{A}^{\mathbf{Y}}$ are in the differential domains. From (3) to (6), we can find that if the areas represented by superpixel \mathbf{X}_i does not change in the event, the mapped superpixel $\mathbf{X}_{j'}$, $j' \in \mathcal{N}_{\mathbf{Y}_i}^K$ will be similar as \mathbf{X}_i with a high probability, then the difference between $\Delta_{i,j}^{\mathbf{X}}$, $j \in \mathcal{N}_{\mathbf{X}_i}^K$ and $\Delta_{i,j'}^{\mathbf{X}}$, $j' \in \mathcal{N}_{\mathbf{Y}_i}^K$ will be small, thus leading to small values of the elements in the change vector $f_i^{\mathbf{X}}$. On the contrary, if the areas represented by superpixel \mathbf{X}_i changes in the event, the mapped superpixel is no longer similar to \mathbf{X}_i with a high probability, then the difference between $\Delta_{i,j}^{\mathbf{X}}$, $j \in \mathcal{N}_{\mathbf{X}_i}^K$ and $\Delta_{i,j'}^{\mathbf{X}}$, $j' \in \mathcal{N}_{\mathbf{Y}_i}^K$ will be large, thus leading to large values of the elements in the change vector $f_i^{\mathbf{X}}$. Also, the superpixel \mathbf{Y}_i and change vector $f_i^{\mathbf{Y}}$ has a similar relationship. Therefore, we can find that the $f_i^{\mathbf{X}}$ and $f_i^{\mathbf{Y}}$ have different representations for the changed and unchanged superpixels, that is, the changed and unchanged superpixels can be distinguished.

2.2.3. K-selection

In the structure transformation (3), (4) and change vector calculation (6), we can find that the number K of the KNN graph plays an important role. Obviously, a very small K is not appropriate, which will make the graph ($G_{\mathbf{X}}$ and $G_{\mathbf{Y}}$) less informative, and makes the change measurement ($f_i^{\mathbf{X}}$ and $f_i^{\mathbf{Y}}$) not robust enough. On the contrary, a particular large K is also not appropriate, which tends to over connect the graph and leads to confusion in the change measurement. For example, in the extreme case of $K = N_s$, the $f_i^{\mathbf{X}}$ and $f_i^{\mathbf{Y}}$ will always be equal to 0, regardless of whether \mathbf{X}_i changes.

Therefore, we need to choose a suitable k_i for each vertex instead of a fixed K as in the common KNN graph. Here, we propose a K -adaptive strategy similar to the one in IRG-McS [43] to pursue the goal of “each superpixel is connected to as many truly similar vertices as possible”.

Step 1. Set $k_{\max} = \lceil \sqrt{N_s} \rceil$ with $\lceil \cdot \rceil$ represents the rounding up operation, and construct the KNN graph of $G_{\mathbf{X}}$ with $K = k_{\max}$.

Step 2. Calculate the in-degree $di(\mathbf{X}_i) = \sum_{j=1}^{N_s} \mathbf{A}_{j,i}^{\mathbf{X}}$ for each vertex \mathbf{X}_i , that is, compute the number of times \mathbf{X}_i occurs among the k_{\max} nearest-neighbors of all the vertexes.

Step 3. Set $k_{\min} = \lceil \sqrt{N_s}/10 \rceil$, and calculate $k_{\mathbf{X}_i} = \min \{k_{\max}, \max \{di(\mathbf{X}_i), k_{\min}\}\}$. Then, construct the adaptive KNN graph $G_{\mathbf{X}(a)}$ with $k_{\mathbf{X}_i}$ for each vertex \mathbf{X}_i .

Similarly, we can construct the adaptive KNN graph $G_{\mathbf{Y}(a)}$ with $k_{\mathbf{Y}_i}$ for each vertex \mathbf{Y}_i . With this strategy of K -selection, we can select a smaller k for superpixel that belongs to low density and select a larger k for superpixel that belongs to high density.

2.2.4. Robust graph construction

In the backward structure transformations of μ^{bw} and ϑ^{bw} (4), we directly mapped the $\mathbf{A}^{\mathbf{X}}$ and $\mathbf{A}^{\mathbf{Y}}$ to each other's differential domains without considering the status of the neighbors (changed or unchanged). However, the changed neighbors will affect the structure transformation and degrade the performance of change vectors of $f^{\mathbf{X}}$ and $f^{\mathbf{Y}}$ (6).

For the unchanged superpixel \mathbf{X}_i , if one of its K NNs changed in the event, i.e., the superpixel $\mathbf{X}_{j'}$, $j' \in \mathcal{N}_{\mathbf{X}_i}^K$ is changed, then the mapped \mathbf{Y}_i and $\mathbf{Y}_{j'}$ are dissimilar (belongs to different objects). This brings an unstable $\Delta_{i,j'}^{\mathbf{Y}}$ with large elements in the backward transformation $\mu^{bw}(\mathbf{X}_i)$ (4) of the unchanged superpixel \mathbf{X}_i , and thus makes the value of element in the $f_i^{\mathbf{Y}}$ larger. On the contrary, for the changed superpixel \mathbf{X}_i , if one of its K NN changed the same way as \mathbf{X}_i , that is \mathbf{X}_i and $\mathbf{X}_{j'}$, $j' \in \mathcal{N}_{\mathbf{X}_i}^K$ changed to the same category, then the mapped \mathbf{Y}_i and $\mathbf{Y}_{j'}$ are similar (belongs to same objects). This brings an unstable $\Delta_{i,j'}^{\mathbf{Y}}$ with small elements in the backward transformation $\mu^{bw}(\mathbf{X}_i)$ (4) of the changed superpixel \mathbf{X}_i , and thus makes the value of the element in the $f_i^{\mathbf{Y}}$ smaller. From the above analysis, we can find that whether for the changed or unchanged superpixel \mathbf{X}_i , its changed K NNs will make the change vector $f_i^{\mathbf{Y}}$ less discriminative. Also, for the superpixel \mathbf{Y}_i and the change vector $f_i^{\mathbf{X}}$, we have the same conclusion.

To reduce the negative impact of the changed neighbors, we need to eliminate them in the backward structure transformation (4). However, we cannot identify in advance which superpixels are changed, so we employ an iterative framework to complete the elimination of changed neighbors in the backward structure transformation. Therefore, we propagate the detection result generated by CRF segmentation of the previous round (in Section III) back to the structure transformation of the next round.

Let the previous round result of CRF segmentation in Section III be the index subset \mathcal{T} of changed superpixels and the index subset \mathcal{S} of unchanged superpixels. With the K -adaptive strategy and unchanged index subset \mathcal{S} , we can construct one adaptive KNN graph $G_{\mathbf{X}(a)} = \{V_{\mathbf{X}(a)}, B_{\mathbf{X}(a)}\}$ and one adaptive-robust KNN graph $G_{\mathbf{X}(a-r)} = \{V_{\mathbf{X}(a-r)}, B_{\mathbf{X}(a-r)}\}$ for image \mathbf{X} as

$$V_{\mathbf{X}(a)} = V_{\mathbf{X}(a-r)} = \{\mathbf{X}_i | i = 1, \dots, N_s\},$$

$$B_{\mathbf{X}(a)} = \left\{ (\mathbf{X}_i, \mathbf{X}_j) | i = 1, \dots, N_s; j \in \mathcal{N}_{\mathbf{X}_i}^{k_{\mathbf{X}_i}} \right\}, \quad (7)$$

$$B_{\mathbf{X}(a-r)} = \left\{ (\mathbf{X}_i, \mathbf{X}_j) | i = 1, \dots, N_s; j \in \mathcal{N}_{\mathbf{X}_i}^{k_{\mathbf{X}_i}^r} \right\}.$$

Here, $\mathcal{N}_{\mathbf{X}_i}^{k_{\mathbf{X}_i}}$ represents the position set of $k_{\mathbf{X}_i}$ NNs of \mathbf{X}_i in the whole image \mathbf{X} with the adaptive $k_{\mathbf{X}_i}$ calculated by **Steps 1-3**, and $\mathcal{N}_{\mathbf{X}_i}^{k_{\mathbf{X}_i}^r}$ represents the position set of $k_{\mathbf{X}_i}^r$ NNs of \mathbf{X}_i in the unchanged set $\{\mathbf{X}_j | j \in \mathcal{S}\}$ with the adaptive $k_{\mathbf{X}_i}^r$ calculated by **Steps 1-3** with $k_{\max} = \lceil \sqrt{\text{card}(\mathcal{S})} \rceil$ and $k_{\min} = \lceil \sqrt{\text{card}(\mathcal{S})}/10 \rceil$, where $\text{card}(\cdot)$ denotes the cardinality of a set. Then, we can obtain the corresponding adjacent matrices of $\mathbf{A}^{\mathbf{X}(a)}$ and $\mathbf{A}^{\mathbf{X}(a-r)}$. For the post-event image \mathbf{Y} , we can construct the adaptive $G_{\mathbf{Y}(a)} = \{V_{\mathbf{Y}(a)}, B_{\mathbf{Y}(a)}\}$ and adaptive-robust $G_{\mathbf{Y}(a-r)} = \{V_{\mathbf{Y}(a-r)}, B_{\mathbf{Y}(a-r)}\}$, and obtain the corresponding adjacent matrices of $\mathbf{A}^{\mathbf{Y}(a)}$ and $\mathbf{A}^{\mathbf{Y}(a-r)}$ in a similar way.

2.2.5. Robust structure transformation and DI calculation

By using the structure based graphs of $G_{\mathbf{X}(a)}$, $G_{\mathbf{X}(a-r)}$, $G_{\mathbf{Y}(a)}$ and $G_{\mathbf{Y}(a-r)}$, the robust forward transformation of (3) can be rewritten as

$$\begin{aligned}\mu^{fw}(\mathbf{X}_i) &= \frac{1}{k_{\mathbf{X}_i}} \sum_{j \in \mathcal{N}_{\mathbf{X}_i}^{k_{\mathbf{X}_i}}} |\Delta_{i,j}^{\mathbf{X}}|, \\ \vartheta^{fw}(\mathbf{Y}_i) &= \frac{1}{k_{\mathbf{Y}_i}} \sum_{j \in \mathcal{N}_{\mathbf{Y}_i}^{k_{\mathbf{Y}_i}}} |\Delta_{i,j}^{\mathbf{Y}}|.\end{aligned}\quad (8)$$

The robust backward transformation of (4) can be rewritten as

$$\begin{aligned}\mu^{bw}(\mathbf{X}_i) &= \frac{1}{k'_{\mathbf{X}_i}} \sum_{j \in \mathcal{N}_{\mathbf{X}_i}^{k'_{\mathbf{X}_i}}} |\Delta_{i,j}^{\mathbf{Y}}|, \\ \vartheta^{bw}(\mathbf{Y}_i) &= \frac{1}{k'_{\mathbf{Y}_i}} \sum_{j \in \mathcal{N}_{\mathbf{Y}_i}^{k'_{\mathbf{Y}_i}}} |\Delta_{i,j}^{\mathbf{X}}|.\end{aligned}\quad (9)$$

Then the robust change vectors of $f_i^{\mathbf{X}}$ and $f_i^{\mathbf{Y}}$ can be calculated by

$$\begin{aligned}f_i^{\mathbf{X}} &= \left| \mu^{fw}(\mathbf{X}_i) - \vartheta^{bw}(\mathbf{Y}_i) \right| = \left| \sum_{j=1}^{N_S} \left(\frac{A_{i,j}^{\mathbf{X}(a)}}{k_{\mathbf{X}_i}} - \frac{A_{i,j}^{\mathbf{Y}(a-r)}}{k'_{\mathbf{Y}_i}} \right) |\Delta_{i,j}^{\mathbf{X}}| \right|, \\ f_i^{\mathbf{Y}} &= \left| \mu^{bw}(\mathbf{X}_i) - \vartheta^{fw}(\mathbf{Y}_i) \right| = \left| \sum_{j=1}^{N_S} \left(\frac{A_{i,j}^{\mathbf{Y}(a)}}{k_{\mathbf{Y}_i}} - \frac{A_{i,j}^{\mathbf{X}(a-r)}}{k'_{\mathbf{X}_i}} \right) |\Delta_{i,j}^{\mathbf{Y}}| \right|.\end{aligned}\quad (10)$$

Then, by assigning the change features of $f_i^{\mathbf{X}}$ and $f_i^{\mathbf{Y}}$ to the pixels located in Λ_i , we can obtain the $\text{DI}^{\mathbf{X}} \in \mathbb{R}^{H \times W \times (3C_{\mathbf{X}})}$ and $\text{DI}^{\mathbf{Y}} \in \mathbb{R}^{H \times W \times (3C_{\mathbf{Y}})}$ respectively with

$$\left. \begin{aligned}\text{DI}^{\mathbf{X}}(h, w, c) &= f_i^{\mathbf{X}}(c) \\ \text{DI}^{\mathbf{Y}}(h, w, c) &= f_i^{\mathbf{Y}}(c)\end{aligned} \right\} \text{if } (h, w) \in \Lambda_i. \quad (11)$$

3. CRF segmentation

Once the DIs are obtained by calculating the structure difference, the CD problem can be regarded as an image segmentation problem. In the proposed method, a CRF model is designed to label superpixels as “changed” or “unchanged” by considering both the unary component and pairwise components.

