Pattern Recognition 177 (2026) 113275

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Leveraging image transformation and optical flow for heterogeneous change detection under co-registration errors

Yuli Sun [a,b], Lin Lei [a], Gangyao Kuang [a,*]

[a] College of Electronic Science and Technology, National University of Defense Technology, Changsha, China
[b] College of Aerospace Science and Engineering, National University of Defense Technology, Changsha, China

ARTICLE INFO

Keywords:
Change detection
Co-registration errors
Heterogeneous images
Image transformation
Optical flow

ABSTRACT

Heterogeneous change detection (HeCD) is a significant yet challenging task in remote sensing. Most existing methods rely on well-registered images, making them vulnerable in real-world scenarios. To address this issue, we propose a unified model that jointly integrates graph-based image transformation and optical flow-based image registration, named GiToF, for HeCD task in the presence of co-registration errors. Unlike conventional approaches that separately perform registration and change detection, GiToF formulates them into a collaborative framework and employs three tailored constraints to guide the unified model: the structural consistency for image transformation of unchanged image, the Lucas-Kanade optical flow constraint for image registration of matched image, and the prior sparsity for changed image. This unified model allows GiToF to mitigate the impact of co-registration errors on image transformation and change estimation via optical flow-based matching, while image transformation and change estimation ensure that the brightness constancy and smooth motion assumptions of optical flow are satisfied, which are typically violated due to image heterogeneity and drastic changes in the HeCD problem. Therefore, GiToF provides a more robust HeCD solution by seamlessly coupling image transformation, registration, and change estimation. Extensive experiments on five benchmark datasets demonstrate that GiToF achieves superior performance and robustness compared with state-of-the-art methods under co-registration errors.

1. Introduction

1.1. Background

Heterogeneous change detection (HeCD) in remote sensing refers to the technique of extracting information on land cover changes by comparing multiple images of the same area captured at different times by different sensors, e.g., optical sensor and synthetic aperture radar (SAR). With the continuous advancement of remote sensing related technology, HeCD has become an increasingly interesting topic [1], as it can break through the limitation of traditional homogeneous change detection (HoCD) that relies on images from the same sensor with similar imaging conditions [2,3].

HeCD is particularly crucial for emergency response in the event of natural disasters such as earthquakes and floods [4], where homogeneous multitemporal images are difficult to obtain urgently due to imaging constraints, such as unfavorable lighting and weather conditions (e.g., disasters accompanied torrential rain, cloudy skies, dense fog, etc.), and satellite re-entry cycle limitations [5,6]. In this case, HeCD can utilize emergency-acquired post-event SAR images (leveraging the all-weather, all-time imaging capabilities of SAR) and pre-event optical images (given the abundant historical resources of optical imagery), to rapidly detect the affected areas and analyze the disaster situations.

1.2. Related work

Due to the fact that the images used in HeCD are acquired from different sensors under varying imaging conditions, the same land cover features are represented quite differently in heterogeneous images, which presents significant challenges for HeCD [7].

First, the heterogeneity of multitemporal images prevents HeCD from utilizing direct comparison methods to extract change information, as is done in HoCD [8,9]. Consequently, the primary challenge in HeCD is how to establish connections between heterogeneous images and make them comparable, which is the focus of current research in the field [10].

Existing HeCD methods can generally be divided into five categories based on how they make heterogeneous images comparable: (i) Image

---

[*] Corresponding author.
E-mail address: kuanggangyao@nudt.edu.cn (G. Kuang).

https://doi.org/10.1016/j.patcog.2026.113275
Received 1 June 2025; Received in revised form 22 January 2026; Accepted 7 February 2026

classification-based methods perform the post-categorization comparisons by segmenting heterogeneous images into the same land cover categories and then comparing the segmentation results to identify changes [11]. (ii) Similarity metric-based methods typically assume certain relationships between heterogeneous images that are independent of imaging conditions [12], and use these imaging-modal invariance assumptions to construct similarity metrics to measure changes [13,14]. (iii) Deep feature transformation-based methods use labeled data to train Siamese or pseudo-Siamese networks [15,16], transforming heterogeneous images into a shared latent feature space for comparison [17]. (iv) Image translation-based methods employ traditional regression functions [18–20] or utilize generative adversarial networks (GANs) and their variants [21] to transform one image into another image domain, performing style migration to convert heterogeneous images into homogeneous ones for comparison [22,23]. (v) Energy model-based methods typically utilize Markov random fields (MRF) to directly model the HeCD problem, which flexibly convert the change label assignment task into an energy minimization problem by constructing reasonable constraints [24–26].

Second, the significant differences in imaging conditions between multitemporal images pose substantial challenges for achieving precise registration of heterogeneous images. It is well known that image registration is a critical preprocessing step in change detection problems. In previous research, it has often been assumed that the multitemporal images being compared are already well-registered, wherein pixels at the same positions in the images correspond to the same geographical regions. Simultaneously, image registration and change detection are often treated as two independent research tasks. Currently, the registration of homogeneous images from the same sensor is well-established, which often leads to the neglect of co-registration errors in HoCD. However, the registration of heterogeneous images acquired from different sensors with varying imaging conditions poses a significant challenge in itself [27,28].[1] A considerable amount of research is still focused on improving the accuracy of heterogeneous image registration, such as the variational inference based methods [28,32], adjustable threshold and variational mixture transformation method [30], dynamic threshold calculation strategy based method [33]. However, for heterogeneous images involving large areas, high resolution, and complex scenarios [34,35], achieving pixel-level matching accuracy remains highly challenging [36,37]. Consequently, it is essential to consider the impact of co-registration errors in the HeCD task. Unfortunately, most of the current research on HeCD methods focus primarily on the first aspect of “making heterogeneous images comparable”, often overlooking the practical issue of residual misregistration, which can severely affect detection performance.

Currently, only a handful of studies jointly consider change detection and image registration tasks in tandem, but all of these studies focus solely on HoCD, without addressing the more complex challenges associated with HeCD. For example, in the case of HoCD with SAR images, a sparsity-driven joint change detection and registration method has been proposed [38], which involves creating a patch dictionary and using sparse representation to determine the optimal matches, with the reconstruction error serving as the indicator of change. Song et al. [39] proposed an optimization framework that decomposes the observed image into three components: an affine transformation based wrapped image, a sparse representation of the changes, and the accompanying image noise. For HoCD of optical images, an MRF-based energy model has been proposed [40], which incorporates three energy terms: change detection, registration, and a coupling term. Mesquita et al. [41] trained an end-to-end network for detecting changes in optical aerial images with co-registration errors by extracting common features for both change detection and registration tasks. On the other hand, Zhou et al. [42] found that most of the deep features extracted by the change detection network and the image registration network exhibit high cosine similarity, and they proposed a unified network designed to handle the separate tasks of change detection and registration, aiming to minimize training and memory requirements. However, this network neither accounts for change detection in the presence of co-registration errors nor considers heterogeneous images. Clearly, existing research primarily focuses on HoCD with misregistration, overlooking the more complex challenges posed by HeCD with co-registration errors.

1.3. Motivations and contributions

To fill the gap and enhance the practical usability of HeCD, this paper focuses on detecting changes between heterogeneous images under coarse registration, i.e., imperfect registration with residual errors. The goal of our work is not to simply enhance fine-grained registration performance, but to mitigate its influence on change detection by introducing a joint model that integrates change detection with registration, thereby reducing the reliance on precise pixel-level correspondence.

We first analyze the impact of co-registration errors on HeCD, particularly in image translation-based HeCD. On the one hand, misregistration directly causes change detection errors at the boundaries, that is, at the junctions of different land cover types, co-registration errors can make unchanged areas appear as if they have changed. For instance, when two adjacent land covers (such as forest and water bodies) are not accurately aligned, change detection algorithms may mistakenly identify these regions as having undergone change, resulting in false alarms. On the other hand, due to the distorted spatial relationships, misregistration indirectly affects the accuracy of image transformation, which is a crucial step in HeCD for enabling the comparison of heterogeneous images. It is clear that, given the unique nature of HeCD, the impact of co-registration errors is amplified, presenting a complex, overarching challenge that affects the entire workflow and undermines the reliability and accuracy of change detection results. Furthermore, the image registration process is also affected by the changed regions, i.e., the two images cannot be properly aligned in areas where changes have occurred. Therefore, it is crucial to consider how to filter out the influence introduced by these changed areas during the image registration process. It is evident that change detection and image registration are highly coupled, which presents significant challenges for HeCD under misregistration conditions.

In this paper, an unsupervised HeCD method for imperfectly registered images by leveraging **G**raph-based **i**mage **T**ransformation and **o**ptical **F**low (referred to as GiToF) is proposed. Specifically, GiToF jointly addresses the change detection and registration tasks by combining image translation-based comparison with Lucas-Kanade (LK) optical flow [43] based registration. In the HeCD problem with co-registration errors, relying solely on image transformation or optical flow methods is ineffective. Firstly, as analyzed above, co-registration errors significantly impact the image transformation process. Secondly, the fundamental assumptions of the Lucas-Kanade optical flow method are difficult to satisfy in HeCD problems: the brightness constancy assumption is violated due to the heterogeneity of the images, as the same object may appear significantly different in images captured by different sensors under varying imaging conditions; additionally, changes in the images disrupt the small-motion and spatial coherence constraints, as land cover changes are typically significant and drastic.

---

[1] Although some recent studies have achieved pixel-level or even sub-pixel registration accuracy on certain benchmark datasets, this level of precision is often not attainable in practical scenarios [29]. On the one hand, these methods typically require abundant, well-aligned training data, which are difficult to obtain in real-world applications. On the other hand, many of these methods are evaluated on small images (e.g., 256×256 or 512×512) from flat or homogeneous regions, where geometric deformation is limited. In contrast, real-world applications frequently involve large-scale scenes with significant terrain variation and structural complexity [30,31], where it is difficult to model global deformation using a single transformation function. Under such conditions, local misalignments are common and hard to eliminate.

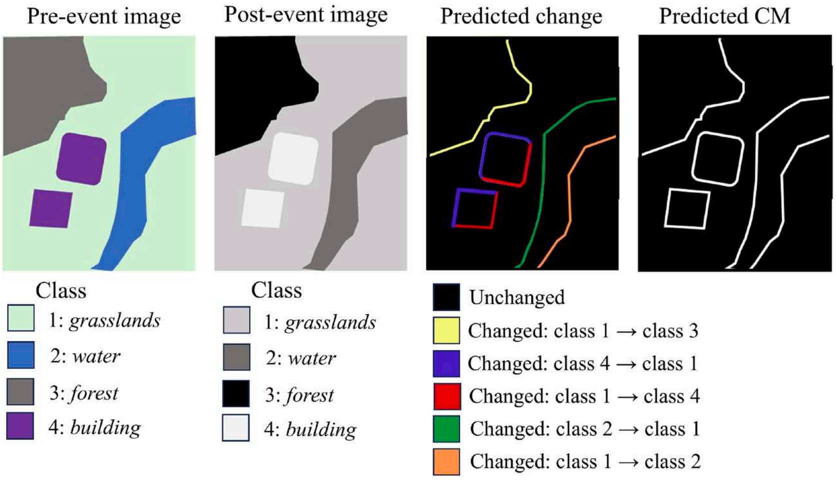


Fig. 1. An example about the influence of co-registration errors on HeCD. There are four classes of objects in pre- and post-event images, and no real change has happened. Due to the 1-pixel shift in both the horizontal and vertical directions, 5 kinds of false changes appear.

