



## Global structure graph mapping for multimodal change detection

Te Han, Yuqi Tang, Yuzeng Chen, Bin Zou & Huihui Feng

To cite this article: Te Han, Yuqi Tang, Yuzeng Chen, Bin Zou & Huihui Feng (2024) Global structure graph mapping for multimodal change detection, International Journal of Digital Earth, 17:1, 2347457, DOI: [10.1080/17538947.2024.2347457](https://doi.org/10.1080/17538947.2024.2347457)

To link to this article: <https://doi.org/10.1080/17538947.2024.2347457>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 09 May 2024.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



# Global structure graph mapping for multimodal change detection

Te Han<sup>a</sup>, Yuqi Tang<sup>a,b</sup>, Yuzeng Chen<sup>c</sup>, Bin Zou<sup>a,b</sup> and Huihui Feng<sup>a,b</sup>

<sup>a</sup>School of Geosciences and Info-Physics, Central South University, Changsha, People's Republic of China; <sup>b</sup>Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring (Central South University), Ministry of Education, Changsha, People's Republic of China; <sup>c</sup>School of Geodesy and Geomatics, Wuhan University, Wuhan, Hubei, People's Republic of China

## ABSTRACT

Multimodal change detection (MCD) combines multiple remote sensing data sources to realize surface change monitoring, which is essential for disaster evaluation and environmental monitoring. However, due to the 'incomparable' features in multimodal data, traditional change detection methods for unimodal (homogenous) data no longer apply. To address this issue, this paper proposes a novel MCD method with global structure graph mapping (GSGM) which extracts the 'comparable' structural features between multimodal datasets and constructs a global structure graph (GSG) to express the structure information for each of the multi-temporal images, which are then cross-mapped to the other data domain. The change intensity (CI) is determined by measuring the change of GSGs after mapping and the differences between GSGs and mapped GSGs. The forward and backward CI maps (CIMs) are then fused with the latent low-rank representation method (LLRR), and the change map (CM) is obtained by threshold segmentation. Experiments on five multimodal and four unimodal datasets demonstrate the effectiveness and robustness of the proposed method (source code is made available at <https://github.com/rshante0426/GSGM>).

## ARTICLE HISTORY

Received 27 December 2023  
Accepted 19 April 2024

## KEYWORDS

Change detection; multi-source data; structural feature; global structure graph; image fusion

## 1. Introduction

### 1.1. Background

Change detection (CD) detects and analyzes changes on the Earth's surface by acquiring remote sensing images from different times in the same geographical area (Han et al. 2024; Reba and Seto 2020). It is widely used in urban studies (Y. Tang and Zhang 2017; Y. Tang et al. 2024; Y. Chen et al. 2022; Y. Chen et al. 2022; Y. Chen et al. 2024), forest resource monitoring (Wei et al. 2023), and disaster warning and rescue (Brunner, Lemoine, and Bruzzone 2010; D. Wang et al. 2022). Nowadays, the majority of CD techniques are used for unimodal CD (UCD), such as homogenous multispectral, hyperspectral, and synthetic aperture radar (SAR) images (Bovolo and Bruzzone 2007; Saha, Bovolo, and Bruzzone 2019; H. Tang, Wang, and Zhang 2022). These images are acquired from the same sensor and have the same image

**CONTACT** Yuqi Tang ✉ [yqtang@csu.edu.cn](mailto:yqtang@csu.edu.cn) School of Geosciences and Info-Physics, Central South University, 410083, Changsha, People's Republic of China; Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring (Central South University), Ministry of Education, 410083, Changsha, People's Republic of China

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

characteristics, allowing for the extraction of change information through direct comparison. However, in many practical applications, it is difficult to obtain a set of high-quality unimodal images due to the influence of satellite performance, the environment, etc. The surge in remote sensing satellites has facilitated the integration of multimodal satellite data (Fan, Hou, and Shi 2021; Rajakumar and Satheeskumaran 2022). This integration, which leverages the fusion of imagery from various sensors, has significantly boosted the Earth observation capabilities of remote sensing images (Hong et al. 2023). As a result, it has attracted considerable interest within the academic community, highlighting the potential for more comprehensive and detailed insights into our planet's conditions (Hong et al. 2023; Hong et al. 2024). Moreover, it is difficult for UCD methods to achieve satisfactory CD results when the data comes from different satellites.

Multimodal CD (MCD) is defined by its analysis of pre-event and post-event images that originate from different sensor types or distinct satellites. By harnessing the power of diverse satellite imagery, MCD can combine optical images with SAR images, or compare images from different satellites such as Sentinel-2 and GaoFen-2. The essence of MCD lies in the use of pre-event and post-event data acquired by sensors that vary in nature or platform, which introduces a range of advantages over UCD include: 1) it is capable of concurrent image capture, enabling the fusion of complementary information from multiple sources (e.g. the spectral and textural information from optical images, and the all-weather imaging capabilities of SAR); 2) obtaining pre- and post-event images within a compressed timeframe reduces disaster response latency. As a result, MCD has become a popular research topic. In recent years, it has been explored through various methods, which can be divided into three categories based on their principles: image space transformation, feature space unification, and similarity measure.

1) *Image space transformation*. To transform multimodal images to each other's image space and extract the change information, most of these methods build the spatial transformation relationship between the images using a statistical model. (Z. Liu et al. 2018) constructed the image space transformation relationship between optical and SAR images by manually selecting samples from unchanged regions and obtained results by directly comparing the transformed images with the original images. To reduce the reliance on a priori information, (Luppino et al. 2019)) established a multimodal image space transformation relationship by creating pseudo-samples using affinity matrices and four regression models. Additionally, deep learning-based MCD methods have been proposed, such as the coupling translation network (CPTN) (Gong et al. 2019) and adversarial cyclic encoders network (ACE) (Luppino et al. 2022). However, due to the differences between multimodal images and the reliance on the collection of training samples or pseudo-samples, there are still significant variations between the original and transformed images.

2) *Feature space unification*. Multimodal images obtained using various sensors possess dissimilar imaging characteristics. Thus, to enable direct comparison, the images can be transformed into a unified feature space. This allows identical geographical objects across multimodal images to share common spectral features. This type of method can be divided into category feature space unification and latent feature space unification. Category feature space unification, also known as image classification-based methods, assigns the same category label to the same image objects, such as hierarchical extreme learning machine classification (HELMC) (Han et al. 2021) and multitemporal segmentation and compound classification (MS-CC) area (L. Wan, Xiang, and You 2019). However, these methods are limited by the classification accuracy of multimodal images, particularly SAR images affected by coherent speckle noise. Latent feature space unification enables the comparison of multimodal images by projecting them into a unified feature space, and is mainly based on deep learning, such as the symmetric convolutional coupling network (SCCN) (J. Liu et al. 2018),

approximately symmetrical deep neural network (ASDNN) (Zhao et al. 2017), two-stage joint feature learning (TSJFL) (Han, Tang, and Chen 2022), iterative feature mapping network (IFMN) (Zhan et al. 2018), commonality autoencoder change detection (CACD) (Wu et al. 2022), conditional generative adversarial network (CGAN) (Gong et al. 2017), logarithmic transformation feature learning (LTFL) (Zhan et al. 2018), etc. However, the existence of changes makes it challenging to uniformly match the spectral properties of all image objects in multimodal images.

3) *Similarity measure*. This type of method distinguishes between changed and unchanged image pixels by measuring the similarity of certain features of a multimodal image, with a lower similarity indicating a higher probability of change. (Mercier, Moser, and Serpico 2008) constructed the relationships of unchanged image pixels based on copula theory and extracted change information using the Kullback-Leibler distance. Wan *et al.* (L. Wan, Zhang, and You 2018) detected the changed regions by comparing the ‘ranked histogram’ features of optical and SAR images. (Alberga 2009) obtained the changed regions by calculating the similarity of multimodal images using mutual information. Additionally, some researchers have calculated the similarity of the grayscale features of multimodal images by constructing models, such as the multidimensional statistical model (MSM) (Prendes et al. 2015), energy-based model (EBM) (Touati and Mignotte 2018), convolution model-based mapping (CMBM) (Touati, Mignotte, and Dahmane 2019), multimodal change detection Markovian model (M3CD) (Touati, Mignotte, and Dahmane 2020), and graph-based fusion (GBF) (Jimenez-Sierra et al. 2020). Despite the potential of existing methods to accurately measure change information in complex image change scenes, the construction of appropriate features is often challenging. Furthermore, the similarity measurement between multimodal images is further complicated by the presence of significant noise in SAR images, leading to an increased false detection rate.

## 1.2. Motivation and contribution

Despite the considerable differences in imaging features between multimodal images, some researchers have proposed that there are common structural features that can be expressed through a graph based on the self-similarity attributes of the image. In Sun, Lei, Li, et al. (2021), a non-local patch-based graph (NLPG) is constructed in two images separately, and the changes in the NLPG are produced by cross-mapping the pre- and post-temporal images. Based on the NLPG, they proposed a novel graph difference metric formula to calculate the divergence between the structure graphs of multimodal images. Additionally, Mignotte (2020) extracted the structural fractal features of the image and mapped them to the domain of the other image, then employed a fractal decoding strategy to obtain the change information and combined it with Markovian segmentation to identify the changed regions. Subsequently, the iterative robust graph and Markovian co-segmentation method (IRG-McS) was proposed (Sun et al. 2021), which acquires a graph structure by updating the graph, obtains the optimal difference graph through iteration, and combines it with Markovian segmentation to obtain the change map (CM). Moreover, the sparse-constrained adaptive structure consistency-based method (SCASC) builds a structural consistency image regression model and uses the Markov segmentation model to differentiate changed and unchanged regions in different images (Sun et al. 2022). Chen et al. (2022) used graph autoencoders to characterize the structural features of multimodal images and extract the changed regions by comparing the structural relationships across different modalities.

Although these methods demonstrate the efficacy of the structure graph method for MCD, some problems remain: 1) Existing methods, when constructing structural graphs,

predominantly concentrate on the average pixel differences between image patches, overlooking the overall similarity of the patches. This focus can amplify noise and outliers, leading to inaccurate similarity measurements. Additionally, these methods often rely on non-local information, which does not capture the full context of the image, including the layout and relationships between objects, resulting in an incomplete representation of image structural features. 2) There is a lack of adequate consideration for structural changes, with a uniform weight applied to all image patches for change measurement. This approach fails to account for the nuanced differences between structural features and their mapped counterparts within the same image domain. Different image patches, representing various land cover types such as buildings, vegetation, and water bodies, respond differently to changes, and their significance varies. A uniform weight does not consider this diversity, which can lead to over-sensitivity or neglect of certain land cover types, ultimately affecting the overall performance of CD. 3) The current focus on local feature fusion in change intensity maps (CIMs) overlooks the importance of global features. While local features are prone to noise and outliers, which can result in false positives, global features offer more stable information that enhances the robustness of CD.

In this paper, we propose a global structure graph mapping (GSGM)-based MCD method that fully utilizes the global structure information of images and improves the accuracy of CD. It segments the image into overlapping patches and forms a global structure graph (GSG) that optimally employs global patches to represent the image's structural features. Vertices in GSG mainly consist of target and vertex patches (for convenience, image patches in GSG besides the target ones are referred to as 'vertex patches' in this paper). The proposed GSGM exploits the overall similarity between target and vertex patches to construct a robust structure graph for each of the multimodal images, which are mapped to each other's image domains to explore the change information of multimodal images. Subsequently, a weighted formula is established to calculate the change intensity (CI) of the image, which stems from both the GSGs post-mapping and the disparities between the GSGs and their corresponding mapped versions within the same image domain. Furthermore, GSGM fuses the forward and backward CIMs using the low-rank representation fusion technique to enhance the expression of CI, and the final CM is obtained by threshold segmentation. In summary, the contributions of this paper are as follows:

- 1) The similarity at the patch level between image patches is assessed, offering a nuanced understanding of their relationship. A GSG is introduced to encapsulate the image's structural features, providing a holistic view of the image's layout and structure. This approach, which focuses on both local details and global composition, allows for a comprehensive representation of image content. The GSG evaluates both similarities and dissimilarities among vertices, which aids in more accurate change localization.
- 2) A novel weighted structural difference formula has been developed to more precisely capture the significance of individual image patches in change measurement. This method assigns variable weights to each patch, enabling a more granular assessment of structural changes. Furthermore, the changes in the GSGs after mapping, as well as discrepancies between the GSGs and mapped GSGs in the same image domain, are considered, enhancing the robustness of the change measurement approach.
- 3) The methodology also incorporates low-rank representation to fuse forward and backward CIMs, integrating global features such as the image's overall structure or background patterns with local features that capture detailed information. This integration results in a more robust CIM.
- 4) The proposed GSGM is fully unsupervised and does not require any labeled data, and the effectiveness of GSGM is validated by experiments conducted across nine datasets, where it was compared with state-of-the-art methods.