CRF is an undirected graphical model to estimate probability distribution conditioned on observation. Formally, let $\mathcal{G} = (\mathcal{V}, \mathcal{B})$ be a graph constructed on random variables $\mathbf{Z} = \{Z_1, \dots, Z_N\}$ with labels $\mathbf{L} = \{L_1, \dots, L_N\}$. Then (\mathbf{Z}, \mathbf{L}) is a CRF when the probability of labeled variables \mathbf{L} conditioned on observed variables \mathbf{Z} obeys the Markov property. CRF models the conditional probability of a label sequence given the observed data sequence [51], as follows:

$$P(\mathbf{L}|\mathbf{Z}) = \frac{1}{J(\mathbf{L})} \prod_{c \in C} \Phi_c(L_c, \mathbf{Z}), \quad (12)$$

where c is a clique and C is the set of all cliques, $J(\mathbf{L})$ is the normalization constant, and $\Phi_c(L_c, \mathbf{Z})$ is the potential function defined on clique c , such as unary, pairwise potential, and even high-order potentials.

In the proposed CRF model, the following notations and definitions are used.

1) Given the observation field $\mathbf{Z} = \{Z_i | i \in \mathcal{I}\}$ with $Z_i = [f_i^{\mathbf{X}}; f_i^{\mathbf{Y}}] \in \mathbb{R}^{(3C_{\mathbf{X}}+3C_{\mathbf{Y}})}$ being the stacked change vector and representing the change features of the i -th superpixel, where $\mathcal{I} = \{1, 2, \dots, N_S\}$ is the set of superpixel indices.

2) $\mathbf{L} = \{L_i | i \in \mathcal{I}\}$ is the corresponding label field with $L_i \in \{0, 1\}$, where $L_i = 0$ means that the region of Λ_i is unchanged and $L_i = 1$ means that the region of Λ_i is changed in the event.

The proposed CRF for MCD is formulated by considering the unary potential, local spatially adjacent neighbor (LSAN) based pairwise potential, and global spectrally similar neighbor (GSSN)

based pairwise potential as follows

$$P(\mathbf{L}|\mathbf{Z}) = \frac{1}{J(\mathbf{L})} \exp \left\{ - \underbrace{\sum_{i \in \mathcal{I}} \varphi_i(\mathbf{Z}, L_i)}_{\text{unary}} - \alpha \underbrace{\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{N}_i^{LSAN}} \phi_{ij}(\mathbf{Z}, L_i, L_j)}_{\text{LSAN pairwise}} \right. \\ \left. - \beta \underbrace{\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{N}_i^{GSSN}} \psi_{ij}(\mathbf{Z}, L_i, L_j)}_{\text{GSSN pairwise}} \right\}, \quad (13)$$

where φ_i is the unary potential that represents the single relationship between the change vector and its corresponding label, ϕ_{ij} is the LSAN based pairwise potential that constructs the interaction between the i -th node and its spatial neighborhood defined by \mathcal{N}_i^{LSAN} , and ψ_{ij} is the GSSN based pairwise potential that constructs the interaction between the i -th node and its spectrally similar neighborhood (K NNs) defined by \mathcal{N}_i^{GSSN} , and $\alpha, \beta > 0$ are two balancing parameters to control the weights of these potentials. The Gibbs energy for the CRF model is given as

$$\begin{aligned}E(\mathbf{L}|\mathbf{Z}) &= \underbrace{\sum_{i \in \mathcal{I}} \varphi_i(\mathbf{Z}, L_i)}_{E_{\text{unary}}} + \alpha \underbrace{\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{N}_i^{LSAN}} \phi_{ij}(\mathbf{Z}, L_i, L_j)}_{E_{\text{LSAN}}} \\ &\quad + \beta \underbrace{\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{N}_i^{GSSN}} \psi_{ij}(\mathbf{Z}, L_i, L_j)},\end{aligned}\quad (14)$$

which consists of E_{unary} , E_{LSAN} and E_{GSSN} . Next, we introduce the designed potentials for the MCD problem.

3.1. Unary potential

The unary potential considers the probability of a superpixel being labeled as changed or unchanged, ignoring the influence of other superpixels. Therefore, some discriminative classifiers including support vector machine (SVM), FCM, and logistic regression, can be used for the unary potential. In the MCD, it is difficult to estimate changes with a fixed distribution model because MCD involves images with different modalities. The soft clustering method of FCM is suitable for the unary potential without prior assumption of the distribution of changed/unchanged class.

Specifically, FCM is employed on the change matrices $f^{\mathbf{X}}$ and $f^{\mathbf{Y}}$ separately to minimize the following objective functions

$$\begin{aligned}\mathcal{E}^{\mathbf{X}} &= \sum_{c=0}^{N_S} \sum_{i=1}^{N_S} (u_{i,c}^{\mathbf{X}})^m \|f_i^{\mathbf{X}} - v_c^{\mathbf{X}}\|_2^2, \\ \mathcal{E}^{\mathbf{Y}} &= \sum_{c=0}^{N_S} \sum_{i=1}^{N_S} (u_{i,c}^{\mathbf{Y}})^m \|f_i^{\mathbf{Y}} - v_c^{\mathbf{Y}}\|_2^2,\end{aligned}\quad (15)$$

where $v_c^{\mathbf{X}} \in \mathbb{R}^{3C_{\mathbf{X}}}$ (and $v_c^{\mathbf{Y}} \in \mathbb{R}^{3C_{\mathbf{Y}}}$) is the center of the c -th cluster, $u_{i,c}^{\mathbf{X}} \in [0, 1]$ (and $u_{i,c}^{\mathbf{Y}} \in [0, 1]$) is the membership grade of i -th superpixel in cluster c (i.e., $c = 0$ represents the unchanged class, and $c = 1$ represents the changed class) with the constraint $\sum_{c=0}^1 u_{i,c}^{\mathbf{X}} = 1$

(and $\sum_{c=0}^1 u_{i,c}^{\mathbf{Y}} = 1$), $m > 1$ is the exponent for the fuzzy partition matrix that controls the amount of fuzzy overlap between clusters, with larger values indicating a greater degree of overlap.

Once the membership values of $u_{i,c}^{\mathbf{X}}$ and $u_{i,c}^{\mathbf{Y}}$ for each superpixel are calculated by N_{fcm} iteration, the unary potential can be con-

structed as

$$\varphi_i(\mathbf{Z}, L_i) = \sum_{c=0}^1 \delta(L_i = c) (-\log(u_{i,c}^{\mathbf{X}}) - \log(u_{i,c}^{\mathbf{Y}})), \quad (16)$$

where $\delta(\cdot)$ being 1 if the specified condition inside parenthesis holds, and 0 otherwise.

3.2. LSAN based pairwise potential

Based on the reasonable assumption that spatially adjacent nodes may be very similar and thus should share the same label with high probability, the LSAN is constructed by taking the contextual information into consideration, which can reduce the salt-and-pepper noise in the CM. First, an R -adjacency neighbor system is constructed for LSAN, that is, if two superpixels (located at Λ_i and Λ_j) intersect or the spatial distance between their center points is less than R , these two superpixels are marked as the R -adjacency neighbors of each other and denoted as $i \in N_j^{LSAN}$ (or $j \in N_i^{LSAN}$). Because the average size of superpixel generated by the SLIC segmentation is around HW/N_S , we set $R = 2\sqrt{HW/N_S}$ for simplicity.

The commonly used pairwise potential for CD usually takes into account both spatial and spectral information, similar to the following form

$$\phi_{ij}(\mathbf{Z}, L_i, L_j) = \frac{\delta(L_i \neq L_j)}{d(\Lambda_i, \Lambda_j)} \exp\left(-\frac{D_{i,j}^{\mathbf{X}}}{2\sigma_{\mathbf{X}}^2}\right) \exp\left(-\frac{D_{i,j}^{\mathbf{Y}}}{2\sigma_{\mathbf{Y}}^2}\right), \quad (17)$$

where $\sigma_{\mathbf{X}}^2$ and $\sigma_{\mathbf{Y}}^2$ are two normalization parameters, $d(\Lambda_i, \Lambda_j)$ is the Euclidean spatial distance between the i -th and j -th superpixels.

However, this kind of pairwise potential in (17) does not consider the specificity of MCD task. For example, when the i -th superpixel \mathbf{X}_i and j -th superpixel \mathbf{X}_j in the pre-event image are very similar (i.e., $D_{i,j}^{\mathbf{X}}$ is very small), but the i -th superpixel \mathbf{Y}_i and the j -th superpixel \mathbf{Y}_j in the post-event image are very different (i.e., $D_{i,j}^{\mathbf{Y}}$ is very large), then the probability that labels of L_i and L_j are different should be high, and $\phi_{ij}(\mathbf{Z}, L_i, L_j)$ should be small. And as the similarity between \mathbf{X}_i and \mathbf{X}_j becomes stronger and the difference between \mathbf{Y}_i and \mathbf{Y}_j becomes greater, the probability that the labels of L_i and L_j are different should be higher, and $\phi_{ij}(\mathbf{Z}, L_i, L_j)$ should be smaller. However, it is clear that the cost of (17) does not satisfy this requirement.