To address these challenges, we propose a joint model by decomposing the well-registered post-event image ($\mathbf{Y}_r$) into unchanged translated image ($\mathbf{Y}_t$) and changed image ($\boldsymbol{\Delta}$) caused by the event as $\mathbf{Y}_r = \mathbf{Y}_t - \boldsymbol{\Delta}$. We then employ a structural consistency constraint to ensure that the unchanged translated image and the original pre-event image ($\mathbf{X}$) share the same topological structure in the graph. Additionally, we apply the Lucas-Kanade optical flow method to solve for the displacements ($\mathbf{s}$) between the original and well-registered post-event images, expressed as $\mathbf{Y}_r = \mathbf{Y}\circ\mathbf{s}$, which satisfies the brightness constancy and spatial coherence assumptions. Finally, we use prior sparsity knowledge to constrain the changed image with sparseness penalty term. In this way, we integrate image transformation, image matching, and change estimation into a unified model, allowing these three sub-processes to mutually reinforce each other. By iteratively solving the model, GiToF gradually reduces the impact of transformation errors, co-registration errors, and changes on other sub-processes, ultimately improving the performance of change detection. The main contributions of this work can be summarized as follows:

- We propose GiToF, a unified joint model for unsupervised change detection of heterogeneous images with co-registration errors. GiToF couples graph-based image transformation with optical flow-based registration in a collaborative framework, thereby improving the robustness of HeCD in real-world applications where accurate co-registration is difficult to achieve.
- We formulate image transformation, registration, and change estimation within a unified framework that promotes collaboration among the sub-tasks, enabling GiToF to reduce the impact of co-registration errors through optical flow-based matching, while ensuring that the brightness constancy and smooth motion assumptions of optical flow are satisfied, which are typically violated in heterogeneous images with drastic changes.
- We introduce a multi-scale coarse-to-fine strategy with MRF-based fusion, which improves robustness and detection quality in three aspects: enhancing the ability of the Lucas-Kanade optical flow to handle large displacements; enabling progressive refinement of both matching and change detection results; and leveraging multi-scale fusion of changed images to further boost detection performance.

2. The influence of co-registration errors on HeCD

Given a pair of heterogeneous images acquired at different times t_1 and t_2 respectively, the pre-event image $\mathbf{X} \in \mathbb{R}^{M\times N\times C_x}$ belongs to domain $\mathcal{X}$, and the post-event image $\mathbf{Y} \in \mathbb{R}^{M\times N\times C_y}$ belongs to $\mathcal{Y}$, where M, N, C_x and C_y denote the height, width, and channel numbers of $\mathbf{X}$ and $\mathbf{Y}$, respectively. By utilizing the inherent geographical coordinate references of remote sensing images, the heterogeneous images are coarsely registered without very large-angle rotations, or scale differences.

We define the pixels in images as $x(m,n,c)$ and $y(m,n,c)$, and indicate the unchanged and changed hypotheses by H_0 and H_1, respectively. In traditional HoCD with well-registered images, since the pre- and post-event images are acquired by the same sensor, $x(m,n,c)$ and $y(m,n,c)$ exhibit similar and distinct statistical properties under the H_0 and H_1 conditions, respectively. In this case, changes can be detected using direct comparison methods, such as the change vector analysis (CVA) [44] in HoCD of multispectral images and the ratio/log-ratio operators in HoCD of SAR images [45]. However, in HeCD with imperfectly registered images, on the one hand, the assumption that $x(m,n,c)$ and $y(m,n,c)$ follow similar statistical properties under the condition H_0 is no longer valid, making direct comparison impossible; on the other hand, the spatial misalignment caused by co-registration errors further complicates their comparison. Specifically, the impact of co-registration errors on HeCD manifests in the image comparison and image transformation as follows.

Firstly, we disregard the influence of co-registration errors on image transformation and focus solely on their influence on image comparison, assuming that heterogeneous images can be accurately transformed into homogeneous images for comparison. As illustrated in Fig. 1, there is a 1-pixel translation error in both the horizontal and vertical directions between the pre- and post-event images. This translation error results in two main flaws in change detection: (i) co-registration errors directly introduce detection errors at boundaries. Specifically, at the junctions of different land cover types, co-registration errors may cause unchanged areas to appear as if they have changed. Generally, the larger the co-registration error and the more “fragmented” the objects contained in the images, the more serious the false detections will be. (ii) It implicitly increases the diversity of detected changes, elevating the variance of changed pixels in the difference image (DI), which in turn significantly affects both the extraction of changed regions and the classification of changes.

Secondly, we focus on the influence of co-registration errors on image transformation. In HeCD, the heterogeneous images need be transformed into the same domain to enable comparison, e.g., translating the pre-event image to the domain of post-event image as $\mathcal{X}^{M\times N\times C_x} \rightarrow \mathcal{Y}^{M\times N\times C_y}$. Most image transformation methods are not pixel-independent, meaning that each pixel is inevitably affected by others, such as spatially neighboring pixels or feature-similar pixels. The incorrect mappings and distorted spatial relationships caused by misregistration make the unmatched pixels interact with each other during

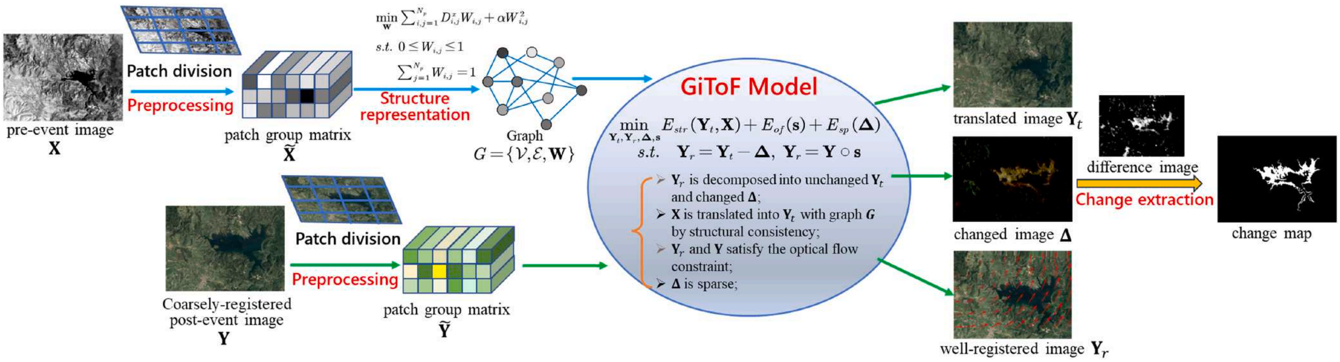


Fig. 2. Framework of the proposed GiToF for HeCD with co-registration errors.

the image transformation process, further exacerbating the difficulty of distinguishing between changed and unchanged areas. It can be found that due to the unique nature of HeCD, the impact of co-registration errors is amplified, presenting a complex, overarching challenge that affects the entire workflow and diminishes the reliability and accuracy of change detection results. In the experimental section, we present the results of HeCD with co-registration errors (e.g. Fig. 7), from which it can be seen that misregistration seriously affect the performance of change detection. This effect is not only observed as false detections at the object boundaries (which could potentially be mitigated through post-processing techniques), but more critically, it also leads to a degradation in overall change detection performance, as illustrated by the false detections in non-edge areas, which are hard to eliminate through post-processing.

A natural question may arise: why not perform registration first and then conduct change detection? What advantages does considering registration and change detection simultaneously offer over this step-by-step approach?

We argue that, **firstly**, the accuracy of heterogeneous image registration is a well-recognized challenge. Current research struggles to achieve rigorous matching for heterogeneous images of arbitrarily complex scenes, particularly when dealing with high resolution, large areas, terrain variations, and geometric distortions (such as those found in SAR images). In such cases, co-registration errors can sometimes exceed tens of pixels. **Secondly**, the interference of changes on registration complicates the process further. Specifically, the presence of changes disrupts image registration, as it is inherently impossible to establish matching relationships in regions that have changed. This necessitates the removal of changed regions during the image registration process. **Lastly**, high-precision image registration algorithms can be highly complex, yet subpixel-level accuracy is not always required for any arbitrary change detection tasks. In practice, it is often sufficient to ensure that the change detection process is robust to co-registration errors, while maintaining the timeliness of the algorithm for real-world applications, such as disaster assessment and dynamic monitoring.

Therefore, these three points drive us to explore the integration of change detection and registration into a single model. *The goal is not to achieve more precise registration, but rather to ensure that reliable change detection results can still be obtained even in the presence of large co-registration errors*, thereby achieving strong robustness and accuracy against such discrepancies.

3. Methodology

The framework of the GiToF for HeCD under misregistration is illustrated in Fig. 2. GiToF involves of three steps: 1) preprocessing, 2) building and solving the joint image transformation and optical flow model, and 3) fusing and segmenting the DIs to extract the changed regions.

In the change detection task, it is required that the multitemporal images are comparable and spatially aligned within the same image or feature domain. In the optical flow-based image registration task, it is necessary for the multitemporal images to satisfy the brightness constancy assumption and the small and smooth motion assumption. However, in the case of HeCD task, both of these assumptions are violated: the multitemporal images are captured from different sensors under varying acquisition conditions, and there are drastic changes between the images.

To meet these constraints, the proposed joint framework utilizes an image decomposition model as

$$\begin{aligned} \min_{\mathbf{Y}_t,\mathbf{Y}_r,\mathbf{\Delta},\mathbf{s}} \ & E_{str}\left(\mathbf{Y}_t,\mathbf{X}\right) + E_{of}(\mathbf{s}) + E_{sp}(\mathbf{\Delta}) \\ s.t.\ \ & \mathbf{Y}_r = \mathbf{Y}_t - \mathbf{\Delta},\ \mathbf{Y}_r = \mathbf{Y} \circ \mathbf{s} \end{aligned} \tag{1}$$

where E_{str} is the structural consistency based regularization (5), E_{of} is the Lucas-Kanade optical flow based regularization (8), and E_{sp} is the prior sparsity based regularization (11).

GiToF decomposes the matched post-event image $\mathbf{Y}_r$ into the unchanged translated image $\mathbf{Y}_t$ and the changed image $\mathbf{\Delta}$ caused by event as $\mathbf{Y}_r = \mathbf{Y}_t - \mathbf{\Delta}$, and uses three regularization terms for this decomposition: 1) E_{str} is used to ensure that the unchanged translated image $\mathbf{Y}_t$ and the original pre-event image $\mathbf{X}$ maintain the same topological structure in the similarity graph; 2) E_{of} is used for solving the displacement $\mathbf{s}$ between original and matched post-event images as $\mathbf{Y}_r = \mathbf{Y} \circ \mathbf{s}$; 3) E_{sp} is based on the fact that only a small portion of regions undergo changes in the real-world scenarios, while the majority of regions remain unchanged. Next, we describe each of these constraints.

3.1. Graph based image transformation

In the proposed GiToF, it transforms the pre-event image into the domain of post-event image, defined as $\mathbf{Y}_t = \mathcal{M}(\mathbf{X}) : \mathcal{X}^{M\times N\times C_x} \to \mathcal{Y}^{M\times N\times C_y}$ with $\mathcal{M}$ denoting the translation function. Recently, we have explored image transformation methods based on structural consistency [10,19], which can be expressed as follows: when imaging the same scene with different types of sensors, although the acquired images may differ in appearance, the topological relationships between objects within the images remain consistent. The structural consistency constraint can be achieved by constructing a graph model that represents the structure of the original image, and then ensuring that the transformed image retains the same structural characteristics within this graph model.

We first divide the pre-event image $\mathbf{X}$ into small non-overlapping square patches with size $p \times p$, and uses replication padding operation for empty region at the edges when dimensions do not evenly divide. Then, these patches are vectorized and stacked into a patch group matrix (PGM), denoted as $\tilde{\mathbf{X}} \in \mathbb{R}^{C_x p^2 \times N_p}$ with $N_p = \lceil M/p \rceil \cdot \lceil N/p \rceil$ and $\lceil \cdot \rceil$ representing the up rounding. For the translated image $\mathbf{Y}_t$, we can also divide it into square patches and obtain the PGM of $\tilde{\mathbf{Y}}_t \in \mathbb{R}^{C_y p^2 \times N_p}$ with the same operation. We denote this operation of obtaining a PGM from an image as $\mathcal{P}$, and the inverse operation, which transforms a PGM back to an image, as $\mathcal{P}^{-1}$.

