

# Sparse-Constrained Adaptive Structure Consistency-Based Unsupervised Image Regression for Heterogeneous Remote-Sensing Change Detection

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

**Abstract**—Change detection of heterogeneous multitemporal satellite images is an important and challenging topic in remote sensing. Since the imaging mechanisms of heterogeneous sensors are different, it is not possible to directly compare heterogeneous images to detect changes as in the homogeneous images. To address this challenge, we propose an unsupervised image regression-based change detection method based on the structure consistency. The proposed method first adaptively constructs a similarity graph to represent the structure of a pre-event image, then uses the graph to translate the pre-event image to the domain of the post-event image, and then computes the difference image. Finally, a superpixel-based Markovian segmentation model is designed to segment the difference image into changed and unchanged classes. The proposed adaptive structure consistency-based image regression model can not only alleviate the impact of noise and changed pixels on the regression process by using the structure-based transformation, but also easily distinguish between changed and unchanged classes in the difference image by using the prior sparse knowledge of changes. Experimental results on six different datasets demonstrate the effectiveness of the proposed method by comparing with some state-of-the-art methods.

**Index Terms**—Graph, heterogeneous data, image regression, sparse regularization, structure consistency, unsupervised change detection.

## I. INTRODUCTION

### A. Background

CHANGE detection (CD) is a technique to recognize changes in imagery acquired on the same geographical location but at different times [1]. CD is a very important topic in satellite remote sensing (RS), which has been found a wide range of applications in damage assessment [2], agricultural surveys [3], and environmental monitoring [4].

Manuscript received May 18, 2021; revised July 27, 2021; accepted August 31, 2021. Date of publication September 17, 2021; date of current version January 21, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61701508. (*Corresponding author: Lin Lei*)

Yuli Sun, Lin Lei, Ming Li, and Gangyao Kuang are with the College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China (e-mail: sunyuli@mail.ustc.edu.cn; aleilei@163.com; liming17@nudt.edu.cn; kuangyeats@hotmail.com).

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

Digital Object Identifier 10.1109/TGRS.2021.3110998

Currently, most of the past CD techniques assume that the satellite images are homogeneous, that is, images are acquired from the same sensor, such as CD of SAR images [5], optical images [6], and hyperspectral images [7]. However, the assumption of homogeneity does not apply to many practical examples and applications, especially when different sensors are involved. This is a significant limitation in emergency response to sudden events, such as earthquakes and floods. In such kind of scenario, the pre-event SAR image is usually unavailable due to the limited coverage and high cost, and the high-quality post-event optical image often cannot be obtained due to the adverse light and weather conditions [2]. Therefore, the CD method based on the heterogeneous images, that is, images are acquired from different sensors, becomes particularly important in these scenarios. On the one hand, heterogeneous CD allows to fully utilize the capabilities of all available sensors. But, on the other hand, it poses additional technical challenges that different sensors measure different physical quantities of the same object, resulting in different characteristics in the heterogeneous images, which makes it impossible to directly compare heterogeneous images to detect changes as in homogeneous CD.

Generally, the existing methods of heterogeneous CD can be classified into supervised and unsupervised according to whether labeled data is required or not and also into pixel-based and object-based according to the basic analysis unit used in image comparison. Since the heterogeneous images cannot be directly compared due to the different imaging mechanisms, the goal of heterogeneous CD method is to transform the “incomparable” images to a common space where they are “comparable.” According to the transformed space, the heterogeneous CD methods can also be roughly classified into three categories.

1) *Classification-Based Method*: These methods transform the heterogeneous images to a common category space by classifying the images separately and then compare the classification results to detect the changes, such as the post-classification comparison method [8], the multitemporal segmentation, and compound classification method (MS-CC) [9], [10]. The classification-based method is intuitive and easy to be implemented, but it has the following two

weaknesses: first, the CD performance depends on the accuracy of image classification, however, precise classification is difficult, especially for SAR images that are affected by severe speckle noise; second, the CD granularity depends on the fineness of image classification, that is, only changes between categories can be detected, while changes within the same category and outside all categories cannot be detected.

2) *Similarity-Based Method*: These methods usually transform the heterogeneous images to a feature space, where similarity measures with assumed imaging-modality-invariant properties are used to measure the change level. They can be further divided into traditional similarity-based and deep learning-based methods. The former manually constructs the connection between heterogeneous images, such as the pixel pair (PP) method [11], sorted histogram distance (SHD) [12], Kullback–Leibler (KL) distance with copula theory [13], manifold distance of local joint distributions [14], and nonlocal pixel pairwise energy-based model (NLPEM) [15]. The latter compares the images in a common latent feature space, which is learned by using the deep neural networks (DNNs) with unchanged samples of heterogeneous images, such as the symmetric convolutional coupling network (SCCN) [16], anomaly feature learning-based deep sparse residual model (AFL-DSR) [17], and logarithmic transformation feature learning network (LT-FL) [18]. For these similarity-based methods, the modal-invariant features are very important, which directly affects the CD performance. Therefore, such methods often face two challenges: first, when the scene is complex or the noise in the image is severe (especially the speckle noise of SAR images), the manually constructed connections of traditional similarity-based methods may no longer applicable; second, when the unchanged samples for training is not enough or the samples are mixed with wrong samples (changed samples), the learned features of deep learning-based methods may no longer characterize the relationship between heterogeneous images, resulting in a dramatic decrease in the CD performance.

3) *Regression-Based Method*: These methods map one image to the domain of the other image. In [19], homogeneous pixel transformation (HPT) uses kernel regression on a sample of K-nearest neighbor (KNN) pixels to set up mappings between the input images. The regression process of HPT is supervised with unchanged training data. To avoid the requirement of the labeled data, an unsupervised affinity matrix-based image regression (AM-IR) is proposed [20], which uses the AMD to identify pixels that are likely to be unchanged as pseudo-training data and then selects four regression methods to carry out the transformation: Gaussian process regression (GPR), support vector regression (SVR), random forest regression (RFR), and the HPT. In addition, some researchers also use the neural network to complete the image translation. In [21], a conditional generative adversarial network (cGAN)-based heterogeneous CD method is proposed, which constrains a cGAN-based network to translate the optical image to the SAR image and an approximation network to transform the original SAR image to the translated one. A coupling translation network (CPTN) is proposed in [22], which uses a coupling variational autoencoder (VAE)

to extract a shared-latent space for images and then uses a coupled GAN to translate the heterogeneous images. In [23], two deep image translation methods with an AMD-based change prior are proposed: the X-Net with two fully convolutional networks and the adversarial cyclic encoders network (ACE-Net) with two autoencoders whose code spaces are aligned by adversarial training. Note that the regression processes of these traditional methods or deep translation networks need to be trained with unchanged pairs of heterogeneous data. Therefore, these regression-based methods either require labeled training set under the supervision mode (such as HPT), the pseudo-training set/change prior to guide the training process (such as AM-IR, X-Net, ACE-Net), or a complex coarse-to-fine filtering process to construct the pseudo-training set (such as cGAN and CPTN).

### B. Motivation

Recently, the structure consistency between heterogeneous images has been explored by researchers and used for heterogeneous CD [24]–[28], which is based on the widespread self-similarity property of the images. With the assumption that heterogeneous images share the same structure information, two similarity-based methods have been proposed: the nonlocal patch similarity graph-based method (NPSG) [25] constructs KNN graph for each patch and then compares the graphs to measure the structure consistency, and the adaptive local structure consistency-based method (ALSC) [26] adaptively constructs graph to represent the local structure of each patch and calculates the difference image (DI) by graph projection. Based on the self-similarity property, two regression-based methods have also been proposed: the fractal projection and Markovian segmentation-based method (FPMS) [27] maps the pre-event image to the domain of the postevent image by fractal projection, which consists a fractal encoding step of the pre-event image and a fractal decoding step of the postevent image; the patch similarity graph matrix-based method (PSGM) [28] first learns a self-expression matrix for one image and then computes the regression image by multiplying the other image with this learned matrix.

These self-similarity-based heterogeneous change detection methods [25]–[28] have inspired us that the image structure can be represented by constructing a graph, and this graph can be further used to complete the image regression based on structure consistency. In this article, we propose a sparse-constrained adaptive structure consistency-based image regression method for heterogeneous CD (named SCASC for short). Specifically, we first divide the multitemporal images into superpixels and construct an adaptive probabilistic graph (APG) to represent the structure information of pre-event image, which is similar to ALSC. Then, we use the APG to complete the image regression by solving an optimization model based on the structure consistency and sparse prior knowledge of changes. Finally, a Markov random field (MRF) segmentation model is used to compute the binary change map (CM), which can improve the detection accuracy by considering the spatial information of neighbor superpixels.

The main differences between SCASC and the previous self-similarity and regression-based methods are as follows.

- 1) Different from NPSG [25] and ALSC [26] that construct the graph one by one for each patch within its local search window with fixed number of neighbors, this article constructs the APG for the whole image with adaptive neighbors, which is more robust. At the same time, different from ALSC and NPSG that directly compare the graphs to detect the changes, which may lose some image information such as some small objects and some areas with blurred edge, the proposed method uses the APG to compute the regression image and compares the images in the same domain to reduce information loss.
- 2) Different from FPMS [27] that uses the fractal encoding and fractal projection to complete the structure representation and image translation, respectively, SCASC constructs the APG to represent the image structure and then translates the image by a regression model. Different from PSGM [28] that learns the self-expression matrix through a complicated iterative solving process, the APG of SCASC can be learned by a closed-form solution. More importantly, the image regression models of PSGM and SCASC are also very different. PSGM is based on the assumption that the regression image can be self-expressed through the self-expression matrix of pre-event image, while SCASC uses the similarity relationship-based graph Laplacian regularization to compute the regression image.
- 3) A main challenge in the unsupervised regression-based heterogeneous CD method is to avoid that noise and changed pixels affect the regression. Different from the common translation-based methods which aim to learn a luminance transformation function to map one image to the domain of the other image, such as HPT [19] and AM-IR [20], the proposed SCASC uses the structure consistency between heterogeneous images to compute the regression image which is more robust to noise. Moreover, SCASC decomposes the original image of the target domain into the regression image and the changed image in the optimization model, which can reduce the impact of changed pixels on the regression process.
- 4) Different from other methods that use the individual pixels or square patches [24]–[28], the proposed SCASC uses the superpixels as the basic unit, which brings two main advantages: first, it can maintain the structure and edge of the object and contain the context information, because each superpixel is internally homogeneous (belongs to the same object); second, it can also reduce the computational complexity, especially for the large-scale very-high-resolution (VHR) images.

### C. Contribution

The main contributions of this work can be summarized as follows.

- 1) An unsupervised, object-based, and image regression-based heterogeneous CD method is proposed based on the structure consistency, which adaptively constructs a graph to represent the image structure and then uses the graph to translate the pre-event image to the domain of postevent image without requiring any training set (or pseudo-training set).

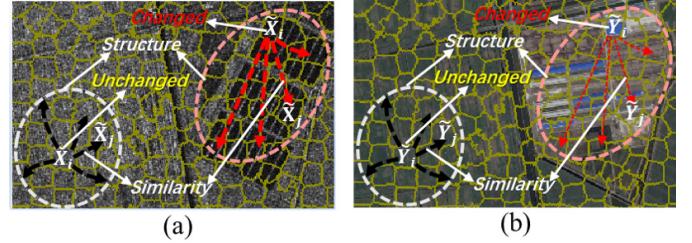


Fig. 1. Structure consistency of heterogeneous images. (a) SAR image. (b) Optical image. The thickness of the connecting line indicates the degree of similarity between image parts ( $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$ ,  $\tilde{\mathbf{Y}}_i$  and  $\tilde{\mathbf{Y}}_j$ ). The structure is defined as the similarity relationships between the image parts. In the SAR image and the optical image, the structure of the unchanged part is consistent, while the changed part is not.