Here, we propose a novel LSAN based pairwise potential for CRF, which takes into account not only the spatial continuity, but also the similarity relationship of the original multitemporal images in the MCD task. It is defined as

$$\phi_{ij}(\mathbf{Z}, L_i, L_j) = \frac{\delta(L_i \neq L_j)}{d(\Lambda_i, \Lambda_j)} \times \begin{cases} \exp\left(-\frac{D_{i,j}^{\mathbf{X}}}{2\sigma_{\mathbf{X}}^2}\right) \exp\left(-\frac{D_{i,j}^{\mathbf{Y}}}{2\sigma_{\mathbf{Y}}^2}\right); & \text{if } D_{i,j}^{\mathbf{X}} \leq \sigma_{\mathbf{X}}^2, D_{i,j}^{\mathbf{Y}} \leq \sigma_{\mathbf{Y}}^2 \\ \exp\left(\frac{D_{i,j}^{\mathbf{X}}}{2\sigma_{\mathbf{X}}^2} - 1\right) \exp\left(-\frac{D_{i,j}^{\mathbf{Y}}}{2\sigma_{\mathbf{Y}}^2}\right); & \text{if } D_{i,j}^{\mathbf{X}} \leq \sigma_{\mathbf{X}}^2, D_{i,j}^{\mathbf{Y}} > \sigma_{\mathbf{Y}}^2 \\ \exp\left(-\frac{D_{i,j}^{\mathbf{X}}}{2\sigma_{\mathbf{X}}^2}\right) \exp\left(\frac{D_{i,j}^{\mathbf{Y}}}{2\sigma_{\mathbf{Y}}^2} - 1\right); & \text{if } D_{i,j}^{\mathbf{X}} > \sigma_{\mathbf{X}}^2, D_{i,j}^{\mathbf{Y}} \leq \sigma_{\mathbf{Y}}^2 \\ \exp(-1); & \text{if } D_{i,j}^{\mathbf{X}} > \sigma_{\mathbf{X}}^2, D_{i,j}^{\mathbf{Y}} > \sigma_{\mathbf{Y}}^2 \end{cases}, \quad (18)$$

where the normalization parameters of $\sigma_{\mathbf{X}}^2$ and $\sigma_{\mathbf{Y}}^2$ are set to

be $\sigma_{\mathbf{X}}^2 = \frac{\sum_{i=1}^{N_S} \sum_{j \in N_i^{LSAN}} D_{i,j}^{\mathbf{X}}}{\sum_{i=1}^{N_S} \text{card}(N_i^{LSAN})}$ and $\sigma_{\mathbf{Y}}^2 = \frac{\sum_{i=1}^{N_S} \sum_{j \in N_i^{LSAN}} D_{i,j}^{\mathbf{Y}}}{\sum_{i=1}^{N_S} \text{card}(N_i^{LSAN})}$, representing the average neighborhood feature difference over the whole image. From (18) and the LSAN based energy function $E_{LSAN} = \sum_{i \in \mathcal{I}} \sum_{j \in N_i^{LSAN}} \phi_{ij}(\mathbf{Z}, L_i, L_j)$, we can find that E_{LSAN} gives a penalty for the discontinuity of $L_i \neq L_j$ in four cases, as shown in Fig. 2. **Case**

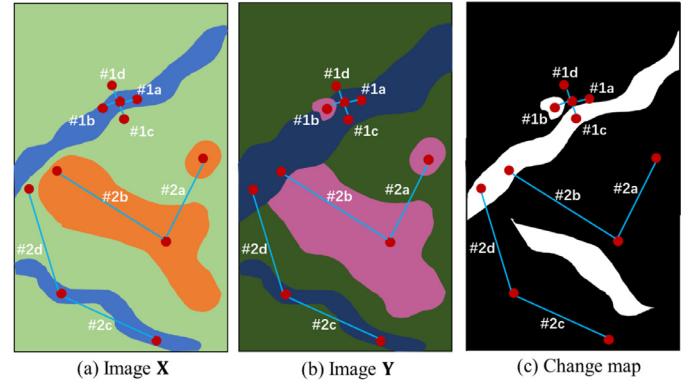


Fig. 2. Illustrations of LSAN and GSSN. Four cases in LSAN: #1a, #1b, #1c, and #1d. Four cases in GSSN: #2a, #2b, #2c, and #2d.

#1a: a large value when \mathbf{X}_i and \mathbf{X}_j , \mathbf{Y}_i and \mathbf{Y}_j are both similar; **Case #1b:** a small value when \mathbf{X}_i and \mathbf{X}_j are similar but \mathbf{Y}_i and \mathbf{Y}_j are dissimilar; **Case #1c:** a small value when \mathbf{X}_i and \mathbf{X}_j are dissimilar but \mathbf{Y}_i and \mathbf{Y}_j are similar; **Case #1d:** a median value when \mathbf{X}_i and \mathbf{X}_j , \mathbf{Y}_i and \mathbf{Y}_j are both dissimilar, which means that the i -th and j -th superpixels are not spectrally closely related to each other, so the relationship between their labels is also ambiguous.

3.3. GSSN based pairwise potential

For each superpixel in the image, its neighbors include not only neighbors on the local space (the R -adjacency neighbors), but also spectrally similar neighbors (the K NNs). Therefore, in the proposed CRF model, we constructs the interaction between the i -th node and its spectrally similar neighborhood.

By using the K -adaptive graphs of $G_{\mathbf{X}(a)}$ and $G_{\mathbf{Y}(a)}$, the spectrally similar neighborhood N_i^{GSSN} for the i -th nodes can be defined as $N_i^{GSSN} = N_i^{k_{\mathbf{X}_i}} \cup N_i^{k_{\mathbf{Y}_i}}$, which means that j is a spectrally similar neighbor of i (i.e., $j \in N_i^{GSSN}$) as long as $(\mathbf{X}_i, \mathbf{X}_j)$ is an edge of $G_{\mathbf{X}(a)}$ (i.e., $A_{i,j}^{\mathbf{X}(a)} = 1$) or $(\mathbf{Y}_i, \mathbf{Y}_j)$ is an edge of $G_{\mathbf{Y}(a)}$ (i.e., $A_{i,j}^{\mathbf{Y}(a)} = 1$). Fig. 2 illustrates the constraints between nodes in GSSN.

1) We first investigate the relationship between the i -th node and its neighbors $j \in N_i^{k_{\mathbf{X}_i}}$. As \mathbf{X}_j belongs to the truly similar neighbors of \mathbf{X}_i by using the K -adaptive strategy, we can assume that \mathbf{X}_j and \mathbf{X}_i represent the same kind of object. **Case #2a:** if \mathbf{Y}_j and \mathbf{Y}_i also represent the same kind of object (\mathbf{Y}_j and \mathbf{Y}_i are very

similar), we have that the i -th node and j -th node should belong to the same class label of unchanged or changed, that is, the probability of $L_i = L_j$ is higher. **Case #2b:** on the contrary, if \mathbf{Y}_j and \mathbf{Y}_i are dissimilar (they may represent different kinds of objects), we have that the i -th node and j -th node may belong to the different class labels, that is, the probability of $L_i = L_j$ is lower.

2) Second, for the i -th node and its neighbors $j \in \mathcal{N}_{\mathbf{Y}_i}^{k_{\mathbf{Y}_i}}$, we have a similar deduction. **Case #2c:** if \mathbf{X}_j and \mathbf{X}_i are very similar, then the i -th node and j -th node should belong to the same class label (i.e., $L_i = L_j$). **Case #2d:** on the contrary, if \mathbf{X}_j and \mathbf{X}_i are dissimilar, then the i -th node and j -th node may belong to the different class labels (i.e., $L_i \neq L_j$).

Based on these constraints, we construct the GSSN based pairwise potential $\psi_{ij}(\mathbf{Z}, L_i, L_j)$ as follows

$$\psi_{ij}(\mathbf{Z}, L_i, L_j) = \begin{cases} \omega_{ij}^{\mathbf{Y}} \delta(L_i \neq L_j) + \varpi_{ij}^{\mathbf{Y}} \delta(L_i = L_j); \\ \omega_{ij}^{\mathbf{X}} \delta(L_i \neq L_j) + \varpi_{ij}^{\mathbf{X}} \delta(L_i = L_j); \\ (\omega_{ij}^{\mathbf{X}} + \omega_{ij}^{\mathbf{Y}}) \delta(L_i \neq L_j) + (\varpi_{ij}^{\mathbf{X}} + \varpi_{ij}^{\mathbf{Y}}) \delta(L_i = L_j); \end{cases}$$

where the auxiliary functions of $\omega_{ij}^{\mathbf{Y}}$, $\omega_{ij}^{\mathbf{X}}$, $\varpi_{ij}^{\mathbf{Y}}$ and $\varpi_{ij}^{\mathbf{X}}$ are defined as

$$\begin{aligned} \omega_{ij}^{\mathbf{Y}} &= \begin{cases} \exp\left(-\frac{D_{i,j}^{\mathbf{Y}}}{\rho_{\mathbf{Y}_i}^2 + \rho_{\mathbf{Y}_j}^2}\right); & \text{if } D_{i,j}^{\mathbf{Y}} \leq \frac{\rho_{\mathbf{Y}_i}^2 + \rho_{\mathbf{Y}_j}^2}{2} \\ \exp\left(-\frac{1}{2}\right); & \text{otherwise} \end{cases}, \\ \omega_{ij}^{\mathbf{X}} &= \begin{cases} \exp\left(-\frac{D_{i,j}^{\mathbf{X}}}{\rho_{\mathbf{X}_i}^2 + \rho_{\mathbf{X}_j}^2}\right); & \text{if } D_{i,j}^{\mathbf{X}} \leq \frac{\rho_{\mathbf{X}_i}^2 + \rho_{\mathbf{X}_j}^2}{2} \\ \exp\left(-\frac{1}{2}\right); & \text{otherwise} \end{cases}, \\ \varpi_{ij}^{\mathbf{Y}} &= \begin{cases} 2 \exp\left(-\frac{1}{2}\right) - \exp\left(-\frac{D_{i,j}^{\mathbf{Y}}}{\rho_{\mathbf{Y}_i}^2 + \rho_{\mathbf{Y}_j}^2}\right); & \text{if } D_{i,j}^{\mathbf{Y}} \leq \frac{\rho_{\mathbf{Y}_i}^2 + \rho_{\mathbf{Y}_j}^2}{2} \\ \exp\left(-\frac{1}{2}\right); & \text{otherwise} \end{cases}, \\ \varpi_{ij}^{\mathbf{X}} &= \begin{cases} 2 \exp\left(-\frac{1}{2}\right) - \exp\left(-\frac{D_{i,j}^{\mathbf{X}}}{\rho_{\mathbf{X}_i}^2 + \rho_{\mathbf{X}_j}^2}\right); & \text{if } D_{i,j}^{\mathbf{X}} \leq \frac{\rho_{\mathbf{X}_i}^2 + \rho_{\mathbf{X}_j}^2}{2} \\ \exp\left(-\frac{1}{2}\right); & \text{otherwise} \end{cases}, \end{aligned} \quad (20)$$

with $\rho_{\mathbf{X}_i}^2 = \frac{1}{k_{\mathbf{X}_i}} \sum_{j \in \mathcal{N}_{\mathbf{X}_i}^{k_{\mathbf{X}_i}}} D_{i,j}^{\mathbf{X}}$ and $\rho_{\mathbf{Y}_i}^2 = \frac{1}{k_{\mathbf{Y}_i}} \sum_{j \in \mathcal{N}_{\mathbf{Y}_i}^{k_{\mathbf{Y}_i}}} D_{i,j}^{\mathbf{Y}}$ represents the average feature difference over the K NNs of i -th node in the pre-event image and post-event image, respectively.

