Fourier domain structural relationship analysis for unsupervised multimodal change detection

Hongruixuan Chen[a], Naoto Yokoya[a,b,*] and Marco Chini[c]

[a]Graduate School of Frontier Sciences, The University of Tokyo, Chiba, 277-8561, Japan
[b]RIKEN Center for Advanced Intelligence Project (AIP), RIKEN, Tokyo, 103-0027, Japan
[c]Luxembourg Institute of Science and Technology (LIST), Belvaux, 4450, Luxembourg

ARTICLE INFO

Keywords:
Change detection
Multimodal remote sensing images
Fourier domain
Structural relationship
Graph spectral convolution

ABSTRACT

Change detection on multimodal remote sensing images has become an increasingly interesting and challenging topic in the remote sensing community, which can play an essential role in time-sensitive applications, such as disaster response. However, the modal heterogeneity problem makes it difficult to compare the multimodal images directly. This paper proposes a Fourier domain structural relationship analysis framework for unsupervised multimodal change detection (FD-MCD), which exploits both modality-independent local and nonlocal structural relationships. Unlike most existing methods analyzing the structural relationship in the original domain of multimodal images, the three critical parts in the proposed framework are implemented on the (graph) Fourier domain. Firstly, a local frequency consistency metric calculated in the Fourier domain is proposed to determine the local structural difference. Then, the nonlocal structural relationship graphs are constructed for pre-change and post-change images. The two graphs are then transformed to the graph Fourier domain, and high-order vertex information is modeled for each vertex by graph spectral convolution, where the Chebyshev polynomial is applied as the transfer function to pass K-hop local neighborhood vertex information. The nonlocal structural difference map is obtained by comparing the filtered graph representations. Finally, an adaptive fusion method based on frequency-decoupling is designed to effectively fuse the local and nonlocal structural difference maps. Experiments conducted on five real datasets with different modality combinations and change events show the effectiveness of the proposed framework.

1. Introduction

1.1. Background

The development of remote sensing technology has provided strong data support for detecting land-cover changes using various kinds of remote sensing images, such as very-high-resolution optical images, multispectral images, and synthetic aperture radar (SAR) data. Under this circumstance, multimodal change detection has attracted a growing interest in the remote sensing community since it can allow us to alleviate the assumption of unimodal and co-calibrated data sources in change detection (Touati et al., 2020b), thereby exploiting a large number of heterogeneous remote sensing images. The advantages of expanding the data source are evident in terms of timely and comprehensive access to land-cover changes, making this technique of practical significance in high temporal resolution monitoring and response to emergent events, such as earthquakes, wildfires, and floods (Adriano et al., 2021; Ienco et al., 2019; Vetrivel et al., 2018).

Multimodal change detection aims at detecting changes of objects or phenomena from multitemporal remote sensing images with different modalities. Compared to unimodal change detection, relatively few research works have been carried out in this field despite its undeniable practical potential. In multimodal change detection, the pre-change and post-change images could be acquired by different sensors or by the same sensor but with different acquisition modes (Sun et al., 2022; Touati et al., 2020b). This means that the pre-change and post-change images could have different statistical distributions, channel numbers, and noise levels, known as the modal heterogeneity problem. Therefore, it is difficult to detect land-cover changes accurately from multimodal images using mature paradigms designed for unimodal change detection (Gil-Yepes et al., 2016; Hou et al., 2021; Hussain et al., 2013; Wu et al., 2022; Zhu, 2017). According to whether or not annotation information is given to the change detector, the existing multimodal change detection methods can be divided into supervised, semi-supervised, and unsupervised ones. Annotating labels for supervised multimodal change detection models is time-consuming and often needs expert knowledge (Chen et al., 2020; Lei et al., 2022b; Wu et al., 2021a). This step is not necessary in the case of unsupervised multimodal change detection, which aims at transforming multimodal images to a new domain where the modal heterogeneity can be significantly reduced (which is also the scope of this paper).

We divide the existing unsupervised multimodal change detection methods into four types based on the transformation method and the domain type, namely 1) modality translation-based methods, 2) feature learning-based methods, 3) classification-based methods, and 4) similarity measurement based methods. Firstly, modality translation methods mainly employ adversarial learning or style transfer approaches to project pre-change/post-change image from

Manuscript submitted on December 22, 2022.
*Corresponding author
ORCID(s):

its modality to the modality of post-change/pre-change image to alleviate modal heterogeneity. After that, an unsupervised change detector can be directly applied to detect changes from transformed images. Some typical and effective models, such as conditional generative adversarial network (cGAN) (Niu et al., 2019), image-to-image translation and cycle consistency (Luppino et al., 2022a,b), the CutMix transformation (Radoi, 2022), and style transfer learning networks (Jiang et al., 2020), have been studied and applied. Also, some classic models, like the concentric circular invariant convolution model (Touati et al., 2019b) and traditional regression models (Luppino et al., 2019), show sound transformation effects. However, the computational overhead of this kind of method is generally quite large.

Secondly, the feature learning-based methods aim at finding a high-level feature space in which the features extracted from multimodal images can be compared directly. Due to the powerful learning capabilities of deep learning models, most such methods currently apply or design unsupervised deep learning models to find such a feature space. Some commonly used models include deep belief network (Zhao et al., 2017), denoising autoencoder (Liu et al., 2018a,b; Zhan et al., 2018; Zhang et al., 2016), sparse autoencoder (Touati et al., 2020a), and convolutional autoencoder (Wu et al., 2021b). A fundamental assumption of these methods is that changed pixels only occupy a small part of multimodal image pairs. Based on this assumption, these methods can minimize the difference between paired features directly for learning the common feature space. Nonetheless, the optimization process will be compromised if the changed pixels occupy a large proportion.

Thirdly, the classification-based methods use the class domain instead of a high-dimensional feature space as the common domain. These approaches first classify multimodal image pairs, and then the obtained classification results can be directly compared to detect changes, such as the post-classification comparison method (Camps-Valls et al., 2008), compound classification method (Wan et al., 2019a,b), and multidimensional evidential reasoning method (Liu et al., 2014). Since unsupervised classification models struggle to obtain very accurate classification results, the classification-based methods suffer from the classification error accumulation problem (Wu et al., 2017a,b). Although some advanced classification models have been proposed to address this problem, like hierarchical extreme learning machine (Han et al., 2021) and self-paced convolutional network (Li et al., 2021), the problem still persists.

Finally, the similarity measure-based methods transform the multimodal images to a constructed feature space based on some modality-independent relationships, in which the constructed modality-independent features are used to distinguish the changed and unchanged areas, such as sorted histogram (Wan et al., 2018), homogeneous pixel transformation method (Liu et al., 2018b), local spatial-temporal gradient (Touati et al., 2019a), affinity matrix distance (AMD) (Luppino et al., 2019), pixel pair method (Kwan et al., 2019; Touati et al., 2020b), difference prior information (Jimenez-Sierra et al., 2020, 2022), and patch similarity graph matrix (PSGM) (Sun et al., 2021c). The main advantages of such approaches are that they usually do not require a priori assumptions about change information and are intuitive and easy to implement in practice. We follow the research lines of this type of method in this paper.

1.2. Motivations and contributions

The existing similarity measures can be broadly divided into two types based on the modality-independent relationships they utilize. We call the metric utilizing the local structural relationship (see Figure 1-(a)) as the local similarity metric and the nonlocal structural relationship (see Figure 1-(b)) as the nonlocal similarity metric, respectively. Despite the promising results obtained by these approaches based on local and nonlocal similarity metrics, there are still some problems as follows.

Firstly, the local similarity metrics exploit the inherent modality-independent similarity relationship of pixels within the local area of remote sensing images. However, most of the local similarity metrics are calculated in the original domain of remote sensing images, which makes the relationship function of these metrics still modality-relevant. Due to the modal heterogeneity in the original domain, applying the same relationship function does not yield accurate detection effects. Thus, it is often inevitable to design the associated relationship functions for different modalities or to select different parameters for the same relationship function. For example, in (Luppino et al., 2019), the proposed AMD needs to determine the bandwidth of the Gaussian kernel function according to the imaging modality.

Secondly, the nonlocal similarity metrics exploit the relationship between areas within the image, inspired by the self-similarity property of images (Buades et al., 2005). In (Mignotte, 2020), self-similarity was first introduced into unsupervised multimodal change detection for modality translation. After that, Sun *et al.* constructed the K-nearest neighbour (KNN) graph data, where each vertex is connected to its K most similar vertices, to represent the nonlocal structural relationship and compare the similarity between constructed graphs to detect changes (Sun et al., 2021b). Based on this idea, a series of representative works have been proposed, including the patch similarity graph matrix (PSGM) (Sun et al., 2021c), nonlocal patch similarity graph (NPSG) (Sun et al., 2021b), iterative robust graph and Markcovian co-segmentation (IRGMcS) (Sun et al., 2021a), and structured graph-based image regression and Markov segmentation (HGIR-MRF) (Sun et al., 2022), and structural relationship graph convolutional autoencoder (SR-GCAE) (Chen et al., 2022). Although these methods have achieved decent performance, they have two major problems. Firstly, each KNN graph only contains the one-hop local neighborhood vertices of each vertex and ignores the second-hop and more distant vertex information. We refer to the current method as a first-order KNN graph-based method. More importantly, these methods just simply compare the edges

or vertices of the KNN graph and do not fully consider the structural information contained in the graph.

Furthermore, most of these methods employ only one of the local or nonlocal structural relationships. It is reasonable to assume that better detection results can be obtained by comprehensively utilizing both relationships. However, a consequent problem is how to effectively fuse the detection results based on the two structural relationships to obtain more robust detection results.

Whereas all of the above methods are performed in the original domain of multimodal images, we propose to uniformly address the above challenges by exploring local and nonlocal structural relationships and fusing the associated detection results over the (graph) Fourier domains. To address the first problem, we propose a local frequency consistency metric based on the observation that the modal heterogeneity in the original domain of multimodal images will be eliminated in the Fourier domain. Thus, we can directly compare the frequency features of local areas between multimodal images to determine the change level. For the problems of nonlocal similarity metrics, we construct nonlocal structural relationship graphs for pre-change and post-change images. After transforming the two graphs to the graph Fourier domain, we process them with a parameter-free graph spectral convolution model to fully extract graph information. Finally, we propose to decouple the obtained local and nonlocal structural difference maps into low- and high-frequency components using Fourier transform and adaptively fuse different components with different rules to obtain more robust difference maps.

The main contribution of this paper can be summarized as follows:

- An unsupervised structural relationship analysis framework is presented for multimodal change detection, which could provide a perspective from the Fourier domain for multimodal change detection.
- A simple but effective local frequency consistency metric calculated on the Fourier domain is proposed to measure the local structural difference. This metric can also be served as the robust prior for subsequent change detection methods.
- A graph spectral convolution model is developed to model the high-order vertex information of graph data on the graph Fourier domain for more effective nonlocal structural relationship analysis.
- A frequency-decoupling change information adaptive fusion method is proposed to fuse local and nonlocal structural difference maps adaptively.

The rest of the paper is organized as follows. Section 2 describes the preliminary knowledge of our method. Section 3 elaborates on the proposed method. Section 4 provides the experimental results and the analysis. Finally, the conclusions are drawn in Section 5.

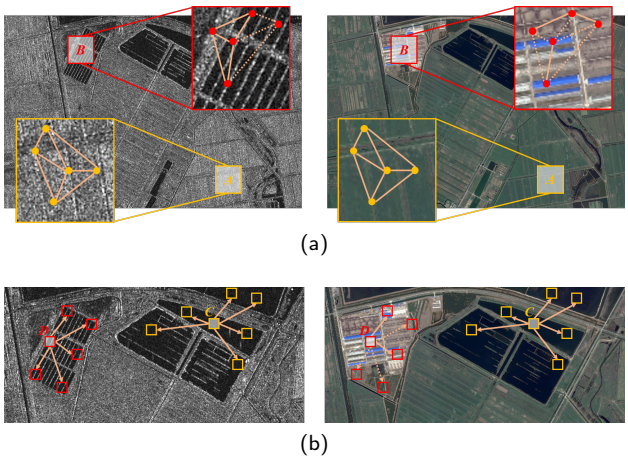


Figure 1: Two modality-independent structural relationships inherent in multimodal images. (a) Local structural relationship. (b) Nonlocal structural relationship. Here, the solid line implies high similarity, and the dotted line implies low similarity.

2. Relevant knowledge

2.1. Structural relationship in multimodal data

Even though it is difficult to compare multimodal remote sensing images directly, there are still some inherent modality-independent relationships in multimodal images that can be used to detect land-cover changes. We utilize the local structural relationship and nonlocal structural relationship inherent in multimodal images in this work, as shown in Figure 1.

Figure 1-(a) shows the local structural relationship between a pre-change SAR image and a post-change optical image. For the unchanged area A, the similarity relationship between pixels within the area A remains close to the same for both images, despite their significant modal differences. However, for changed area B, the similarity relationship between pixels within the area is significantly inconsistent between the two images due to the occurrence of a change event. Therefore, we can detect the land-cover changes from multimodal images by comparing the consistency level of the local structural relationship. Some typical local similarity metrics include affinity matrix distance (Luppino et al., 2019) and local spatial-temporal gradient difference (Touati et al., 2019a).