- 2) The proposed adaptive structure consistency-based image regression model can alleviate the impact of noise and changed pixels on the regression process by using the structure-based transformation. The sparse regularization of changes is taken into account in the regression model to make the translated image more accurate and make the DI easier to distinguish between changed and unchanged pixels.
- 3) A superpixel-based MRF segmentation model is designed to calculate the final CM, which combines the change information and spatial context information of DI to improve the detection accuracy.
- 4) Experimental results demonstrate the effectiveness of the proposed method by comparing with some state-of-the-art (SOTA) methods (source code of the proposed method is made available at <https://github.com/yulisun/SCASC>).

The rest of this article is structured as follows. Section II describes the details of the proposed SCASC for heterogeneous CD. Section III presents the experimental results and some discussions. Finally, Section IV concludes this article.

## II. METHODOLOGY

We consider a pair of co-registered heterogeneous RS images obtained by different sensors before and after an event, which are denoted as  $\tilde{\mathbf{X}} \in \mathbb{R}^{M \times N \times C_x}$  in domain  $\mathcal{X}$  and  $\tilde{\mathbf{Y}} \in \mathbb{R}^{M \times N \times C_y}$  in domain  $\mathcal{Y}$ , and their pixels are denoted as  $\tilde{x}(m, n, c)$  and  $\tilde{y}(m, n, c)$ , respectively. Here,  $M$ ,  $N$ , and  $C_x$  ( $C_y$ ) represent the height, width, and number of bands of the image  $\tilde{\mathbf{X}}$  ( $\tilde{\mathbf{Y}}$ ), respectively. Since different imaging mechanisms result in that heterogeneous images cannot be directly compared, we need to find the connection between heterogeneous images to make them comparable.

As illustrated in Fig. 1, each image is divided into small parts with the same segmentation form. For the SAR image, if  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  represent the same (or different) kind of objects, and neither of them changes during the event, then  $\tilde{\mathbf{Y}}_i$  and  $\tilde{\mathbf{Y}}_j$  also represent the same (or different) kind of objects in the optical image. We use the similarity relationships between the target part ( $\tilde{\mathbf{X}}_i$ ) and other parts ( $\tilde{\mathbf{X}}_j$ ) to represent the structure of this target part ( $\tilde{\mathbf{X}}_i$ ). This nonlocal similarity within the image itself would eliminate the discrepancy between the heterogeneous images. Therefore, the structure of  $\tilde{\mathbf{X}}_i$  can

be well preserved by the unchanged  $\tilde{\mathbf{Y}}_i$ , as shown by the unchanged part in Fig. 1. On the contrary, if  $\tilde{\mathbf{X}}_i$  has changed in the event, the structure of  $\tilde{\mathbf{X}}_i$  is no longer preserved by  $\tilde{\mathbf{Y}}_i$ , showing that the similarity relationships between  $\tilde{\mathbf{Y}}_i$  and  $\tilde{\mathbf{Y}}_j$  are quite different from that of  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$ , as illustrated by the changed part in Fig. 1. We can find that the similarity relationship-based structure consistency is quite imaging modality invariant, and it can be used to measure the change level by calculating the difference between the structures of pre-event image  $\tilde{\mathbf{X}}$  and postevent image  $\tilde{\mathbf{Y}}$ . Therefore, there are two main problems to be considered: how to represent the image structure and how to use the structure consistency to detect changes.

### A. Superpixel Segmentation and Feature Extraction

In the proposed method, image block represented the same kind of object is chosen as the basic analysis unit rather than the individual pixel or square patch, which can be obtained by superpixel segmentation. We select the simple liner iterative clustering (SLIC) method [29] to obtain the segmented superpixels. For the optical pre-event image  $\tilde{\mathbf{X}}$  with R, G, and B bands, the original SLIC is directly employed to generate superpixels, which first transforms the RGB color images to the CIELAB color space to compute the Euclidean type color distance. For the multispectral image  $\tilde{\mathbf{X}}$  with  $C_X > 3$ , we first use the principal component analysis (PCA) method to reduce its dimensionality so that each pixel contains three principal components and then SLIC is used to segment the image without the CIELAB space transformation. For the SAR images that are usually considered to be contaminated by multiplicative speckle noise with the gamma distribution, directly using the Euclidean color distance is not appropriate, because adaptation to the noise distribution is essential for robust distance criterion. Inspired by several similarity criteria proposed in [30], the following color distances can be used to generate superpixels instead of Euclidean distance:

$$\begin{aligned} d_c &= \log\left(\frac{\tilde{x}_i + \tilde{x}_j}{2\sqrt{\tilde{x}_i\tilde{x}_j}}\right) \\ d_c &= (\log(\tilde{x}_i) - \log(\tilde{x}_j))^2 \end{aligned} \quad (1)$$

where  $\tilde{x}_i$  and  $\tilde{x}_j$  are intensity values of two pixels of the SAR image. Then, the generated superpixels have adaptive sizes and shapes and adhere well to the edges.

When we divide the image  $\tilde{\mathbf{X}}$  into  $N_S$  superpixels  $\tilde{\mathbf{X}}_i$ ,  $i = 1, \dots, N_S$  by using SLIC, we can obtain the segmentation map  $\Lambda = \{\Lambda_i | i = 1, \dots, N_S\}$ , where  $\bigcup_{i=1}^{N_S} \Lambda_i = \{(m, n) | m = 1, \dots, M; n = 1, \dots, N\}$  and  $\Lambda_i \cap \Lambda_j = \emptyset$  if  $i \neq j$ . Then, we map the segmentation map  $\Lambda$  to image  $\tilde{\mathbf{Y}}$  to obtain  $N_S$  superpixels  $\tilde{\mathbf{Y}}_i$ ,  $i = 1, \dots, N_S$  as

$$\tilde{\mathbf{Y}}_i = \{\tilde{y}(m, n, c) | (m, n) \in \Lambda_i; c = 1, \dots, C_Y\}. \quad (2)$$

Then, the pair of superpixels (such as  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{Y}}_i$ ) correspond to the same area.

Once the superpixel segmentation is completed, different features representing different information can be extracted from the superpixel, such as the spectral (intensity)-, textural-,

and spatial information-based features. Denote the feature extraction operator as  $\mathcal{F}$ , we can obtain the feature vectors of  $\mathbf{X}_i = \mathcal{F}(\tilde{\mathbf{X}}_i)$  and  $\mathbf{Y}_i = \mathcal{F}(\tilde{\mathbf{Y}}_i)$ . In this article, the mean, median, and variance of each band are selected as the superpixel features for its simplicity. Therefore, we have the feature matrices of  $\mathbf{X} \in \mathbb{R}^{3C_X \times N_S}$  and  $\mathbf{Y} \in \mathbb{R}^{3C_Y \times N_S}$  by stacking the feature vectors.

### B. Adaptive Structure-Based Graph

In order to capture the structure information of the image  $\tilde{\mathbf{X}}$ , we construct a probabilistic graph  $S^{\mathbf{X}}$ . For the  $i$ th superpixel  $\tilde{\mathbf{X}}_i$ , all the superpixels  $\{\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_{N_S}\}$  can be connected to  $\tilde{\mathbf{X}}_i$  as neighbors with probabilities  $S_{i,j}^{\mathbf{X}}$ ,  $j = 1, \dots, N_S$ . Intuitively, a small distance between  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  should be assigned a larger probability  $S_{i,j}^{\mathbf{X}}$  [31]. Therefore, we can solve the following minimization problem to achieve this goal:

$$\begin{aligned} \min_{S_{i,j}^{\mathbf{X}}} \sum_{j=1}^{N_S} \text{dist}_{i,j}^{\mathbf{X}} S_{i,j}^{\mathbf{X}} + \gamma_i (S_{i,j}^{\mathbf{X}})^2 \\ \text{s.t. } 0 \leq S_{i,j}^{\mathbf{X}} \leq 1, \quad \sum_{j=1}^{N_S} S_{i,j}^{\mathbf{X}} = 1 \end{aligned} \quad (3)$$

where  $\text{dist}_{i,j}^{\mathbf{X}} = \|\mathbf{X}_i - \mathbf{X}_j\|_2^2$  is the distance between two superpixels  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  by using the feature vectors, and  $\gamma_i \geq 0$  is a balance parameter. If we choose  $\gamma_i = 0$ , problem (1) has a trivial solution, that is, only the nearest superpixel (the superpixel  $\tilde{\mathbf{X}}_j$  with the minimum  $\text{dist}_{i,j}^{\mathbf{X}}$ ) can be the neighbor of  $\tilde{\mathbf{X}}_i$  with probability 1 and all the other superpixels cannot be connected with  $\tilde{\mathbf{X}}_i$  ( $S_{i,j}^{\mathbf{X}} = 0$  for other  $\tilde{\mathbf{X}}_j$ ). On the other hand, if we solve problem (1) without the prior distance information term, that is, we choose  $\gamma_i \rightarrow 0$ , the optimal solution of (1) is that all the superpixels can be the neighbor of  $\tilde{\mathbf{X}}_i$  with the same probability  $1/N_S$ .

Denote  $\text{dist}_i^{\mathbf{X}} \in \mathbb{R}^{N_S}$  and  $S_i^{\mathbf{X}} \in \mathbb{R}^{N_S}$  as the distance vector and similarity vector, whose  $j$ th elements are  $\text{dist}_{i,j}^{\mathbf{X}}$  and  $S_{i,j}^{\mathbf{X}}$ , respectively. Then, problem (3) can be rewritten as

$$\min_{S_i^{\mathbf{X}} = 1, 0 \leq S_{i,j}^{\mathbf{X}} \leq 1} \left\| S_i^{\mathbf{X}} + \frac{1}{2\gamma_i} \text{dist}_i^{\mathbf{X}} \right\|_2^2 \quad (4)$$

where  $\mathbf{1}_{N_S} \in \mathbb{R}^{N_S}$  is a column vector with all elements being 1. We use the method of Lagrangian multiplier to change (4) to

$$\begin{aligned} \mathcal{L}(S_i^{\mathbf{X}}, \eta, \zeta) &= \left\| S_i^{\mathbf{X}} + \frac{1}{2\gamma_i} \text{dist}_i^{\mathbf{X}} \right\|_2^2 - \eta(\mathbf{1}_{N_S}^T S_i^{\mathbf{X}} - 1) - \zeta^T S_i^{\mathbf{X}} \end{aligned} \quad (5)$$

where  $\eta$  and  $\zeta \in \mathbb{R}^{N_S}$  are the Lagrangian multipliers. By using the Karush-Kuhn-Tucker condition, we can obtain the closed-form solution for  $S_i^{\mathbf{X}}$  as

$$S_{i,j}^{\mathbf{X}} = \left( -\frac{\text{dist}_{i,j}^{\mathbf{X}}}{2\gamma_i} + \eta \right)_+. \quad (6)$$

From (6), we can find that  $S_i^{\mathbf{X}}$  is a sparse vector, that is, only the  $k$  nearest neighbors of  $\tilde{\mathbf{X}}_i$  could have a connection with  $\tilde{\mathbf{X}}_i$ . We rank  $\text{dist}_i^{\mathbf{X}}$  in ascending order as

$\text{dist}_{i,(1)}^{\mathbf{X}}, \text{dist}_{i,(2)}^{\mathbf{X}}, \dots, \text{dist}_{i,(N_S)}^{\mathbf{X}}$ . Suppose that the optimal  $S_i^{\mathbf{X}}$  has only  $k_i$  nonzero elements, we obtain

$$\begin{cases} -\frac{\text{dist}_{i,(k_i)}^{\mathbf{X}}}{2\gamma_i} + \eta > 0 \\ -\frac{\text{dist}_{i,(k_i+1)}^{\mathbf{X}}}{2\gamma_i} + \eta \leq 0 \\ \sum_{h=1}^{k_i} \left( -\frac{\text{dist}_{i,(h)}^{\mathbf{X}}}{2\gamma_i} + \eta \right) = 1. \end{cases} \quad (7)$$

Then, we can obtain the following inequality for  $\gamma_i$ :

$$\begin{cases} \gamma_i > \frac{k_i}{2} \text{dist}_{i,(k_i)}^{\mathbf{X}} - \frac{1}{2} \sum_{h=1}^{k_i} \text{dist}_{i,(h)}^{\mathbf{X}} \\ \gamma_i \leq \frac{k_i}{2} \text{dist}_{i,(k_i+1)}^{\mathbf{X}} - \frac{1}{2} \sum_{h=1}^{k_i} \text{dist}_{i,(h)}^{\mathbf{X}}. \end{cases} \quad (8)$$

Then, we can find that the regularization parameter  $\gamma_i$  can be replaced by the number of neighbors  $k_i$  when we set  $\gamma_i$  to be

$$\gamma_i = \frac{k_i}{2} \text{dist}_{i,(k_i+1)}^{\mathbf{X}} - \frac{1}{2} \sum_{h=1}^{k_i} \text{dist}_{i,(h)}^{\mathbf{X}}. \quad (9)$$

Then, we can obtain the optimal  $S_i^{\mathbf{X}}$  as

$$S_{i,(j)}^{\mathbf{X}} = \begin{cases} \frac{\text{dist}_{i,(k_i+1)}^{\mathbf{X}} - \text{dist}_{i,(j)}^{\mathbf{X}}}{k_i \text{dist}_{i,(k_i+1)}^{\mathbf{X}} - \sum_{h=1}^{k_i} \text{dist}_{i,(h)}^{\mathbf{X}}}, & j \leq k_i \\ 0, & j > k_i. \end{cases} \quad (10)$$

Therefore, the tuning of regularization parameter  $\gamma_i$  becomes the tuning of the number of neighbors  $k_i$ , which is more intuitive (it has explicit meaning) and easier (it is an integer).