In (19) and (20), we use the relationships between the values of $D_{i,j}^{\mathbf{X}}$ and $(\rho_{\mathbf{X}_i}^2 + \rho_{\mathbf{X}_j}^2)/2$, $D_{i,j}^{\mathbf{Y}}$ and $(\rho_{\mathbf{Y}_i}^2 + \rho_{\mathbf{Y}_j}^2)/2$ to evaluate whether \mathbf{X}_i and \mathbf{X}_j , \mathbf{Y}_i and \mathbf{Y}_j are similar, respectively. For the case #2a, when $D_{i,j}^{\mathbf{Y}} \leq (\rho_{\mathbf{Y}_i}^2 + \rho_{\mathbf{Y}_j}^2)/2$, we consider that \mathbf{Y}_i and \mathbf{Y}_j belong to the same kind of object, and then $\omega_{ij}^{\mathbf{Y}}$ is larger than $\exp(-1/2)$, and $\varpi_{ij}^{\mathbf{Y}}$ is smaller than $\exp(-1/2)$, resulting in a larger penalty for $L_i \neq L_j$ in $\psi_{ij}(\mathbf{Z}, L_i, L_j)$ for the case #2a. However, for the case #2b, there is an ambiguity in measuring the dissimilarity, that is, we can infer that \mathbf{Y}_i and \mathbf{Y}_j most probably belong to the same kind of object when $D_{i,j}^{\mathbf{Y}}$ is smaller than $(\rho_{\mathbf{Y}_i}^2 + \rho_{\mathbf{Y}_j}^2)/2$, but we cannot directly conclude that \mathbf{Y}_i and \mathbf{Y}_j belong to different kinds of objects when $D_{i,j}^{\mathbf{Y}}$ is greater than $(\rho_{\mathbf{Y}_i}^2 + \rho_{\mathbf{Y}_j}^2)/2$. In this case, we assign the same value to $\omega_{ij}^{\mathbf{Y}} = \exp(-1/2)$ and $\varpi_{ij}^{\mathbf{Y}} = \exp(-1/2)$ due to this ambiguity, resulting an equal penalty for $L_i \neq L_j$ and $L_i = L_j$ in $\psi_{ij}(\mathbf{Z}, L_i, L_j)$ for the case #2b. In the same way, for cases the of #2c and #2d, we have the same analysis.

By substituting the equations of $\delta(L_i = L_j) + \delta(L_i \neq L_j) = 1$ and $\omega_{ij}^{\mathbf{X}} + \varpi_{ij}^{\mathbf{X}} = \omega_{ij}^{\mathbf{Y}} + \varpi_{ij}^{\mathbf{Y}} = 2 \exp(-1/2)$ into (19), and eliminating the constant irrelevant items (not relevant to L_i and L_j), the GSSN

based pairwise potential $\psi_{ij}(\mathbf{Z}, L_i, L_j)$ of (19) can be simplified as

$$\psi_{ij}(\mathbf{Z}, L_i, L_j) = \delta(L_i \neq L_j)$$

$$\times \begin{cases} 2\left(\omega_{ij}^{\mathbf{Y}} - \exp\left(-\frac{1}{2}\right)\right); & \text{if } j \in \mathcal{N}_{\mathbf{X}_i}^{k_{\mathbf{X}_i}} \setminus \mathcal{N}_{\mathbf{Y}_i}^{k_{\mathbf{Y}_i}} \\ 2\left(\omega_{ij}^{\mathbf{X}} - \exp\left(-\frac{1}{2}\right)\right); & \text{if } j \in \mathcal{N}_{\mathbf{Y}_i}^{k_{\mathbf{Y}_i}} \setminus \mathcal{N}_{\mathbf{X}_i}^{k_{\mathbf{X}_i}} \\ 2\left(\omega_{ij}^{\mathbf{X}} + \omega_{ij}^{\mathbf{Y}} - 2 \exp\left(-\frac{1}{2}\right)\right); & \text{if } j \in \mathcal{N}_{\mathbf{X}_i}^{k_{\mathbf{X}_i}} \cap \mathcal{N}_{\mathbf{Y}_i}^{k_{\mathbf{Y}_i}} \end{cases} \quad (21)$$

$$\begin{aligned} &\text{if } j \in \mathcal{N}_{\mathbf{X}_i}^{k_{\mathbf{X}_i}} \setminus \mathcal{N}_{\mathbf{Y}_i}^{k_{\mathbf{Y}_i}} \\ &\text{if } j \in \mathcal{N}_{\mathbf{Y}_i}^{k_{\mathbf{Y}_i}} \setminus \mathcal{N}_{\mathbf{X}_i}^{k_{\mathbf{X}_i}}, \\ &\text{if } j \in \mathcal{N}_{\mathbf{X}_i}^{k_{\mathbf{X}_i}} \cap \mathcal{N}_{\mathbf{Y}_i}^{k_{\mathbf{Y}_i}} \end{aligned} \quad (19)$$

3.4. Inference by the graph cut algorithm

By substituting the unary potential φ_i of (16), LSAN based ϕ_{ij} of (18) and GSSN based ψ_{ij} of (21) into (14), we can obtain the Gibbs energy function $E(\mathbf{L}|\mathbf{Z})$ of the proposed CRF model. Then the maximum a posteriori (MAP) labeling of the CRF is given by

$$\mathbf{L}^* = \arg \max_{\mathbf{L}} P(\mathbf{L}|\mathbf{Z}) = \arg \min_{\mathbf{L}} E(\mathbf{L}|\mathbf{Z}). \quad (22)$$

The energy minimization of (22) can be solved efficiently by using the graph cuts. In the proposed IST-CRF, the min-cut/max-flow algorithm [52] is selected to solve the energy minimization problem. Once the optimal \mathbf{L}^* is calculated, we can obtain the changed index subset $\mathcal{T} = \{i | L_i^* = 1; i \in \mathcal{I}\}$ and the unchanged index subset $\mathcal{S} = \{i | L_i^* = 0; i \in \mathcal{I}\}$, which should be propagated back to the structure transformation process in the next iteration.

Finally, the binary CM can be computed as

$$\text{CM}(h, w) = L_i^*, \quad (h, w) \in \Lambda_i. \quad (23)$$

The overall framework of the proposed IST-CRF is summarized in [Algorithm 1](#), which uses an iterative framework to combine the

Algorithm 1 IST-CRF.

Input: Images of \mathbf{X} and \mathbf{Y} , parameters of N_S , N_{Iter} , α and β .
Preprocessing: Superpixel segmentation and feature extraction.
 Implement the Superpixel co-segmentation to obtain Λ .
 Extract the features to obtain the feature matrices $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$.

Main iteration loop of IST-CRF:

Set initial index subset as $\mathcal{S}^0 = \{1, 2, \dots, N_S\}$.
 for $i = 1, 2, \dots, N_{\text{Iter}}$ do

1. **Structure transformation:**

Construct $G_{\mathbf{X}(a)}$, $G_{\mathbf{X}(a-r)}$, $G_{\mathbf{Y}(a)}$ and $G_{\mathbf{Y}(a-r)}$ with \mathcal{S}^{i-1} .
 Implement the structure transformations of μ^{fw} , ϑ^{fw} , μ^{bw} , ϑ^{bw} .
 Calculate the change vectors of $f^{\mathbf{X}}$ and $f^{\mathbf{Y}}$.

2. CRF segmentation:

Calculate the potentials of φ_i , ϕ_{ij} and ψ_{ij} .
 Solve the CRF model to obtain \mathcal{S}^i and \mathcal{T}^i .
 End for

Output: Compute the final change map.

structure transformation and CRF segmentation.

4. Experiments and discussions

In this section, we first test the proposed IST-CRF on different real data sets, and compare the IST-CRF with some SOTA methods.

Table 1

Description of the five heterogeneous data sets.

Dataset	Sensor	Size (pixels)	Date	Location	Event (& Spatial resolution)
#1	Landsat-5/Google Earth	300 × 412 × 1(3)	Sept. 1995 - July 1996	Sardinia, Italy	Lake expansion (30m.)
#2	Pleiades/WorldView2	2000 × 2000 × 3(3)	May 2012 - July 2013	Toulouse, France	Construction (0.52m.)
#3	Landsat-5/Landsat-8	1534 × 808 × 7(10)	Aug. 2011 - June 2013	Texas, USA	Forest fire (30m.)
#4	Radarsat-2/Google Earth	593 × 921 × 1(3)	June 2008 - Sept. 2012	Shuguang Village, China	Building construction (8m.)
#5	Landsat-8/Sentinel-1A	875 × 500 × 11(3)	Jan. 2017 - Feb. 2017	Sutter County, USA	Flooding (\approx 15m.)

Then, some detailed discussions about the parameters and computational complexity are provided.

4.1. Data set description and quantitative evaluation metrics

Five heterogeneous data sets are used to verify the effectiveness of the proposed MCD method, as listed in **Table 1**. These data sets contain different types of modalities: multisensor (cross-sensor) images, e.g., images obtained from Landsat-5 (near-infrared band) and Google Earth (RGB bands) in Dataset #1, images obtained from Pleiades and WorldView2 in Dataset #2, and images obtained from Landsat-5 and Landsat-8 in Dataset #3¹; multisource images, e.g., images obtained from Radarsat-2 and Google Earth in Dataset #4, and images obtained from Landsat-8 and Sentinel-1A in Dataset #5². These data sets also contain different resolutions (varying from 0.52 to 30m), different image sizes (varying from 300 to 2000 pixels in width or length) and different change types (such as flooding, fire, and construction), which can test the performance of MCD methods under different conditions.

To assess the performance of IST-CRF, two types of criteria are employed. First, we evaluate the DI by the receiver operating characteristic (ROC) curve and the precision-recall (PR) curve, which are plotted by using the true positive rate (TPR, also known as the recall rate) versus the false positive rate (FPR), and the precision rate versus the recall rate, respectively. The areas under the ROC curve and PR curve are called AUR and AUP, respectively. Second, we evaluate the final CM by the overall accuracy (OA), Kappa coefficient (Kc), and F1-measure (Fm). These criteria are computed as $TPR = Recall = TP/(TP + FN)$, $FPR = FP/(TN + FP)$, Precision = $TP/(TP + FP)$, OA = $(TP + TN)/(TP + TN + FP + FN)$, Fm = $(2 \times Precision \times Recall) / (Precision + Recall)$, and $Kc = (OA - PRE) / (1 - PRE)$ with

$$PRE = \frac{(TP+FN)(TP+FP)+(TN+FP)(TN+FN)}{(TP+TN+FP+FN)^2}, \quad (24)$$

where TP, FP, TN, and FN represent the true positives, false positives, true negatives, and false negatives, respectively.

4.2. Experimental results

As reported in [Algorithm 1](#), the main parameters of IST-CRF are the number of superpixels N_S , the maximum number of iterations N_{iter} , the balance parameters of α and β . For all the experiments, we fix $N_S = 5000$, $N_{iter} = 7$, vary $\alpha, \beta \in \{1, 3, 5, 7, 9, 11, 13\}$ and select the best one as the result. We will analyze the impact of these parameters in detail in subsection IV-C.

[Figs. 3 \(d\)-\(f\)](#) show the gray images of \tilde{DI}^X , \tilde{DI}^Y , \tilde{DI}^{fuse} generated by IST-CRF on different data sets, which are computed by $\tilde{DI}^X(h, w) = \|DI^X(h, w, :) \|_2$, $\tilde{DI}^Y(h, w) = \|DI^Y(h, w, :) \|_2$, and $\tilde{DI}^{fuse}(h, w) = \tilde{DI}^X(h, w) / mean(\tilde{DI}^X) + \tilde{DI}^Y(h, w) / mean(\tilde{DI}^Y)$, respectively. Some outliers in these DIs are smoothed for better

Table 2

AUR and AUP of DIs generated by IST-CRF on the heterogeneous data sets.

Measures	DIs	Datasets				
		#1	#2	#3	#4	#5
AUR	\tilde{DI}^X	0.860	0.751	0.963	0.908	0.907
	\tilde{DI}^Y	0.897	0.713	0.942	0.983	0.867
	\tilde{DI}^{fuse}	0.891	0.756	0.959	0.978	0.912
AUP	\tilde{DI}^X	0.521	0.475	0.848	0.402	0.403
	\tilde{DI}^Y	0.522	0.383	0.678	0.797	0.174
	\tilde{DI}^{fuse}	0.622	0.453	0.823	0.766	0.332

Table 3

Quantitative measures of binary CMs generated by IST-CRF on the heterogeneous data sets.