## 2. Methodology

Consider two co-registered multimodal images  $\mathbf{X} = \{x(i, j, b) | 1 \leq i \leq H, 1 \leq j \leq W, 1 \leq b \leq B_X\}$  and  $\mathbf{Y} = \{y(i, j, b) | 1 \leq i \leq H, 1 \leq j \leq W, 1 \leq b \leq B_Y\}$ . Here,  $H$  and  $W$  denote the length and width of the images, and  $B_X$  and  $B_Y$  represent the number of bands in the image  $\mathbf{X}$  and  $\mathbf{Y}$  respectively. Recognizing that multimodal images  $\mathbf{X}$  and  $\mathbf{Y}$  are from distinct data domains, a direct comparison of their spectral characteristics is impractical. However, in regions that remain unchanged, they share consistent structural attributes. To explore these attributes, we utilize GSG encoding techniques, which encode the structural properties of  $\mathbf{X}$  and  $\mathbf{Y}$  into their respective graphs, more accurately capturing the global structural features of the images. By creating GSGs that capture the inherent structural similarities and differences between  $\mathbf{X}$  and  $\mathbf{Y}$ , we can effectively extract the nuances of changes in multimodal imagery. These structural attributes are manifested in the connectivity patterns between vertices and neighbor vertices within the graph. Changes in a region will consequently modify the connectivity patterns of the corresponding vertices. In constructing the structural graph, each image patch is considered a distinct vertex, encapsulating features such as color and texture that represent specific image areas. The relationships among these vertices unveil the images' patterns, hierarchies, and spatial arrangements. Comparing the GSGs of  $\mathbf{X}$  and  $\mathbf{Y}$  allows us to quantify the changes between the multimodal images. As illustrated in the framework (Figure 1), the proposed method encompasses three key stages: (1) construction of a GSG using image patches as the fundamental processing units; (2) evaluation of image CI through the forward/backward mapping of the GSG; and (3) generation of a CM through CIMs fusion and threshold segmentation.

### 2.1. GSG construction

Consider a target patch in the image  $\mathbf{X}$ , which is defined as  $\mathbf{p}_{(i,j)}^{\mathbf{X}} = \{\mathbf{X}(i - p_s : i + p_s, j - p_s : j + p_s, 1 : B_X)\}$ ,  $1 \leq i \leq H$ ,  $1 \leq j \leq W$ , whose width is  $w_s = 2p_s + 1$ . To make full use of the structural features of images, a GSG  $G_{(i,j)}$  is constructed to represent the global structural features of

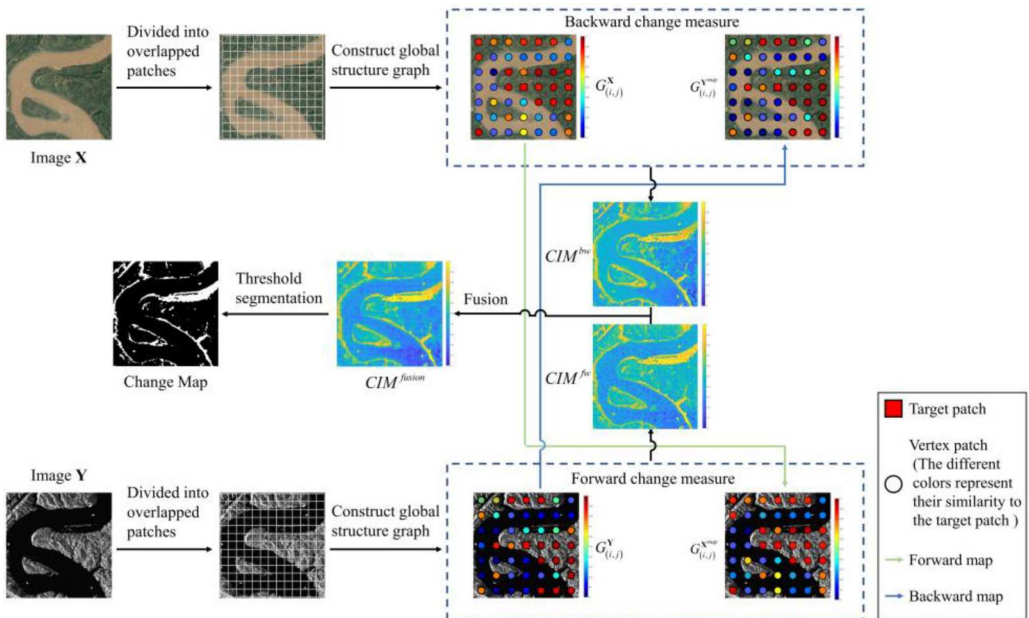


Figure 1. Framework of GSGM-based MCD method.

the images in this paper. Taking the target patch  $\mathbf{p}_{(i,j)}^{\mathbf{X}}$  as an example, its GSG  $G_{(i,j)}^{\mathbf{X}}$  is defined as follows:

$$\begin{aligned}
 G_{(i,j)}^{\mathbf{X}} &= \{V_{(i,j)}^{\mathbf{X}}, E_{(i,j)}^{\mathbf{X}}, w_{(i,j)}^{\mathbf{X}}\} \\
 V_{(i,j)}^{\mathbf{X}} &= \left\{ \begin{array}{l} \mathbf{p}_{(i,j)}^{\mathbf{X}}, \mathbf{p}_{(m,n)}^{\mathbf{X}}; (m, n) \in \Omega_{(m,n)}^{\mathbf{X}}, \\ (i, j) \neq (m, n) \end{array} \right\} |V_{(i,j)}^{\mathbf{X}}| = N \\
 E_{(i,j)}^{\mathbf{X}} &= \{(\mathbf{p}_{(i,j)}^{\mathbf{X}}, \mathbf{p}_{(m,n)}^{\mathbf{X}}) | \mathbf{p}_{(i,j)}^{\mathbf{X}}, \mathbf{p}_{(m,n)}^{\mathbf{X}} \in V_{(i,j)}^{\mathbf{X}}\} \\
 w_{\mathbf{X}}(\mathbf{p}_{(i,j)}^{\mathbf{X}}, \mathbf{p}_{(m,n)}^{\mathbf{X}}) &= \text{sim}(\mathbf{p}_{(i,j)}^{\mathbf{X}}, \mathbf{p}_{(m,n)}^{\mathbf{X}}) \forall (\mathbf{p}_{(i,j)}^{\mathbf{X}}, \mathbf{p}_{(m,n)}^{\mathbf{X}}) \in E_{(i,j)}^{\mathbf{X}}
 \end{aligned} \tag{1}$$

where  $V_{(i,j)}^{\mathbf{X}}$  denotes the vertex of GSG  $G_{(i,j)}^{\mathbf{X}}$ , and each vertex patch  $\mathbf{p}_{(m,n)}^{\mathbf{X}}$  is connected to target patch  $\mathbf{p}_{(i,j)}^{\mathbf{X}}$  through edge  $E_{(i,j)}^{\mathbf{X}}$ . In GSG  $G_{(i,j)}^{\mathbf{X}}$ , the number of vertex patches is  $N$ . The similarity between each vertex patch  $\mathbf{p}_{(m,n)}^{\mathbf{X}}$  and target patch  $\mathbf{p}_{(i,j)}^{\mathbf{X}}$  denotes its connection weight  $w_{(i,j)}^{\mathbf{X}}$  and  $\text{sim}(\cdot)$  represents the similarity calculation operation.  $\Omega_{(m,n)}^{\mathbf{X}}$  is an index set of vertex patch  $\mathbf{p}_{(m,n)}^{\mathbf{X}}$  arranged in descending order according to the similarity between vertex patch  $\mathbf{p}_{(m,n)}^{\mathbf{X}}$  and target patch  $\mathbf{p}_{(i,j)}^{\mathbf{X}}$ .

Similarly, GSG  $G_{(i,j)}^{\mathbf{Y}}$  of image  $\mathbf{Y}$  can also be constructed.

## 2.2. Patch similarity measure

The high correlation between adjacent pixels in remote sensing images has been widely acknowledged. However, traditional methods for measuring image patch similarity are based on the calculation of the difference between independent pixels, which makes them vulnerable to the influence of noise, thus leading to inaccurate similarity measurements. To address the similarity measurement of pixel difference, this paper introduces the structural similarity index measure (SSIM) (Z. Wang et al. 2004), which evaluates image similarity by taking into account image brightness, contrast, and structure. SSIM is defined as follows:

$$SSIM(x, y) = [l(x, y)]^{\alpha} [c(x, y)]^{\beta} [s(x, y)]^{\gamma} \tag{2}$$

where

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{3}$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{4}$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{5}$$

where  $l(x, y)$ ,  $c(x, y)$ , and  $s(x, y)$  represent the brightness comparison function, contrast comparison function, and structure comparison function of image patches  $x$  and  $y$  respectively;  $\alpha > 0$ ,  $\beta > 0$ , and  $\gamma > 0$  are the adjust parameters;  $\mu_x$  and  $\mu_y$  are the mean values of  $x$  and  $y$  respectively;  $\sigma_x$  and  $\sigma_y$  are the variance of  $x$  and  $y$  respectively;  $\sigma_{xy}$  denotes the covariance of  $x$  and  $y$ ; and  $C_1$ ,  $C_2$ , and  $C_3$  are constant to keep  $l(x, y)$ ,  $c(x, y)$ , and  $s(x, y)$  stable. The value of SSIM ranges from 0 to 1. The greater the value, the more similar the two patches. It is generally posited that  $l(x, y)$ ,  $c(x, y)$ , and  $s(x, y)$  make equal contributions to SSIM, thereby leading to  $\alpha = \beta = \gamma = 1$  being conventionally set at unity (Z. Wang et al. 2004).



### 2.3. CI measure

GSGs  $G_{(i,j)}^X$  and  $G_{(i,j)}^Y$  are constructed in images  $X$  and  $Y$  respectively, and express the structural features of the images together with the changes between them. However, due to the differences in the imaging characteristics of multimodal images, the changes between GSGs cannot be effectively measured by a direct comparison. To accurately measure the differences in multimodal images and circumvent the imaging discrepancies between them, GSGs  $G_{(i,j)}^X$  and  $G_{(i,j)}^Y$  are mapped to image domains  $\mathcal{Y}$  and  $\mathcal{X}$  respectively to obtain their maps  $G_{(i,j)}^{Xmap}$  and  $G_{(i,j)}^{Ymap}$  for the comparison of multimodal images in the same image domain. As an example, GSG  $G_{(i,j)}^X$  is mapped into image  $Y$  (forward mapping), and its mapped GSG  $G_{(i,j)}^{Xmap}$  is defined as follows:

$$\begin{aligned} G_{(i,j)}^{Xmap} &= \{V_{(i,j)}^{Xmap}, E_{(i,j)}^{Xmap}, w_{(i,j)}^{Xmap}\} \\ V_{(i,j)}^{Xmap} &= \left\{ \mathbf{p}_{(\tilde{m}, \tilde{n})}^Y; (\tilde{m}, \tilde{n}) \in \Omega_{(m,n)}^X, \right. \\ &\quad \left. (i, j) \neq (\tilde{m}, \tilde{n}) \right\} | V_{(i,j)}^{Xmap} | = N \\ E_{(i,j)}^{Xmap} &= \{(\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(\tilde{m}, \tilde{n})}^Y) | \mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(\tilde{m}, \tilde{n})}^Y \in V_{(i,j)}^{Xmap}\} \\ w_{Xmap}(\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(\tilde{m}, \tilde{n})}^Y) &= sim(\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(\tilde{m}, \tilde{n})}^Y) \forall (\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(\tilde{m}, \tilde{n})}^Y) \in E_{(i,j)}^{Xmap} \end{aligned} \quad (6)$$

By forward mapping, the connectivity between the vertices of mapped GSG  $G_{(i,j)}^{Xmap}$  is consistent with GSG  $G_{(i,j)}^X$ , but the patches of the mapped GSG  $G_{(i,j)}^{Xmap}$  are all represented using the pixel values of image  $Y$ . This can indicate that GSG  $G_{(i,j)}^{Xmap}$  and GSG  $G_{(i,j)}^Y$  are in the same image domain, thus the CI of the region where the target patch  $\mathbf{p}_{(i,j)}^Y$  is located can be measured. In this paper, we consider that the structural differences between  $G_{(i,j)}^{Xmap}$  and  $G_{(i,j)}^Y$  are mainly composed of the following two parts:

1) *Intrinsic change of GSG  $G_{(i,j)}^{Xmap}$* . If the region represented by target patch  $\mathbf{p}_{(i,j)}^Y$  remains unchanged, the internal pattern of the mapped GSG  $G_{(i,j)}^{Xmap}$  is stable; that is, the similarity of vertex patches to target patch  $\mathbf{p}_{(i,j)}^Y$  is stable. Otherwise, once the region represented by target patch  $\mathbf{p}_{(i,j)}^Y$  changes, the similarity of vertex patches to target patch  $\mathbf{p}_{(i,j)}^Y$  would change, and the change information would express the intrinsic change of GSG  $G_{(i,j)}^{Xmap}$ . This change is calculated as follows:

$$dif_{(i,j)}^{Y1} = \frac{1}{N} \left( \sum_{\substack{(m,n)^s \in \Omega_{(m,n)}^Y \\ (\tilde{m}, \tilde{n})^s \in \Omega_{(\tilde{m}, \tilde{n})}^X}} |sim(\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(m,n)^s}^Y) - sim(\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(\tilde{m}, \tilde{n})^s}^Y)| \right) \quad (7)$$

where  $(m, n)^s \in \Omega_{(m,n)}^Y$  denotes the index of vertex patch  $\mathbf{p}_{(m,n)^s}^Y$  that ranks  $s$ -th in similarity to target patch  $\mathbf{p}_{(i,j)}^Y$  in vertex  $V_{(i,j)}^Y$ , and  $(\tilde{m}, \tilde{n})^s \in \Omega_{(\tilde{m}, \tilde{n})}^X$  represents the index of vertex patch  $\mathbf{p}_{(\tilde{m}, \tilde{n})^s}^Y$  that ranks  $s$ -th in similarity to target patch  $\mathbf{p}_{(i,j)}^X$  in vertex  $V_{(i,j)}^{Xmap}$ . The change in the region represented by target patch  $\mathbf{p}_{(i,j)}^Y$  can be calculated from the difference in the similarity of vertex patches  $\mathbf{p}_{(m,n)^s}^Y$  and  $\mathbf{p}_{(\tilde{m}, \tilde{n})^s}^Y$  with target patch  $\mathbf{p}_{(i,j)}^Y$ . For instance, if vertex patch  $\mathbf{p}_{(m,n)^s}^X$  in GSG  $G_{(i,j)}^X$  exhibits a high similarity to target patch  $\mathbf{p}_{(i,j)}^X$ , and the region represented by target patch  $\mathbf{p}_{(i,j)}^Y$  has changed, vertex patch  $\mathbf{p}_{(\tilde{m}, \tilde{n})^s}^Y$  in GSG  $G_{(i,j)}^{Xmap}$  will display a low similarity to target patch  $\mathbf{p}_{(i,j)}^Y$ , resulting in a diminished value for  $sim(\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(\tilde{m}, \tilde{n})^s}^Y)$ . Conversely, in GSG  $G_{(i,j)}^Y$ , vertex patch  $\mathbf{p}_{(m,n)^s}^Y$  and target patch  $\mathbf{p}_{(i,j)}^Y$  exhibit a high similarity, and  $sim(\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(m,n)^s}^Y)$  has a greater value. Consequently, the disparity between



$sim(\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(\tilde{m},\tilde{n})^s}^Y)$  and  $sim(\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(m,n)^s}^Y)$  is more noticeable, accentuating the change of the region represented by target patch  $\mathbf{p}_{(i,j)}^Y$ .

