Iterative Global Mapping-Local Searching for Heterogeneous Change Detection with Unregistered Images

Yuli Sun[1,2] · **Junzheng Wu**[3] · **Han Zhang**[3] · **Zhang Li**[2] · **Lin Lei**[1] · **Gangyao Kuang**[1]

Received: 30 September 2024

Abstract Heterogeneous change detection (HeCD) is a highly valuable yet challenging task in remote sensing. However, existing HeCD methods primarily focus on well-registered images, without considering unregistered heterogeneous image pairs, which are more common in real-world applications. In this paper, we first analyze why unregistered images significantly complicate the HeCD problem: they not only cause boundary errors in change detection but also undermine the image transformation process required for making heterogeneous images comparable. These effects are further amplified by the intrinsic complexity of HeCD. In light of this, we propose an unsupervised iterative global mapping-local searching method (IGmLs) for HeCD subject to registration errors. Specifically, IGmLs utilizes the global mapping to transform images into a common structural space to enable the comparison of heterogeneous images, and uses the local searching to reduce the direct influence of registration errors on change metrics, which is based on the analysis that misregistration would cause an unwanted increment on the change metric in unchanged region. Then, IGmLs builds a Markov random field (MRF) model to combine the global mapping and local searching processes, which enhances the robustness to misregistrations by considering spatial correlations. Finally, an iterative framework is employed to backpropagate the matching and changing results to refine the global mapping and local searching processes, which further eliminates the indirect influence of changes and misregistration on the image transformation and change metrics. Extensive experiments on five datasets have demonstrated the effectiveness of the proposed IGmLs. The codes will be released at `https://github.com/yulisun/IGmLs`.

Keywords Change detection · heterogeneous data · registration errors · multimodal · remote sensing images

1 Introduction

Change detection (CD) of remote sensing images refers to the technique of extracting change information by comparing multiple images acquired in the same geographical area but at different times [51]. Remote sensing observation technology is capable of monitoring the Earth's surface over extended periods, on a large scale, and at regular intervals. Consequently, CD is one of the earliest and most widely used research topics in remote sensing technology [76, 9], which has been extensively applied in the fields of environmental monitoring [65], agricultural surveys [36], urban studies [1], and disaster assessment [77].

With the advancement of aerospace and imaging techniques, an increasing number of remote sensing images can be captured, providing richer data resources for CD techniques while simultaneously setting higher demands. For instance, CD techniques must be capable of detecting changes using not only homogeneous images (abbreviated as HoCD) captured by the same sensor [10], but also heterogeneous images (abbreviated as HeCD) captured by different sensors [34, 56], to fully leverage the advantages of different sensors [60], such as synthetic aperture radar (SAR) and optical sensors. Compared to HoCD, HeCD is more flexible and tailored to practical needs, offering several advantages [54, 17, 11]. For example, HeCD can rapidly extract change information by utilizing any available pre- and post-change images from different sources when homogeneous images are unavailable due to imaging conditions such as weather and light [74, 41], which is particularly essential in the case of sudden events (such as earthquakes and floods). HeCD can also improve the temporal resolution and extend the timeframe of time series monitoring by incorporating heterogeneous images [3].

Despite the significant practical value of HeCD, it also presents additional challenges in that heterogeneous images are from different sources and possess distinctly different image characteristics, including varying spectral characteristics, radiometric properties, spatial resolutions, and so on.

First, the image heterogeneity prevents detecting changes by directly comparing images as in HoCD. To enable the comparison of heterogenous images, previous researches have been devoted to establish the correlation between images from different sources and transform them into the same

---

✉ Corresponding author: Yuli Sun, Gangyao Kuang
1 College of Electronic Science and Technology, National University of Defense Technology, China
2 College of Aerospace Science and Engineering, National University of Defense Technology, China
3 Northwest Institute of Nuclear Technology, China

domain for analysis [53, 18]. These efforts have facilitated the development of HeCD methods, which can be broadly categorized into three types according to the common domain of comparison [54]: (i) image classification-based methods that transform heterogeneous images into the same land cover category space, (ii) image translation-based methods that transform one image into the domain of another, (iii) and feature transformation-based methods that transform heterogeneous images into a common constructed or learned feature domain. Therefore, current research on HeCD primarily continues to cope with the following three challenges: how to establish universal associations between heterogeneous images to make them widely applicable to various HeCD tasks; how to achieve homogenized representations of heterogeneous images, i.e., how to accurately transform heterogeneous images into the same domain; and how to improve the difference metric's discriminability between changed/unchanged regions.

Second, achieving accurate geometric registration of heterogeneous images presents considerable challenges. In the context of CD problem, image registration serves as an essential preprocessing step before detecting changes [8]. It is commonly assumed in CD that the compared multitemporal images are already well-registered, where pixels occupying the same positions correspond to identical geographical areas. In practical applications, image registration and CD are typically treated as two independent processes [36]. Currently, the registration technique for homogeneous image is very mature, so the influence of the registration error is seldom considered separately in the HoCD problem. However, the heterogeneous image registration remains a prominent and challenging research topic [69]. Current algorithms for heterogeneous image registration are mostly tailored for small to medium-sized and simple scene images, making it arduous to achieve pixel-level registration accuracy in complex scenarios such as high-resolution, large-scale images with pronounced terrain variations [72]. Therefore, it is imperative to consider HeCD problem under the conditions with potential image registration errors.

Currently, only a few approaches have been proposed to simultaneously address the CD and registration problems, all of which focus solely on the HoCD. For SAR images, Nguyen et al. [46] and Song et al. [52] have proposed two optimization models for jointly tackling CD and registration tasks, where the former considers the shift transformation between SAR images and the latter considers the affine transformation. For multispectral images, Vakalopoulou et al. [61] have exploited a decomposed interconnected Markov random field (MRF) over two graphs, where one graph corresponds to the CD term and the other graph corresponds to the registration term. In [62], they have extended this approach to incorporate registration, CD, and classification within a graph-based MRF model.

To fill the gap and promote the practice application of HeCD, in this study we first analyze the influence of registration errors on HeCD. On one hand, misregistration may lead to change detection errors at the boundaries between different objects, such as false changes at boundaries due to the misalignment of identical objects as illustrated in Fig. 1, which also occurs in HoCD [7]. On the other hand, misregistration can introduce ambiguity in heterogeneous image transformation, thereby affecting the overall performance of HeCD. As previously mentioned, HeCD necessitates transforming heterogeneous images into a common domain to enable comparison, while most of these transformations are not pixel-wise independent, i.e., each pixel will be influenced by other pixels (e.g., neighboring pixels [43, 31] or nonlocal homogeneous pixels [57, 55]) during the transformation process. As a result, the unregistered regions further exacerbate the difficulty of translating heterogeneous images or transforming features, ultimately degrading the accuracy of HeCD.

To address these challenges, we propose a novel iterative global mapping-local searching based method (IGmLs) for HeCD considering image registration errors. Specifically, IGmLs employs two scales to segment images into superpixels: a fine scale for registration and change detection, and a coarse scale for image transformation (it is relatively more robust to registration errors). IGmLs then utilizes the global mapping approach to transform heterogenous images into a common structural space at coarse scales, which extracts the similarity relationships between fine-scale superpixels and each coarse-scale superpixel (can be regarded as feature template) as structure features, and utilizes graph mapping to compute the change metrics. Next, IGmLs uses a locale sliding window to search for the initial best matching position and minimum matching cost for each fine-scale superpixel. Subsequently, an MRF model is built to combine the results of global mapping and local searching to identify initial changed superpixels. Finally, an iterative framework backpropagates the initial matching and changing results to refine the aforementioned global mapping and local searching processes, which brings two benefits: eliminating the changed superpixels in the feature template to reduce the change confusion, and updating the matching position of feature template to enhance the accuracy of image transformation. Overall, IGmLs implements an iterative coarse-to-fine filtering and correction process.

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

- We conducted an analysis to elucidate how misregistration affects HeCD in terms of image transformation and feature comparison, and subsequently proposed an unsupervised iterative global mapping-local searching method for HeCD with unregistered images. To the best of our knowledge, this is the first work that provides a detailed analysis of the impact of registration errors on HeCD and proposes a method to address it.
- We proposed an iterative coarse-to-fine filtering framework that involves global mapping to enable the comparison of heterogeneous images, uses local searching to reduce the direct influence of misregistration on the change metrics, and utilizes the iterative back propagation to further eliminate the indirect influence of changes and registration errors on the image transformation and change metrics.
- We built an MRF model that fuses the global mapping and local searching results, and incorporate constraints such as spatial consistency, prior sparsity, change and displacement constraints into the MRF model, thereby enhancing the robustness of change detection against misregistrations.

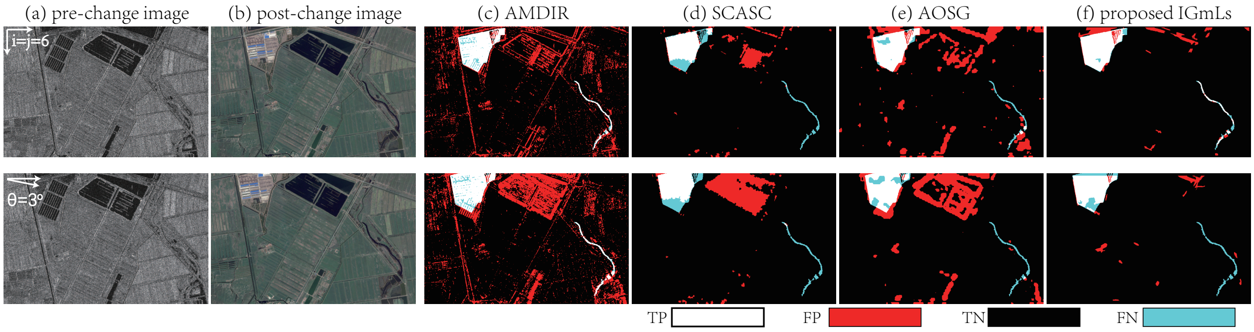


Fig. 1: An example of HeCD with unregistered images. (a) and (b) correspond to the pre- and post-change images, where exists a registration error of 6 pixels each horizontally and vertically in the first row, and a registration error of 3° rotation in the second row. (c)-(f) correspond to the detection results generated by AMDIR [31], SCASC [55], AOSG [20] and the proposed IGmLs. In the change map, white, red, black, and cyan mark true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), respectively.

- We conducted extensive experiments with state-of-the-art methods on five datasets and validated the effectiveness of the proposed method. The codes will be released at `https://github.com/yulisun/IGmLs`.

2 Related work

This paper focuses on the problem of heterogeneous change detection in the presence of image registration errors. Therefore, this section reviews related work on heterogeneous change detection, multimodal image matching, and change detection with misregistration.

2.1 Heterogeneous change detection

Heterogeneous images often exhibit different imaging characteristics and show different representations for the same object. Therefore, the primary task of HeCD is to transform heterogeneous images into a common domain to make them comparable, and existing HeCD methods can be broadly categorized into three types: image classification, image translation, and feature transformation-based methods.

1) Image classification-based methods transform heterogeneous images into the same land cover class space by using classifiers, and then compare the classification results to detect changes, such as the post-classification comparison method [49], evidence reasoning-based method [29], hierarchical compound classification method [63], hierarchical extreme learning machine-based classification network [19]. While these methods are relatively intuitive and can be flexibly designed for various application tasks, they still face some limitations: (i) error accumulation during the classification process of heterogeneous images will lead to a decrease in change detection accuracy; (ii) classification algorithms often rely on large amounts of high-quality training samples, which increases the cost of manual annotation.

2) Image translation-based methods transform heterogeneous images into the same image domain as either the pre-change or post-change image, and then compare the homogeneous original image and translated image. Examples include the homogeneous pixel transformation method [28], fractal project method [43], affinity matrix distance based image regression [31], and structured graph based regression method [58, 57], which achieve pixel/superpixel conversion in unchanged regions by manually constructing correlations between the heterogeneous images. Some other methods perform image style migration with the help of generative adversarial networks (GAN) or their variants [39], such as the robust fusion-based GAN [64], copula mixtures base CycleGAN [23], and multi-domain constrained heterogeneous translation network [66]. Through image transformation, the diversity of image data before and after a change event can be enhanced, providing richer visual information and more comprehensive change information [33, 32]. However, these algorithms primarily focus on the spectral/radiometric characteristics of the images, making it challenging to establish an accurate transformation model for heterogeneous images and escape the influence of image noise.

3) Feature transformation-based methods transform the heterogeneous images into a common feature domain, and then identify the changes by comparing the extracted features. For example, the traditional similarity metrics-based methods often manually extract common features from heterogeneous images and calculate their differences to measure changes, such as the spatial self-similarity feature [81], topological structure feature [56], local and non-local Fourier domain features [12]. Some deep learning-based methods usually employ the Siamese or pseudo-Siamese networks to transform heterogeneous images into the same latent feature space [68, 27], such as the self-guided autoencoders [50], iterative training sample augmentation based network [35], structural graph convolutional autoencoder [13], object-guided transformer [11]. This kind of methods can leverage the powerful learning capabilities of deep neural networks to construct a common feature space [67], and they can also utilize the general frameworks or foundational large models from computer vision (CV) or natural language processing (NLP) to design networks that are well-suited for remote sensing change detection task [78, 75].

Although existing HeCD methods have gained considerable progress and achieved relatively good change detection performance, they all assume that the registration of heterogeneous images is safe from errors. However, this assumption often contradicts practical application scenarios.

2.2 Multimodal image registration

Automatic registration of multimodal remote sensing images (e.g., optical and SAR images) remains a challenging problem due to the differences in geometry and intensity [37, 38]. In the last few years, a number of multimodal image registration methods have been developed, which can be roughly partitioned into two categories: feature-based registration methods and area-based registration methods.