We construct a probabilistic graph $G = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$ to characterize the structure feature of pre-event image $\mathbf{X}$ by setting each square patch as vertex with $\mathcal{V} = \left\{\tilde{\mathbf{X}}_1, \ldots, \tilde{\mathbf{X}}_{N_p}\right\}$, and connecting it with neighboring patches with weights $\mathbf{W}$ as

$$\begin{aligned} &\min_{\mathbf{W}} \sum_{i,j=1}^{N_p} D^x_{i,j} W_{i,j} + \alpha W^2_{i,j} \\ &s.t.\ 0 \le W_{i,j} \le 1,\ \sum_{j=1}^{N_p} W_{i,j} = 1 \end{aligned} \tag{2}$$

where $D^x_{i,j}$ represents the distance between the ith and jth patches of pre-event image, and α is a tuning parameter for adjusting the neighborhood number. According to [19], the closed-form optimal solution of problem (2) is

$$W_{i,(j)} = \begin{cases} \frac{D^x_{i,(k+1)} - D^x_{i,(j)}}{k D^{t1}_{i,(k+1)} - \sum_{h=1}^{k} D^x_{i,(h)}}, j \le k \\ 0, \qquad j > k \end{cases} \tag{3}$$

where $D^x_{i,(j)}$ denotes the jth smallest distance in the vector $\mathbf{D}^x_i$, and the symbol (j) in $D^x_{i,(j)}$ denotes the index of the jth smallest value. In order to choose an appropriate k for each vertex, we directly apply the k-selection strategy in [19], which has also been shown to effectively characterize image structure in other studies [13,46]. Although the constructed probabilistic graph can be viewed as a k-nearest neighbors (KNN) graph in terms of connectivity, it differs from conventional KNN graphs in that both the neighbor selection and the edge weights are adaptively determined according to the underlying data distribution.

The structure consistency between the pre-event image $\mathbf{X}$ and the translated image $\mathbf{Y}_t$ requires that they share the same similarity relationships in the graph G. That is, if the square patches $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{X}}_j$ connect by G are very similar (with large edge weight $W_{i,j}$), then the square patches $(\tilde{\mathbf{Y}}_t)_i$ and $(\tilde{\mathbf{Y}}_t)_j$ in translated image $\mathbf{Y}_t$ should be also very similar. Therefore, we can obtain the following regularization

$$E_{str}\left(\mathbf{Y}_t, \mathbf{X}\right) = \sum_{i,j=1}^{N_p} \left\| (\tilde{\mathbf{Y}}_t)_i - (\tilde{\mathbf{Y}}_t)_j \right\|_2^2 W_{i,j} \tag{4}$$

which is similar as the first term in the graph construction model (2). Defining $\mathbf{L}^x = \mathbf{D}^g - \left(\mathbf{W} + \mathbf{W}^T\right)/2$ as the graph Laplacian matrix with $\mathbf{D}^g$ being the diagonal degree matrix with $D^g_{i,i} = \sum_{j=1}^{N_p} \left(W_{i,j} + W_{j,i}\right)/2$, the structural consistency based regularization (4) can be rewritten as

$$E_{str}\left(\mathbf{Y}_t, \mathbf{X}\right) = 2\mathrm{Tr}\left(\tilde{\mathbf{Y}}_t \mathbf{L}^x (\tilde{\mathbf{Y}}_t)^T\right) \tag{5}$$

where $\mathrm{Tr}(\cdot)$ denotes the matrix trace.

3.2. Optical flow based registration

In the optical flow based image registration task, it requires the multitemporal images to satisfy the brightness constancy assumption and the small and smooth motion assumption, which are both violated in the HeCD task due to the heterogeneity of the images and the presence of changes. Therefore, in the proposed GiToF, we first translate the pre-event $\mathbf{X}$ into $\mathbf{Y}_t$ and then divide $\mathbf{Y}_t$ into the well-registered image $\mathbf{Y}_r$ and changed image $\mathbf{\Delta}$ as $\mathbf{Y}_t = \mathbf{Y}_r + \mathbf{\Delta}$. Afterward, we proceed to search for the registration between the $\mathbf{Y}_r$ and $\mathbf{Y}$ as $\mathbf{Y}_r = \mathbf{Y} \circ \mathbf{s}$, with $\mathbf{s} = (\mathbf{u}, \mathbf{v})$ denoting the displacements. Then, for each pixel, we have

$$y_r(m, n, c) = y\left(m + u_{m,n}, n + v_{m,n}, c\right) \tag{6}$$

where $\mathbf{u} \in \mathbb{R}^{M \times N}$ and $\mathbf{v} \in \mathbb{R}^{M \times N}$ represent the horizontal and vertical flow components, respectively. By assuming the displacements are small, with the local Taylor series approximations and abbreviating $u_{m,n}, v_{m,n}$ to u, v for simplifying the notation, we have

$$y(m + u, n + v, c) = y(m, n, c) + u F^c_u + v F^c_v \tag{7}$$

where F^c_u and F^c_v denote the partial derivatives of $F^c_u = \frac{\partial y(m,n,c)}{\partial m}$ and $F^c_v = \frac{\partial y(m,n,c)}{\partial n}$.

The Lucas-Kanade method assumes the displacements are approximately constant within a neighborhood, and calculates the flow vector by the least squares principle of the following minimization problem

$$\begin{aligned} E_{of}(u, v) &= \sum_{c=1}^{C_y} N_w * \left(\left(y(m + u, n + v, c) - y_r(m, n, c)\right)^2\right) \\ &= \sum_{c=1}^{C_y} N_w * \left(\left(u F^c_u + v F^c_v + F^c_t\right)^2\right) \end{aligned} \tag{8}$$

where N_w denotes a standard Gaussian template with size $w \times w$, and $F^c_t = y(m, n, c) - y_r(m, n, c)$. By setting $\frac{\partial E_{of}}{\partial u} = 0$ and $\frac{\partial E_{of}}{\partial v} = 0$ and using similar notation as in [37], we have

$$\begin{pmatrix} N_w * (\mathbf{F}_u \bullet \mathbf{F}_u) & N_w * (\mathbf{F}_u \bullet \mathbf{F}_v) \\ N_w * (\mathbf{F}_u \bullet \mathbf{F}_v) & N_w * (\mathbf{F}_v \bullet \mathbf{F}_v) \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -N_w * (\mathbf{F}_u \bullet \mathbf{F}_t) \\ -N_w * (\mathbf{F}_v \bullet \mathbf{F}_t) \end{pmatrix} \tag{9}$$

where

$$\begin{aligned} \mathbf{F}_u \bullet \mathbf{F}_u &= \textstyle\sum_{c=1}^{C_y} \left(F^c_u\right)^2 \\ \mathbf{F}_u \bullet \mathbf{F}_v &= \textstyle\sum_{c=1}^{C_y} \left(F^c_u F^c_v\right) \\ \mathbf{F}_u \bullet \mathbf{F}_t &= \textstyle\sum_{c=1}^{C_y} \left(F^c_u F^c_t\right) \end{aligned} \tag{10}$$

3.3. Prior sparsity term

In practical change detection problems, we have prior knowledge about the sparsity of changes, that is, only a small portion of the regions undergo changes in real-world scenarios, while the majority remain unchanged. This prior sparsity is widely utilized in other change detection methods [19,25,26]. Intuitively, the $\ell_{2,0}$-norm based $\|\mathbf{\Delta}\|_{2,0}$ that equals to the number of changed pixels should be chosen as the sparse penalty function. However, considering the non-convexity of the $\ell_{2,0}$-norm, we instead use its convex relaxation of $\ell_{2,1}$-norm regularization as

$$E_{sp}(\mathbf{\Delta}) = \textstyle\sum_{m,n=1}^{N_p} \left(\sum_{c=1}^{C_y} (\Delta(m, n, c))^2\right)^{1/2} \tag{11}$$

where $E_{sp}(\mathbf{\Delta})$ degenerates to the ℓ_1-norm regularization when $C_y = 1$.

3.4. Objective function and optimization details

By incorporating the regularization terms of E_{str} (5), E_{of} (8), E_{sp} (11), and substituting them into (1), and introducing a Lagrange multiplier $\tilde{\mathbf{R}} \in \mathbb{R}^{C_y p^2 \times N_p}$, we have augmented Lagrangian function

$$\begin{aligned} \Theta\left(\mathbf{Y}_t, \mathbf{Y}_r, \mathbf{\Delta}, \mathbf{s}\right) = {} & 2\mathrm{Tr}\left(\tilde{\mathbf{Y}}_t \mathbf{L}^x (\tilde{\mathbf{Y}}_t)^T\right) + E_{of}(\mathbf{u}, \mathbf{v}) + \lambda E_{sp}(\mathbf{\Delta}) + \beta \left\|\mathbf{Y}_r - \mathbf{Y} \circ \mathbf{s}\right\|_F^2 \\ & + \frac{\mu}{2} \left\|\tilde{\mathbf{Y}}_t - \tilde{\mathbf{Y}}_r - \tilde{\mathbf{\Delta}}\right\|_F^2 + \mathrm{Tr}\left(\tilde{\mathbf{R}}^T \left(\tilde{\mathbf{Y}}_t - \tilde{\mathbf{Y}}_r - \tilde{\mathbf{\Delta}}\right)\right) \end{aligned} \tag{12}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix, $\lambda, \beta, \mu > 0$ are penalty parameters, $\tilde{\mathbf{Y}}_t, \tilde{\mathbf{Y}}_r, \tilde{\mathbf{\Delta}}$ are the patch group matrices corresponding to $\mathbf{Y}_t, \mathbf{Y}_r, \mathbf{\Delta}$ respectively, and these matrices can be converted into each other using $\mathcal{P}$ and $\mathcal{P}^{-1}$. We utilize the alternating direction method of multipliers (ADMM) for the minimization model (12). Given the current points $\left(\mathbf{Y}^k_t, \mathbf{Y}^k_r, \mathbf{\Delta}^k, \mathbf{s}^k\right)$ at the kth iteration, we have the following five subproblems.

1) $\mathbf{Y}_t$-subproblem. Minimizing the function Θ w.r.t. $\mathbf{Y}_t$ is given by

$$\min_{\mathbf{Y}_t} 2\mathrm{Tr}\left(\tilde{\mathbf{Y}}_t \mathbf{L}^x (\tilde{\mathbf{Y}}_t)^T\right) + \frac{\mu}{2} \left\|\tilde{\mathbf{Y}}_t - \tilde{\mathbf{Y}}^k_r - \tilde{\mathbf{\Delta}}^k + \tilde{\mathbf{R}}^k / \mu\right\|_F^2 \tag{13}$$

which has the closed-form solution of

$$\tilde{\mathbf{Y}}_t^{k+1} = \left(\mu\tilde{\mathbf{Y}}_r^k + \mu\tilde{\boldsymbol{\Delta}}^k - \tilde{\mathbf{R}}^k\right) \times \left(4\mathbf{L}^x + \mu\mathbf{I}_{N_p}\right)^{-1} \tag{14}$$

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

2) $\mathbf{Y}_r$-subproblem. Minimizing the function Θ w.r.t. $\mathbf{Y}_r$ is given by

$$\min_{\mathbf{Y}_r} \beta\left\|\mathbf{Y}_r - \mathbf{Y}\circ\mathbf{s}^k\right\|_F^2 + \frac{\mu}{2}\left\|\tilde{\mathbf{Y}}_r - \tilde{\mathbf{Y}}_t^{k+1} + \tilde{\boldsymbol{\Delta}}^k - \tilde{\mathbf{R}}^k/\mu\right\|_F^2 \tag{15}$$

which has the closed-form solution of

$$\tilde{\mathbf{Y}}_r^{k+1} = \frac{2\beta\mathcal{P}\left(\mathbf{Y}\circ\mathbf{s}^k\right) + \mu\tilde{\mathbf{Y}}_t^{k+1} - \mu\tilde{\boldsymbol{\Delta}}^k + \tilde{\mathbf{R}}^k}{2\beta + \mu} \tag{16}$$

that is, for each pixel, we have

$$y_r^{k+1}(m,n,c) = \frac{2\beta y\left(m + u_{m,n}^k, n + v_{m,n}^k, c\right)}{2\beta + \mu} + \frac{\mu y_t^{k+1}(m,n,c) - \mu\Delta^k(m,n,c) + R^k(m,n,c)}{2\beta + \mu} \tag{17}$$

3) $\mathbf{s}$-subproblem. Based on the Lucas-Kanade optical flow method, $\mathbf{s}^{k+1} = \left(\mathbf{u}^{k+1}, \mathbf{v}^{k+1}\right)$ can be updated by solving (9) with $\mathbf{Y}_r^{k+1}$.