It is worth noting that the top and bottom similarity vertex patches are often sensitive to the change in the region represented by the target patch. This is because, once the region represented by target patch  $\mathbf{p}_{(i,j)}^X$  changes, the similarity between the most and least similar vertex patch  $\mathbf{p}_{(m,n)}^X$  in mapped graph  $G_{(i,j)}^{X^{map}}$  and target patch  $\mathbf{p}_{(i,j)}^Y$  will change with high probability. To further highlight changes, we employ a weighting approach to augment the change in the similarity of vertex patch  $\mathbf{p}_{(m,n)^s}^Y$  to target patch  $\mathbf{p}_{(i,j)}^Y$ , so (7) can be reformulated as follows:

$$dif_{(i,j)}^{Y1} = \frac{1}{N} \left( \sum_{\substack{(m,n)^s \in \Omega_{(m,n)}^Y \\ (\tilde{m},\tilde{n})^s \in \Omega_{(\tilde{m},\tilde{n})}^X}} |\beta_{(m,n)^s}^{Y1} \times sim(\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(m,n)^s}^Y) - \beta_{(\tilde{m},\tilde{n})^s}^{Y1} \times sim(\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(\tilde{m},\tilde{n})^s}^Y)| \right) \quad (8)$$

$$\beta_{(m,n)^s}^{Y1} = \exp \left( \lambda \times \left| sim(\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(m,n)^s}^Y) - \frac{1}{N} \left( \sum_{(m,n)^s \in \Omega_{(m,n)}^Y} sim(\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(m,n)^s}^Y) \right) \right| \right) \quad (9)$$

$$\beta_{(\tilde{m},\tilde{n})^s}^{Y2} = \exp \left( \lambda \times \left| sim(\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(\tilde{m},\tilde{n})^s}^Y) - \frac{1}{N} \left( \sum_{(\tilde{m},\tilde{n})^s \in \Omega_{(\tilde{m},\tilde{n})}^X} sim(\mathbf{p}_{(i,j)}^Y, \mathbf{p}_{(\tilde{m},\tilde{n})^s}^Y) \right) \right| \right) \quad (10)$$

where  $\beta_{(m,n)^s}^{Y1}$  and  $\beta_{(\tilde{m},\tilde{n})^s}^{Y2}$  donate the variation weight of vertex patches  $\mathbf{p}_{(m,n)^s}^Y$  and  $\mathbf{p}_{(\tilde{m},\tilde{n})^s}^Y$  respectively; and  $\lambda$  is the weight coefficient of  $\beta_{(m,n)^s}^{Y1}$  and  $\beta_{(\tilde{m},\tilde{n})^s}^{Y2}$ .

2) *Difference between mapped GSG  $G_{(i,j)}^{X^{map}}$  and GSG  $G_{(i,j)}^Y$ .* Once a region in the image has changed, the structural features of mapped GSG  $G_{(i,j)}^{X^{map}}$  and GSG  $G_{(i,j)}^Y$  will be dissimilar, and the structural difference between mapped GSG  $G_{(i,j)}^{X^{map}}$  and GSG  $G_{(i,j)}^Y$  will reflect the change in the region. This structural difference can be calculated as follows:

$$dif_{(i,j)}^{Y2} = \exp(\lambda) - \frac{1}{N} \left( \sum_{\substack{(m,n)^s \in \Omega_{(m,n)}^Y \\ (\tilde{m},\tilde{n})^s \in \Omega_{(\tilde{m},\tilde{n})}^X}} |\beta_{(m,n)^s}^{Y1} \times sim(\mathbf{p}_{(m,n)^s}^Y, \mathbf{p}_{(\tilde{m},\tilde{n})^s}^Y)| \right) \quad (11)$$

To be more specific, if the region represented by target patch  $\mathbf{p}_{(i,j)}^Y$  remains unchanged, then vertex patches  $\mathbf{p}_{(m,n)^s}^Y$  and  $\mathbf{p}_{(\tilde{m},\tilde{n})^s}^Y$  are likely to retain a high similarity to target patch  $\mathbf{p}_{(i,j)}^Y$ , resulting  $insim(\mathbf{p}_{(m,n)^s}^Y, \mathbf{p}_{(\tilde{m},\tilde{n})^s}^Y)$  being a large value and  $dif_{(i,j)}^{Y2}$  being a small value. However, if the region represented by target patch  $\mathbf{p}_{(i,j)}^Y$  changes, the similarity of vertex patch  $\mathbf{p}_{(\tilde{m},\tilde{n})^s}^Y$  to target patch  $\mathbf{p}_{(i,j)}^Y$  will likely change, causing  $sim(\mathbf{p}_{(m,n)^s}^Y, \mathbf{p}_{(\tilde{m},\tilde{n})^s}^Y)$  to be a small value and  $dif_{(i,j)}^{Y2}$  to be a large value with a high probability, thus signifying a significant change in the region represented by target patch  $\mathbf{p}_{(i,j)}^Y$ .

The final structural difference  $dif_{(i,j)}^Y$  of the region represented by target patch  $\mathbf{p}_{(i,j)}^Y$  is the aggregate of the two structural differences mentioned above:

$$dif_{(i,j)}^Y = dif_{(i,j)}^{Y1} + dif_{(i,j)}^{Y2} \quad (12)$$

The mean value of structural difference  $F_{(s,t)}^Y$  of all encompassed image pixels  $(s, t)$ ,  $1 \leq s \leq H$ ,  $1 \leq t \leq W$  is taken as the change information of pixels  $(s, t)$ , and the forward

mapping CIM  $CIM^{fw}$  is calculated as follows:

$$CIM_{(s,t)}^{fw} = \frac{1}{|F_{(s,t)}^Y|} \sum_{dif_{(i,j)}^Y \in F_{(s,t)}^Y} dif_{(i,j)}^Y \quad (13)$$

In the same way, the backward mapping CIM  $CIM^{bw}$  can be obtained by mapping GSG  $G_{(i,j)}^Y$  to image domain  $\mathcal{X}$ .

Setting the target patch's step size to  $\Delta_p = p_s$  will accelerate the calculation, and the search step size of the vertex patch is set to  $\Delta_v = \lfloor \alpha_{\Delta_v} \times \min(H/2, W/2) \rfloor$ , where  $\alpha_{\Delta_v}$  donates the step size factor which regulates the step size, while  $\lfloor \cdot \rfloor$  is the downward integer operation.

## 2.4. CIMs fusion

The CI for each image pixel is represented by CIMs  $CIM^{fw}$  and  $CIM^{bw}$  which are generated from image domains  $D^{fw}$  and  $D^{bw}$  respectively. The change intensities of the same index in CIMs  $CIM^{fw}$  and  $CIM^{bw}$  can differ because images  $\mathbf{X}$  and  $\mathbf{Y}$  are from distinct data modalities. Therefore, in order to obtain robust CIMs,  $CIM^{fw}$  and  $CIM^{bw}$  must be fused, and the global and local structural features of CIMs  $CIM^{fw}$  and  $CIM^{bw}$  are considered for CIMs fusion.

Inspired by latent low-rank representation (LLRR) (G. Liu and Yan 2011), we decompose CIMs  $CIM^{fw}$  and  $CIM^{bw}$  into a low-rank part (global structural), significance part (local structural), and image noise part. The CIM decomposition is obtained by the following optimization function:

$$\begin{aligned} \min_{Z,L,E} \quad & \|Z\|_* + \|L\|_* + \lambda \|E\|_1 \\ \text{s.t.} \quad & D = DZ + LD + E \end{aligned} \quad (14)$$

where  $D$  is the CIM,  $Z$  represents the low rank coefficient,  $L$  denotes the significance coefficient,  $\lambda$  is a positive parameter, and  $\|\cdot\|_*$  and  $\|\cdot\|_1$  are the nuclear norm and  $l_1$  norm respectively.

To facilitate the solution, (14) is equivalent to the following equation:

$$\begin{aligned} \min_{Z,L,J,S,E} \quad & \|J\|_* + \|S\|_* + \lambda \|E\|_1 \\ \text{s.t.} \quad & D = DZ + LD + E, Z = J, L = S \end{aligned} \quad (15)$$

Based on an augmented Lagrange multiplier (ALM), (15) can be converted into an unconstrained augmented Lagrange function:

$$\begin{aligned} \mathcal{L}(J, S, Z, L, E, \eta) = & \|J\|_* + \|S\|_* + \lambda \|E\|_1 \\ & + \text{tr}(\eta_1^T (D - DZ - LD - E)) \\ & + \text{tr}(\eta_2^T (Z - J)) + \text{tr}(\eta_3^T (L - S)) \\ & + \frac{\mu}{2} (\|D - DZ - LD - E\|_F^2) \\ & + \frac{\mu}{2} (\|Z - J\|_F^2 + \|L - S\|_F^2) \end{aligned} \quad (16)$$

where  $\text{tr}(\cdot)$  and  $\|\cdot\|_F$  are the trace and Frobenius norm of the matrix,  $\eta$  is the Lagrange multiplier, and  $\mu$  is the penalty parameter. (16) achieves the update of  $J$ ,  $S$ ,  $Z$ ,  $L$ , and  $E$  respectively by fixing other parameters.

With the above calculation, the low-rank coefficients  $Z^{fw}$ ,  $Z^{bw}$  and significance coefficients  $L^{fw}$ ,  $L^{bw}$  of CIMs  $CIM^{fw}$  and  $CIM^{bw}$  can be obtained. Mean weighted fusion is used for the low-rank part of the CIMs in order to preserve the low-rank information:

$$CIM_Z^{fusion} = \frac{1}{2} (Z^{fw} \times CIM^{fw} + Z^{bw} \times CIM^{bw}) \quad (17)$$

**Table 1.** GSGM framework.

GSGM

**Input:** images  $\mathbf{X}$  and  $\mathbf{Y}$ , parameters of  $p_s$ ,  $\lambda$ ,  $\alpha_{\Delta_v}$  and  $\zeta$ **1. GSG Construction:**

Calculation of patch similarity using SSIM;

Construct GSGs  $G_{(ij)}^X$  and  $G_{(ij)}^Y$ .**2. CI Measure:**Calculation of intrinsic change of GSG  $G_{(ij)}^{X^{map}}$ ;Calculation of difference between mapped GSG  $G_{(ij)}^{X^{map}}$  and GSG  $G_{(ij)}^Y$ ;Calculation of forward CIM  $CIM^{fw}$  and backward CIM  $CIM^{bw}$ .**3. CIMs fusion:**

Fusion of low-rank part (global structure);

Fusion of significance part (local structure).

**4. CM Generation:**

Compute the binary CM with (20).