1) Feature-based registration methods usually extract the salient features from images and establish feature correspondences according to the defined matching criterion, which mainly contain four steps: detecting keypoints, constructing descriptors, matching features, and computing transformation parameters [25]. In these methods, the scale-invariant feature transform (SIFT) [30] and its variations are widely used for keypoint detection, such as SAR-SIFT [15], principal component analysis SIFT (PCA-SIFT) [22], Affine-SIFT [45], and adaptive binning SIFT [48]. To address the nonlinear radiation distortions between multimodal images, some researchers have explored the phase congruency between images, such as the radiation-variation insensitive feature transform (RIFT) [24] and locally normalized image feature transform (LNIFT) [26]. The essence of feature-based registration methods lies in the abstract representation of multimodal images by extracting features, thereby achieving efficient registration. These methods can be applied to image pairs with significant geometric distortions, but their registration accuracy is limited by the performance of keypoint detection and feature description.

2) Area-based registration methods, also known as template matching, search for the patch with the highest similarity to the template within a sliding window. The key of these methods lies in the similarity metric. Commonly used similarity metrics mainly include sum of squared difference (SSD), normalized cross correlation (NCC), and mutual information (MI). In addition to spatial domain-based metrics, some phase correlation methods utilize the Fourier transform to measure similarity in the frequency domain, such as the histogram of orientated phase congruency (HOPC) [70] and channel features of orientated gradients (CFOG) [69]. Some learning-based methods utilize deep neural networks to extract high-level features and then find the optimal match using traditional similarity metric [71, 79], while some other approaches directly searching matches using end-to-end mode [73]. Generally, area-based registration methods are relatively intuitive and simple, without requiring complex pre-processing of the original multimodal images. However, this type of methods are sensitive to nonlinear transformations, and require georeferencing to improve registration accuracy. Furthermore, due to the exhaustive search in optimal matching, these methods require significant memory expenditure and have high computational complexity.

As previously discussed, some HeCD methods and image registration methods require calculating the similarity between heterogeneous images or patches, such as feature transformation-based HeCD and template matching-based image registration. Therefore, they share certain common requirements in terms of feature extraction and similarity metrics.

2.3 Methods combining change detection and registration

Currently, only a few studies have focused on the CD under misregistration conditions. In [4], Bovolo et al. have proposed a multiscale strategy to reduce registration noise in CD of multispectral images, which exploits the magnitude and direction information of spectral change vectors at different resolution levels, and uses the spectral difference image calculated at the lower resolution to reduce registration noise in CD at higher resolutions. Sundaresan et al. [59] have evaluated two CD algorithms (image differencing algorithm and MRF-based algorithm) in the presence of registration errors. Their experimental results indicate that the spatial correlation-based MRF model is more robust than image differencing, with the maximum misregistration limited to 1 pixel of root mean square error (RMSE).

A few other methods have considered jointly solving the registration and HoCD problem. Nguyen et al. [46] have proposed a sparsity-driven joint image registration and CD method for SAR images, which constructs a dictionary by using the searched patches in the reference image, then finds the best matching by sparse representation and takes the prediction error as the change measurement. Song et al. [52] have introduced an optimization model for the CD and registration of SAR images, which decomposes the sensed image into a wrapped image with affinity transformation, a sparse changed image, and image noise. For HoCD of multispectral images, Vakalopoulou et al. [61, 62] have proposed a decomposed interconnected graphical model, which induced an MRF-based energy model consisting of a registration energy term, a change detection energy term, and a coupling energy term. They have exploited eleven similarity metrics for the registration and CD, such as the SSD, NCC, NMI, and others. However, all of these similarity metrics are only suitable for homogeneous multispectral images but not for heterogeneous images. Diego et al. [40] have extended this MRF framework to multi-sensor scenes and proposed a mid-level sensor-invariant representation for multimodal registration. In [42], Mesquita et al. have employed convolutional neural networks (CNN) for scene-level CD in optical aerial images with registration errors, which extracts shared features for both CD and registration from a Siamese CNN encoder. Recently, Zhou et al. [80] have proposed a unified network for the two independent tasks of CD and registration from the purpose of reducing training and memory resources, but this network does not account for CD in the presence of registration errors.

Consequently, it can be seen that current research focuses only on the HoCD with misregistration, neglecting the more complex HeCD with misregistration. To fill this gap, this paper proposes a novel iterative global mapping-local searching method for HeCD with unregistered images.

3 The influence of misregistration on HeCD

Given two multitemporal remote sensing images of $\tilde{\mathbf{X}} \in \mathbb{R}^{M_1 \times N_1 \times B_1}$ and $\tilde{\mathbf{Y}} \in \mathbb{R}^{M_2 \times N_2 \times B_2}$, with pixels denoted as $\tilde{x}_{m,n,b}$ and $\tilde{y}_{m,n,b}$ respectively, acquired by different sensors over the same area at times t_1 and t_2, which are coarsely registered by utilizing the inherent geographical coordinate

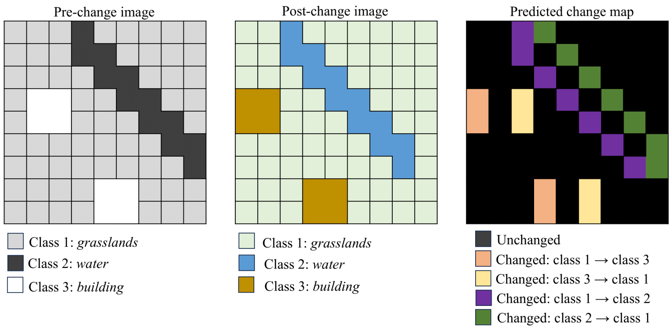


Fig. 2: A example about the influence of registration errors on HeCD. There are three classes of objects in pre- and post-change images, and no change between two images. Due to the 1-pixel registration error, 4 kinds of false changes appear in the change map.

references of remote sensing images. Therefore, there are no significant translations, rotations, or scale differences between $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$.

In order to analyze the influence of registration errors in HeCD, we use a simple example shown in Fig. 2, where there is a 1-pixel translation error between the pre-change and post-change images. On the one hand, from Fig. 2 we can find that misregistration will result in change detection errors at the boundaries between different kinds of objects. Generally, the more fragmented the objects included in the images, the greater the error caused by misregistration. Furthermore, there are two additional implicit effects: **(i)** a significant increase in the number of change classes and in the variance of changed pixels due to the introduction of numerous false changes (e.g., pseudo-changes between different classes, as illustrated in Fig. 2), which affects the extraction and categorization of actual changes; and **(ii)** a general increase in the variance of unchanged pixels caused by imperfect alignment of pixels representing the same objects across multitemporal images, which leads to confusion between changed and unchanged regions.

Moreover, the specificity of the HeCD problem exacerbates the errors introduced by misregistration. In HeCD, heterogeneous images must be transformed into a common feature or image domain to enable meaningful comparison for detecting changes. However, most of these transformations are not performed independently pixel-by-pixel. Instead, each pixel is inevitably influenced by other pixels, such as those in local spatial neighborhoods or non-local spectrally similar regions. Consequently, unregistered pixels propagate their effects to others during the transformation process, further increasing the difficulty of feature alignment in heterogeneous images. This leads to an error amplification process, resulting in heightened confusion between changed and unchanged regions.

From the above analysis, the influence of misregistration on HeCD is full-process and highly complex, making the resulting change detection errors difficult to eliminate through post-processing. Its impact is not limited to misdetections at object boundaries (which can sometimes be alleviated) but, more critically, includes global false alarms in non-boundary regions, which are particularly difficult to remove and severely degrade the final results, as illustrated in Fig. 1. To address this issue, we propose an iterative coarse-to-fine filtering and correction framework. This approach uses a global mapping process to enable the comparison of heterogeneous images, utilizes a local searching process to mitigate the direct CD errors caused by misregistration, and then employs an iterative framework to progressively reduce the indirect errors arising from inaccurate image transformation during the global mapping process due to misregistration.

4 Methodology

The proposed IGmLs method for HeCD with unregistered images is illustrated in Fig. 3, which comprises four steps: 1) pre-processing, 2) global mapping for transforming heterogeneous images and calculating the change metrics, 3) local searching for finding the initial matching position and initial changed regions, 4) and iterative filtering through backpropagation.

4.1 Superpixel segmentation and feature extraction

With the development of positioning related technologies, there are no significant translations, rotations, or scale differences between coarsely registered images by using geographic coordinates. Therefore, the registration errors between pixels in heterogeneous images are usually within a controllable range of a few dozens of pixels.

We choose superpixel as the basic unit of analysis, which helps to reduce computational complexity and offers advantages in detecting meaningful changes. Additionally, the superpixel-based CD methods are more robust to the registration errors compared to pixel-based CD methods [21]. In this paper, we segment images into superpixels at two scales: a fine scale for registration and change detection, and a coarse scale for heterogeneous image transformation, which is relatively robust to registration errors.

The Gaussian mixture model base superpixel (GMMSP) method [2] is employed to segment the images, which efficiently generates superpixels that closely conform to object boundaries while maintaining linear complexity with respect to the number of pixels. For optical images consisting of RGB channels, the original GMMSP is directly applied; for multispectral images with more than three channels, the principle component analysis method (PCA) is utilized to reduce their channel dimension to three before performing GMMSP segmentation; for SAR images, the generalized likelihood ratio based distance in [14] for multiplicative speckle noise with Gamma distribution is used to replace the Euclidean distance in original GMMSP. Using GMMSP or modified GMMSP, we first segment the image $\tilde{\mathbf{X}}$ with two scales and obtain two segmentation maps of Λ^f (fine scale) and Λ^c (coarse scale), and two sets of N_f and N_c segmented superpixels with $N_f > N_c$, defined as

$$\begin{aligned}\tilde{\mathbf{X}}_i^f &= \left\{ \tilde{x}_{m,n,b} \middle| (m,n) \in \Lambda_i^f, b = 1, \cdots, B_1 \right\}, \ i = 1, \cdots, N_f, \\ \tilde{\mathbf{X}}_j^c &= \left\{ \tilde{x}_{m,n,b} | (m,n) \in \Lambda_j^c, b = 1, \cdots, B_1 \right\}, \ j = 1, \cdots, N_c.\end{aligned} \tag{1}$$

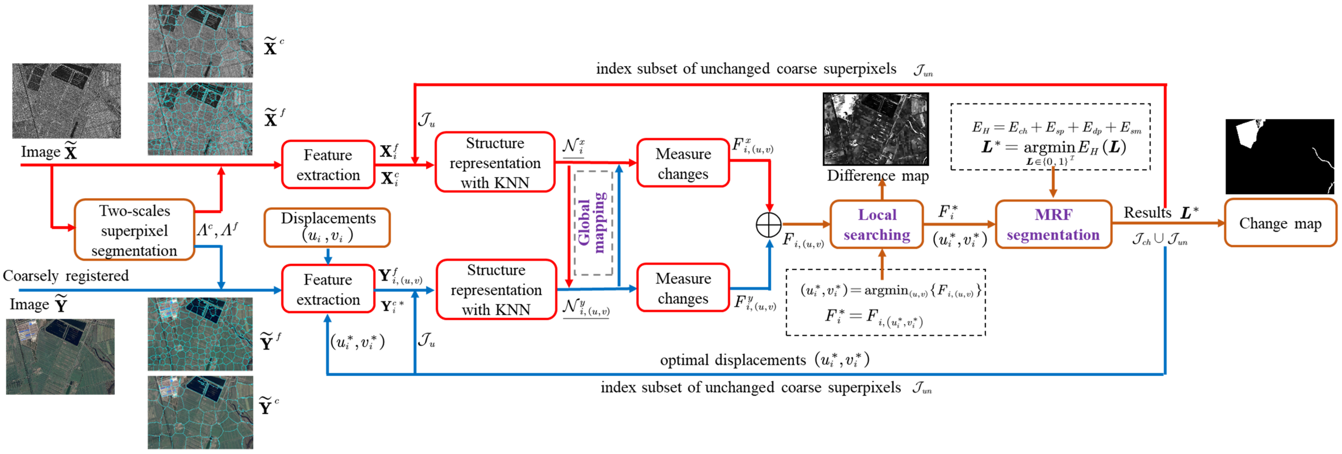


Fig. 3: Framework of the proposed IGmLs for HeCD with coarsely registered images.

We project these two superpixel segmentation maps Λ^f, Λ^c to the other image $\tilde{\mathbf{Y}}$, we have two mapped superpixels as

$$\begin{aligned} \tilde{\mathbf{Y}}_i^f &= \left\{ \tilde{y}_{m,n,b} \middle| (m,n) \in \Lambda_i^f, b = 1, \cdots, B_2 \right\}, \ i = 1, \cdots, N_f \\ \tilde{\mathbf{Y}}_j^c &= \left\{ \tilde{y}_{m,n,b} \middle| (m,n) \in \Lambda_j^c, b = 1, \cdots, B_2 \right\}, \ j = 1, \cdots, N_c \end{aligned} \tag{2}$$

Subsequently, for each superpixel, we extract the three quartiles, the mean, and the variance of the pixel values from each band as spectral features for simplicity (other feature choices are also possible). This allows us to obtain the feature matrices of $\mathbf{X}^f \in \mathbb{R}^{5B_1 \times N_f}$, $\mathbf{X}^c \in \mathbb{R}^{5B_1 \times N_c}$, $\mathbf{Y}^f \in \mathbb{R}^{5B_2 \times N_f}$, and $\mathbf{Y}^c \in \mathbb{R}^{5B_2 \times N_c}$ for different sets of superpixels from different images, where each column represents the feature vector of the corresponding superpixel.

