

Patch Similarity Graph Matrix-Based Unsupervised Remote Sensing Change Detection With Homogeneous and Heterogeneous Sensors

Yuli Sun¹, Lin Lei¹, Xiao Li¹, Xiang Tan, and Gangyao Kuang, *Senior Member, IEEE*

Abstract—Change detection (CD) of remote sensing images is an important and challenging topic, which has found a wide range of applications in many fields. In particular, one of the main challenges is to detect changes between heterogeneous images, where the difference in imaging mechanism makes it difficult to carry out a direct comparison. In this article, we propose an unsupervised CD framework based on the patch similarity graph matrix (PSGM), which assumes that the patch similarity graph structure of each homogeneous or heterogeneous image is consistent if no change occurs. First, it learns the PSGM of one image based on the self-expressive property, which can be interpreted as containing the edges of the fully connected graphs with each image patch as a vertex. Then, the change level depends on how much one image still conforms to the similarity graph structure learned from the other image. Meanwhile, the change map can be further optimized by using the prior sparse knowledge that only a small part of the image changed and most areas remain unchanged. Experiments with both homogeneous and heterogeneous data sets demonstrate the effective performance of the proposed PSGM-based CD method.

Index Terms—Heterogeneous data, similarity graph matrix, sparse regularization, unsupervised change detection (CD).

I. INTRODUCTION

A. Background

THE change detection (CD) of remote sensing images is a process of analyzing the differences between images to identify changes of an object or phenomenon that have occurred in the same geographical area at different times. CD is a very important topic in remote sensing images processing, which plays a key role in many real-world applications, such as disaster relief, agriculture survey, urban planning, and military monitoring [1]–[3].

Generally, CD techniques can be divided into two categories based on homogeneous and heterogeneous images according to imaging modalities, where homogeneous images refer to

images coming from the same imaging modality satellite sensor(s), e.g., radar or optical sensors, and the heterogeneous images refer to images acquired from sensors with disparate imaging modalities. In this article, we propose a new method for remote sensing CD, which can be applied to both homologous and heterologous CD tasks (with special attention to heterogeneous CD).

The procedure of traditional CD in homologous images (both synthetic aperture radar (SAR) or optical images) can be divided into three steps: image preprocessing, difference image (DI) generation, and analysis of DI [4]. According to the second and third steps, the current CD in homologous images can be roughly grouped into four categories: 1) image arithmetical operation: image differencing (usually for optical images) and image-ratio/log-ratio (usually for SAR images) [5] have been widely used to highlight changes; 2) image transformation: the original data set is transformed into another space to highlight changes [6], [7]; 3) image classification: postclassification methods compare results of classification to generate CD maps; and 4) other advanced methods: new techniques, such as the deep learning-based methods, have been also introduced into CD task [8]–[13].

Recently, the heterogeneous CD has attracted a growing interest due to the huge amount of heterogeneous data from the existing earth-observing satellites with the development of remote sensing technology. Especially, it is of great practical significance for immediate evaluation and emergency disasters. In such scenarios (e.g., earthquake or flood), the preevent SAR image is usually unavailable, and the preevent optical image can be obtained from archived data of remote sensing platforms, whereas only the postevent SAR image can be available because of the weather constraints, such as cloud or sunlight. Despite its undeniable potential, there are relatively few works that have been devoted to heterogeneous CD tasks compared with the homogeneous CD task. As different sensors provide different descriptions on the same truth and show different characteristics, the heterogeneous CD invalidates the assumptions of the same measured physical quantities, the same signature class, and the same data statistical behavior (an example is shown in Fig. 1¹), which are often used in

Manuscript received January 31, 2020; revised July 2, 2020; accepted July 28, 2020. Date of publication August 13, 2020; date of current version May 21, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61701508. (Yuli Sun and Lin Lei contributed equally to this work.) (Corresponding author: Lin Lei.)

The authors are with the College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China (e-mail: sunyuli@mail.ustc.edu.cn; alaleilin@163.com; lxcherishm@126.com; tanxiang@nudt.edu.cn; kuangyeats@hotmail.com).

Color versions of one or more of the figures in this article are available online at <https://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2020.3013673

¹The first pair of optical-SAR images is provided by Image Processing Research Group, University Federico II of Naples, at www.grip.unina.it.

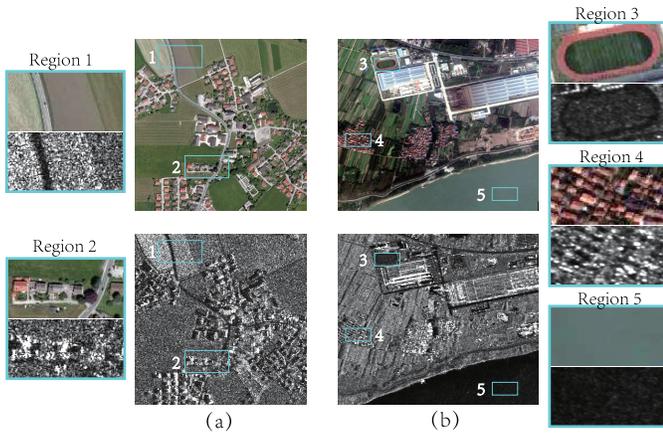


Fig. 1. Two pairs of optical-SAR images. (a) Images acquired over Rosenheim, Germany. (b) Images acquired over Wuhan, China. Both SAR images are acquired by TerraSAR-X, and both optical images are provided by Google Earth Pro. From these two pairs of optical-SAR images and five enlarged regions, we can see that different sensors provide different descriptions on the same truth and show different characteristics.

homologous CD methods. Therefore, it is more difficult to obtain the DI for the heterogeneous CD. According to the different methods for generating the binary map, the existing heterogeneous CD methods can be roughly divided into three categories.

The first is the classification-based methods [14]–[17]. These methods first classify the images, transform the heterogeneous images into the same category space, and then compare them to detect changes. Among them, postclassification comparison (PCC) is the most widely used method [14], [15]. However, PCC often suffers from error propagation or error accumulation [18]. To alleviate this problem, the method based on multitemporal segmentation and compound classification (MS-CC) [16], [17] have been proposed recently. However, as shown in these articles, image segmentation will affect the accuracy of CD, especially the segmentation is very difficult in the SAR image.

The second is the deep learning-based methods [19]–[23]. Because of the outstanding capabilities in efficient feature learning, the convolutional neural networks (CNNs) and stacked denoising autoencoders (SDAEs) have been used in heterogeneous CD to infer spatial information from the data and learn new representations, such as the symmetric convolutional coupling network (SCCN) [19] and approximately symmetrical deep neural network (ASDNN) [20]. Meanwhile, the translation network can also be used to translate the two heterogeneous images into homogeneous ones, such as the conditional generative adversarial network (cGAN) [23]. However, there are two major flaws in these deep learning-based methods: the process of training networks is time-consuming, and it requires a lot of pairs of unchanged heterogeneous images to construct training set, which needs a high cost of manual operation in practice when the method uses supervision mode or needs a more complicated training process to construct an iterative coarse to fine filter for the training samples when it uses unsupervision mode.

The third is the traditional DI-based methods that use other methods to obtain the DI different from the deep

learning-based methods and can be roughly divided into parametric- [24]–[26], nonparametric- [27]–[32], and the regression-based [33], [34] approaches. In parametric techniques, a mixture (or set) of multivariate distributions are generally used to model the joint statistics or the dependencies between the heterogeneous images, such as the copula theory and Kullback–Leibler (KL) distance-based method [24], local joint distributions and manifold learning-based method [25], and the Markov model for multimodal change detection (M3CD) method [26]. In nonparametric techniques, some imaging modality-invariant operators are designed to highlight the changes, such as the different forms of pixel pair difference used in [28]–[30] and sparse representations of two coupled dictionaries learned from each observed image [31], [32]. In regression-based techniques, an operation similar to image-to-image translation is used to predict how every pixel in one image domain would appear in the other image domain, such as the supervised homogeneous pixel transformation (HPT) method [33] and the unsupervised image regression framework with pseudotraining data [34]. For the traditional DI-based methods, two points are very important: one is how to accurately explore the relationship between heterogeneous images; the other is how to design change detection operators to make full use of these relationships. Therefore, this kind of method usually faces such difficulties: when the scene is complex or the noise in the image is very serious (especially the speckle noise of SAR image), these relationships are no longer applicable or the designed operators cannot fully represent the relationships between heterogeneous images, resulting in a sharp deterioration in CD performance.

B. Motivation

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

Inspired by the self-expressive property [35], we propose a new method based on the patch similarity graph matrix (PSGM) for both homogeneous and heterogeneous CD tasks (more focus on heterogeneous CD), which belongs to the family of above-cited traditional DI-based methods. It is well known that a small patch in the image can be approximately reconstructed by selecting similar patches in an extended search window (or the whole image) based on pathwise similarity, which is also the basis of the widely used nonlocal denoising algorithms, such as probabilistic patch-based (PPB) algorithm [36] and block-matching 3-D (BM3D) [37]. Here, we employ this similarity-based reconstruction to build the relationship between different images.

Although the same object shows different measurements under different sensors resulting in incomparable pixel values of heterogeneous images, the similarity relationship between

patches in one image domain should be consistent with that in another image domain. In other words, this patch similarity between heterogeneous images can be transferred, and then, the change level will depend on how much the preevent (postevent) image still conforms to the similarity graph structure learned from the postevent (preevent) image. Therefore, the proposed similarity graph matrix-based CD method mainly consists of two parts. First, it calculates the corresponding PSGM for each input image (preevent or postevent), which can be a single channel or multichannel, such as different polarization modes in SAR images or different bands in the optical image. For the similar matrix learning of multichannel images, we use the strategy of multiview fusion. Second, it measures the change level to obtain a preliminary DI by multiplying one image with the learned PSGM from the other image and calculating the difference. Furthermore, the prior information of sparse change is employed to optimize the preliminary DI. This is because, in the preevent and postevent images, most of the areas are unchanged, while the proportion of the changed parts is very small. Then, with this sparsity-induced DI, the binary CM can be treated as an image binary segmentation problem, and many methods for homogeneous CD can be used to extract the changes, such as the Otsu thresholding, principal component analysis with K-means (PCA KM) clustering [38], and the Gabor wavelets with two-level fuzzy c-means clustering (GaborTLC) [7].

Compared with other CD methods, the main advantages of the proposed model are as follows.

- 1) It is based on the widespread nature of image self-similarity, so it has strong flexibility to process a wide variety of different images (homogeneous or heterogeneous) without the main drawbacks of parameter models that require an explicit data distribution knowledge (along with a corresponding complex parameter estimation process), and it can be applied to some complex scenes, where some other hypothetical relationships are no longer applicable as shown in Section V.
- 2) It is unsupervised, and it can overcome the drawbacks of some models that require a large and representative training set (along with a heavily supervised training procedure or a complicated prescreening process of the training set).
- 3) Because it uses patches as processing units rather than individual pixels, the context information is naturally used during the CD process. At the same time, the sparse prior of changes is also employed to obtain the DI, so the proposed method can reduce the impact of noise and get a cleaner and more robust DI.

C. Contribution

The main contributions of this work are summarized as follows.

- 1) We propose a novel CD framework based on the PSGM with the self-expressive property, which measures the change level between the preevent and postevent images from how much one image still conforms to the graph structure of the other image.

- 2) The proposed method can be applied to both heterologous and homologous CD tasks. For the heterologous images, it can establish the relationship between them through a similar graph structure. For the homogeneous CD, because it makes use of the similarity between image patches and the sparse prior knowledge of changes, it can alleviate the impact of noise, especially in the SAR image, so it can obtain a better DI than the direct difference method or ratio/log-ratio method.
- 3) The proposed CD framework is completely unsupervised, and it makes no rigorous assumption, which can be easily adapted for the data across different observation spaces and extended to other heterogeneous image processing problems.

D. Outline and Notation

The overall structure of the study takes the form of six sections, including this introductory section. Section II describes the related similarity learning techniques. Section III introduces the proposed method. Section IV is concerned with some properties and extensions of the proposed method. Section V presents the numerical results. In the end, we provide our conclusion in Section VI.

Here, we define our notation. We define the ℓ_1 -norm, $\ell_{2,1}$ -norm, and Frobenius norm of a matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$ as $\|\mathbf{X}\|_1 = \sum_{i=1}^M \sum_{j=1}^N |x_{i,j}|$, $\|\mathbf{X}\|_{2,1} = \sum_{j=1}^N (\sum_{i=1}^M x_{i,j}^2)^{1/2}$, and $\|\mathbf{X}\|_F = (\sum_{i=1}^M \sum_{j=1}^N x_{i,j}^2)^{1/2}$, respectively. We define the nuclear norm of matrix \mathbf{X} as $\|\mathbf{X}\|_* = \sum_i \sigma_i$, where σ_i is the i th singular value of \mathbf{X} . \mathbf{X}^T is defined as the transpose of \mathbf{X} . $\mathbf{X} \geq 0$ means that all the elements of \mathbf{X}^T are nonnegative. \mathbf{I}_N represents an $N \times N$ identity matrix.

II. PRELIMINARIES

In this section, we give a brief review of some related similarity learning techniques, given a data matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$ with n data points and m features. According to the self-expressive property, each data point can be represented as a linear combination of other points. In the widely used locally linear embedding (LLE) [39], it reconstructs the original data by expressing each data point by its K nearest neighbors, which can be written as

$$\mathbf{x}_i \approx \sum_{j \in N(i)} \mathbf{x}_j z_{j,i} \quad (1)$$

where $N(i)$ represents the neighborhood and $z_{j,i}$ denotes the weight between the j th and i th data points. If \mathbf{x}_j is quite similar to \mathbf{x}_i , the weight coefficient $z_{j,i}$ should be large, and vice versa. From this point of view, the $z_{j,i}$ can be regarded as the similarity between the data points, and thus, $\mathbf{Z} = [z_{j,i}] \in \mathbb{R}^{n \times n}$ is also called similarity graph matrix [40]. Rather than using the predefined neighborhood, \mathbf{Z} can be obtained by solving the following optimization problem:

$$\min_{\mathbf{Z}} \gamma \|\mathbf{X} - \mathbf{XZ}\|_F^2 + f(\mathbf{Z}) \quad (2)$$

where $f(\mathbf{Z})$ is a regularizer of \mathbf{Z} and γ is a balancing parameter. As (2) uses all data points and determines the

\mathbf{Z} automatically according to the optimization result, it is supposed to capture the global structure information of \mathbf{X} . This model has drawn significant attention and achieved impressive performance in many applications, including face recognition [41], subspace clustering [42], and semisupervised learning [43].

III. PSGM-BASED CD

We consider two coregistered heterogeneous images before (pre) and after (post) an event denoted as $\tilde{\mathbf{X}} = \{\tilde{x}(m, n, c) | 1 \leq m \leq M, 1 \leq n \leq N, 1 \leq c \leq C_X\}$ and $\tilde{\mathbf{Y}} = \{\tilde{y}(m, n, c) | 1 \leq m \leq M, 1 \leq n \leq N, 1 \leq c \leq C_Y\}$, which are acquired on the same geographical area by different sensors at different times, respectively. Here, M , N , and C_X (C_Y) are the height, the width, and the number of channels of images, respectively. As aforementioned in Section I, it is impossible to directly compare their pixel values for detecting changes on heterogeneous images, which is due to that different object characteristics are described by different sensors. For this heterogeneous CD, a nature idea is to find a bridge (relationship) between two images, which is consistent in unchanged areas regardless of their completely different low-dimensional pixel values and is obviously different in changed areas.

Based on the self-expressive property, a small patch in the image can be approximately reconstructed by other similar patches within this image. To exploit the nonlocal self-similarity, first, image $\tilde{\mathbf{X}}$ is divided into a number of nonoverlapped squared patches with size $p \times p$, and then, we vectorize and stack them into patch group matrix (PGM) \mathbf{X} with the size $p^2 \times N_p \times C_X$, where $N_p = \lceil M/p \rceil \cdot \lceil N/p \rceil$ with $\lceil \cdot \rceil$ representing an up rounding operation. For the image $\tilde{\mathbf{Y}}$, we follow the same operation to obtain \mathbf{Y} with the size $p^2 \times N_p \times C_Y$. In this way, each patch pair of the original images $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ is still in the same position relationship in the new PGM \mathbf{X} and \mathbf{Y} . Then, for the preevent image $\tilde{\mathbf{X}}$, we describe the global geometry of these patches by linear coefficients that reconstructing each patch from its similar patches, which can be characterized by the similarity graph matrix \mathbf{Z} between patches with $\mathbf{X} = \mathbf{XZ}$. The PSGM \mathbf{Z} can be interpreted as containing the edges of the fully connected graphs with each patch in $\tilde{\mathbf{X}}$ (or each column in \mathbf{X}) as a vertex. Therefore, the core idea of the proposed CD method is that if there is no change between the two images, their patch similarity should be consistent, and then, their PSGM can be transferred as $\mathbf{Y} = \mathbf{YZ}$; conversely, if changes occur within the patch area, then the patch similarity graph structure will change. This can be illustrated in Fig. 2. Intuitively, the more the changes occur in the i th patch \mathbf{X}_i , the less structure of this patch will be conformed by the corresponding patch \mathbf{Y}_i , and then, the more error $|\mathbf{Y}_i - \sum_{j=1}^{N_p} \mathbf{Y}_j z_{j,i}|$ occurs in the reconstruction process. Hence, the change level can be quantified by how dissimilar between the original image \mathbf{Y} (or \mathbf{X}) and the reconstructed image \mathbf{YZ} (or \mathbf{XZ}) based on the learned PSGM from the other image \mathbf{X} (or \mathbf{Y}).