Furthermore, to preserve the significance features of the CIMs and highlight the changed regions, the significance part is fused using the square weighted fusion method:

$$CIM_L^{fusion} = \frac{1}{2}((L^{fw} \times CIM^{fw})^2 + (L^{bw} \times CIM^{bw})^2) \quad (18)$$

The fused CIM  $CIM^{fusion}$  is calculated as follows:

$$CIM^{fusion} = CIM_Z^{fusion} + CIM_Z^{fusion} \quad (19)$$

## 2.5. CM generation

To further distinguish the changed and unchanged regions, the CM is obtained using the threshold segmentation method proposed in (Hou, Wang, and Liu 2017):

$$CM_{(s,t)} = \begin{cases} 1, & \text{if } CIM_{(s,t)}^{fusion} \geq \zeta \times \text{mean}(CIM^{fusion}) \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

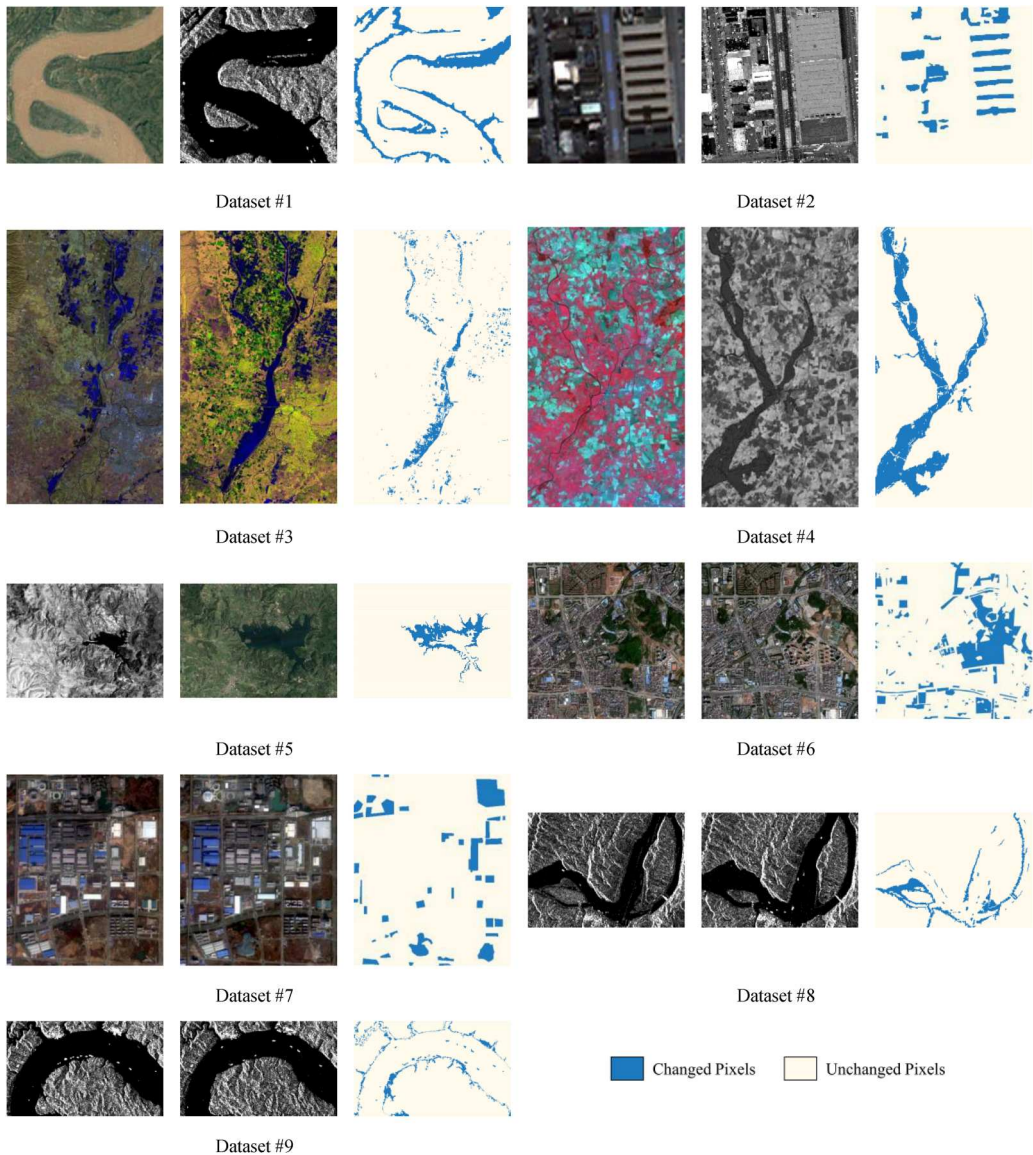
where  $\zeta$  represents the user adjustment threshold parameter and  $\text{mean}(\cdot)$  donates the mean value operation. The GSGM framework is summarized in Table 1:

## 3. Experiments and discussion

In this section, nine experimental datasets are introduced and used in an accuracy evaluation with several comparison methods, followed by the experimental results and parameter analysis of the proposed GSGM.

### 3.1. Datasets

The effectiveness of the proposed GSGM in MCD and its applicability in UCD are validated using five multimodal datasets (datasets #1–#5) and four unimodal datasets (datasets #6–#9). Images from different satellite types are included in the multimodal datasets, such as dataset #1 (optical and SAR images), dataset #2 (LiDAR and optical images), datasets #3 and #5 (images with the same sensor type but from different sensors), and dataset #4 (optical and normalized difference vegetation index (NDVI) images). The unimodal data consists of two optical images and two SAR images (Figure 2). All images underwent preprocessing, such as optical imagery (radiometric calibration, geometric



**Figure 2.** Datasets #1–#9; each dataset from left to right is the pre-event image, post-event image, and reference CM.

calibration, atmospheric correction, etc.) and SAR imagery (radiometric calibration, geometric calibration, filtering, interferometric processing, etc.). These preprocessing steps are intended to eliminate or reduce influence from sensors, platforms, or the environment in order to make the images more suitable for further analysis.

As shown in Table 2, the datasets used in this study had a large time window (the earliest image was acquired in December 1999 and the most recent in May 2018), a variety of sensor images with resolutions ranging from 2 m (dataset #6) to 30 m (dataset #5) in different regions of the world, and image sizes ranging from  $300 \times 412$  to  $3,500 \times 2,000$  covering multiple change scenarios (e.g. urban construction, flooding, river expansion, etc.). The robustness and efficiency of the proposed GSGM may be demonstrated by the nine datasets.

**Table 2.** Descriptions of datasets. In the ‘size’ column, the numbers within the parentheses indicate the number of bands for the imagery of the second time phase. For example, in dataset #1,  $600 \times 600 \times 3(1)$  signifies that the imagery from the first time phase has 3 bands, while the imagery from the second time phase has 1 band.

Dataset	Sensor	Size(pixels)	Date	Location	Event (& Spatial resolution)
#1	Google Earth/ Sentinel-1	$600 \times 600 \times 3(1)$	Dec. 1999–Nov. 2017	Chongqing, China	River expansion (10 m)
#2	QuickBird/LiDAR	$400 \times 400 \times 3(1)$	Nov. 2007–June 2011	San Francisco, USA	Urban construction (0.5 m)
#3	Landsat-8/Sentinel-1	$3,500 \times 2,000 \times 11(3)$	Jan. 2017–Feb. 2017	Sutter County, USA	Flooding ( $\approx 15$ m)
#4	Spot/NDVI	$990 \times 554 \times 3(1)$	1999–2000	Gloucester, England	Flooding ( $\approx 25$ m)
#5	Landsat-5/Google Earth	$300 \times 412 \times 1(3)$	Sept. 1995–July 1996	Sardinia, Italy	Lake expansion (30 m)
#6	WorldView-3	$1,431 \times 1,431 \times 3$	2010–2015	Shenzhen, China	Urban construction (2 m)
#7	Zi-Yuan 3	$458 \times 559 \times 3$	2014–2016	Wuhan, China	Urban construction (5.8 m)
#8	Sentinel-1	$516 \times 700 \times 1$	Nov. 2017–May 2018	Chongqing, China	River expansion (10 m)
#9	Sentinel-1	$898 \times 1,500 \times 1$	Nov. 2017–May 2018	Chongqing, China	River expansion (10 m)

### 3.2. Evaluation metrics

Empirical receiver operating characteristics (ROC) curves are used to evaluate the effectiveness of the CIMs produced by each comparison method. ROC curves represent the estimated pixel-wise probability of detection (PD) as a function of the probability of false alarm (PFA) by varying the binary segmentation threshold. The area under the curve (AUC) of the ROC curves is determined as the evaluation metric. For binary CM, Overall accuracy (OA), Kappa coefficient (KC), and F1-measure (F1) are used as the evaluation metrics:

$$OA = (TP + TN)/N \quad (21)$$

$$KC = (OA - PRE)/(1 - PRE) \quad (22)$$

where

$$PRE = \frac{(TP + FN)(TP + FP) + (TN + FP)(TN + FN)}{N_p^2} \quad (23)$$

$$F1 = (2 * P * R)/(P + R) \quad (24)$$

where

$$P = TP/(TP + FP) \quad (25)$$

$$R = TP/(TP + FN) \quad (26)$$

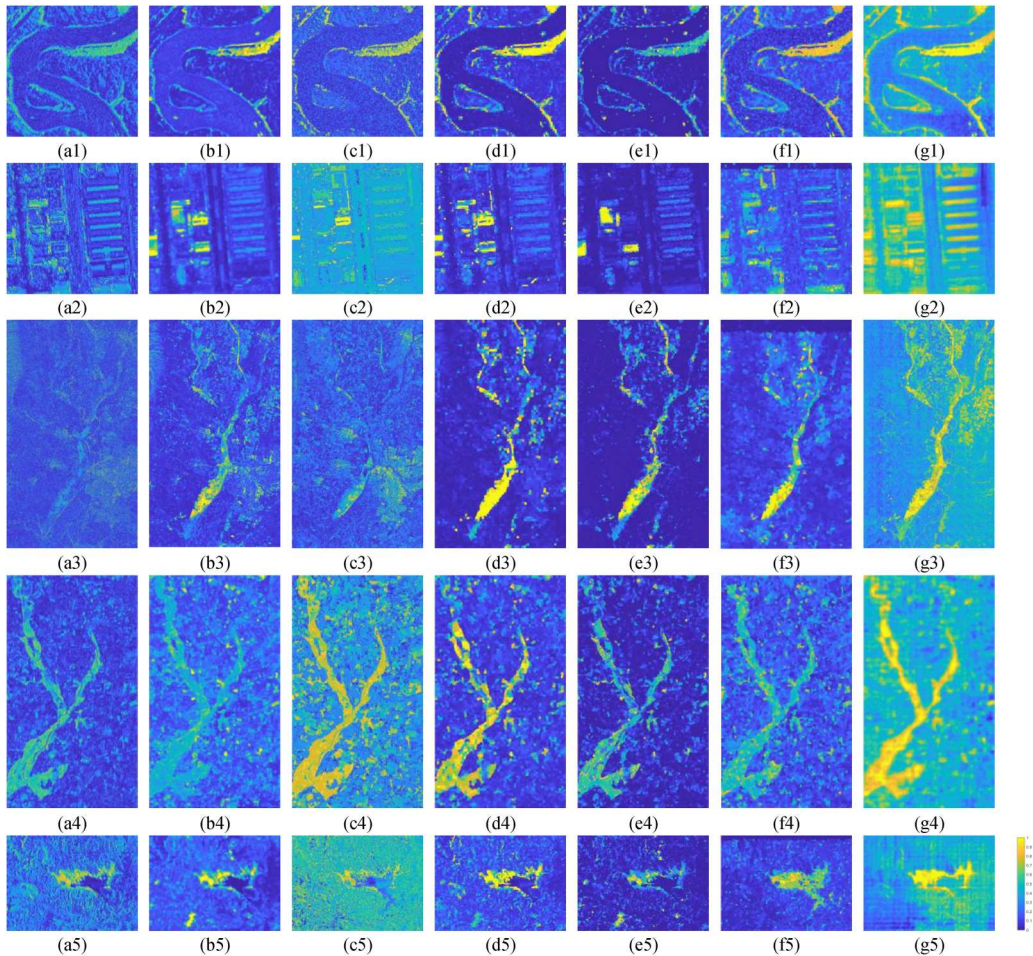
where  $N_p$  represents the total number of pixels in the image; TP means true positives, representing the number of correctly detected changed pixels; TN means true negatives, representing the number of correctly detected unchanged pixels; FP means false positives, representing the number of incorrectly detected changed pixels; and FN means false negatives, indicating the number of incorrectly detected unchanged pixels.



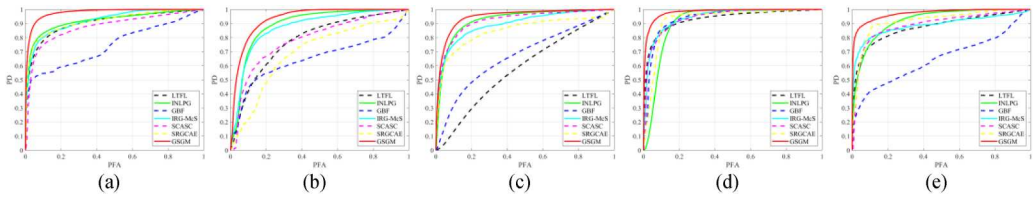
### 3.3. Comparison methods

In order to prove the effectiveness of the proposed GSGM, five novel methods are selected as comparison methods:

- 1) LTFL (Zhan et al. 2018): Logarithmic transformation feature learning framework (LTFL) displays the stack denoising autoencoder to extract high-dimensional features from remote sensing images, and selects reliable samples to train a neural network classifier to distinguish between changed and unchanged regions according to the classifier.
- 2) INLPG (Sun et al. 2022): Improved nonlocal patch-based graph (INLPG) builds a K-nearest neighbor (KNN) graph in the two temporal images respectively, and compares the KNN graph in the same image domain to extract the change information.
- 3) GBF (Jimenez-Sierra et al. 2020): Graph-based fusion (GBF) uses the Laplacian matrix of the regularized graph to minimize the graph similarity of the two temporal images and highlight the changed regions.
- 4) IRG-McS (Sun et al. 2021): Iterative robust graph and Markovian co-segmentation models (IRG-McS) construct a KNN graph to represent the structure of each image, and the CIM is



**Figure 3.** CIMs of different methods on multimodal datasets: (a1–a5) LTFL; (b1–b5) INLPG; (c1–c5) GBF; (d1–d5) IRG-McS; (e1–e5) SCASC; (f1–f5) SRGCAE; (g1–g5) GSGM.



**Figure 4.** ROC curves on (a) dataset #1, (b) dataset #2, (c) dataset #3, (d) dataset #4 and (e) dataset #5.

obtained through the cross-mapping of the graph. On the basis of Markovian co-segmentation, the construction of the KNN graph and change information are iteratively optimized.

- 5) SCASC (Sun et al. 2022): Sparse-constrained adaptive structure consistency (SCASC) is used to construct a regression model of the structure consistency image, and the prior sparse knowledge from the CIM is used to distinguish between changed and unchanged regions.
- 6) SRGCAE (H. Chen et al. 2022): Structural relationship graph convolutional autoencoder (SRGCAE) measures the similarity of multimodal image structural relationships to extract the changes.