From (3), we can find that the probabilistic graph  $S^{\mathbf{X}}$  is a KNN type graph with the probability calculation criteria (10). The number of neighbors  $k_i$  plays an important role in the graph  $S^{\mathbf{X}}$ . A very small  $k_i$  is not robust enough for the graph, whereas a very large  $k_i$  tends to overconnect the graph and leads to confusion, that is, there may be no  $k_i$  really similar superpixels for  $\tilde{\mathbf{X}}_i$  and then some superpixels that do not really represent the same object as  $\tilde{\mathbf{X}}_i$  will also be connected to  $\tilde{\mathbf{X}}_i$ . Therefore, we need to choose a suitable  $k_i$  for each superpixel instead of a common neighborhood number  $k$  as follows.

*Step 1:* Set  $k_{\max} = \lceil (N_S)^{1/2} \rceil$  and  $k_{\min} = \lceil (N_S)^{1/2} / 10 \rceil$  with  $\lceil \cdot \rceil$  representing the rounding up operation, and find the  $k$  nearest neighbor of each superpixel with  $k = k_{\max}$ .

*Step 2:* Calculate the in-degree  $\text{di}(\tilde{\mathbf{X}}_i)$  of each superpixel  $\tilde{\mathbf{X}}_i$ , that is, the number of times  $\tilde{\mathbf{X}}_i$  occurs among the  $k$  nearest neighbor of all superpixels.

*Step 3:* Set  $k_i = \min\{k_{\max}, \max\{\text{di}(\tilde{\mathbf{X}}_i), k_{\min}\}\}$  for each superpixel  $\tilde{\mathbf{X}}_i$  and construct the graph  $S^{\mathbf{X}}$  by using (10).

In this strategy,  $k_{\max}$  and  $k_{\min}$  are sets based on the following desirable properties: first,  $k_{\max}$  should be large enough to accurately measure the “popularity” of each superpixel in Steps 1 and 2 and small enough to represent the real structure of objects with the most superpixels ( $k_{\max}$  should be smaller than the number of superpixels that represent the most dominant object), while  $k_{\min}$  should be large enough to preserve

TABLE I  
IMPLEMENTATION STEPS OF ALGORITHM 1

Algorithm 1. The construction of adaptive structure based graph $S^{\mathbf{X}}$
<b>Input:</b> The feature matrix $\mathbf{X}$ .
1. Calculating the adaptive $k_i$ for the $i$ -th superpixel.
2. Sorting the distance vector $\text{dist}_i^{\mathbf{X}}$ .
3. Calculating the adaptive similarity vector $S_i^{\mathbf{X}}$ by using (10).
<b>Output:</b> The graph $S^{\mathbf{X}}$ .

the structure of objects with the least superpixels ( $k_{\min}$  should be larger than the number of superpixels that represent the least dominant object); second,  $k_{\max}$  and  $k_{\min}$  should be data-dependent, that is, the value of  $k_i$  should increase with the size of the dataset ( $N_S$ ), but at a slower rate than  $N_S$  itself. With this strategy of selecting neighbor number, we can choose a larger  $k_i$  for superpixel that belongs to high density (with larger  $\text{di}(\tilde{\mathbf{X}}_i)$ ) and choose a smaller  $k_i$  for superpixel that belongs to low density (with smaller  $\text{di}(\tilde{\mathbf{X}}_i)$ ).

The construction of adaptive structure-based graph  $S^{\mathbf{X}}$  is listed in Algorithm 1 of Table I, where we can find that  $S^{\mathbf{X}}$  is data-dependent in two aspects: adaptive neighbors selection and adaptive probability calculation.

### C. Sparse-Constrained Adaptive Structure Consistency-Based Image Regression

We define the regression function as  $\mathcal{M}$  and the regression image of  $\tilde{\mathbf{X}}$  in domain  $\mathcal{Y}$  as  $\tilde{\mathbf{Z}} \in \mathbb{R}^{M \times N \times C_Y}$ , that is,  $\tilde{\mathbf{Z}} = \mathcal{M}(\tilde{\mathbf{X}})$ . Because we use superpixel as the basic unit, we need to find the mapping function for feature matrices  $\mathbf{X} \rightarrow \mathbf{Z}$ , where  $\mathbf{Z} \in \mathbb{R}^{3C_Y \times N_S}$  is the feature matrix of regression image  $\tilde{\mathbf{Z}}$  as  $\mathbf{Z} = \mathcal{F}(\tilde{\mathbf{Z}})$ .

As illustrated in Fig. 1, the structure consistency between  $\tilde{\mathbf{X}}$  and its regression image  $\tilde{\mathbf{Z}}$  requires that they have the same similarity relationships between superpixels. Specifically, if  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  are very similar (represent the same kind of objects),  $\tilde{\mathbf{Z}}_i$  and  $\tilde{\mathbf{Z}}_j$  are also very similar (with small distance); on the contrary, if  $\tilde{\mathbf{X}}_i$  and  $\tilde{\mathbf{X}}_j$  are very different (represent the different kinds of objects),  $\tilde{\mathbf{Z}}_i$  and  $\tilde{\mathbf{Z}}_j$  are also very different (with large distance). Since the structure information (similarity relationships) of  $\tilde{\mathbf{X}}$  is characterized by the adaptive structure-based graph  $S^{\mathbf{X}}$ , then we have the following constraint:

$$\min_{\mathbf{Z}} \sum_{i,j=1}^{N_S} \|\mathbf{Z}_i - \mathbf{Z}_j\|_2^2 S_{i,j}^{\mathbf{X}}. \quad (11)$$

Denote the degree matrix  $\mathbf{D}_S \in \mathbb{R}^{N_S \times N_S}$  as a diagonal matrix with the  $i$ -th diagonal element being  $\sum_{j=1}^{N_S} (S_{i,j}^{\mathbf{X}} + S_{j,i}^{\mathbf{X}})/2$ , and the Laplacian matrix  $\mathbf{L}_S$  as  $\mathbf{L}_S = \mathbf{D}_S - ((\mathbf{S}^{\mathbf{X}} + (\mathbf{S}^{\mathbf{X}})^T)/2)$ , then we have

$$\sum_{i,j=1}^{N_S} \|\mathbf{Z}_i - \mathbf{Z}_j\|_2^2 S_{i,j}^{\mathbf{X}} = 2\text{Tr}(\mathbf{Z}\mathbf{L}_S\mathbf{Z}^T) \quad (12)$$

where  $\text{Tr}(\cdot)$  denotes the trace of a matrix.

On the other hand, the regression image  $\tilde{\mathbf{Z}}$  can be regarded as a part of postevent image  $\tilde{\mathbf{Y}}$ , that is, image  $\tilde{\mathbf{Y}}$  can be decomposed into a regression image  $\tilde{\mathbf{Z}}$  and a changed image  $\tilde{\Delta}$ . Then we have  $\mathbf{Y} = \mathbf{Z} - \Delta$ , where  $\Delta \in \mathbb{R}^{3C_Y \times N_S}$

is the changed feature matrix. Based on the fact that most of the objects are unchanged in the event and only a small part of objects are changed, then this prior sparsity can be used to regularize (12) as follows:

$$\min_{\mathbf{Z}, \Delta} 2\text{Tr}(\mathbf{Z}\mathbf{L}_S\mathbf{Z}^T) + \lambda\|\Delta\|_{2,1} \quad \text{s.t. } \mathbf{Y} = \mathbf{Z} - \Delta \quad (13)$$

where  $\lambda > 0$  is a regularization parameter. The  $\ell_{2,1}$ -norm regularization term  $\|\Delta\|_{2,1}$  is defined as  $\|\Delta\|_{2,1} = \sum_{i=1}^{N_S} \|\Delta_i\|_2$ , which is a convex relaxation of the original  $\ell_{2,0}$ -norm of  $\|\Delta\|_{2,0}$ .

By using the alternating direction method of multipliers (ADMM), the augmented Lagrangian function of (13) can be written as

$$\begin{aligned} \mathcal{L}(\mathbf{Z}, \Delta, \mathbf{W}) = & 2\text{Tr}(\mathbf{Z}\mathbf{L}_S\mathbf{Z}^T) + \text{Tr}(\mathbf{W}^T(\mathbf{Z} - \mathbf{Y} - \Delta)) \\ & + \frac{\mu}{2}\|\mathbf{Z} - \mathbf{Y} - \Delta\|_F^2 + \lambda\|\Delta\|_{2,1} \end{aligned} \quad (14)$$

where  $\mathbf{W} \in \mathbb{R}^{3C_Y \times N_S}$  is a Lagrange multiplier, and  $\mu > 0$  is a penalty parameter. Then, the alternating direction method can be used to solve the minimization problem of (14).

First, given the current points  $(\mathbf{Z}^k, \Delta^k, \mathbf{W}^k)$  at the  $k$ th iteration, the  $\mathbf{Z}$ -subproblem can be formulated as

$$\begin{aligned} \mathbf{Z}^{k+1} = \arg \min_{\mathbf{Z}} & \left\{ 2\text{Tr}(\mathbf{Z}\mathbf{L}_S\mathbf{Z}^T) + \text{Tr}((\mathbf{W}^k)^T \mathbf{Z}) \right. \\ & \left. + \frac{\mu}{2}\|\mathbf{Z} - \mathbf{Y} - \Delta^k\|_F^2 \right\}. \end{aligned} \quad (15)$$

It can be solved by taking its first-order derivative to zero, then we can update  $\mathbf{Z}$  as

$$\mathbf{Z}^{k+1} = (\mu\mathbf{Y} + \mu\Delta^k - \mathbf{W}^k) \times (4\mathbf{L}_S + \mu\mathbf{I}_{N_S})^{-1} \quad (16)$$

where  $\mathbf{I}_{N_S} \in \mathbb{R}^{N_S \times N_S}$  represents an identity matrix.