Figure 1-(b) illustrates the nonlocal structural relationship between a pre-change SAR image and a post-change optical image. The nonlocal structural relationship is inspired by the self-similarity property of images, which has been widely applied in image restoration and denoising (Buades et al., 2005, 2010; Dabov et al., 2007). Firstly, some most similar areas can be found for both area C and area D. Clearly, since area C is unchanged, this similarity relationship between area C and its most similar areas will remain consistent in both images. However, in area D, where a change event happens, it is evident that this similarity relationship cannot be maintained between both images. Similar to the local structural relationship, we can compare

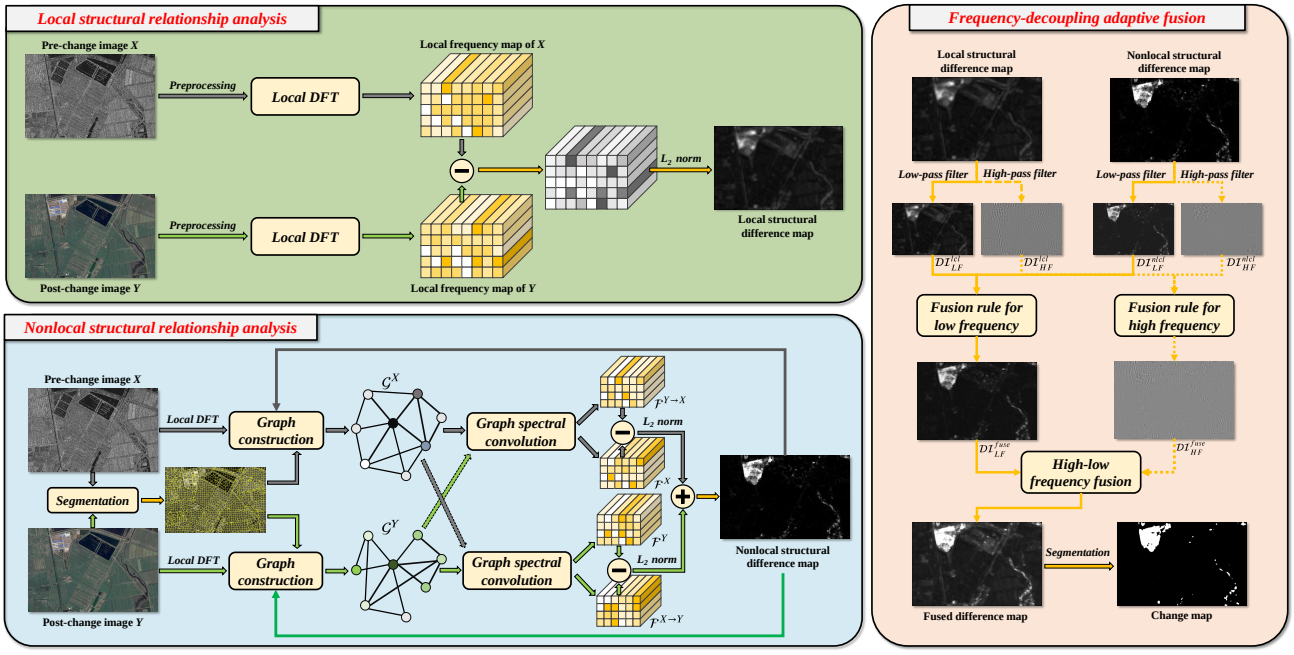


Figure 2: The overview of our Fourier domain structural relationship analysis framework.

the consistency level of the nonlocal structural relationship as an indirect measure of the change level.

Both structural relationships are computed within the remote sensing images and are, therefore, modality-independent. There have been some approaches that use these two relationships separately for change detection. In this paper, we explore both relationships in the (graph) Fourier domain for multimodal change detection rather than the original domain as the existing methods.

2.2. Fourier transform and its variants

The Fourier transform is a mathematical transformation that decomposes functions that depend on space or time into functions that depend on spatial frequency or temporal frequency. Given an integrable function $f(t)$, Fourier transform is defined as

$$\hat{f}(\xi) = \mathcal{F}_t[f](\xi) = \int_{-\infty}^{\infty} f(t)e^{-i2\pi\xi t}dt, \tag{1}$$

where the variable t often represents time and the transform variable ξ represents frequency. The transform of $f(t)$ at frequency ξ is given by the complex number $\hat{f}(\xi)$. All values of $\xi \in \mathbb{R}$ produce the Fourier domain function.

If $f(t)$ is analytic, it can be expressed as a combination of complex exponentials of all possible frequencies (Stein and Shakarchi, 2010), which is called inverse Fourier transform, given by

$$f(t) = \mathcal{F}_{\xi}^{-1}[\hat{f}](t) = \int_{-\infty}^{\infty} \hat{f}(\xi)e^{i2\pi\xi t}d\xi. \tag{2}$$

For digital image processing, we usually consider a discrete variant of the Fourier transform, the discrete Fourier transform (DFT). It is a sampled Fourier transform and therefore does not contain all frequencies forming the signal but only a set of samples that is large enough to fully describe the spatial domain image.

For a two dimensional image I with a size of $H \times W$, the two dimensional discrete Fourier transform is formed as

$$\begin{aligned}\hat{I}(u,v) &= \mathcal{DF}_{h,w}[I](u,v) \\ &= \sum_{h=0}^{H-1}\sum_{w=0}^{W-1} I(h,w)e^{-\imath 2\pi\left(\frac{u}{H}h+\frac{v}{W}w\right)},\end{aligned} \tag{3}$$

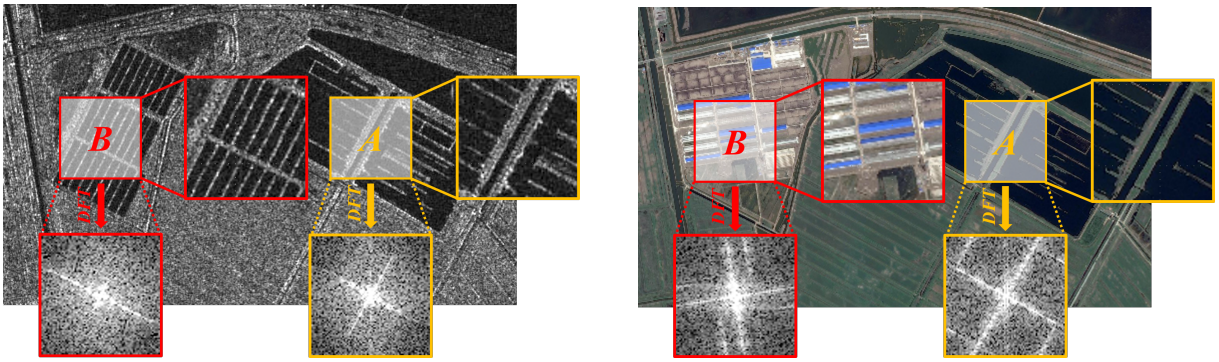
where $I(h,w)$ is the image in the original domain and the exponential term is the basis function corresponding to each point $\hat{I}(u,v)$ in the Fourier domain.

Also, the Fourier image $\hat{I}$ can be retransformed to the spatial domain, given by:

$$\begin{aligned}I(h,w) &= \mathcal{DF}_{u,v}^{-1}[\hat{I}](h,w) \\ &= \frac{1}{HW}\sum_{u=0}^{H-1}\sum_{v=0}^{W-1} \hat{I}(u,v)e^{\imath 2\pi\left(\frac{h}{N}u+\frac{w}{N}v\right)}.\end{aligned} \tag{4}$$

In addition, the complex exponential $e^{i2\pi\xi t}$ in Fourier transform are actually the eigenfunctions of the one dimensional Laplacian operator (Shuman et al., 2013). Therefore, the graph Fourier transform can be defined analogously when processing graph data.

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the vertex set with the vertex number of $|\mathcal{V}| = N$ and $\mathcal{E}$ is the edge set, a graph signal $f : \mathcal{V} \to \mathbb{R}$ is a function defined on the vertices of $\mathcal{G}$. The signal f maps every vertex $\{v_i\}_{i=1,\dots,N} \in \mathcal{V}$ to a real



(a) (b)

Figure 3: Illustration of a changed area and an unchanged area in a multimodal image-pair in original and Fourier domains. (a) SAR image in T_1. (b) Optical image in T_2. In both images, the unchanged area A shows the modality difference in the original domain but becomes similar in the Fourier domain. In comparison, the changed area B keeps an obvious difference in both domains.

number (or vector) $f(i)$. Then, the graph Fourier transform is formulated as

$$\hat{f}\left(\lambda_l\right) = \mathcal{GF}_i[f]\left(\lambda_l\right) = \sum_{i=1}^{N} f(i) u_l^T(i), \tag{5}$$

where λ_l is l-th eigenvalue of the graph Laplacian matrix ordered as $0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \ldots \leq \lambda_N := \lambda_{\max}$ and u_l^T is the transpose of corresponding eigenvector, also called graph Fourier basis.

The inverse graph Fourier transform is expressed by

$$f(i) = \mathcal{GF}_l^{-1}[\hat{f}](i) = \sum_{l=1}^{N} \hat{f}\left(\lambda_l\right) u_l(i). \tag{6}$$

Similar to the Fourier transform, the graph Fourier transform provides a way to represent graph signals in two different domains, i.e., the vertex domain and the graph Fourier domain.

3. Methodology

In this section, an unsupervised framework for multimodal change detection is proposed by analyzing structural relationships on the (graph) Fourier domain. As shown in Figure 2, the proposed framework consists of three major parts: 1) local structural relationship analysis in the Fourier domain; 2) nonlocal structural relationship analysis in the graph Fourier domain; 3) frequency-decoupling change information adaptive fusion.

3.1. Local structural relationship analysis in Fourier domain

Given a pair of co-registered multimodal remote sensing images, the pre-change image acquired in T_1 with modality $\mathcal{X}$ is denoted as $X \in \mathbb{R}^{H\times W\times C_X}$, and the post-change image acquired in T_2 with modality $\mathcal{Y}$ is denoted as $Y \in \mathbb{R}^{H\times W\times C_Y}$. The pixels in the two images are denoted as $x(h,w,c)$ and $y(h,w,c)$, respectively. Here, H, W, and C_X/C_Y are the height, width, and channel number of X/Y, respectively.

Generally, most of the existing local similarity metrics can be formulated as

$$\mathcal{DI}^{lcl}(h,w) = \sum_{(m,n)\in\mathcal{W}_{h,w}} \mathcal{D}(\Phi^{\mathcal{X}}(X(h,w), X(m,n)), \Phi^{\mathcal{Y}}(Y(h,w), Y(m,n))), \tag{7}$$

where $\mathcal{DI}^{lcl}(h,w)$ is the level of local structural difference between X and Y in the location (h,w), $\mathcal{W}_{h,w}$ is a local square window centered in (h,w), $D(\cdot,\cdot)$ is a distance function, $\Phi^{\mathcal{X}}(\cdot)$ and $\Phi^{\mathcal{Y}}(\cdot)$ is the relationship function measuring the relationship of pixels in $\mathcal{W}_{h,w}$ with modality $\mathcal{X}$ and $\mathcal{Y}$, respectively.

We can obtain different local similarity metrics by bringing the specific functions into Equation (7). For example, using the Gaussian kernel function as the relationship function and the L_1 distance as the distance metric, we can obtain the affinity matrix distances proposed in (Luppino et al., 2019, 2022b).

However, the relationship function in Equation (7) is still modality-relevant. Using the same relationship function, i.e., $\Phi^{\mathcal{X}}(\cdot,\cdot) = \Phi^{\mathcal{Y}}(\cdot,\cdot)$ often does not lead to very accurate detection results due to the modal heterogeneity in the original domain of multimodal images. Thus, it is often unavoidable to design different relationship functions according to the modalities or select suitable parameters in the same relationship functions for different modalities.

Therefore, our motivation in this part is to find a suitable relationship function that can avoid the selection of parameters and is robust enough for remote sensing images with different imaging modalities. Specifically, we argue that although multimodal images vary considerably in their original domain, the inner land-cover objects of unchanged areas vary spatially at similar distribution. For example, in an agricultural area, the regular distribution of crops over the spatial domain should be similar under different imaging modalities. Alternatively, an urban area shows a specific

distribution pattern due to the presence of regular regions defined by artifacts (like roads and buildings), which is more or less maintained in the two imaging modalities. Another example is that the waters show a smooth distribution of pixels in different imaging modalities. The pattern of these spatial distributions will be directly reflected in the changing intensity of their pixel values in the spatial direction, that is, the frequency of the images. Thus, we argue that the same unchanged area in different imaging modalities will have close frequencies.

Figure 3 shows an unchanged area A and a changed area B with two modalities in the original domain and Fourier domain (amplitude image), respectively. Intuitively, although the original image of area A shows the apparent difference in pixel values caused by different imaging modalities, thereby making direct comparison difficult, the patterns of frequency images of area A in the two modalities are close that direct comparison seems possible. In comparison, both original and frequency images of changed area B show apparent differences between the two modalities.