The proposed method consists of five steps: 1) preprocessing; 2) PSGM learning; 3) DI generation; 4) DI optimization

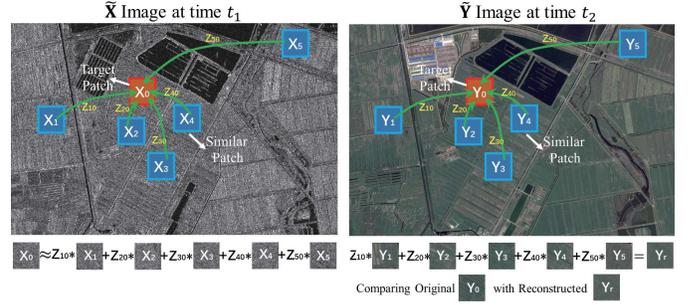


Fig. 2. Illustration of patch similarity graph structure transmission.

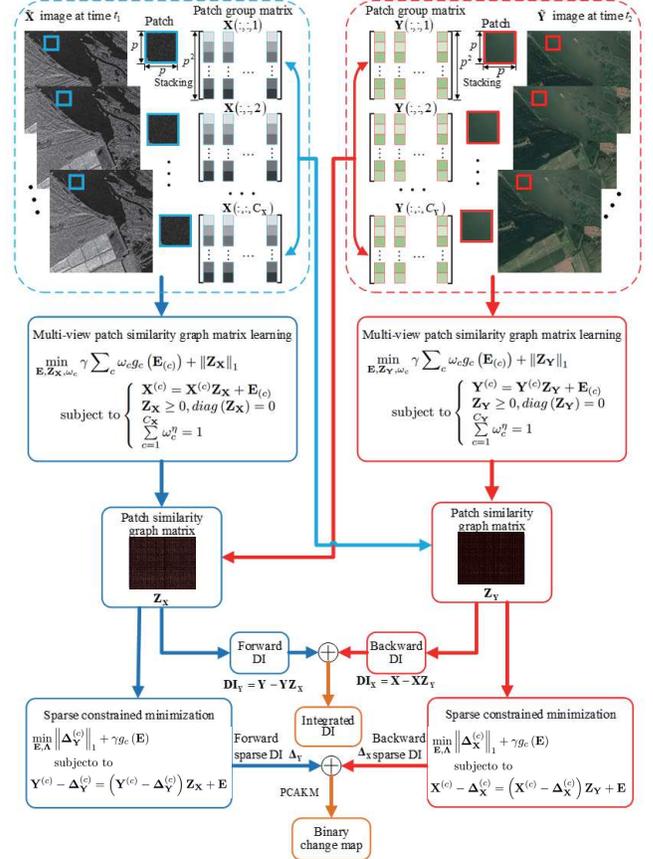


Fig. 3. Framework of the proposed change detection method.

with sparsity regularization; and 5) binary CM generation with clustering method. The framework is illustrated in Fig. 3.

A. Preprocessing

In the preprocessing, the main task is to generate the PGM \mathbf{X} and \mathbf{Y} from the original images $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ using the above-described procedure, where the patch pairs of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ at the same location are stacked with a one-to-one correspondence at the same column in \mathbf{X} and \mathbf{Y} . Moreover, two aspects need to be paid attention to in this process. First, for SAR images with large dynamic range, logarithmic transformation can be used to enhance the structures with small intensity values. Second, when the noise is serious, some filtering operations can be used. This is mainly for the SAR image, as well known,

the imaging quality of the optical sensor is generally better than that of the SAR sensor. SAR is inherently affected by speckle noise [44], which strongly degrades the appearance of images visually and diminishes the performances of subsequent automated tasks. Thus, when the SAR image suffers from serious speckle noise, some classical and useful filters, such as the Lee filter [45], the Frost filter [46], and the Kuan filter [47], can be selected to get better visual effect images and make the relationship between image patches more obvious. These two additional operations of logarithmic transformation and despeckling filter can be added automatically as needed by using simple discrimination.

B. PSGM Learning

To capture the global structure information of patches in the preevent image, we need to solve the PSGM \mathbf{Z}_X with the preprocessed \mathbf{X} . Here, we consider two cases: the simple single-channel case and the complicated multichannel case.

For the simple case, when the preevent image is single channel, that is, $C_X = 1$, we use the following optimization for calculating \mathbf{Z}_X :

$$\begin{aligned} & \min_{\mathbf{E}, \mathbf{Z}_X} \gamma g(\mathbf{E}) + \|\mathbf{Z}_X\|_1 \\ & \text{s.t. } \mathbf{X} = \mathbf{X}\mathbf{Z}_X + \mathbf{E}, \mathbf{Z}_X \geq 0, \text{diag}(\mathbf{Z}_X) = 0 \end{aligned} \quad (3)$$

where $\gamma > 0$ is a balancing parameter and \mathbf{E} represents unknown corruption. $\|\mathbf{Z}_X\|_1$ is the sparse penalty term, which is based on the fact that each patch can be reconstructed by a few similar patches and other patches with little similarity have no contribution to this reconstruction. $g(\mathbf{E})$ represents the data fidelity term, and it can be squared Frobenius norm, ℓ_1 -norm, or $\ell_{2,1}$ -norm. Especially, if the noise obeys the Gaussian distribution, $\|\mathbf{E}\|_F^2$ is more suitable; if random corruption, such as the impulse noise, is assumed in the data, $\|\mathbf{E}\|_1$ is usually adopted; and $\|\mathbf{E}\|_{2,1}$ is more suitable to characterize sample-specific corruptions and outliers. Here, we also restrict the reconstruction weights \mathbf{Z}_X to be nonnegative for ease of interpretation, and the constraint $\text{diag}(\mathbf{Z}_X) = 0$ means that all the diagonal elements of \mathbf{Z}_X are zero, which is used to avoid the numerically trivial solution $\mathbf{Z}_X = \mathbf{I}$.

Problem (3) can be efficiently solved by the alternating direction method of multipliers (ADMM). To make the objective function (3) separable, we first introduce an auxiliary variable $\mathbf{J} \in \mathbb{R}^{N_p \times N_p}$ and rewrite the model (3) as its equivalent form with auxiliary matrix

$$\begin{aligned} & \min_{\mathbf{E}, \mathbf{J}} \gamma g(\mathbf{E}) + \|\mathbf{J}\|_1 \\ & \text{s.t. } \mathbf{X} = \mathbf{X}\mathbf{Z}_X + \mathbf{E}, \mathbf{J} = \mathbf{Z}_X, \mathbf{Z}_X \geq 0, \text{diag}(\mathbf{Z}_X) = 0. \end{aligned} \quad (4)$$

The corresponding augmented Lagrangian function of (4), which is a combination of the Lagrangian and quadratic penalty functions, can be written as

$$\begin{aligned} L(\mathbf{E}, \mathbf{J}, \mathbf{Z}_X, \mathbf{W}_1, \mathbf{W}_2, \mu) &= \gamma g(\mathbf{E}) + \|\mathbf{J}\|_1 + \text{Tr}(\mathbf{W}_1^T (\mathbf{J} - \mathbf{Z}_X)) \\ &+ \text{Tr}(\mathbf{W}_2^T (\mathbf{X} - \mathbf{X}\mathbf{Z}_X - \mathbf{E})) \\ &+ \frac{\mu}{2} \{ \|\mathbf{J} - \mathbf{Z}_X\|_F^2 + \|\mathbf{X} - \mathbf{X}\mathbf{Z}_X - \mathbf{E}\|_F^2 \} \\ &\text{s.t. } \mathbf{Z}_X \geq 0, \text{diag}(\mathbf{Z}_X) = 0 \end{aligned} \quad (5)$$

where \mathbf{W}_1 and \mathbf{W}_2 are Lagrange multipliers, and $\mu > 0$ is a penalty parameter. Then, we can use the alternating direction method (ADM) to solve the minimization of (5) by separating it into \mathbf{E} -subproblem, \mathbf{J} -subproblem, and \mathbf{Z}_X -subproblem.

First, given the current points $(\mathbf{E}^k, \mathbf{J}^k, \mathbf{Z}_X^k, \mathbf{W}_1^k, \mathbf{W}_2^k)$ at the k th iteration, the minimization of (5) with respect to \mathbf{E} can be formulated as

$$\begin{aligned} \mathbf{E}^{k+1} = \arg \min_{\mathbf{E}} \{ & \gamma g(\mathbf{E}) + \text{Tr}((\mathbf{W}_2^k)^T (\mathbf{X} - \mathbf{X}\mathbf{Z}_X^k - \mathbf{E})) \\ & + \frac{\mu}{2} \|\mathbf{X} - \mathbf{X}\mathbf{Z}_X^k - \mathbf{E}\|_F^2 \}. \end{aligned} \quad (6)$$

This can be solved by the proximal operator as

$$\mathbf{E}^{k+1} = \text{prox}_{\frac{\gamma}{\mu} g}(\mathbf{Q}^k) \quad (7)$$

where $\mathbf{Q}^k = \mathbf{X} - \mathbf{X}\mathbf{Z}_X^k + (\mathbf{W}_2^k/\mu)$, and the proximal operation is defined as

$$\text{prox}_{\alpha g}(\mathbf{D}) = \arg \min_{\mathbf{X}} g(\mathbf{X}) + \frac{1}{2\alpha} \|\mathbf{X} - \mathbf{D}\|_F^2. \quad (8)$$

Depending on different regularization strategies, we have different closed-form solutions for updating \mathbf{E}^{k+1} . If we choose $g(\mathbf{E}) = \|\mathbf{E}\|_F^2$, we have

$$\mathbf{E}^{k+1} = \frac{\mathbf{Q}^k}{1 + 2\gamma/\mu}. \quad (9)$$

If we choose $g(\mathbf{E}) = \|\mathbf{E}\|_1$, we can update \mathbf{E}^{k+1} as

$$\mathbf{E}^{k+1} = \text{shrink} \left\{ \mathbf{Q}^k, \frac{\gamma}{\mu} \right\} \quad (10)$$

where $\text{shrink}(\mathbf{X}, \lambda)$ denotes the elementwise soft shrinkage operator given in [48] as

$$[\text{shrink}(\mathbf{X}, \lambda)]_{i,j} = \text{sign}(x_{i,j}) \max\{|x_{i,j}| - \lambda, 0\}. \quad (11)$$

If we choose $g(\mathbf{E}) = \|\mathbf{E}\|_{2,1}$ in [49], we have

$$[\mathbf{E}^{k+1}]_{:,i} = \max \left\{ \|\mathbf{Q}_{:,i}\|_2 - \frac{\gamma}{\mu} \right\} \frac{\mathbf{Q}_{:,i}}{\|\mathbf{Q}_{:,i}\|_2} \quad (12)$$

where it follows the convention $0 \cdot (0/0) = 0$.

Second, with the fixed $(\mathbf{E}^{k+1}, \mathbf{Z}_X^k, \mathbf{W}_1^k, \mathbf{W}_2^k)$, the minimization of (5) with respect to \mathbf{J} can be formulated as

$$\begin{aligned} \mathbf{J}^{k+1} = \arg \min_{\mathbf{J}} \{ & \|\mathbf{J}\|_1 + \text{Tr}((\mathbf{W}_1^k)^T (\mathbf{J} - \mathbf{Z}_X^k)) \\ & + \frac{\mu}{2} \|\mathbf{J} - \mathbf{Z}_X^k\|_F^2 \}. \end{aligned} \quad (13)$$

Then, \mathbf{J} can be updated elementwisely by

$$\mathbf{J}^{k+1} = \text{shrink} \left\{ \mathbf{Z}_X^k - \frac{\mathbf{W}_1^k}{\mu}, \frac{1}{\mu} \right\}. \quad (14)$$

Third, with the fixed $(\mathbf{E}^{k+1}, \mathbf{J}^{k+1}, \mathbf{W}_1^k, \mathbf{W}_2^k)$, we can update \mathbf{Z}_X^{k+1} as

$$\begin{aligned} \mathbf{Z}_X^{k+1} &= \arg \min_{\mathbf{Z}_X} \{ \text{Tr}((\mathbf{W}_1^k)^T (\mathbf{J}^{k+1} - \mathbf{Z}_X)) \\ &+ \text{Tr}((\mathbf{W}_2^k)^T (\mathbf{X} - \mathbf{X}\mathbf{Z}_X - \mathbf{E}^{k+1})) \\ &+ \frac{\mu}{2} \{ \|\mathbf{J}^{k+1} - \mathbf{Z}_X\|_F^2 + \|\mathbf{X} - \mathbf{X}\mathbf{Z}_X - \mathbf{E}^{k+1}\|_F^2 \} \}. \end{aligned} \quad (15)$$

TABLE I
MAIN IMPLEMENTATION STEPS OF PSGM LEARNING
FOR SINGLE-CHANNEL IMAGE

Algorithm 1. PSGM for single channel
Input: Patch group matrix \mathbf{X} , parameters $\gamma > 0$, $\mu > 0$.
Initialize: Random matrix \mathbf{Z}_X , and $\mathbf{W}_1, \mathbf{W}_2 = 0$.
Repeat:
1: Update \mathbf{E}^{k+1} through (9), (10), (12) according to different regularization strategies.
2: Update \mathbf{J}^{k+1} through (14).
3: Update \mathbf{Z}_X^{k+1} through (16); set $\mathbf{Z}_X \leftarrow \max\{\mathbf{Z}_X, 0\}$, $\text{diag}(\mathbf{Z}_X) \leftarrow 0$.
4: Update the Lagrangian multipliers through (17) and (18).
Until stopping criterion is met.

By setting its first-order derivative to zero, we can solve (15) by

$$\mathbf{Z}_X^{k+1} = (\mathbf{I}_{N_p} + \mathbf{X}^T \mathbf{X})^{-1} \times \left(\mathbf{X}^T (\mathbf{X} - \mathbf{E}^{k+1}) + \mathbf{J}^{k+1} + \frac{\mathbf{W}_1^k + \mathbf{X}^T \mathbf{W}_2^k}{\mu} \right). \quad (16)$$

Finally, we update the Lagrangian multipliers as

$$\mathbf{W}_1^{k+1} = \mathbf{W}_1^k + \mu (\mathbf{J}^{k+1} - \mathbf{Z}_X^{k+1}) \quad (17)$$

$$\mathbf{W}_2^{k+1} = \mathbf{W}_2^k + \mu (\mathbf{X} - \mathbf{X} \mathbf{Z}_X^{k+1} - \mathbf{E}^{k+1}). \quad (18)$$

The complete procedure is outlined in Table I. As a matter of fact, due to the columnwise independence property of \mathbf{Z}_X , problem (5) can be easily accelerated by columnwisely parallel solution.

For the multichannel case, that is, $C_X > 1$, we take PSGM \mathbf{Z}_X learning as a multiview fusion problem. Then, for each channel of \mathbf{X} , denote $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(C_X)}$ as the C_X views of the same area, where the c th view matrix $\mathbf{X}^{(c)} \in \mathbb{R}^{p^2 \times N_p}$ is defined as $\mathbf{X}^{(i)} = \mathbf{X}(:, :, c)$. It is obvious that a straightforward way is to stack up these channels into a new one and then put it into the single-channel model (3). However, in this way, the importance and uniqueness of different channels are ignored, and when some unreliable channels are considered, the learning performance will even degraded seriously. Here, we combine these views linearly with autoweighting parameters ω_c ($c = 1, \dots, C_X$). For all views, we learn a unified PSGM, which can be formulated by the following optimization problem:

$$\begin{aligned} \min_{\mathbf{E}, \mathbf{Z}_X, \omega_c} \quad & \gamma \sum_c \omega_c g_c(\mathbf{E}_{(c)}) + \|\mathbf{Z}_X\|_1 \\ \text{s.t.} \quad & \mathbf{X}^{(c)} = \mathbf{X}^{(c)} \mathbf{Z}_X + \mathbf{E}_{(c)}, \quad \mathbf{Z}_X \geq 0, \quad \text{diag}(\mathbf{Z}_X) = 0 \\ & \sum_{c=1}^{C_X} \omega_c^\eta = 1, \quad \omega_c \geq 0 \end{aligned} \quad (19)$$

where parameter $\eta \in (0, 1)$ is used to control the distribution of weights for different views.