Second, with the fixed points  $(\mathbf{Z}^{k+1}, \Delta^k, \mathbf{W}^k)$ , the  $\Delta$ -subproblem can be formulated as

$$\begin{aligned} \Delta^{k+1} = \arg \min_{\Delta} & \left\{ \lambda\|\Delta\|_{2,1} - \text{Tr}((\mathbf{W}^k)^T \Delta) \right. \\ & \left. + \frac{\mu}{2}\|\mathbf{Z}^{k+1} - \mathbf{Y} - \Delta\|_F^2 \right\} \end{aligned} \quad (17)$$

which can be solved by the proximal operator as

$$\Delta^{k+1} = \text{prox}_{\frac{\lambda}{\mu}\|\cdot\|_{2,1}}(\mathbf{Q}^{k+1}) \quad (18)$$

where  $\mathbf{Q}^{k+1} = \mathbf{Z}^{k+1} - \mathbf{Y} + (\mathbf{W}^k/\mu)$ , and the proximal operator is defined as

$$\text{prox}_{af}(\mathbf{b}) = \arg \min_{\mathbf{x}} f(\mathbf{x}) + \frac{1}{2\alpha}\|\mathbf{x} - \mathbf{b}\|_F^2. \quad (19)$$

The closed-form solution of (18) can be obtained as in [32]

$$\Delta_i^{k+1} = \max \left\{ \|\mathbf{Q}_i^{k+1}\|_2 - \frac{\lambda}{\mu}, 0 \right\} \frac{\mathbf{Q}_i^{k+1}}{\|\mathbf{Q}_i^{k+1}\|_2} \quad (20)$$

where we follow the convention  $0 \cdot (0/0) = 0$ .

Third, with the fixed points  $(\mathbf{Z}^{k+1}, \Delta^{k+1}, \mathbf{W}^k)$ , the Lagrangian multiplier can be updated as

$$\mathbf{W}^{k+1} = \mathbf{W}^k + \mu(\mathbf{Z}^{k+1} - \mathbf{Y} - \Delta^{k+1}). \quad (21)$$

The ADMM framework for the minimization problem of (13) is summarized in Algorithm 2 of Table II, where  $N_{iter}$  is the maximum number of iterations,  $\xi^k$  is the relative difference

TABLE II  
MAIN IMPLEMENTATION STEPS OF SCASC-BASED IMAGE REGRESSION

Algorithm 2. SCASC based image regression

---

<b>Input:</b>	Feature matrix $\mathbf{Y}$ , adaptive structure based graph $S^X$ , parameters $\lambda, \mu, \xi^0$ and $N_{iter}$ .
<b>Initialize:</b>	Matrices $\Delta^0 = 0, \mathbf{W}^0 = 0$ .
<b>Main iteration loop:</b>	
for $k = 0, 1, \dots, N_{iter}$ do	
<b>Z-updating:</b> compute $\mathbf{Z}^{k+1}$ according to (16).	
<b>Δ-updating:</b> compute $\Delta^{k+1}$ according to (20).	
<b>W-updating:</b> compute $\mathbf{W}^{k+1}$ according to (21).	
Exit criterion: compute $\xi^{k+1} = \ \Delta^{k+1} - \Delta^k\ _2 / \ \Delta^{k+1}\ _2$	
if $\xi^{k+1} < \xi^0$ then	
exit	
End if	
End for	
<b>Output:</b> Matrices $\mathbf{Z}^{k+1}, \Delta^{k+1}$ .	

---

between two iterative results. The exit criterion  $\xi^k < \xi^0$  means that there is no longer any appreciable changes in the iteration and the algorithm runs into convergence.

Once the regression feature matrix  $\mathbf{Z}$  is solved, we can use it to obtain the regression image  $\tilde{\mathbf{Z}}$  by extracting the mean features in  $\mathbf{Z}$

$$\tilde{z}(m, n, c) = z_{c,i}; (m, n) \in \Lambda_i, c = 1, \dots, C_Y, i = 1, \dots, N_S \quad (22)$$

where  $\tilde{z}(m, n, c)$  is the pixel value of  $\tilde{\mathbf{Z}}$ , and  $z_{c,i}$  is the mean feature of each band in  $\mathbf{Z}$ .

#### D. DI Generation and Binary CM Calculation

By using the SCASC image regression model (13), we can obtain the changed feature matrix  $\Delta$ . Then, we can obtain the DI as follows:

$$\text{DI}(m, n) = \|\Delta_i\|_2; (m, n) \in \Lambda_i, i = 1, \dots, N_S. \quad (23)$$

Then, the thresholding method or clustering method can be used to obtain the final CM. For example, the Otsu threshold method [33], K-means clustering [34], or fuzzy c-means (FCM) clustering [35] is widely used to segment the DI. In this article, we employ the MRF model to compute the binary CM, which can incorporate the change and spatial contextual information of DI.

The binary segmentation problem of CM calculation can be regarded as a superpixel-labeling problem. Suppose  $\mathbf{L} = \{L_i | i = 1, \dots, N_S\}$  represents the label set of superpixels, in which  $L_i = 0$  means that the region of  $\Lambda_i$  has not changed and  $L_i = 1$  means that it has changed. In accordance with the MRF approach, the superpixel-labeling is equivalent to minimizing an energy function  $E(\mathbf{L})$ , which is the log likelihood of the posterior distribution of MRF [36]

$$\mathbf{L} = \arg \min_{\mathbf{L}} \{E(\mathbf{L}) = \alpha E_c(\mathbf{L}) + (1 - \alpha) E_s(\mathbf{L})\} \quad (24)$$

where  $E_c(\mathbf{L})$  is the change energy term,  $E_s(\mathbf{L})$  is the spatial energy term, and  $\alpha \in (0, 1)$  is a balanced parameter.

The  $R$ -adjacency neighbor system is constructed for  $E_s(\mathbf{L})$ , that is, if two superpixels (located in  $\Lambda_i$  and  $\Lambda_j$ ) intersect or

the 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  $i \in \mathcal{N}_j^R$  (or  $j \in \mathcal{N}_i^R$ ). As the size of each superpixel generated by SLIC is around  $MN/N_S$ , we set  $R = 2(MN/N_S)^{1/2}$  for the  $R$ -adjacency neighbor system. Then, the spatial energy term  $E_s(\mathbf{L})$  of MRF is defined as

$$\begin{aligned} E_s(\mathbf{L}) &= \sum_{i=1}^{N_S} \sum_{j \in \mathcal{N}_i^R} \varepsilon_s(L_i, L_j) \\ \varepsilon_s(L_i, L_j) &= \frac{\exp\left(-\frac{\|\Delta_i - \Delta_j\|_2^2}{2\sigma^2}\right)}{d(\Lambda_i, \Lambda_j)} \delta(L_i - L_j) \end{aligned} \quad (25)$$

where  $\sigma^2 = ((\sum_{i=1}^{N_S} \sum_{j \in \mathcal{N}_i^R} \|\Delta_i - \Delta_j\|_2^2) / (\sum_{i=1}^{N_S} |\mathcal{N}_i^R|))$  is a normalization parameter representing the average feature distance,  $d(\Lambda_i, \Lambda_j)$  is the Euclidean spatial distance between the centers of two superpixels, and  $\delta(\cdot)$  is the function defined as  $\delta(x) = 1$  if  $x \neq 0$  and  $\delta(x) = 0$  if  $x = 0$ .

The change energy term  $E_c(\mathbf{L})$  is defined as

$$\begin{aligned} E_c(\mathbf{L}) &= \sum_{i=1}^{N_S} \varepsilon_c(L_i) \\ \varepsilon_c(L_i) &= \begin{cases} \max\{-\log(\|\Delta_i\|_2^2/2T), 0\}; & \text{if } L_i = 1 \\ \min\{-\log(1 - \|\Delta_i\|_2^2/2T), \omega\}; & \text{if } L_i = 0 \end{cases} \end{aligned} \quad (26)$$

where  $T$  is the Otsu thresholding parameter on the vector  $\{\|\Delta_i\|_2^2 | i = 1, \dots, N_S\}$ , and  $\omega$  is set to be slightly larger than the maximal sum of the spatial energy term for each superpixel as

$$\omega = \log 2 + \max_{i=1, \dots, N_S} \sum_{j \in \mathcal{N}_i^R} \varepsilon_c(L_i, L_j). \quad (27)$$

From (26), we can find that when  $\|\Delta_i\|_2^2 < T$ ,  $\varepsilon_c(L_i)|_{L_i=1}$  is larger than  $\varepsilon_c(L_i)|_{L_i=0}$ , which makes the  $i$ th superpixel tend to be labeled as unchanged class; when  $\|\Delta_i\|_2^2 > T$ ,  $\varepsilon_c(L_i)|_{L_i=1}$  is smaller than  $\varepsilon_c(L_i)|_{L_i=0}$ , which makes the  $i$ th superpixel tend to be labeled as changed class. At the same time, the maximum and minimum in (26) are used to make  $\varepsilon_c(L_i)$  have nonnegative and upper bound constraints when  $\|\Delta_i\|_2^2 > 2T$ . In addition, this type of  $\varepsilon_c(L_i)$  can be regarded as derived from the assumption that  $\|\Delta_i\|_2^2$  obeys a uniform distribution between 0 and  $2T$ .

With these defined energy functions, the energy minimization problem of (24) can be implemented via the min-cut/max-flow algorithm [37]. Once the optimal  $\mathbf{L}$  is solved, we can obtain the binary CM as

$$\text{CM}(m, n) = L_i, \quad \text{if } (m, n) \in \Lambda_i. \quad (28)$$

### III. EXPERIMENTAL ANALYSIS AND DISCUSSION

In this section, experiments are conducted on different heterogeneous datasets to verify the effectiveness of the proposed SCASC.<sup>1</sup>

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

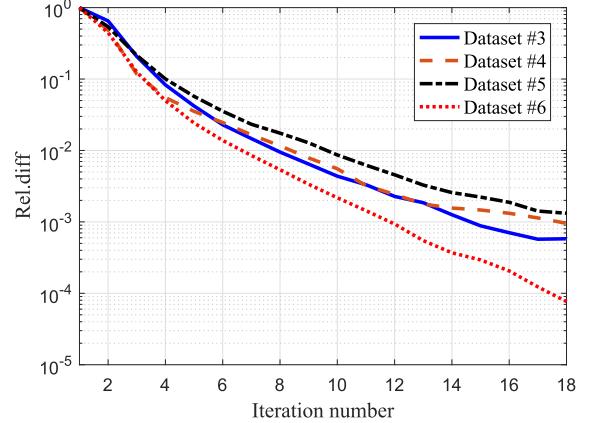


Fig. 2. Convergence performance of Algorithm 1.

#### A. Heterogeneous Datasets and Quantitative Measures

To validate the proposed heterogeneous CD method, six real heterogeneous datasets are employed, as listed in Table III. These datasets reflect different types of heterogeneity: multisensor optical images (same sensor type but with different sensors, e.g., #3 and #4); and multisource images (different sensor types, e.g., #5 and #6). These datasets cover different resolutions (varying from 0.52 to 30 m), different image sizes (varying from 300 to 2000 pixels in width or length), and different types of events (such as flooding, fire, and construction), which can evaluate the robustness of SCASC in different CD tasks.

To evaluate the performance of SCASC, two types of quantitative measures are employed. First, the DI can be evaluated by the receiver-operating characteristics (ROC) curve and the precision-recall (PR) curve, which are plotted by using the true positive (TP) rate (TPR) versus the false positive (FP) rate (FPR) and the precision rate versus the recall rate, respectively. The TPR (also known as the recall rate), FPR, and precision rate are computed as  $\text{TPR} = \text{TP}/(\text{TP} + \text{FN})$ ,  $\text{FPR} = \text{FP}/(\text{TN} + \text{FP})$ , and  $\text{Precision} = \text{TP}/(\text{TP} + \text{FP})$  respectively, where TP, TN, FN, and FP represent the TPs, true negatives (TNs), false negative (FNs), and FPs, respectively. Second, the final CM can be evaluated by the percentage of correct classification (PCC), the F1 score, and the Kappa coefficient (KC) are computed as  $\text{PCC} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})$ ,  $\text{F1} = (2 \times \text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall})$ , and  $\text{KC} = (\text{PCC} - \text{PRE})/(1 - \text{PRE})$  with

$$\text{PRE} = \frac{(\text{TP} + \text{FN})(\text{TP} + \text{FP}) + (\text{TN} + \text{FP})(\text{TN} + \text{FN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})^2}. \quad (29)$$

#### B. Parameter Analysis

The main parameters of SCASC are: the number of superpixels  $N_S$  in the superpixel segmentation process, the regularization parameter  $\lambda$ , maximum iterations  $N_{\text{iter}}$ , and relative difference of termination  $\zeta^0$  in the ADMM framework of regression process; the balance parameter  $\alpha$  in the MRF segmentation process.