### 3.4. Experiments

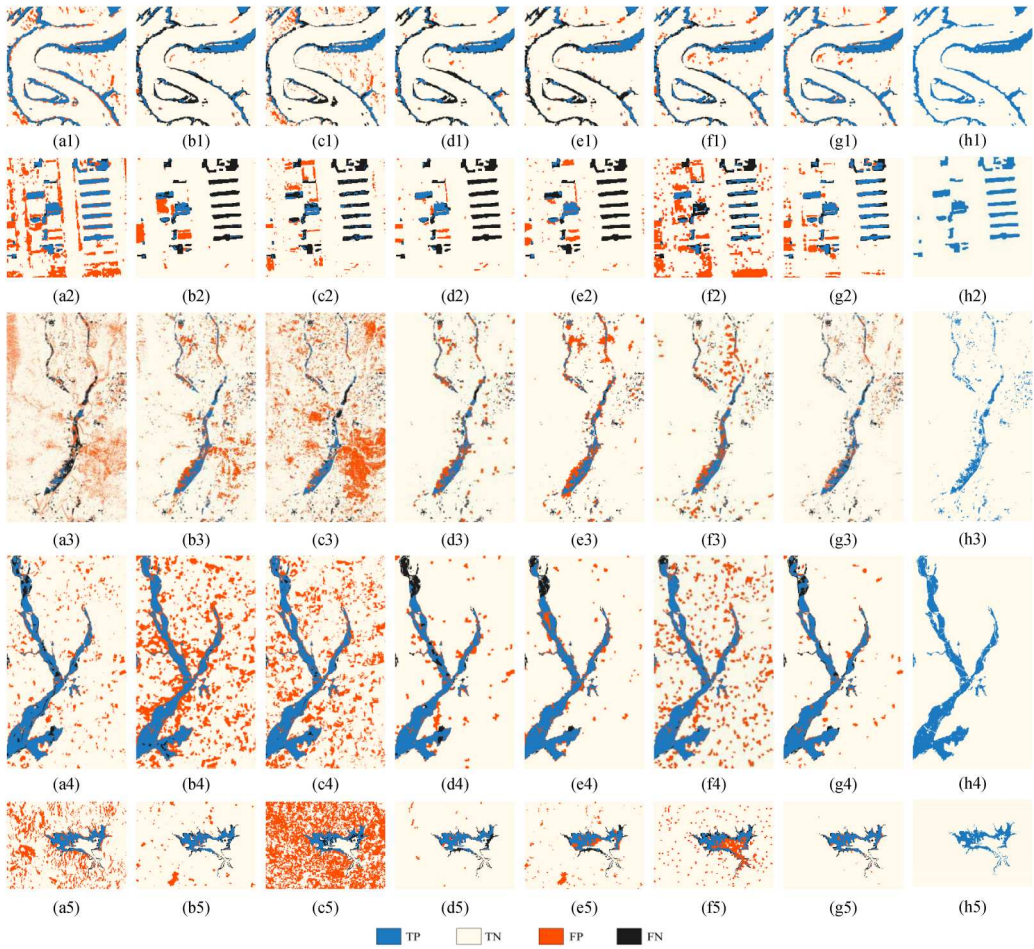
#### 3.4.1. Experiments on multimodal datasets

Figure 3 depicts the CIMs of various methods on multimodal datasets #1–5. The comparison methods in dataset #1 generally highlight the changed regions, but the intensity of changes varies significantly across the various changed regions. Among the comparative methods, except for SRGCAE, all other methods failed to detect changes in the right half of dataset #2. The CIMs of the other comparison methods did a better job of displaying the changed regions in dataset #3 than INLPG and GBF, although there are some gaps in the changed regions. The CI difference between the changed and unchanged regions of the CIMs of INLPG in dataset #4 is not significant. The other comparison methods can almost completely highlight the regions that have changed in dataset #4, but they have more ‘pseudo-change’ patches in the unchanged regions. IRG-MCS and SCASC show good CIMs in dataset #5, but some FPs remain. GSGM can effectively reflect the structural changes of multimodal data by constructing the GSG. The ROC curves of the CIMs for different methods on datasets #1–#5 are plotted in Figure 4, and it can be seen that the proposed GSGM achieves optimal ROC curves. Therefore, the CIMs of GSGM on datasets #1–#5 can effectively highlight the changed regions while preserving their internal integrity, as well as having the fewest ‘pseudo-change’ patches in the unchanged regions.

The CIMs of different methods on the multimodal datasets are shown in Figure 5. In dataset #1, LTFL has more FPs due to the presence of shadow differences in the land parts of the images. INLPG, GBF, IRG-MCS, SCASC, and SRGCAE overcome the effects of image shadowing and have fewer FPs, but there are large

numbers of FNs. GSGM could obtain the most complete changed regions while suppressing image shadows. Dataset #2 mostly captures changes in an urban environment. None of the other comparison methods, excluding LTFL, were able to identify the changes in the right half of the region in dataset #2, and GSGM has fewer FPs than LTFL. Datasets #3 and #4 show changes as a result of flooding. Visually, LTFL, INLPG, and GBF all show significant FPs, whereas IRG-MCS, SCASC, and SRGCAE have comparatively fewer FPs, and GSGM comes closest to the reference CM overall. The challenges with dataset #5 are that the changed regions are rather minor, as well as the shadowing of the two images. In the land region, LTFL, INLPG, and GBF yield more FPs, while IRG-MCS, SCASC, and SRGCAE produce comparatively fewer FPs. The proposed GSGM has the fewest FPs and detects the most complete changed regions. The accuracy assessments of the





**Figure 5.** CMs of different methods on multimodal datasets: (a1–a5) LTFL; (b1–b5) INLPG; (c1–c5) GBF; (d1–d5) IRG-McS; (e1–e5) SCASC; (f1–f5) SRGCAE; (g1–g5) GSGM; (h1–h5) reference CM.

comparison methods on multimodal datasets #1–#3 and #4–#5 are shown in [Tables 3](#) and [4](#) respectively, with the greatest accuracy denoted in bold. It is clear that GSGM obtains the highest AUC, OA, KC, and F1 values across all multimodal datasets, demonstrating its efficiency in identifying changes across multimodal images based on global structural features.

### 3.4.2. Experiments on unimodal datasets

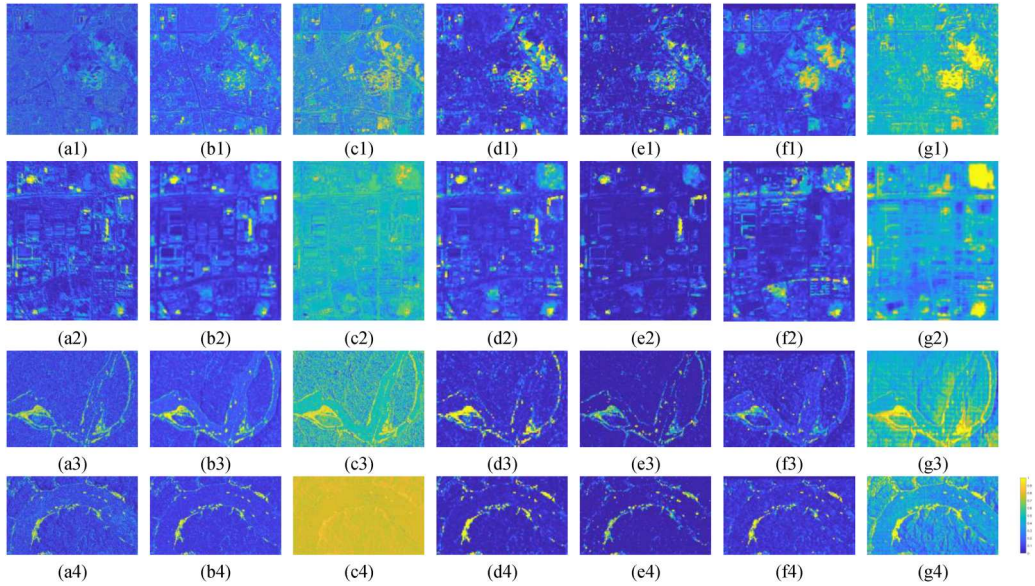
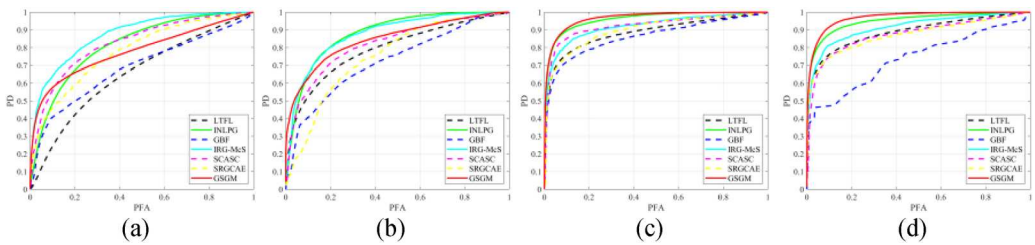
Four unimodal datasets were used to confirm the effectiveness of the proposed method on unimodal datasets. The CIMs of different methods on the unimodal datasets are shown in [Figure 6](#). In dataset #6, LTFL fails to highlight the changed regions, and GBF highlights the unchanged regions, although it also highlights the changed regions. The other comparison methods have lower CI in the unchanged regions, but the highlighted changed regions are not internally connected. In dataset #7, GBF does not clearly highlight the changed regions, while the other comparison methods only partially highlight the changed regions, and only GSGM completely highlights the changed region in the upper right corner of the image. Although the other methods could basically highlight the changed regions in datasets #8 and #9, there are some voids inside the changed regions, and GBF fails to highlight the changed regions in these datasets. As illustrated in [Figure 6](#) (f1–f4), the changed regions can be desirably highlighted by GSGM with the best internal connectivity. [Figure 7](#)

**Table 3.** Quantitative measures of binary CMs on multimodal datasets. The bolded font indicates the highest value.

Methods	Dataset #1			Dataset #2			Dataset #3					
	AUC	OA	KC	F1	AUC	OA	KC	F1	AUC	OA	KC	F1
LTL	0.9204	0.9204	0.7119	0.7579	0.7799	0.8182	0.4329	0.5210	0.5801	0.8504	0.0631	0.1240
INLPG	0.9590	0.9083	0.5635	0.6092	0.8916	0.8778	0.2093	0.2695	0.9240	0.9059	0.3730	0.4128
GBF	0.7881	0.8989	0.5530	0.6109	0.6502	0.8686	0.1984	0.2669	0.6710	0.7915	0.1093	0.1737
IRG-McS	0.9307	0.9128	0.6022	0.6483	0.8732	0.8933	0.2460	0.2916	0.8950	0.9469	0.4703	0.4975
SCASC	0.8668	0.8955	0.5069	0.5599	0.7842	0.8803	0.2083	0.2656	0.9168	0.9381	0.4585	0.4888
SRGCAE	0.9157	0.9223	0.6812	0.7259	0.6931	0.7835	0.1696	0.2854	0.8495	0.9376	0.4246	0.4557
GSGM	<b>0.9815</b>	<b>0.9472</b>	<b>0.7988</b>	<b>0.8301</b>	<b>0.9339</b>	<b>0.9118</b>	<b>0.5715</b>	<b>0.6212</b>	<b>0.9480</b>	<b>0.9571</b>	<b>0.5195</b>	<b>0.5419</b>

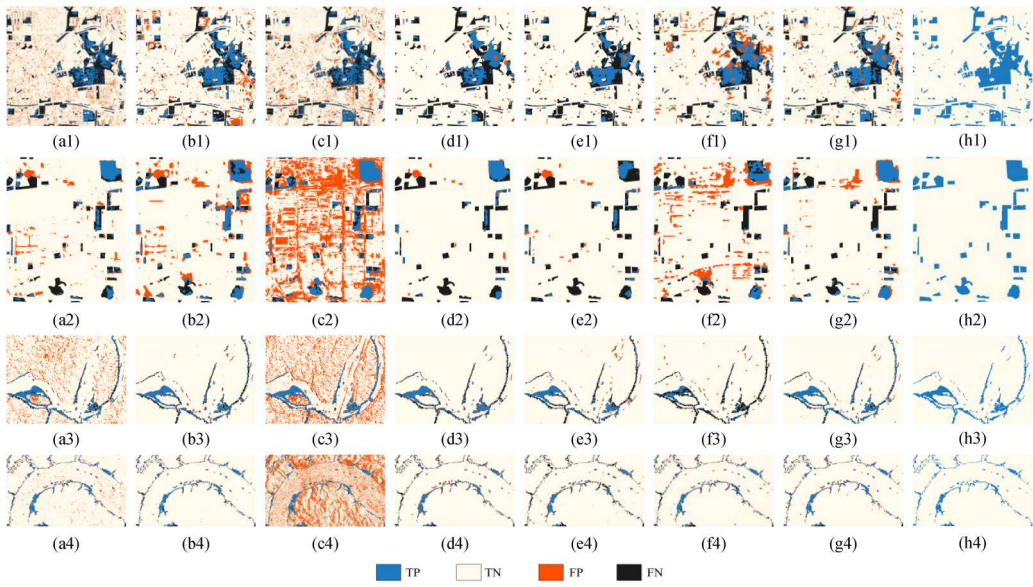
**Table 4.** Quantitative measures of binary CMs on multimodal datasets. The bolded font indicates the highest value.