Similar to the single-channel case, we select the ADMM to solve this minimization problem. By introducing the auxiliary variable $\mathbf{J} \in \mathbb{R}^{N_p \times N_p}$ and using the Lagrange multipliers \mathbf{W}_1 and $\mathbf{W}_{2(c)}$, the augmented Lagrangian function of (19)

can be written as

$$\begin{aligned} \min_{\mathbf{E}_{(c)}, \mathbf{J}, \mathbf{Z}_X, \omega_c} \quad & \left\{ L(\mathbf{E}_{(c)}, \mathbf{J}, \mathbf{Z}_X, \omega_c, \mathbf{W}_1, \mathbf{W}_{2(c)}) \right. \\ & := \|\mathbf{J}\|_1 + \text{Tr}(\mathbf{W}_1^T (\mathbf{J} - \mathbf{Z}_X)) + \frac{\mu}{2} \|\mathbf{J} - \mathbf{Z}_X\|_F^2 \\ & + \sum_c \gamma \omega_c g_c(\mathbf{E}_{(c)}) \\ & + \text{Tr}(\mathbf{W}_{2(c)}^T (\mathbf{X}^{(c)} - \mathbf{X}^{(c)} \mathbf{Z}_X - \mathbf{E}_{(c)})) \\ & \left. + \sum_c \frac{\mu}{2} \|\mathbf{X}^{(c)} - \mathbf{X}^{(c)} \mathbf{Z}_X - \mathbf{E}_{(c)}\|_F^2 \right\} \\ \text{s.t.} \quad & \sum_{c=1}^{C_X} \omega_c^\eta = 1 \end{aligned} \quad (20)$$

we can solve (20) by alternatively updating $\mathbf{E}_{(c)}$, \mathbf{J} , \mathbf{Z}_X , and ω_c while fixing other variables as constant. Given the current points $(\mathbf{E}_{(c)}, \mathbf{J}, \mathbf{Z}_X, \omega_c, \mathbf{W}_1, \mathbf{W}_{2(c)})$ at the k th iteration, the updating of $\mathbf{E}_{(c)}$, \mathbf{J} , and \mathbf{Z}_X is similar to Algorithm 1 in Table I, which can be written as

$$\begin{aligned} \mathbf{E}_{(c)}^{k+1} = \arg \min_{\mathbf{E}_{(c)}} \quad & \left\{ \gamma \omega_c^k g_c(\mathbf{E}_{(c)}) + \frac{\mu}{2} \|\mathbf{X}^{(c)} - \mathbf{X}^{(c)} \mathbf{Z}_X^k - \mathbf{E}_{(c)}\|_F^2 \right. \\ & \left. + \text{Tr}((\mathbf{W}_{2(c)}^k)^T (\mathbf{X}^{(c)} - \mathbf{X}^{(c)} \mathbf{Z}_X^k - \mathbf{E}_{(c)})) \right\} \end{aligned} \quad (21)$$

$$\mathbf{J}^{k+1} = \arg \min_{\mathbf{J}} \left\{ \|\mathbf{J}\|_1 + \text{Tr}((\mathbf{W}_1^k)^T (\mathbf{J} - \mathbf{Z}_X^k)) + \frac{\mu}{2} \|\mathbf{J} - \mathbf{Z}_X^k\|_F^2 \right\} \quad (22)$$

$$\begin{aligned} \mathbf{Z}_X^{k+1} = \arg \min_{\mathbf{Z}_X} \quad & \left\{ \text{Tr}((\mathbf{W}_1^k)^T (\mathbf{J}^{k+1} - \mathbf{Z}_X)) + \frac{\mu}{2} \|\mathbf{J}^{k+1} - \mathbf{Z}_X\|_F^2 \right. \\ & + \sum_c \text{Tr}((\mathbf{W}_{2(c)}^k)^T (\mathbf{X}^{(c)} - \mathbf{X}^{(c)} \mathbf{Z}_X - \mathbf{E}_{(c)}^{k+1})) \\ & \left. + \frac{\mu}{2} \|\mathbf{X}^{(c)} - \mathbf{X}^{(c)} \mathbf{Z}_X - \mathbf{E}_{(c)}^{k+1}\|_F^2 \right\}. \end{aligned} \quad (23)$$

For $\mathbf{E}_{(c)}^{k+1}$ updating, it can be solved by the proximal operator as

$$\mathbf{E}_{(c)}^{k+1} = \text{prox}_{\frac{\gamma}{\mu} \omega_c^k g_c}(\mathbf{Q}_{(c)}^k) \quad (24)$$

where $\mathbf{Q}_{(c)}^k = \mathbf{X}^{(c)} - \mathbf{X}^{(c)} \mathbf{Z}_X^k + (\mathbf{W}_{2(c)}^k / \mu)$, and the closed-form solutions of different g_c can be obtained as (9), (10), and (12) by replacing γ with $\gamma \omega_c^k$.

The \mathbf{J}^{k+1} updating is the same as in Algorithm 1, which can be updated elementwisely by (14).

The \mathbf{Z}_X^{k+1} updating can be solved by setting its first-order derivative to zero, and then, we have

$$\begin{aligned} \mathbf{Z}_X^{k+1} = \left(\mathbf{I}_{N_p} + \sum_c (\mathbf{X}^{(c)})^T \mathbf{X}^{(c)} \right)^{-1} \\ \times \left(\mathbf{J}^{k+1} + \frac{\mathbf{W}_1^k}{\mu} + \sum_c (\mathbf{X}^{(c)})^T (\mathbf{X}^{(c)} - \mathbf{E}_{(c)}) \right. \\ \left. + \frac{(\mathbf{X}^{(c)})^T \mathbf{W}_{2(c)}^k}{\mu} \right). \end{aligned} \quad (25)$$

Then, we consider the autoweighting parameters ω_c . With the fixed $(\mathbf{E}_{(c)}, \mathbf{J}, \mathbf{Z}_X, \mathbf{W}_1, \mathbf{W}_{2(c)})$, the optimization of (20) with respect to ω_c can be simplified as the following problem:

$$\min \sum_c \omega_c g_c(\mathbf{E}_{(c)}) \quad \text{s.t.} \quad \sum_{c=1}^{C_X} \omega_c^\eta = 1, \quad \omega_c \geq 0. \quad (26)$$

TABLE II
MAIN IMPLEMENTATION STEPS OF PSGM LEARNING
FOR MULTICHANNEL IMAGE

Algorithm 2. PSGM for multi-channel image
Input: A set of patch group matrices $\{\mathbf{X}^{(c)}\}_{c=1}^{C_X}$, parameters $\gamma > 0, \mu > 0, \eta \in (0, 1)$.
Initialize: Random matrix \mathbf{Z}_X , and $\mathbf{W}_1, \mathbf{W}_2 = 0, \omega_c = 1/C_X$.
Repeat:
1: Update \mathbf{E}^{k+1} through (24) according to different regularization strategies.
2: Update \mathbf{J}^{k+1} through (14).
3: Update \mathbf{Z}_X^{k+1} through (25); set $\mathbf{Z}_X \leftarrow \max\{\mathbf{Z}_X, 0\}$, $\text{diag}(\mathbf{Z}_X) \leftarrow 0$.
4: Calculate the reconstruction error $h_c = g_c(\mathbf{E}^{k+1})$.
5: Update the weighting parameters ω_c^{k+1} through (28).
6: Update the Lagrangian multipliers.
Until stopping criterion is met.

The Lagrange function of (26) is

$$L(\omega_c) = \sum_c \omega_c h_c + \lambda \left(1 - \sum_c \omega_c^\eta \right) \quad (27)$$

where $h_c = g_c(\mathbf{E}^{(c)})$. By using the Karush–Kuhn–Tucker (KKT) condition $(\partial L(\omega_c)/\partial \omega_c) = 0$ and the constraint $\sum_{c=1}^{C_X} \omega_c^\eta = 1$, the optimal solution of ω_c is

$$\omega_c = h_c^{-\frac{1}{\eta-1}} \left(\sum_c h_c^{-\frac{\eta}{\eta-1}} \right)^{-\frac{1}{\eta}}. \quad (28)$$

From (28), we can find that the channel with smaller reconstruction error (h_c) will get larger weighting coefficient, which means that a reliable channel will play a greater role in the PGSM learning process. We can also observe that when $\eta \rightarrow 1$, we will get equal weights, and when $\eta \rightarrow 0$, we will set the channel whose h_c is the smallest to 1 and set the other channels to 0. In Table II, we provide the complete algorithm for the multichannel PSGM learning.

C. DI Generation

When we obtain the PSGM \mathbf{Z}_X that represents the structure information of patches in the preevent image, we can measure the change level by comparing the structure consistency of the postevent image with that of the preevent image. For the single-channel postevent image $\tilde{\mathbf{Y}}$, that is, $C_Y = 1$, we have the forward DI as $DI_Y = |\mathbf{Y} - \mathbf{Y}\mathbf{Z}_X|$; for the multichannel $\tilde{\mathbf{Y}}$, we calculate the reconstruction error of each channel as $DI_{Y^{(c)}} = |\mathbf{Y}^{(c)} - \mathbf{Y}^{(c)}\mathbf{Z}_X|$, and then, we fuse these errors to obtain the DI as $[DI_Y]_{i,j} = (\sum_{c=1}^{C_Y} ([DI_{Y^{(c)}}]_{i,j})^2)^{1/2}$. After the reverse operations of vectorizing and stacking, we reshape the DI_Y with size $p^2 \times N_p$ back to the DI $\tilde{D}\tilde{I}_Y$ with size $M \times N$.

The backward DI $\tilde{D}\tilde{I}_X$ that associates the preevent image $\tilde{\mathbf{X}}$ with the postevent image PSGM \mathbf{Z}_Y can be similarly obtained as the generation of the forward DI. Then, the integrated DI can be given by

$$\tilde{D}\tilde{I} = \tilde{D}\tilde{I}_X + \tilde{D}\tilde{I}_Y. \quad (29)$$

D. DI Optimization With Sparsity Regularization

Based on the fact that most of the areas in the image are unchanged and only a small part of them changed, then we can use this prior sparsity to optimize the DI.

For the forward comparison, let the changed part be $\Delta_Y^{(c)}$ for each channel $\mathbf{Y}^{(c)}$; then, with the learned \mathbf{Z}_X , we have

$$\mathbf{Y}^{(c)} - \Delta_Y^{(c)} = (\mathbf{Y}^{(c)} - \Delta_Y^{(c)})\mathbf{Z}_X + \mathbf{E} \quad (30)$$

where \mathbf{E} represents the unknown corruption. According to sparse constraints of $\Delta_Y^{(c)}$, we have the following optimization problem:

$$\begin{aligned} \min_{\mathbf{E}, \Lambda} \quad & \|\Lambda\|_1 + \gamma g_c(\mathbf{E}) \\ \text{s.t.} \quad & \mathbf{Y}^{(c)} - \Delta_Y^{(c)} = (\mathbf{Y}^{(c)} - \Delta_Y^{(c)})\mathbf{Z}_X + \mathbf{E}, \quad \Lambda = \Delta_Y^{(c)}. \end{aligned} \quad (31)$$

By using the ADMM, the augmented Lagrangian function of (31) can be formulated as

$$\begin{aligned} L(\mathbf{E}, \Lambda, \Delta_Y^{(c)}, \mathbf{W}_1, \mathbf{W}_2) &= \|\Lambda\|_1 + \gamma g_c(\mathbf{E}) + \text{Tr}(\mathbf{W}_2^T (\Lambda - \Delta_Y^{(c)})) \\ &+ \text{Tr}(\mathbf{W}_1^T ((\mathbf{Y}^{(c)} - \Delta_Y^{(c)}) (\mathbf{I}_{N_p} - \mathbf{Z}_X) - \mathbf{E})) \\ &+ \frac{\mu}{2} \{ \|(\mathbf{Y}^{(c)} - \Delta_Y^{(c)}) (\mathbf{I}_{N_p} - \mathbf{Z}_X) - \mathbf{E}\|_F^2 + \|\Lambda - \Delta_Y^{(c)}\|_F^2 \} \end{aligned} \quad (32)$$

where \mathbf{W}_1 and \mathbf{W}_2 are Lagrange multipliers. Given the current point $(\mathbf{E}^k, \Lambda^k, \Delta_Y^{(c)k}, \mathbf{W}_1^k, \mathbf{W}_2^k)$, the updating scheme is

$$\mathbf{E}^{k+1} = \text{prox}_{\frac{1}{\mu} g_c} \left((\mathbf{Y}^{(c)} - \Delta_Y^{(c)k}) (\mathbf{I}_{N_p} - \mathbf{Z}_X) + \frac{\mathbf{W}_1^k}{\mu} \right) \quad (33)$$

$$\Lambda^{k+1} = \text{shrink} \left\{ \Delta_Y^{(c)k} - \frac{\mathbf{W}_2^k}{\mu}, \frac{1}{\mu} \right\} \quad (34)$$

$$\begin{aligned} \Delta_Y^{(c)k+1} &= \left(\Lambda^{k+1} + \mathbf{S}_1 - \mathbf{E}^{k+1} (\mathbf{I}_{N_p} - \mathbf{Z}_X)^T \right. \\ &\quad \left. + \frac{\mathbf{W}_1^k (\mathbf{I}_{N_p} - \mathbf{Z}_X)^T + \mathbf{W}_2^k}{\mu} \right) \mathbf{S}_2 \end{aligned} \quad (35)$$

$$\mathbf{W}_1^{k+1} = \mathbf{W}_1^k + \mu ((\mathbf{Y}^{(c)} - \Delta_Y^{(c)k+1}) (\mathbf{I}_{N_p} - \mathbf{Z}_X) - \mathbf{E}^{k+1}) \quad (36)$$

$$\mathbf{W}_2^{k+1} = \mathbf{W}_2^k + \mu (\Lambda^{k+1} - \Delta_Y^{(c)k+1}) \quad (37)$$

where $\mathbf{S}_1 = \mathbf{Y}^{(c)} (\mathbf{I}_{N_p} - \mathbf{Z}_X) (\mathbf{I}_{N_p} - \mathbf{Z}_X)^T$ and $\mathbf{S}_2 = (\mathbf{I}_{N_p} + (\mathbf{I}_{N_p} - \mathbf{Z}_X) (\mathbf{I}_{N_p} - \mathbf{Z}_X)^T)^{-1}$ in the $\Delta_Y^{(c)k+1}$ updating (35), and the \mathbf{E}^{k+1} updating with the proximal operator has different closed-form solutions for g_c as in (7) based on different strategies. Since the minimization problem (32) is convex, the convergence of the ADMM can be guaranteed [50]. Meanwhile, although the calculation form for $\Delta_Y^{(c)k+1}$ updating appears to be complicated, some formulas remain unchanged during the update process, such as \mathbf{S}_1 and \mathbf{S}_2 , so they can be calculated off-line and stored in advance to reduce the computational cost. After getting the change part $\Delta_Y^{(c)}$ of each channel, we have the forward sparsity-induced DI Δ_Y as

$$[\Delta_Y]_{i,j} = \left(\sum_{c=1}^{C_Y} ([\Delta_Y^{(c)}]_{i,j})^2 \right)^{1/2}. \quad (38)$$

In summary, we present the alternating iterative algorithm for minimization problem (32) in Table III (Algorithm 3).

TABLE III
MAIN IMPLEMENTATION STEPS FOR GENERATING
SPARSITY-INDUCED DI

Algorithm 3. ADMM for sparsity induced DI
Input: A set of patch group matrices $\{\mathbf{Y}^{(c)}\}_{c=1}^{C_Y}$, learned PGM \mathbf{Z}_X , parameters $\gamma > 0, \mu > 0$.
Initialize: Random matrix $\Delta_Y^{(c)}$, and $\mathbf{W}_1, \mathbf{W}_2 = 0$.
For each channel, Repeat:
1: Update \mathbf{E}^{k+1} through (33) according to different regularization strategies.
2: Update Λ^{k+1} through (34).
3: Update $\Delta_Y^{(c)k+1}$ through (35).
4: Update the Lagrangian multipliers through (36) and (37).
Until stopping criterion is met.
Output: Difference image Δ_Y through (38).