TABLE III  
DESCRIPTION OF THE SIX HETEROGENEOUS DATASETS

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	Landsat-5/EO-1 ALI	1534 × 808 × 7(10)	Aug. 2011 - Sept. 2011	Texas, USA	Forest fire (30m.)
#3	Pleiades/WorldView2	2000 × 2000 × 3(3)	May 2012 - July 2013	Toulouse, France	Construction (0.52m.)
#4	Spot/NDVI	990 × 554 × 3(1)	1999 - 2000	Gloucester, England	Flooding ( $\approx$ 25m.)
#5	Radarsat-2/Google Earth	593 × 921 × 1(3)	June 2008 - Sept. 2012	Shuguang Village, China	Building construction (8m.)
#6	Landsat-8/Sentinel-1A	875 × 500 × 11(3)	Jan. 2017 - Feb. 2017	Sutter County, USA	Flooding ( $\approx$ 15m.)

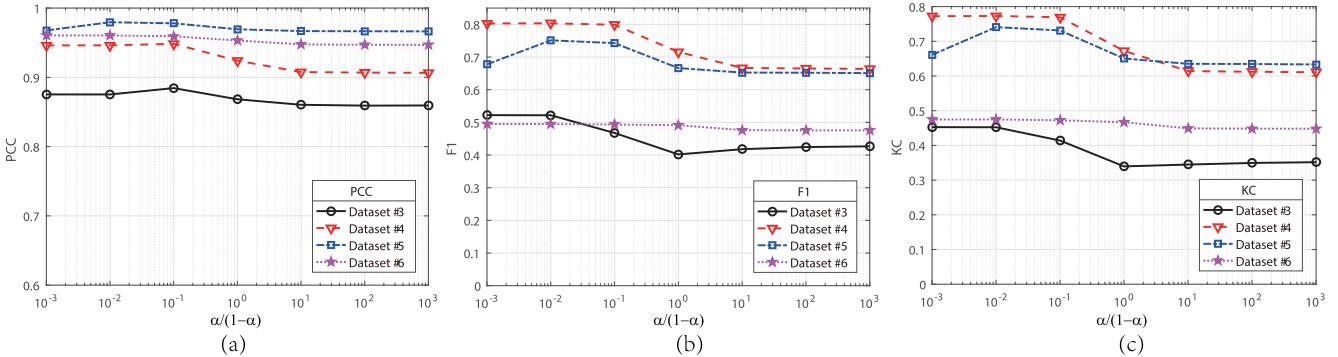


Fig. 3. Influences of parameter  $\alpha$  on the SCASC performance. (a) PCC. (b) F1. (c) KC.

Generally,  $N_S$  has two effects on the algorithm: first, it affects the detection granularity of the algorithm; second, it determines the complexity of the algorithm as analyzed in Section III-E2. Large  $N_S$  brings small superpixels and improves the detection granularity, but also increases the computational complexity (as reported in Table VII in Section III-E2). In this article, we simply set  $N_S = 10^4$  as a compromise choice.

To investigate the effect of  $N_{\text{iter}}$  and  $\xi^0$  on the ADMM framework in Table II, we plot the convergence curves of Algorithm 2 in Fig. 2, which is measured by relative difference  $\zeta^k$  versus iteration number  $k$ . It is clear that the ADMM converges quickly, and the exit criterion of  $N_{\text{iter}} = 10$  or  $\xi^0 = 0.01$  is enough for the algorithm, while taking into account the computational time.

For the parameter  $\lambda$  in model (13), it controls the strength of sparse regularization, which should be selected according to the proportion of the changed component (i.e., sparsity level). Generally, if the fewer the regions in the image that has been changed, the larger we should select the  $\lambda$ . Alternatively, if we know in prior the size of the changed regions (the prior sparsity level  $\tau$ ), we can rewrite the  $\ell_{2,1}$ -norm constraint model (13) as the  $\ell_{2,0}$ -norm constraint model as

$$\min_{\mathbf{Z}} 2\text{Tr}(\mathbf{Z}\mathbf{L}_S\mathbf{Z}^T) \quad \text{s.t. } \mathbf{Y} = \mathbf{Z} - \Delta, \quad \|\Delta\|_{2,0} \leq \tau \quad (30)$$

which can be approximately solved by using Algorithm 2, but replacing the  $\Delta$ -subproblem with the following minimization:

$$\Delta^{k+1} = \arg \min_{\Delta} \left\{ \|\Delta - \mathbf{Q}^{k+1}\|_F^2 \right\} \quad \text{s.t. } \|\Delta\|_{2,0} \leq \tau. \quad (31)$$

It can be solved by the hard thresholding operator as

$$\Delta_i^{k+1} = \begin{cases} \mathbf{Q}_i^{k+1}, & \text{if } i \in p^\tau \\ 0, & \text{if } i \notin p^\tau \end{cases} \quad (32)$$

where  $p^\tau$  is the top  $\tau$  values' indices vector of  $\{\|\mathbf{Q}_i^{k+1}\|_2 | i = 1, \dots, N_S\}$  with descending order.

For the parameter  $\alpha$  in the MRF model (24), it is used to balance the change energy term and spatial energy term. To investigate the effect of  $\alpha$  on the MRF segmentation model, we plot the accuracy curves of segmentation results in Fig. 3, which is measured by PCC and KC versus  $\alpha/(1-\alpha)$  from  $10^{-3}$  to  $10^3$  with the ratio of 10. When  $\alpha$  becomes smaller, the spatial energy term  $E_s(\mathbf{L})$  takes up a larger proportion in  $E(\mathbf{L})$ , making the segmented CM smoother; conversely, when  $\alpha$  becomes larger, the change energy term  $E_c(\mathbf{L})$  takes up a larger proportion in  $E(\mathbf{L})$ , making the segmented CM more similar to the image segmented by Otsu thresholding method. From Fig. 3, it can be seen that the segmentation performance is better when the value of  $\alpha/(1-\alpha)$  is between  $10^{-3}$  and  $10^{-2}$ , so we set  $\alpha = 0.05$  in this article.

In summary, for all the experiments, we set the  $N_S = 10^4$  for the superpixel segmentation process,  $\lambda = 0.1$ ,  $N_{\text{iter}} = 10$ , and  $\xi^0 = 0.01$  for the ADMM framework of regression process,  $\alpha = 0.05$  for the MRF segmentation process.

### C. Image Regression Performance of SCASC

In the first experiment, we test SCASC on the Datasets #1 and #2<sup>2</sup> to verify the effectiveness of structure consistency-based image regression. Both Datasets #1 and #2 contain one pre-event image and two post-event images, where one of the postevent images is homogeneous to the pre-event image and the other is heterogeneous, as shown in Fig. 4(a)-(c). For the Dataset #1, the two near-infrared (NIR) band images are acquired by Landsat-5 in September 1995 and July 1996,

<sup>2</sup>Dataset #2 is provided by Volpi *et al.* [38] and kindly made available at <https://sites.google.com/site/michelevolpiresearch/codes/cross-sensor>.

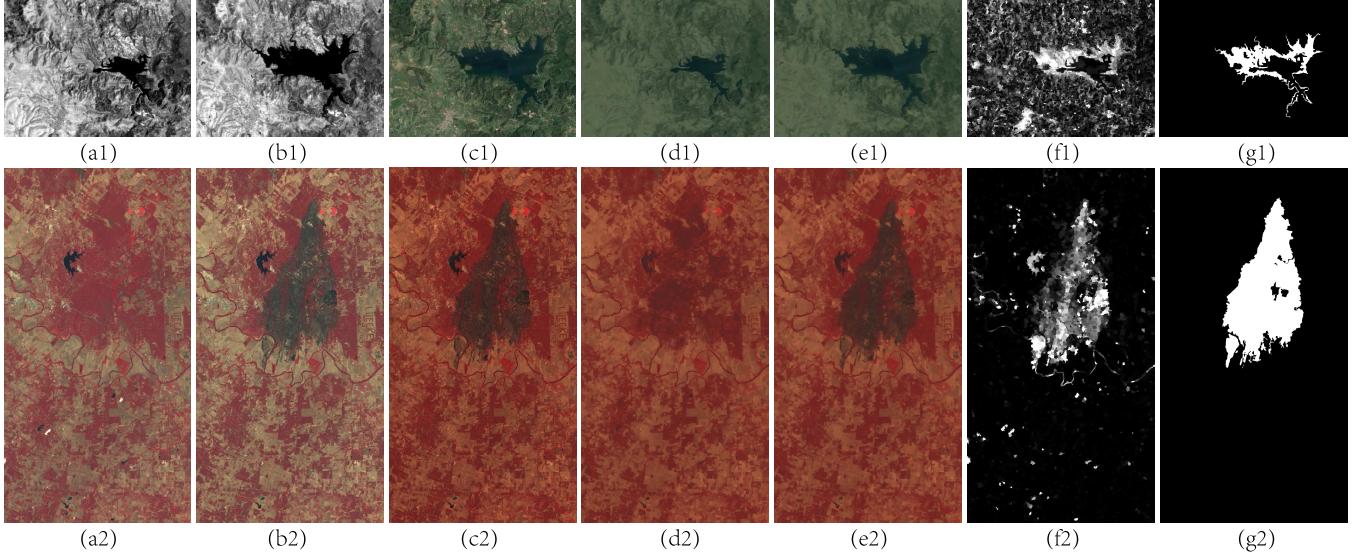


Fig. 4. Regression images of SCASC on Datasets #1 and #2. (From left to right) (Top row) (a1) pre-event image  $\tilde{\mathbf{X}}_{\text{NIR}-t_1}$ , (b1) postevent image  $\tilde{\mathbf{X}}_{\text{NIR}-t_2}$ , (c1) postevent image  $\tilde{\mathbf{Y}}_{\text{opt}-t_2}$ , (d1) regression image  $\tilde{\mathbf{Z}}_{\text{NIR}-t_1}$ , (e1) regression image  $\tilde{\mathbf{Z}}_{\text{NIR}-t_2}$ , (f1) DI generated by SCASC with  $\tilde{\mathbf{X}}_{\text{NIR}-t_1}$  and  $\tilde{\mathbf{Y}}_{\text{opt}-t_2}$ , (g1) ground truth of Dataset #1. (From left to right) (Bottom row) (a2) Pre-event image  $\tilde{\mathbf{X}}_{\text{L5}-t_1}$ , (b2) postevent image  $\tilde{\mathbf{X}}_{\text{L5}-t_2}$ , (c2) postevent image  $\tilde{\mathbf{Y}}_{\text{ALI}-t_2}$ , (d2) regression image  $\tilde{\mathbf{Z}}_{\text{L5}-t_1}$ , (e2) regression image  $\tilde{\mathbf{Z}}_{\text{L5}-t_2}$ , (f2) DI generated by SCASC with  $\tilde{\mathbf{X}}_{\text{L5}-t_1}$  and  $\tilde{\mathbf{Y}}_{\text{ALI}-t_2}$ , and (g2) ground truth of Dataset #2.

and another postevent optical image is obtained from Google Earth in July 1996. We denote these images of Datasets #1 as  $\tilde{\mathbf{X}}_{\text{NIR}-t_1}$ ,  $\tilde{\mathbf{X}}_{\text{NIR}-t_2}$  and  $\tilde{\mathbf{Y}}_{\text{opt}-t_2}$ , respectively. For the Dataset #2, the first two multispectral images are acquired by Landsat-5 in August 2011 and September 2011, and another postevent multispectral image is acquired by the Advanced Land Imager (ALI) from the Earth Observing (EO-1) mission in September 2011. Similarly, the pre-event and two postevent images of Datasets #2 are denoted as  $\tilde{\mathbf{X}}_{\text{L5}-t_1}$ ,  $\tilde{\mathbf{X}}_{\text{L5}-t_2}$  and  $\tilde{\mathbf{Y}}_{\text{ALI}-t_2}$ , respectively.