Methods	Dataset #4				Dataset #5			
	AUC	OA	KC	F1	AUC	OA	KC	F1
LTFL	0.9322	0.9195	0.6689	0.7145	0.8754	0.8078	0.2849	0.3549
INLPG	0.9121	0.7829	0.4134	0.5162	0.9270	0.9503	0.6128	0.6392
GBF	0.9388	0.8272	0.4827	0.5695	0.6484	0.4667	0.0344	0.1418
IRG-McS	0.9427	0.9358	0.7136	0.7502	0.8883	0.9700	0.7242	0.7401
SCASC	0.9394	0.9503	0.7762	0.8046	0.8889	0.9472	0.5958	0.6238
SRGCAE	0.9266	0.8728	0.5724	0.6400	0.9174	0.9129	0.4737	0.5164
GSGM	<b>0.9806</b>	<b>0.9565</b>	<b>0.7937</b>	<b>0.8184</b>	<b>0.9716</b>	<b>0.9762</b>	<b>0.7860</b>	<b>0.7986</b>

**Figure 6.** CIMs of different methods on unimodal datasets: (a1–a4) LTFL; (b1–b4) INLPG; (c1–c4) GBF; (d1–d4) IRG-McS; (e1–e4) SCASC; (f1–f4) SRGCAE; (g1–g4) GSGM.**Figure 7.** ROC curves on (a) dataset #6, (b) dataset #7, (c) dataset #8 and (d) dataset #9.

shows the ROC curves of the CIMs of different methods on the unimodal datasets, where it can be seen that the proposed GSGM achieves approximate results with the comparison methods in dataset #7 and optimal ROC curves in datasets #8 and #9.

**Figure 8** illustrates the CIMs of different methods on the unimodal datasets. Datasets #6 and #7 reflect the changes caused by urban construction with complex feature types, such as buildings, vegetation, roads, etc. Visually, in dataset #6, LTFL and INLPG have more FPs, while GBF, IRG-McS, SCASC, and SRGCAE have fewer FPs, but generate a large number of FNs. In dataset #7, GBF,



**Figure 8.** CMs of different methods on unimodal datasets: (a1–a4) LTFL; (b1–b4) INLPG; (c1–c4) GBF; (d1–d4) IRG-McS; (e1–e4) SCASC; (f1–f4) SRGCAE; (g1–g4) GSGM; (h1–h4) reference CM.

**Table 5.** Quantitative measures of binary CMs on unimodal datasets. The bolded font indicates the highest value.

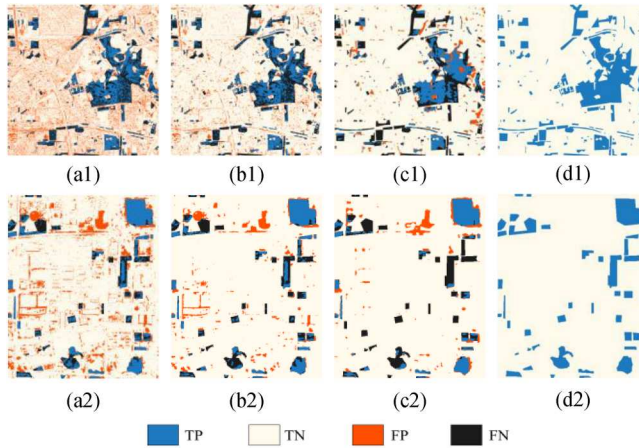
Method	Dataset #6				Dataset #7			
	AUC	OA	KC	F1	AUC	OA	KC	F1
LTFL	0.6520	0.8072	0.3255	0.4413	0.7917	0.9051	0.4092	0.4598
INLPG	0.8166	0.8398	0.4374	0.5335	<b>0.8802</b>	0.8851	0.4500	0.5138
GBF	0.6861	0.8127	0.3707	0.4852	0.7238	0.6240	0.1439	0.2845
IRG-McS	<b>0.8709</b>	<b>0.8697</b>	0.4307	0.4892	0.8757	0.9203	0.4024	0.4365
SCASC	0.8255	0.8659	0.4230	0.4855	0.8182	0.9090	0.3035	0.3407
SRGCAE	0.7787	0.8338	0.3759	0.4724	0.7528	0.8354	0.1958	0.2874
GSGM	0.7725	0.8633	<b>0.5154</b>	<b>0.5971</b>	0.8468	<b>0.9212</b>	<b>0.4929</b>	<b>0.5340</b>

LTFL, and INLPG have considerable FPs, and IRG-MCS and SCASC have more FNs. In comparison to these methods, the proposed GSGM has the most comprehensive detection results. The difficulty of datasets #8 and #9 lies in the fact that the changed regions are relatively minor and there is intense noise in the images. LTFL and GBF do not effectively suppress the image noise, and generate more FPs. INLPG, IRG-MCS, SCASC, and SRGCAE overcome the effects of image noise to a certain extent, but fail to detect some minor changes. Visually, the proposed GSGM is unaffected by image noise and can detect minor changes. Owing to the change metric of the global structural features of the image, the CIMs of the proposed GSGM can more accurately distinguish

**Table 6.** Quantitative measures of binary CMs on the unimodal datasets. The bolded font indicates the highest value.

Method	Dataset #8				Dataset #9			
	AUC	OA	KC	F1	AUC	OA	KC	F1
LTFL	0.8848	0.8774	0.3891	0.4506	0.8894	0.9550	0.6188	0.6428
INLPG	0.9594	0.9583	0.6508	0.6720	0.9498	0.9659	0.6448	0.6621
GBF	0.8615	0.7826	0.2946	0.3789	0.7372	0.6636	0.1185	0.2087
IRG-McS	0.9196	0.9556	0.6305	0.6533	0.9141	0.9591	0.5579	0.5782
SCASC	0.9008	0.9513	0.5881	0.6128	0.8486	0.9528	0.4745	0.4974
SRGCAE	0.8984	0.9350	0.3187	0.3422	0.8708	0.9592	0.5212	0.5399
GSGM	<b>0.9664</b>	<b>0.9603</b>	<b>0.7137</b>	<b>0.7351</b>	<b>0.9689</b>	<b>0.9687</b>	<b>0.7018</b>	<b>0.7183</b>





**Figure 9.** The first and second rows correspond to the CMs of dataset #6 and dataset #7, respectively. (a1-a2) CVA; (b1-b2) DCVA; (c1-c2) GSGM; (d1-d2) reference CM.

**Table 7.** Quantitative measures of binary CMs on unimodal dataset #6 and dataset #7. The bolded font indicates the highest value.

Method	Dataset #6			Dataset #7		
	OA	KC	F1	OA	KC	F1
CVA	0.7650	0.3168	0.4609	0.8558	0.3999	0.4770
DCVA	0.8394	0.4263	0.5224	0.9093	<b>0.4960</b>	<b>0.5464</b>
GSGM	<b>0.8633</b>	<b>0.5154</b>	<b>0.5971</b>	<b>0.9212</b>	0.4929	0.5340

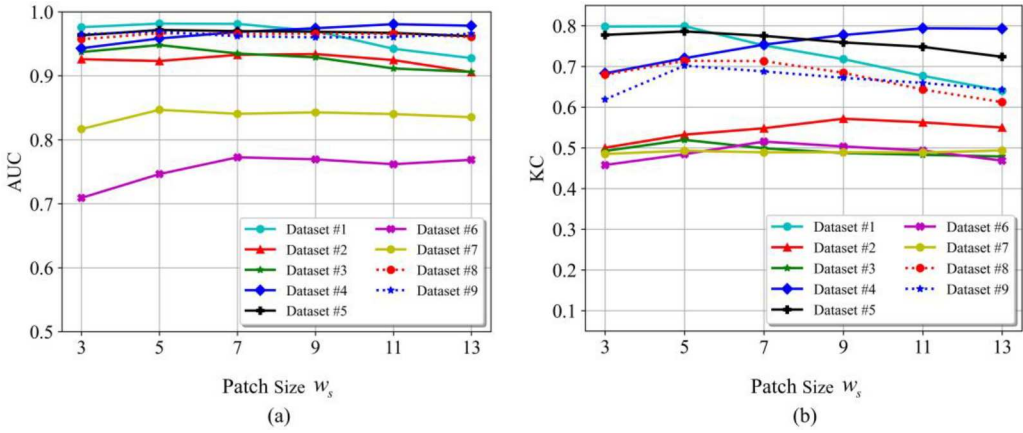
the changed and unchanged regions, while the CD result of GSGM is closest to the reference CM. From the accuracy evaluation of CMs on the unimodal datasets in Tables 5 and 6, it can be seen that the KC and F1 of GSGM are the highest among these five comparison methods, indicating that GSGM is applicable not only to MCD, but also to UCD.

To further validate the efficacy of the proposed GSGM on unimodal datasets #6 and #7, we utilized two established UCD methods: Change Vector Analysis (CVA) (Bovolo and Bruzzone 2007) and Deep Change Vector Analysis (DCVA) (Saha, Bovolo, and Bruzzone 2019). As depicted in Figure 9, the results from both datasets show that CMs of CVA are prone to significant salt-and-pepper noise, a consequence of its pixel-centric approach to CD. In contrast, DCVA, which incorporates deep features from the imagery, is more effective at reducing such noise. The precision evaluation detailed in Table 7 reveals that DCVA not only surpasses CVA in accuracy but also equals the performance of GSGM on dataset #7. However, on dataset #6, DCVA's performance is slightly inferior to GSGM, underscoring the robustness of the proposed GSGM across different unimodal datasets.

### 3.5. Discussion

#### 3.5.1. Experiment setting and parameter analysis

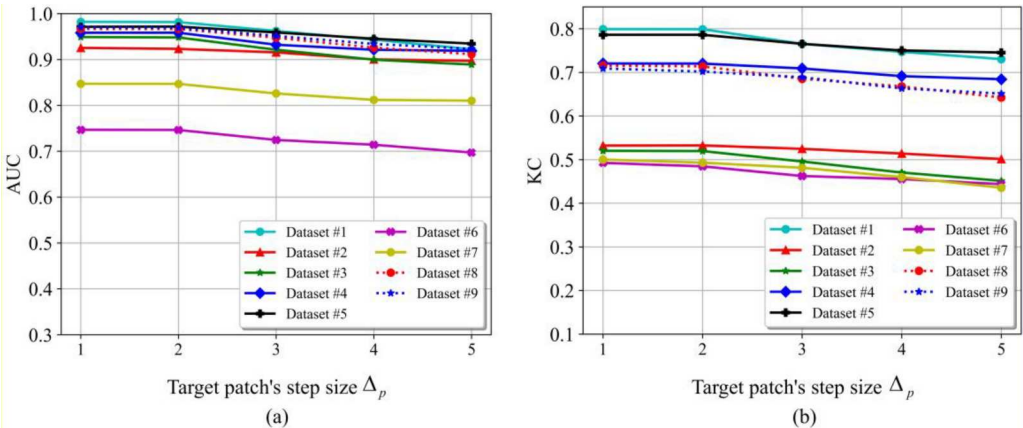
Based on the ROC curves, threshold parameter  $\zeta$  is set between 1.3 and 2.0 for all datasets to obtain satisfactory CMs. Step coefficient  $\alpha_{\Delta_s}$ , patch size  $w_s$  and target patch's step size  $\Delta_p$  are the main parameters of GSGM. Step coefficient  $\alpha_{\Delta_v}$  controls the sparsity of the vertex patches; the greater the value of  $\alpha_{\Delta_v}$ , the more numerous the vertex patches. Considering the accuracy and efficiency of the method, this paper fixes  $\alpha_{\Delta_v}$  to 0.1, and discusses the influence of patch size  $w_s$  and target patch's step size  $\Delta_p$  on the accuracy of GSG.



**Figure 10.** Influence of patch size  $w_s$  on GSGM performance: (a) AUC- $w_s$  curves; (b) KC- $w_s$  curves.

1) *Patch size  $w_s$ .* Weight coefficient  $\lambda$  is fixed to 2, and patch size  $w_s$  is set from 3 to 13 with an interval of 2. Figure 10 shows the influence of patch size  $w_s$  on GSGM. It can be seen that datasets #1, #3, #5, #7, #8, and #9 reach the maximum AUC and KC when  $w_s$  is set to 5, and datasets #2, #4, and #6 reach the maximum AUC and KC when  $w_s$  is set to 9, 11, and 7 respectively. This is because datasets #1, #3, #5, #7, #8, and #9 contain relatively small changed regions, so the patch size needs to be set smaller to ensure the detection of small changed regions, while the changed regions of datasets #2, #4, and #6 are relatively large, so the patch size needs to be set larger to ensure the complete detection of changed regions. Therefore, this paper suggests that when the changed regions are relatively small, patch size  $w_s$  should be set to 5, and when the changed regions are relatively large, patch size  $w_s$  should be set to a larger value (e.g. 9 or 11).

2) *Target patch's step size  $\Delta_p$ .* To evaluate the influence of the target patch's step size  $\Delta_p$  on the performance of the proposed GSGM, we maintained the patch size  $w_s$  at a constant value of 5 and systematically adjusted  $\Delta_p$  from 1 to 5. Figure 11 shows that the GSGM's accuracy is maintained at a high level for  $\Delta_p$  values of 1 and 2. However, there is a discernible decrease in accuracy as  $\Delta_p$  continues to increase. This decline is attributed to the sparser distribution of image patches and the consequent reduction in the robustness of the extracted structural features with larger  $\Delta_p$  values.



**Figure 11.** Influence of target patch's step size  $\Delta_p$  on GSGM performance: (a) AUC- $\Delta_p$  curves; (b) KC- $\Delta_p$  curves.

**Table 8.** Computational time (seconds) of the proposed GSGM on datasets #3 and #5 under different values of  $\Delta_p$ .