Similarly, we can obtain the backward sparsity-induced DI Δ_X by using Algorithm 3 with the preevent image \mathbf{X} and the learned \mathbf{Z}_Y . Then, the sparsity-induced integrated DI can be given by

$$\Delta = \Delta_X + \Delta_Y. \quad (39)$$

E. Binary CM Generation

After the sparse DI is obtained, the CD can be treated as an image binary segmentation problem. The clustering method is widely used for its simplicity and effectiveness, which is a process of grouping a set of features into meaningful clusters. In order to extract the features, PCA [38], compressed projection [51], and Gabor wavelets [7] are often used. Then, the clustering methods, such as K-means clustering [52] and fuzzy c-means (FCM) clustering [53], are often employed to obtain the final result. In this article, we select the PCAKM algorithm [38] to obtain the final binary CM.

IV. EXTENSIONS

In this section, we discuss some properties of the proposed CD method and simply extend this model to other remote sensing image processing problems, which can enlarge the application scope of the proposed algorithm.

A. Complexity Analysis

The main computational complexity of the proposed PSGM-based CD method is concentrating on the process of PSGM learning (Algorithm 1 or 2) and DI optimization with sparsity regularization (Algorithms 3). Because the iterative schemes of these algorithms are similar, we only analyze one of them and take Algorithm 1 as an example.

- 1) *Updating E*: As the proximal operation has closed-form solutions for these three types of $g(\mathbf{E})$, updating \mathbf{E} requires $\mathcal{O}(p^2 N_p^2)$ for the matrix multiplication.
- 2) *Updating J*: As \mathbf{J} can be updated element wisely by the soft shrinkage operator in (14), it requires $\mathcal{O}(N_p^2)$.
- 3) *Updating Z_X in (16)*: The matrix inversion of $(\mathbf{I}_{N_p} + \mathbf{X}^T \mathbf{X})^{-1}$ requires $\mathcal{O}(N_p^3)$, the matrix multiplication in brackets of $\mathbf{X}^T (\mathbf{X} - \mathbf{E})$ requires $\mathcal{O}(p^2 N_p^2)$, and the matrix multiplication outside the brackets with the

inverse matrix requires $\mathcal{O}(N_p^3)$. Then, updating \mathbf{Z}_X needs $\mathcal{O}(N_p^3)$. However, as $(\mathbf{I}_{N_p} + \mathbf{X}^T \mathbf{X})^{-1}$ can be simplified by the Sherman–Morrison–Woodbury formula as

$$(\mathbf{I}_{N_p} + \mathbf{X}^T \mathbf{X})^{-1} = \mathbf{I}_{N_p} - \mathbf{X}^T (\mathbf{I}_{p^2} + \mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \quad (40)$$

and it can be calculated off-line in advance, the computational cost of updating \mathbf{Z}_X can be greatly reduced.

- 4) *Updating Lagrangian Multipliers W₁ and W₂*: It requires $\mathcal{O}(p^2 N_p^2)$.

Although the complexity of the algorithm is very high in the abovementioned theoretical analysis, which requires $\mathcal{O}(N_p^3)$, some acceleration methods can be used, such as the accelerated linearized ADMM [54]. Meanwhile, as mentioned earlier, the PSGM learning process can also be accelerated by adopting the columnwisely parallel solution.

B. Extending to Low-Rank Model

To exploit the structure information of the patches, we solve the PSGM with sparse regularization, which encourages patch reconstruction from the similar patches. Meanwhile, take the preevent image for example, for each patch in $\tilde{\mathbf{X}}$ (e.g., column \mathbf{X}_i), it has many similar patches in $\tilde{\mathbf{X}}$ (e.g., columns \mathbf{X}_j), and then, the reconstruction forms of these similar patches should also be similar, that is, the i th and j th columns in \mathbf{Z}_X have similar expressions. From this, we can infer that the PSGM \mathbf{Z}_X should have low rank property. Replacing sparse regularization $\|\mathbf{Z}_X\|_1$ with low rank regularization $\|\mathbf{Z}_X\|_*$, model (3) can be rewritten as

$$\begin{aligned} \min_{\mathbf{E}, \mathbf{Z}_X} \gamma g(\mathbf{E}) + \|\mathbf{Z}_X\|_* \\ \text{s.t. } \mathbf{X} = \mathbf{X} \mathbf{Z}_X + \mathbf{E}, \mathbf{Z}_X \geq 0, \text{diag}(\mathbf{Z}_X) = 0 \end{aligned} \quad (41)$$

which can be solved by using the same framework as Algorithm 1, except that the \mathbf{J} -subproblem becomes

$$\begin{aligned} \mathbf{J}^{k+1} = \arg \min_{\mathbf{J}} \{ \|\mathbf{J}\|_* + \text{Tr}((\mathbf{W}_1^k)^T (\mathbf{J} - \mathbf{Z}_X^k)) \\ + \frac{\mu}{2} \|\mathbf{J} - \mathbf{Z}_X^k\|_F^2 \}. \end{aligned} \quad (42)$$

Then, we can update \mathbf{J}^{k+1} by the singular value thresholding function as [55]

$$\mathbf{J}^{k+1} = \mathbf{U} \text{shrink} \left\{ \Sigma, \frac{1}{\mu} \right\} \mathbf{V}^T \quad (43)$$

where $\mathbf{U} \Sigma \mathbf{V}^T$ is the singular value decomposition (SVD) of $\mathbf{Z}_X^k - (\mathbf{W}_1^k / \mu)$, and the soft shrinkage operator is defined in (11). Then, we can obtain a low-rank PSGM. However, since SVD is required to compute (43), it increases the computational cost, which makes the method limited to small-scale images.

C. Extending to Other Applications

First, as the PSGM represents the structure information of patches, it can obviously be applied to remote sensing image clustering. Take the optical image for example, once we learn the PSGM \mathbf{Z} by using Algorithm 2, and then, we can obtain the symmetrized matrix $\mathbf{W} = (\mathbf{Z} + \mathbf{Z}^T) / 2$, and the spectral

clustering algorithm [56] can be applied to \mathbf{W} to get the patch-based clustering results. This model is different from the graph-based clustering method proposed in [57], where it takes each pixel as a vertex and the corresponding values of all spectral bands as a feature vector. However, here, we take each patch as a vertex and the pixel values as a feature vector, so this method can be applied to unsupervised single or multichannel remote sensing image classification.

Second, the PSGM can also be used in the remote sensing image regression. Suppose that we have some patch pairs-based training data set as $\mathbf{T} = \{(\mathbf{X}_i, \mathbf{Y}_i) | i \in \Gamma_t\}$, and then, we need to predict the rest patches $\{\mathbf{Y}_j | j \in \Gamma_p\}$ with known $\{\mathbf{X}_j | j \in \Gamma_p\}$. One possible method is to learn the PSGM \mathbf{Z}_X with model (3) but set the j th row ($j \in \Gamma_p$) of the \mathbf{Z}_X to be zero. We can use $\hat{\mathbf{Y}}_j = \sum_{i \in \Gamma_t} \mathbf{Y}_i z_{j,i}$ to obtain the preliminary predictive value of \mathbf{Y}_j and then further employ other correction processing to make it closer to the real value.

V. EXPERIMENTAL ANALYSIS

In this section, simulations are performed to demonstrate the proposed conclusions and evaluate the performance of the proposed PSGM-based CD method. The data fidelity term $g(\mathbf{E})$ is selected according to different situations. In general, for the optical image, we choose the squared Frobenius norm; for the SAR image, we choose the ℓ_1 -norm. For the ADMM frameworks, we stop the program when a maximum of 100 iterations, or a relative difference of 10^{-5} is reached; for other parameters used in the proposed method, we will stay in Section V-D for detailed analysis.

A. Experiments on Heterogeneous Data Sets

To evaluate the efficiency of the proposed model, we first validate it on three real pairs of heterogeneous remote sensing images.

- 1) The first heterogeneous data set consists of two near-infrared (NIR) band images and one optical image (Sardinia, Italy), as shown in Fig. 4. The two NIR band images, with $300 \times 412 \times 1$ pixels, are acquired by Landsat-5 in September 1995 and July 1996, respectively; whereas the optical image, with $300 \times 412 \times 3$ pixels, is obtained from Google Earth with red, green, and blue bands in July 1996. The data set represents the expansion of the Lake Mulargia, as shown in Fig. 4(d).
- 2) The second data set is a pair of SAR/optical satellite images (Shuguang Village, Dongying, China), as shown in Fig. 5. The SAR image, with $593 \times 921 \times 1$ pixels, is taken by the Radarsat-2 with the C-band in June 2008; whereas the optical image, with $593 \times 921 \times 3$ pixels, is acquired from Google Earth with red, green, and blue bands in September 2012. The data set represents the changes of land use in the farmland, as shown in Fig. 5(c), which is generated by a manual annotation that combines expert knowledge and prior information.
- 3) The third data set is a pair of SAR/optical satellite images (Wuhan, China), as shown in Fig. 6. The SAR image, with $495 \times 503 \times 1$ pixels, is taken by the Radarsat-2 with the C-band in June 2008; whereas the

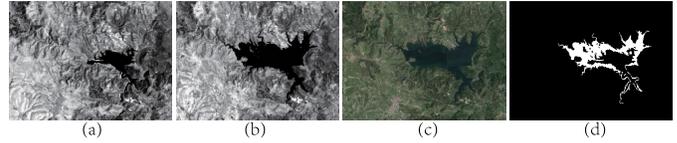


Fig. 4. Sardinia data set. (a) NIR band image acquired in September 1995. (b) NIR band image acquired in July 1996. (c) Optical image acquired in July 1996. (d) Ground-truth image representing the changes between (a) and (c).



Fig. 5. Shuguang data set. (a) SAR image acquired in June 2008. (b) Optical image acquired in September 2012. (c) Ground-truth image.

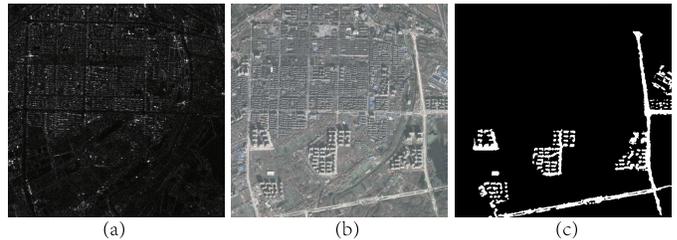


Fig. 6. Wuhan data set. (a) SAR image acquired in June 2008. (b) Optical image acquired in November 2011. (c) Ground-truth image.

optical image, with $495 \times 503 \times 3$ pixels, is acquired from Google Earth with red, green, and blue bands in November 2011. The data set represents the changes in new buildings and roads, as shown in Fig. 6(c), which is generated through the on-the-spot investigation.

In the first experiment, we test the proposed model on the Sardinia and Shuguang data sets to verify the effectiveness of the PSGM. We first show that for the images of the same object obtained from different sources, although their pixel values representing distinct characteristics are different, their image patch similarity structure is consistent. Using the Sardinia data set, we set the NIR image acquired in July 1996 to be \mathbf{X} and set the optical image acquired in the same time to be \mathbf{Y} (denote the red, green, and blue bands as $\mathbf{Y}^{(1)}$, $\mathbf{Y}^{(2)}$, and $\mathbf{Y}^{(3)}$, respectively). Fig. 7 shows the original \mathbf{X} , \mathbf{Y} and the reconstructed \mathbf{XZ}_X , \mathbf{YZ}_X , \mathbf{XZ}_Y , and \mathbf{YZ}_Y in detail. By comparing \mathbf{X} and \mathbf{Y} with the \mathbf{XZ}_X and \mathbf{YZ}_Y in Fig. 7, it can be shown that the learned PSGM fully contains the structure information of the original image, and it is a good proof of self-similarity as the error of reconstruction $\mathbf{X} - \mathbf{XZ}_X$ (and $\mathbf{Y} - \mathbf{YZ}_Y$) is very small. By comparing \mathbf{X} , \mathbf{Y} with \mathbf{XZ}_Y and \mathbf{YZ}_X in Fig. 7, we can find that the patch similarity structure between heterogeneous images is consistent. For further illustration of PSGM, Fig. 8 shows the histograms of the white regions in Fig. 7 (white rectangle boxes) as a comparison. From Fig. 8, we can find that the PSGM can establish a connection between heterogeneous

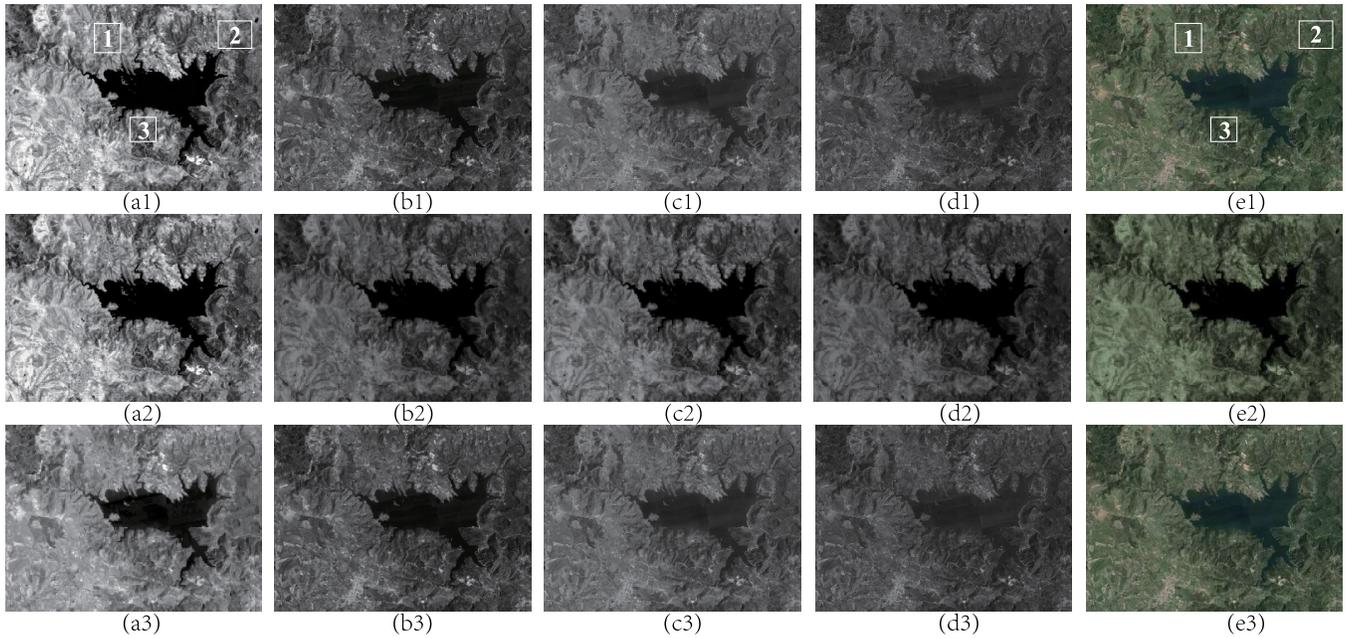


Fig. 7. PSGM reconstructed images on the Sardinia data set. (a1)–(e1) Original X , $Y^{(1)}$, $Y^{(2)}$, $Y^{(3)}$, and Y , respectively. (a2)–(e2) PSGM Z_X reconstructed XZ_X , $Y^{(1)}Z_X$, $Y^{(2)}Z_X$, $Y^{(3)}Z_X$, and YZ_X , respectively. (a3)–(e3) PSGM Z_Y reconstructed XZ_Y , $Y^{(1)}Z_Y$, $Y^{(2)}Z_Y$, $Y^{(3)}Z_Y$, and YZ_Y , respectively.