Datasets	TN	TP	FP	FN	OA	Kc	Fm
#1	0.926	0.047	0.013	0.015	0.973	0.758	0.773
#2	0.824	0.073	0.025	0.079	0.897	0.529	0.585
#3	0.907	0.069	0.006	0.019	0.976	0.836	0.849
#4	0.950	0.037	0.004	0.009	0.987	0.848	0.855
#5	0.931	0.021	0.025	0.023	0.952	0.447	0.472
Average	-	-	-	-	0.957	0.683	0.706

display. As can be seen from [Figs. 3\(d\)-\(f\)](#), the DIs obtained by IST-CRF are able to highlight the changes very well, which demonstrates the effectiveness of the structure transformation. [Fig. 4](#) plots the ROC curves and PR curves of these DIs, and [Table 2](#) lists the corresponding AUR and AUP. From [Figs. 3–4](#) and [Table 2](#), we can see the robustness of the structure transformation, which can establish connections between multimodal images and be applied to different types of modalities. On the other hand, by comparing the DIs of each data set, we can also find that there are differences between \tilde{DI}^X and \tilde{DI}^Y , due to the fact that they are computed on different domains as illustrated by [\(10\)](#), [\(11\)](#). Therefore, it is necessary to fuse these two DIs for two reasons: first, since we do not know in advance the changed areas in practice, it is impossible to judge which DI is better (highlighting the changes better); second, these two DIs contain complementary change information, and accurate fusion can improve the CD performance. However, directly fusing these \tilde{DI}^X and \tilde{DI}^Y from different domains may not work very well. For example, in Datasets #3 and #4, the fused \tilde{DI}^{fuse} performs worse than \tilde{DI}^X and \tilde{DI}^Y respectively, resulting in a smaller AUR and AUP in [Table 2](#). Therefore, in order to avoid the information loss caused by direct fusion of \tilde{DI}^X and \tilde{DI}^Y , we use the CRF model to incorporate them into the observation field, and fuse them in the segmentation process instead of the commonly used fusion first and then segmentation.

[Fig. 3 \(g\)](#) shows the binary CM obtained by IST-CRF on all the data sets, and [Table 3](#) lists the corresponding criteria of TN, TP, FP, FN, OA, Kc and Fm. From [Fig. 3\(g\)](#) and [Table 3](#), we can see that the proposed IST-CRF can detect the changed region very well under different data sets with different types of modalities, which demonstrates the effectiveness of the IST-CRF. The average OA, Kc and Fm obtained by IST-CRF on these five heterogeneous data sets are about 0.957, 0.683, and 0.706, respectively.

¹ Dataset #3 is provided by Volpi et al. [\[22\]](#) and made available at <https://sites.google.com/site/michelevolpiresearch/codes/cross-sensor>.

² Dataset #5 is provided by Luppino et al. [\[12\]](#) and made available at <https://sites.google.com/view/luppino>.

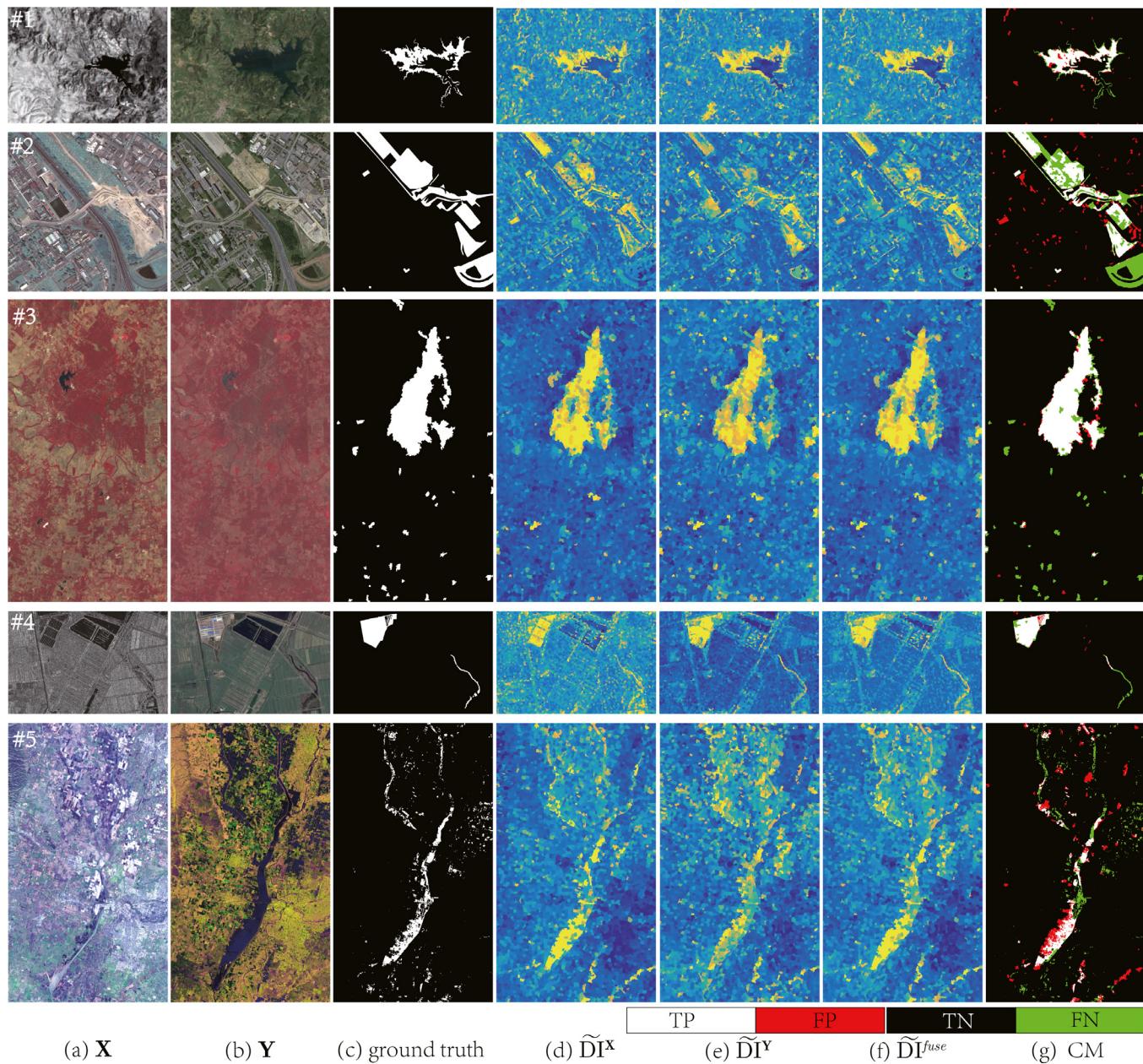


Fig. 3. DIS and binary CMs of IST-CRF on heterogeneous data sets. From top to bottom, they correspond to Datasets #1 to #5, respectively. From left to right are: (a) pre-event image; (b) post-event image; (c) the ground truth; (d) $\widetilde{\mathbf{DI}}^X$; (e) $\widetilde{\mathbf{DI}}^Y$; (f) $\widetilde{\mathbf{DI}}^{fuse}$; (g) binary CM of IST-CRF. In the binary CM, White: true positives (TP); Red: false positives (FP); Black: true negatives (TN); Green: false negatives (FN).

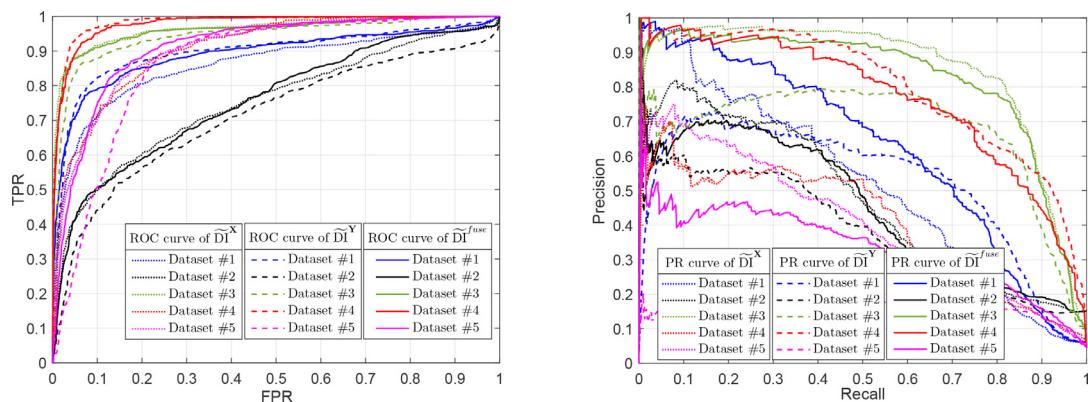


Fig. 4. ROC curves (left) and PR curves (right) of IST-CRF generated DIS.

Table 4

Accuracy rate of CMs generated by different methods on different data sets. The results of these comparison methods are reported by their original published papers, except results indicated with † are reported by [59] (they are consistent with their open source codes in [22]). Italicized and underlined marks are used for deep learning based methods.

Dataset #1	OA
<u>DFR-MT</u> [53]	0.975
<u>CACFL</u> [26]	0.975
IST-CRF	0.972
IRG-McS [43]	0.971
<u>ALSC</u> [54]	0.965
M3CD [42]	0.964
PSGM [13]	0.961
MDS [55]	0.942
<u>AFL-DSR</u> [25]	0.929
FPMS [41]	0.928
RMN [57]	0.847
Dataset #2	OA
IST-CRF	0.897
IRG-McS [43]	0.882
<u>AFL-DSR</u> [25]	0.880
RMN [57]	0.877
M3CD [42]	0.862
NLPEM [56]	0.853
FPMS [41]	0.838
Dataset #3	OA
IST-CRF	0.976
<u>DCCAE</u> [58]	0.943
<u>DCCA</u> [59]	0.939
†KCCA [22]	0.917
†CCA [22]	0.772
Dataset #4	OA
IST-CRF	0.987
<u>DPFL</u> [60]	0.987
<u>X-Net</u> [14]	0.984
IRG-McS [43]	0.983
<u>ACE-Net</u> [14]	0.982
<u>AFL-DSR</u> [25]	0.980
<u>CACFL</u> [26]	0.979
PSGM [13]	0.977
<u>SCCN</u> [32]	0.976
NPSG [45]	0.975
MDS [55]	0.967
<u>LT-FL</u> [23]	0.964
ALSC [54]	0.963
FPMS [41]	0.942
RMN [57]	0.884
Dataset #5	OA
IRG-McS [43]	0.959
IST-CRF	0.952
FPMS [41]	0.952
<u>DPFL</u> [60]	0.945
ALSC [54]	0.944
NPSG [45]	0.941
AMD-IR [12]	0.933
<u>SSL</u> [30]	0.924
<u>ACE-Net</u> [14]	0.915
<u>X-Net</u> [14]	0.911

A comparison with different SOTA approaches is summarized in Table 4, including DFR-MT [53], CACFL [26], IRG-McS [43], ALSC [54], M3CD [42], PSGM [13], MDS [55], AFL-DSR [25], FPMS [41], NLPEM [56], RMN [57], DCCAE [58], DCCA [59], kCCA [22], CCA [22], DPFL [60], X-Net [14], ACE-Net [14], SCCN [32], NPSG [45], LT-FL [23], AMD-IR [12], SSL [30]. Among these comparison approaches, DFR-MT, CACFL, AFL-DSR, DCCAE, DCCA, DPFL, X-Net, ACE-Net, SCCN, LT-FL, and SSL are deep learning based methods. For the sake of fairness, we directly quote the results of the corresponding data sets in their original published papers in Table 4.

As can be seen in Table 4, the IST-CRF consistently yields better or very competitive accuracy by comparing with these SOTA approaches, which again demonstrates the effectiveness of the proposed iterative structure transformation and CRF segmentation based method.