Datasets	Size	$\Delta_p = 1$	$\Delta_p = 2$	$\Delta_p = 3$	$\Delta_p = 4$	$\Delta_p = 5$
Dataset #3	$3,500 \times 2,000 \times 11(3)$	10208.53	2550.78	1233.76	831.25	663.65
Dataset #5	$300 \times 412 \times 1(3)$	317.23	76.19	36.35	23.12	16.44

**Table 9.** Quantitative measures of different graphs on datasets #1–#9.

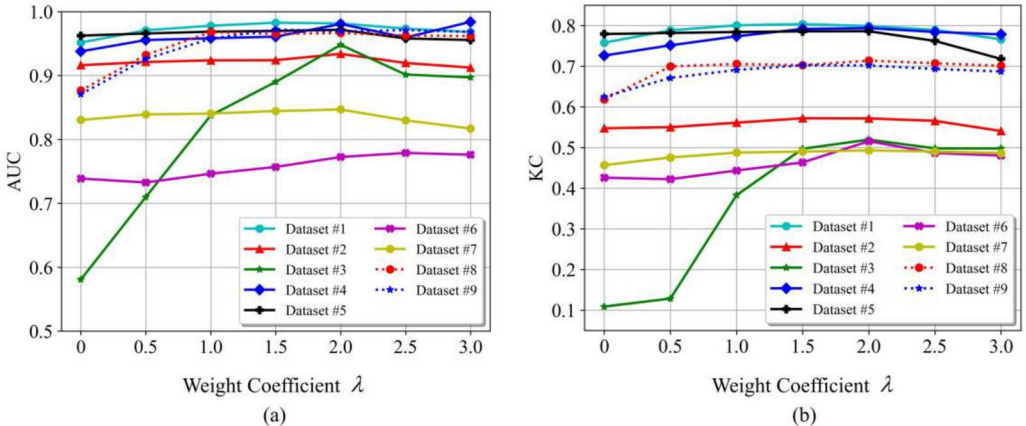
GSG	#1	#2	#3	#4	#5	#6	#7	#8	#9
AUC	0.9734	0.9233	0.9248	0.9745	0.9703	0.7567	0.7997	0.9031	0.8653
✓	0.9815	0.9339	0.9480	0.9806	0.9716	0.7725	0.8468	0.9664	0.9689
KC	0.7844	0.5672	0.4865	0.7606	0.7569	0.4210	0.4443	0.5446	0.5475
✓	0.7988	0.5715	0.5195	0.7937	0.7860	0.5154	0.4929	0.7137	0.7018

On the other hand, a  $\Delta_p$  value that is too small leads to an overabundance of image patches, which in turn amplifies the computational burden of the algorithm. Given these considerations, this study recommends setting  $\Delta_p$  equal to  $(w_s - 1)/2$ , which simplifies the process and optimizes computational efficiency without compromising accuracy.

3) *Computation time.* To illustrate the computational efficiency of the proposed GSGM, we fix the patch size  $w_s$  at 5, and use datasets #3 and #5, which represent the largest and smallest in size, respectively (refer to Table 7). Our implementation of GSGM utilizes MATLAB 2020a on a Windows desktop with an AMD Ryzen 7 3800X 8-Core Processor (3.89 GHz) and 64 GB of RAM. As demonstrated in Table 8, GSGM's computational time is inversely proportional to the increase in the step size parameter  $\Delta_p$ . Figure 11 further confirms that GSGM maintains significant accuracy when  $\Delta_p$  ranges from 1 to 4. To optimize for reduced computation time without compromising accuracy, it is recommended to set a larger  $\Delta_p$  value. Moreover, computational efficiency can be enhanced by refining the graph construction process, such as improving the efficiency of similarity measurements between image patches, or by leveraging parallel computing to distribute tasks across multiple cores or nodes.

### 3.5.2. Ablation study

The proposed GSGM consists of three main tasks: GSG construction, CI measure, and CIMs fusion. The effectiveness of the core components of these three tasks (GSG, weight change metrics (WCM), and LLRR) will be discussed separately.

**Figure 12.** Influence of weight coefficient  $\lambda$  on GSGM performance; (a) AUC- $\lambda$  curves; (b) KC- $\lambda$  curves.



**Table 10.** Effectiveness of CI measurement.

	$dif_{(i,j)}^1$	$dif_{(i,j)}^2$	#1	#2	#3	#4	#5	#6	#7	#8	#9
AUC		✓	0.9603	0.9126	0.9046	0.9601	0.9585	0.7412	0.8253	0.9624	0.9548
	✓		0.9688	0.9235	0.8998	0.9652	0.9692	0.7510	0.8163	0.9643	0.9606
KC	✓	✓	0.9815	0.9339	0.9480	0.9806	0.9716	0.7725	0.8468	0.9664	0.9689
		✓	0.7619	0.5211	0.4788	0.7512	0.7583	0.4723	0.4615	0.6989	0.6812
	✓		0.7694	0.5488	0.4756	0.7567	0.7821	0.4812	0.4527	0.7001	0.6885
	✓	✓	0.7988	0.5715	0.5195	0.7937	0.7860	0.5154	0.4929	0.7137	0.7018

**Table 11.** Quantitative measures of different DIs fusion methods on datasets #1–#9.

	MWF	WF	LLRR	#1	#2	#3	#4	#5	#6	#7	#8	#9
AUC	✓			0.9708	0.9209	0.8932	0.9716	0.9559	0.7581	0.8304	0.9658	0.9678
		✓		0.9709	0.9215	0.8971	0.9718	0.9558	0.7583	0.8306	0.9659	0.9680
KC			✓	0.9815	0.9339	0.9480	0.9806	0.9716	0.7725	0.8468	0.9664	0.9689
	✓			0.7972	0.5207	0.4924	0.7676	0.7626	0.4602	0.4797	0.7068	0.6788
		✓		0.7976	0.5208	0.4964	0.7682	0.7627	0.4611	0.4804	0.7080	0.6798
			✓	0.7988	0.5715	0.5195	0.7937	0.7860	0.5154	0.4929	0.7137	0.7018

1) *The effectiveness of GSG.* To validate the efficacy of the GSG, the top 10% of vertex patches most similar to each target patch are utilized to construct the graph, following INLPG (Sun et al. 2022). The results in Table 9 demonstrate that GSG attained the highest AUC and KC values across all datasets, with AUC increasing by 3.58% and KC by 11.79% on average respectively. This proves that GSG could better express the structural features of the images.

2) *The effectiveness of WCM.* Weight coefficient  $\lambda$  affects the contribution of each vertex patch to the CI of the target patch, and weight coefficient  $\lambda$  is set to 0, 0.5, 1, 1.5, 2, 2.5, and 3 respectively. For datasets #1, #3, #5, #7, #8, and #9, patch size  $w_s$  is set to 5, and for datasets #2, #4, and #6, patch size  $w_s$  is set to 9, 11, and 7 respectively. Figure 12 shows the influence of weight coefficient  $\lambda$  on GSGM. As can be seen from Figure 11, most datasets reach the best AUC and KC values when weight coefficient  $\lambda$  is 2.0; only some datasets (#1 and #2) achieve the maximum AUC and KC values when  $\lambda$  is 1.5; and when weight coefficient  $\lambda$  is set to 0, all datasets obtain lower accuracy. This indicates the effectiveness of the WCM. For simplicity, this paper suggests setting weight coefficient  $\lambda$  to 2.0.

3) *The effectiveness of the CI measure.* In this paper, we explored two distinct methods for measuring changes: the Intrinsic change of the GSG (referred to as  $dif_{(i,j)}^1$ ) and the difference between the mapped GSG and the GSG (referred to as  $dif_{(i,j)}^2$ ). To assess the individual impacts of  $dif_{(i,j)}^1$  and  $dif_{(i,j)}^2$ , we conducted two separate experiments, each focusing on one method for analyzing changes in multimodal imagery. The outcomes of these ablation experiments are detailed in Table 10. Table 10 reveals that utilizing both  $dif_{(i,j)}^1$  and  $dif_{(i,j)}^2$  together, rather than relying solely on either  $dif_{(i,j)}^1$  or  $dif_{(i,j)}^2$ , led to a notable enhancement in performance metrics. Specifically, there is an average increase of 2.38% and 2.14% of AUC, and 6.67% and 5.3% of KC, respectively. These findings suggest that the combined use of GSG's Intrinsic change and the difference between mapped GSG and GSG significantly improves the representation of change information and enhances the precision of CD.

4) *The effectiveness of LLRR.* For CIMs fusion, mean weighted fusion (MWF) and wavelet fusion (Pu 2000) are utilized as controls. As evident in Table 11, the proposed GSGM continued to achieve optimal results, elevating the AUC values by 2.80% and 1.63% while improving KC values by 4.56% and 4.37% over MWF and WF respectively. These findings validate that the proposed method can effectively enhance the representation of change information and improve detection accuracy through fusing the low-rank and significance components of the CIMs.

## 4. Conclusion

In order to address the issue of ‘incomparability’ caused by the large imaging differences of multimodal images, this paper proposes a global structure graph mapping (GSGM)-based MCD method. GSGM constructs the GSGs of multimodal images to represent the structural information, then compares them by cross-mapping the GSGs into the same image domain. A new weighted change metric formula is built to obtain robust change information, the forward and backward CIMs are fused using low-rank representation to obtain a robust CIM, and the CM are obtained through threshold segmentation. Experiments on five multimodal datasets and four unimodal datasets demonstrate the superiority and robustness of the proposed method.

The proposed GSGM has certain limitations. Specifically, there is room for the enrichment and diversification of the data types currently used. Furthermore, the method’s detection accuracy requires further improvement, especially for complex change scenes (such as datasets #3 and #6). In future research, we will focus on developing the diversity of evaluated datasets to enhance the CD capabilities across various image types. For example, we will continue to consider the polarization mode and image quality of SAR imagery (including scalloping and speckle effects), and their impact on image structural features. Building upon this, we will take into account the physical characteristics of SAR imagery to enhance its applicability to MCD tasks within complex scenarios. In order to further reduce the impact of imaging conditions (such as atmospheric effects, thermal noise, and illumination) on the quality of remote sensing images, and to enhance the accuracy of constructing GSGs, we will consider further improvements in the quality of remote sensing images in future research (such as spectral calibration for hyperspectral imagery (Hong et al. 2019)). Meanwhile, refining graph construction to improve the representation of structural image features also holds promise for strengthening the method’s versatility.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This study was supported by the National Natural Science Foundation of China (Grant 41971313); National Natural Science Foundation of China (Grant 42271411); Scientific Research Innovation Project for Graduate Students in Hunan Province (No. CX20220169); and Research Project on Monitoring and Early Warning Technologies for the Implementation of Land Use Planning in Guangzhou City (2020B0101130009).

## Data availability statement

Datasets #1–#2 and #8–#9 that support this study are available at [<https://github.com/rshante0426/MCD-datasets>]; datasets #3–#5 that support this study are available at [<https://www-labs.iro.umontreal.ca/~mignotte/>]; and datasets #6–#7 that support this study are available at [<https://github.com/MinZHANG-WHU/FDCNN?tab=readme-ov-file>].

## References

- Alberga, Vito. 2009. “Similarity Measures of Remotely Sensed Multi-Sensor Images for Change Detection Applications.” *Remote Sensing* 1 (3): 122–143. <https://doi.org/10.3390/rs1030122>.
- Bovolo, Francesca, and Lorenzo Bruzzone. 2007. “A Theoretical Framework for Unsupervised Change Detection Based on Change Vector Analysis in the Polar Domain.” *IEEE Transactions on Geoscience and Remote Sensing* 45 (1): 218–236. <https://doi.org/10.1109/TGRS.2006.885408>.
- Brunner, Dominik, Guido Lemoine, and Lorenzo Bruzzone. 2010. “Earthquake Damage Assessment of Buildings Using VHR Optical and SAR Imagery.” *IEEE Transactions on Geoscience and Remote Sensing* 48 (5): 2403–2420. <https://doi.org/10.1109/TGRS.2009.2038274>.