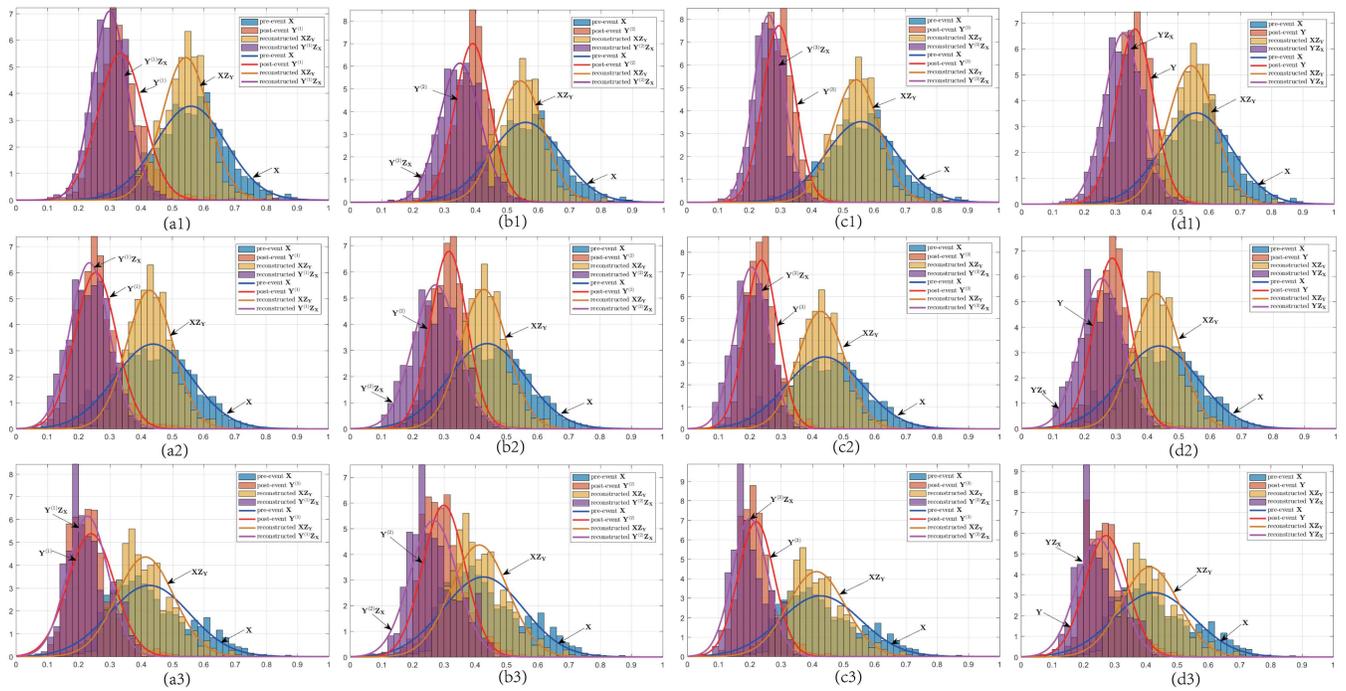


Fig. 8. Histograms comparison of the reconstructed image on the Sardinia data set. (From top to bottom row) Histograms of regions 1–3 indicated by the white rectangular boxes in Fig. 7. (a1)–(c3) Comparisons of the histograms in each band of Y . (d1)–(d3) Comparisons of the histograms with the gray scale of Y .

images, which also proves that it can be used for remote sensing image regression as mentioned in Section IV.

Then, we show that for the multitemporal images obtained from different sources, the PSGM reconstructed image can keep consistent for unchanging areas and has strong differentiation for changing areas. Similar to Figs. 7 and 8, Figs. 9 and 10 show the PSGM reconstructed images and

the detailed histograms for the Shuguang data set, where the preevent SAR image is set to be X and the postevent optical image is set to be Y (denote the red, green, and blue bands as $Y^{(1)}$, $Y^{(2)}$, and $Y^{(3)}$, respectively). From Fig. 9, it can be seen that the PSGM reconstructed images YZ_X and XZ_Y are similar with the original images Y and X in the unchanged areas, respectively; whereas this similarity no longer exists

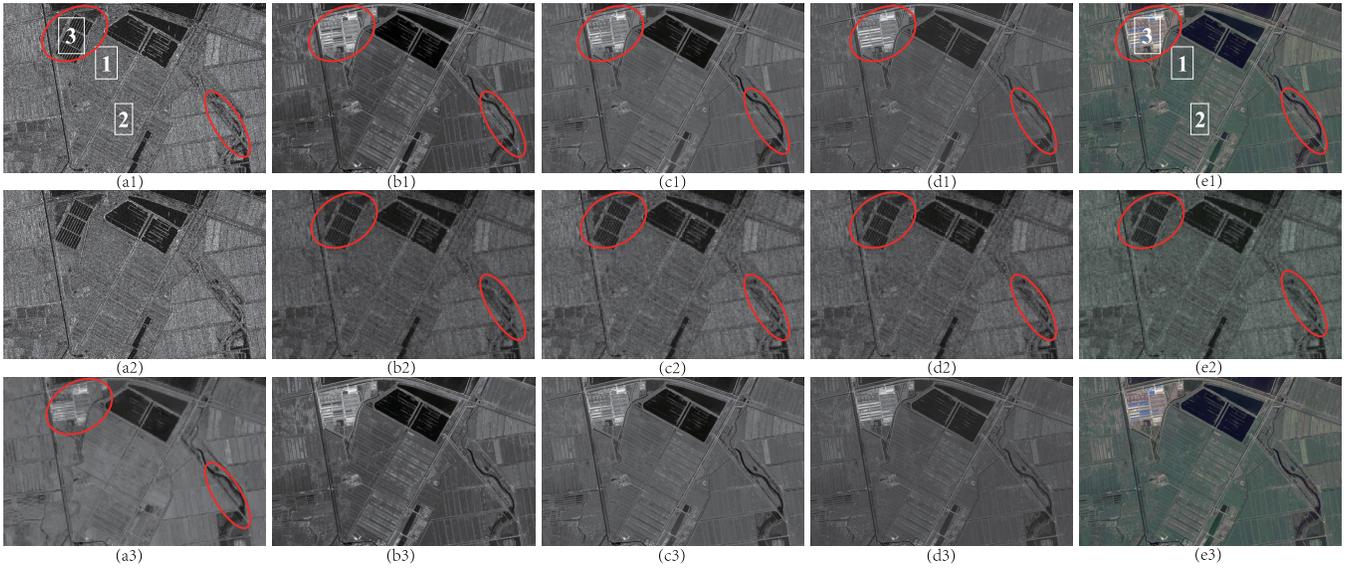


Fig. 9. PSGM reconstructed images on the Shuguang data set. (a1)–(e1) Original X , $Y^{(1)}$, $Y^{(2)}$, $Y^{(3)}$, and Y , respectively. (a2)–(e2) PSGM Z_X reconstructed XZ_X , $Y^{(1)}Z_X$, $Y^{(2)}Z_X$, $Y^{(3)}Z_X$, and YZ_X , respectively. (a3)–(e3) PSGM Z_Y reconstructed XZ_Y , $Y^{(1)}Z_Y$, $Y^{(2)}Z_Y$, $Y^{(3)}Z_Y$, and YZ_Y , respectively.

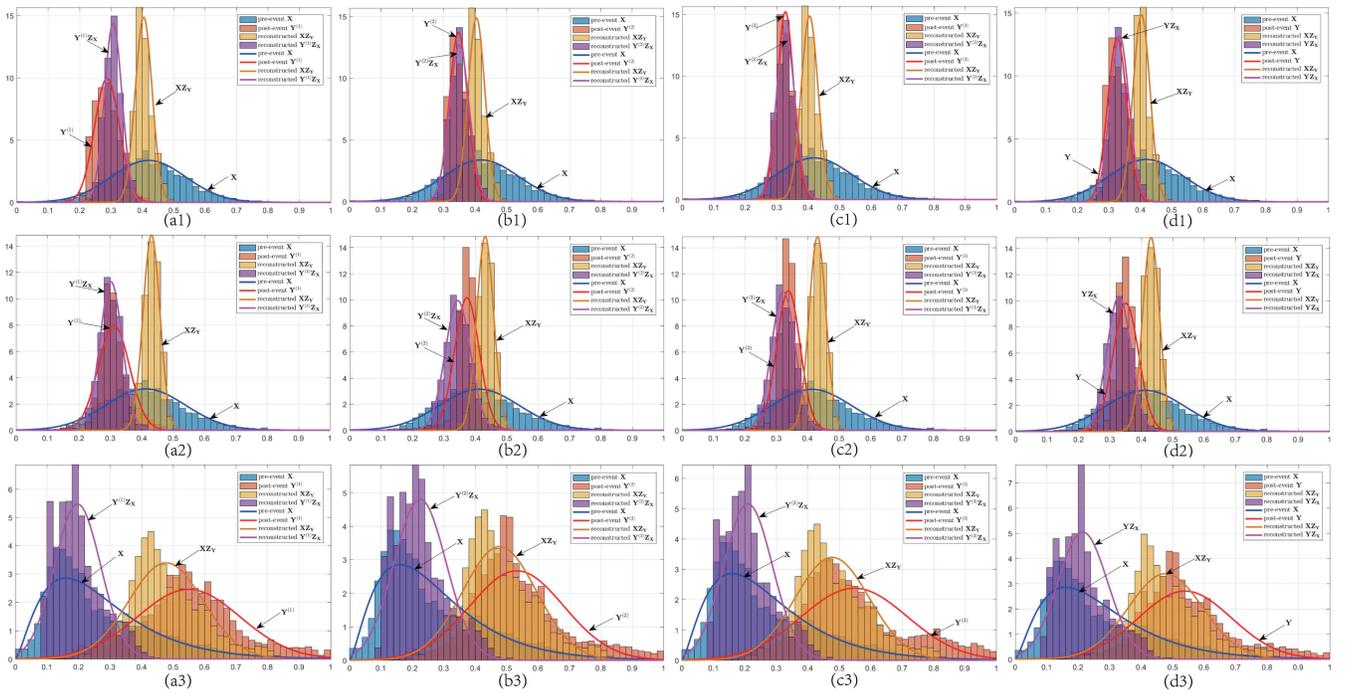


Fig. 10. Histograms comparison of the reconstructed image on the Shuguang data set. (From top to bottom row) Histograms of regions 1–3 indicated by white rectangular boxes in Fig. 9. (a1)–(c3) Comparisons of the histograms in each band of Y . (d1)–(d3) Comparisons of the histograms with the gray scale of Y .

in the changed areas (indicated by red ellipses). This can be further proved in Fig. 10, the histograms of YZ_X (or XZ_Y) and Y (or X) are very close in the selected unchanged regions (1 and 2), as shown in the first and second rows. However, the histograms of YZ_X (or XZ_Y) and Y (or X) are quite different in the selected changed regions (3), as shown in the third row of Fig. 10, which means that the reconstruction errors between YZ_X (or XZ_Y) and Y (or X) can be used to measure the change level between the heterogeneous images.

In the second experiment, we test the PSGM-based CD method on the abovementioned three heterogeneous data sets to obtain the DI and the binary CM. For the comparison methods, we choose the KL divergence-based method [58], the pixels' pair (PP) method [30], [59], and the affinity matrix distance (AMD) method [34] to compare the generation of DI. The PP method computes differences between pixels in each image separately, and then, the difference scores are compared between images in pair to generate the DI. The AMD method

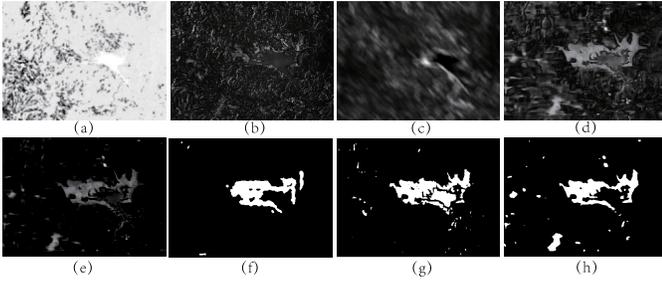


Fig. 11. DIs and binary CMs on the Sardinia data set. (a) DI of KL. (b) DI of PP. (c) DI of AMD. (d) DI of PSGM. (e) Sparse-induced DI of PSGM. (f) Binary CM of M3CD. (g) Binary CM of SCCN. (h) Binary CM of PSGM-PCA KM.

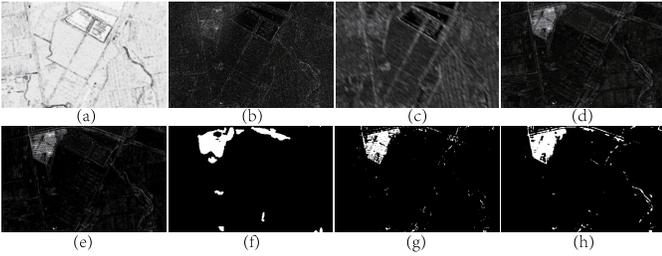


Fig. 12. DIs and binary CMs on the Shuguang data set. (a) DI of KL. (b) DI of PP. (c) DI of AMD. (d) DI of PSGM. (e) Sparse-induced DI of PSGM. (f) Binary CM of M3CD. (g) Binary CM of SCCN. (h) Binary CM of PSGM-PCA KM.

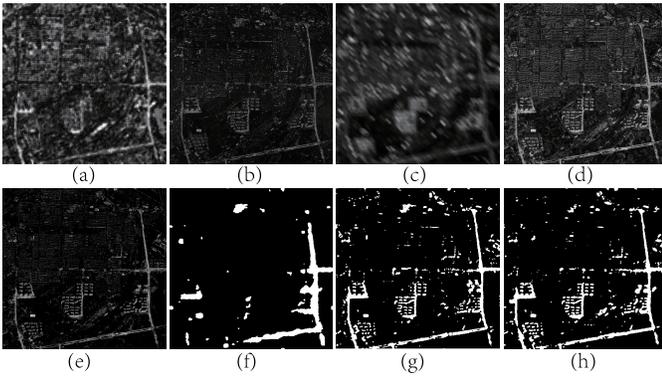


Fig. 13. DIs and binary CMs on the Wuhan data set. (a) DI of KL. (b) DI of PP. (c) DI of AMD. (d) DI of PSGM. (e) Sparse-induced DI of PSGM. (f) Binary CM of M3CD. (g) Binary CM of SCCN. (h) Binary CM of PSGM-PCA KM.

computes the patch-based affinity matrices \mathbf{A}^X and \mathbf{A}^Y for two heterogeneous images separately, where each element $\mathbf{A}_{i,j}^X = \exp\{-\frac{(\|p_i^X - p_j^X\|_2^2)}{(h^2)}\}$ represents the affinity between the i th and j th pixels in the patch. Then, the Frobenius norm $f = \|\mathbf{A}^X - \mathbf{A}^Y\|_F$ is computed and assigned to all the pixels in the patch, and the final change possibility of each pixel can be obtained by averaging all the patch distances f that covering this pixel. The performance of the DIs generated by comparing methods can be assessed by the empirical receiver operating characteristics (ROCs) curves, representing the estimated pixelwise probability of detection (PD) as a function of the probability of false alarm (PFA). In order to evaluate the binary CM generated by PSGM-PCA KM, we also choose the

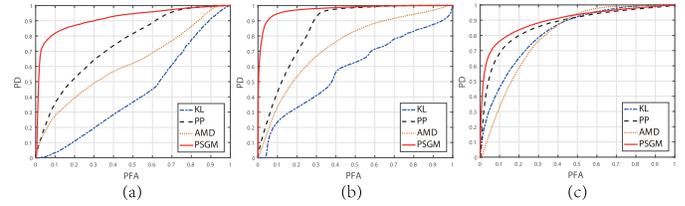


Fig. 14. ROC curves of DIs generated by different methods on the heterogeneous data sets. (a) Sardinia. (b) Shuguang. (c) Wuhan.

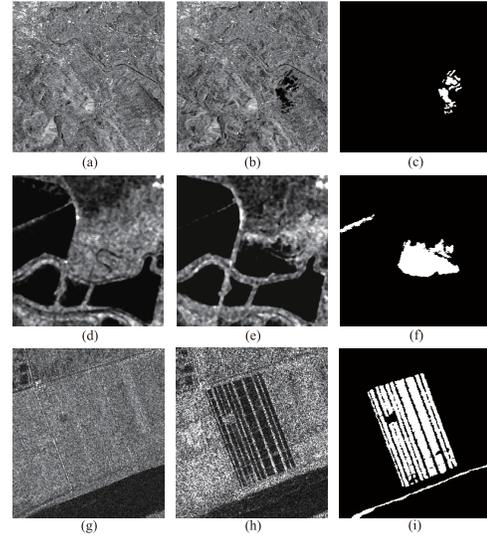


Fig. 15. Bern, San Francisco, and Yellow River data sets. (Top row) Bern data set. (a) Image acquired in April 1999. (b) Image acquired in May 1999. (c) Ground-truth image representing the changes between (a) and (b). (Middle row) San Francisco data set. (d) Image acquired in August 2003. (e) Image acquired in May 2004. (f) Ground-truth image representing the changes between (d) and (e). (Bottom row) Yellow River data set. (g) Image acquired in June 2008. (h) Image acquired in June 2009. (i) Ground-truth image representing the changes between (g) and (h).

M3CD method [26] and the deep learning-based SCCN [19] for comparison. We use some quantitative parameters to evaluate the final binary CM: false negatives (FNs), false positives (FPs), the overall error (OE), the percentage correct classification (PCC), and the Kappa coefficient (Kappa).

Figs. 11–13 show the DIs and binary CMs of all comparing methods on the Sardinia, Shuguang, and Wuhan data sets, respectively. Fig. 14 shows the ROC curves on different data sets of the comparing methods. By comparing the PSGM-based DI with other DIs in Figs. 11–13, we can find that the PSGM highlights the changing part more than other methods, which can also be shown in Fig. 14. Especially, in the Sardinia and Shuguang data sets, the proposed PSGM can achieve a much larger area under the curve than other methods, as shown in Fig. 14(a) and (b). In addition, by comparing (d) and (e) subfigures of Figs. 11–13, we can see that the sparsity-induced DI of PSGM can effectively reduce the impact of noise and other interference items and produce a difference map closer to the real change.