Based on the above observation, it is natural to introduce the discrete Fourier transform to measure the local structural relationship of multimodal images in the Fourier domain. Since amplitude information contains components of all frequencies (Proakis, 2001; Wong, 2011), we here calculate the amplitude information of remote sensing images, i.e., $\Phi^{\mathcal{X}}(\cdot) = \Phi^{\mathcal{Y}}(\cdot) = |\mathcal{DF}[\cdot]|$ with

$$|\mathcal{DF}[I]|(u,v) = \sqrt{Re\left(\hat{I}(u,v)\right) + Im\left(\hat{I}(u,v)\right)}, \quad (8)$$

where $Re\left(\hat{I}\right)$ and $Im\left(\hat{I}\right)$ are real and imaginary parts of the Fourier image, respectively. Then, we propose a local similarity metric in the Fourier domain, called local frequency consistency (LFC). If $C_X = C_Y$, the local structural difference can be measured by LFC as

$$\mathcal{DI}^{lcl}(h,w) = \frac{1}{|\mathcal{W}|}\left\|\left|\mathcal{DF}_{\mathcal{W}_{h,w}}[X]\right| - \left|\mathcal{DF}_{\mathcal{W}_{h,w}}[Y]\right|\right\|_F, \quad (9)$$

where $\|\cdot\|_F$ is the Frobenius norm. If $C_X \neq C_Y$, we execute the principal component analysis (PCA) algorithm first so that X and Y have the same number of channels and then calculate the local frequency consistency.

In addition to being used in our framework to calculate local structural differences, the proposed metric can also be used as a change prior to other change detection models, as we will elaborate in Section 3.4.

3.2. Nonlocal structural relationship analysis in graph Fourier domain

Since the nonlocal structural relationship is the relationship between image areas, the unit for analysis in this part is not a pixel but an image object, also called superpixel. Firstly, the image co-segmentation algorithm based on the fractal net evolution approach (FNEA) (Baatz, 2000) is performed on the stacked image of X and Y to get a unified segmentation map Ω as

$$\begin{cases} \Omega = \{\Omega_i \mid i = 1, 2, \ldots, N_o\} \\ \Omega_i \cap \Omega_j = \varnothing \ \text{ if } i \neq j \\ \bigcup_{i=1}^{N_o} \Omega_i = \{(h,w) \mid h = 1, \ldots, H; w = 1, \ldots, W\} \end{cases} \quad (10)$$

where N_o is the number of the obtained objects.

According to the pixel position index of Ω_i, the i-th object in X and Y can be defined as

$$\begin{cases} O_i^X = \left\{x(h,w,c) \mid (h,w) \in \Omega_i, c = 1, 2, \cdots, C_X\right\} \\ O_i^Y = \left\{y(h,w,c) \mid (h,w) \in \Omega_i, c = 1, 2, \cdots, C_Y\right\} \end{cases}. \quad (11)$$

Next, different kinds of features are extracted from the image objects to represent their information. Previous methods often adopted the statistical quantities of spectral information (Sun et al., 2021a, 2022). Based on our argument and observation in Section 3.1 that the modality heterogeneity of multimodal images in the original domain could be reduced in the Fourier domain, we propose to utilize the mean of local frequencies of pixels in the object as the feature vector. Thus, we have the feature vectors of O_i^X and O_i^Y as

$$\begin{cases} \tilde{\mathcal{O}}_i^X = \frac{1}{|\Omega_i|}\sum_{(h,w)\in\Omega_i} \text{vec}\left(\left|\mathcal{DF}_{\mathcal{W}_{h,w}}[X]\right|\right) \\ \tilde{\mathcal{O}}_i^Y = \frac{1}{|\Omega_i|}\sum_{(h,w)\in\Omega_i} \text{vec}\left(\left|\mathcal{DF}_{\mathcal{W}_{h,w}}[Y]\right|\right) \end{cases} \quad (12)$$

where $\tilde{\mathcal{O}}_i^X \in \mathbb{R}^{C_X|\mathcal{W}|}$ and $\tilde{\mathcal{O}}_i^Y \in \mathbb{R}^{C_Y|\mathcal{W}|}$ are the local frequency features of O_i^X and O_i^Y, respectively, and vec $(\cdot)$ is vectorization operation transforming local frequency image into vector. We further denote the feature matrices of X and Y consisting of $\tilde{\mathcal{O}}_i^X$ and $\tilde{\mathcal{O}}_i^Y$ as $\tilde{\mathcal{O}}^X = \left[\tilde{\mathcal{O}}_1^X, \tilde{\mathcal{O}}_2^X, \cdots, \tilde{\mathcal{O}}_{N_o}^X\right] \in \mathbb{R}^{N_o \times C_X|\mathcal{W}|}$ and $\tilde{\mathcal{O}}^Y = \left[\tilde{\mathcal{O}}_1^Y, \tilde{\mathcal{O}}_2^Y, \cdots, \tilde{\mathcal{O}}_{N_o}^Y\right] \in \mathbb{R}^{N_o \times C_Y|\mathcal{W}|}$, respectively.

Then, we build graph data to represent the nonlocal structural relationship between X and Y. Taking the pre-change image X as an example, we build a nonlocal structural relationship weighted graph $\mathcal{G}^X$ as

$$\begin{cases} \mathcal{G}^X = \left\{\mathcal{V}^X, \mathcal{E}^X, W^X\right\} \\ \mathcal{V}^X = \left\{O_i^X \mid i = 1, ..., N_o\right\} \\ \mathcal{E}^X = \left\{e_{i,j}^X = \left(O_i^X, O_j^X\right) \mid i, j = 1, ..., N_o\right\} \\ W^X = \left\{w_{e_{i,j}^X} \mid e_{i,j}^X \in \mathcal{E}^X\right\} \end{cases} \quad (13)$$

where $\mathcal{V}^X$ is the vertex set consisting of image objects of X, $\mathcal{E}^X$ is the edge set, and each edge connects two vertices, $w_{e_{i,j}^X}$ measures the weight of edge $e_{i,j}^X$ connecting vertex i and j by

a truncation Gaussian kernel function as

$$w_{e_{i,j}^X} = \begin{cases} \exp\left(-\phi^X D_{i,j}^X\right), & w_{e_{i,j}^X} \geq \tau^X \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

where $\phi^X > 0$ is the bandwidth of Gaussian kernel function, $D_{i,j}^X$ is the distance between O_i^X and O_j^X in the Fourier domain, defined as $D_{i,j}^X = \|\tilde{O}_i^X - \tilde{O}_j^X\|_2^2$, and τ^X is a truncation threshold. For post-change image Y, we can also construct such a graph like $\mathcal{G}^X$ in Equation (13), denoted as $\mathcal{G}^Y = \{\mathcal{V}^Y, \mathcal{E}^Y, W^Y\}$.

To fully exploit the structural information contained in $\mathcal{G}^X$ and $\mathcal{G}^Y$, we propose to analyze them in graph Fourier domain, instead of simply comparing their edges or vertices as previous works (Sun et al., 2021a,b). Specifically, considering the nonlocal structural relationship graph of X, its normalized graph Laplacian matrix is defined as

$$\hat{L}^X = I - \left(D^X\right)^{-\frac{1}{2}} W^X \left(D^X\right)^{-\frac{1}{2}}, \tag{15}$$

where I is the identity matrix and D^X is a diagonal matrix of W^X, namely $D_{i,i}^X = \sum_{j=1}^{N_o} W_{i,j}^X$.

Since $\hat{L}^X$ is a real symmetric positive semi-definite matrix, it has a complete set of orthonormal eigenvectors $\{U_l^X\}_{l=1}^{N_o} \in \mathbb{R}^{N_o}$, i.e., the graph Fourier basis (Shuman et al., 2013), and associated real non-negative eigenvalues $\{\lambda_l^X\}_{l=1}^{N_o}$ ordered as $\lambda_1^X \leq \lambda_2^X \leq \cdots \leq \lambda_{N_o}^X$. $\hat{L}_X$ is diagonalized by the graph Fourier basis as $\hat{L}_X = \mathcal{U}^X \Lambda^X \left(\mathcal{U}^X\right)^T$, where $\Lambda^X = \text{diag}\left(\lambda_1^X, \lambda_2^X, \cdots, \lambda_{N_o}^X\right) \in \mathbb{R}^{N_o \times N_o}$ is the diagonal matrix of eigenvalues and graph Fourier bases $\mathcal{U}^X = \left[U_1^X, U_2^X, \cdots, U_{N_o}^X\right] \in \mathbb{R}^{N_o \times N_o}$.

The graph Fourier transform of $\mathcal{G}^X$ is given by $\hat{\mathcal{O}}^X = \mathcal{U}^T \tilde{\mathcal{O}}^X$, and the inverse transform is $\tilde{\mathcal{O}}^X = \mathcal{U} \hat{\mathcal{O}}^X$. Then, with graph Laplacian eigenvectors as the basis, the convolutional operation can be defined on the graph Fourier domain (Chung, 1997; Shuman et al., 2013) to capture the graph structural information of $\mathcal{G}^X$ as

$$\mathcal{F}^X = \mathcal{U}^X \mathcal{H}\left(\Lambda^X\right) \left(\mathcal{U}^X\right)^T \tilde{\mathcal{O}}^X, \tag{16}$$

where $\mathcal{H}\left(\Lambda^X\right)$ is the graph Fourier filters based on Λ^X, defined as $\mathcal{H}\left(\Lambda^X\right) = \text{diag}\left(h\left(\lambda_1^X\right), h(\lambda_2^X), \cdots, h(\lambda_{N_o}^X)\right)$, $h(\cdot)$ is the transfer function. When the graph Fourier filter is a K-order polynomial of eigenvalues, the filtered signal at vertex i equals a linear combination of the components of the input signal at vertices within a K-hop local neighborhood of vertex i (Shuman et al., 2013). Therefore, the graph Fourier filter in our work is designed as a K-order polynomial with the form $h(\lambda_l^X) = \sum_{k=0}^{K} a_k \left(\lambda_l^X\right)^k$ to capture the high-order vertex information. One such polynomial, traditionally used in graph signal processing, is the Chebyshev polynomial. Based on it, the graph Fourier filter used in our work is

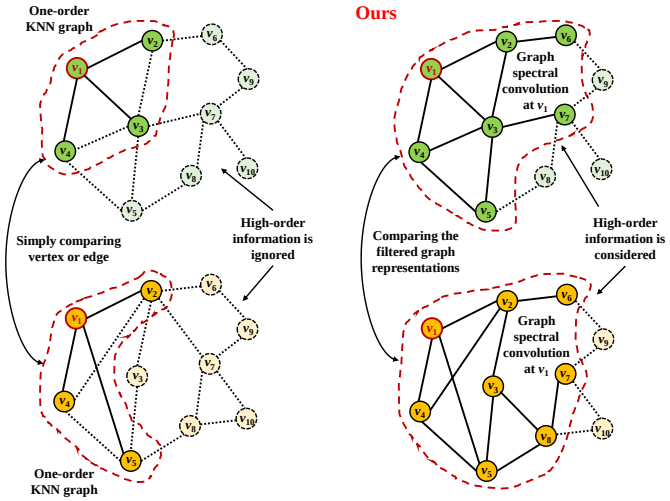


Figure 4: Illustration of the difference between the existing one-order KNN graph-based methods and ours in the case of considering 2-hop local neighborhood information.

expressed as

$$\mathcal{H}\left(\Lambda^X\right) = \sum_{k=0}^{K} T_k\left(\Lambda^X\right), \tag{17}$$

where Chebyshev polynomial T_k of order k is computed by the recurrence relation $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$ with $T_0 = 1$ and $T_1 = x$.

Often, only the first S instead of full graph Fourier bases and associated eigenvalues are used for convolution because they carry the most of smooth geometry of the graph (Bruna et al., 2013). Thus, Equation (16) is rewritten as

$$\mathcal{F}^X = \mathcal{U}_{1:S}^X \left(\sum_{k=0}^{K} T_k\left(\Lambda_{1:S}^X\right)\right) \left(\mathcal{U}_{1:S}^X\right)^T \tilde{\mathcal{O}}^X, \tag{18}$$

where $\mathcal{U}_{1:S}^X = \left[U_1^X, U_2^X, \cdots, U_S^X\right] \in \mathbb{R}^{N_o \times S}$ and $\Lambda_{1:S}^X = \left[\lambda_1^X, \lambda_2^X, \cdots, \lambda_S^X\right] \in \mathbb{R}^{S \times S}$, and the value of S depends upon the intrinsic regularity of $\mathcal{G}^X$.

Similarly, we have the same process for $\mathcal{G}^Y$ as

$$\mathcal{F}^Y = \mathcal{U}_{1:S}^Y \left(\sum_{k=0}^{K} T_k\left(\Lambda_{1:S}^Y\right)\right) \left(\mathcal{U}_{1:S}^Y\right)^T \tilde{\mathcal{O}}^Y. \tag{19}$$

Through Equation (18) and Equation (19), the structural information of $\mathcal{G}^X$ and $\mathcal{G}^Y$ can be fully encoded into $\mathcal{F}^X$ and $\mathcal{F}^Y$ for the subsequent measurement of nonlocal structural difference.