4.3. Discussion

4.3.1. The iterative framework

In order to reduce the influence of changed superpixels on the image transformation, we use an iterative framework to combine the image transformation and image segmentation to eliminate the unstable neighbors (changed superpixels detected by CRF segmentation) in the backward structure transformation of (9). To verify the effectiveness of the iterative framework, we further investigate the process of DI calculation. Figs. 5(c1) and 5(d1) show the proportions of changed superpixels in the K NNs of each superpixel in the graphs of $G_{X(a)}$ and $G_{Y(a)}$, respectively. It is clear that some superpixels in $G_{Y(a)}$ contain a lot of changed superpixels in their NNs, as shown in the middle of Fig. 5(d1). As illustrated in subsection II-B, these changed NNs will make the change vector f_i^X less discriminative, which in turn makes the two DIs (\tilde{D}^X_i , \tilde{D}^Y_i) of the initial iteration perform very differently, as shown in Figs. 5(a2) and 5(b2). Figs. 5(a2)-(d2) show the \tilde{D}^X_i of the initial, second, fourth, and sixth iterations; and Figs. 5(a3)-(d3) show the \tilde{D}^Y_i of these iterations. Fig. 6 plots the ROC curve and PR curve of these DIs generated by IST-CRF with different iterations (iter = 1, 2, 4, 6). From Figs. 5–6, we can find significant improvements in the quality of \tilde{D}^X_i , where their corresponding AUR values are 0.774, 0.844, 0.857, and 0.860, AUP values are 0.148, 0.379, 0.464, and 0.501, respectively. Furthermore, the detection accuracy can be improved as can be seen from Figs. 5(a4)–(d4), where their corresponding Fm values are 0.671, 0.750, 0.758, and 0.771, respectively.

4.3.2. Parameter analysis

As reported in the Algorithm 1, the main parameters of the IST-CRF are the superpixel number N_S , maximum iterations N_{iter} , balance parameters of α and β .

Generally, a large N_S will generate small superpixels, which can improve the detection granularity, but it will also increase the computational complexity. For the N_{iter} , based on our experiments, the most obvious benefit of the iterative framework occurs after the first iteration, and then gradually tends to be stable. Considering the detection accuracy and computational complexity (computing environment), we fix the $N_S = 5000$ and $N_{iter} = 7$ as a compromise choice in our experiments.

For the parameters of α and β , they are used to balance the unary potential based energy E_{unary} , LSAN based energy E_{LSAN} and GSSN based energy E_{GSSN} in the CRF segmentation model (14). A higher value of α indicates a smoother CM, and a higher value of β acquires more consistent results within the similar neighborhood of both images. A sensitivity analysis is conducted to measure the impact of these balance parameters on the IST-CRF. In Fig. 7, we verify α and β from 0 to 14 with an interval of 1. Two remarks can be observed in the Fig. 7: first, the IST-CRF achieves good results for a fairly large range of both α and β as shown in the red area of the Fig. 7, which indicates that the algorithm is certain robust to the balance parameters; second, in these datasets, the OA value at the origin position ($\alpha = \beta = 0$) is always smaller, and the coordinates where the maximum OA values are located are not on the axis, i.e., neither α nor β is 0 at this time, which shows the simultaneous effectiveness of LSAN and GSSN based pairwise potentials. However, how to obtain the optimal α and β is still an unsolved problem. Here we give a relatively effec-

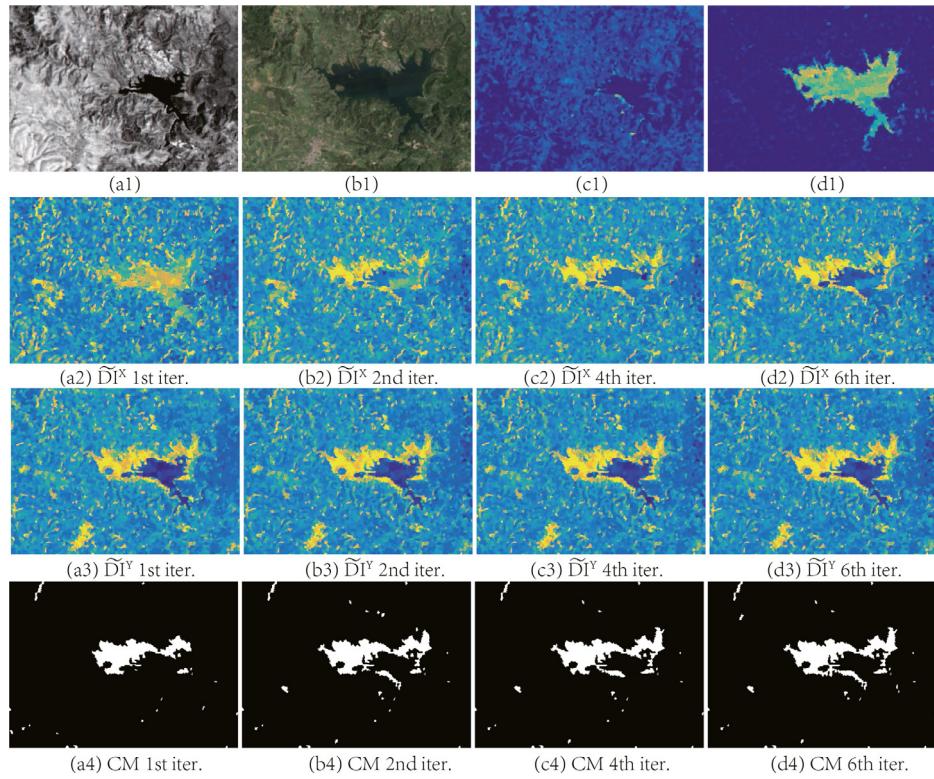


Fig. 5. DIs and binary CMs generated by IST-CRF with different iterations on Dataset #1. (a1)-(b1) the pre-event and post-event images, respectively. (c1)-(d1) the proportions of changed superpixels in $G_{X(a)}$ and $G_{Y(a)}$, respectively. From the second to the fourth row, they correspond to the $\tilde{D}I^X$, $\tilde{D}I^Y$ and binary CM of IST-CRF. From left to right are the results generated by the first iteration (1st iter.), second iteration (2nd iter.), fourth iteration (4th iter.), and sixth iteration (6th iter.) of IST-CRF, respectively.

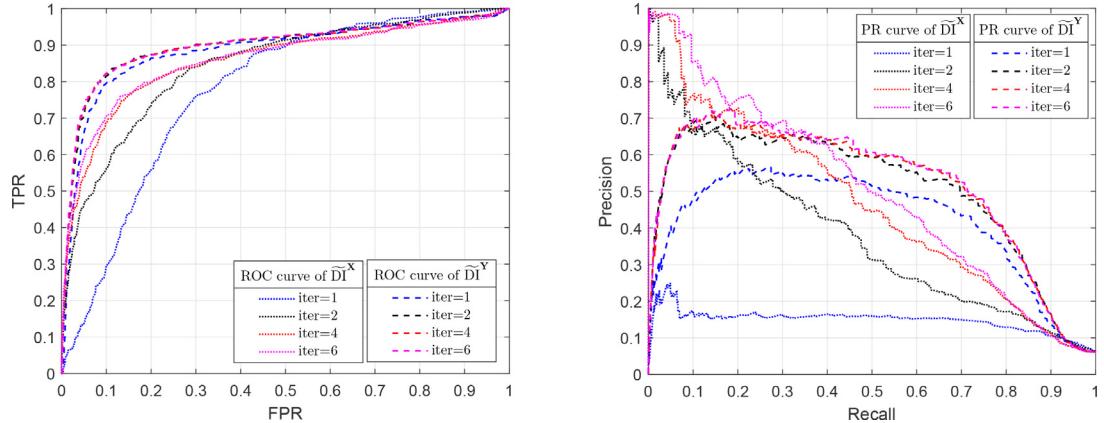


Fig. 6. ROC curves (left) and PR curves (right) of IST-CRF generated DIs with different iterations (iter = 1, 2, 4, 6) on Dataset #1.

tive empirical strategy by setting $\alpha = \frac{\sum_{i \in \mathcal{I}} \varphi_i(\mathbf{Z}, L_i=1)}{\sum_{i \in \mathcal{I}} \sum_{j \in N_i^{LSAN}} \phi_{ij}(\mathbf{Z}, L_i, L_j \neq L_i)}$ and $\beta = \frac{\sum_{i \in \mathcal{I}} \varphi_i(\mathbf{Z}, L_i=1)}{\sum_{i \in \mathcal{I}} \sum_{j \in N_i^{GSSN}} \psi_{ij}(\mathbf{Z}, L_i, L_j \neq L_i)}$ in each iteration, which can usually obtain relatively satisfactory results according to our experience.

4.3.3. Computational time

The main computational complexity of the IST-CRF is concentrating on the pre-processing (superpixel segmentation and feature extraction), structure transformation and CRF segmentation (using min-cut/maxflow algorithm [52]).

Pre-processing. The complexity of SLIC is linear in the number of pixels in the image $\mathcal{O}(HW)$ as reported in [47]. The average number of pixels within each superpixel is HW/N_S , then the complexity of mean and variance feature extraction is

around $\mathcal{O}((C_X + C_Y)HW)$, the median feature extraction is around $\mathcal{O}((C_X + C_Y)HW \log(HW/N_S))$.

Structure transformation. The complexity of calculating distance matrices \mathbf{D}^X and \mathbf{D}^Y is $\mathcal{O}((3C_X + 3C_Y)N_S^2/2)$, the complexity of sorting the distance matrix by column to construct the KNN graphs is $\mathcal{O}(N_S^2 \log N_S)$ by using some accelerated sorting algorithms, such as the Block sort or Tree sort.

CRF segmentation. The complexity of FCM in the calculation of unary potential (15), (16) is $\mathcal{O}(nm^2N_{fcm})$ [61], where $n = N_S$ is the number of data point in change matrices f^X and f^Y , $m = C_X$ (or $m = C_Y$) is the dimension of the data point in f^X (or f^Y), $c = 2$ is the number of clusters (changed/unchanged), and N_{fcm} is the number of iterations in FCM ($N_{fcm} = 100$ in our experiments). The complexity of computing E_{LSAN} and E_{GSSN} are around $\mathcal{O}(N_{LSAN})$ and $\mathcal{O}(N_{GSSN})$ respectively, where N_{LSAN} represents the number of edges

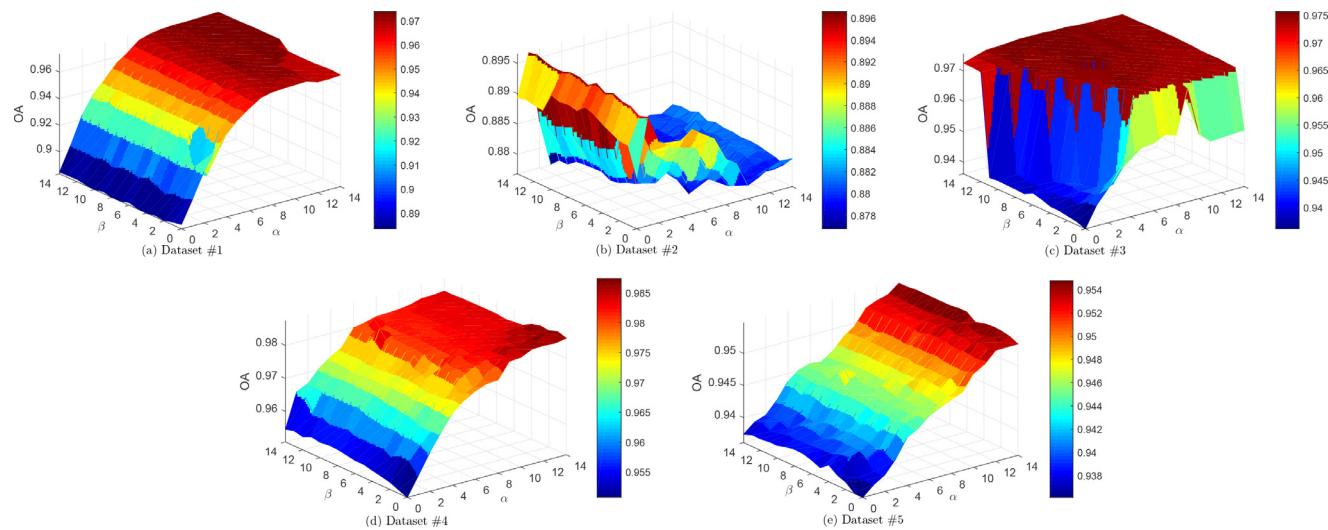


Fig. 7. Sensitivity analysis of parameters α and β in IST-CRF for different data sets: (a) Dataset #1; (b) Dataset #2; (c) Dataset #3; (d) Dataset #4; (e) Dataset #5.