As shown in Figs. 11–13, the binary CM produced by PSGM is closer to the reference image, preserves more details than M3CD and SCCN, and, thus, leads to a much less FN and

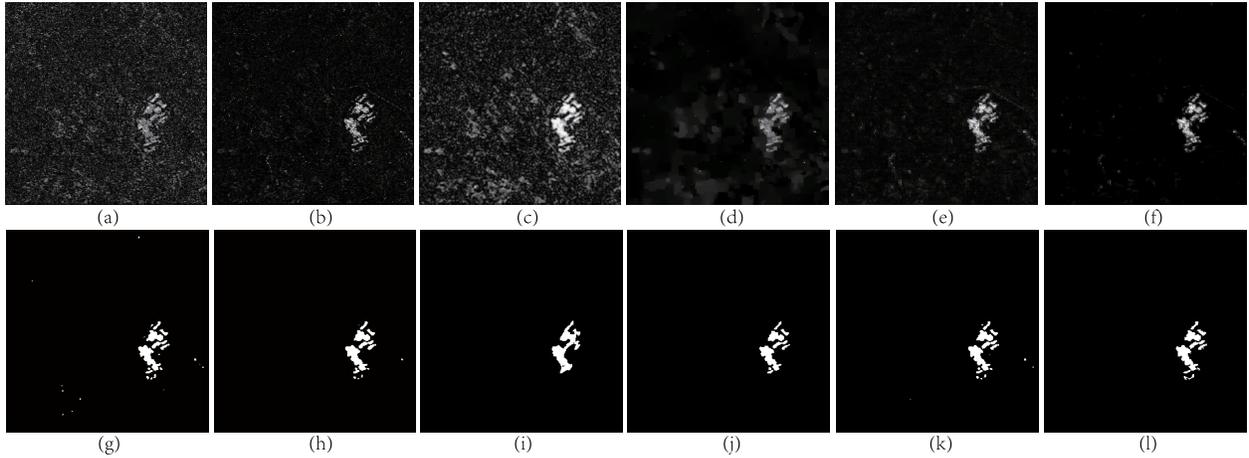


Fig. 16. DIs and binary CMs on the Bern data set. (a) DI of direct difference. (b) DI of log-ratio. (c) DI of mean-ratio. (d) DI of SDCD. (e) DI of PSGM. (f) Sparse-induced DI of PSGM. (g) Binary CM of PCAKM. (h) Binary CM of GaborTLC. (i) Binary CM of Ms-CapsNet. (j) Binary CM of CWNN. (k) Binary CM of PSGM-PCAKM. (l) Binary CM of PSGM-CWNN.

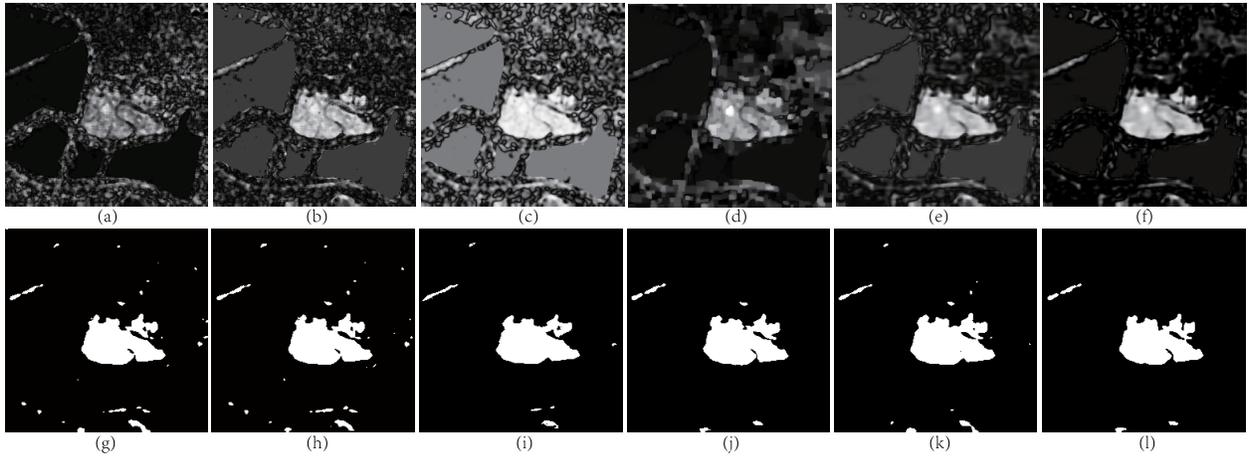


Fig. 17. DIs and binary CMs on the San Francisco data set. (a) DI of direct difference. (b) DI of log-ratio. (c) DI of mean-ratio. (d) DI of SDCD. (e) DI of PSGM. (f) Sparse-induced DI of PSGM. (g) Binary CM of PCAKM. (h) Binary CM of GaborTLC. (i) Binary CM of Ms-CapsNet. (j) Binary CM of CWNN. (k) Binary CM of PSGM-PCAKM. (l) Binary CM of PSGM-CWNN.

TABLE IV
QUANTITATIVE MEASURES OF BINARY CMs ON THE
HETEROGENEOUS DATA SETS

Datasets	Methods	FN	FP	OE	PCC(%)	Kappa
Sardinia	M3CD	2565	2012	4577	96.30	0.6690
	SCCN	1814	3199	5013	95.97	0.6771
	PSGM	1926	2895	4821	96.10	0.6821
Shuguang	M3CD	8041	12728	20769	96.20	0.6017
	SCCN	9726	5652	15378	97.18	0.6520
	PSGM	5262	7539	12801	97.66	0.7438
Wuhan	M3CD	7706	8784	16490	93.38	0.4758
	SCCN	3281	11314	14595	94.14	0.6108
	PSGM	2067	9627	11694	95.32	0.6858

larger Kappa value, as listed in Table IV. From Figs. 11–14 and Table IV, we can find that the PSGM model gains effective performance on the heterogeneous CD task.

B. Experiments on the Homogeneous Data Sets

1) *Experiments on Homogeneous SAR Data Sets:* Here, we test the proposed PSGM-based CD method on three real pairs of homogeneous SAR images.

The first data set of SAR images is from Bern, Switzerland, in April and May 1999, acquired by the ERS-2 SAR sensor with the size of 301×301 pixels. Fig. 15(a) and (b) shows the two multitemporal SAR images, respectively. Fig. 15(c) shows the ground truth of the CD map, which represents that River Aare inundated parts of Bern, Thun, and the whole airport in Bern.

The second data set is from San Francisco, in August 2003 and May 2004, acquired by the ERS-2 SAR sensor with the size of 256×256 pixels, as shown in Fig. 15(d) and (e), respectively. Fig. 15(f) shows the reference image generated by integrating prior information with photograph interpretation.

The third data set is from Yellow River Estuary, China, in June 2008 and June 2009, acquired by the Radarsat-2 SAR sensor with the size of 289×257 pixels, as shown in Fig. 15(g) and (h), respectively. Fig. 15(i) shows the reference image that is created by integrating prior information with photograph interpretation based on the input images in Fig. 15(g) and (h).

For the homogeneous SAR data sets, we select the common direct difference, log-ratio, mean-ratio [60], and

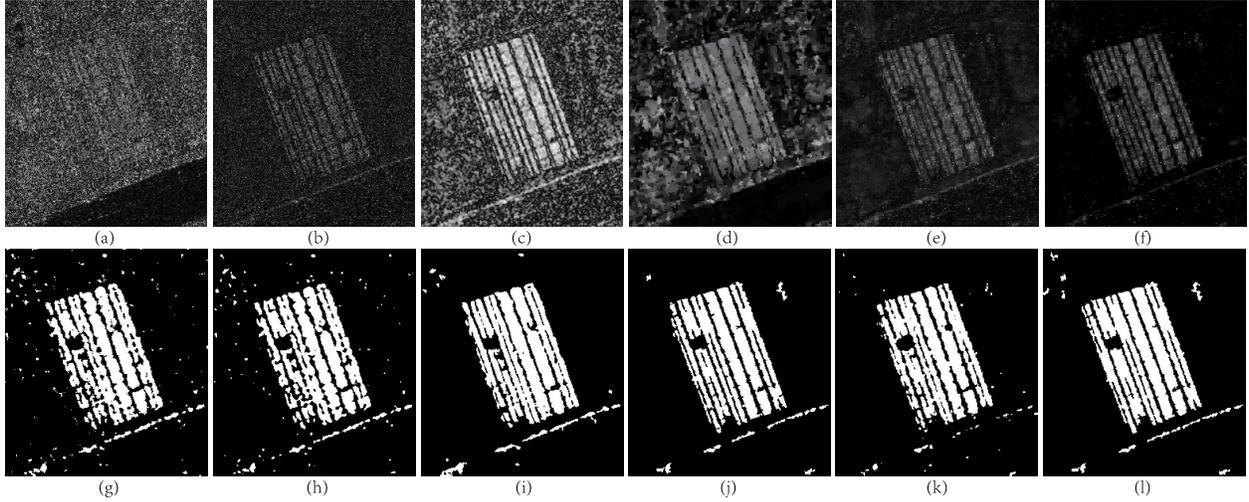


Fig. 18. DIs and binary CMs on the Yellow River data set. (a) DI of direct difference. (b) DI of log-ratio. (c) DI of mean-ratio. (d) DI of SDCD. (e) DI of PSGM. (f) Sparse-induced DI of PSGM. (g) Binary CM of PCAKM. (h) Binary CM of GaborTLC. (i) Binary CM of Ms-CapsNet. (j) Binary CM of CWNN. (k) Binary CM of PSGM-PCAKM. (l) Binary CM of PSGM-CWNN.

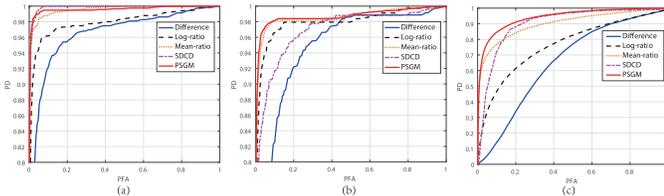


Fig. 19. ROC curves of DIs generated by different methods on the homogeneous SAR data sets. (a) Bern. (b) San Francisco. (c) Yellow River.

sparsity-driven CD (SDCD) [61] methods for comparing the DI and select PCAKM [38], GaborTLC [7], multiscale capsule network (Ms-CapsNet) [62], and convolutional-wavelet neural network (CWNN) [10] for comparing the binary CM. In addition, inspired by the pseudotraining sets used in the unsupervised hyperspectral image CD [11], [12], we also try to use the sparsity-induced DI of PSGM to construct a better pseudodata set for CWNN (PSGM-CWNN) to test the application of PSGM in deep learning networks.

Figs. 16–18 show the DIs, sparsity-induced DIs, and binary CMs of different methods on the homogeneous SAR data sets, and Table V lists the corresponding values of evaluation criteria. As shown in Figs. 16–18, the PSGM-based DI can alleviate the impact of noise, which is mainly due to that it reduces some noise during the PSGM learning process as the patch-based learning can relax the influence of speckle noise. The ROC curves on the San Francisco and Yellow River data sets in Fig. 19 also show that the PSGM-based $\tilde{D}\tilde{I}$, with the largest area under the curve, is better than others. By comparing (e) and (f) subfigures of Figs. 16–18, we can find that the sparsity-induced Δ can effectively reduce the adverse effects of noise and highlight the changed parts. From Table V, we can find that the PSGM-based PCAKM can provide better results when comparing with PCAKM and GaborTLC. Moreover, it increases the Kappa coefficient by 2.47% and 5.99% comparing with the log-ratio-based PCAKM on the San Francisco and Yellow River data sets, respectively.

TABLE V
QUANTITATIVE MEASURES OF BINARY CMs ON
THE HOMOGENEOUS SAR DATA SETS

Datasets	Methods	FN	FP	OE	PCC(%)	Kappa
Bern	PCAKM	145	158	303	99.67	0.8679
	GaborTLC	158	122	280	99.69	0.8753
	PSGM-PCAKM	154	131	285	99.69	0.8738
	Ms-CapsNet	276	235	511	94.44	0.7719
	CWNN	230	85	315	99.65	0.8528
	PSGM-CWNN	156	103	259	99.71	0.8838
San Francisco	PCAKM	378	701	1079	98.35	0.8798
	GaborTLC	324	763	1087	98.34	0.8802
	PSGM-PCAKM	256	603	859	98.69	0.9045
	Ms-CapsNet	419	359	778	98.81	0.9100
Yellow River	CWNN	295	437	732	99.88	0.9170
	PSGM-CWNN	312	319	631	99.04	0.9275
	PCAKM	2656	2149	4805	93.53	0.7784
	GaborTLC	2706	1659	4365	94.12	0.7954
Yellow River	PSGM-PCAKM	2360	1065	3425	95.39	0.8383
	Ms-CapsNet	1786	1631	3417	95.40	0.8440
	CWNN	1694	712	2406	96.76	0.8875
	PSGM-CWNN	1129	1133	2262	96.95	0.8972

As shown in Figs. 16–18, the PSGM-based PCAKM can efficiently reduce the isolated error points caused by the speckle noise and obtain cleaner binary CMs, which once again validates the advantages of the PSGM and sparsity regularization model. Meanwhile, by comparing the performance of CWNN and PSGM-CWNN, we can find that the sparsity-induced DI of PSGM can also be applied to unsupervised deep learning method to generate a better pseudotraining set and then improve the detection performance of the original method, which also extends the application scope of the proposed PSGM method.

2) *Experiments on Homogeneous Optical Data Set:* The optical data set is from Changsha, China, in March 2016 and March 2017, acquired by the GaoFen-2 Optical sensor (red, green, and blue bands) with the size of $400 \times 400 \times 3$ pixels, as shown in Fig. 20(a) and (b), respectively. Fig. 20(m) shows the reference image representing the changes caused by the construction of the subway, which is generated through the on-the-spot investigation.

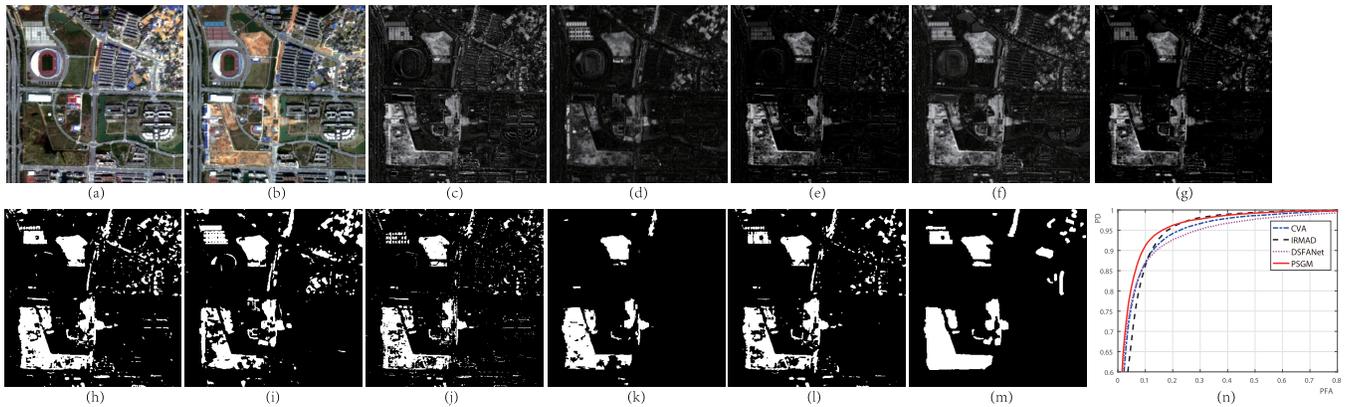


Fig. 20. Changsha data set. (a) Image acquired in March 2016. (b) Image acquired in March 2017. (c) DI of CVA. (d) DI of IRMAD. (e) DI of DSFANet. (f) DI of PSGM. (g) Sparse-induced DI of PSGM. (h) Binary CM of CVA. (i) Binary CM of IRMAD. (j) Binary CM of DSFANet. (k) Binary CM of GCNCD. (l) Binary CM of PSGM. (m) Ground-truth image. (n) ROC curve of different DIs.

TABLE VI
QUANTITATIVE MEASURES OF BINARY CMs ON
THE HOMOGENEOUS OPTICAL DATA SET

Methods	FN	FP	OE	PCC(%)	Kappa
CVA	3067	8124	11191	93.01	0.7257
IRMAD	5694	7143	12837	91.98	0.6636
DSFANet	6826	3897	10723	93.30	0.6932
GCNCD	7751	1009	8760	94.53	0.7276
PSGM	4958	4845	9803	93.87	0.7351

For the homogeneous optical data set, we choose the change vector analysis (CVA) method [63], iteratively reweighted multivariate alteration detection (IRMAD) method [64], and deep slow feature analysis network (DSFANet) [65] as the DI comparison methods and choose PCAKM to generate the corresponding binary CMs of DIs generated by CVA, IRMAD, and PSGM for comparing. Meanwhile, we also select the semisupervised CD with a graph convolutional network (GCNCD) [66] as a comparison.