4) $\boldsymbol{\Delta}$-subproblem. Minimizing the function Θ w.r.t. $\boldsymbol{\Delta}$ is given by

$$\min_{\boldsymbol{\Delta}} \lambda\left(\sum_{c=1}^{C_y}(\Delta(m,n,c))^2\right)^{1/2} + \frac{\mu}{2}\sum_{c=1}^{C_y}\left(\Delta(m,n,c) - Q^k(m,n,c)\right)^2 \tag{18}$$

where $\mathbf{Q}^k = \mathbf{Y}_t^{k+1} - \mathbf{Y}_r^{k+1} + \frac{\mathbf{R}^k}{\mu}$. The closed-form solution is

$$\boldsymbol{\Delta}^{k+1}(m,n,:) = \max\left\{\left\|\mathbf{Q}^k(m,n,:)\right\|_2 - \frac{\lambda}{\mu}, 0\right\}\frac{\mathbf{Q}^k(m,n,:)}{\left\|\mathbf{Q}^k(m,n,:)\right\|_2} \tag{19}$$

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

5) $\tilde{\mathbf{R}}$-subproblem. Finally, the Lagrangian multiplier is updated by

$$\tilde{\mathbf{R}}^{k+1} = \tilde{\mathbf{R}}^k + \mu\left(\tilde{\mathbf{Y}}_t^{k+1} - \tilde{\mathbf{Y}}_r^{k+1} - \tilde{\boldsymbol{\Delta}}^{k+1}\right) \tag{20}$$

It can be observed that the minimization model (12) is convex, and each subproblem admits a closed-form solution, which theoretically guarantees convergence. However, since the LK optical flow is employed as an approximation in (1), its convergence is only conditional. Theoretically, the LK method exhibits local convergence under certain conditions: (i) the initial displacement estimate is sufficiently close to the true motion, satisfying the small-displacement assumption of the linearized brightness constancy constraint; and (ii) the local image gradient matrix is non-singular, ensuring stable solution of the least-squares problem. For regions with weak texture or large displacements, convergence cannot be guaranteed, and the solution may become unstable. To mitigate this, GiToF employs the coarse-to-fine multi-scale strategy with L levels, offering four key benefits.

First, it enables the LK optical flow method to handle large displacements effectively; second, it leverages the characteristic of change detection being relatively robust to misregistration at the coarse scale, making initialization easier and more accurate; third, it allows for a gradual refinement of the registration, image transformation, and change detection results; and fourth, it provides multi-scale changed images that can be fused to yield more accurate change detection outcomes. Specifically, GiToF first creates a pyramidal structure of down sampled images, then solve the minimization of (12) to roughly estimate the displacements, translated image, changed image at a coarse level, then incrementally refines them over the scales, as illustrated in Fig. 3.

3.5. Change extraction

Once the multi-sale changed images, $\left\{\boldsymbol{\Delta}^i\right\}_{i=1}^L$, are computed from the minimization model (12), we can obtain multi-sale DI as

$$DI_{m,n}^i = \left(\sum_{c=1}^{C_y}(\Delta^i(m,n,c))^2\right)^{1/2}, \ i = 1, \ldots, L \tag{21}$$

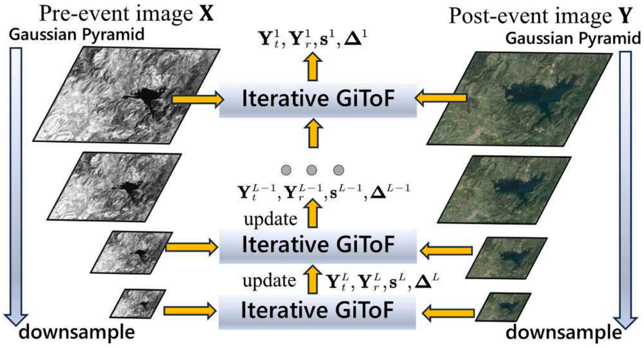


Fig. 3. The multi-scale strategy of GiToF.

The simplest approach is to first fuse and then segment these DIs to obtain the CM, e.g., linear fusion coupled with threshold-based segmentation, which may lead to insufficient utilization of multi-scale information. Instead of fusion followed by segmentation, we used an MRF-based fusion segmentation method.

To reduce salt-and-pepper noise, we choose the superpixel as the basic unit in the CM by co-segmenting the pre-event image $\mathbf{X}$ and registered post-event image $\mathbf{Y}_r$ similar as the preprocessing in [24], resulting N_{cs} superpixels with co-segmentation map Λ. The binary CM computation problem can be regarded as superpixel label assignments, that is, for each superpixel Λ_j, the label $B_j = 0$ indicates unchanged and $B_j = 1$ indicates changed. According to the MRF theory, the label assignments can be modeled as the energy minimization problem

$$\mathbf{B}^* = \underset{\mathbf{B}\in\{0,1\}^{N_{cs}}}{\arg\min} \ \eta\mathcal{J}_c(\mathbf{B}) + (1-\eta)\mathcal{J}_s(\mathbf{B}) \tag{22}$$

where $\mathcal{J}_c$ and $\mathcal{J}_s$ represent the change energy and spatial energy functions respectively, and $0 < \eta < 1$ is a weighting parameter.

To construct the function $\mathcal{J}_c$, we first up-sample the multi-scale DI^i, $i = 1, \ldots, L$, to the size of $M \times N$, and compute the average change level of the j-superpixel at ith scale as

$$v_j^i = \text{mean}\left(DI_{m,n}^i\right), (m,n) \in \Lambda_j \tag{23}$$

and then normalize change level as $v_j^i = \min\left\{v_j^i/2T_i, 1\right\}$, where T_i is the threshold of Otsu method [47] on the v^i at the ith scale. Then, we define $\mathcal{J}_c$ similar as in [5]

$$\mathcal{J}_c(\mathbf{B}) = \sum_{j=1}^{N_{cs}}\sum_{i=1}^{L}\varphi_c\left(B_j, v_j^i\right) \tag{24}$$

where φ_c is defined as

$$\varphi_c(b,a) = \begin{cases} -\left(\frac{1-a}{a}\right)^{\varepsilon}\log(a), \text{if } b = 1 \\ -\log(1-a), \ \text{if } b = 0 \end{cases} \tag{25}$$

with $\varepsilon > 0$ being a parameter that controls the rate of imbalance penalty. As analyzed in [5], this type of $\mathcal{J}_c$ utilizes the penalty function φ_c similar to the focal loss function, taking into account the imbalance between changed and unchanged classes in the HeCD problem.

For the function $\mathcal{J}_s$, we directly utilize the form in SCASC [19] to push the spatially adjacent superpixels be assigned the same labels, which can incorporate the contextual information of each superpixel. With these energy terms, the model (22) can be efficiently implemented by graph cut algorithm, and the final CM can be obtained.

4. Experiments and discussions

In this section, we evaluate the performance of the proposed GiToF,[2] comparing it with twelve SOTA HeCD methods: HPT [20], AMDIR [18],

---

[2] The codes will be released at https://github.com/yulisun/GiToF

Table 1
Imperfectly registered heterogeneous datasets.

Dataset	Sensor	Image size (pixels)	Date	Location	Spatial resolution	Change event	RMSE (pixels)
#1	Landsat-5 Google Earth	$300 \times 412 \times 1$ $300 \times 412 \times 3$	Sept. 1995 July 1996	Sardinia, Italy	30m	Lake expansion	10.23
#2	Pleiades WorldView2	$2000 \times 2000 \times 3$ $2000 \times 2000 \times 3$	May 2012 July 2013	Toulouse, France	0.52m	Construction	24.44
#3	Radarsat-2 Google Earth	$593 \times 921 \times 1$ $593 \times 921 \times 3$	June 2008 Sept. 2012	Shuguang Village, China	8m	Construction	21.75
#4	TerraSAR-X Pleiades	$4404 \times 2604 \times 1$ $4404 \times 2604 \times 1$	Feb. 2009 July 2013	Toulouse, France	2m	Construction	37.01
#5	Google Earth Gaofen-3	$1500 \times 2000 \times 3$ $1500 \times 2000 \times 1$	Mar. 2018 Dec. 2019	Beijing City, China	1m	Land changes	14.46

FPMS [48], CICM [2], SCASC [19], USSD [12], AOSG [13], GSGM [46], GCLN [7], CFRL [8], CGSL [9], and ITSA [3]. For these comparison methods, we employed their publicly available codes with the recommended parameter settings, and generated the change detection results under the same computational environment.

4.1. Datasets and evaluation criteria

We conduct experiments on five datasets, as illustrated in Fig. 4. Datasets #1-#4 are obtained by adding affine transformations to the original well-registered heterogenous image pairs [48],[3] and the root mean square error (RMSE) of misregistration are 10.23, 24.44, 21.75 and 37.01 respectively. Dataset #5 is obtained by performing image matching algorithm of RIFT [27] on the original georeferenced coarsely-registered images, and the RMSE of residual misregistration is 14.46. The registration errors in Datasets #1-#4 primarily reflect common global transformation errors, which are typically caused by inaccurate global transformation parameters estimated by automatic registration algorithms. In contrast, Dataset #5 contains local nonlinear deformation errors, which are more complex and frequently encountered in real-world applications. Table 1 lists the details of these datasets.

We evaluate the performance of different methods using two types of metrics. First, the receiver operating characteristic (ROC) curve and precision-recall (PR) curve, along with the area under ROC curve (AUR) and the area under PR curve (AUP), are employed to assess the DI. Second, to assess the quality of CM, we employ various colors to label true positive (TP), false positive (FP), true negative (TN), and false negative (FN) on the CM as a visual display, and compute the overall accuracy (OA), Kappa coefficient (κ) and F1-score as quantitative assessments.

For all the experimental results of GiToF, we set the multiscale parameter as $L = \log_2 \left\lfloor \frac{\sqrt{MN}}{100} \right\rfloor$, set the patch size of each sale as $p^i = \max\left\{5, \left\lceil \frac{\sqrt{MN}}{100\times 2^{i-1}} \right\rceil\right\}$, with $i = 1, \ldots, L$, which ensures the number of square patches at each scale is less than 10^4. In addition, we set the maximum number of iterations as $N_{iter} = 10$, the weighting parameters as $\lambda = 0.1$, $\beta = 0.5$ for the minimization model (12) and $\eta = 0.05$ for the MRF model (22). These parameters will be discussed in Section 4.3.3.