Since $\mathcal{G}^X$ and $\mathcal{G}^Y$ are constructed in different modalities, we cannot compare $\mathcal{F}^X$ and $\mathcal{F}^Y$ directly to measure the nonlocal structural difference. Instead, we can map the nonlocal structural relationship in each graph to the modality of another graph to eliminate the modal heterogeneity. Since the structural information of $\mathcal{G}^X$ and $\mathcal{G}^Y$ can be represented by its graph Fourier bases and associated eigenvalues, we

propose to perform graph spectral convolution on two graphs with each other's graph Fourier filters as

$$\begin{cases} \mathcal{F}^{Y\to X} = \mathcal{U}^{Y}_{1:S}\left(\sum_{k=0}^{K} T_k\left(\Lambda^{Y}_{1:S}\right)\right)\left(\mathcal{U}^{Y}_{1:S}\right)^{T}\tilde{\mathcal{O}}^{X} \\ \mathcal{F}^{X\to Y} = \mathcal{U}^{X}_{1:S}\left(\sum_{k=0}^{K} T_k\left(\Lambda^{X}_{1:S}\right)\right)\left(\mathcal{U}^{X}_{1:S}\right)^{T}\tilde{\mathcal{O}}^{Y} \end{cases}. \quad (20)$$

Subsequently, the nonlocal structural difference can be measured by

$$\begin{cases} d^{\mathcal{X}}_{\Omega_i} = \frac{1}{C_X\,|\mathcal{W}|}\sqrt{\sum_{j=1}^{C_X|\mathcal{W}|}\left(\mathcal{F}^{X}_{i,j} - \mathcal{F}^{Y\to X}_{i,j}\right)^2} \\ d^{\mathcal{Y}}_{\Omega_i} = \frac{1}{C_Y\,|\mathcal{W}|}\sqrt{\sum_{j=1}^{C_Y|\mathcal{W}|}\left(\mathcal{F}^{Y}_{i,j} - \mathcal{F}^{X\to Y}_{i,j}\right)^2} \\ \mathcal{DI}^{nlcl}(h,w) = d^{\mathcal{X}}_{\Omega_i} + d^{\mathcal{Y}}_{\Omega_i},\ (h,w)\in\Omega_i, i = 1,2,\cdots,N_o \end{cases} \quad (21)$$

where $d^{\mathcal{X}}_{\Omega_i}$ and $d^{\mathcal{Y}}_{\Omega_i}$ indicates the nonlocal structural difference of area Ω_i under the modality $\mathcal{X}$ and $\mathcal{Y}$, respectively; the two difference values are summed to get a more robust result; then, the nonlocal structural difference map $\mathcal{DI}^{nlcl}$ can be obtained by assigning the summed difference value back to the pixels belonging to Ω_i. Figure 4 intuitively illustrates the difference between our method with the existing one-order KNN graph-based methods.

Finally, the vertex where the change event occurs needs to be removed from $\mathcal{E}^X$ and $\mathcal{E}^Y$ for more stable detection (Sun et al., 2021a). Here, we first adopt the detection result obtained by LFC as the prior to indicate which vertex is changed in advance. Thus, $\mathcal{E}^X$ and $\mathcal{E}^Y$ should be modified as

$$\begin{aligned} \mathcal{E}^X &= \left\{\left(O^X_i, O^X_j\right) \mid i = 1,...,N_o, j\in\mathcal{T}\right\}, \\ \mathcal{E}^Y &= \left\{\left(O^Y_i, O^Y_j\right) \mid i = 1,...,N_o, j\in\mathcal{T}\right\}, \end{aligned} \quad (22)$$

where $\mathcal{T}$ is the set containing the index of unchanged vertices. An iterative strategy is then designed to get $\mathcal{T}$ from the detection results of the last round. This process is illustrated by gray and green arrows from the difference map to the graph construction blocks in Figure 2.

3.3. Frequency-decoupling adaptive change information fusion

After we get the local structural difference map $\mathcal{DI}^{lcl}$ and nonlocal structural difference map $\mathcal{DI}^{nlcl}$, we fuse them to generate a more robust difference map. Here, instead of fusing $\mathcal{DI}^{lcl}$ and $\mathcal{DI}^{nlcl}$ by a simple addition operation, we propose a frequency-decoupling adaptive fusion strategy.

Firstly, we perform discrete Fourier transform to transform $\mathcal{DI}^{lcl}$ and $\mathcal{DI}^{nlcl}$ to the Fourier domain. A low-pass filter and a high-pass filter are separately applied on the amplitudes of $\mathcal{DI}^{lcl}$ and $\mathcal{DI}^{nlcl}$ to obtain their low- and high-frequency components. Then, we transform the components back to the original domain, obtaining four maps, denoted as $\mathcal{DI}^{lcl}_{LF}$, $\mathcal{DI}^{lcl}_{HF}$, $\mathcal{DI}^{nlcl}_{LF}$, and $\mathcal{DI}^{nlcl}_{HF}$, corresponding to the low- and high-frequency maps of $\mathcal{DI}^{lcl}$ and the low- and high-frequency maps of $\mathcal{DI}^{nlcl}$, respectively.

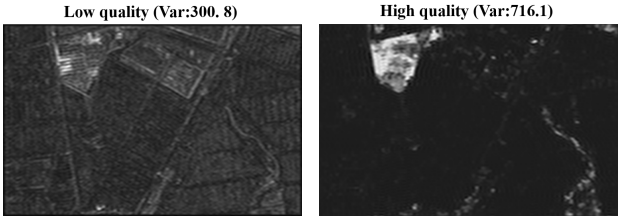


Figure 5: Example of the low-frequency components of difference maps with (a) low quality and (b) high quality.

Subsequently, we fuse the low- and high-frequency components based on different rules. For low-frequency components $\mathcal{DI}^{lcl}_{LF}$ and $\mathcal{DI}^{nlcl}_{LF}$, we fuse them adaptively according to their variance in change intensity as

$$\mathcal{DI}^{fuse}_{LF} = \frac{\sigma^2_{\mathcal{DI}^{lcl}_{LF}}\mathcal{DI}^{lcl}_{LF} + \sigma^2_{\mathcal{DI}^{nlcl}_{LF}}\mathcal{DI}^{nlcl}_{LF}}{\sigma^2_{\mathcal{DI}^{lcl}_{LF}} + \sigma^2_{\mathcal{DI}^{lcl}_{LF}}}, \quad (23)$$

where $\sigma^2_{\mathcal{DI}}$ is the variance in change intensity of difference map. The basic idea behind the fusion of low-frequency components is that the difference map with high quality should count on more weights during the fusion process. We observe that in high-quality difference maps, changed and unchanged pixels tend to show a high contrast. This high contrast can be reflected in the variance of the pixel values of the difference map. An example is given in Figure 5 to intuitively show the above idea.

For high-frequency components, we use the following rule to fuse them adaptively (Piella, 2003):

$$\mathcal{DI}^{fuse}_{HF}(h,w) = \begin{cases} \mathcal{DI}^{lcl}_{HF}(h,w), \sigma^{lcl}_{\mathcal{W}}(h,w) \le \sigma^{nlcl}_{\mathcal{W}}(h,w) \\ \mathcal{DI}^{nlcl}_{HF}(h,w), \sigma^{lcl}_{\mathcal{W}}(h,w) > \sigma^{nlcl}_{\mathcal{W}}(h,w) \end{cases} \quad (24)$$

where $\sigma_{\mathcal{W}}(h,w)$ is the local standard deviation of the high-frequency components in a window $\mathcal{W}$ centering at the location (h,w). The high-frequency components reflect the information about the salient features of differece map such as the boundary of the changed area. Some noise and voids in changed areas will show larger $\sigma_{\mathcal{W}}$ than the local background. Thus, the above rule can eliminate these pixels and refine the final detection result.

Next, $\mathcal{DI}^{fuse}_{LF}$ and $\mathcal{DI}^{fuse}_{HF}$ are transformed to the Fourier domain. The low-frequency area of the amplitude of $\mathcal{DI}^{fuse}_{LF}$ and the high-frequency area of the amplitude of $\mathcal{DI}^{fuse}_{HF}$ are

combined and transformed back to the spatial domain to get the final difference map $\mathcal{DI}^{final}$.

Finally, once $\mathcal{DI}^{final}$ is generated, a threshold segmentation algorithm can be executed on $\mathcal{DI}^{final}$ to classify each pixel to non-change class ω_{nc} or change class ω_c:

$$\mathcal{CM}(h,w) = \begin{cases} \omega_{nc}, & \mathcal{DI}^{final}(h,w) \leq T \\ \omega_c, & \text{otherwise} \end{cases} \tag{25}$$

where $\mathcal{CM}$ means the binary change map and T is a threshold value, which can be obtained by threshold segmentation algorithms, like expectation maximum approach (Bruzzone and Diego Fernàndez Prieto, 2000; Moon, 1996) and Otsu (Otsu, 1979).

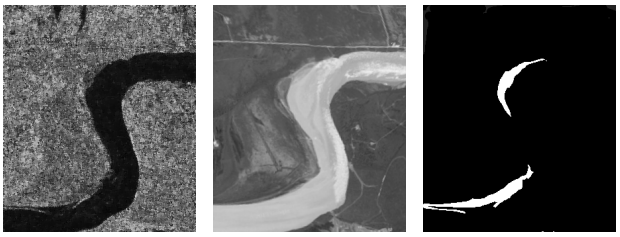
3.4. Some possible extensions and applications

Our framework, from the perspective of (graph) Fourier domain, explores the local and nonlocal structural relationships inherent in multimodal images and fuse associated detection results. Moreover, the three essential parts are not limited to our framework. We present here some of their possible extensions and applications.

1) Local frequency consistency metric. In addition to directly detecting land-cover changes using this metric, it can provide a preliminary result for other change detection methods. In Section 3.2, we utilize the detection results obtained by the local frequency consistency metric as the prior information to provide the initial unchanged vertices in the iterative nonlocal structural relationship analysis framework. Also, it can provide prior information for other unsupervised multimodal change detection frameworks. For example, both modality translation methods and feature learning-based methods need to exclude the effect of changed pixels during their optimization stage. Our metric can provide robust unchanged candidates $\mathcal{T}$ to train modality transformation models $\hat{Y} = \mathcal{R}^{\mathcal{X}\to\mathcal{Y}}(X)$ and $\hat{X} = \mathcal{R}^{\mathcal{Y}\to\mathcal{X}}(Y)$, where $\mathcal{R}^{\mathcal{X}\to\mathcal{Y}}(\cdot)$ and $\mathcal{R}^{\mathcal{Y}\to\mathcal{X}}(\cdot)$ are the transformation models. Similarly, $\mathcal{T}$ can also be applied to guide the optimization stage of feature learning-based methods in the form like $\min \mathcal{L} = \min \text{dist}\left(F_X^{\mathcal{T}}, F_Y^{\mathcal{T}}\right)$, where loss function $\mathcal{L}$ is minimized to shrink the distance between features of possible unchanged areas $F_X^{\mathcal{T}}$ and $F_Y^{\mathcal{T}}$. Alternatively, the detection results from our metric are used directly as pseudo-labels for training the deep learning models as change detector. In Section 4.4.1, we compare our metric with some representative local similarity metrics and show its effectiveness in providing robust unchanged information for an image regression model.

2) Nonlocal structural relationship analysis in graph Fourier domain. In our framework, we propose to analyze the nonlocal structural relationship graphs of multimodal images on the graph Fourier domain with the graph spectral convolution instead of simply comparing one-order KNN graphs. The graph spectral convolution employed here is parameter-free. For the purpose of extracting more representative graph features, we can consider extending the current filters to the form with learnable parameters θ to further strengthen its ability, i.e., $\mathcal{H}(\Lambda) \to \mathcal{H}_\theta(\Lambda)$. Then, the existing methods and paradigms of graph convolutional networks can be introduced (Kipf and Welling, 2016; Park et al., 2019). For learning the parameters θ in the unsupervised case, self-supervised learning like graph data reconstruction (Chen et al., 2022; Park et al., 2019) or training using the pseudo-labels (Wang et al., 2022) obtained by our local frequency metric can be alternatively considered.

3) Frequency-decoupling adaptive change information fusion. In our framework, this method is proposed to fuse the local and nonlocal structural difference maps effectively. However, the proposed image fusion strategy or some of the rules involved therein can be applied to other change detection frameworks involving difference map fusion. For example, it is common to use the ensembling learning strategy to obtain multiple difference maps and then fuse them to obtain a more robust difference map in practice (Du et al., 2013; Wang et al., 2018). Our rule of weighting the difference maps according to their variance in change intensity can be an effective alternative compared to directly adding these difference maps together. Also, this decoupling-and-fusing strategy is well suited to dealing with scenarios involving noise, such as SAR image change detection.



(a) (b) (c)

Figure 6: River dataset. (a) SAR image in T_1. (b) Panchromatic image in T_2. (c) Ground truth.

$\mathcal{T}$ from the detection results of the last round.

(a) (b) (c)

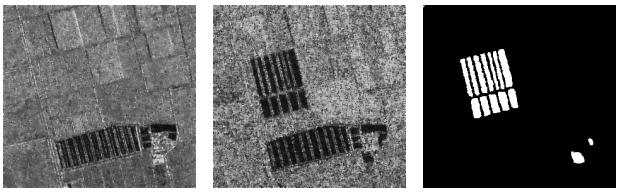
Figure 7: Shuguang dataset. (a) SAR image in T_1. (b) Multispectral image in T_2. (c) Ground truth.

4. Experiments and analysis

4.1. Datasets description

To verify the effectiveness of our framework, we conduct experiments on five multimodal datasets with different kinds of modal combinations and change events.

In the first dataset, called the River dataset, the image in T_1 is a SAR image captured by Radarsat-2 and the image in T_2 is a panchromatic image captured by Landsat-7. They have a size of 343×291 pixels. Figure 6 shows the pre-change and post-change images and the change reference

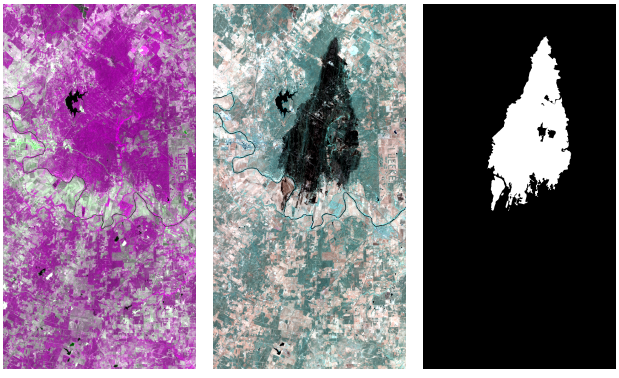


(a) (b) (c)

Figure 8: Farmland dataset. (a) SAR image with one look in T_1. (b) SAR image with four looks in T_2. (c) Ground truth.