In this way, we have two superpixel sets of different scales: the fine scale set with N_f superpixels, and the coarse scale set with N_c superpixels. For the fine scale, the superpixels $\tilde{\mathbf{X}}_i^f$ have a high probability of being internally homogeneous, i.e., the pixels within each superpixel $\tilde{\mathbf{X}}_i^f$ represent the same object. However, the interior of the mapped superpixel $\tilde{\mathbf{Y}}_i^f$ may not be homogeneous due to two factors: object boundary errors caused by registration errors, and changes in the object itself. For the coarse scale, the pixels within each coarse superpixel $\tilde{\mathbf{X}}_j^c$ are likely to represent the same type of object, and differences in internal homogeneity degree between $\tilde{\mathbf{X}}_j^c$ and $\tilde{\mathbf{Y}}_j^c$ caused by registration errors are reduced. This is because coarse segmentation allows for robustness against misregistration than fine segmentation. From a certain perspective, multiscale superpixel segmentation can be viewed as down-sampling the image. For example, a 1-pixel shift in the original image can be seen as a subpixel shift of 0.1 with downscaling factor of 10, then the impact of pixel shift on the coarse scale is equivalent to only N_c/N_f of its impact on the fine scale.

Based on the characteristics of superpixels at different scales and the influence of misregistration on HeCD analyzed in Section 3, we employ the fine-scale superpixels as the basic units for registration and change detection by utilizing their more precise boundaries and finer object segmentation, and employ the coarse-scale superpixels for heterogeneous image transformation by utilizing their robustness against registration errors. Specifically, **(i)** the registration process can employ the more accurate edge contour information of fine-scale superpixels during template matching, by setting $\tilde{\mathbf{X}}_i^f$ as the template and sliding the $\tilde{\mathbf{Y}}_i^f$ within a search window in image $\tilde{\mathbf{Y}}$ to find the optimal matching; **(ii)** the change detection process can utilize the finer-scale superpixel segmentation results to detect more subtle land cover changes, thereby improving change detection accuracy; **(iii)** the image transformation process can use the coarse-scale superpixels that are more robust to misregistration to transform heterogeneous images into a common structural space, by calculating the similarity between each fine-scale superpixel and coarse-scale superpixel within the same image.

4.2 Global mapping for change metrics

In order to compare the heterogeneous images to detect changes, it is necessary to transform them into a common space. Our main hypothesis is based on the structural consistency property between heterogeneous images, which can be referred as follows [53, 56]: although the heterogeneous images acquired by different sensors over the same area are distinct in terms of their presentation, the similarity relationships between objects inside the images are consistent. Structural consistency signifies that object maintains the same similarity relationships across images captured by different sensors under varying conditions, except when it changes to another kind of object.

In this paper, we characterize the image topological structure by calculating the similarity between each fine-scale superpixel and each coarse-scale superpixel within the image. For instance, if both the i-th fine superpixel and j-th coarse superpixel ($\tilde{\mathbf{X}}_i^f$ and $\tilde{\mathbf{X}}_j^c$) in the pre-change image represent the same kind of objects (e.g., grass), showing that their features are very similar, and if both of them remain unchanged during the event, then the corresponding i-th fine superpixel and j-th coarse superpixel ($\tilde{\mathbf{Y}}_i^f$ and $\tilde{\mathbf{Y}}_j^c$) in the post-change image after registration correction also represent the same kind of objects (e.g., grass), maintaining similar features. In light of this, we encode each image in terms of the internal similarities between fine- and coarse-scale superpixels.

For the image $\tilde{\mathbf{X}}$, we first compute the feature distance vector $\mathbf{D}_i^x = \left\{ D_{i,j}^x \right\}_{j=1}^{N_c}$ for the i-th fine superpixel $\tilde{\mathbf{X}}_i^f$ as:

$$D_{i,j}^x = \left\| \mathbf{X}_i^f - \mathbf{X}_j^c \right\|_2^2,\ i = 1, \cdots, N_f;\ j = 1, \cdots, N_c. \tag{3}$$

Similarly, for the image $\tilde{\mathbf{Y}}$, we compute the feature distance vector $\mathbf{D}_i^y = \left\{ D_{i,j}^y \right\}_{j=1}^{N_c}$ for the i-th fine superpixel $\tilde{\mathbf{Y}}_i^f$ as

$$D_{i,j}^y = \left\| \mathbf{Y}_i^f - \mathbf{Y}_j^c \right\|_2^2,\ i = 1, \cdots, N_f;\ j = 1, \cdots, N_c. \tag{4}$$

A simple and intuitive idea is to directly compare the feature distance vectors, such as $|\mathbf{D}_i^x - \mathbf{D}_i^y|$, or the similarity metrics induced by the distance, to measure the similarity relationship differences between $\tilde{\mathbf{X}}_i^f$ and $\tilde{\mathbf{Y}}_i^f$, i.e., regarding as the change level of the i-th superpixel or the change probability of the region represented by Λ_i^f. However, since the feature distances of $D_{i,j}^x$ and $D_{i,j}^y$ are calculated on different domains as in (3) and (4), directly comparing these distances/similarities will lead to confusion of heterogeneous data. In this paper, we address this issue by employing global mapping to measure the change level.

First, we sort the feature distance vector $\mathbf{D}_i^x$ in ascending order for each fine superpixel $\tilde{\mathbf{X}}_i^f$ to obtain the index set of its k-nearest neighbors (KNN) in the coarse superpixel set within image $\tilde{\mathbf{X}}$, denoted as $\mathcal{N}_i^x = \{ j | D_{i,j}^x \text{ is one of the } k - smallest \text{ values in } \mathbf{D}_i^x,\ j = 1, \cdots, N_c \}$. If an appropriate k is chosen, then $\tilde{\mathbf{X}}_i^f$ and its KNN, i.e. $\tilde{\mathbf{X}}_j^c, j \in \mathcal{N}_i^x$, represent the same kind of objects. In this paper, we set $k = \lceil \sqrt{N_c} \rceil$ for simplicity, where $\lceil \cdot \rceil$ denotes the upward rounding. With these KNN sets, we can define an affinity matrix $\mathbf{A}^x \in \mathbb{R}^{N_f \times N_c}$ to depict these similarity relationships as: $A_{i,j}^x = \begin{cases} 1, \text{ if } j \in \mathcal{N}_i^x \\ 0, \text{ otherwise} \end{cases}$.

In the same way, we can obtain the KNN index set $\mathcal{N}_i^y$ of each fine superpixel $\tilde{\mathbf{Y}}_i^f$ in the coarse superpixel set within the post-change image, and the corresponding affinity matrix $\mathbf{A}^y \in \mathbb{R}^{N_f \times N_c}$, denoted as: $A_{i,j}^y = \begin{cases} 1, \text{ if } j \in \mathcal{N}_i^y \\ 0, \text{ otherwise} \end{cases}$.

Then, we map the index set $\mathcal{N}_i^x$ of pre-change image to the coarse superpixel set of the post-change image, and compare the difference between the original and mapped feature distances as the change metric

$$\begin{aligned} F_i^y &= \frac{1}{k} \sum_{j' \in \mathcal{N}_i^x} D_{i,j'}^y - \frac{1}{k} \sum_{j \in \mathcal{N}_i^y} D_{i,j}^y \\ &= \frac{1}{k} \sum_{j=1}^{N_c} \left[\mathbf{D}^y \odot (\mathbf{A}^x - \mathbf{A}^y) \right]_{i,j}, \end{aligned} \tag{5}$$

where $\odot$ represents the Hadamard product. Similarly, we map the index set $\mathcal{N}_i^y$ of post-change image to the coarse superpixel set of the pre-change image, and compute another change metric as

$$\begin{aligned} F_i^x &= \frac{1}{k} \sum_{j' \in \mathcal{N}_i^y} D_{i,j'}^x - \frac{1}{k} \sum_{j \in \mathcal{N}_i^x} D_{i,j}^x \\ &= \frac{1}{k} \sum_{j=1}^{N_c} \left[\mathbf{D}^x \odot (\mathbf{A}^y - \mathbf{A}^x) \right]_{i,j}. \end{aligned} \tag{6}$$

Four aspects of these two change metrics of $\mathbf{F}^x$ and $\mathbf{F}^y$ are worth noting, taking the F_i^y of (5) as an example:

First, F_i^y compares the distances calculated within the same image domain, thereby avoiding the confusion that arises from heterogeneous data.

Second, without considering registration errors, F_i^y can measure whether the similarity relationships of $\tilde{\mathbf{X}}_i^f$ change or not, i.e., it can detect changes. For example, consider $\tilde{\mathbf{X}}_i^f$ and its KNN $\tilde{\mathbf{X}}_{j'}^c, j' \in \mathcal{N}_i^x$ (representing the same kind of objects), if both of them are unchanged during the event, then the mapped $\tilde{\mathbf{Y}}_i^f$ and $\tilde{\mathbf{Y}}_{j'}^c$ also represent the same kind of objects, resulting in small values of $D_{i,j'}^y$ and F_i^y in (5) for the unchanged i-th region Λ_i^f; conversely, if $\tilde{\mathbf{X}}_i^f$ is changed and $\tilde{\mathbf{X}}_{j'}^c$ is unchanged during the event, then the mapped $\tilde{\mathbf{Y}}_i^f$ and $\tilde{\mathbf{Y}}_{j'}^c$ represent different kinds of objects, resulting in large vales of $D_{i,j'}^y$ and F_i^y in (5) for the changed i-th region Λ_i^f. Thanks to the sparsity of changes in practical application, it is reasonable to assume that most of the KNN for each $\tilde{\mathbf{X}}_{j'}^c$ is unchanged. Therefore, F_i^y can be used to roughly estimate the initial change level of the i-th region of Λ_i^f.

Third, image registration errors will reduce the distinguishability of F_i^y between changed and unchanged regions. Specifically, under perfect registration, unchanged regions Λ_i^f and $\Lambda_{j'}^c$, $(j' \in \mathcal{N}_i^x)$ produce highly similar features, yielding a small value of $D_{i,j'}^y$. However, when misregistration occurs, the spatial correspondence between $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ is partially violated. Even if the physical scene has not changed, $\tilde{\mathbf{Y}}_i^f$ no longer aligns with the same object as $\tilde{\mathbf{Y}}_{j'}^c$, introducing artificial differences in the feature space. As a result, $D_{i,j'}^y$ is increased due to misregistration, which in turn leads to a higher F_i^y for unchanged regions, thereby reducing the discriminability between changed and unchanged regions.

Fourth, the changed superpixel in the KNN set of $\tilde{\mathbf{X}}_{j'}^c, j' \in \mathcal{N}_i^x$ also affects the distinguishability of F_i^y with respect to changed and unchanged. For instance, if $\tilde{\mathbf{X}}_i^f$ is unchanged and its KNN $\tilde{\mathbf{X}}_{j'}^c$ is changed, then the mapped $\tilde{\mathbf{Y}}_i^f$ and $\tilde{\mathbf{Y}}_{j'}^c$ will be very different, resulting a large element of $D_{i,j'}^y$ for the unchanged Λ_i^f; conversely, if $\tilde{\mathbf{X}}_i^f$ is changed and its KNN $\tilde{\mathbf{X}}_{j'}^c$ is also changed to the same change type as $\tilde{\mathbf{X}}_i^f$, then the mapped $\tilde{\mathbf{Y}}_i^f$ and $\tilde{\mathbf{Y}}_{j'}^c$ will be very similar, resulting a small element of $D_{i,j'}^y$ for the changed Λ_i^f. Although this influence is controllable due to the small percentage of changed regions, i.e., it guarantees the validity of the initial F_i^y in (5), we still need to filter out this influence to improve the accuracy of change metric. Since we can't know in advance which regions are changed, we employ an iterative framework to gradually filter out the negative effects of changed regions by propagating the initial detection results back into the change metric, as described in Subsection 4.5.

4.3 Local searching

We utilize a local searching process to find the initial matching position and identify the initial changed regions. Using the fine-scale superpixels of $\tilde{\mathbf{X}}_i^f$ as the template, we slide the corresponding $\tilde{\mathbf{Y}}_i^f$ within a $(2w+1) \times (2w+1)$ search window with a step size w_s to obtain candidate superpixels $\tilde{\mathbf{Y}}_{i,(u,v)}^f$ as

$$\begin{aligned} \Lambda_{i,(u,v)}^f &= \left\{ (m,n) \,|m = m' + uw_s, n = n' + vw_s, (m',n') \in \Lambda_i^f \right\} \\ \tilde{\mathbf{Y}}_{i,(u,v)}^f &= \left\{ \tilde{y}_{m,n,b} |\, (m,n) \in \Lambda_{i,(u,v)}^{fs}, b = 1, \cdots, B_2 \right\}, \end{aligned}$$

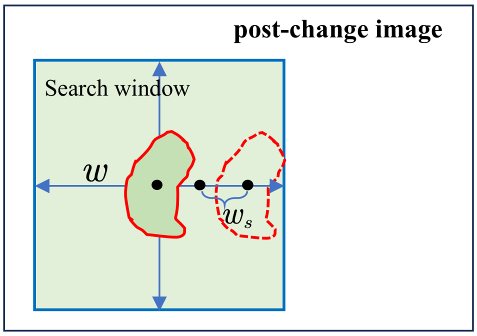


Fig. 4: Illustration of local searching window and step.

$$(7)$$

where $-\lceil\frac{w}{w_s}\rceil \leq u, v \leq \lceil\frac{w}{w_s}\rceil$. We easily have $\Lambda_i^f = \Lambda_{i,(0,0)}^f$ and $\tilde{\mathbf{Y}}_i^f = \tilde{\mathbf{Y}}_{i,(0,0)}^f$.