Fig. 4(d) and (e) shows the regression images of  $\tilde{\mathbf{Z}}_{\text{NIR}-t_1}$ ,  $\tilde{\mathbf{Z}}_{\text{NIR}-t_2}$  and  $\tilde{\mathbf{Z}}_{\text{L5}-t_1}$ ,  $\tilde{\mathbf{Z}}_{\text{L5}-t_2}$  by using (22), which are calculated by translating  $\tilde{\mathbf{X}}_{\text{NIR}-t_1}$ ,  $\tilde{\mathbf{X}}_{\text{NIR}-t_2}$  and  $\tilde{\mathbf{X}}_{\text{L5}-t_1}$ ,  $\tilde{\mathbf{X}}_{\text{L5}-t_2}$  to the domains of  $\tilde{\mathbf{Y}}_{\text{opt}-t_2}$  and  $\tilde{\mathbf{Y}}_{\text{ALI}-t_2}$ , respectively. From Fig. 4(d) and (e), it can be found that the structure consistency can be used to implement the image regression, that is, to transform the image acquired by one sensor into a similar image acquired by another sensor. By comparing  $\tilde{\mathbf{Z}}_{\text{NIR}-t_2}$  and  $\tilde{\mathbf{Y}}_{\text{opt}-t_2}$ ,  $\tilde{\mathbf{Z}}_{\text{L5}-t_2}$  and  $\tilde{\mathbf{Y}}_{\text{ALI}-t_2}$ , we can find that the structures of  $\tilde{\mathbf{X}}_{\text{NIR}-t_2}$  and  $\tilde{\mathbf{X}}_{\text{L5}-t_2}$  represented by  $S^{\tilde{\mathbf{X}}_{\text{NIR}-t_2}}$  and  $S^{\tilde{\mathbf{X}}_{\text{L5}-t_2}}$  can be well conformed by  $\tilde{\mathbf{Y}}_{\text{NIR}-t_2}$  and  $\tilde{\mathbf{Y}}_{\text{L5}-t_2}$ , respectively. Meanwhile, by comparing  $\tilde{\mathbf{Z}}_{\text{NIR}-t_1}$  and  $\tilde{\mathbf{Y}}_{\text{opt}-t_2}$ ,  $\tilde{\mathbf{Z}}_{\text{L5}-t_1}$  and  $\tilde{\mathbf{Y}}_{\text{ALI}-t_2}$ , we can find that the structure consistency is no longer maintained in the changed areas, which can be used to measure the change level. Fig. 4(f) shows the DIs generated by SCASC with the pre-event  $\tilde{\mathbf{X}}_{\text{NIR}-t_1}$ ,  $\tilde{\mathbf{X}}_{\text{L5}-t_1}$  and the postevent  $\tilde{\mathbf{Y}}_{\text{opt}-t_2}$ ,  $\tilde{\mathbf{Y}}_{\text{ALI}-t_2}$ , respectively. It can be seen that the DIs can well distinguish between changed and unchanged areas.

#### D. Comparing With Some SOTA Methods

In order to evaluate the performance of SCASC, we select the recently proposed M3CD<sup>3</sup> [39], FPMS<sup>4</sup> [27], NPSG<sup>5</sup> [25],

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

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

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

ALSC [26], and PSGM [28] for comparison. Fig. 5(a) and (b) shows the heterogeneous pre-event and postevent images of Datasets #3 to #6, and Fig. 5(c) shows the regression image of SCASC by transforming the pre-event image into the domain of postevent image. Fig. 5(d) shows the DI of SCASC calculated by using (23), and their corresponding ROC and PR curves are plotted in Fig. 6. From the regression images of Fig. 5(c) and DIs of Fig. 5(d), we can see that the proposed SCASC can well establish the connections between heterogeneous images by using the structure consistency and project one image to the domain of the other image. In Fig. 6, the areas under the ROC curves (AUR) of Datasets #3 to #6 are 0.793, 0.936, 0.968, and 0.887, respectively, and the areas under the PR curves (AUP) of Datasets #3 to #6 are 0.458, 0.636, 0.695, and 0.447, respectively.

Fig. 5(e)–(j) shows the binary CMs generated by different methods on Datasets #3 to #6. By comparing these CMs in Fig. 5, we can find that the changed and unchanged regions are well detected by SCASC with relatively small FPs and FNs. In addition, we also list the quantitative measures of these CMs in Table IV, the highest and second-highest scores are highlighted in bold and underlined, respectively. As can be seen in Table IV, the proposed method achieves very good results (optimal or suboptimal), which verifies the effectiveness of the sparse-constrained structure consistency-based regression model and MRF-based segmentation method.

In addition, in order to further compare the performance of the proposed method, some other representative and SOTA methods other than M3CD, FPMS, NPSG, PSGM, and ALSC are selected for comparison, such as NLPEM [15], SCCN [16], AFL-DSR [17], LT-FT [18], HPT [19], AM-IR [20], reliable mixed-norm (RMN)-based method [40], multidimensional evidential reasoning (MDER)-based method [41], and multidimensional scaling (MDS)-based method [42], as summarized in Table V (we directly quote the results of the corresponding

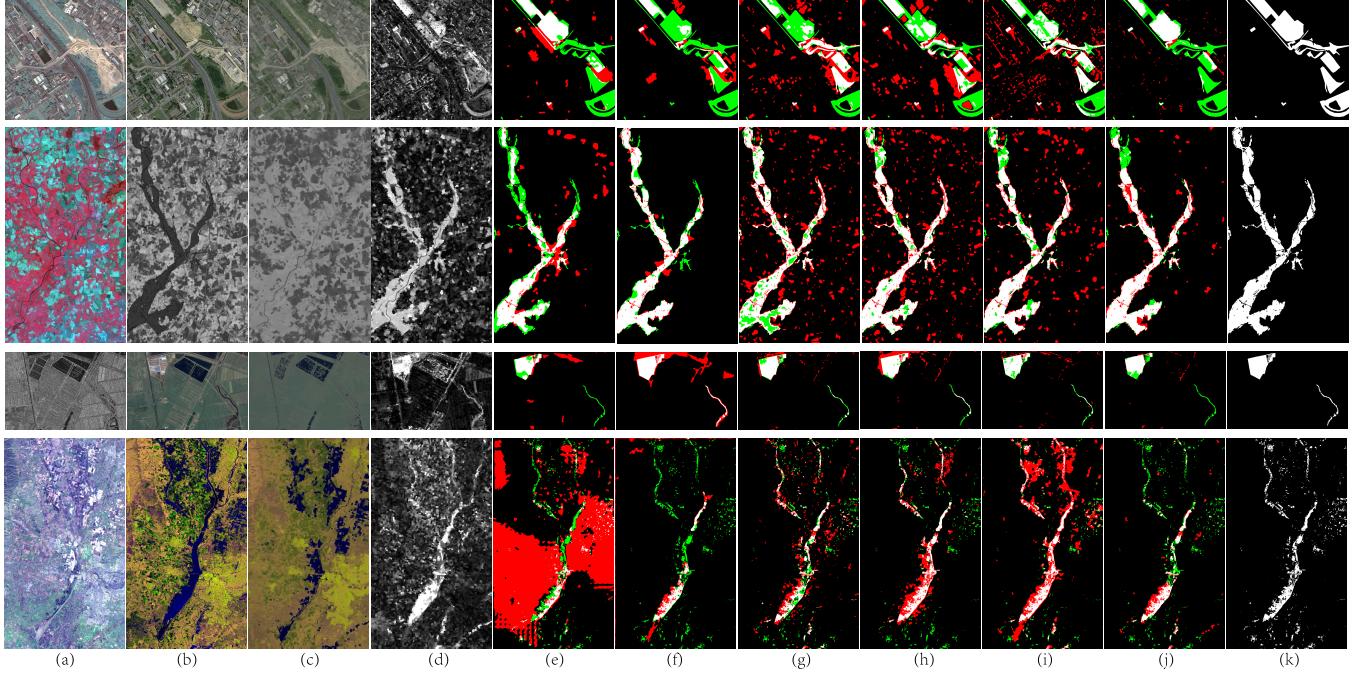


Fig. 5. DI of SCASC and binary CMs of different methods on heterogeneous datasets. From top to bottom, they correspond to Datasets #3 to #6, respectively. (From left to right) (a) Pre-event image, (b) postevent image, (c) regression image of SCASC, (d) DI of SCASC, (e) binary CM of M3CD, (f) binary CM of FPMS, (g) binary CM of NPSG, (h) binary CM of ALSC, (i) binary CM of PSGM, (j) binary CM of SCASC, (k) ground truth. In the binary CM, White: TPs; Red: FPs; Black: TNs; Green: FNs.

TABLE IV  
QUANTITATIVE MEASURES OF BINARY CMs ON THE HETEROGENEOUS DATASETS. THE HIGHEST AND SECOND-HIGHEST SCORES ARE HIGHLIGHTED IN BOLD AND UNDERLINED, RESPECTIVELY

Methods	Dataset #3			Dataset #4			Dataset #5			Dataset #6			Average		
	PCC	F1	KC												
M3CD	<u>0.863</u>	0.481	0.405	0.915	0.636	0.588	0.962	0.622	0.602	0.575	0.077	0.021	0.829	0.454	0.404
FPMS	0.838	0.296	0.215	<b>0.962</b>	<b>0.837</b>	<b>0.816</b>	0.938	0.597	0.569	0.947	0.356	0.329	<b>0.921</b>	0.522	0.482
NPSG	0.830	0.446	0.346	0.902	0.663	0.608	0.975	0.742	0.729	0.941	0.449	0.419	0.912	0.575	0.526
ALSC	0.815	0.422	0.312	0.907	0.693	0.641	0.963	0.688	0.669	0.944	0.498	0.470	0.907	0.575	0.523
PSGM	0.857	<b>0.558</b>	<b>0.473</b>	0.922	0.719	0.675	<u>0.977</u>	<b>0.756</b>	<b>0.744</b>	0.908	0.422	0.383	0.916	<u>0.614</u>	0.569
SCASC	<b>0.892</b>	0.516	0.464	0.949	<u>0.800</u>	<u>0.771</u>	<b>0.979</b>	0.751	0.741	<b>0.961</b>	<b>0.500</b>	<b>0.479</b>	<b>0.945</b>	<b>0.642</b>	<b>0.614</b>

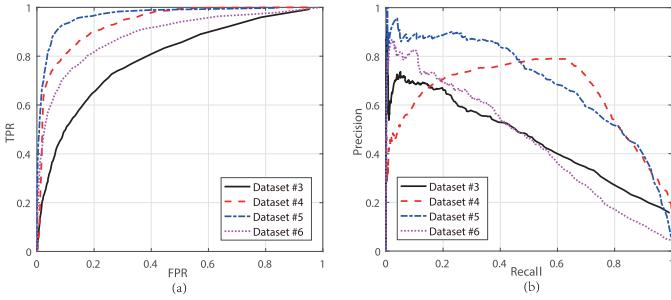


Fig. 6. ROC and PR curves of SCASC generated DIs. (a) ROC curves. (b) PR curves.

datasets in their original published articles). Among these comparison methods, SCCN, AFL-DSR, and LT-FT are deep learning-based methods. From Table V, we can find that the SCASC is able to achieve higher or competitive accuracy rates and to obtain consistently good results across different

datasets. The average PCC of SCASC on Datasets #3 to #6 is about 94.5%.