(a) (b) (c)

Figure 9: Sardinia dataset. (a) NIR image in T_1. (b) Multispectral image in T_2. (c) Ground truth.



(a) (b) (c)

Figure 10: Texas dataset. (a) Multispectral image with seven spectral bands in T_1. (b) Multispectral image with ten spectral bands in T_2. (c) Ground truth.

map, respectively. The change event in the River dataset was bank erosion.

The second dataset is the Shuguang dataset, as shown in Figure 7. In this dataset, the pre-change SAR image was acquired in 2008 and the post-change optical image was acquired in 2012, with the same size of 593 ×921 pixels after preprocessing. The major change event in the Shuguang dataset is the building construction on farmland.

The third dataset is composed of two SAR images taken by Radarsat-2 and the associated reference map with the same size of 291×306 pixels, as shown in Figure 8. Note that even though the pre-change and post-change images were imaged by the same sensor, they are single-look and four-look, respectively, and therefore show the modal difference caused by different noise levels. The change event in the Farmland dataset is farmland reclamation.

The next dataset is the Sardinia dataset. Figure 9 shows the multimodal image-pair and associated ground truth with the size of 300×412. The pre-change image is a NIR image captured by Landsat-5, and the post-change image was obtained from Google Earth with three bands. The main change event in this dataset is the lake overflow.

The last dataset, called Texas, has two multispectral images with different numbers of bands, as shown in Figure 10. The image in T_1 has seven spectral bands imaged by Landsat-5, and the image in T_2 has ten spectral bands imaged by EO-1 ALI. Both images are 1534×808 pixels in size, showing the changes in a forest area of Texas, USA, due to wildfires.

The descriptions of the above five multimodal datasets are summarized in Table 1.

4.2. Experimental setting

In our experiment, the overall framework is implemented with Python[1]. Several hyperparameters influence the performance of the proposed FD-MCD, including the local window size w of $\mathcal{W}_{h,w}$ in local frequency consistency, the number of graph Fourier basis S in graph spectral convolution and the order number of Chebyshev polynomial K. The values of these hyperparameters in the experiments are listed in Table 2. The specific influence of these parameters at different values is discussed in Section 4.4. In addition, the fast Fourier transform is applied to reduce the time complexity of discrete Fourier transform from $O\left(N^2\right)$ to $O\left(N\log\left(N\right)\right)$.

To show the superiority of the proposed framework, we compare it with the state-of-the-art (SOTA) methods on the five datasets. Firstly, we take six recent unsupervised multimodal change detection approaches whose codes are open-sourced as the comparison methods:

1. RIF (Touati et al., 2019b) transforms the image from its own modality to the modality of another image by a circular invariant convolution model. The conjugate gradient approach is applied to optimize the parameters of the convolution model.
2. FPMS (Mignotte, 2020) utilizes self-similarity properties to map the pre-change image to the domain of the post-change image and calculate the difference map. The Markov random field is applied to generate the change map from the difference map.
3. M3CD (Touati et al., 2020b) is based on a pixel-pair-based similarity metric. The parameters of a likelihood model are estimated by the standard iterative conditional estimation framework.
4. CCLMRF (Mignotte, 2022) depends on a spatially adaptive class conditional likelihood, which is adaptive to the given imaging modality pair. After estimating the parameters of the likelihood model, a change detector is performed.
5. NPSG (Sun et al., 2021b) builds one-order KNN graphs with patches as vertices for multimodal images

[1]Source code for this work will be available at https://github.com/ChenHongruixuan/FDMCD

Table 1
Descriptions of the five multimodal datasets

Name	Imaging Sensor	Image Size	Location	Date	Event
River	Radarsat-2/Landsat-7	343×291×1/1	Yellow River, China	June 2008/Spet. 2010	River flood
Shuguang	Radarsat-2/Google Earth	593×921×1/3	Dongying, China	June 2008/Spet. 2012	Constructions
Farmland	Radarsat-2 (single/four look)	291×306×1/1	Eastern China	June 2008/June 2009	Farmland reclaim
Sardinia	Landsat-5/Google Earth	300×412×1/3	Sardinia, Italy	Sept. 1995/July 1996	Lake overflow
Texas	Landsat-5/EO-1 ALI	1534×808×7/10	Texas, USA	Aug. 2011/Sept. 2011	Forest fire

Table 2
Values of some important hyperparameter in the proposed framework on the five datasets.

Dataset	w	S	K
River	17	100	3
Shuguang	19	100	3
Farmland	5	20	2
Sardinia	19	75	2
Texas	17	50	3

and maps the structure of one-order KNN graphs from one image to another to avoid heterogeneity leakage and measure the change intensity.

6. IRGMcS (Sun et al., 2021a) also constructs the one-order KNN graph to measure the nonlocal structural similarity. Superpixel segmentation and iterative Markovian model are employed in IRGMcS for acceleration and post-processing, respectively.

Apart from these six methods, in Table 8, we further list the Kappa coefficients of other SOTA methods on each dataset, which is reported in their original literature. These methods cover the four categories mentioned in Section 1.1. In this way, the superiority of the proposed approach can be fully verified.

Five evaluation criteria are employed to comprehensively evaluate the performance of different methods, including false alarm rate (FA), missing alarm rate (MA), overall accuracy (OA), F1 score (F1), and Kappa coefficient (KC). Additionally, receiver operating characteristics (ROC) and precision-recall (PR) curves are drawn to evaluate the quality of the obtained difference maps, which can exclude the effect of the threshold segmentation method.

4.3. Change detection performance

Figure 11 shows the change maps obtained by our framework and the six comparison methods on the five datasets. The six comparison methods can accurately detect land-cover changes for some modality combinations but cannot achieve accurate detection results on all five datasets. For example, RIF, a modality translation method, can obtain change maps with good visual effects on Shuguang and Sardinia datasets. However, it fails to detect land-cover changes in the River dataset. Also, NPSG and IRGMcS yield accurate change maps in some change detection conditions, such as on the Shuguang and River datasets with a modality combination of SAR-optical and change events with artificial and natural changes. However, they fail to accurately detect land-cover changes caused by wildfire in the modality combination of multispectral-multispectral in the Texas dataset. In contrast, the proposed method can

Table 3
Accuracy assessment for different models on the River dataset. The best results are highlighted in bold, and the second-best results are underlined.

Method	FA (%)	MA (%)	OA (%)	F1	KC
RIF	95.95	74.60	78.44	0.0699	0.0157
FPMS	81.00	**7.90**	87.22	0.3150	0.2768
M3CD	92.26	32.22	71.17	0.1388	0.0865
CCLMRF	77.65	21.64	90.62	0.3477	0.3137
NPSG	37.25	35.94	97.65	0.6335	0.6213
IRGMcS	39.18	26.15	97.66	0.6665	0.6544
FD-MCD	**39.11**	11.43	**97.82**	**0.7216**	**0.7107**

Table 4
Accuracy assessment for different models on the Shuguang dataset. The best results are highlighted in bold, and the second-best results are underlined.

Method	FA (%)	MA (%)	OA (%)	F1	KC
RIF	44.09	16.43	96.22	0.6699	0.6507
FPMS	62.29	**0.80**	92.43	0.5464	0.5141
M3CD	47.61	35.20	95.64	0.5794	0.5569
CCLMRF	68.39	47.81	92.61	0.3937	0.3569
NPSG	38.19	8.86	97.00	0.7366	0.7213
IRGMcS	**11.72**	30.28	98.18	0.7791	0.7698
FD-MCD	21.08	16.04	**98.23**	**0.8136**	**0.8044**

Table 5
Accuracy assessment for different models on the Farmland dataset. The best results are highlighted in bold, and the second-best results are underlined.

Method	FA (%)	MA (%)	OA (%)	F1	KC
RIF	47.54	6.45	94.60	0.6722	0.6453
FPMS	32.61	7.29	96.91	0.7805	0.7643
M3CD	42.48	5.07	95.55	0.7164	0.6938
CCLMRF	40.91	**2.45**	95.86	0.7360	0.7149
NPSG	32.00	9.89	96.90	0.7751	0.7588
IRGMcS	18.34	17.57	97.86	0.8204	0.8090
FD-MCD	**13.83**	7.30	**98.69**	**0.8931**	**0.8862**

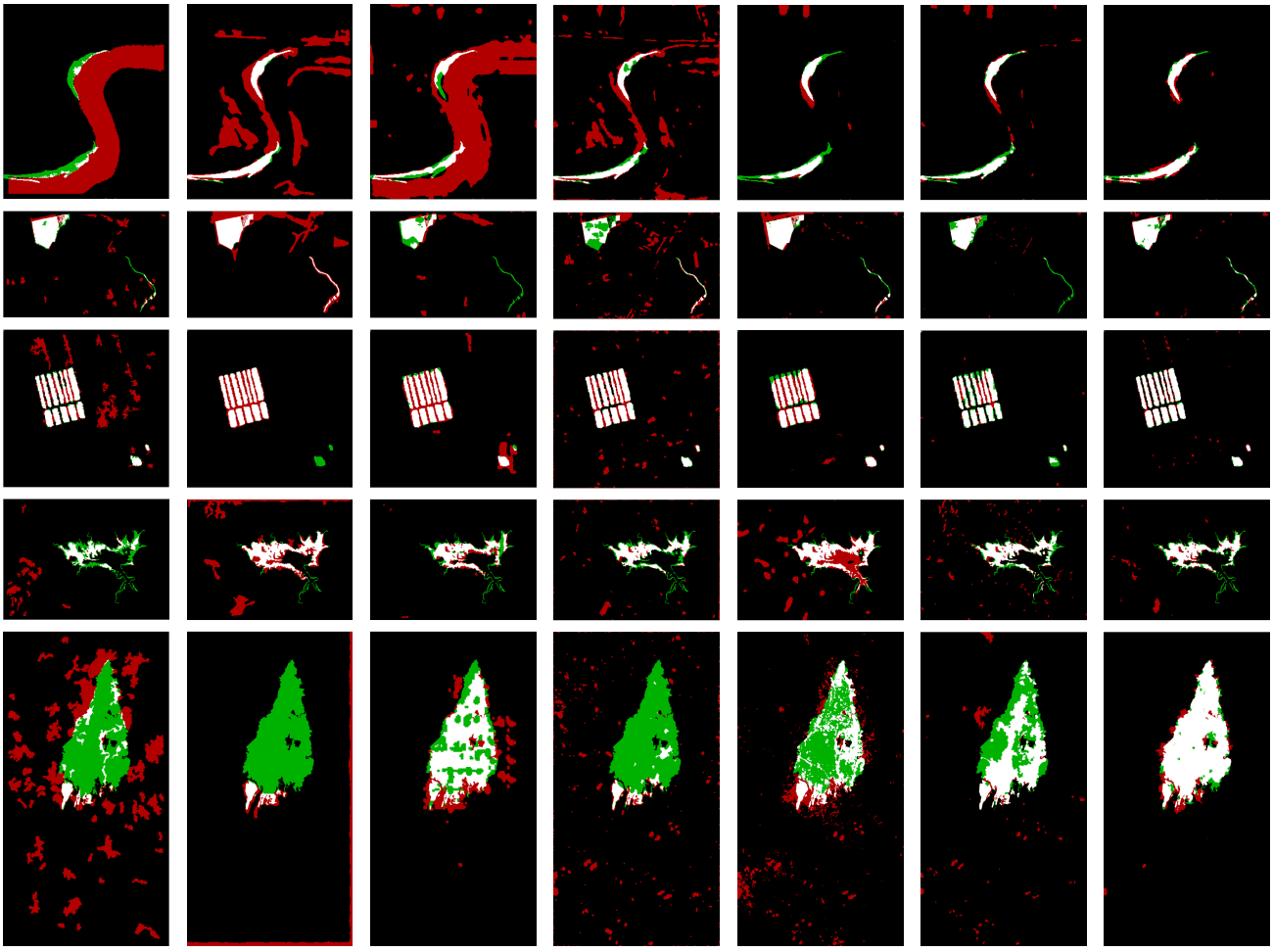


Figure 11: Change maps obtained by the six comparison methods and the proposed method on the five datasets. (a) RIF. (b) FPMS. (c) M3CD. (d) CCLMRF. (e) NPSG. (f) IRGMcS. (g) FD-MCD. From top to bottom, change maps on the River dataset, Shuguang dataset, Farmland dataset, Sardinia dataset, and Texas dataset are shown. In change maps, white indicates true positives (TP); black indicates true negatives (TN); red indicates false positives (FP); green indicates false negatives (FN).

Table 6
Accuracy assessment for different models on the Sardinia dataset. The best results are highlighted in bold, and the second-best results are underlined.

Method	FA (%)	MA (%)	OA (%)	F1	KC
RIF	45.67	56.86	94.25	0.4809	0.4509
FPMS	57.07	**12.26**	92.05	0.5765	0.5382
M3CD	28.45	33.63	96.30	0.6886	0.6690
CCLMRF	35.50	35.62	95.61	0.6444	0.6210
NPSG	56.60	13.32	92.20	0.5784	0.5406
IRGMcS	35.85	29.50	95.75	0.6717	0.6490
FD-MCD	**26.24**	27.42	**96.72**	**0.7316**	**0.7142**

Table 7
Accuracy assessment for different models on the Texas dataset. The best results are highlighted in bold, and the second-best results are underlined.