Table 5
Computational time (seconds) of each process of IST-CRF.

Data sets	N_S	t_{pre}	t_{st}	t_{seg}	t_{total}
Dataset #1 $300 \times 412 \times 1(3)$	5000	0.54	2.90	3.21	6.84
	10000	1.00	7.20	8.02	16.57
	20000	2.05	17.60	21.77	42.24
Dataset #2 $2000 \times 2000 \times 3(3)$	5000	3.23	3.08	3.53	10.02
	10000	3.80	7.72	8.34	20.24
	20000	5.27	20.30	23.59	50.16

in the R -adjacency neighbor system for LSAN and N_{GSSN} represents the number of edges in the spectrally similar neighbor system for GSSN. The theoretical complexity and empirical complexity of min-cut/max-flow algorithm have been studied in [52], that is, the theoretical complexity of the worst-case is $\mathcal{O}(2(N_{LSAN} + N_{GSSN})N_S^2)$. However, its empirical complexity is relatively low on typical problem instances in vision, as shown in the examples of [52] and the Table 5.

Table 5 reports the computational time of each process of IST-CRF with different N_S on Datasets #1 and #2, which is performed in MATLAB 2016a running on a Windows Laptop with Intel Core i9-10980HK CPU and 64 GB of RAM. In Table 5, t_{pre} , t_{st} , and t_{seg} represent the computational times spent in the pre-processing, structure transformation, and CRF segmentation respectively, t_{total} represents the total running time of IST-CRF. As can be seen in Table 5, the running time of IST-CRF is influenced by N_S , and IST-CRF is very efficient because it uses superpixel as the basic unit of analysis.

5. Conclusion

In this paper, an image transformation based method named IST-CRF is proposed for the unsupervised MCD problem, which contains three highlights. First, to make the heterogeneous images “comparable”, IST-CRF transforms the multitemporal images into the same differential domain with explicit expressions by using graph based structure transformations. Second, to make full use of the spectral-spatial information, a CRF segmentation model is designed by incorporating the change information based unary potential, LSAN and GSSN based pairwise potentials. Third, to reduce the influence of changes and thus improve the detection accuracy, an iterative framework is used to combine the structure transformation and CRF segmentation. The effectiveness and adaptability

of the proposed method are validated on five real data sets acquired from different modalities.

In the proposed IST-CRF, we first use the graph to capture the structure information and then complete the image transformation. However, the graph is constructed based on a single segmentation scale. In our future work, the multi-scale graph construction and fusion strategy will be further explored to improve the detection accuracy. In addition, we use the coarse-to-refine iterative framework to alleviate the influence of changes on the transformation process, but it also increases the computational time. In the future, we will try to develop an end-to-end model that can directly output the change results to further improve the detection efficiency.

Declaration of Competing Interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled “Iterative Structure Transformation and Conditional Random Field based Method for Unsupervised Multimodal Change Detection”.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61971426, 12171481 and 4210010534, and in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2022JQ-694. The authors would like to thank the editors and anonymous reviewers for their 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* 10 (6) (1989) 989–1003, doi:[10.1080/01431168908903939](https://doi.org/10.1080/01431168908903939).
- [2] N. Longbotham, F. Pacifici, T. Glenn, A. Zare, M. Volpi, D. Tuia, E. Christophe, J. Michel, J. Inglada, J. Chanussot, Q. Du, Multi-modal change detection, application to the detection of flooded areas: outcome of the 2009–2010 data fusion contest, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 5 (1) (2012) 331–342, doi:[10.1109/JSTARS.2011.2179638](https://doi.org/10.1109/JSTARS.2011.2179638).
- [3] Z. Lv, T. Liu, J.A. Benediktsson, N. Falco, Land cover change detection techniques: very-high-resolution optical images: a review, *IEEE Geosci. Remote Sens. Mag.* (2021) 2–21, doi:[10.1109/MGRS.2021.3088865](https://doi.org/10.1109/MGRS.2021.3088865).