In the change metrics of F_i^y (5) and F_i^x (6), the registration errors are not taken into account. By replacing $\tilde{\mathbf{Y}}_i^f$ in (5) and (6) with $\tilde{\mathbf{Y}}_{i,(0,0)}^f$, we calculate the change metric of $\mathbf{F}$ for different shifts (u, v) as

$$\begin{aligned}
&F_{i,(u,v)}^y = \\
&\sum_{j' \in \mathcal{N}_i^x} \left\| \mathbf{Y}_{i,(u,v)}^f - \mathbf{Y}_{j'}^c \right\|_2^2 - \sum_{j \in \mathcal{N}_{i,(u,v)}^y} \left\| \mathbf{Y}_{i,(u,v)}^f - \mathbf{Y}_j^c \right\|_2^2, \\
&F_{i,(u,v)}^x = \sum_{j' \in \mathcal{N}_{i,(u,v)}^y} \left\| \mathbf{X}_i^f - \mathbf{X}_{j'}^c \right\|_2^2 - \sum_{j \in \mathcal{N}_i^x} \left\| \mathbf{X}_i^f - \mathbf{X}_j^c \right\|_2^2, \\
&F_{i,(u,v)} = F_{i,(u,v)}^y / B_2 + F_{i,(u,v)}^x / B_1,
\end{aligned} \tag{8}$$

where $\mathbf{Y}_{i,(u,v)}^f$ denotes the feature vector of the superpixel $\tilde{\mathbf{Y}}_{i,(u,v)}^f$, and $\mathcal{N}_{i,(u,v)}^y$ denotes the KNN index set of shifted superpixel $\tilde{\mathbf{Y}}_{i,(u,v)}^f$ in the coarse superpixel set $\left\{\tilde{\mathbf{Y}}_j^c\right\}_{j=1}^{N_c}$ within the post-change image.

As analyzed in Section 3 and Section 4.2, the registration error will increase the value of change metric under unchanged region. Therefore, for each region represented by Λ_i^f, we identify its change level as the minimum value F_i^* in the local searching window and record the corresponding displacements (u_i^*, v_i^*) as

$$\begin{aligned}
&(u_i^*, v_i^*) = \underset{-\lceil\frac{w}{w_s}\rceil \leq u,v \leq \lceil\frac{w}{w_s}\rceil}{\operatorname{argmin}} \left\{F_{i,(u,v)}\right\}, \\
&F_i^* = F_{i,(u_i^*, v_i^*)}.
\end{aligned} \tag{9}$$

By assigning the change level F_i^* to the pixels within Λ_i^f, we can obtain the difference image (DI) as

$$\mathrm{DI}\,(m,n) = F_i^*, \quad (m,n) \in \Lambda_i^f, \; i = 1, \cdots, N_f. \tag{10}$$

Fig. 4 illustrates the local searching process. In this process, we introduce a search step size in the template matching, which reduces the search space for candidate superpixels from $(2w+1) \times (2w+1)$ to $\left(2\lceil\frac{w}{w_s}\rceil + 1\right) \times \left(2\lceil\frac{w}{w_s}\rceil + 1\right)$, thereby compressing the search space scale by a factor of nearly w_s^2. It's worth noting that, although such a search step w_s is not typically adopted in template matching-based image registration methods, it is feasible for change detection problems. This distinction arises from the inherent characteristics of these two tasks:

(i) Image registration requires precise tie-points to solve the transformation model (e.g., affine transformation), where minor matching errors in the tie-points are amplified across the transformation model, necessitating highly accurate matches for each tie-point. Therefore, stepwise search is generally unsuitable for image registration, which requires pixel-by-pixel search accuracy.

(ii) Change detection, in contrast, aims to identify changes within individual superpixel regions, where the change detection results have minimal influence on each other. Moreover, superpixel-based object-change detection is relatively robust to small registration errors (e.g., within 3 pixels). Consequently, a small search step can significantly improve the efficiency of the proposed IGmLs while causing only a slight loss of accuracy.

4.4 MRF based change label assignment

We define the change label vector as $\boldsymbol{L} \in \mathbb{R}^{N_f}$, where $L_i \in \{0, 1\}$ denotes the label of fine superpixel regions represented by Λ_i^f. Specifically, we divide the index set $\mathcal{I} = \{i = 1, \cdots, N_f\}$ into changed subset $\mathcal{C} = \{i | L_i = 1, i \in \mathcal{I}\}$ and unchanged subset $\mathcal{U} = \{i | L_i = 0, i \in \mathcal{I}\}$. Then, we build the label assignment problem as an MRF based energy minimization problem with

$$\boldsymbol{L}^* = \underset{\boldsymbol{L} \in \{0,1\}^{\mathcal{I}}}{\arg\min} \left\{E_H(\boldsymbol{L}) := E_{ch} + E_{sp} + E_{dp} + E_{sm}\right\}, \tag{11}$$

where E_H denotes the hybrid energy functions, E_{ch}, E_{sp}, E_{dp}, and E_{sm} denote the change-based, sparse-based, displacement-based, and smooth-based energy functions, respectively.

4.4.1 Change-based energy

E_{ch} is defined as

$$E_{ch} = \sum_{i \in \mathcal{I}} (1 - L_i)\, F_i^*. \tag{12}$$

Intuitively, the larger the change metric F_i^*, the higher the probability that Λ_i^f will be labeled as changed ($L_i = 1$) for the purpose of minimizing E_{ch}.

4.4.2 Sparse-based energy

E_{sp} is defined as

$$E_{sp} = \lambda \|\boldsymbol{L}\|_0 = \lambda \sum_{i \in \mathcal{I}} L_i, \tag{13}$$

where $\lambda > 0$ is a weighting parameter. E_{sp} is based on the sparse priori knowledge that usually only a small portion of the region changes in practical CD problems. Moreover, by combining E_{ch} in (12) and E_{sp} in (13), we can obtain

$$E_{ch} + E_{sp} = \sum_{i \in \mathcal{I}} (\lambda - F_i^*)\, L_i + \sum_{i \in \mathcal{I}} F_i^*. \tag{14}$$

Minimizing this function $E_{ch}+E_{sp}$ with respect to $\boldsymbol{L}$, we have

$$L_i^* = \begin{cases} 1, \text{ if } F_i^* \geq \lambda \\ 0, \text{ if } F_i^* < \lambda \end{cases}. \tag{15}$$

Therefore, the weighting parameter λ can be regarded as a threshold for segmenting the change level vector $\mathbf{F}^*$. In light of this, we set λ to be the Otsu threshold parameter [47] on $\mathbf{F}^*$ as $\lambda = Th\left(\mathbf{F}^*\right)$, thereby avoiding the effort of manually selecting the parameter.

4.4.3 Displacement-based energy

E_{dp} is used to penalize big matching displacements of the unchanged regions, defined as

$$\begin{aligned} E_{dp} &= \alpha \sum_{i\in\mathcal{U}} \phi_i, \\ \phi_i &= \max\left\{ \sqrt{\left(u_i^* w_s\right)^2 + \left(v_i^* w_s\right)^2} - t, 0 \right\}, \end{aligned} \tag{16}$$

where $\alpha > 0$ is a weighting parameter, and t serves as a tolerance margin reflecting the expected accuracy of the initial coarse image registration (e.g., based on geographic coordinates). In this paper, we set t to be the search size $t = w$ for simplicity.

4.4.4 Smooth-based energy

E_{sm} is used to penalize the inconsistencies in change labels within the spatial neighborhoods. We define an R-adjacent neighborhood as follows: if two regions, represented by Λ_i^f and Λ_j^f are either connected or the spatial distance between their center points is less than R, then $\langle i,j \rangle$ are marked as spatial neighbors of each other, denoted as $i \in \mathcal{N}_j^R$, or $j \in \mathcal{N}_i^R$. Given that the average size of Λ_i^f is MN/N_f, we set $R = 2\sqrt{MN/N_f}$ for simplicity. Then, we construct the E_{sm} as

$$\begin{aligned} E_{sm} &= \beta \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{N}_i^R} \varphi_{i,j} \left| L_i - L_j \right|, \\ \varphi_{i,j} &= \frac{e^{-\left(F_i^*-F_j^*\right)^2/2\sigma^2}}{dist\left(\Lambda_i^f, \Lambda_j^f\right)}, \end{aligned} \tag{17}$$

where $\beta > 0$ is the weighting parameter, $dist\left(\Lambda_i^f, \Lambda_j^f\right)$ denotes the Euclidean spatial distance between the center of Λ_i^f and Λ_j^f, and σ denotes the normalization parameter with $\sigma = \frac{\sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{N}_i^R} \left(F_i^*-F_j^*\right)^2}{\sum_{i\in\mathcal{I}} \left|\mathcal{N}_i^R\right|}$.

4.4.5 Hybrid energy functions

By substituting the energy functions of (12), (13), (16), and (17) into (11), and using $\mathcal{U} = \{i | L_i = 0, i \in \mathcal{I}\}$, we have

$$E_H = \sum_{i\in\mathcal{I}} \left\{ \left(1-L_i\right)\left(F_i^* + \alpha\phi_i\right) + \lambda L_i + \sum_{j\in\mathcal{N}_i^R} \beta\varphi_{i,j} \left| L_i - L_j \right| \right\} \tag{18}$$

We further set the weighting parameters of α and β as

$$\begin{aligned} \alpha &= \frac{\alpha' \sum_{i\in\mathcal{I}} F_i^*}{\sum_{i\in\mathcal{I}} \phi_i}, \\ \beta &= \frac{\beta' \sum_{i\in\mathcal{I}} F_i^*}{\sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{N}_i^R} \varphi_{i,j}}, \end{aligned} \tag{19}$$

where α' and β' are easier to adjust.

In the E_H of (19), we can observe that the energy functions are antagonistic to each other. The first term tends to favor a changed label, i.e., $\boldsymbol{L} = \mathbf{1}$, while the second term tends to favor an unchanged label, i.e., $\boldsymbol{L} = \mathbf{0}$. The last term promotes smoothness, aiming for consistent labels within neighborhoods, i.e., $\boldsymbol{L} = \mathbf{1}$ or $\boldsymbol{L} = \mathbf{0}$.

The energy minimization problem of (11) can be solved via the graph cuts algorithm [5, 6], and then the initial change map (CM) can be computed as

$$\mathrm{CM}\left(m,n\right) = L_i^*, \quad \left(m,n\right) \in \Lambda_i^f, \; i = 1, \cdots, N_f. \tag{20}$$

4.5 Iterative process

In the proposed IGmLs, we use an iterative coarse-to-fine filtering and correction framework to gradually reduce the influence caused by registration errors and changed regions on the global-mapping and local-searching processes. Suppose that the displacements and change labels of the current iteration are $\left(u_i^*, v_i^*\right)$, $i = 1, \cdots, N_f$ of (9) and $\boldsymbol{L}^*$ of (11) respectively, we will update global mapping process in the next iteration as follows.

(i) Identify the index sets $\mathcal{J}_{ch}$, $\mathcal{J}_{un}$ of changed and unchanged coarse superpixels with the initial CM of (20) as

$$\begin{cases} j \in \mathcal{J}_{ch}, \text{ for } \forall \left(m,n\right) \in \Lambda_j^c, \text{if } \exists \, \mathrm{CM}\left(m,n\right) = 1 \\ j \in \mathcal{J}_{un}, \text{ otherwise} \end{cases}. \tag{21}$$

That is, as long as there is a element in Λ_j^c, $j = 1, \cdots, N_c$ that has changed, we mark this Λ_j^c as changed coarse region.

(ii) Update the shifted post-change image with the displacements $\left(u_i^*, v_i^*\right)$ as

$$\tilde{y}_{m,n,b}^* = \tilde{y}_{m+u_i^* w_s, n+v_i^* w_s, b}, \; \left(m,n\right) \in \Lambda_i^f, \tag{22}$$

and obtain the shifted coarse superpixel as

$$\tilde{\mathbf{Y}}_j^{c*} = \left\{ \tilde{y}_{m,n,b}^* \middle| \left(m,n\right) \in \Lambda_j^c, b = 1, \cdots, B_2 \right\}. \tag{23}$$

Then, we extract the feature matrix of shifted coarse superpixel as $\mathbf{Y}^{c*}$.

(iii) Re-compute the change metric of $\mathbf{F}$ in (8) by replacing $\mathbf{Y}^c$, $\mathcal{N}_i^x$ and $\mathcal{N}_{i,(u,v)}^y$ with $\mathbf{Y}^{c*}$, $\underline{\mathcal{N}}_i^x$ and $\underline{\mathcal{N}}_{i,(u,v)}^y$, respectively. Here, $\underline{\mathcal{N}}_i^x$ denotes the KNN index set of fine superpixel $\tilde{\mathbf{X}}_i^f$ in the unchanged coarse superpixel set of $\left\{ \tilde{\mathbf{X}}_j^c | j \in \mathcal{J}_{un} \right\}$, and $\underline{\mathcal{N}}_{i,(u,v)}^y$ denotes the KNN index set of shifted fine superpixel $\tilde{\mathbf{Y}}_{i,(u,v)}^f$ in the unchanged shifted coarse superpixel set of $\left\{ \tilde{\mathbf{Y}}_j^{c*} | j \in \mathcal{J}_{un} \right\}$.

(iv) Finally, compute the $\left(u_i^*, v_i^*\right)$ and $\mathbf{F}^*$ with (9), and then calculate the $\boldsymbol{L}^*$ of (11) with the updated $\left(u_i^*, v_i^*\right)$ and $\mathbf{F}^*$.