### E. Discussion

1) *Effectiveness of Adaptive Probabilistic Graph:* We have constructed an adaptive structure-based graph  $S^X$  in Algorithm 1 (Table I) to capture the structure information of the image  $\tilde{X}$ . To verify the effectiveness of the proposed adaptive  $k$ -selection strategy-based APG, we compare it with the widely used  $k$ -nearest neighbors graph (KNNG) and the  $k$  fixed APG. Specifically, the KNNG computes the  $S^X$  as

$$S_{i,j}^X = \begin{cases} \exp(-\text{dist}_{i,j}^X), & j \in \mathcal{N}_{\tilde{X}_i}^k \\ 0, & \text{otherwise} \end{cases} \quad (33)$$

where  $\mathcal{N}_{\tilde{X}_i}^k$  represents the position set of the KNN of  $\tilde{X}_i$  by sorting all elements in the distance vector  $\text{dist}_i^X$  and taking out

TABLE V  
ACCURACY RATE OF CMS GENERATED BY DIFFERENT METHODS ON DIFFERENT DATASETS. THE RESULTS OF THESE COMPARISON METHODS ARE REPORTED BY THEIR ORIGINAL PUBLISHED ARTICLES

Dataset #3	PCC	Dataset #4	PCC	Dataset #5	PCC	Dataset #6	PCC
<b>SCASC</b>	<b>0.892</b>	HPT <sup>[19]</sup>	0.957-0.964	AFL-DSR <sup>[17]</sup>	0.980	<b>SCASC</b>	<b>0.961</b>
AFL-DSR <sup>[17]</sup>	0.880	<b>SCASC</b>	<b>0.949</b>	SCCN <sup>[16]</sup>	0.976	AM-IR <sup>[20]</sup>	0.933
RMN <sup>[40]</sup>	0.877	AFL-DSR <sup>[17]</sup>	0.836	MDS <sup>[42]</sup>	0.967		
NLPEM <sup>[15]</sup>	0.853	MDER <sup>[41]</sup>	0.818	LT-FL <sup>[18]</sup>	0.964		
				RMN <sup>[40]</sup>	0.884		

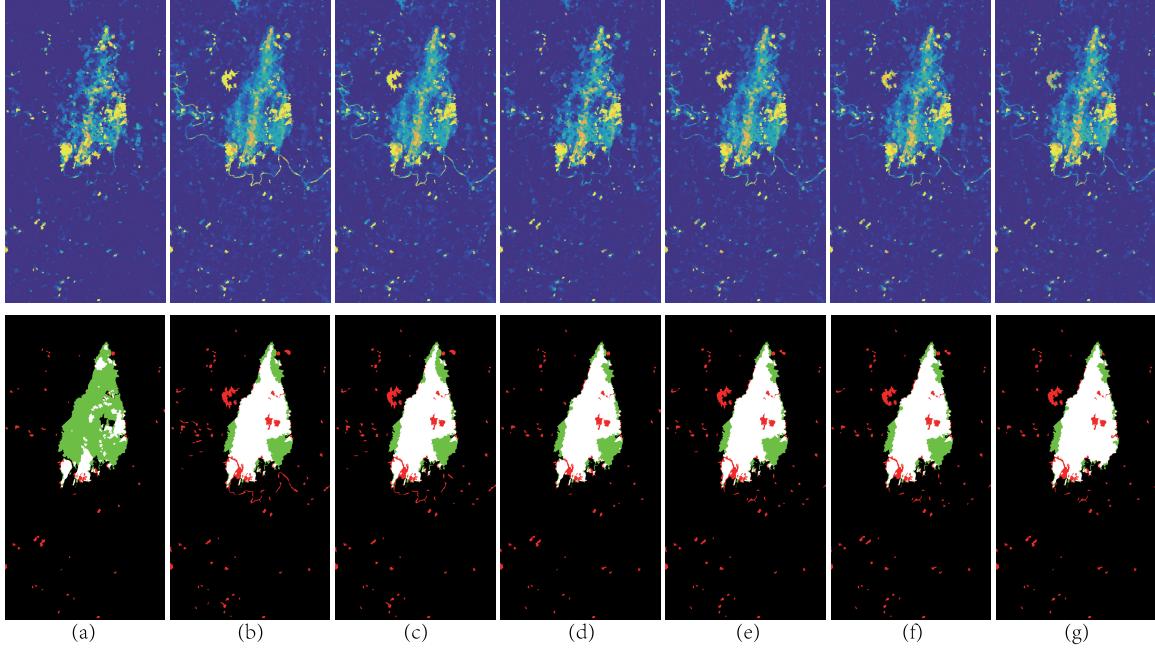


Fig. 7. (Top row) DIs and (bottom row) CMs generated by SCASC on Dataset #2 with different graphs. (a) KNNG with fixed  $k = k_{\min}$ . (b) KNNG with fixed  $k = k_{\max}$ . (c) KNNG with fixed  $k = k_{\text{mean}}$ . (d) APG with fixed  $k = k_{\min}$ . (e) APG with fixed  $k = k_{\max}$ . (f) APG with fixed  $k = k_{\text{mean}}$ . (g) APG with adaptive  $k$ .

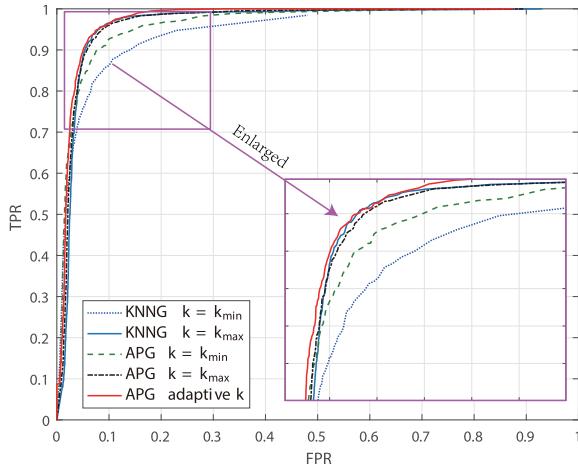


Fig. 8. ROC curves of DI generated by SCASC with different graphs on Dataset #2.

the  $k$  smallest elements. We further normalize the weights as  $S_{i,j}^X \leftarrow S_{i,j}^X / \sum_{j=1}^{N_S} S_{i,j}^X$  to make  $\sum_{j=1}^{N_S} S_{i,j}^X = 1$ .

Fig. 7 shows the DIs and CMs generated on Dataset #2 (with  $\tilde{\mathbf{X}}_{L5-t1}$  and  $\tilde{\mathbf{Y}}_{ALI-t2}$ ) by SCASC with the proposed adaptive

TABLE VI  
AUR OF DIS GENERATED BY SCASC WITH DIFFERENT GRAPHS ON DIFFERENT DATASETS

Graphs	$k$ -value	Datasets					
		#1	#2	#3	#4	#5	#6
k fixed KNNG	$k_{\min}$	0.881	0.940	0.785	0.916	0.951	0.867
	$k_{\max}$	0.891	0.964	0.769	0.931	0.958	0.874
	$k_{\text{mean}}$	0.890	0.965	0.780	0.931	0.960	0.873
k fixed APG	$k_{\min}$	0.881	0.961	0.777	0.920	0.953	0.871
	$k_{\max}$	0.892	0.967	0.787	0.933	0.962	0.877
	$k_{\text{mean}}$	0.891	0.967	0.785	0.934	0.963	0.881
APG with adaptive $k$		0.894	0.972	0.793	0.936	0.967	0.886

$k$ -based APG and  $k$  fixed KNNG, APG of  $k = k_{\min}$ ,  $k = k_{\max}$ ,  $k = k_{\text{mean}}$ , where  $k_{\text{mean}} = (1/N_S) \sum_{i=1}^{N_S} k_i$  is the mean value of the adaptive  $k$ . Fig. 8 plots the corresponding ROC curves. Table VI lists the AUR of DIs generated by SCASC with graphs on Datasets #1 to #6. From Figs. 7 and 8 and Table VI, we can find that: 1) the APG usually performs better than the KNNG in SCASC, which is due to the fact that the constraint of  $\sum_{j=1}^{N_S} \text{dist}_{i,j}^X S_{i,j}^X$  in the APG construction process

TABLE VII  
COMPUTATIONAL TIME (SECONDS) OF EACH PROCESS OF SCASC

$N_S$	Dataset #1 (with size $300 \times 412$ )					Dataset #3 (with size $2000 \times 2000$ )				
	$t_{p1}$	$t_{p2}$	$t_{p3} (t_{p3}^*)$	$t_{p4}$	$t_{total} (t_{total}^*)$	$t_{p1}$	$t_{p2}$	$t_{p3} (t_{p3}^*)$	$t_{p4}$	$t_{total} (t_{total}^*)$
5000	0.30	0.13	3.35 (4.20)	1.12	5.09 (5.92)	2.87	0.12	3.48 (5.01)	1.58	8.43 (9.95)
10000	0.64	0.36	20.20 (17.71)	4.26	25.65 (23.16)	3.21	0.36	19.61 (18.58)	4.20	27.77 (26.74)
20000	1.37	0.99	142.82 (71.47)	14.96	160.26 (88.91)	4.19	1.08	143.92 (70.27)	14.23	163.77 (90.11)

(3) is consistent with the constraint of  $\sum_{i,j=1}^{N_S} \|\mathbf{Z}_i - \mathbf{Z}_j\|_2^2 S_{i,j}^X$  in the image regression model (13); 2) the proposed adaptive  $k$  strategy can improve the performance of APG by selecting appropriate  $k$  for each vertex, resulting in better a DI (a higher AUR) than the  $k$  fixed KNNG and APG whether  $k$  takes the maximum value  $k_{max}$ , the minimum value  $k_{min}$ , or the mean value  $k_{mean}$ , which ultimately improves detection performance as shown in Fig. 7. This demonstrates the superiority of the proposed adaptive structure-based graph  $S^X$ .

2) *Complexity Analysis*: The main computational complexity of the proposed SCASC is concentrating on the pre-processing (superpixel segmentation and feature extraction), constructing the adaptive structure-based graph (Algorithm 1), solving the image regression model (13) (Algorithm 2) and the MRF segmentation (24) by using min-cut/max-flow algorithm [37].

a) *Preprocessing*: the complexity of the simple linear iterative clustering (SLIC) method is linear in the number of pixels in the image  $\mathcal{O}(MN)$ , which is reported in [29]. The average number of pixels within each superpixel is  $MN/N_S$ , then the complexity of mean and variance feature extraction is around  $\mathcal{O}((C_X + C_Y)MN)$ , the median feature extraction is around  $\mathcal{O}((C_X + C_Y)MN \log(MN/N_S))$ .

b) *Algorithm 1*: first, calculating the distances between all the superpixels requires  $\mathcal{O}(3C_X N_S^2/2)$ . Second, sorting the distance vector  $\text{dist}_i^X$  requires  $\mathcal{O}(N_S \log N_S)$  by using some accelerated sorting algorithms, such as the Block sort or Tree sort. Then calculating the adaptive  $k$  requires  $\mathcal{O}(N_S^2 \log N_S)$ . Third, calculating the closed-form similarity vector  $S_i^X$  by using (10) requires  $\mathcal{O}(N_S)$ . Therefore, constructing the adaptive structure-based graph (Algorithm 1) requires  $\mathcal{O}(N_S^2(3C_X/2 + \log N_S + 1))$ .

c) *Algorithm 2*: first, updating  $\mathbf{Z}$  with (16). The matrix inversion of  $(4\mathbf{L}_S + \mu\mathbf{I}_{N_S})^{-1}$  requires  $\mathcal{O}(N_S^3)$ , the matrix multiplication requires  $\mathcal{O}(3C_Y N_S^2)$ . Then, updating  $\mathbf{Z}$  needs  $\mathcal{O}(N_S^3)$ . Second, updating  $\Delta$  with (20). As  $\Delta$  can be solved column-wise by using the closed-form proximal operator, then it requires  $\mathcal{O}(3C_Y N_S)$ . Third, updating  $\mathbf{W}$  with (21). It requires  $\mathcal{O}(3C_Y N_S)$  to update the Lagrange multiplier.

d) *MRF segmentation*: first, calculating the spatial energy term  $E_s(\mathbf{L})$  and change energy term  $E_c(\mathbf{L})$  requires  $\mathcal{O}(N_R)$  and  $\mathcal{O}(N_S)$ , respectively, where  $N_R$  is the number of edges in the  $R$ -adjacency neighbor system. Second, the theoretical complexity and empirical complexity of min-cut/max-flow algorithm has been studied in [37], that is, the theoretical complexity of the worst case is  $\mathcal{O}(2N_R N_S^2)$ . However, its empirical complexity is relatively low on typical problem instances in vision, which can also be seen in Table VII.