- Chen, Yuzeng, Yuqi Tang, Te Han, Yuwei Zhang, Bin Zou, and Huihui Feng. 2022. "RAMC: A Rotation Adaptive Tracker with Motion Constraint for Satellite Video Single-Object Tracking." *Remote Sensing* 14 (13): 3108. <https://doi.org/10.3390/rs14133108>.
- Chen, Yuzeng, Yuqi Tang, Yi Xiao, Qiangqiang Yuan, Yuwei Zhang, Fengqing Liu, Jiang He, and Liangpei Zhang. 2024. "Satellite Video Single Object Tracking: A Systematic Review and an Oriented Object Tracking Benchmark." *ISPRS Journal of Photogrammetry and Remote Sensing* 210 (April): 212–240. <https://doi.org/10.1016/j.isprsjrs.2024.03.013>.
- Chen, Yuzeng, Yuqi Tang, Zhiyong Yin, Te Han, Bin Zou, and Huihui Feng. 2022. "Single Object Tracking in Satellite Videos: A Correlation Filter-Based Dual-Flow Tracker." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15: 6687–6698. <https://doi.org/10.1109/JSTARS.2022.3185328>.
- Chen, Hongruixuan, Naoto Yokoya, Chen Wu, and Bo Du. 2022. "Unsupervised Multimodal Change Detection Based on Structural Relationship Graph Representation Learning." *IEEE Transactions on Geoscience and Remote Sensing* 60: 1–18. <https://doi.org/10.1109/TGRS.2022.3229027>.
- Fan, Qiufeng, Fanbo Hou, and Feng Shi. 2021. "A Fusion Method for Infrared and Visible Images Based on Iterative Guided Filtering and Two Channel Adaptive Pulse Coupled Neural Network." *International Journal of Image and Data Fusion* 12 (1): 23–47. <https://doi.org/10.1080/19479832.2020.1814877>.
- Gong, Maoguo, Xudong Niu, Tao Zhan, and Mingyang Zhang. 2019. "A Coupling Translation Network for Change Detection in Heterogeneous Images." *International Journal of Remote Sensing* 40 (9): 3647–3672. <https://doi.org/10.1080/01431161.2018.1547934>.
- Gong, Maoguo, Xudong Niu, Puzhao Zhang, and Zhetao Li. 2017. "Generative Adversarial Networks for Change Detection in Multispectral Imagery." *IEEE Geoscience and Remote Sensing Letters* 14 (12): 2310–2314. <https://doi.org/10.1109/LGRS.2017.2762694>.
- Han, Te, Yuqi Tang, and Yuzeng Chen. 2022. "Heterogeneous Image Change Detection Based on Two-Stage Joint Feature Learning." In *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 3215–3218. <https://doi.org/10.1109/IGARSS46834.2022.9883323>.
- Han, Te, Yuqi Tang, Xin Yang, Zefeng Lin, Bin Zou, and Huihui Feng. 2021. "Change Detection for Heterogeneous Remote Sensing Images with Improved Training of Hierarchical Extreme Learning Machine (HELM)." *Remote Sensing* 13 (23): 4918. <https://doi.org/10.3390/rs13234918>.
- Han, Te, Yuqi Tang, Bin Zou, and Huihui Feng. 2024. "Unsupervised Multimodal Change Detection Based on Adaptive Optimization of Structured Graph." *International Journal of Applied Earth Observation and Geoinformation* 126 (February): 103630. <https://doi.org/10.1016/j.jag.2023.103630>.
- Hong, Danfeng, Jing Yao, Chenyu Li, Deyu Meng, Naoto Yokoya, and Jocelyn Chanussot. 2023. "Decoupled-and-Coupled Networks: Self-Supervised Hyperspectral Image Super-Resolution With Subpixel Fusion." *IEEE Transactions on Geoscience and Remote Sensing* 61: 1–12. <https://doi.org/10.1109/TGRS.2023.3324497>.
- Hong, Danfeng, Naoto Yokoya, Jocelyn Chanussot, and Xiao Xiang Zhu. 2019. "An Augmented Linear Mixing Model to Address Spectral Variability for Hyperspectral Unmixing." *IEEE Transactions on Image Processing* 28 (4): 1923–1938. <https://doi.org/10.1109/TIP.2018.2878958>.
- Hong, Danfeng, Bing Zhang, Xuyang Li, Yuxuan Li, Chenyu Li, Jing Yao, Naoto Yokoya, et al. 2024. "SpectralGPT: Spectral Remote Sensing Foundation Model". <https://doi.org/10.1109/TPAMI.2024.3362475>.
- Hong, Danfeng, Bing Zhang, Hao Li, Yuxuan Li, Jing Yao, Chenyu Li, Martin Werner, Jocelyn Chanussot, Alexander Zipf, and Xiao Xiang Zhu. 2023. "Cross-City Matters: A Multimodal Remote Sensing Benchmark Dataset for Cross-City Semantic Segmentation Using High-Resolution Domain Adaptation Networks." *Remote Sensing of Environment* 299 (December): 113856. <https://doi.org/10.1016/j.rse.2023.113856>.
- Hou, Bin, Yunhong Wang, and Qingjie Liu. 2017. "Change Detection Based on Deep Features and Low Rank." *IEEE Geoscience and Remote Sensing Letters* 14 (12): 2418–2422. <https://doi.org/10.1109/LGRS.2017.2766840>.
- Jimenez-Sierra, David Alejandro, Hernán Darío Benítez-Restrepo, Hernán Darío Vargas-Cardona, and Jocelyn Chanussot. 2020. "Graph-Based Data Fusion Applied To: Change Detection and Biomass Estimation in Rice Crops." *Remote Sensing* 12 (17): 2683. <https://doi.org/10.3390/rs12172683>.
- Liu, Jia, Maoguo Gong, Kai Qin, and Puzhao Zhang. 2018. "A Deep Convolutional Coupling Network for Change Detection Based on Heterogeneous Optical and Radar Images." *IEEE Transactions on Neural Networks and Learning Systems* 29 (3): 545–559. <https://doi.org/10.1109/TNNLS.2016.2636227>.
- Liu, Zhunga, Gang Li, Gregoire Mercier, You He, and Quan Pan. 2018. "Change Detection in Heterogeneous Remote Sensing Images via Homogeneous Pixel Transformation." *IEEE Transactions on Image Processing* 27 (4): 1822–1834. <https://doi.org/10.1109/TIP.2017.2784560>.
- Liu, Guangcan, and Shuicheng Yan. 2011. "Latent Low-Rank Representation for Subspace Segmentation and Feature Extraction." In 2011 *International Conference on Computer Vision*, 1615–1622. Barcelona, Spain: IEEE. <https://doi.org/10.1109/ICCV.2011.6126422>.

- Luppino, Luigi Tommaso, Filippo Maria Bianchi, Gabriele Moser, and Stian Normann Anfinsen. 2019. "Unsupervised Image Regression for Heterogeneous Change Detection." *IEEE Transactions on Geoscience and Remote Sensing* 57 (12): 9960–9975. <https://doi.org/10.1109/TGRS.2019.2930348>.
- Luppino, Luigi Tommaso, Michael Kampffmeyer, Filippo Maria Bianchi, Gabriele Moser, Sebastiano Bruno Serpico, Robert Jenssen, and Stian Normann Anfinsen. 2022. "Deep Image Translation With an Affinity-Based Change Prior for Unsupervised Multimodal Change Detection." *IEEE Transactions on Geoscience and Remote Sensing* 60: 1–22. <https://doi.org/10.1109/TGRS.2021.3056196>.
- Mercier, G., G. Moser, and S. B. Serpico. 2008. "Conditional Copulas for Change Detection in Heterogeneous Remote Sensing Images." *IEEE Transactions on Geoscience and Remote Sensing* 46 (5): 1428–1441. <https://doi.org/10.1109/TGRS.2008.916476>.
- Mignotte, Max. 2020. "A Fractal Projection and Markovian Segmentation-Based Approach for Multimodal Change Detection." *IEEE Transactions on Geoscience and Remote Sensing* 58 (11): 8046–8058. <https://doi.org/10.1109/TGRS.2020.2986239>.
- Prendes, Jorge, Marie Chabert, Frederic Pascal, Alain Giros, and Jean-Yves Tourneret. 2015. "A New Multivariate Statistical Model for Change Detection in Images Acquired by Homogeneous and Heterogeneous Sensors." *IEEE Transactions on Image Processing* 24 (3): 799–812. <https://doi.org/10.1109/TIP.2014.2387013>.
- Pu, Tian. 2000. "Contrast-Based Image Fusion Using the Discrete Wavelet Transform." *Optical Engineering* 39 (8): 2075. <https://doi.org/10.1117/1.1303728>.
- Rajakumar, C., and S. Satheeskumaran. 2022. "Singular Value Decomposition and Saliency - Map Based Image Fusion for Visible and Infrared Images." *International Journal of Image and Data Fusion* 13 (1): 21–43. <https://doi.org/10.1080/19479832.2020.1864786>.
- Reba, Meredith, and Karen C. Seto. 2020. "A Systematic Review and Assessment of Algorithms to Detect, Characterize, and Monitor Urban Land Change." *Remote Sensing of Environment* 242 (June): 111739. <https://doi.org/10.1016/j.rse.2020.111739>.
- Saha, Sudipan, Francesca Bovolo, and Lorenzo Bruzzone. 2019. "Unsupervised Deep Change Vector Analysis for Multiple-Change Detection in VHR Images." *IEEE Transactions on Geoscience and Remote Sensing* 57 (6): 3677–3693. <https://doi.org/10.1109/TGRS.2018.2886643>.
- Sun, Yuli, Lin Lei, Dongdong Guan, and Gangyao Kuang. 2021. "Iterative Robust Graph for Unsupervised Change Detection of Heterogeneous Remote Sensing Images." *IEEE Transactions on Image Processing* 30: 6277–6291. <https://doi.org/10.1109/TIP.2021.3093766>.
- Sun, Yuli, Lin Lei, Dongdong Guan, Ming Li, and Gangyao Kuang. 2022. "Sparse-Constrained Adaptive Structure Consistency-Based Unsupervised Image Regression for Heterogeneous Remote-Sensing Change Detection." *IEEE Transactions on Geoscience and Remote Sensing* 60: 1–14. <https://doi.org/10.1109/TGRS.2021.3110998>.
- Sun, Yuli, Lin Lei, Xiao Li, Hao Sun, and Gangyao Kuang. 2021. "Nonlocal Patch Similarity Based Heterogeneous Remote Sensing Change Detection." *Pattern Recognition* 109 (January): 107598. <https://doi.org/10.1016/j.patcog.2020.107598>.
- Sun, Yuli, Lin Lei, Xiao Li, Xiang Tan, and Gangyao Kuang. 2022. "Structure Consistency-Based Graph for Unsupervised Change Detection With Homogeneous and Heterogeneous Remote Sensing Images." *IEEE Transactions on Geoscience and Remote Sensing* 60: 1–21. <https://doi.org/10.1109/TGRS.2021.3053571>.
- Tang, Huakang, Honglei Wang, and Xiaoping Zhang. 2022. "Multi-Class Change Detection of Remote Sensing Images Based on Class Rebalancing." *International Journal of Digital Earth* 15 (1): 1377–1394. <https://doi.org/10.1080/17538947.2022.2108921>.
- Tang, Yuqi, Xin Yang, Te Han, Fangyan Zhang, Bin Zou, and Huihui Feng. 2024. "Enhanced Graph Structure Representation for Unsupervised Heterogeneous Change Detection." *Remote Sensing* 16 (4): 721. <https://doi.org/10.3390/rs16040721>.
- Tang, Yuqi, and Liangpei Zhang. 2017. "Urban Change Analysis with Multi-Sensor Multispectral Imagery." *Remote Sensing* 9 (3): 252. <https://doi.org/10.3390/rs9030252>.
- Touati, Redha, and Max Mignotte. 2018. "An Energy-Based Model Encoding Nonlocal Pairwise Pixel Interactions for Multisensor Change Detection." *IEEE Transactions on Geoscience and Remote Sensing* 56 (2): 1046–1058. <https://doi.org/10.1109/TGRS.2017.2758359>.
- Touati, Redha, Max Mignotte, and Mohamed Dahmane. 2019. "Multimodal Change Detection Using a Convolution Model-Based Mapping." In *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 1–6. Istanbul, Turkey: IEEE. <https://doi.org/10.1109/IPTA.2019.8936127>.
- Touati, Redha, Max Mignotte, and Mohamed Dahmane. 2020. "Multimodal Change Detection in Remote Sensing Images Using an Unsupervised Pixel Pairwise-Based Markov Random Field Model." *IEEE Transactions on Image Processing* 29: 757–767. <https://doi.org/10.1109/TIP.2019.2933747>.
- Wan, Ling, Yuming Xiang, and Hongjian You. 2019. "A Post-Classification Comparison Method for SAR and Optical Images Change Detection." *IEEE Geoscience and Remote Sensing Letters* 16 (7): 1026–1030. <https://doi.org/10.1109/LGRS.2019.2892432>.

- Wan, L., T. Zhang, and H. J. You. 2018. "Multi-Sensor Remote Sensing Image Change Detection Based on Sorted Histograms." *International Journal of Remote Sensing* 39 (11): 3753–3775. <https://doi.org/10.1080/01431161.2018.1448481>.
- Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. 2004. "Image Quality Assessment: From Error Visibility to Structural Similarity." *IEEE Transactions on Image Processing* 13 (4): 600–612. <https://doi.org/10.1109/TIP.2003.819861>.
- Wang, Decheng, Feng Zhao, Hui Yi, Yanan Li, and Xiangning Chen. 2022. "An Unsupervised Heterogeneous Change Detection Method Based on Image Translation Network and Post-Processing Algorithm." *International Journal of Digital Earth* 15 (1): 1056–1080. <https://doi.org/10.1080/17538947.2022.2092658>.
- Wei, Xuexin, Yang Liu, Lin Qi, Jilong Chen, Guoqin Wang, Linxiu Zhang, and Ronggao Liu. 2023. "Monitoring Forest Dynamics in Africa during 2000–2020 Using a Remotely Sensed Fractional Tree Cover Dataset." *International Journal of Digital Earth* 16 (1): 2212–2232. <https://doi.org/10.1080/17538947.2023.2220613>.
- Wu, Yue, Jiaheng Li, Yongzhe Yuan, A. K. Qin, Qi-Guang Miao, and Mao-Guo Gong. 2022. "Commonality Autoencoder: Learning Common Features for Change Detection from Heterogeneous Images." *IEEE Transactions on Neural Networks and Learning Systems* 33 (9): 4257–4270. <https://doi.org/10.1109/TNNLS.2021.3056238>.
- Zhan, Tao, Maoguo Gong, Xiangming Jiang, and Shuwei Li. 2018. "Log-Based Transformation Feature Learning for Change Detection in Heterogeneous Images." *IEEE Geoscience and Remote Sensing Letters* 15 (9): 1352–1356. <https://doi.org/10.1109/LGRS.2018.2843385>.
- Zhan, Tao, Maoguo Gong, Jia Liu, and Puzhao Zhang. 2018. "Iterative Feature Mapping Network for Detecting Multiple Changes in Multi-Source Remote Sensing Images." *ISPRS Journal of Photogrammetry and Remote Sensing* 146 (December): 38–51. <https://doi.org/10.1016/j.isprsjprs.2018.09.002>.
- Zhao, Wei, Zhirui Wang, Maoguo Gong, and Jia Liu. 2017. "Discriminative Feature Learning for Unsupervised Change Detection in Heterogeneous Images Based on a Coupled Neural Network." *IEEE Transactions on Geoscience and Remote Sensing* 55 (12): 7066–7080. <https://doi.org/10.1109/TGRS.2017.2739800>.