The framework of the proposed IGmLs is summarized as in Algorithm 1. Briefly, in the preprocessing, IGmLs obtains

```
Algorithm 1. IGmLs for HeCD with unregistered images.
Input: Images of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$, parameters of $N_f$,$N_c$, $\alpha'$ and $\beta'$.
Preprocessing:
  Implement GMMSP on $\tilde{\mathbf{X}}$ to obtain $\Lambda^f$, $\Lambda^c$, $\tilde{\mathbf{X}}_i^f$ and $\tilde{\mathbf{X}}_i^c$.
  Extract the features to obtain $\mathbf{X}^f$ and $\mathbf{X}^c$.
  Obtain the shifted fine-scale $\tilde{\mathbf{Y}}_{i,(u,v)}^{fs}$ and features $\mathbf{Y}_{i,(u,v)}^{fs}$.
Main iteration loop of IGmLs:
  Set the initial displacements as $(u_i^*, v_i^*) = (0,0)$.
  Set the initial index subset as $\mathcal{J}_{un} = \{1, 2, \cdots, N_c\}$.
  for $iter = 1, \cdots, N_{iter}$ do
    1. DI generation:
    Find the $\underline{\mathcal{N}}_i^x$ of $\tilde{\mathbf{X}}_i^f$ in the unchanged $\tilde{\mathbf{X}}^c$ with $\mathcal{J}_{un}$.
    Generate the shifted coarse $\mathbf{Y}^{c*}$ with $(u_i^*, v_i^*)$.
    Find the $\underline{\mathcal{N}}_{i,(u,v)}^y$ of $\tilde{\mathbf{Y}}_{i,(u,v)}^f$ in the unchanged $\tilde{\mathbf{Y}}^{c*}$ with $\mathcal{J}_{un}$.
    Calculate the change metric of $\mathbf{F}$ for different shifts $(u, v)$.
    Update the displacements $(u_i^*, v_i^*)$ with (9).
    Calculate the minimum value $F_i^*$ with (9).
    Compute the DI with (10).
    2. Change label assignment:
    Construct the energy functions to obtain $E_H$.
    Solve the minimization problem (11) to obtain $\boldsymbol{L}^*$.
    Compute the CM with (20).
    Update the index set $\mathcal{J}_{un}$ with (21).
  End for
Output: The change map with (20).
```

the superpixel segmentation results at both coarse and fine scales; in the global-mapping process, IGmLs transforms heterogeneous images into a common structural space to make them comparable, which contains KNN construction and change metric calculation; in the local-searching process, IGmLs initially eliminates the influence of registration errors on the change metric; in the change extraction process, IGmLs employs an MRF model to calculate the initial change regions; and finally, IGmLs utilizes the iterative refinement framework to further eliminate the influence of changes and misregistration on the change metrics, thereby yielding more accurate change detection results.

5 Experiments and discussions

5.1 Experimental setting

5.1.1 Datasets

To test the performance of the proposed IGmLs, we employ five datasets listed in Table 1. Datasets #1-#3 are obtained by first adding affine transformations to the original well-registered heterogenous image pairs provided by [43] [1] and latter performing traditional image registration method of RIFT [24], the RMSE of misregistration are 24.51, 15.02, 28.97 respectively, where all the pixels in the image pairs are taken as ground control points (GCPs) for calculating the RMSE given affine transformations and RIFT registration. Datasets #4-#5 are obtained by performing RIFT on the original georeferenced matched images, and the RMSE of misregistration are 8.78 and 14.46 respectively, where 49 manually selected GCPs are used for calculating the RMSE. Fig. 5 shows the pre-change and post-change images, the ground truth, and registration performance of these datasets.

5.1.2 Evaluation Metrics

We evaluate the performance of algorithm in detecting changes by examining DI and CM, which can be quantitatively assessed by Precision-Recall (PR) curve and confusion matrix, respectively. Specifically, we use the PR curve and area under curve (AUC) to assess the DI, and employ various colors to label true positive (TP), false positive (FP), true negative (TN), and false negative (FN) on the CM, and compute the overall accuracy (OA), Kappa coefficient (κ) and F1-score for the CM.

5.1.3 Comparison methods

Since there have been no studies on HeCD under misregistration to date, we can only choose thirteen SOTA methods for common HeCD task as comparison, including HPT [28], AMDIR [31], IRG-McS [53], SCASC [55], CICM [60], FPMS [43], USSD [81], AOSG [20], GSGM [18], LPEM [54], CFRL [27], CGSL [68], and ITSA [35]. For these comparison methods, we directly use their publicly available codes with default parameters to generate the change detection results.

5.1.4 Experimental settings

For all the experiment results of IGmLs, we set superpixel numbers as $N_f = 2500$ and $N_c = 500$, set the search step size as $w_s = 3$, and fix the balancing parameters as $\alpha' = 0.01$ and $\beta' = 2$ for the energy model E_H. We adjust the search window size w for different datasets based on the RMSE of misregistration. Specifically, we set $w = 21, 15, 24, 9, 15$ for Datasets #1-#5. These parameters will be discussed in Section 5.3.

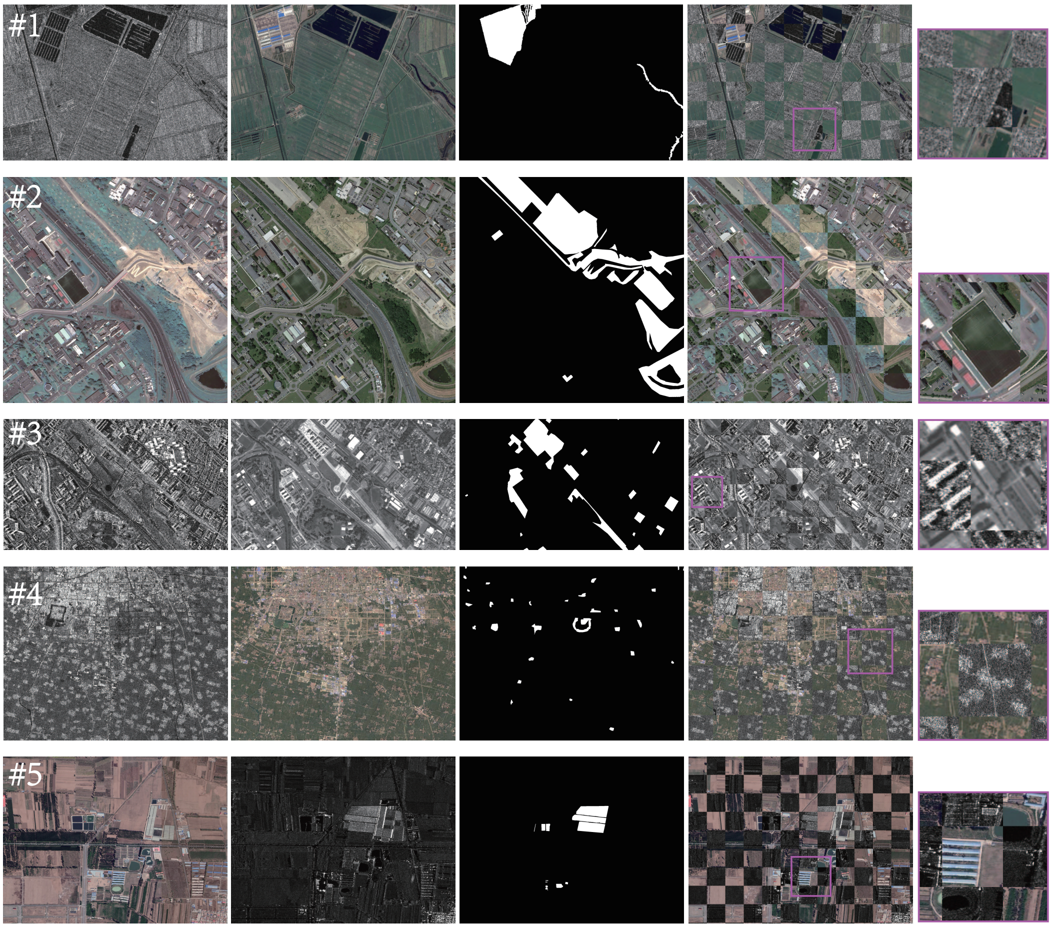
5.2 Experimental results

Fig. 6 shows DIs generated by the proposed IGmLs on different datasets by using the change metric (8)-(10). Two facets can be noted: first, the DIs can reflect change information between the unregistered heterogeneous images, i.e., the discriminative ability between changed and unchanged regions in the DIs is notable, validating the effectiveness of the iterative global mapping-local searching framework. Second, the DI assigns overly small pixels values to some truly changed areas, i.e., it tends to lead to missed detections in these regions, e.g., the upper part of DI on Dataset #2 in Fig. 6(b). The reason for this phenomenon lies in the fact that (9) uses the minimum value of F_i^* in the local searching window, which would make some of the real changes to be treated as registration errors. To mitigate this phenomenon, in the later MRF based energy model of (11), we introduce the displacement-based energy E_{dp} and smooth-based energy E_{sm}, which can help the CM to detect these changes that are ignored in the DI.

To evaluate the detection performance of different methods, Fig. 7 shows the CMs of these comparison methods, where the TP, FP, TN and FN are marked with different colors for better visual comparison. Fig. 2 lists the corresponding quantitative evaluation indicators of these CMs. From these results, it can be found that HPT, AMDIR, USSD, AOSG,

---

[1] The well-registered images of Datasets #1-#3 are available at `http://www-labs.iro.umontreal.ca/~mignotte`

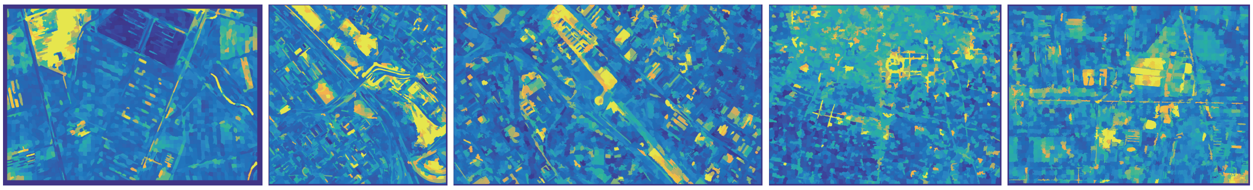


(a) pre-change (b) post-change (c) ground truth (d) registration performance

Fig. 5: Dataset: from top to bottom, they are corresponding to Datasets #1 to #5 respectively. (a) pre-change image; (b) post-change image; (c) ground truth; (d) presentation of misregistration.

Table 1: Description of the heterogeneous datasets.

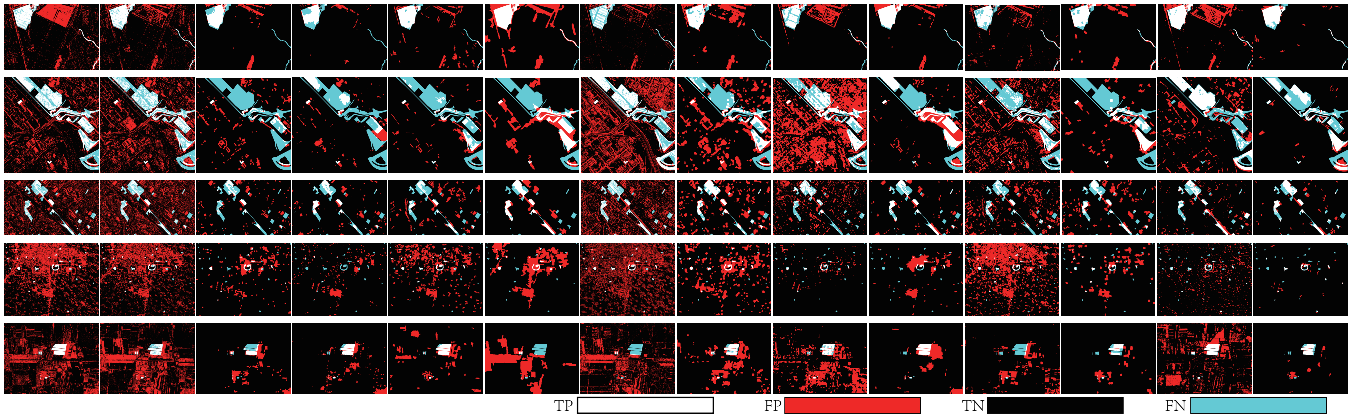
Dataset	Sensor	Image size (pixels)	Date	Location	Spatial resolution	Change event	RMSE (pixels)
#1	Radarsat-2 Google Earth	593 × 921 × 1 593 × 921 × 3	June 2008 Sept. 2012	Shuguang Village, China	8m	Construction	24.51
#2	Pleiades WorldView2	2000 × 2000 × 3 2000 × 2000 × 3	May 2012 July 2013	Toulouse, France	0.52m	Construction	15.02
#3	TerraSAR-X Pleiades	4404 × 2604 × 1 4404 × 2604 × 1	Feb. 2009 July 2013	Toulouse, France	2m	Construction	28.97
#4	Sentinel-1A Google Earth	904 × 1168 × 1 904 × 1168 × 3	Jan. 2016 Dec. 2016	Shangqiu City, China	20m	Lake and building changes	8.93
#5	Google Earth Gaofen-3	1500 × 2000 × 3 1500 × 2000 × 1	Mar. 2018 Dec. 2019	Beijing City, China	1m	Land changes	14.46



(a) Dataset #1 (b) Dataset #2 (c) Dataset #3 (d) Dataset #4 (e) Dataset #5

Fig. 6: DIs obtained by IGmLs: from (a) to (e) are the DIs on Datasets #1 to #5.

GSGM and CFRL have relatively more false positives as shown in the red region of CM in Fig. 7, e.g., in the Datasets #2, #4 and #5. In contrast, IRG-McS, SCASC, CICM, FPMS and CGSL have more miss detections as shown in the cyan region of CM, e.g., in the Datasets #2 and #3. Generally speaking, by comparing the FP and FN in Fig. 7, it can be seen that among these 13 comparison methods, CICM, FPMS, SCASC, ITSA, CGSL and LPEM are relatively less affected by registration errors. This might be attributed to the fact that the CICM and FPMS employ the downsampling operations and consider the spatial relationships between multitemporal images, ITSA uses an iterative training strategy to progres-



(a) HPT (b) AMDIR (c) IRG-McS (d) SCASC (e) CICM (f) FPMS (g) USSD (h) AOSG (i) GSGM (j) LPEM (k) CFRL (l) CGSL (m) ITSA (n) IGmLs

Fig. 7: CMs generated by different methods. From top to bottom, they are corresponding to Datasets #1 to #5, respectively. From (a) to (k) are the change maps generated by (a) HPT [28], (b) AMDIR [31], (c) IRG-McS [53], (d) SCASC [55], (e) CICM [60], (f) FPMS [43], (g) USSD [81], (h) AOSG [20], (i) GSGM [18], (j) LPEM [54], (k) CFRL [27], (l) CGSL [68], (m) ITSA [35], and (n) the proposed IGmLs. In the change map, white, red, black, and cyan mark true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), respectively.