4.2. Quantitative and qualitative results

4.2.1. Difference images

Fig. 5 displays the DIs generated by different methods across five datasets, except for the comparison methods that directly output CM without generating DI. It can be seen that due to the presence of co-registration errors, the DIs obtained by these comparison methods are affected by the confusion between changed and unchanged areas. As analyzed in Section 2, co-registration errors increase the change intensity of DI in unchanged regions, resulting in changes being submerged within unchanged areas, as shown in Fig. 5(a)-(g). In addition, different comparison methods exhibit varying sensitivity to co-registration errors. For example, HPT, AMDIR, USSD, and GSGM are relatively more sensitive to the misregistration, especially on Datasets #2, #4, and #5. On one hand, these datasets have higher resolutions and relatively larger RMSE of co-registration errors; on the other hand, these methods are based on pixel value regression (HPT and AMDIR) and pixel feature comparison (USSD and GSGM), meaning they treat pixels as the basic processing unit. In contrast, FPMS and SCASC are relatively robust to co-registration errors and can achieve higher quality DI, as shown in Fig. 5(c) and (d). This is because FPMS uses a downsampling operation in its preprocessing step, though this comes at the cost of granularity in detecting small changes. SCASC treats superpixel as the basic processing unit, and such object-based change detection methods are inherently more robust to co-registration errors than pixel-based methods. However, compared with these benchmarking methods, the proposed GiToF can overcome the influence of misregistration on change detection, significantly improving the quality of DI. As shown in Fig. 5(h), the DI generated by GiToF is more discriminative between changes and unchanges, effectively suppressing the change confusion caused by co-registration errors. Additionally, it also exhibits a certain level of sparsity, which can be attributed to the sparse regularization term $E_{sp}(\mathbf{\Delta})$ used in the joint minimization model.

To better assess DIs quality, Fig. 6 shows their ROC and PR curves, and Table 2 reports the corresponding quantitative indicator scores, i.e., the AUR and AUP of area under these curves. Across these five datasets, GiToF achieves significantly higher AUP values than other comparison methods. For example, the average AUR and AUP values obtained by GiToF are 0.875 and 0.566, respectively, which are 4.0% and 13.6% higher than those of the second-ranked FPMS, and 4.6% and 22.5% higher than those of the third-ranked SCASC.

4.2.2. Change maps

To more fully assess the ability of GiToF in detecting changes under co-registration errors, we compare the CMs generated by different methods as shown in Fig. 7. In the Dataset #1, all methods, except AMDIR and USSD, show some ability to detect changes. This is mainly attributed to two factors: first, the relatively small co-registration error of the heterologous images in Dataset #1, i.e., an RMSE of 10.23 pixels; and second, the low resolution of the Dataset #1 images (30m), the simplicity of the scenes (containing fewer land cover classes), and the significant single changes (lake expansion). These factors collectively reduce the impact of co-registration errors on change detection. However, the performance of these methods degrades when the co-registration error increases. For example, in a similar Dataset #3 that consists of low-resolution images (8m) with fewer land cover classes, but with larger co-registration errors (RMSE of 21.75 pixels), many false detection are generated, as seen with methods such as HPT, FPMS, AOSG, GSGM, and GCLN. In Datasets

---

[3] The original well-registered images of Datasets #1-#4 are available at http://www-labs.iro.umontreal.ca/~mignotte

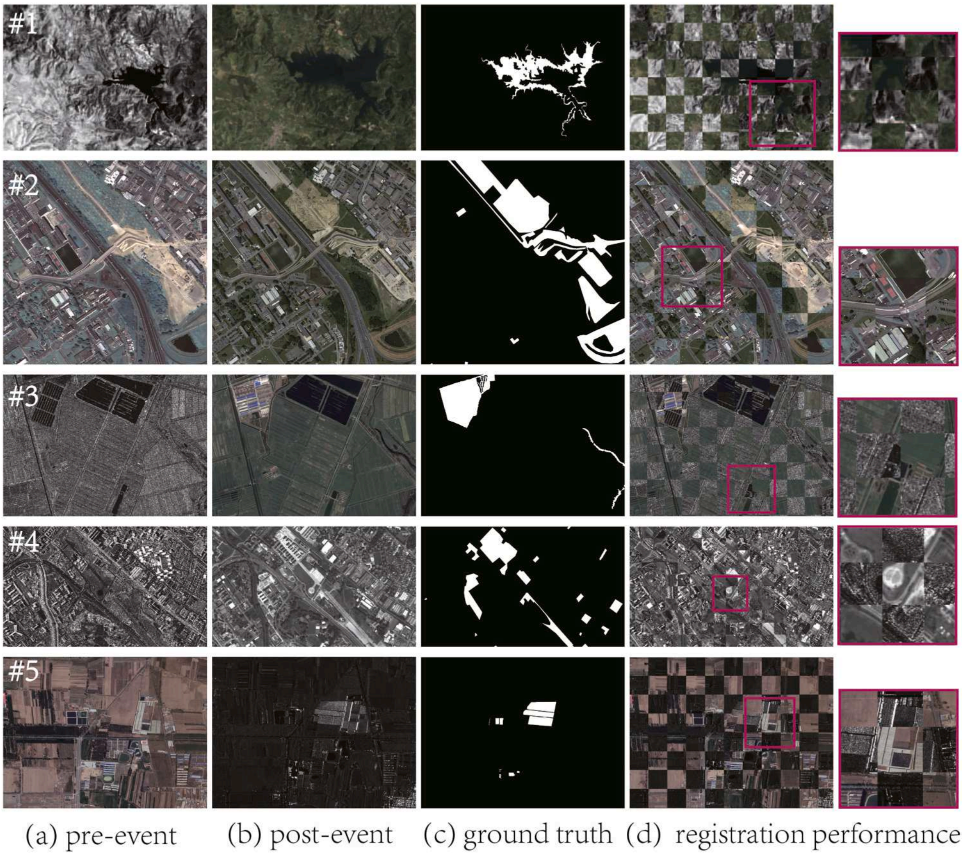


Fig. 4. Dataset: from top to bottom, they are corresponding to Datasets #1 to #5 respectively. (a) pre-event image; (b) post-event image; (c) ground truth; (d) presentation of misregistration.

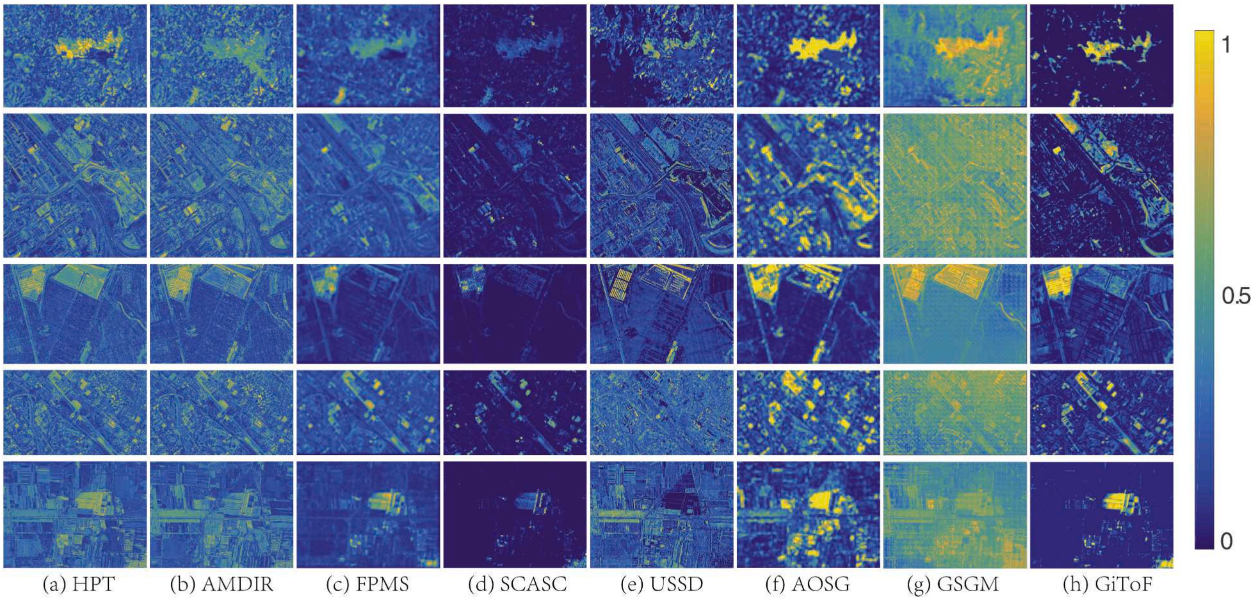
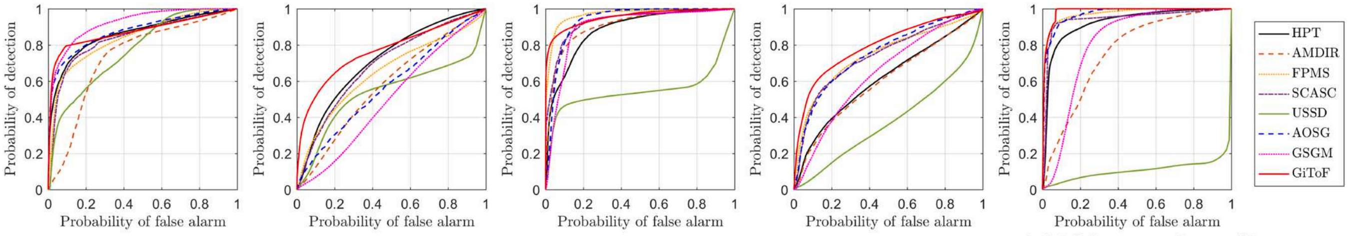


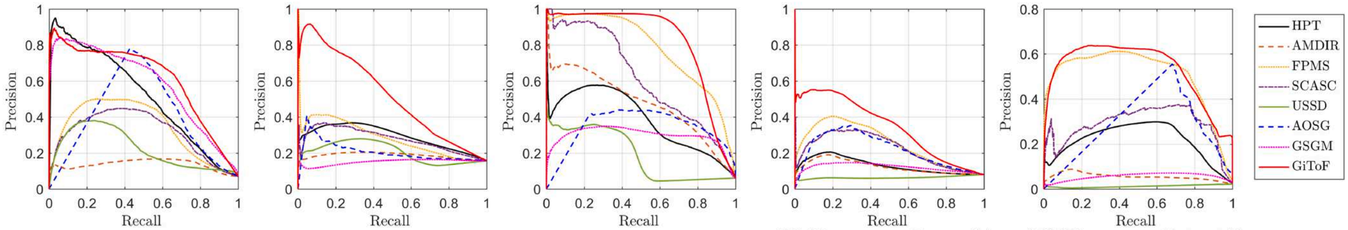
Fig. 5. DIs obtained by different methods on Datasets #1-#5 (from top row to bottom row). From left to right are the DIs generated by (a) HPT, (b) AMDIR, (c) FPMS, (d) SCASC, (e) USSD, (f) AOSG, (g) GSGM, and (h) the proposed GiToF. In the DIs, the brighter yellow indicates a higher change probability, and darker blue indicates a higher unchange probability. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
AUR and AUP on different datasets. The best and second best scores are marked in underline and **bold**, respectively.

Methods	Dataset #1		Dataset #2		Dataset #3		Dataset #4		Dataset #5		Average	
	AUR	AUP	AUR	AUP	AUR	AUP	AUR	AUP	AUR	AUP	AUR	AUP
HPT [20]	0.861	0.514	**0.714**	**0.291**	0.885	0.384	0.613	0.129	0.924	0.213	0.800	0.306
AMDIR [18]	0.733	0.136	0.581	0.186	0.909	0.474	0.602	0.122	0.763	0.055	0.717	0.194
FPMS [48]	0.840	0.337	0.649	0.282	0.968	**0.781**	0.746	**0.248**	**0.972**	**0.503**	**0.835**	**0.430**
SCASC [19]	0.838	0.310	0.692	0.276	0.929	0.614	0.742	0.221	0.943	0.285	0.829	0.341
USSD [12]	0.772	0.219	0.540	0.203	0.551	0.190	0.380	0.063	0.110	0.012	0.470	0.137
AOSG [13]	0.871	0.393	0.575	0.204	0.935	0.339	**0.754**	0.215	0.970	0.278	0.821	0.286
GSGM [46]	0.921	**0.585**	0.502	0.150	0.900	0.294	0.647	0.120	0.816	0.055	0.757	0.241
proposed GiToF	**0.874**	0.615	0.768	0.520	**0.954**	0.821	0.796	0.346	0.985	0.529	0.875	0.566



(a1) ROC curves on Dataset #1 (b1) ROC curves on Dataset #2 (c1) ROC curves on Dataset #3 (d1) ROC curves on Dataset #4 (e1) ROC curves on Dataset #5



(a2) PR curves on Dataset #1 (b2) PR curves on Dataset #2 (c2) PR curves on Dataset #3 (d2) PR curves on Dataset #4 (e2) PR curves on Dataset #5

Fig. 6. ROC curves (top row) and PR curves (bottom row) of DIs on different datasets.