Although the complexity of the proposed method is very high in the above-mentioned theoretical analysis, which requires  $\mathcal{O}(N_S^3)$  for each iteration in Algorithm 2, it can be accelerated by two strategies. First, since  $(4\mathbf{L}_S + \mu\mathbf{I}_{N_S})^{-1}$  is not updated during the iteration framework, we can calculate it off-line in advance to reduce the complexity. Obviously, this still requires a large amount of computation when  $N_S$  is very large. Second, since matrix  $\mathbf{L}_S$  is a Laplacian matrix that is sparse, real, symmetric, and positive definite, the linear system of  $\mathbf{Z}^{k+1}(4\mathbf{L}_S + \mu\mathbf{I}_{N_S}) = (\mu\mathbf{Y} + \mu\Delta^k - \mathbf{W}^k)$  can be solved efficiently by using iterative solvers, such as the conjugate gradient (CG) method. In addition, the preconditioned CG (PCG) method can also be used to accelerate the CG method by using some preconditioners [43], such as Jacobi, incomplete Cholesky (IC), and successive overrelaxation (SOR). In practice, when  $N_S > 10^4$ , we recommend the second strategy of using the PCG method with the IC preconditioner (IC-PCG).

Table VII reports the computational time of each process of SCASC with different  $N_S$ , which are performed in MATLAB 2016a running on a Windows desktop with Intel(R) Core(TM) i7-8700K CPU and 32GB of RAM. In Table VII,  $t_{p1}$  to  $t_{p4}$  represent the computational time spent in the superpixel segmentation and feature extraction, construction of adaptive structure-based graph (Algorithm 1), SCASC-based image regression (Algorithm 2), and image segmentation with min-cut/max-flow algorithm, respectively.  $t_{p3}^*$  represents the computational time of Algorithm 2 by using the IC-PCG for updating  $\mathbf{Z}$ , and  $t_{total}$  and  $t_{total}^*$  represent the total computational time of SCASC without and with IC-PCG, respectively. From Table VII, we can find that first, the number of superpixels is the main factor affecting the running time rather than the image size; second, the IC-PCG can effectively reduce the running time when  $N_S$  is very large.

#### IV. CONCLUSION

In this article, we propose an unsupervised, object-based, and image regression-based CD method for heterogeneous RS images, which is based on the fact that heterogeneous images share the same structure information for the same ground object. The proposed SCASC first extracts the image structural information by efficiently constructing a robust and data-dependent graph, which adaptively selects neighbors and adaptively assigns weights for each node. With the structure consistency property, SCASC uses the structure-based graph to translate one image to the domain of the other image by a sparse-constrained image regression model. Finally, an MRF segmentation model is designed to improve the accuracy

of CM by combining the change information and spatial contextual information. Experimental results on different heterogeneous datasets demonstrate the effectiveness of SCASC.

In the SCASC, the sparsity of changes is used as a regularization term in the regression model. A future work would be to explore more prior knowledge of the changes to improve the regression results obtained by Algorithm 2, such as the gradient sparsity of DI. Also, the distribution model of DI obtained from the image regression process needs to be further investigated, so as to design a more accurate segmentation model.

#### ACKNOWLEDGMENT

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

#### REFERENCES

- [1] A. Singh, “Review article digital change detection techniques using remotely-sensed data,” *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989.
- [2] D. Brunner, G. Lemoine, and L. Bruzzone, “Earthquake damage assessment of buildings using VHR optical and SAR imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2403–2420, May 2010.
- [3] L. Gueguen and R. Hamid, “Toward a generalizable image representation for large-scale change detection: Application to generic damage analysis,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3378–3387, Jun. 2016.
- [4] R. E. Kennedy *et al.*, “Remote sensing change detection tools for natural resource managers: Understanding concepts and tradeoffs in the design of landscape monitoring projects,” *Remote Sens. Environ.*, vol. 113, no. 7, pp. 1382–1396, Jul. 2009.
- [5] X. Zhang, H. Su, C. Zhang, X. Gu, X. Tan, and P. M. Atkinson, “Robust unsupervised small area change detection from SAR imagery using deep learning,” *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 79–94, Mar. 2021.
- [6] Z. Y. Lv, T. F. Liu, P. Zhang, J. A. Benediktsson, T. Lei, and X. Zhang, “Novel adaptive histogram trend similarity approach for land cover change detection by using bitemporal very-high-resolution remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9554–9574, Dec. 2019.
- [7] Q. Wang, Z. Yuan, Q. Du, and X. Li, “GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2019.
- [8] W. Zhou, A. Troy, and M. Grove, “Object-based land cover classification and change analysis in the Baltimore metropolitan area using multi-temporal high resolution remote sensing data,” *Sensors*, vol. 8, no. 3, pp. 1613–1636, Mar. 2008.
- [9] L. Wan, Y. Xiang, and H. You, “A post-classification comparison method for SAR and optical images change detection,” *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1026–1030, Jul. 2019.
- [10] L. Wan, Y. Xiang, and H. You, “An object-based hierarchical compound classification method for change detection in heterogeneous optical and SAR images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9941–9959, Dec. 2019.
- [11] C. Kwan, B. Ayhan, J. Larkin, L. Kwan, S. Bernabé, and A. Plaza, “Performance of change detection algorithms using heterogeneous images and extended multi-attribute profiles (EMAPs),” *Remote Sens.*, vol. 11, no. 20, p. 2377, Oct. 2019.
- [12] L. Wan, T. Zhang, and H. J. You, “Multi-sensor remote sensing image change detection based on sorted histograms,” *Int. J. Remote Sens.*, vol. 39, no. 11, pp. 3753–3775, Mar. 2018.
- [13] G. Mercier, G. Moser, and S. B. Serpico, “Conditional copulas for change detection in heterogeneous remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1428–1441, May 2008.
- [14] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J. Y. Tourneret, “A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors,” *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 799–812, Mar. 2015.
- [15] R. Touati and M. Mignotte, “An energy-based model encoding nonlocal pairwise pixel interactions for multisensor change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 1046–1058, Feb. 2018.
- [16] J. Liu, M. Gong, K. Qin, and P. Zhang, “A deep convolutional coupling network for change detection based on heterogeneous optical and radar images,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 545–559, Mar. 2018.
- [17] R. Touati, M. Mignotte, and M. Dahmane, “Anomaly feature learning for unsupervised change detection in heterogeneous images: A deep sparse residual model,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 588–600, 2020.
- [18] T. Zhan, M. Gong, X. Jiang, and S. Li, “Log-based transformation feature learning for change detection in heterogeneous images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 9, pp. 1352–1356, Sep. 2018.
- [19] Z. Liu, G. Li, G. Mercier, Y. He, and Q. Pan, “Change detection in heterogeneous remote sensing images via homogeneous pixel transformation,” *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1822–1834, Apr. 2018.
- [20] L. T. Luppino, F. M. Bianchi, G. Moser, and S. N. Anfinsen, “Unsupervised image regression for heterogeneous change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9960–9975, Dec. 2019.
- [21] X. Niu, M. Gong, T. Zhan, and Y. Yang, “A conditional adversarial network for change detection in heterogeneous images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 45–49, Jan. 2019.
- [22] M. Gong, X. Niu, T. Zhan, and M. Zhang, “A coupling translation network for change detection in heterogeneous images,” *Int. J. Remote Sens.*, vol. 40, no. 9, pp. 3647–3672, May 2019.
- [23] L. T. Luppino *et al.*, “Deep image translation with an affinity-based change prior for unsupervised multimodal change detection,” *IEEE Trans. Geosci. Remote Sens.*, early access, Feb. 17, 2021, doi: [10.1109/TGRS.2021.3056196](https://doi.org/10.1109/TGRS.2021.3056196).
- [24] Y. Sun, L. Lei, X. Li, X. Tan, and G. Kuang, “Structure consistency-based graph for unsupervised change detection with homogeneous and heterogeneous remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, early access, Feb. 2, 2021, doi: [10.1109/TGRS.2021.3053571](https://doi.org/10.1109/TGRS.2021.3053571).
- [25] Y. Sun, L. Lei, X. Li, H. Sun, and G. Kuang, “Nonlocal patch similarity based heterogeneous remote sensing change detection,” *Pattern Recognit.*, vol. 109, Jan. 2021, Art. no. 107598.
- [26] L. Lei, Y. Sun, and G. Kuang, “Adaptive local structure consistency-based heterogeneous remote sensing change detection,” *IEEE Geosci. Remote Sens. Lett.*, early access, Nov. 26, 2020, doi: [10.1109/LGRS.2020.3037930](https://doi.org/10.1109/LGRS.2020.3037930).
- [27] M. Mignotte, “A fractal projection and Markovian segmentation-based approach for multimodal change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8046–8058, Nov. 2020.
- [28] Y. Sun, L. Lei, X. Li, X. Tan, and G. Kuang, “Patch similarity graph matrix-based unsupervised remote sensing change detection with homogeneous and heterogeneous sensors,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 4841–4861, Jun. 2021.
- [29] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [30] C.-A. Deledalle, L. Denis, and F. Tupin, “How to compare noisy patches? Patch similarity beyond Gaussian noise,” *Int. J. Comput. Vis.*, vol. 99, no. 1, pp. 86–102, 2012.
- [31] F. Nie, X. Wang, and H. Huang, “Clustering and projected clustering with adaptive neighbors,” in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2014, pp. 977–986.
- [32] J. Yang, W. Yin, Y. Zhang, and Y. Wang, “A fast algorithm for edge-preserving variational multichannel image restoration,” *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 569–592, 2009.
- [33] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
- [34] J. A. Hartigan and M. A. Wong, “Algorithm AS 136: A K-means clustering algorithm,” *J. Roy. Stat. Soc. C (Appl. Statist.)*, vol. 28, no. 1, pp. 100–108, 1979.
- [35] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. Boston, MA, USA: Springer, 1981.

- [36] R. Szeliski *et al.*, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1068–1080, Jun. 2008.
- [37] Y. Boykov and V. Kolmogorov, "An experimental comparison of mincut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.
- [38] M. Volpi, G. Camps-Valls, and D. Tuia, "Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis," *J. Photogram. Remote Sens.*, vol. 107, pp. 50–63, Sep. 2015.
- [39] R. Touati, M. Mignotte, and M. Dahmane, "Multimodal change detection in remote sensing images using an unsupervised pixel pairwise based Markov random field model," *IEEE Trans. Image Process.*, vol. 29, pp. 757–767, Aug. 2019.
- [40] R. Touati, M. Mignotte, and M. Dahmane, "A reliable mixed-norm-based multiresolution change detector in heterogeneous remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3588–3601, Sep. 2019.
- [41] Z.-G. Liu, G. Mercier, J. Dezert, and Q. Pan, "Change detection in heterogeneous remote sensing images based on multidimensional evidential reasoning," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 168–172, Jan. 2014.
- [42] R. Touati, M. Mignotte, and M. Dahmane, "Change detection in heterogeneous remote sensing images based on an imaging modality-invariant MDS representation," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3998–4002.
- [43] F. Nar, A. Ozgur, and A. N. Saran, "Sparsity-driven change detection in multitemporal SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 7, pp. 1032–1036, Jul. 2016.



**Yuli Sun** received the B.S. degree from Xiangtan University, Xiangtan, China, in 2009, and the M.S. degree from the University of Science and Technology of China, Hefei, China, in 2014. He is currently pursuing the Ph.D. degree in information and communication engineering with the College of Electronic Science, National University of Defense Technology, Changsha, China.

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 the National University of Defense Technology, Changsha, China, in 2008.

She is currently a Professor with the School of Electronic Science, National University of Defense Technology. Her research interests include computer vision, remote sensing image interpretation, and data fusion.



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

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



**Ming Li** received the B.S. and M.S. degrees from Central South University, Changsha, China, in 2013 and 2017, respectively. He is currently pursuing the Ph.D. degree with the College of Electronic Science, National University of Defense Technology, Changsha.

His research interests cover remote sensing object detection, image retrieval, and change detection.



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

He is currently a Professor with the School of Electronic Science, National University of Defense Technology. His research interests include remote sensing, SAR image processing, change detection, SAR ground moving target indication, and classification with polar metric SAR images.