Table 2: OA, κ and F1 of CMs. The best and second best scores are marked in **red** and **blue**, respectively.

Methods	Dataset #1			Dataset #2			Dataset #3			Dataset #4			Dataset #5			Average		
	OA	κ	F1	OA	κ	F1	OA	κ	F1	OA	κ	F1	OA	κ	F1	OA	κ	F1
HPT [28]	0.822	0.305	0.373	0.783	0.352	0.479	0.738	0.121	0.229	0.671	0.032	0.055	0.770	0.116	0.151	0.757	0.185	0.257
AMDIR [31]	0.880	0.415	0.468	0.715	0.190	0.354	0.694	0.084	0.200	0.722	0.035	0.058	0.734	0.073	0.111	0.749	0.159	0.238
IRG-McS [53]	0.946	0.585	0.614	0.803	0.055	0.150	0.899	0.282	0.337	0.926	0.133	0.152	0.941	0.225	0.249	0.903	0.256	0.300
SCASC [55]	**0.957**	0.599	0.621	0.833	0.187	0.265	0.913	0.380	0.427	0.943	0.022	0.041	**0.952**	**0.406**	**0.425**	**0.920**	0.319	0.356
CICM [60]	0.943	0.621	0.651	0.843	0.202	0.269	0.881	0.239	0.304	0.864	0.078	0.099	0.933	0.339	0.362	0.893	0.296	0.337
FPMS [43]	0.885	0.467	0.516	0.814	0.256	0.364	**0.921**	**0.503**	**0.546**	0.862	0.073	0.095	0.836	0.009	0.023	0.864	0.257	0.309
USSD [81]	0.917	0.390	0.434	0.641	0.166	0.354	0.719	0.121	0.231	0.695	0.035	0.058	0.780	0.010	0.022	0.750	0.139	0.220
AOSG [20]	0.906	0.470	0.515	0.734	0.112	0.271	0.847	0.329	0.403	0.855	0.085	0.106	0.897	0.220	0.249	0.848	0.243	0.309
GSGM [18]	0.907	0.425	0.471	0.536	0.054	0.279	0.848	0.225	0.304	**0.972**	**0.217**	**0.230**	0.812	0.117	0.151	0.815	0.207	0.287
LPEM [54]	0.948	**0.659**	**0.686**	**0.851**	0.400	0.487	0.917	0.430	0.475	0.932	0.076	0.095	0.945	0.397	0.417	0.919	**0.393**	**0.432**
CFRL [27]	0.929	0.505	0.542	0.728	0.135	0.297	0.831	0.156	0.243	0.676	0.024	0.047	0.906	0.013	0.047	0.814	0.167	0.235
CGSL [68]	0.949	0.634	0.661	0.837	0.264	0.351	0.879	0.302	0.367	0.930	0.133	0.152	0.962	0.387	0.404	0.911	0.344	0.387
ITSA [35]	0.896	0.482	0.529	0.835	**0.414**	**0.513**	0.916	0.449	0.495	0.935	0.111	0.129	0.809	0.141	0.175	0.878	0.319	0.368
IGmLs	**0.977**	**0.806**	**0.818**	**0.895**	**0.524**	**0.580**	**0.939**	**0.553**	**0.585**	**0.986**	**0.395**	**0.402**	**0.980**	**0.593**	**0.603**	**0.955**	**0.574**	**0.598**

sively reduce false detections, while SCASC, CGSL and LPEM are superpixel-based HeCD methods and also applies smooth constraints in the CM generation. These operations can mitigate the influence of small misregistration on HeCD to some extent, but this is far from satisfactory when faced with large registration errors. On the contrary, by inspecting the CM obtained by the proposed IGmLs in Fig. 7(n), it can be demonstrated that IGmLs can well resist the registration error on HeCD.

From Table 2, IGmLs gains the highest scores (OA, κ, F1) on all the evaluated datasets and shows significant improvement over the other methods. The average κ and F1 of IGmLs are 0.574 and 0.598 respectively, which are 0.181 and 0.166 higher than that of the second-ranked LPEM respectively. This outstanding performance can be attributed to the following three aspects: first, IGmLs uses the local searching process to reduce the influence of registration errors on the structural difference metric of the global mapping process; second, the displacement-based energy and smooth-based energy are employed in the MRF segmentation model for CM calculation, which further enhances the robustness to registration errors by considering spatial correlations; third, IGmLs utilizes the iterative coarse-to-fine filtering framework to backpropagate the matching and changing results, which further eliminates the influence of changes and misregistration on the change metrics, thereby refining the HeCD results. The combination of these three aspects enables the proposed IGmLs to more accurately portray the HeCD problem under misregistration conditions.

5.3 Discussions

5.3.1 Ablation study

IGmLs comprises three key processes: global mapping, local searching, and iterative refinement. Given that global mapping is essential for establishing a common structural space and enabling meaningful comparison of heterogeneous images, the ablation study primarily focuses on local searching and iterative refinement. Table 3 reports the average evaluation metrics of the CMs generated by IGmLs with and without these two processes across all datasets, showing that removing either component noticeably degrades accuracy. These findings confirm that all components are indispensable for IGmLs to handle challenging unregistered heterogeneous images effectively. We next present and analyze the contributions of local searching and iterative refinement in detail.

(ii) The effectiveness of local searching: As analyzed in Section 3 and Section 4.3, the registration errors directly cause change detection errors at the boundaries between different kinds of objects and increase the variance of both changed and unchanged pixels, which in turn lead to poor discriminability

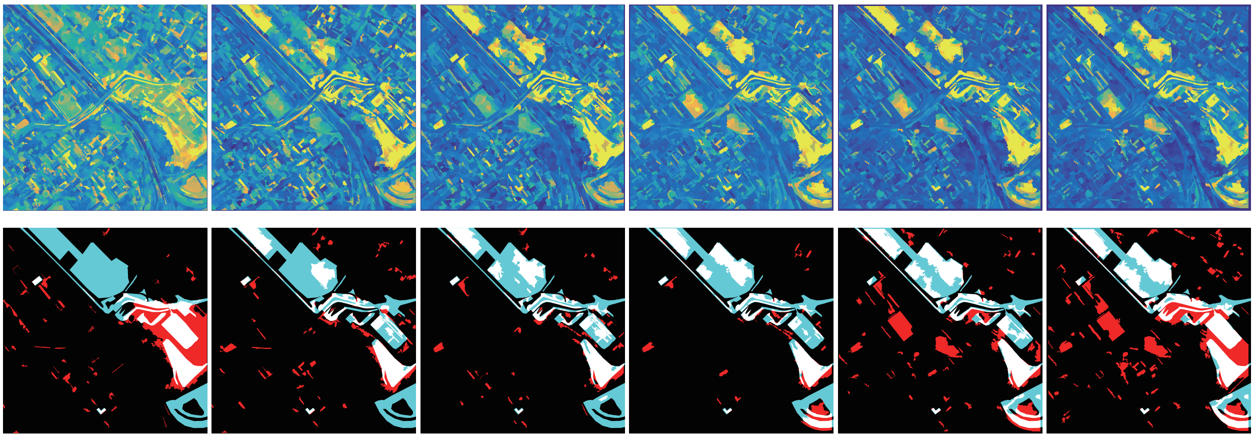

(a) w=0 (b) w=5 (c) w=10 (d) w=15 (e) w=20 (f) w=25

Fig. 8: DIs (first row) and CMs (second row) generated by IGmLs with different searching window sizes on Dataset #2. (a) $w = 0$, (b) $w = 5$, (c) $w = 10$, (d) $w = 15$, (e) $w = 20$, (f) $w = 25$.

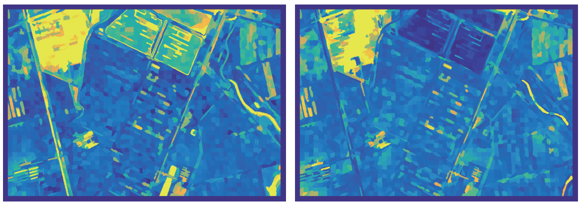
Table 3: Ablation study of IGmLs measured by the average scores across all datasets, with "GM", "LS" and "IR" denoting the global mapping, local searching, and iterative refinement, respectively.

Settings	OA	κ	F1
GM	0.913	0.286	0.314
GM + LS	0.925	0.441	0.482
GM + IR	0.921	0.322	0.357
GM + LS + IR	0.955	0.574	0.598

Table 4: AUC and F1 of IGmLs with different searching window sizes and step sizes on Dataset #2.

Indicators	Window sizes of local searching with fixed $w_s = 3$					
	$w = 0$	$w = 5$	$w = 10$	$w = 15$	$w = 20$	$w = 25$
AUC of DI	0.323	0.452	0.568	0.583	0.592	0.599
F1 of CM	0.358	0.455	0.553	0.580	0.584	0.591
Indicators	Step sizes of local searching with fixed $w = 15$					
	$w_s = 1$	$w_s = 2$	$w_s = 3$	$w_s = 4$	$w_s = 5$	$w_s = 6$
AUC of DI	0.595	0.589	0.583	0.575	0.571	0.568
F1 of CM	0.588	0.583	0.580	0.572	0.561	0.554

between changed/unchanged regions in the DI. IGmLs uses the local searching process to find the minimum value of change metric to reduce the direct influence of misregistration. Fig. 8 shows the DIs and CMs generated by IGmLs with different searching window sizes, and Table 4 reports the corresponding quantitative evaluation indicators (AUC and F1) of these DIs and CMs. From Fig. 8 and Table 4, we can find that as the window size w of local searching increases, the change/unchange distinguishability in DI is stronger, and also leads to better CM. However, it is worth noting that when w becomes too large, it can also introduce certain drawbacks: on one hand, it enlarges the searching space, thereby increasing the algorithm's complexity; on the other hand, it makes IGmLs more difficult to detect subtle changes, i.e., it will confuse real changes with errors caused by misregistration. Therefore, the choice of w needs to be a compromise between algorithm efficiency, the level of misregistration in the multitemporal images, and the granularity of the changes to be detected. In addition, within the local searching process, we introduce a search step size w_s in the template matching, which reduces the search space roughly by a factor of w_s^2. Table 4 reports the performance of IGmLs with different w_s values. Considering both computational complexity and detection accuracy, we recommend setting w_s between 2 and 4 as a compromise choice.


(a) the 1st DI (b) the 2nd DI

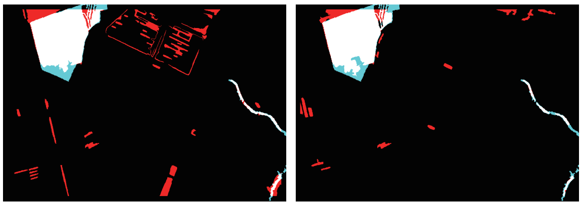

(c) the 1st CM (d) the 2nd CM

Fig. 9: DIs and CMs generated by the first and second iterations of IGmLs on Dataset #1. (a) DI of the first iteration; (b) DI of the second iteration; (c) CM of the first iteration; (d) CM of the second iteration.

(ii) The effectiveness of iterative refinement: In order to demonstrate the effectiveness of iterative coarse-to-fine filtering framework employed in IGmLs, we will compare the DIs and CMs generated by the first and second iterations of IGmLs. From Fig. 9(a) and (b), we can find that the DI of second iteration is well able to alleviate the change confusion in the DI of first iteration caused by two factors: the confusion on the change metric of (8) caused by the changed coarse superpixel in the KNN sets of $\tilde{\mathbf{X}}_{j'}^{c}, j' \in \mathcal{N}_i^x$ and $\tilde{\mathbf{Y}}_{j'}^{c}, j' \in \mathcal{N}_{i,(u,v)}^y$, and the confusion on the structure representation caused by the registration errors on feature template $\mathbf{Y}^c$ of coarse superpixels. Fig. 10 shows the PR curves of the DIs generated in the first and second iterations, where the corresponding areas under PR curves are 0.565 and 0.675, respectively. Fig. 9(c) and (d) show the CMs of these two iterations, from which it can be seen that the CM after

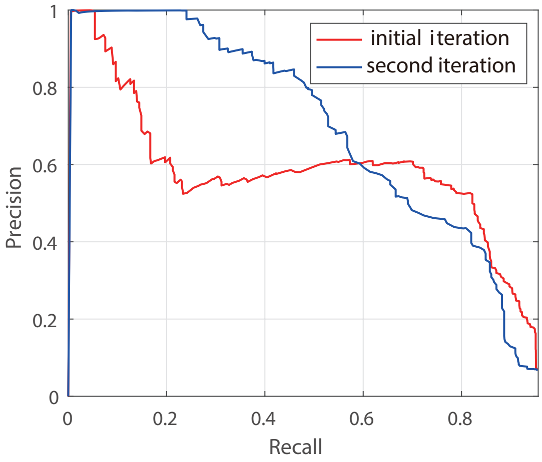


Fig. 10: Precision-Recall curves of the DI generated by the first and second iterations of IGmLs on Dataset #1.