- [4] F. Bovolo, L. Bruzzone, A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain, *IEEE Trans. Geosci. Remote Sens.* 45 (1) (2007) 218–236, doi:[10.1109/TGRS.2006.885408](https://doi.org/10.1109/TGRS.2006.885408).
- [5] H. Li, T. Celik, N. Longbotham, W.J. Emery, Gabor feature based unsupervised change detection of multitemporal sar images based on two-level clustering, *IEEE Geosci. Remote Sens. Lett.* 12 (12) (2015) 2458–2462.
- [6] H. Chen, C. Wu, B. Du, L. Zhang, L. Wang, Change detection in multisource VHR images via deep siamese convolutional multiple-layers recurrent neural network, *IEEE Trans. Geosci. Remote Sens.* 58 (4) (2020) 2848–2864, doi:[10.1109/TGRS.2019.2956756](https://doi.org/10.1109/TGRS.2019.2956756).
- [7] Z. Lv, G. Li, Z. Jin, J.A. Benediktsson, G.M. Foody, Iterative training sample expansion to increase and balance the accuracy of land classification from VHR imagery, *IEEE Trans. Geosci. Remote Sens.* 59 (1) (2021) 139–150, doi:[10.1109/TGRS.2020.2996064](https://doi.org/10.1109/TGRS.2020.2996064).
- [8] G. Mercier, G. Moser, S.B. Serpico, Conditional copulas for change detection in heterogeneous remote sensing images, *IEEE Trans. Geosci. Remote Sens.* 46 (5) (2008) 1428–1441, doi:[10.1109/TGRS.2008.916476](https://doi.org/10.1109/TGRS.2008.916476).
- [9] G. Moser, S.N. Anfinsen, L.T. Luppino, S.B. Serpico, Change detection with heterogeneous remote sensing data: From semi-parametric regression to deep learning, in: IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, 2020, pp. 3892–3895, doi:[10.1109/IGARSS39084.2020.9323831](https://doi.org/10.1109/IGARSS39084.2020.9323831).
- [10] Z.-G. Liu, G. Qiu, G. Mercier, Q. Pan, A transfer classification method for heterogeneous data based on evidence theory, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* (2019) 1–13, doi:[10.1109/TSMC.2019.2945808](https://doi.org/10.1109/TSMC.2019.2945808).
- [11] Z. Liu, G. Li, G. Mercier, Y. He, Q. Pan, Change detection in heterogeneous remote sensing images via homogeneous pixel transformation, *IEEE Trans. Image Process.* 27 (4) (2018) 1822–1834, doi:[10.1109/TIP.2017.2784560](https://doi.org/10.1109/TIP.2017.2784560).
- [12] L.T. Luppino, F.M. Bianchi, G. Moser, S.N. Anfinsen, Unsupervised image regression for heterogeneous change detection, *IEEE Trans. Geosci. Remote Sens.* 57 (12) (2019) 9960–9975, doi:[10.1109/TGRS.2019.2930348](https://doi.org/10.1109/TGRS.2019.2930348).
- [13] Y. Sun, L. Lei, X. Li, X. Tan, G. Kuang, Patch similarity graph matrix-based unsupervised remote sensing change detection with homogeneous and heterogeneous sensors, *IEEE Trans. Geosci. Remote Sens.* 59 (6) (2021) 4841–4861, doi:[10.1109/TGRS.2020.3013673](https://doi.org/10.1109/TGRS.2020.3013673).
- [14] L.T. Luppino, M. Kampffmeyer, F.M. Bianchi, G. Moser, S.B. Serpico, R. Jenssen, S.N. Anfinsen, Deep image translation with an affinity-based change prior for unsupervised multimodal change detection, *IEEE Trans. Geosci. Remote Sens.* (2021) 1–22, doi:[10.1109/TGRS.2021.3056196](https://doi.org/10.1109/TGRS.2021.3056196).
- [15] X. Niu, M. Gong, T. Zhan, Y. Yang, A conditional adversarial network for change detection in heterogeneous images, *IEEE Geosci. Remote Sens. Lett.* 16 (1) (2019) 45–49, doi:[10.1109/LGRS.2018.2868704](https://doi.org/10.1109/LGRS.2018.2868704).
- [16] M. Gong, X. Niu, T. Zhan, M. Zhang, A coupling translation network for change detection in heterogeneous images, *Int J Remote Sens.* 40 (9) (2019) 3647–3672.
- [17] X. Jiang, G. Li, Y. Liu, X.-P. Zhang, Y. He, Change detection in heterogeneous optical and SAR remote sensing images via deep homogeneous feature fusion, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 13 (2020) 1551–1566, doi:[10.1109/JSTARS.2020.2983993](https://doi.org/10.1109/JSTARS.2020.2983993).
- [18] J.R. Jensen, E.W. Ramsey, H. Mackey Jr, E. Christensen, R. Sharitz, Inland wetland change detection using aircraft MSS data, *Photogramm Eng Remote Sensing* 53 (5) (1987) 521–529.
- [19] L. Wan, Y. Xiang, H. You, An object-based hierarchical compound classification method for change detection in heterogeneous optical and SAR images, *IEEE Trans. Geosci. Remote Sens.* 57 (12) (2019) 9941–9959, doi:[10.1109/TGRS.2019.2930322](https://doi.org/10.1109/TGRS.2019.2930322).
- [20] Y. Wu, Z. Bai, Q. Miao, W. Ma, Y. Yang, M. Gong, A classified adversarial network for multi-spectral remote sensing image change detection, *Remote Sens. (Basel)* 12 (13) (2020), doi:[10.3390/rs12132098](https://doi.org/10.3390/rs12132098).
- [21] J. Prendes, M. Chabert, F. Pascual, A. Giros, J.-Y. Tourneret, A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors, *IEEE Trans. Image Process.* 24 (3) (2015) 799–812, doi:[10.1109/TIP.2014.2387013](https://doi.org/10.1109/TIP.2014.2387013).
- [22] M. Volpi, G. Camps-Valls, D. Tuia, Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis, *ISPRS J. Photogramm. Remote Sens.* 107 (2015) 50–63, doi:[10.1016/j.isprsjprs.2015.02.005](https://doi.org/10.1016/j.isprsjprs.2015.02.005). Multitemporal remote sensing data analysis
- [23] T. Zhan, M. Gong, X. Jiang, S. Li, Log-based transformation feature learning for change detection in heterogeneous images, *IEEE Geosci. Remote Sens. Lett.* 15 (9) (2018) 1352–1356, doi:[10.1109/LGRS.2018.2843385](https://doi.org/10.1109/LGRS.2018.2843385).
- [24] J. Liu, W. Zhang, F. Liu, L. Xiao, A probabilistic model based on bipartite convolutional neural network for unsupervised change detection, *IEEE Trans. Geosci. Remote Sens.* (2021) 1–14, doi:[10.1109/TGRS.2021.3071347](https://doi.org/10.1109/TGRS.2021.3071347).
- [25] R. Touati, M. Mignotte, M. Dahmane, Anomaly feature learning for unsupervised change detection in heterogeneous images: a deep sparse residual model, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 13 (2020) 588–600, doi:[10.1109/JSTARS.2020.2964409](https://doi.org/10.1109/JSTARS.2020.2964409).
- [26] Y. Wu, J. Li, Y. Yuan, A.K. Qin, Q.-G. Miao, M.-G. Gong, Commonality autoencoder: learning common features for change detection from heterogeneous images, *IEEE Trans. Neural Netw. Learn. Syst.* (2021) 1–14, doi:[10.1109/TNNLS.2021.3056238](https://doi.org/10.1109/TNNLS.2021.3056238).
- [27] Y. Qin, Z. Niu, F. Chen, B. Li, Y. Ban, Object-based land cover change detection for cross-sensor images, *Int J. Remote Sens.* 34 (19) (2013) 6723–6737.
- [28] X. Jiang, G. Li, X.-P. Zhang, Y. He, A semisupervised siamese network for efficient change detection in heterogeneous remote sensing images, *IEEE Trans. Geosci. Remote Sens.* (2021) 1–18, doi:[10.1109/TGRS.2021.3061686](https://doi.org/10.1109/TGRS.2021.3061686).
- [29] H. Li, M. Gong, M. Zhang, Y. Wu, Spatially self-paced convolutional networks for change detection in heterogeneous images, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 14 (2021) 4966–4979, doi:[10.1109/JSTARS.2021.3078437](https://doi.org/10.1109/JSTARS.2021.3078437).
- [30] Y. Chen, L. Bruzzone, Self-supervised change detection in multiview remote sensing images, *IEEE Trans. Geosci. Remote Sens.* 60 (2022) 1–12, doi:[10.1109/TGRS.2021.3089453](https://doi.org/10.1109/TGRS.2021.3089453).
- [31] M. Gong, P. Zhang, L. Su, J. Liu, Coupled dictionary learning for change detection from multisource data, *IEEE Trans. Geosci. Remote Sens.* 54 (12) (2016) 7077–7091, doi:[10.1109/TGRS.2016.2594952](https://doi.org/10.1109/TGRS.2016.2594952).
- [32] J. Liu, M. Gong, K. Qin, P. Zhang, A deep convolutional coupling network for change detection based on heterogeneous optical and radar images, *IEEE Trans. Neural Netw. Learn. Syst.* 29 (3) (2018) 545–559, doi:[10.1109/TNNLS.2016.2636227](https://doi.org/10.1109/TNNLS.2016.2636227).
- [33] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Trans. Syst. Man Cybern.* 9 (1) (1979) 62–66.
- [34] J.A. Hartigan, M.A. Wong, Algorithm as 136: a k-means clustering algorithm, *Journal of the royal statistical society. series c (applied statistics)* 28 (1) (1979) 100–108.
- [35] J.C. Bezdek, R. Ehrlich, W. Full, FCM: the fuzzy c-means clustering algorithm, *Computers & geosciences* 10 (2–3) (1984) 191–203.
- [36] Z. Li, W. Shi, P. Lu, L. Yan, Q. Wang, Z. Miao, Landslide mapping from aerial photographs using change detection-based Markov random field, *Remote Sens. Environ.* 187 (2016) 76–90, doi:[10.1016/j.rse.2016.10.008](https://doi.org/10.1016/j.rse.2016.10.008).
- [37] M. Hao, M. Zhou, J. Jin, W. Shi, An advanced superpixel-based markov random field model for unsupervised change detection, *IEEE Geosci. Remote Sens. Lett.* 17 (8) (2020) 1401–1405, doi:[10.1109/LGRS.2019.2948660](https://doi.org/10.1109/LGRS.2019.2948660).
- [38] L. Zhou, G. Cao, Y. Li, Y. Shang, Change detection based on conditional random field with region connection constraints in high-resolution remote sensing images, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 9 (8) (2016) 3478–3488, doi:[10.1109/JSTARS.2016.2514610](https://doi.org/10.1109/JSTARS.2016.2514610).
- [39] P. Lv, Y. Zhong, J. Zhao, L. Zhang, Unsupervised change detection based on hybrid conditional random field model for high spatial resolution remote sensing imagery, *IEEE Trans. Geosci. Remote Sens.* 56 (7) (2018) 4002–4015, doi:[10.1109/TGRS.2018.2819367](https://doi.org/10.1109/TGRS.2018.2819367).
- [40] S. Shi, Y. Zhong, J. Zhao, P. Lv, Y. Liu, L. Zhang, Land-use/land-cover change detection based on class-prior object-oriented conditional random field framework for high spatial resolution remote sensing imagery, *IEEE Trans. Geosci. Remote Sens.* (2020) 1–16, doi:[10.1109/TGRS.2020.3034373](https://doi.org/10.1109/TGRS.2020.3034373).
- [41] M. Mignotte, A fractal projection and Markovian segmentation-based approach for multimodal change detection, *IEEE Trans. Geosci. Remote Sens.* 58 (11) (2020) 8046–8058, doi:[10.1109/TGRS.2020.2986239](https://doi.org/10.1109/TGRS.2020.2986239).
- [42] R. Touati, M. Mignotte, M. Dahmane, Multimodal change detection in remote sensing images using an unsupervised pixel pairwise-based Markov random field model, *IEEE Trans. Image Process.* 29 (2020) 757–767, doi:[10.1109/TIP.2019.2933747](https://doi.org/10.1109/TIP.2019.2933747).
- [43] Y. Sun, L. Lei, D. Guan, G. Kuang, Iterative robust graph for unsupervised change detection of heterogeneous remote sensing images, *IEEE Trans. Image Process.* 30 (2021) 6277–6291, doi:[10.1109/TIP.2021.3093766](https://doi.org/10.1109/TIP.2021.3093766).
- [44] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3-d transform-domain collaborative filtering, *IEEE Trans. Image Process.* 16 (8) (2007) 2080–2095, doi:[10.1109/TIP.2007.901238](https://doi.org/10.1109/TIP.2007.901238).
- [45] Y. Sun, L. Lei, X. Li, H. Sun, G. Kuang, Nonlocal patch similarity based heterogeneous remote sensing change detection, *Pattern Recognit.* 109 (2021) 107598, doi:[10.1016/j.patcog.2020.107598](https://doi.org/10.1016/j.patcog.2020.107598).
- [46] Y. Sun, L. Lei, X. Li, X. Tan, G. Kuang, Structure consistency-based graph for unsupervised change detection with homogeneous and heterogeneous remote sensing images, *IEEE Trans. Geosci. Remote Sens.* (2021) 1–21, doi:[10.1109/TGRS.2021.3053571](https://doi.org/10.1109/TGRS.2021.3053571).
- [47] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Ssstrunk, Slic superpixels compared to state-of-the-art superpixel methods, *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (11) (2012) 2274–2282, doi:[10.1109/TPAMI.2012.120](https://doi.org/10.1109/TPAMI.2012.120).
- [48] C.-A. Deledalle, L. Denis, F. Tupin, How to compare noisy patches? patch similarity beyond Gaussian noise, *Int. J. Comput. Vis.* 99 (1) (2012) 86–102.
- [49] S. Huang, Z. Kang, I.W. Tsang, Z. Xu, Auto-weighted multi-view clustering via kernelized graph learning, *Pattern Recognit.* 88 (2019) 174–184, doi:[10.1016/j.patcog.2018.11.007](https://doi.org/10.1016/j.patcog.2018.11.007).
- [50] Z. Kang, C. Peng, Q. Cheng, X. Liu, X. Peng, Z. Xu, L. Tian, Structured graph learning for clustering and semi-supervised classification, *Pattern Recognit.* 110 (2021) 107627, doi:[10.1016/j.patcog.2020.107627](https://doi.org/10.1016/j.patcog.2020.107627).
- [51] J.D. Lafferty, A. McCallum, F.C.N. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, in: *ICML*, 2001, pp. 282–289.
- [52] Y. Boykov, V. Kolmogorov, An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (9) (2004) 1124–1137, doi:[10.1109/TPAMI.2004.60](https://doi.org/10.1109/TPAMI.2004.60).
- [53] P. Zhang, M. Gong, L. Su, J. Liu, Z. Li, Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images, *ISPRS J. Photogramm. Remote Sens.* 116 (2016) 24–41, doi:[10.1016/j.isprsjprs.2016.02.013](https://doi.org/10.1016/j.isprsjprs.2016.02.013).
- [54] L. Lei, Y. Sun, G. Kuang, Adaptive local structure consistency-based heterogeneous remote sensing change detection, *IEEE Geosci. Remote Sens. Lett.* (2020) 1–5, doi:[10.1109/LGRS.2020.3037930](https://doi.org/10.1109/LGRS.2020.3037930).
- [55] R. Touati, M. Mignotte, M. Dahmane, Change detection in heterogeneous remote sensing images based on an imaging modality-invariant MDS representation, in: 2018 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 3998–4002, doi:[10.1109/ICIP.2018.8451184](https://doi.org/10.1109/ICIP.2018.8451184).

- [56] R. Touati, M. Mignotte, An energy-based model encoding nonlocal pairwise pixel interactions for multisensor change detection, *IEEE Trans. Geosci. Remote Sens.* 56 (2) (2018) 1046–1058, doi:[10.1109/TGRS.2017.2758359](https://doi.org/10.1109/TGRS.2017.2758359).
- [57] R. Touati, M. Mignotte, M. Dahmane, A reliable mixed-norm-based multiresolution change detector in heterogeneous remote sensing images, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 12 (9) (2019) 3588–3601, doi:[10.1109/JSTARS.2019.2934602](https://doi.org/10.1109/JSTARS.2019.2934602).
- [58] Y. Zhou, H. Liu, D. Li, H. Cao, J. Yang, Z. Li, Cross-sensor image change detection based on deep canonically correlated autoencoders, in: *International Conference on Artificial Intelligence for Communications and Networks*, Springer, 2019, pp. 251–257.
- [59] J. Yang, Y. Zhou, Y. Cao, L. Feng, Heterogeneous image change detection using deep canonical correlation analysis, in: 2018 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 2917–2922, doi:[10.1109/ICPR.2018.8545885](https://doi.org/10.1109/ICPR.2018.8545885).
- [60] M. Yang, L. Jiao, F. Liu, B. Hou, S. Yang, M. Jian, Dpfl-nets: deep pyramid feature learning networks for multiscale change detection, *IEEE Trans Neural Netw Learn Syst* (2021) 1–15, doi:[10.1109/TNNLS.2021.3079627](https://doi.org/10.1109/TNNLS.2021.3079627).
- [61] P. Hore, L.O. Hall, D.B. Goldgof, Single pass fuzzy c means, in: 2007 IEEE International Fuzzy Systems Conference, 2007, pp. 1–7, doi:[10.1109/FUZZY.2007.4295372](https://doi.org/10.1109/FUZZY.2007.4295372).

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

Lin Lei received the Ph.D. degree in Information and Communication Engineering from National University of Defense Technology, Changsha, China, in 2008. She is currently 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 works in the High-Tech Institute of Xi'an. His research interests include synthetic aperture radar image processing, machine learning, and computation vision in remote sensing applications.

Junzheng Wu received the B.S. degree from Tsinghua University, Beijing, China, in 2008, and the M.S. degree from Northwestern Polytechnic University, Xi'an, China, in 2013. Currently, He is working toward the Ph.D. degree in information and communication engineering in the College of Electronic Science, National University of Defense Technology, since 2019. His research interests cover machine learning and remote sensing image processing.

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