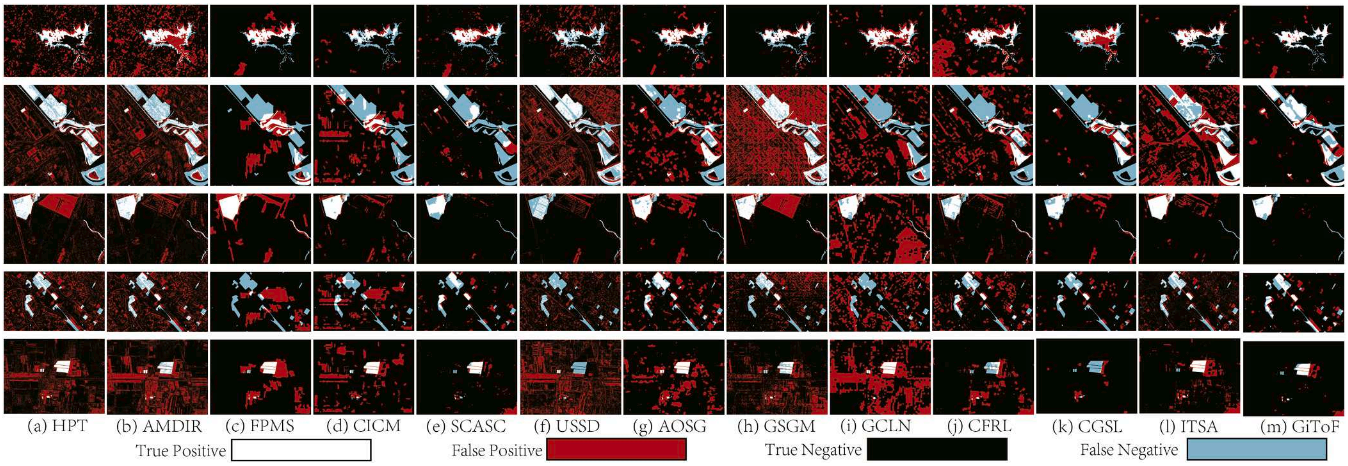


Fig. 7. CMs generated by different methods on Datasets #1-#5 (from top row to bottom row). From left to right are the CMs generated by (a) HPT, (b) AMDIR, (c) FPMS, (d) CICM, (e) SCASC, (f) USSD, (g) AOSG, (h) GSGM, (i) GCLN, (j) CFRL, (k) CGSL, (l) ITSA, and (m) the proposed GiToF. In the CM, white, red, black, and cyan mark TP, FP, TN, and FN, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

#2, #4, and #5, where the resolution of the pre- and post-event images is higher and the image scenes are more complex, the detection performance of most comparison methods is poorer, which is caused by three factors: first, the higher resolution increases the heterogeneity of the images, making HeCD more challenging (even for well-registered images); second, the co-registration errors on these datasets are also relatively large (e.g., the RMSE of the Datasets #2 and #4 are 24.44 and 37.01 pixels, respectively); and third, the images contain more objects and are more fragmented, as analyzed in Section 2, and the impact of the co-registration error on HeCD is more severe.

Among these comparison methods, SCASC performs relatively better as shown in Fig. 7(e). This is because SCASC is an image regression algorithm based on structural consistency, which is more robust to image noise and heterogeneity. Additionally, SCASC uses superpixels as the basic unit, making it less sensitive to co-registration errors. The use of operations such as downsampling (CICM, FPMS), superpixel

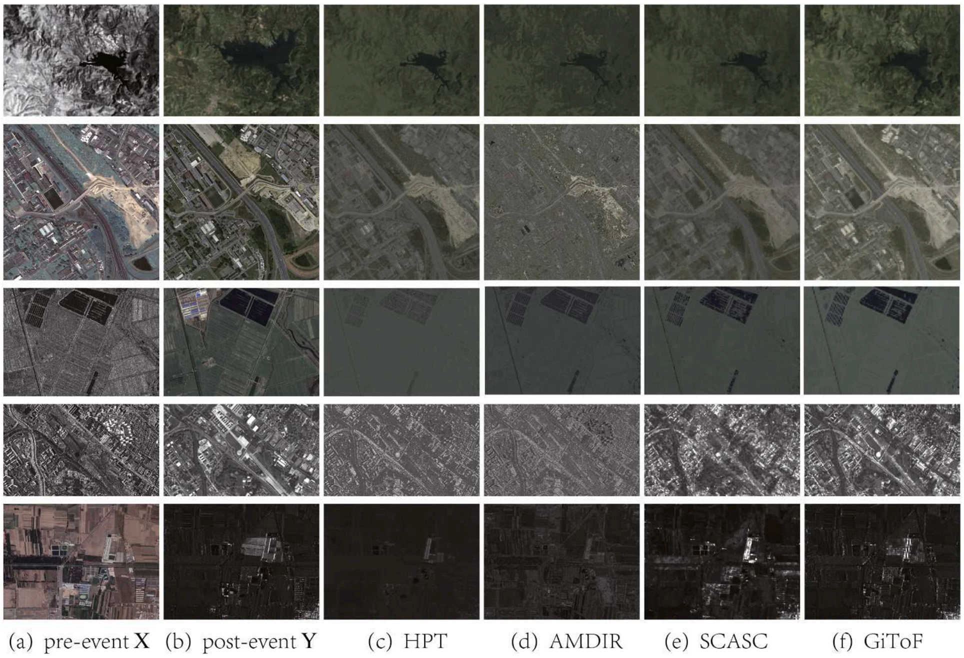


Fig. 8. The transformed images on Datasets #1-#5 (from top row to bottom row). From left to right are the (a) original pre-event image, (b) original post-event image, (c) transformed image of HPT, (d) transformed image of AMDIR, (e) transformed image of SCASC, (f) transformed image of GiToF.

segmentation (SCASC), and spatial smoothing (SCASC, AOSG) can mitigate the influence of small misregistration on HeCD to some extent, but this is far from satisfactory when faced with large co-registration errors. On the contrary, by observing the CM obtained from the proposed GiToF in Fig. 7(m), it can be found that GiToF exhibits significantly fewer missed detections and false alarms across datasets with different resolutions and scenes. Furthermore, GiToF achieves satisfactory results even under large co-registration errors, which demonstrates the effectiveness of GiToF against the misregistration on HeCD.

Table 3 reports corresponding quantitative indicator scores of the CMs, i.e., the OA, κ, F1. Across these five datasets, GiToF achieves the highest scores than other comparison methods. For example, the average κ and F1 scores obtained by GiToF are 0.580 and 0.613 respectively, surpassing the second-ranked SCASC by 16.6% and 14.7%, respectively. This outstanding performance can be attributed to the following three aspects: first, the GiToF integrates image transformation, image registration, and change estimation into a unified HeCD model that promotes collaboration among the sub-tasks; second, the multi-scale coarse-to-fine strategy allows the Lucas-Kanade optical flow method to effectively handle large co-registration errors; third, the MRF-based fusion segmentation strategy can integrate multi-scale changed images to improve the CM performance.

4.3. Discussion

4.3.1. Image transformation

Fig. 8 shows the transformed images of the image transformation based HeCD methods of HPT, AMDIR, SCASC and the proposed GiToF, where HPT is a supervised method that uses 40% of the unchanged pixel pairs as training set. By comparing these transformed images with the original images, one can find that: first, in Datasets #1 and #3, these methods roughly accomplish the image style migration, i.e., transforming the pre-event image to the domain of post-event image. However, in other datasets with higher image resolution and larger co-registration errors, their transformation is greatly discounted, such as the AMDIR in Dataset#2, HPT, AMDIR and SCASC in Datasets #4 and #5. Second, the proposed GiToF is not only able to complete the image transformation in complex scenarios of Datasets #2, #4 and #5, but its transformed image is also closer to the original images, that is, the style of transformed image is more consistent with the target post-event image, and the structure of transformed image remains consistent with the original pre-event image, as shown in the Fig. 8(f). In addition, although at first glance, the transformed images of SCASC and GiToF appear to display similar patterns, there are significant performance differences between them: first, GiToF exhibits more accurate transformations, which can be visually observed, especially in Datasets #1, #2 and #4; second, the transformed images of SCASC are still imperfectly registered, whereas GiToF not only performs image transformation but also completes image registration. This is why there is such a large performance gap in their respective DIs, as illustrated by Figs. 5, 6 and Table 2.

4.3.2. Image registration

GiToF is a joint framework that combines graph-based image transformation with the Lucas-Kanade optical flow method for the HeCD. Fig. 9 shows the visualizing results of the flow fields generated by different multimodal image registration methods on Datasets #1 and #5, including SRIF [34], RIFT [27], LNIFT [35], GeFolki [36], OSflow [37] and the proposed GiToF, where GeFolki and OSflow are representative optical flow based methods. As can be seen Fig. 9, although SRIF, RIFT and LNIFT yield smooth optical flow results by finding matching points and then calculating the global affine matrix for image registration, there is a significant disparity between them and the ground truth, especially in the Dataset #5. Among the optical flow based methods, the proposed GiToF gives a smoother and more accurate optical flow field than GeFolki and OSflow, which is attributed to the fact that GiToF converts heterogeneous image matching to homogeneous image matching while taking into account the influence of changed regions. The RMSE of co-registration errors of these methods on these two datasets are: SRIF: 4.55, 1.82; RIFT: 2.80, 8.76; LNIFT: 3.91, 222.35; GeFolki: 3.03, 36.94; OSflow: 3.36, 3.77; and GiToF: 2.40, 2.79, respectively.

4.3.3. Parameter analysis

The main parameters used in GiToF are the multiscale parameter L, the maximum number of iterations N_{iter}, and the weighting parameters of λ, β and η.