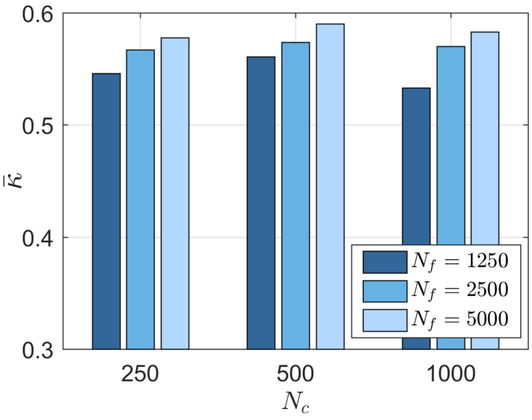


Fig. 11: Sensitivity analysis of the numbers of superpixels.

iterative filtering eliminates many false detections, leading to an improvement of the Kappa coefficient of CM from 0.705 to 0.806.

5.3.2 Parameters analysis

The main parameters used in IGmLs are the superpixel numbers of N_f and N_c, the number of neighbors k, the weighting parameters of α' and β' in the MRF based energy model.

In general, the numbers of fine and coarse superpixels should be selected according to the spatial resolution of images and the timeliness requirement of HeCD task. As N_f and N_c increase, the computational complexity also increases as discussed in the next subsection, and N_c has a greater impact on computational complexity than N_f. Conversely, smaller superpixel numbers can improve the change detection efficiency, however, excessively small N_f and N_c may cause the segmented superpixels to contain multiple classes of objects inside, leading to inaccurate image structure representation and thereby affecting the accuracy of change detection. As shown in Fig. 11, with N_c fixed, increasing N_f improves average Kappa values ($\bar{\kappa}$) by generating smaller fine superpixels and enhancing granularity. In contrast, with N_f fixed, an excessively large N_c lowers $\bar{\kappa}$ because over-segmented coarse superpixels lose their robustness to registration errors. Therefore, this paper simply set $N_f = 2500$ and $N_c = 500$ as a compromise.

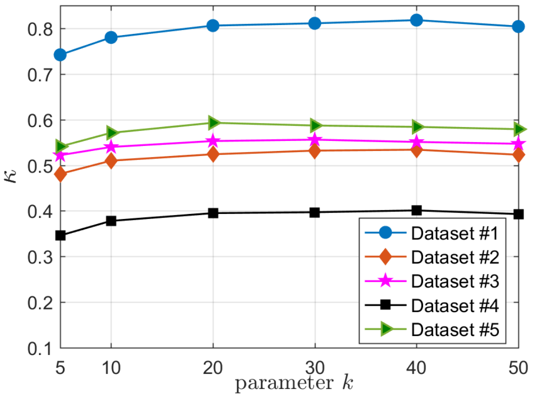


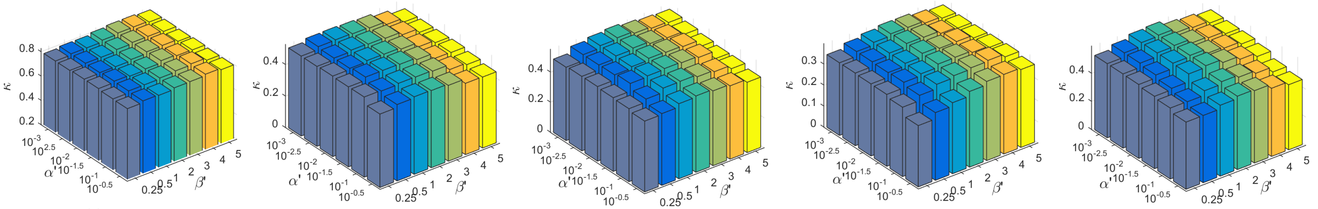
Fig. 12: Sensitivity analysis of parameter k.

For the number of neighbors k, a very small k may not provide sufficient robustness for KNN, failing to adequately capture the similarity relationships between superpixels. Conversely, a very large k may lead to over-connections, where dissimilar superpixels are linked, reducing the discriminability of the change metric and increasing computational cost. Considering $N_c = 500$ in IGmLs, we tested $k = 5, 10, 20, 30, 40, 50$, and evaluated the corresponding κ values of the CMs, as shown in Fig. 12. The results indicate that IGmLs is robust to the choice of k, with the average Kappa across $k = 10$ to 50 ranging from 0.556 to 0.578. Accordingly, we set $k = \lceil \sqrt{N_c} \rceil$ in IGmLs for simplicity, balancing accuracy and computational cost, following empirical guidelines from density estimation [16] and KNN classification [44].

For the penalty parameters of α' and β', they are applied to tune the weights of displacement-based energy E_{dp} and smooth-based energy E_{sm} in the (18). The E_{dp} tends to obtain changed label with $L_i = 1$ as $\alpha' \to \infty$, the E_{sm} tends to get smooth results, i.e., aiming for consistent labels of $\boldsymbol{L} = \mathbf{1}$ or $\boldsymbol{L} = \mathbf{0}$ as $\beta' \to \infty$. We show parameters sensitivity of IGmLs on α' and β' in Fig. 13, where we adjust the α' from $10^{-7/2}$ to $10^{-1/2}$ and β' from 0.25 to 5, respectively. From Fig. 13, we can find that IGmLs obtains favorable κ for a certain range of values of α' and β', which indicates its robustness to weighting parameters. As can be seen from Fig. 13, the recommended ranges of parameters are $\left[10^{-3}, 10^{-3/2}\right]$ and $[1, 3]$ for α' and β', respectively.

5.3.3 Complexity analysis

The main computational complexity of the IGmLs is concentrating on the pre-processing, DI generation and change label assignment as reported in the framework of Algorithm 1. In the pre-processing, the complexity of superpixel segmentation of GMMSP is $\mathcal{O}\left(M_1 N_1\right)$ as illustrated in [2]. In the DI generation, computing the feature distances between fine-scale and coarse-scale superpixels in image $\tilde{\mathbf{X}}$ requires $\mathcal{O}\left(B_1 N_f N_c\right)$, computing the feature distances between shifted fine-scale superpixels in the local searching window and coarse-scale superpixels in image $\tilde{\mathbf{Y}}$ requires $\mathcal{O}\left(4\lceil \frac{w}{w_s} \rceil^2 B_2 N_f N_c\right)$, finding the KNN index set for all the fine-scale superpixels in image $\tilde{\mathbf{X}}$ requires $\mathcal{O}\left(N_f N_c \log N_c\right)$, finding the KNN index set for all the shifted fine-scale superpixels in the local searching window in image $\tilde{\mathbf{Y}}$ approximately requires $\mathcal{O}\left(4\lceil \frac{w}{w_s} \rceil^2 N_f N_c \log N_c\right)$, computing the change metric (8)



(a) Dataset #1 (b) Dataset #2 (c) Dataset #3 (d) Dataset #4 (e) Dataset #5

Fig. 13: Sensitivity analysis of parameters α' and β'.

with different shifts requires $\mathcal{O}\left(4\lceil\frac{w}{w_s}\rceil^2 N_f\sqrt{N_c}\right)$. In the MRF based change label assignment, constructing the energy functions of E_{ch}, E_{sp} and E_{dp} approximately requires $\mathcal{O}\left(N_f\right)$, constructing the energy E_{sm} requires $\mathcal{O}\left(N_R\right)$ with $N_R = \sum_{i\in\mathcal{I}}\left|\mathcal{N}_i^R\right|$. Finally, the min-cut/max-flow algorithm is used for solving the energy minimization model, whose empirical complexity is very low on typical problem instances and theoretical complexity is presented in [6]. Generally speaking, the complexity of IGmLs is mainly related to the superpixel numbers of N_f and N_c, the searching window size w, and the step size w_s. Taking Dataset #3 (with image size 4404 × 2604) as an example, the total computational time of IGmLs is 219.3 seconds[2], where the pre-processing takes 6.7 seconds, the DI generation takes 207.7 seconds, and the change label assignment takes 4.9 seconds in IGmLs.

6 Conclusion

This paper focuses on the issue of unsupervised HeCD with unregistered heterogeneous images. We first analyze the influence of registration errors on HeCD task in terms of image transformation and feature comparison. Based on this analysis, the paper proposes an unsupervised iterative global mapping-local searching method, named IGmLs. Specifically, IGmLs segments images into two-scales superpixels: the fine-scale superpixel as the basic unit of change detection, and the coarse-scale superpixel as the feature templates in image transformations. Then, IGmLs utilizes the global mapping approach to transform heterogeneous images into a common structural space to enable the comparison, and uses the local searching process to reduce the direct influence of registration errors on the change metric. Subsequently, IGmLs constructs an MRF model to combine the results of global mapping and local searching to identify the changed regions. Finally, an iterative framework is employed to backpropagate the matching and changing results to refine the global mapping and local searching processes, which further eliminates the indirect influence of changes and misregistration on the image transformation and change metrics, thereby improving the accuracy of HeCD. The experimental results verify the effectiveness of the proposed method. Future work will explore integrating this iterative framework into other HeCD methods to further improve robustness to registration errors. Given that research on HeCD under misregistration conditions is still scarce, the approach proposed in this paper is expected to help fill this gap and facilitate the practical application of HeCD.

Acknowledgment

The authors would like to thank Prof. Max Mignotte (University of Montreal), Dr. Redha Touati (Polytechnique Montreal) and Dr. Luigi Tommaso Luppino (UiT the Arctic University of Norway) for sharing their codes and data sets. This work was supported by the National Natural Science Foundation of China under Grant No. 62401577; the Natural Science Foundation of Hunan Province of China under Grant No. 2024JJ6466; the China Postdoctoral Science Foundation under Grant No. 2024M754301 and No. 2025T181178.

Conflict of interest The authors declare that they have no conflict of interest.

Data Availability The datasets analysed during the current study are available from the corresponding author on reasonable request. Public Code availability: `https://github.com/yulisun/IGmLs`.

References

1. Ban, Y., Yousif, O.A.: Multitemporal spaceborne sar data for urban change detection in china. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **5**(4), 1087–1094 (2012) 1
2. Ban, Z., Liu, J., Cao, L.: Superpixel segmentation using gaussian mixture model. IEEE Transactions on Image Processing **27**(8), 4105–4117 (2018) 5, 14
3. Bovolo, F., Bruzzone, L.: The time variable in data fusion: A change detection perspective. IEEE Geoscience and Remote Sensing Magazine **3**(3), 8–26 (2015) 1
4. Bovolo, F., Bruzzone, L., Marchesi, S.: A multiscale technique for reducing registration noise in change detection on multitemporal vhr images. In: 2007 International Workshop on the Analysis of Multi-temporal Remote Sensing Images, pp. 1–6. IEEE (2007) 4
5. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient nd image segmentation. International journal of computer vision **70**(2), 109–131 (2006) 9
6. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(9), 1124–1137 (2004) 9, 15
7. Brown, K., Foody, G., Atkinson, P.: Modelling geometric and misregistration error in airborne sensor data to enhance change detection. International Journal of Remote Sensing **28**(12), 2857–2879 (2007) 2
8. Bruzzone, L., Bovolo, F.: A novel framework for the design of change-detection systems for very-high-resolution remote sensing images. Proceedings of the IEEE **101**(3), 609–630 (2013) 2
9. Chang, H., Wang, P., Diao, W., Xu, G., Sun, X.: Remote sensing change detection with bitemporal and differential feature interactive perception. IEEE Transactions on Image Processing **33**, 4543–4555 (2024) 1
10. Chang, H., Wang, P., Diao, W., Xu, G., Sun, X.: A triple-branch hybrid attention network with bitemporal feature joint refinement for

---

[2] IGmLs is performed in MATLAB 2016a running on a Windows Laptop with Intel Core i9-10980HK CPU and 64 GB of RAM.

remote-sensing image semantic change detection. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–16 (2024) 1