Method	FA (%)	MA (%)	OA (%)	F1	KC
RIF	87.75	85.01	79.53	0.1348	0.0200
FPMS	85.13	94.30	86.49	0.0824	0.0249
M3CD	32.45	34.63	92.97	0.6644	0.6252
CCLMRF	73.43	90.77	87.63	0.1370	0.0869
NPSG	46.78	55.86	89.93	0.4826	0.4273
IRGMcS	17.87	46.91	93.78	0.6449	0.6125
FD-MCD	**13.29**	**7.98**	**97.65**	**0.8929**	**0.8797**

generate accurate change maps with few FP and FN pixels in all five datasets with different modality combinations, spatial resolutions, and change events, showing its effectiveness in unsupervised multimodal change detection.

We further draw the ROC and PR curves of the difference maps obtained by our framework on the five datasets, as shown in Figure 12. The distributions of these curves indicate that the changed pixels are well distinguished from the unchanged ones in the obtained difference maps.

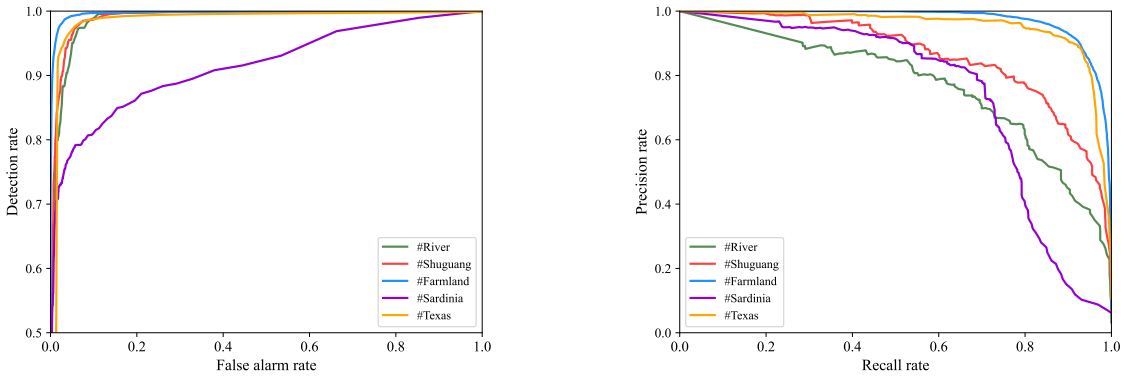


Figure 12: ROC and PR curves of the difference maps generated by the proposed method on the five datasets. (a) ROC curves. (b) PR curves.

Then, we report the corresponding quantitative results in Table 4 to Table 7. The two modality translation methods, RIF and FPMS, can successfully project an image from one modality to another under specific modality combinations and change events, such as in the Shuguang and Farmland datasets. However, both methods fail to detect changes in the Texas dataset, only obtaining KCs of 0.020 and 0.025, respectively. This is because both methods assume a small proportion of the area of change in multimodal image-pair, whereas the proportion of the changed area in the Texas dataset is large, thereby affecting the regression effect.

M3CD relies on a local similarity metric based on a pixel-pair relationship model, which compares the difference of pixel-pair in pre-change and post-change images and can therefore be considered modality-independent. However, this metric is not robust enough for different modality combinations and change events. For example, M3CD only yields a KC of 0.0865 in the River dataset.

NPSG shows more robust performance in the five datasets compared to the above methods. However, NPSG does not fully use the structural relationships in the constructed KNN graphs, and the one-order KNN graphs may not adequately cope with the complex detection conditions in multimodal remote sensing images. For example, it has both high FA (46.78%) and MA (55.86%) rates on the Texas dataset. IRGMcS achieved higher accuracy than NPSG on all five datasets by improving the one-order KNN graph construction step of NPSG and introducing MRF model for post-processing. However, it still fails to detect a large proportion of changed areas in the Texas dataset, resulting in a large MA rate (46.91%).

In comparison, our method, on the (graph) Fourier domain, simultaneously exploits local and nonlocal structural relationships and adaptively fuses the local and nonlocal structural difference maps, achieving the highest OA, F1, and KC on all five datasets, demonstrating the effectiveness of our motivation.

Apart from the six comparison methods, we also list the KC obtained by some other SOTA methods in Table 8 to further evaluate the performance of the proposed framework. Among these methods in Table 8, CACD (Wu et al., 2021b), X-Net (Luppino et al., 2022b), ACE-Net (Luppino et al., 2022b), SCCN (Liu et al., 2018a), ASDNN (Zhao et al., 2017), DCNN (Li et al., 2019), pt-CDN (Yang et al., 2019), DCNet (Gao et al., 2019), SARDNN (Gong et al., 2016), CAAE (Luppino et al., 2022a) are deep learning-based methods. Note that some of these SOTA methods adopt the post-processing step, such as the conditional random field model in UIR (Luppino et al., 2019) and MRF model in HGIR-MRF (Sun et al., 2022). On the other hand, our approach seeks to perform unsupervised multimodal change detection from a (graph) Fourier domain perspective. Achieving the highest accuracy on each dataset is not our main aim. Therefore, we do not employ these post-processing models to maintain the integrity of our framework for analysis over the Fourier domain. Even so, our method can still show competitive performance in the five datasets and outperform these SOTA methods in some datasets.

4.4. Discussion

FD-MCD contains three major steps: local structural relationship analysis on the Fourier domain, nonlocal structural relationship analysis on the graph Fourier domain, and frequency decoupling change information adaptive fusion. Figure 13 shows the performance contribution of each step in our framework on the five datasets. In this subsection, we will further analyze each step.

4.4.1. Analysis of local frequency consistency metric

The size w of the local square window $\mathcal{W}_{m,n}$ represents the range for each pixel to calculate the local frequency consistency. It will affect the performance of local structural similarity measurement. In Figure 14-(a), as w increases, the performance of local structural similarity measurement in the Fourier domain is enhanced. However, the optimal value

Table 8
Comparison in KC of the SOTA change detection models on the five datasets. KC reported here comes from their original papers. These methods cover the four categories introduced in Section 1.1. Our method is highlighted in bold. Note that some methods use post-processing models, while ours do not.

River	KC	Shuguang	KC	Farmland	KC
FD-MCD	**0.7107**	**FD-MCD**	**0.8044**	pt-CDN (Yang et al., 2019)	0.8883
CACD (Wu et al., 2021b)	0.6720	HGIR-MRF (Sun et al., 2022)	0.779	**FD-MCD**	**0.8862**
ASDNN (Zhao et al., 2017)	0.6218	PSGM (Sun et al., 2021c)	0.7438	DCNet (Gao et al., 2019)	0.8833
SCCN (Liu et al., 2018a)	0.6154	CACD (Wu et al., 2021b)	0.7320	DCNN (Li et al., 2019)	0.8709
SMCFCA (Wu et al., 2021b)	0.5532	X-Net (Luppino et al., 2022b)	0.696	SARDNN (Gong et al., 2016)	0.8692
PCC (Liu et al., 2018a)	0.5064	ACE-Net (Luppino et al., 2022b)	0.689	SCCN (Liu et al., 2018a)	0.8438
Sardinia	KC	Texas	KC		
FD-MCD	**0.7142**	UIR (Luppino et al., 2022b)	0.914		
ALSC (Lei et al., 2022a)	0.6983	CAAE (Luppino et al., 2022a)	0.885		
PSGM (Sun et al., 2021c)	0.6821	**FD-MCD**	**0.8797**		
CAAE (Luppino et al., 2022a)	0.598	X-Net (Luppino et al., 2022b)	0.767		
AFL-DSR (Touati et al., 2020a)	0.5596	ACE-Net (Luppino et al., 2022b)	0.720		
RMN (Touati et al., 2019a)	0.3668	KCCA (Volpi et al., 2015)	0.65		

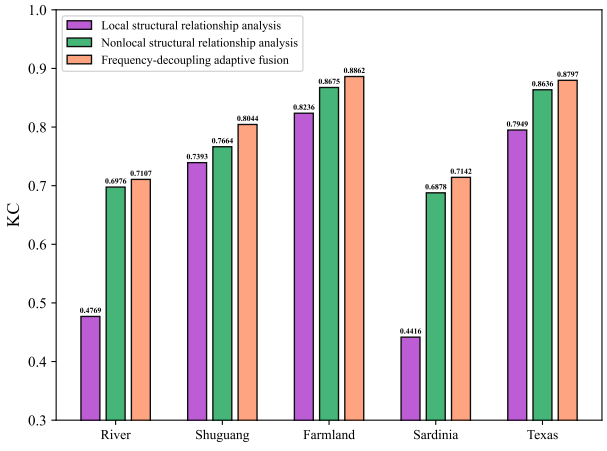


Figure 13: The performance of different steps in the proposed framework.

of w varies across datasets according to the type of modality combination, the spatial resolution of images, and the scale of change events in the dataset.

To further verify our motivation and the effectiveness of our metric, we compare it with directly performing difference operation on the original domain (SP-Diff) and two representative local similarity metrics calculated in the original domain of multimodal images, i.e., local spatial-temporal gradient (LSTG) (Touati et al., 2019a), affinity matrix distance (AMD) (Luppino et al., 2019, 2022b). Figure 15 shows the difference maps and binary change maps obtained by different local structural similarity metrics. Table 9 further lists the KC and time complexity of these metrics.

Firstly, we compare the detection results of SP-Diff and our metric. Due to the modal heterogeneity, directly comparing images in the original domain cannot accurately detect changed areas, with KCs of 0.1610 and 0.4594 on the two datasets. In comparison, our metric accurately detects most of the changed areas based on the observation that the modal

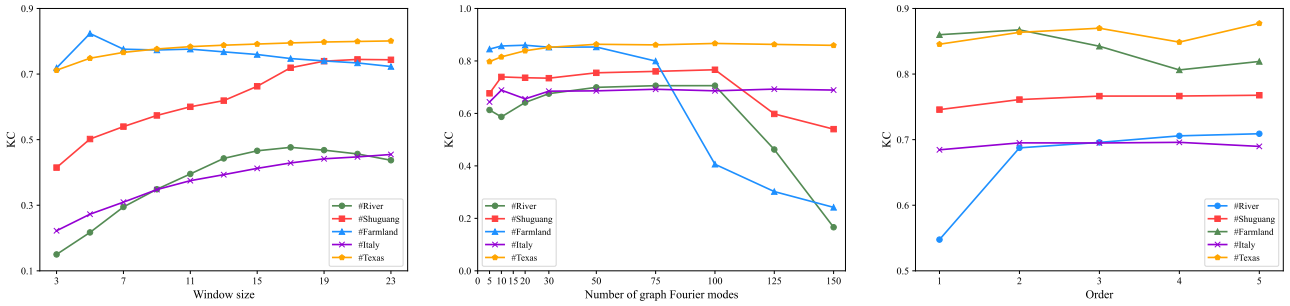
Table 9
Comparison in KC of different kinds of local structural similarity metrics on the two datasets. Here, SP-Diff adopts the same window size as our metric for a fair comparison.

Method	Shuguang	Texas	Time complexity
SP-Diff	0.1610	0.4594	$O(HW)$
LSTG	0.2323	0.4827	$O\left(HWN^2\right)$
AMD	0.2979	0.6125	$O\left(HWN^4\right)$
Ours	**0.7393**	**0.7949**	$O\left(HWN^2\log(N)\right)$

heterogeneity will be greatly reduced in the Fourier domain. Figure 16 visualizes the distributions of multimodal data before and after Fourier transform. Clearly, the distributions of the multimodal images would become close after being transformed into the Fourier domain, even though they vary considerably in their original domain.

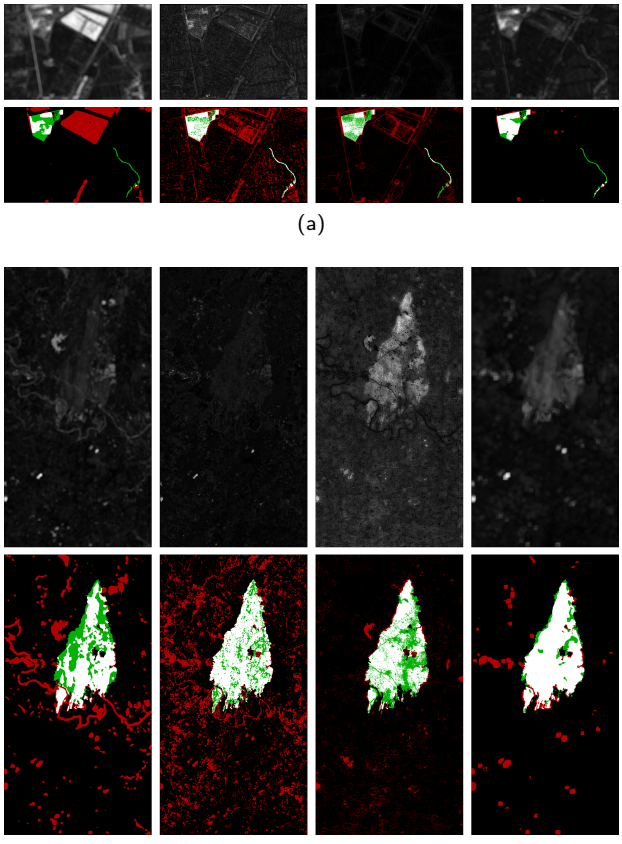
Then, the two local structural similarity metrics achieve better detection results than directly performing the difference operation in the original domain. However, they still cannot compete with our metric. Compared with their results, the change maps obtained by our metric with less noise and more intact changed areas. We also analyze the time complexity of these metrics. The three metrics perform a sliding window operation on the image. LSTG computes the local spatial gradient within each window and has a time complexity of $O\left(N^2\right)$. AMD, on the other hand, needs to compute the Gaussian kernel distance between each pixel and all other pixels within each window and thus has a time complexity of $O\left(N^4\right)$. Our metric performs a fast Fourier transform on the pixels within the window and thus has a time complexity of $O\left(N^2\log(N)\right)$.