The experiment results are shown in Fig. 20 and listed in Table VI, respectively. As can be seen from the results of these comparison algorithms, the improvement effect of PSGM on the homogeneous optical image data set is not as great as that on the SAR image data sets. This is mainly because the noise of the optical image is not as serious as that of the SAR image. However, the PSGM model can still achieve quite competitive results in terms of OE, PCC, and Kappa coefficient, which demonstrates the effectiveness of the proposed method again. Meanwhile, from the DI and binary CM of PSGM, we can see that there are still some isolated error points in them. Inspired by the saliency detection that has shown effective performance in remote sensing [67]–[69] and other fields [70]–[73], we can remove the background pixels by extracting the salient region that probably belongs to the changed object from the DI and then reduce the isolated errors.

3) *Test of Different Noise Levels:* To evaluate the performance of PSGM under different noise conditions, we use a simulated homogeneous SAR data set, as shown in Fig. 21, where Fig. 21(a) and (b) shows the two simulated multitemporal images and Fig. 21(c) shows the ground-truth image.

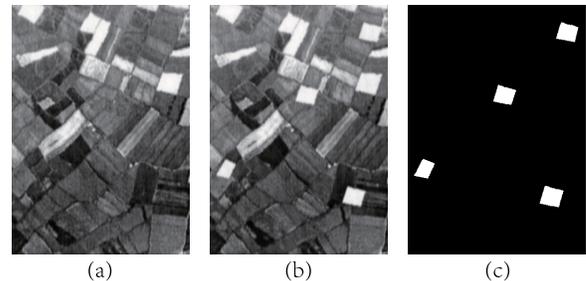


Fig. 21. Simulated data set. (a) and (b) Two simulated multitemporal images, respectively. (c) Ground-truth image.

By adding different levels of multiplicative Gamma distributed speckle noise to the multitemporal images (with $L = 10, 5, 2$), we can obtain different DIs and ROC curves of comparison methods under different noise levels, as shown in Figs. 22 and 23, respectively. From these results, we can find that the higher the noise level in the data set, the more obvious the improvement brought by the PSGM. This is due to that PSGM makes use of the similarity between image patches and the sparse prior knowledge of changes, then it can alleviate the impact of noise.

C. Experiments on Both Homogeneous and Heterogeneous Data Set

In order to further demonstrate the performance of the PSGM in the multimodal data, we will test the PSGM-based method on the complex Texas data set that contains both the homogeneous and heterogeneous data.

The Texas data set consists of three multispectral images²: a pair of preevent and postevent images with $1534 \times 808 \times 7$ pixels from the same sensor, the Landsat 5 TM (denoted as L5t1 and L5t2, respectively), in August 2011 and September 2011, as shown in Fig. 24(a) and (b), respectively; and another postevent image with $1534 \times 808 \times 10$ pixels acquired by the Advanced Land Imager (ALI2) from the Earth Observing mission in September 2011, as shown

²Distributed by LP DAAC, <http://lpdaac.usgs.gov>

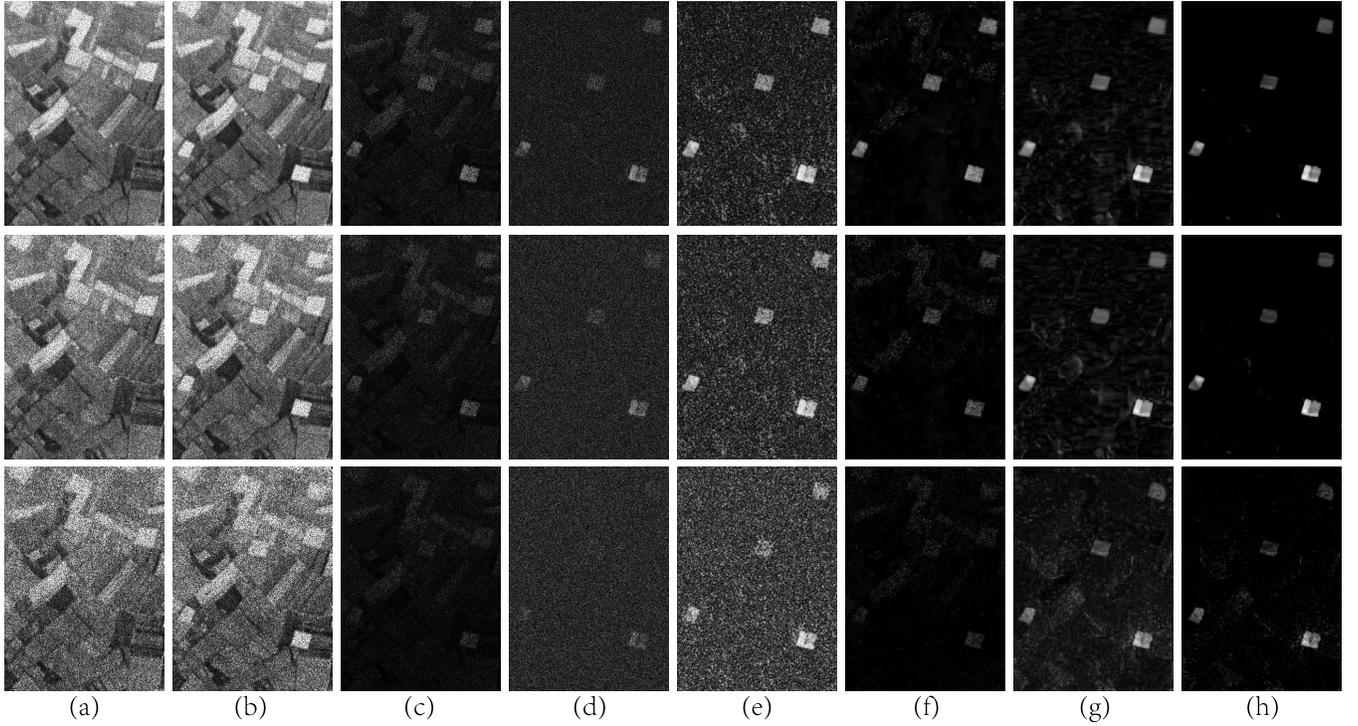


Fig. 22. DI generated by different methods on the simulated data set contaminated by different levels of speckle noise with $L = 10, 5, 2$. (From top to bottom) They correspond to $L = 10, L = 5$, and $L = 2$, respectively. (a) Contaminated preevent image. (b) Contaminated postevent image. (c) DI of direct difference. (d) DI of log-ratio. (e) DI of mean-ratio. (f) DI of SDCD. (g) DI of PSGM. (h) Sparse-induced DI of PSGM.

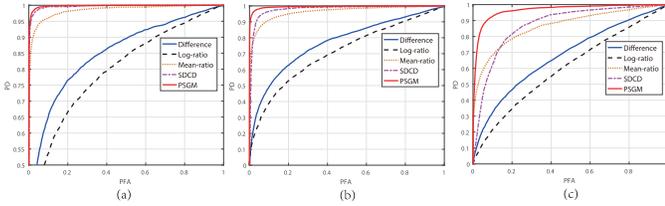


Fig. 23. ROC curves of DIs generated by different methods on the simulated data set contaminated by different levels of speckle noise. (a) $L = 10$. (b) $L = 5$. (c) $L = 2$.

in Fig. 24(c). As the L5t2 and ALIt2 are acquired within a one-day interval, there is no apparent change between them. The change between L5t1 and L5t2 (ALIt2) is provided by Volpi *et al.* [74], which represents a forest fire in Bastrop, TX, USA, as shown in Fig. 24(d).

Define L5t1, L5t2, and ALIt2 as \mathbf{X}_{L5} , \mathbf{Y}_{L5} , and \mathbf{Y}_{ALI} , respectively. Fig. 25 shows each band of the homogeneous (L5t1, L5t2) images: the first row shows $\mathbf{X}_{L5}^{(i)}$, $i = 1, \dots, 7$, and the second row shows $\mathbf{Y}_{L5}^{(i)}$, $i = 1, \dots, 7$. By comparing $\mathbf{X}_{L5}^{(i)}$ and $\mathbf{Y}_{L5}^{(i)}$, we can find that the differences on bands 4–7 are much greater than that on other bands, which is consistent with the sparsity-induced DIs $\Delta_{L5}^{(i)}$ of PSGM, as shown in the third row of Fig. 25. For the heterogeneous images of L5t1 and ALIt2, it is not correct to directly compare $\mathbf{X}_{L5}^{(i)}$ (the first row in Fig. 25) and $\mathbf{Y}_{ALI}^{(j)}$ (the first row in Fig. 26). However, by using the PSGM model, we can obtain the sparsity-induced DIs $\Delta_{ALI}^{(j)}$, as shown in the second row in Fig. 26. At the same time, we also show the final fused DI of homogeneous

(L5t1 and L5t2) and heterogeneous (L5t1 and ALIt2) images in Fig. 24(e) and (f), respectively. From these two fused DIs and sparse DI of each band, we can find that the PSGM-based method can not only detect the changes between heterogeneous images but also detect which bands have changed, which plays an important role in the subsequent tasks of spectral analysis.

D. Parameter Analysis

The next experiment is a test of parameters used in the PSGM-based CD, which consists of two main categories: the parameters used in the PSGM learning process and DI optimization with sparse regularization. Both processes have a balancing parameter γ in the optimization models (3) and (31). This parameter controls the strength of data fidelity and sparse regularization. The selection of γ has an important influence on the performance of these two models. On the one hand, γ should be reasonably small in order to weaken the influence of fitting the corrupted data. On the other hand, if γ is too small, the resulting PSGM \mathbf{Z} (or sparsity-induced DI Δ) may be over regularized, which may bring more errors. In general, when the input image is contaminated by high-level noise, we need to choose a small γ to suppress data fidelity while encouraging the sparsity regularization. Conversely, when the noise level is low, we should choose a large γ to force the data fidelity. In this article, we use an adaptive method to select this parameter during the iterations on the basis of the behavior of the objective function [75] with the initialization $\gamma = 0.01$, which has been proved to be effective.

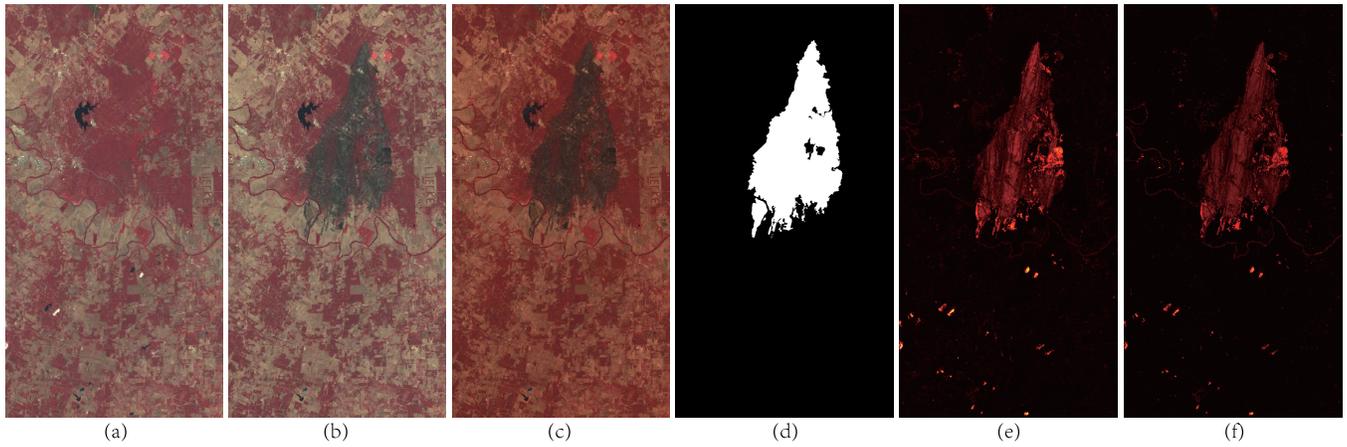


Fig. 24. Texas data set and fused DIs. (a) Image acquired by Landsat 5 TM in August 2011. (b) Image acquired by Landsat 5 TM in September 2011. (c) Image acquired by Advanced Land Imager from the Earth Observing mission in September 2011. (d) Ground-truth image representing the changes between (a) and (b). (e) Fused DI of PSGM on homogeneous images (L5t1 and L5t2). (f) Fused DI of PSGM on heterogeneous images (L5t1 and ALI2).

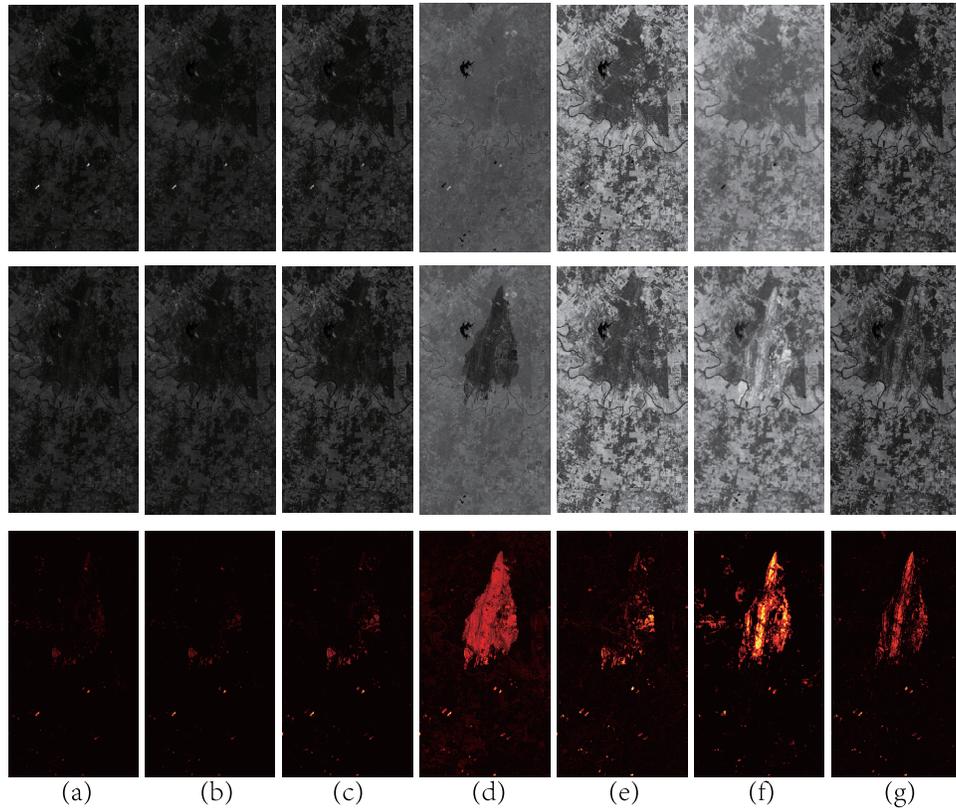


Fig. 25. Each band of L5t1 X_{L5} , L5t2 Y_{L5} and sparsity-induced DIs of $\Delta_{Y_{L5}}$. (From top to bottom) They correspond to X_{L5} , Y_{L5} and $\Delta_{Y_{L5}}$, respectively. (a)–(g) Band 1 to band 7.