Table 3
OA, F1 and κ of CMs. The best and second best scores are marked in underline and **bold**, 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 [20]	0.835	0.320	0.390	0.724	**0.231**	**0.388**	0.813	0.271	0.340	0.717	0.090	0.203	0.773	0.113	0.149	0.772	0.205	0.294
AMDIR [18]	0.707	0.163	0.258	0.652	0.080	0.274	0.874	0.382	0.436	0.717	0.083	0.197	0.689	0.053	0.092	0.728	0.152	0.251
FPMS [48]	0.928	0.524	0.562	0.802	0.181	0.295	0.880	0.436	0.487	0.804	0.009	0.045	0.894	0.253	0.281	0.861	0.281	0.334
CICM [2]	0.939	0.444	0.474	0.755	0.059	0.204	0.940	0.593	0.623	0.795	0.015	0.119	0.866	0.199	0.229	0.859	0.262	0.330
SCASC [19]	0.922	0.481	0.522	**0.833**	0.174	0.252	**0.964**	**0.628**	0.646	**0.902**	**0.328**	**0.377**	**0.969**	**0.517**	**0.531**	**0.918**	**0.426**	**0.466**
USSD [12]	0.851	0.234	0.306	0.623	0.112	0.310	0.918	0.302	0.346	0.773	0.002	0.079	0.644	0.004	0.011	0.762	0.131	0.210
AOSG [13]	0.928	0.523	0.561	0.728	0.091	0.254	0.905	0.451	0.496	0.846	0.261	0.339	0.853	0.187	0.219	0.852	0.303	0.374
GSGM [46]	0.949	0.575	0.602	0.408	0.016	0.263	0.872	0.375	0.430	0.781	0.102	0.204	0.797	0.092	0.128	0.761	0.232	0.326
GCLN [7]	0.899	0.474	0.523	0.679	0.013	0.099	0.703	0.154	0.240	0.697	0.012	0.093	0.650	0.068	0.106	0.725	0.144	0.212
CFRL [8]	0.796	0.262	0.342	0.729	0.078	0.240	0.917	0.440	0.482	0.823	0.168	0.257	0.897	0.113	0.145	0.832	0.212	0.293
CGSL [9]	0.927	0.499	0.538	0.845	0.270	0.348	0.927	0.477	0.515	0.895	0.197	0.252	0.961	0.021	0.041	0.911	0.293	0.339
ITSA [3]	**0.952**	**0.632**	**0.658**	0.647	0.156	0.343	0.944	0.620	**0.649**	0.835	0.184	0.269	0.924	0.319	0.343	0.860	0.382	0.452
proposed GiToF	0.956	0.650	0.674	0.886	0.525	0.590	0.978	0.785	0.797	0.905	0.398	0.451	0.973	0.540	0.553	0.940	0.580	0.613

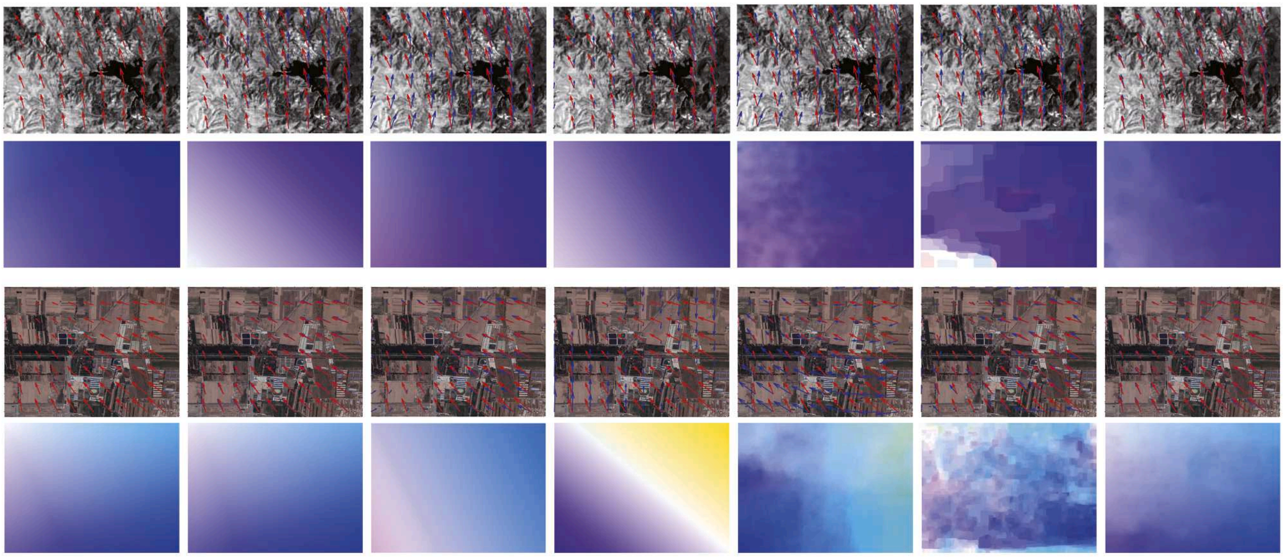

(a) Ground truth (b) SRIF (c) RIFT (d) LNIFT (e) GeFolki (f) OSflow (g) GiToF

Fig. 9. The visualizing results of the flow fields generated by different multimodal image registration methods on Datasets #1 (the first and second rows) and #5 (the third and fourth rows), (a) ground truth, (b) SRIF, (c) RIFT, (d) LNIFT, (e) GeFolki, (f) OSflow, (g) the proposed GiToF. The first and third rows plot arrows with directional components **u** and **v** produced by different methods, where the underline and **bold** arrows indicate the ground truth and those obtained by the comparison methods, respectively. The second and fourth rows show the color-coded optical flow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

It is clear that when a large displacement exists, a larger L should be chosen. For large size remote sensing images, it is not convenient to estimate the initial displacement. In practice, we can select the number of pyramid layers according to image size. In GiToF, we set the $L = \log_2\left\lfloor \frac{\sqrt{MN}}{100} \right\rfloor$, e.g., for 2000×2000 image, we have $L = 4$, resulting in multi-scale images from size 250×250 to 2000×2000. Certainly, when we know that the displacement of the multi-temporal images is very small, it is appropriate to choose a smaller L.

For the maximum number of iterations N_{iter} for solving the minimization model (12), we plot the ADMM convergence curve, which is measured by the relative difference (Rel.diff) $\xi^{k+1} = \frac{\left\|\Delta^{k+1} - \Delta^{k}\right\|_2}{\left\|\Delta^{k+1}\right\|_2}$ versus the iteration number k. As shown in Fig. 10, the ADMM algorithm converges rapidly, and setting the iteration exit criterion as $N_{iter} = 10$ or $\xi_{stop} = 10^{-2}$ proves to be appropriate for the algorithm when balancing change detection accuracy and computational efficiency.

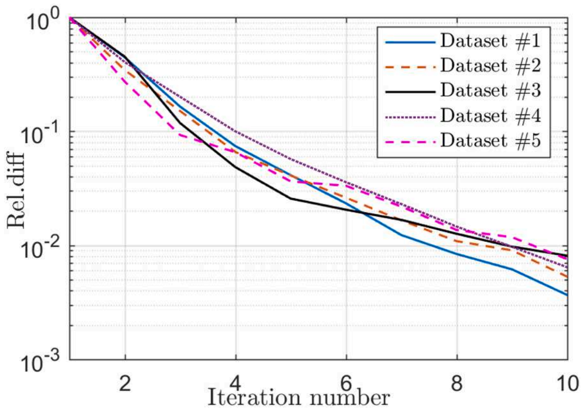


Fig. 10. Convergence performance of ADMM.

For the weighting parameters of λ and β, they are applied to adjust the weights of the prior sparsity based E_{sp} and the image registration term in the minimization function (12), respectively. First, we conduct the ablation analysis. 1) If the GiToF model focuses only on the prior sparsity term, i.e., setting $\lambda \to \infty$, the solution of (12) will be $\boldsymbol{\Delta} = 0$. 2) If the GiToF model ignores the prior sparsity term, i.e., setting $\lambda = 0$,

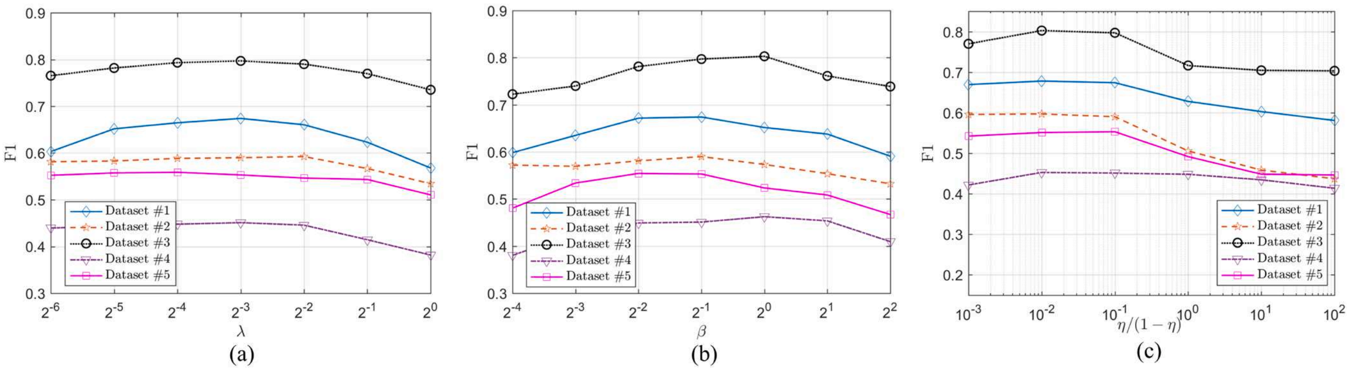


Fig. 11. Sensitivity analysis of parameters (a) λ, (b) β, and (c) η.

then $\mathbf{s} = 0$, $\mathbf{Y}_t = 0$, $\boldsymbol{\Delta} = -\mathbf{Y}$ will be the solution of (12). We show the parameters sensitivity of GiToF on λ and β in Fig. 11, where we plot the $F1$-score of GiToF on all the five datasets with different λ (from 2^{-6} to 2^{0} with fixed $\beta = 0.5$) and β (from 2^{-4} to 2^{2} with fixed $\lambda = 0.1$). From Fig. 11, it can be found that GiToF gives favorable performance to a certain range of λ and β, which demonstrates a certain level of robustness to the weighting parameters. According to Fig. 11, in GiToF we set $\lambda = 0.1$ and $\beta = 0.5$ for simplicity.

In the MRF model (22), η balances the change energy and spatial energy terms. Fig. 11(c) shows the F1-score versus $\eta/(1-\eta)$ from 10^{-3} to 10^{2}. Smaller η emphasizes spatial smoothness, while larger η favors change energy, resembling thresholding segmentation. It can be found that the change detection performance is better when $\eta/(1-\eta)$ is between 10^{-2} and 10^{-1}, so we set $\eta = 0.05$ in this work.

4.3.4. Complexity analysis

The main computational complexity of the GiToF is concentrating on the minimization problem of (12). Taking the finest scale as an example, in each iteration of model (12) minimization process, solving the $\mathbf{Y}_t$-subproblem needs $\mathcal{O}\left(N_p^3\right)$ for the matrix inversion and $\mathcal{O}\left(C_y p^2 N_p^2\right)$ for the matrix multiplication, respectively. However, since the graph Laplacian matrix $\mathbf{L}^x$ is sparse, real, symmetric, and semi-positive definite, then the $\mathbf{Y}_t$ can be efficiently updated by iterative solvers, such as the preconditioned conjugate gradient method. Solving the $\mathbf{Y}_r$-subproblem needs $\mathcal{O}\left(C_y p^2 N_p^2\right)$; solving the $\mathbf{s}$-subproblem needs $\mathcal{O}\left(w^2 C_y M N\right)$; solving the $\boldsymbol{\Delta}$-subproblem needs $\mathcal{O}\left(C_y M N\right)$; and updating the $\tilde{\mathbf{R}}$ needs $\mathcal{O}\left(C_y p^2 N_p\right)$. The execution time (in seconds) of the comparison methods on Dataset #2 (with image size of $2000 \times 2000 \times 3$) are as follows: HPT: 80.8; AMDIR: 121.2; FPMS: 87.6; CICM: 115.7; SCASC: 24.0; USSD: 720.9; AOSG: 188.5; GSGM: 719.3; GCLN: 674.7; CFRL: 4669.2; CGSL: 1233.0; ITSA: 228.6; and GiToF: 112.1, where the hardware configuration includes an Intel Core i7-10870H CPU and an NVIDIA GeForce RTX 3090 GPU. It can be observed that GiToF achieves competitive efficiency while maintaining high detection accuracy. For images with large spatial sizes, image partitioning and parallel processing are recommended in GiToF to improve computational efficiency.