11. Chen, H., Lan, C., Song, J., Broni-Bediako, C., Xia, J., Yokoya, N.: Objformer: Learning land-cover changes from paired osm data and optical high-resolution imagery via object-guided transformer. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–22 (2024) 1, 3
12. Chen, H., Yokoya, N., Chini, M.: Fourier domain structural relationship analysis for unsupervised multimodal change detection. ISPRS Journal of Photogrammetry and Remote Sensing **198**, 99–114 (2023) 3
13. Chen, H., Yokoya, N., Wu, C., Du, B.: Unsupervised multimodal change detection based on structural relationship graph representation learning. IEEE Transactions on Geoscience and Remote Sensing **60**, 1–18 (2022) 3
14. Deledalle, C.A., Denis, L., Tupin, F.: How to compare noisy patches? patch similarity beyond gaussian noise. International journal of computer vision **99**, 86–102 (2012) 5
15. Dellinger, F., Delon, J., Gousseau, Y., Michel, J., Tupin, F.: Sar-sift: a sift-like algorithm for sar images. IEEE Transactions on Geoscience and Remote Sensing **53**(1), 453–466 (2014) 4
16. Fukunaga, K., Hostetler, L.: Optimization of k nearest neighbor density estimates. IEEE Transactions on information theory **19**(3), 320–326 (2003) 14
17. Han, T., Tang, Y., Chen, Y., Yang, X., Guo, Y., Jiang, S.: Sdcgae: Structural difference compensation graph autoencoder for unsupervised multimodal change detection. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–16 (2024) 1
18. Han, T., Tang, Y., Chen, Y., Zou, B., Feng, H.: Global structure graph mapping for multimodal change detection. International Journal of Digital Earth **17**(1), 2347,457 (2024) 2, 10, 12
19. Han, T., Tang, Y., Yang, X., Lin, Z., Zou, B., Feng, H.: Change detection for heterogeneous remote sensing images with improved training of hierarchical extreme learning machine (helm). Remote Sensing **13**(23), 4918 (2021) 3
20. Han, T., Tang, Y., Zou, B., Feng, H.: Unsupervised multimodal change detection based on adaptive optimization of structured graph. International Journal of Applied Earth Observation and Geoinformation **126**, 103,630 (2024) 3, 10, 12
21. Hussain, M., Chen, D., Cheng, A., Wei, H., Stanley, D.: Change detection from remotely sensed images: From pixel-based to object-based approaches. ISPRS Journal of photogrammetry and remote sensing **80**, 91–106 (2013) 5
22. Ke, Y., Sukthankar, R.: Pca-sift: A more distinctive representation for local image descriptors. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., vol. 2, pp. II–II. IEEE (2004) 4
23. Li, C., Li, G., Wang, Z., Wang, X., Varshney, P.K.: Comic: An unsupervised change detection method for heterogeneous remote sensing images based on copula mixtures and cycle-consistent adversarial networks. Information Fusion p. 102240 (2024) 3
24. Li, J., Hu, Q., Ai, M.: Rift: Multi-modal image matching based on radiation-variation insensitive feature transform. IEEE Transactions on Image Processing **29**, 3296–3310 (2020) 4, 10
25. Li, J., Hu, Q., Zhang, Y.: Multimodal image matching: A scale-invariant algorithm and an open dataset. ISPRS Journal of Photogrammetry and Remote Sensing **204**, 77–88 (2023) 4
26. Li, J., Xu, W., Shi, P., Zhang, Y., Hu, Q.: Lnift: Locally normalized image for rotation invariant multimodal feature matching. IEEE Transactions on Geoscience and Remote Sensing **60**, 1–14 (2022) 4
27. Liu, T., Zhang, M., Gong, M., Zhang, Q., Jiang, F., Zheng, H., Lu, D.: Commonality feature representation learning for unsupervised multimodal change detection. IEEE Transactions on Image Processing **34**, 1219–1233 (2025) 3, 10, 12
28. Liu, Z., Li, G., Mercier, G., He, Y., Pan, Q.: Change detection in heterogenous remote sensing images via homogeneous pixel transformation. IEEE Transactions on Image Processing **27**(4), 1822–1834 (2018) 3, 10, 12
29. Liu, Z.g., Mercier, G., Dezert, J., Pan, Q.: Change detection in heterogeneous remote sensing images based on multidimensional evidential reasoning. IEEE Geoscience and Remote Sensing Letters **11**(1), 168–172 (2014) 3
30. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International journal of computer vision **60**, 91–110 (2004) 4
31. Luppino, L.T., Bianchi, F.M., Moser, G., Anfinsen, S.N.: Unsupervised image regression for heterogeneous change detection. IEEE Transactions on Geoscience and Remote Sensing **57**(12), 9960–9975 (2019) 2, 3, 10, 12
32. Luppino, L.T., Hansen, M.A., Kampffmeyer, M., Bianchi, F.M., Moser, G., Jenssen, R., Anfinsen, S.N.: Code-aligned autoencoders for unsupervised change detection in multimodal remote sensing images. IEEE Transactions on Neural Networks and Learning Systems **35**(1), 60–72 (2024) 3
33. Luppino, L.T., Kampffmeyer, M., Bianchi, F.M., Moser, G., Serpico, S.B., Jenssen, R., Anfinsen, S.N.: Deep image translation with an affinity-based change prior for unsupervised multimodal change detection. IEEE Transactions on Geoscience and Remote Sensing **60**, 1–22 (2022) 3
34. Lv, Z., Huang, H., Li, X., Zhao, M., Benediktsson, J.A., Sun, W., Falco, N.: Land cover change detection with heterogeneous remote sensing images: Review, progress, and perspective. Proceedings of the IEEE **110**(12), 1976–1991 (2022) 1
35. Lv, Z., Huang, H., Sun, W., Jia, M., Benediktsson, J.A., Chen, F.: Iterative training sample augmentation for enhancing land cover change detection performance with deep learning neural network. IEEE Transactions on Neural Networks and Learning Systems pp. 1–14 (2024) 3, 10, 12
36. Lv, Z., Liu, T., Benediktsson, J.A., Falco, N.: Land cover change detection techniques: Very-high-resolution optical images: A review. IEEE Geoscience and Remote Sensing Magazine pp. 2–21 (2021) 1, 2
37. Ma, J., Jiang, X., Fan, A., Jiang, J., Yan, J.: Image matching from handcrafted to deep features: A survey. International Journal of Computer Vision **129**, 23–79 (2021) 4
38. Ma, J., Zhao, J., Jiang, J., Zhou, H., Guo, X.: Locality preserving matching. International Journal of Computer Vision **127**, 512–531 (2019) 4
39. Ma, Z., Wang, R., Hao, F., Song, L.: Heterogeneous image change detection based on dual image translation and dual contrastive learning. IEEE Transactions on Geoscience and Remote Sensing pp. 1–14 (2024) 3
40. Marcos, D., Hamid, R., Tuia, D.: Geospatial correspondences for multimodal registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5091–5100 (2016) 4
41. McGregor, I.R., Connette, G., Gray, J.M.: A multi-source change detection algorithm supporting user customization and near real-time deforestation detections. Remote Sensing of Environment **308**, 114,195 (2024) 1
42. Mesquita, D.B., Campos, M.F.M., Nascimento, E.R.: Learning to detect changes in aerial images in the presence of registration errors. IEEE Geoscience and Remote Sensing Letters **19**, 1–5 (2022) 4
43. Mignotte, M.: A fractal projection and markovian segmentation-based approach for multimodal change detection. IEEE Transactions on Geoscience and Remote Sensing **58**(11), 8046–8058 (2020) 2, 3, 10, 12
44. Mitra, P., Murthy, C.A., Pal, S.K.: Unsupervised feature selection using feature similarity. IEEE transactions on pattern analysis and machine intelligence **24**(3), 301–312 (2002) 14
45. Morel, J.M., Yu, G.: Asift: A new framework for fully affine invariant image comparison. SIAM journal on imaging sciences **2**(2), 438–469 (2009) 4
46. Nguyen, L.H., Tran, T.D.: A sparsity-driven joint image registration and change detection technique for sar imagery. In: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 2798–2801. IEEE (2010) 2, 4
47. Otsu, N.: A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics **9**(1), 62–66 (1979) 9
48. Sedaghat, A., Ebadi, H.: Remote sensing image matching based on adaptive binning sift descriptor. IEEE transactions on geoscience and remote sensing **53**(10), 5283–5293 (2015) 4
49. Serra, P., Pons, X., Sauri, D.: Post-classification change detection with data from different sensors: some accuracy considerations. International Journal of Remote Sensing **24**(16), 3311–3340 (2003) 3
50. Shi, J., Wu, T., Qin, A.K., Lei, Y., Jeon, G.: Self-guided autoencoders for unsupervised change detection in heterogeneous remote sensing

images. IEEE Transactions on Artificial Intelligence pp. 1–13 (2024) 3

51. SINGH, A.: Review article digital change detection techniques using remotely-sensed data. International Journal of Remote Sensing **10**(6), 989–1003 (1989) 1
52. Song, S., Jin, K., Zuo, B., Yang, J.: A novel change detection method combined with registration for sar images. Remote Sensing Letters **10**(7), 669–678 (2019) 2, 4
53. Sun, Y., Lei, L., Guan, D., Kuang, G.: Iterative robust graph for unsupervised change detection of heterogeneous remote sensing images. IEEE Transactions on Image Processing **30**, 6277–6291 (2021) 2, 6, 10, 12
54. Sun, Y., Lei, L., Guan, D., Kuang, G., Li, Z., Liu, L.: Locality preservation for unsupervised multimodal change detection in remote sensing imagery. IEEE Transactions on Neural Networks and Learning Systems pp. 1–15 (2024) 1, 2, 10, 12
55. Sun, Y., Lei, L., Guan, D., Li, M., Kuang, G.: Sparse-constrained adaptive structure consistency-based unsupervised image regression for heterogeneous remote sensing change detection. IEEE Transactions on Geoscience and Remote Sensing pp. 1–14 (2022) 2, 3, 10, 12
56. Sun, Y., Lei, L., Li, X., Tan, X., Kuang, G.: Structure consistency-based graph for unsupervised change detection with homogeneous and heterogeneous remote sensing images. IEEE Transactions on Geoscience and Remote Sensing pp. 1–21 (2022) 1, 3, 6
57. Sun, Y., Lei, L., Li, Z., Kuang, G.: Similarity and dissimilarity relationships based graphs for multimodal change detection. ISPRS Journal of Photogrammetry and Remote Sensing **208**, 70–88 (2024) 2, 3
58. Sun, Y., Lei, L., Tan, X., Guan, D., Wu, J., Kuang, G.: Structured graph based image regression for unsupervised multimodal change detection. ISPRS Journal of Photogrammetry and Remote Sensing **185**, 16–31 (2022) 3
59. Sundaresan, A., Varshney, P.K., Arora, M.K.: Robustness of change detection algorithms in the presence of registration errors. Photogrammetric Engineering & Remote Sensing **73**(4), 375–383 (2007) 4
60. Touati, R.: Détection de changement en imagerie satellitaire multimodale. Ph.D. thesis, Université de Montréal (2019) 1, 10, 12
61. Vakalopoulou, M., Karantzalos, K., Komodakis, N., Paragios, N.: Simultaneous registration and change detection in multitemporal, very high resolution remote sensing data. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 61–69 (2015) 2, 4
62. Vakalopoulou, M., Karantzalos, K., Komodakis, N., Paragios, N.: Graph-based registration, change detection, and classification in very high resolution multitemporal remote sensing data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **9**(7), 2940–2951 (2016) 2, 4
63. Wan, L., Xiang, Y., You, H.: An object-based hierarchical compound classification method for change detection in heterogeneous optical and SAR images. IEEE Transactions on Geoscience and Remote Sensing **57**(12), 9941–9959 (2019) 3
64. Wang, J.J., Dobigeon, N., Chabert, M., Wang, D.C., Huang, T.Z., Huang, J.: Cd-gan: A robust fusion-based generative adversarial network for unsupervised remote sensing change detection with heterogeneous sensors. Information Fusion p. 102313 (2024) 3
65. Woodcock, C.E., Loveland, T.R., Herold, M., Bauer, M.E.: Transitioning from change detection to monitoring with remote sensing: A paradigm shift. Remote Sensing of Environment **238**, 111,558 (2020) 1
66. Wu, H., Geng, J., Jiang, W.: Multidomain constrained translation network for change detection in heterogeneous remote sensing images. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–16 (2024) 3
67. Xing, Y., Zhang, Q., Ran, L., Zhang, X., Yin, H., Zhang, Y.: Improving reliability of heterogeneous change detection by sample synthesis and knowledge transfer. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–11 (2024) 3
68. Xu, J., Liu, T., Lei, T., Chen, H., Yokoya, N., Lv, Z., Gong, M.: Cgsl: Commonality graph structure learning for unsupervised multimodal change detection. ISPRS Journal of Photogrammetry and Remote Sensing **229**, 92–106 (2025) 3, 10, 12
69. Ye, Y., Bruzzone, L., Shan, J., Bovolo, F., Zhu, Q.: Fast and robust matching for multimodal remote sensing image registration. IEEE Transactions on Geoscience and Remote Sensing **57**(11), 9059–9070 (2019) 2, 4
70. Ye, Y., Shan, J., Bruzzone, L., Shen, L.: Robust registration of multimodal remote sensing images based on structural similarity. IEEE Transactions on Geoscience and Remote Sensing **55**(5), 2941–2958 (2017) 4
71. Zhang, H., Lei, L., Ni, W., Tang, T., Wu, J., Xiang, D., Kuang, G.: Optical and sar image matching using pixelwise deep dense features. IEEE Geoscience and Remote Sensing Letters **19**, 1–5 (2022) 4
72. Zhang, H., Lei, L., Ni, W., Yang, X., Tang, T., Cheng, K., Xiang, D., Kuang, G.: Optical and sar image dense registration using a robust deep optical flow framework. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **16**, 1269–1294 (2023) 2
73. Zhang, H., Ni, W., Yan, W., Xiang, D., Wu, J., Yang, X., Bian, H.: Registration of multimodal remote sensing image based on deep fully convolutional neural network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **12**(8), 3028–3042 (2019) 4
74. Zhao, B., Sui, H., Liu, J., Shi, W., Wang, W., Xu, C., Wang, J.: Flood inundation monitoring using multi-source satellite imagery: a knowledge transfer strategy for heterogeneous image change detection. Remote Sensing of Environment **314**, 114,373 (2024) 1
75. Zheng, Z., Ermon, S., Kim, D., Zhang, L., Zhong, Y.: Changen2: Multi-temporal remote sensing generative change foundation model. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(2), 725–741 (2025) 3
76. Zheng, Z., Zhong, Y., Ma, A., Zhang, L.: Single-temporal supervised learning for universal remote sensing change detection. International Journal of Computer Vision pp. 1–21 (2024) 1
77. Zheng, Z., Zhong, Y., Wang, J., Ma, A., Zhang, L.: Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters. Remote Sensing of Environment **265**, 112,636 (2021) 1
78. Zheng, Z., Zhong, Y., Zhang, L., Ermon, S.: Segment any change. Advances in Neural Information Processing Systems **37**, 81,204–81,224 (2024) 3
79. Zhou, L., Ye, Y., Tang, T., Nan, K., Qin, Y.: Robust matching for sar and optical images using multiscale convolutional gradient features. IEEE Geoscience and Remote Sensing Letters **19**, 1–5 (2022) 4
80. Zhou, R., Quan, D., Wang, S., Lv, C., Cao, X., Chanussot, J., Li, Y., Jiao, L.: A unified deep learning network for remote sensing image registration and change detection. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–16 (2024) 4
81. Zhu, L., Sun, W., Fan, D., Xing, H., Liu, X.: Unsupervised spatial self-similarity difference-based change detection method for multi-source heterogeneous images. Pattern Recognition **149**, 110,237 (2024) 3, 10, 12