In addition, the detection results obtained by these metrics can serve as change prior for subsequent change detection models to get more accurate results, including image regression (Luppino et al., 2019), probability graph models (Touati et al., 2020b), and deep learning models (Luppino et al., 2022a,b). Here, we further report the performance of



(a) (b) (c)

Figure 14: The relationship between the values of hyperparameters and change detection performance. (a) The size of window $\mathcal{W}$ in local frequency metric. (b) The number of graph Fourier basis S in graph spectral convolution. (c) The order of Chebyshev polynomials K in the transfer function.



(a)

(b)

Figure 15: The difference maps and change maps obtained by different similarity metrics on the (a) Shuguang dataset and (b) Texas dataset. From left to right: results obtained by spectral difference, local spatial-temporal gradient difference, affinity matrix distance, and our local frequency consistency.

our metric as the change prior to a subsequent image regression model. We follow the setting in AMD work (Luppino et al., 2019), adopting a random forest regression model to project one image from one modality to another based on the unchanged samples obtained by our metric. Table 10 shows that the performance of our metric outperforms the AMD

Table 10
Comparison of our local frequency consistency metric and affinity matrix distance (Luppino et al., 2019, 2022b) as change priors for image regression-based multimodal change detection on the Texas dataset

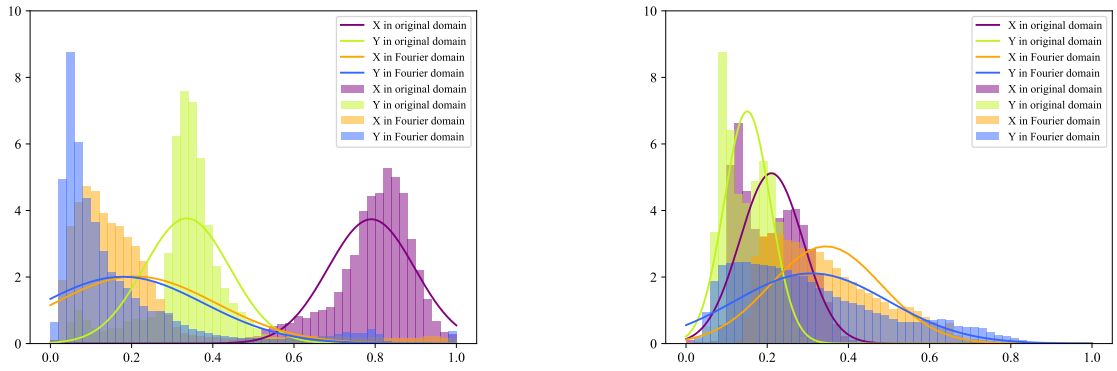
Method	Regression score	KC
AMD	0.9830	0.8553
Ours	0.9823	0.9166

Table 11
Comparison in KC of directly compare (DC) the one-order KNN graphs (Sun et al., 2021a,b) and our analyzing nonlocal structural relationship graph on the graph Fourier domain with graph spectral convolution

Method	River	Shuguang	Farmland	Sardinia	Texas
DC	0.6178	0.6551	0.8270	0.6297	0.8111
Ours	0.6976	0.7664	0.8675	0.6878	0.8636

used as a change prior to the regression model on the Texas dataset. In addition to image regression, our metric can also be used in other deep learning models like feature learning-based methods. Nevertheless, this is beyond the scope of this paper and can be investigated in subsequent work.

Moreover, it is quite challenging to design a metric that can work well for any modality combination. Therefore, our metric may not work in some cases even though it outperforms some representative local structural similarity metrics. This is because the underlying assumption of our local frequency consistency metric is that there is a strong similarity in the spatial distribution (i.e., frequency) of the same land-cover objects in two images with different modalities. Therefore, if the modal heterogeneity between the two images is so significant that this similarity is weak or non-existent, our metric may not yield accurate detection results.



(a) (b)

Figure 16: Comparison in data distribution in the original domain and Fourier domain on the (a) Shuguang dataset and (b) Texas dataset.

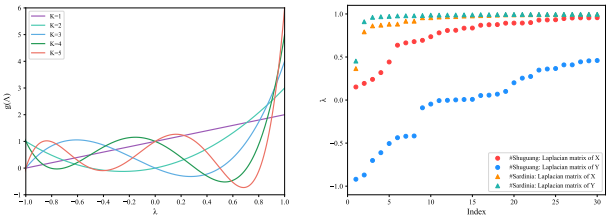
4.4.2. Analysis of graph spectral convolution for nonlocal structural relationship

As mentioned in Section 1.2 and 3.2, these self-similarity-based methods (Sun et al., 2021a,b,c) measure the nonlocal structural similarity by simply comparing the vertex information or edge information of one-order KNN graph. In comparison, our framework constructs a whole nonlocal structural relationship graph to allow the message passing of high-order vertex information and employs graph spectral convolution to extract graph structural information on the graph Fourier domain comprehensively. Table 11 lists the performance of simply comparing one-order KNN graph (Sun et al., 2021a) and our method. Obviously, our method outperforms the simple KNN graph comparison on the five datasets, demonstrating our motivation's effectiveness.

In addition, there are two main parameters in graph spectral convolution designed in our framework, namely the number of graph Fourier basis S and the order of Chebyshev polynomials K.

The number of graph Fourier basis S is an important hyperparameter affecting graph spectral convolution. For graph spectral convolution in the Fourier domain, the eigen-decomposition helps us understand the underlying structure of the graph. Smaller eigenvalues can better explain the structure of the graph data in graph spectral convolution (Saerens et al., 2004). When the number of filters increases to 50, the graph spectral convolution works very well on all five datasets, as shown in Figure 14-(b). The corresponding first 50 eigenvalues on two datasets ordered from small to large are also shown in Figure 17-(b).

K indicates the message passing range for each vertex in the nonlocal structural relationship graph, i.e., within a K-hop local neighborhood of each vertex. The optimal value of K varies on different datasets. Nevertheless, in Figure 14-(c), what is consistent is that as K increases from 1 to 2, the performance of the nonlocal graph spectral convolution increases. It demonstrates the effectiveness of our motivation for extracting high-order vertex information for nonlocal structural relationship analysis.



(a) (b)

Figure 17: (a) Distribution of Chebyshev polynomials for transfer functions and (b) the first fifty eigenvalues of the Laplacian matrix of the global KNN graph on the two datasets.

Table 12
Comparison in KC of direct fusion and frequency decoupling adaptive fusion on the five datasets

Method	River	Shuguang	Farmland	Sardinia	Texas
Direct fusion	0.6952	0.7887	0.8728	0.6772	0.8766
Ours	0.7107	0.8044	0.8862	0.7142	0.8797

4.4.3. Image fusion

Table 12 lists the KCs obtained by our frequency-decoupling adaptive fusion and direct fusion on the five datasets. We could see that the performance of our method is obviously better than direct fusion. This means that decoupling the low- and high-frequency of the difference map with the discrete Fourier transform and fusing each component separately can more effectively fuse the local and nonlocal structural difference information. In addition, on the Sardinia dataset, direct fusion leads to a reduction in final accuracy (0.6878→0.6772) due to the not very high quality of the local structural difference map. In contrast, our

Table 13
Computational time (in seconds) of six comparison models and our framework. Note that as the source codes for the comparison methods are not in the same programming environment, the computational times presented here are for approximate comparison.

Datasets	Image Size	RIF	FPMS	M3CD	CCLMRF	NPSG	IRGMcS	FD-MCD			
								t_{lcl}	t_{nlcl}	t_{fuse}	t_{total}
Farmland	291×306	13.1	5.4	1178.3	21.6	227.3	3.1	5.9	17.6	3.3	26.8
Shuguang	593×921	20.8	13.2	1875.6	40.3	1980.8	29.9	38.3	69.4	20.6	128.3
Texas	1534×808	83.9	20.4	2316.3	46.7	2881.0	46.4	57.7	83.8	47.9	189.4

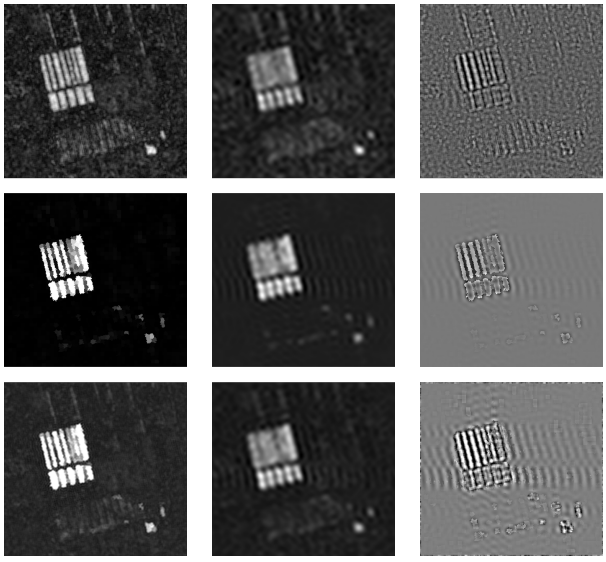


Figure 18: The illustration of components in the frequency-decoupling adaptive fusion on the Farmland dataset. (a) Difference map. (b) Low-frequency component. (c) High-frequency component. From top to bottom, the components of the local structural difference map, the nonlocal structural difference map, and the fused difference map are shown, respectively.

weighting strategy based on the variance of change intensity avoids this problem. Figure 18 shows each component of the local structural difference map, nonlocal structural difference map, and fused difference map on the Farmland dataset.

4.4.4. Time cost

Finally, the running times of FD-MCD and six comparison approaches on three datasets with different sizes are listed in Table 13. Among these methods, RIF, FPMS, M3CD, and CCLMRF are implemented with C++; NPSG and IRGMcS are implemented with MATLAB; our method is implemented with Python. Note that RIF, CCLMRF, M3CD, and FPMS adopt the resampling method to reduce the computational overhead. In Table 13, t_{lcl}, t_{nlcl}, and t_{fuse} represent the computational time spent in the local frequency consistency calculation, nonlocal structural relationship analysis, and frequency-decoupling adaptive fusion, respectively. It can be seen that the nonlocal structural relationship analysis on the graph Fourier domain is the most time-consuming of these three steps. Our whole framework is still a bit time-consuming compared to FPMS and IRGMcS, two approaches that have made improvements in efficiency. However, its computational overhead is acceptable given the good detection results obtained by our framework and the possibility that it can be implemented with a more efficient programming language, such as C++, and executed on more advanced devices.

5. Conclusion

This paper proposes a structural relationship analysis-based method called FD-MCD for unsupervised multimodal change detection. The three critical components of the proposed framework are implemented in the (graph) Fourier domain rather than in the original domain of remote sensing images as in existing methods. Firstly, a modality-independent local frequency consistency metric is presented to calculate the local structural difference between multimodal remote sensing images. Then, two nonlocal structural relationship weighted graphs are constructed for pre-change and post-change images. The structural information in the graphs is explored by parameter-free graph spectral convolution models. The nonlocal structural relationship in each graph is mapped into the other to eliminate the modal heterogeneity. The obtained graph representations are then compared to get the nonlocal structural difference map. Finally, the local and nonlocal structural difference maps are transformed into Fourier domain to decouple the low- and high-frequency components. The low- and high-frequency components will be fused according to their respective rules and then combined to obtain the final difference map. Detailed experiments and analyses demonstrate the effectiveness of our entire framework and each of its components in unsupervised multimodal change detection.

Acknowledgements

This work was supported in part by the JSPS, KAKENHI under Grant Number 22H03609 and JST, FOREST under Grant Number JPMJFR206S.

References

Adriano, B., Yokoya, N., Xia, J., Miura, H., Liu, W., Matsuoka, M., Koshimura, S., 2021. Learning from multimodal and multitemporal earth observation data for building damage mapping. ISPRS J. Photogramm. Remote Sens. 175, 132–143.

Baatz, M., 2000. Multi resolution segmentation: an optimum approach for high quality multi scale image segmentation, in: Beutrage zum AGIT-Symposium. Salzburg, Heidelberg, 2000, pp. 12–23.

Bruna, J., Zaremba, W., Szlam, A., LeCun, Y., 2013. Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203 .

Bruzzone, L., Diego Fernàndez Prieto, 2000. Automatic Analysis of the Difference Image for Unsupervised Change Detection. IEEE Trans. Geosci. Remote Sens. 38, 1171–1182.

Buades, A., Coll, B., Morel, J.M., 2005. A non-local algorithm for image denoising, in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 60–65.

Buades, A., Coll, B., Morel, J.M., 2010. Image denoising methods. a new nonlocal principle. SIAM review 52, 113–147.