Another important parameter in the PSGM learning is the patch size p . To measure the impact of different patch sizes, we test them on the abovementioned data sets and select the Kappa coefficient to evaluate the change detection results, which can describe the global performance of the CD. In Fig. 27, we vary the patch size from 3 to 9 with step 1 for all data sets except for the Shuguang data set. This is because when we set $p = 3$ in the Shuguang data set, the size of corresponding PSGM Z will reach 60786×60786 ,

which will exceed the memory limit. It can be found that the patch size has an important impact on the CD performance, especially for the data sets containing SAR images, as shown in Fig. 27. As the p increases, the Kappa value generally increases first and then decreases. This mainly due to that a very large p makes it very hard to find enough quite similarity patches, which leads to reconstruction errors. However, a very small p is not robust to the speckle noise in the SAR image, and a smaller patch size will lead to larger PSGM Z and

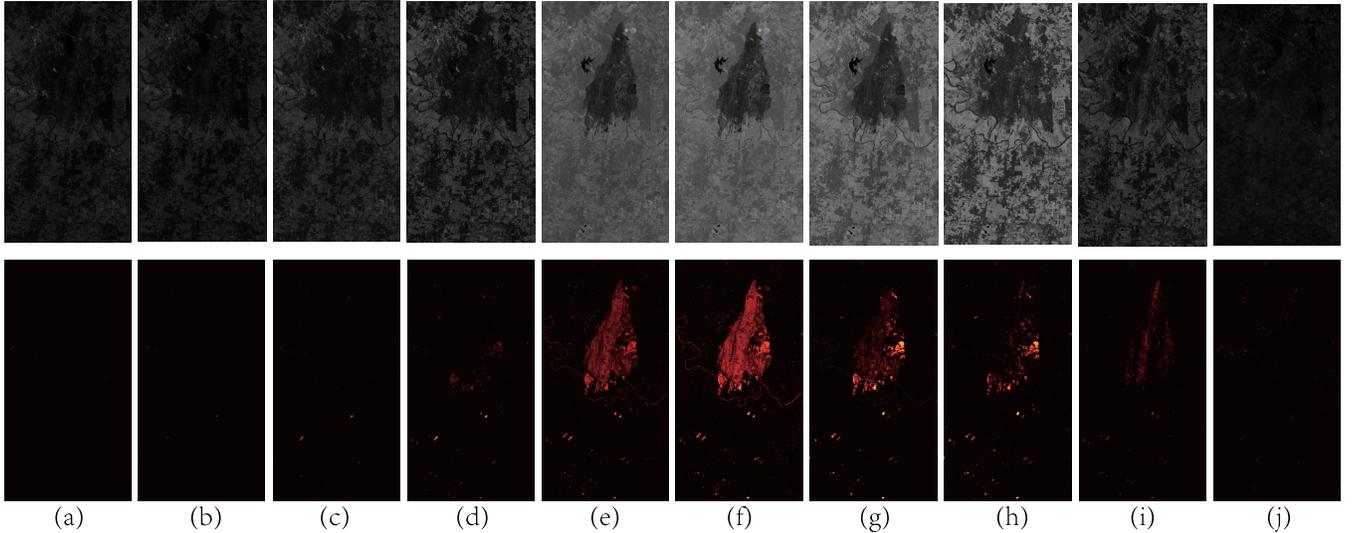


Fig. 26. Each band of ALIt2 \mathbf{Y}_{ALI} and sparsity-induced DIs of $\Delta_{\mathbf{Y}_{\text{ALI}}}$. (From top to bottom) They correspond to \mathbf{Y}_{ALI} and $\Delta_{\mathbf{Y}_{\text{ALI}}}$, respectively. (a)–(j) Band 1 to band 10.

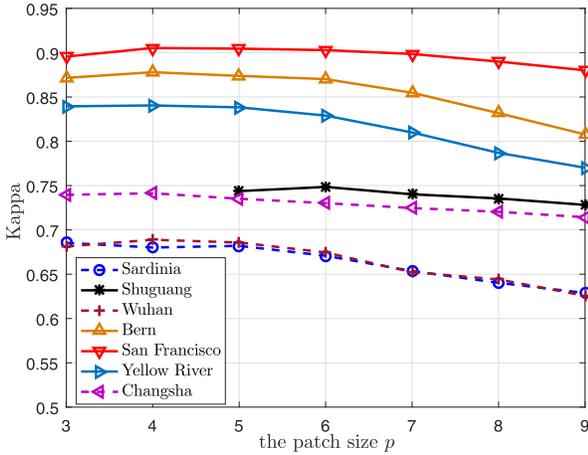


Fig. 27. Change detection performance comparison with different patch sizes p .

require more computation. It is obvious that setting $p = 5$ is a good compromise choice in our experiment. For the selection of this parameter, we suggest that in the case of a high-resolution image, we can choose a larger p ; on the contrary, in the case of a low-resolution image, we can choose a smaller p , and the computational complexity should also be taken into account.

Meanwhile, the performance of the final CD is also influenced by the clustering process. The main parameters of PCAKM are the feature vector size s and the image block size h . In [38], these two parameters are analyzed in detail. Generally, setting $s \in \{3, 5, 7\}$ can meet most of the requirements. In this article, we set s equal to h for convenience in PCA feature extraction. We try different sizes of h and then choose the best result for each method.

E. Computational Time

In Section IV-A, we have analyzed the complexity of the proposed PSGM-based CD method. We report the CPU times of PSGM on different data sets under different conditions, as listed in Table VII, where the “original” refers to PSGM

TABLE VII
COMPUTATIONAL TIME (SECONDS) OF PSGM ON DIFFERENT DATA SETS UNDER DIFFERENT CONDITIONS

Datasets	Image size	p	N_p	Computational time	
				original	accelerated
San Francisco	$256 \times 256 \times 1$	5	2704	16.62	10.75
		7	1369	6.87	5.01
Changsha	$400 \times 400 \times 3$	5	6400	129.93	74.30
		7	3364	34.52	19.49
Shuguang	$593 \times 921 \times 1(3)$	5	22015	2599.60	1357.17
		7	11220	470.46	272.87

using ADMM to calculate the matrix \mathbf{Z} and DI Δ , and the “accelerated” refers to using the accelerated linearized ADMM [54] in these two processes of PSGM. These experiments are performed in MATLAB 2016a running on a desktop with Intel Core i7-8700K CPU, 32 GB of RAM, and 64-bit Windows 7 operating system.

From Table VII, we can see that when the value of N_p is large, the algorithm is time-consuming, which is consistent with our theoretical analysis. Meanwhile, it can also be found that several strategies can be considered to avoid a large amount of computational time: 1) using the accelerated methods in the process of solving the minimization problems, such as the accelerated linearized ADMM [54]; 2) using a large patch size p to reduce the N_p , seeking the balance between computation time and change detection accuracy; 3) dividing the large-scale image into small-scale image, for example, dividing a 2000×2000 image into $17\ 500 \times 500$ images, and then detecting the changes in each small image; and 4) using the superpixel as the basic unit instead of the square patch, which is also our future work.

VI. CONCLUSION

In this article, we present a novel unsupervised remote sensing image CD method, which can also be applied to the challenging heterogeneous CD task. This method is based on the self-expressive property that the similarity between the patches in each image representing the same ground object,

although acquired by different sensors, should be consistent. Therefore, the change level can be measured by how much one image still conforms to the similarity graph structure learned from the other image. Then, the PSGM-based CD method can be implemented in two steps. It first learns the PSGM for each input image, which can be transformed into a convex optimization problem with a sparsity regularization of the similarity matrix and solved by using the ADMM framework. Second, with the learned PSGM, the sparse prior knowledge of the changed part is further used to optimize the preliminary DI. Then, with the sparsity-induced DI, the binary CM can be obtained by clustering method. Experiments with both homogeneous and heterogeneous data sets demonstrate the effective performance of the proposed PSGM model. Meanwhile, we also show that the PSGM model can be extended to other remote sensing image processing problems, which can enlarge the application scope of the proposed algorithm. However, there are two drawbacks of proposed method that it is time-consuming when the size of PSGM is very large and the fusion of the forward and backward DI is very simple as it does not consider the statistical characteristics of images from different domains. Our future work is to develop the fast algorithm by using the superpixel and integrate the statistical information into the process of DI fusion.

ACKNOWLEDGMENT

The authors would like to thank Prof. HengChao Li of Southwest Jiaotong University for providing the codes of GaborTLC [7]. They would also like to thank the editors and anonymous reviewers for their careful reading of an earlier version of this article and constructive suggestions that improved the presentation of this work.

REFERENCES

- [1] Y. Ban and O. A. Yousif, "Multitemporal spaceborne SAR data for urban change detection in China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1087–1094, Aug. 2012.
- [2] J. L. Gil-Yepes, L. A. Ruiz, J. A. Recio, A. Balaguer-Beser, and T. Hermosilla, "Description and validation of a new set of object-based temporal geostatistical features for land-use/land-cover change detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 121, pp. 77–91, Nov. 2016.
- [3] F. Bovolo, C. Marin, and L. Bruzzone, "A hierarchical approach to change detection in very high resolution SAR images for surveillance applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2042–2054, Apr. 2013.
- [4] L. Bruzzone and D. F. Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 452–466, Apr. 2002.
- [5] F. Bovolo and L. Bruzzone, "A detail-preserving scale-driven approach to change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 12, pp. 2963–2972, Dec. 2005.
- [6] Z. Li, W. Shi, H. Zhang, and M. Hao, "Change detection based on Gabor wavelet features for very high resolution remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 783–787, May 2017.
- [7] H.-C. Li, T. Celik, N. Longbotham, and W. J. Emery, "Gabor feature based unsupervised change detection of multitemporal SAR images based on two-level clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2458–2462, Dec. 2015.
- [8] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 125–138, Jan. 2016.
- [9] F. Gao, J. Dong, B. Li, and Q. Xu, "Automatic change detection in synthetic aperture radar images based on PCANet," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1792–1796, Dec. 2016.
- [10] F. Gao, X. Wang, Y. Gao, J. Dong, and S. Wang, "Sea ice change detection in SAR images based on convolutional-wavelet neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1240–1244, Aug. 2019.
- [11] X. Li, Z. Yuan, and Q. Wang, "Unsupervised deep noise modeling for hyperspectral image change detection," *Remote Sens.*, vol. 11, no. 3, p. 258, Jan. 2019.
- [12] 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.
- [13] Z. Yuan, Q. Wang, and X. Li, "ROBUST PCANet for hyperspectral image change detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 4931–4934.
- [14] K. Mubea and G. Menz, "Monitoring land-use change in Nakuru (Kenya) using multi-sensor satellite data," *Adv. Remote Sens.*, vol. 1, no. 3, pp. 74–84, 2012.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] M. Volpi, D. Tuia, F. Bovolo, M. Kanevski, and L. Bruzzone, "Supervised change detection in VHR images using contextual information and support vector machines," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 20, pp. 77–85, Feb. 2013.
- [19] 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.
- [20] W. Zhao, Z. Wang, M. Gong, and J. Liu, "Discriminative feature learning for unsupervised change detection in heterogeneous images based on a coupled neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7066–7080, Dec. 2017.
- [21] T. Zhan, M. Gong, J. Liu, and P. Zhang, "Iterative feature mapping network for detecting multiple changes in multi-source remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 146, pp. 38–51, Dec. 2018.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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, Oct. 2017.
- [28] 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.
- [29] R. Touati, M. Mignotte, and M. Dahmane, "Change detection in heterogeneous remote sensing images based on an imaging modality-invariant MDS representation," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2018, pp. 3998–4002.
- [30] B. Ayhan and C. Kwan, "A new approach to change detection using heterogeneous images," in *Proc. IEEE 10th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, Oct. 2019, pp. 0192–0197.

- [31] M. Gong, P. Zhang, L. Su, and J. Liu, "Coupled dictionary learning for change detection from multisource data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7077–7091, Dec. 2016.
- [32] V. Ferraris, N. Dobigeon, Y. Cavalcanti, T. Oberlin, and M. Chabert, "Coupled dictionary learning for unsupervised change detection between multimodal remote sensing images," *Comput. Vis. Image Understand.*, vol. 189, Dec. 2019, Art. no. 102817.
- [33] 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.
- [34] 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.
- [35] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2790–2797.
- [36] C.-A. Deledalle, L. Denis, and F. Tupin, "Iterative weighted maximum likelihood denoising with probabilistic patch-based weights," *IEEE Trans. Image Process.*, vol. 18, no. 12, pp. 2661–2672, Dec. 2009.
- [37] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
- [38] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and k -means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, Oct. 2009.
- [39] S. T. Roweis, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
- [40] Z. Kang, C. Peng, and Q. Cheng, "Robust subspace clustering via smoothed rank approximation," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 2088–2092, Nov. 2015.
- [41] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 471–478.
- [42] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
- [43] L. Zhuang, Z. Zhou, S. Gao, J. Yin, Z. Lin, and Y. Ma, "Label information guided graph construction for semi-supervised learning," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4182–4192, Sep. 2017.
- [44] J. W. Goodman, "Some fundamental properties of speckle," *J. Opt. Soc. Amer.*, vol. 66, no. 11, p. 1145, Nov. 1976.
- [45] J.-S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, no. 2, pp. 165–168, Mar. 1980.
- [46] V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 2, pp. 157–166, Mar. 1982.
- [47] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-7, no. 2, pp. 165–177, Mar. 1985.
- [48] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009.
- [49] 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, Jan. 2009.
- [50] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.
- [51] B. Hou, Q. Wei, Y. Zheng, and S. Wang, "Unsupervised change detection in SAR image based on Gauss-log ratio image fusion and compressed projection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3297–3317, Aug. 2014.
- [52] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k -means clustering algorithm," *Appl. Statist.*, vol. 28, no. 1, pp. 100–108, 1979.
- [53] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. Boston, MA, USA: Springer, 1981.
- [54] Y. Ouyang, Y. Chen, G. Lan, and E. Pasiliao, "An accelerated linearized alternating direction method of multipliers," *SIAM J. Imag. Sci.*, vol. 8, no. 1, pp. 644–681, Jan. 2015.
- [55] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Jan. 2010.
- [56] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. NIPS*, 2002, pp. 849–856.
- [57] F. de Morsier, M. Borgeaud, V. Gass, J.-P. Thiran, and D. Tuia, "Kernel low-rank and sparse graph for unsupervised and semi-supervised classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3410–3420, Jun. 2016.
- [58] G. Mercier and J. Inglada, "Change detection with misregistration errors and heterogeneous data through the Orfeo toolbox," Institut Telecom, Telecom Bretagne, Rennes, France, Tech. Rep. RR-2008003-ITI, 2008.
- [59] 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.
- [60] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1432–1445, May 2007.
- [61] 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.
- [62] Y. Gao, F. Gao, J. Dong, and H.-C. Li, "SAR image change detection based on multiscale capsule network," *IEEE Geosci. Remote Sens. Lett.*, early access, Mar. 16, 2020, doi: 10.1109/LGRS.2020.2977838.
- [63] W. A. Malila, "Change vector analysis: An approach for detecting forest changes with Landsat," in *Proc. LARS Symp.*, Jan. 1980, p. 385.
- [64] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 463–478, Feb. 2007.
- [65] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9976–9992, Dec. 2019.
- [66] S. Saha, L. Mou, X. X. Zhu, F. Bovolo, and L. Bruzzone, "Semi-supervised change detection using graph convolutional network," *IEEE Geosci. Remote Sens. Lett.*, early access, Apr. 17, 2020, doi: 10.1109/LGRS.2020.2985340.
- [67] J. Geng, X. Ma, X. Zhou, and H. Wang, "Saliency-guided deep neural networks for SAR image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7365–7377, Oct. 2019.
- [68] M. Li, M. Li, P. Zhang, Y. Wu, W. Song, and L. An, "SAR image change detection using PCANet guided by saliency detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 402–406, Mar. 2019.
- [69] C. Li, R. Cong, J. Hou, S. Zhang, Y. Qian, and S. Kwong, "Nested network with two-stream pyramid for salient object detection in optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9156–9166, Nov. 2019.
- [70] R. Cong, J. Lei, H. Fu, M.-M. Cheng, W. Lin, and Q. Huang, "Review of visual saliency detection with comprehensive information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 10, pp. 2941–2959, Oct. 2019.
- [71] R. Cong *et al.*, "An iterative co-saliency framework for RGBD images," *IEEE Trans. Cybern.*, vol. 49, no. 1, pp. 233–246, Jan. 2019.
- [72] R. Cong, J. Lei, H. Fu, Q. Huang, X. Cao, and N. Ling, "HSCS: Hierarchical sparsity based co-saliency detection for RGBD images," *IEEE Trans. Multimedia*, vol. 21, no. 7, pp. 1660–1671, Jul. 2019.
- [73] R. Cong, J. Lei, H. Fu, F. Porikli, Q. Huang, and C. Hou, "Video saliency detection via sparsity-based reconstruction and propagation," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 4819–4831, Oct. 2019.
- [74] M. Volpi, G. Camps-Valls, and D. Tuia, "Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis," *ISPRS J. Photogramm. Remote Sens.*, vol. 107, pp. 50–63, Sep. 2015.
- [75] D. Lazzaro, E. L. Piccolomini, and F. Zama, "A nonconvex penalization algorithm with automatic choice of the regularization parameter in sparse imaging," *Inverse Problems*, vol. 35, no. 8, Aug. 2019, Art. no. 084002.



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



Xiang Tan received the B.S. degree from Xiangtan University, Xiangtan, China, in 2009, and the M.S. degree from the National University of Defense Technology, Changsha, China, in 2011, where he is pursuing the Ph.D. degree in information and communication engineering with the College of Electronic Science.

His research interests cover pattern recognition, multiuser cognitive radio networks, and multiagent reinforcement learning.



Xiao Li received the B.S. degree in electrical engineering and automation from the University of Jinan, Jinan, China, in 2015, and the M.S. degree in control science and engineering from Xiangtan University, Xiangtan, China, in 2018. He is pursuing the Ph.D. degree in information and communication engineering with the National University of Defense Technology, Changsha, China.

His research interests include image processing and pattern recognition, representation and dictionary learning, and computational pathology applications.



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

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