5. Conclusion

This paper focuses on the issue of unsupervised HeCD with imperfectly registered heterogeneous images. First, the influence of co-registration errors on HeCD is discussed, particularly for image translation-based methods. It indicates that registration, image transformation, and change detection are interrelated processes. In light of this, this paper proposes a joint model GiToF that combines graph-based image transformation with the Lucas-Kanade optical flow method. Specifically, GiToF assumes that the matched post-event image can be decomposed into an unchanged translated image and a changed image, and employs three regularization terms for this decomposition: it uses structural consistency to ensure the unchanged translated image retains the same structure as the original pre-event image in the graph; it applies the Lucas-Kanade optical flow constraint between the original and matched post-event images; and it enforces prior sparsity to constrain the sparsity of the changed image. With this joint model, GiToF helps the image transformation process overcome the influence of co-registration errors and changed regions, while also assisting the optical flow method in satisfying the brightness constancy assumption and the small and smooth motion assumption, both of which are typically violated in the HeCD problem due to image heterogeneity and drastic changes. By solving the joint model, the multi-scale changed images can be obtained, which in turn gives the final change detection result by using MRF-based fusion segmentation. The experimental results verify the effectiveness of the proposed method. Since there is currently no research on HeCD under misregistration conditions, we believe that the approach proposed in this paper is expected to fill this gap and promote the practice application of HeCD. While the current method employs square patches as graph vertices, this representation leads to computational complexity challenges in large-scale high-resolution images. In future work, we will explore parallel processing strategy to accelerate the algorithm and make it more scalable for large datasets. Additionally, we will extend GiToF to better cope with the more complex geometric discrepancies encountered in real-world applications, such as large-angle rotations and viewpoint variations between heterogeneous remote sensing images acquired by UAVs and satellites.

CRediT authorship contribution statement

Yuli Sun: Writing – original draft, Methodology, Investigation, Funding acquisition, Conceptualization; **Lin Lei:** Writing – review & editing, Visualization, Validation, Supervision; **Gangyao Kuang:** Writing – review & editing, Project administration, Formal analysis.

Data availability

The data supporting this study are publicly available from open sources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Natural Science Foundation of China [grant number 62401577]; the Natural Science Foundation of

Hunan Province of China [grant number 2024JJ6466]; the China Postdoctoral Science Foundation [grant number 2024M754301]; and the Postdoctoral Fellowship Program of China Postdoctoral Science Foundation [grant number GZC20233545].

References

[1] J. Chen, D. Hou, C. He, Y. Liu, Y. Guo, B. Yang, Change detection with cross-domain remote sensing images: a systematic review, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 17 (2024) 11563–11582.

[2] R. Touati, Détection de Changement En Imagerie Satellitaire Multimodale, Ph.D. thesis, Université de Montréal, 2019.

[3] Z. Lv, H. Huang, W. Sun, M. Jia, J.A. Benediktsson, F. Chen, Iterative training sample augmentation for enhancing land cover change detection performance with deep learning neural network, IEEE Trans. Neural Netw. Learn. Syst. 35 (11) (2024) 15128–15141.

[4] B. Zhao, H. Sui, J. Liu, W. Shi, W. Wang, C. Xu, J. Wang, Flood inundation monitoring using multi-source satellite imagery: a knowledge transfer strategy for heterogeneous image change detection, Remote Sens. Environ. 314 (2024) 114373.

[5] Y. Sun, L. Lei, Z. Li, G. Kuang, Similarity and dissimilarity relationships based graphs for multimodal change detection, ISPRS J. Photogramm. Remote Sens. 208 (2024) 70–88.

[6] I.R. McGregor, G. Connette, J.M. Gray, A multi-source change detection algorithm supporting user customization and near real-time deforestation detections, Remote Sens. Environ. 308 (2024) 114195.

[7] Z. Lv, S. Cheng, L. Xie, J. Li, M. Zhao, A graph contrastive learning network for change detection with heterogeneous remote sensing images, Pattern Recognit. (2025) 112394.

[8] T. Liu, M. Zhang, M. Gong, Q. Zhang, F. Jiang, H. Zheng, D. Lu, Commonality feature representation learning for unsupervised multimodal change detection, IEEE Trans. Image Process. 34 (2025) 1219–1233.

[9] J. Xu, T. Liu, T. Lei, H. Chen, N. Yokoya, Z. Lv, M. Gong, CGSL: commonality graph structure learning for unsupervised multimodal change detection, ISPRS J. Photogramm. Remote Sens. 229 (2025) 92–106.

[10] Y. Sun, L. Lei, X. Li, X. Tan, G. Kuang, Structure consistency-based graph for unsupervised change detection with homogeneous and heterogeneous remote sensing images, IEEE Trans. Geosci. Remote Sens. (2022) 1–21.

[11] L. Wan, Y. Xiang, H. You, An object-based hierarchical compound classification method for change detection in heterogeneous optical and SARimages, IEEE Trans. Geosci. Remote Sens. 57 (12) (2019) 9941–9959.

[12] L. Zhu, W. Sun, D. Fan, H. Xing, X. Liu, Unsupervised spatial self-similarity difference-based change detection method for multi-source heterogeneous images, Pattern Recognit. 149 (2024) 110237.

[13] T. Han, Y. Tang, B. Zou, H. Feng, Unsupervised multimodal change detection based on adaptive optimization of structured graph, Int. J. Appl. Earth Obs. Geoinf. 126 (2024) 103630.

[14] W. Li, X. Wang, G. Li, B. Geng, P.K. Varshney, A copula-guided in-model interpretable neural network for change detection in heterogeneous remote sensing images, IEEE Trans. Geosci. Remote Sens. 63 (2025) 1–17.

[15] T. Liu, Y. Pu, T. Lei, J. Xu, M. Gong, L. He, A.K. Nandi, Hierarchical feature alignment-based progressive addition network for multimodal change detection, Pattern Recognit. 162 (2025) 111355.

[16] X. Hou, Y. Bai, Y. Xie, Y. Zhang, L. Fu, Y. Li, C. Shang, Q. Shen, Self-supervised multimodal change detection based on difference contrast learning for remote sensing imagery, Pattern Recognit. 159 (2025) 111148.

[17] B. Yang, Z. Wang, X. Liu, L. Fang, L. Liu, Reciprocal transformation-based joint deep and broad learning for change detection with heterogeneous images, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–14.

[18] L.T. Luppino, F.M. Bianchi, G. Moser, S.N. Anfinsen, Unsupervised image regression for heterogeneous change detection, IEEE Trans. Geosci. Remote Sens. 57 (12) (2019) 9960–9975.

[19] Y. Sun, L. Lei, D. Guan, M. Li, G. Kuang, Sparse-constrained adaptive structure consistency-based unsupervised image regression for heterogeneous remote sensing change detection, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–14.

[20] Z. Liu, G. Li, G. Mercier, Y. He, Q. Pan, Change detection in heterogenous remote sensing images via homogeneous pixel transformation, IEEE Trans. Image Process. 27 (4) (2018) 1822–1834.

[21] C. Li, G. Li, Z. Wang, X. Wang, P.K. Varshney, COMIC: an unsupervised change detection method for heterogeneous remote sensing images based on copula mixtures and cycle-consistent adversarial networks, Inf. Fusion (2024) 102240.

[22] Z. Zhang, C. Liu, F. Hao, Z. Liu, Style-transfer-based unsupervised change detection from heterogeneous images, IEEE Trans. Aerosp. Electron. Syst. 61 (3) (2025) 6537–6550.

[23] H. Wu, J. Geng, W. Jiang, Multidomain constrained translation network for change detection in heterogeneous remote sensing images, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–16.

[24] Y. Sun, L. Lei, D. Guan, G. Kuang, Z. Li, L. Liu, Locality preservation for unsupervised multimodal change detection in remote sensing imagery, IEEE Trans. Neural Netw. Learn. Syst. 36 (4) (2024) 1–15.

[25] R. Touati, M. Mignotte, M. Dahmane, Multimodal change detection in remote sensing images using an unsupervised pixel pairwise-based markov random field model, IEEE Trans. Image Process. 29 (2020) 757–767.

[26] Y. Sun, L. Lei, Z. Li, G. Kuang, Q. Yu, Detecting changes without comparing images: rules induced change detection in heterogeneous remote sensing images, ISPRS J. Photogramm. Remote Sens. 230 (2025) 241–257.

[27] J. Li, Q. Hu, M. Ai, RIFT: multi-modal image matching based on radiation-variation insensitive feature transform, IEEE Trans. Image Process. 29 (2020) 3296–3310.

[28] X. Ma, S. Xu, J. Zhou, Q. Yang, Y. Yang, K. Yang, S.H. Ong, Point set registration with mixture framework and variational inference, Pattern Recognit. 104 (2020) 107345.

[29] H. Zhang, L. Lei, W. Ni, X. Yang, T. Tang, K. Cheng, D. Xiang, G. Kuang, Optical and SAR image dense registration using a robust deep optical flow framework, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 16 (2023) 1269–1294.

[30] X. Ma, W. Zhao, X. Hao, Y. Yang, K. Yang, Remote sensing image registration with adjustable threshold and variational mixture transformation, IEEE Geosci. Remote Sens. Lett. 17 (5) (2019) 765–769.

[31] D. Xiang, X. Pan, H. Ding, J. Cheng, X. Sun, Two-stage registration of SAR images with large distortion based on superpixel segmentation, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–15.

[32] J. Zhou, X. Ma, L. Liang, Y. Yang, S. Xu, Y. Liu, S.-H. Ong, Robust variational Bayesian point set registration, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 9905–9914.

[33] W. Zhao, X. Ma, L. Liang, L. Liang, Y. Yang, K. Yang, S.H. Ong, Remote sensing image registration based on dynamic threshold calculation strategy and multiple-feature distance fusion, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (10) (2019) 4049–4061.

[34] J. Li, Q. Hu, Y. Zhang, Multimodal image matching: a scale-invariant algorithm and an open dataset, ISPRS J. Photogramm. Remote Sens. 204 (2023) 77–88.

[35] J. Li, W. Xu, P. Shi, Y. Zhang, Q. Hu, LNIFT: locally normalized image for rotation invariant multimodal feature matching, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–14.

[36] G. Brigot, E. Colin-Koeniguer, A. Plyer, F. Janez, Adaptation and evaluation of an optical flow method applied to coregistration of forest remote sensing images, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9 (7) (2016) 2923–2939.

[37] Y. Xiang, F. Wang, L. Wan, N. Jiao, H. You, OS-flow: a robust algorithm for dense optical and SAR image registration, IEEE Trans. Geosci. Remote Sens. 57 (9) (2019) 6335–6354.

[38] L.H. Nguyen, T.D. Tran, A sparsity-driven joint image registration and change detection technique for SAR imagery, in: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2010, pp. 2798–2801.

[39] S. Song, K. Jin, B. Zuo, J. Yang, A novel change detection method combined with registration for SAR images, Remote Sens. Lett. 10 (7) (2019) 669–678.

[40] M. Vakalopoulou, K. Karantzalos, N. Komodakis, N. Paragios, Graph-based registration, change detection, and classification in very high resolution multitemporal remote sensing data, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9 (7) (2016) 2940–2951.

[41] D.B. Mesquita, M.F.M. Campos, E.R. Nascimento, Learning to detect changes in aerial images in the presence of registration errors, IEEE Geosci. Remote Sens. Lett. 19 (2022) 1–5.

[42] R. Zhou, D. Quan, S. Wang, C. Lv, X. Cao, J. Chanussot, Y. Li, L. Jiao, A unified deep learning network for remote sensing image registration and change detection, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–16.

[43] B.D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: IJCAI'81: 7th International Joint Conference on Artificial Intelligence, 2, 1981, pp. 674–679.

[44] F. Bovolo, L. Bruzzone, A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain, IEEE Trans. Geosci. Remote Sens. 45 (1) (2006) 218–236.

[45] G. Moser, S.B. Serpico, Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery, IEEE Trans. Geosci. Remote Sens. 44 (10) (2006) 2972–2982.

[46] T. Han, Y. Tang, Y. Chen, B. Zou, H. Feng, Global structure graph mapping for multimodal change detection, Int. J. Digital Earth 17 (1) (2024) 2347457.

[47] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9 (1) (1979) 62–66.

[48] M. Mignotte, A fractal projection and markovian segmentation-based approach for multimodal change detection, IEEE Trans. Geosci. Remote Sens. 58 (11) (2020) 8046–8058.