Camps-Valls, G., Gómez-Chova, L., Muñoz-Marí, J., Rojo-Álvarez, J.L., Martínez-Ramón, M., 2008. Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. IEEE Trans. Geosci. Remote Sens. 46, 1822–1835.

Chen, H., Wu, C., Du, B., Zhang, L., Wang, L., 2020. Change Detection in Multisource VHR Images via Deep Siamese Convolutional Multiple-Layers Recurrent Neural Network. IEEE Trans. Geosci. Remote Sens. 58, 2848–2864.

Chen, H., Yokoya, N., Wu, C., Du, B., 2022. Unsupervised Multimodal Change Detection Based on Structural Relationship Graph Representation Learning. IEEE Trans. Geosci. Remote Sens. , 1–18.

Chung, F.R., 1997. Spectral graph theory. volume 92. American Mathematical Soc.

Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K., 2007. Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Trans. Image Process. 16, 2080–2095.

Du, P., Liu, S., Xia, J., Zhao, Y., 2013. Information fusion techniques for change detection from multi-temporal remote sensing images. Information Fusion 14, 19–27.

Gao, Y., Gao, F., Dong, J., Wang, S., 2019. Change Detection From Synthetic Aperture Radar Images Based on Channel Weighting-Based Deep Cascade Network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 4517–4529.

Gil-Yepes, J.L., Ruiz, L.A., Recio, J.A., Balaguer-Beser, Á., Hermosilla, T., 2016. 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. 121, 77–91.

Gong, M., Zhao, J., Liu, J., Miao, Q., Jiao, L., 2016. Change Detection in Synthetic Aperture Radar Images Based on Deep Neural Networks. IEEE Trans. Neural Netw. Learn. Syst. 27, 125–138.

Han, T., Tang, Y., Yang, X., Lin, Z., Zou, B., Feng, H., 2021. Change detection for heterogeneous remote sensing images with improved training of hierarchical extreme learning machine (HELM). Remote Sens. 13.

Hou, X., Bai, Y., Li, Y., Shang, C., Shen, Q., 2021. High-resolution triplet network with dynamic multiscale feature for change detection on satellite images. ISPRS J. Photogramm. Remote Sens. 177, 103–115.

Hussain, M., Chen, D., Cheng, A., Wei, H., Stanley, D., 2013. Change detection from remotely sensed images: From pixel-based to object-based approaches. ISPRS J. Photogramm. Remote Sens. 80, 91–106.

Ienco, D., Interdonato, R., Gaetano, R., Ho Tong Minh, D., 2019. Combining sentinel-1 and sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture. ISPRS J. Photogramm. Remote Sens. 158, 11–22.

Jiang, X., Li, G., Liu, Y., Zhang, X.P., He, Y., 2020. Change Detection in Heterogeneous Optical and SAR Remote Sensing Images Via Deep Homogeneous Feature Fusion. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 1551–1566.

Jimenez-Sierra, D.A., Benítez-Restrepo, H.D., Vargas-Cardona, H.D., Chanussot, J., 2020. Graph-based data fusion applied to: Change detection and biomass estimation in rice crops. Remote Sens. 12.

Jimenez-Sierra, D.A., Quintero-Olaya, D.A., Alvear-Muñoz, J.C., Benítez-Restrepo, H.D., Florez-Ospina, J.F., Chanussot, J., 2022. Graph learning based on signal smoothness representation for homogeneous and heterogeneous change detection. IEEE Trans. Geosci. Remote Sens. 60, 1–16.

Kipf, T.N., Welling, M., 2016. Semi-Supervised Classification with Graph Convolutional Networks, pp. 1–14. arXiv:1609.02907.

Kwan, C., Ayhan, B., Larkin, J., Kwan, L., Bernabé, S., Plaza, A., 2019. Performance of change detection algorithms using heterogeneous images and extended multi-attribute profiles (emaps). Remote Sens. 11.

Lei, L., Sun, Y., Kuang, G., 2022a. Adaptive Local Structure Consistency-Based Heterogeneous Remote Sensing Change Detection. IEEE Geosci. Remote Sens. Lett. 19.

Lei, T., Wang, J., Ning, H., Wang, X., Xue, D., Wang, Q., Nandi, A.K., 2022b. Difference enhancement and spatial–spectral nonlocal network for change detection in vhr remote sensing images. IEEE Trans. Geosci. Remote Sens. 60, 1–13.

Li, H., Gong, M., Zhang, M., Wu, Y., 2021. Spatially Self-Paced Convolutional Networks for Change Detection in Heterogeneous Images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 4966–4979.

Li, Y., Peng, C., Chen, Y., Jiao, L., Zhou, L., Shang, R., 2019. A Deep Learning Method for Change Detection in Synthetic Aperture Radar Images. IEEE Trans. Geosci. Remote Sens. 57, 5751–5763.

Liu, J., Gong, M., Qin, K., Zhang, P., 2018a. A Deep Convolutional Coupling Network for Change Detection Based on Heterogeneous Optical and Radar Images. IEEE Trans. Neural Netw. Learn. Syst. 29, 545–559.

Liu, Z., Li, G., Mercier, G., He, Y., Pan, Q., 2018b. Change Detection in Heterogenous Remote Sensing Images via Homogeneous Pixel Transformation. IEEE Trans. Image Process. 27, 1822–1834.

Liu, Z.G., Mercier, G., Dezert, J., Pan, Q., 2014. Change detection in heterogeneous remote sensing images based on multidimensional evidential reasoning. IEEE Geosci. Remote Sens. Lett. 11, 168–172.

Luppino, L.T., Bianchi, F.M., Moser, G., Anfinse, S.N., 2019. Unsupervised Image Regression for Heterogeneous Change Detection. IEEE Trans. Geosci. Remote Sens. 57, 9960–9975.

Luppino, L.T., Hansen, M.A., Kampffmeyer, M., Bianchi, F.M., Moser, G., Jenssen, R., Anfinsen, S.N., 2022a. Code-aligned autoencoders for unsupervised change detection in multimodal remote sensing images. IEEE Trans. Neural Netw. Learn. Syst. , 1–13.

Luppino, L.T., Kampffmeyer, M., Bianchi, F.M., Moser, G., Serpico, S.B., Jenssen, R., Anfinsen, S.N., 2022b. Deep Image Translation with an Affinity-Based Change Prior for Unsupervised Multimodal Change Detection. IEEE Trans. Geosci. Remote Sens. 60. arXiv:2001.04271.

Mignotte, M., 2020. A Fractal Projection and Markovian Segmentation-Based Approach for Multimodal Change Detection. IEEE Trans. Geosci. Remote Sens. 58, 8046–8058.

Mignotte, M., 2022. Mrf models based on a neighborhood adaptive class conditional likelihood for multimodal change detection. AI, Computer Science and Robotics Technology .

Moon, T.K., 1996. The expectation-maximization algorithm. IEEE Signal processing magazine 13, 47–60.

Niu, X., Gong, M., Zhan, T., Yang, Y., 2019. A Conditional Adversarial Network for Change Detection in Heterogeneous Images. IEEE Geosci. Remote Sens. Lett. 16, 45–49.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics 9, 62–66.

Park, J., Lee, M., Chang, H.J., Lee, K., Choi, J.Y., 2019. Symmetric graph convolutional autoencoder for unsupervised graph representation learning, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6519–6528.

Piella, G., 2003. A general framework for multiresolution image fusion: from pixels to regions. Information fusion 4, 259–280.

Proakis, J.G., 2001. Digital signal processing: principles algorithms and applications. Pearson Education India.

Radoi, A., 2022. Generative Adversarial Networks under CutMix Transformations for Multimodal Change Detection. IEEE Geosci. Remote Sens. Lett. 19, 1–5.

Saerens, M., Fouss, F., Yen, L., Dupont, P., 2004. The principal components analysis of a graph, and its relationships to spectral clustering, in: European conference on machine learning, Springer. pp. 371–383.

Shuman, D.I., Narang, S.K., Frossard, P., Ortega, A., Vandergheynst, P., 2013. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Processing Magazine 30, 83–98.

Stein, E.M., Shakarchi, R., 2010. Complex analysis. volume 2. Princeton University Press.

Sun, Y., Lei, L., Guan, D., Kuang, G., 2021a. Iterative Robust Graph for Unsupervised Change Detection of Heterogeneous Remote Sensing Images. IEEE Trans. Image Process. 30, 6277–6291.

Sun, Y., Lei, L., Li, X., Sun, H., Kuang, G., 2021b. Nonlocal patch similarity based heterogeneous remote sensing change detection. Pattern Recognit. 109, 1–16.

Sun, Y., Lei, L., Li, X., Tan, X., Kuang, G., 2021c. Patch Similarity Graph Matrix-Based Unsupervised Remote Sensing Change Detection with Homogeneous and Heterogeneous Sensors. IEEE Trans. Geosci. Remote Sens. 59, 4841–4861.

Sun, Y., Lei, L., Tan, X., Guan, D., Wu, J., Kuang, G., 2022. Structured graph based image regression for unsupervised multimodal change detection. ISPRS J. Photogramm. Remote Sens. 185, 16–31.

Touati, R., Mignotte, M., Dahmane, M., 2019a. A Reliable Mixed-Norm-Based Multiresolution Change Detector in Heterogeneous Remote Sensing Images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 3588–3601.

Touati, R., Mignotte, M., Dahmane, M., 2019b. Multimodal change detection using a convolution model-based mapping, in: Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA), pp. 1–6.

Touati, R., Mignotte, M., Dahmane, M., 2020a. Anomaly Feature Learning for Unsupervised Change Detection in Heterogeneous Images: A Deep Sparse Residual Model. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 13, 588–600.

Touati, R., Mignotte, M., Dahmane, M., 2020b. Multimodal Change Detection in Remote Sensing Images Using an Unsupervised Pixel Pairwise-Based Markov Random Field Model. IEEE Trans. Image Process. 29, 757–767.

Vetrivel, A., Gerke, M., Kerle, N., Nex, F., Vosselman, G., 2018. Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images , and multiple-kernel-learning. ISPRS J. Photogramm. Remote Sens. 140, 45–59.

Volpi, M., Camps-Valls, G., Tuia, D., 2015. Spectral alignment of multitemporal cross-sensor images with automated kernel canonical correlation analysis. ISPRS J. Photogramm. Remote Sens. 107, 50–63.

Wan, L., Xiang, Y., You, H., 2019a. A Post-Classification Comparison Method for SAR and Optical Images Change Detection. IEEE Geosci. Remote Sens. Lett. 16, 1026–1030.

Wan, L., Xiang, Y., You, H., 2019b. An object-based hierarchical compound classification method for change detection in heterogeneous optical and SAR images. IEEE Trans. Geosci. Remote Sens. 57, 9941–9959.

Wan, L., Zhang, T., You, H.J., 2018. Multi-sensor remote sensing image change detection based on sorted histograms. Int. J. Remote Sens. 39, 3753–3775.

Wang, J., Gao, F., Dong, J., Zhang, S., Du, Q., 2022. Change Detection From Synthetic Aperture Radar Images via Graph-Based Knowledge Supplement Network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 1823–1836. arXiv:2201.08954.

Wang, X., Liu, S., Du, P., Liang, H., Xia, J., Li, Y., 2018. Object-based change detection in urban areas from high spatial resolution images based on multiple features and ensemble learning. Remote Sens. 10.

Wong, M.W., 2011. Discrete fourier analysis. volume 5. Springer Science & Business Media.

Wu, C., Chen, H., Du, B., Zhang, L., 2022. Unsupervised change detection in multitemporal vhr images based on deep kernel pca convolutional mapping network. IEEE Trans. Cybern 52, 12084–12098.

Wu, C., Du, B., Cui, X., Zhang, L., 2017a. A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion. Remote Sens. Environ. 199, 241–255.

Wu, C., Zhang, L., Du, B., 2017b. Kernel Slow Feature Analysis for Scene Change Detection. IEEE Trans. Geosci. Remote Sens. 55, 2367–2384.

Wu, J., Li, B., Qin, Y., Ni, W., Zhang, H., Fu, R., Sun, Y., 2021a. A multiscale graph convolutional network for change detection in homogeneous and heterogeneous remote sensing images. Int. J. Appl. Earth Obs. Geoinf. 105, 102615.

Wu, Y., Li, J., Yuan, Y., Qin, A.K., Miao, Q.G., Gong, M.G., 2021b. Commonality Autoencoder: Learning Common Features for Change Detection From Heterogeneous Images. IEEE Trans. Neural Netw. Learn. Syst. , 1–14.

Yang, M., Jiao, L., Liu, F., Hou, B., Yang, S., 2019. Transferred deep learning-based change detection in remote sensing images. IEEE Trans. Geosci. Remote Sens. 57, 6960–6973.

Zhan, T., Gong, M., Jiang, X., Li, S., 2018. Log-based transformation feature learning for change detection in heterogeneous images. IEEE Geosci. Remote Sens. Lett. 15, 1352–1356.

Zhang, P., Gong, M., Su, L., Liu, J., Li, Z., 2016. Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images. ISPRS J. Photogramm. Remote Sens. 116, 24–41.

Zhao, W., Wang, Z., Gong, M., Liu, J., 2017. Discriminative Feature Learning for Unsupervised Change Detection in Heterogeneous Images Based on a Coupled Neural Network. IEEE Trans. Geosci. Remote Sens. 55, 7066–7080.

Zhu, Z., 2017. Change detection using landsat time series: A review of frequencies, preprocessing , algorithms, and applications. ISPRS J. Photogramm. Remote Sens. 130, 370–384.