Abstract—Change detection of heterogeneous remote sensing images is an important and challenging topic in remote sensing for emergency situation resulting from nature disaster. Due to the different imaging mechanisms of heterogeneous sensors, it is difficult to directly compare the images. To address this challenge, we explore an unsupervised change detection method based on adaptive local structure consistency (ALSC) between heterogeneous images in this letter, which constructs an adaptive graph representing the local structure for each patch in one image domain and then projects this graph to the other image domain to measure the change level. This local structure consistency exploits the fact that the heterogeneous images share the same structure information for the same ground object, which is imaging modality-invariant. To avoid the leakage of heterogeneous data, the pixelwise change image is calculated in the same image domain by graph projection. Experiment results demonstrate the effectiveness of the proposed ALSC based change detection method by comparing with some state-of-the-art methods.

Index Terms—Unsupervised change detection, Heterogeneous data, Adaptive local structure, Graph.

I. INTRODUCTION

Change detection (CD) is a process of identifying changes of an object or phenomenon by analyzing the remote sensing images acquired over the same geographical area at different times [1], [2]. Recently, heterogeneous CD has become a growing interest topic that allows the remote sensing community to make full use of the wide range of available earth observation satellites by considering the joint use of heterogeneous satellite images. In particular, the heterogeneous CD has important practical significance for the immediate evaluation of emergency disasters. In this case (such as earthquake or flood), the pre-event synthetic aperture radar (SAR) image is sometimes unavailable and the qualified post-event optical image cannot be obtained due to the adverse atmospheric conditions. Because different sensors provide different physical quantities for the same object and show different characteristics in the image, heterogeneous CD is a challenging task.

Heterogeneous CD is a challenging task since different sensors provide different physical quantities for the same object and show different characteristics in the images. Therefore, different from the algebraic methods (such as difference method, ratio method and log-ratio method) used in the homogeneous CD, it is impossible to compare the heterogeneous images directly to calculate the difference image (DI).

According to the different methods for generating the binary change map (CM), the existing heterogeneous CD methods can be roughly divided into three categories [3]: 1) classification-based methods, such as post-classification comparison (PC-C) [4], multitemporal segmentation and compound classification (MS-CC) [5], [6]; 2) deep-learning based methods, such as symmetric convolutional coupling network (SCCN) [7], logarithmic transformation feature learning (LTFL) based method [8], and conditional generative adversarial network (cGAN) [9]; 3) traditional DI-based methods, such as the local joint distributions with manifold learning based method [10], Markov model for multimodal change detection (M3CD) [11], supervised homogeneous pixel transformation (HPT) method [12], and affinity matrix based image regression (AM-IR) method [13]. The goal of the heterogeneous CD methods is to transform the “incomparable” heterogeneous images into a common “comparable” space, such as the category space of classification-based method, the learned high-dimensional feature space of deep learning-based method, and the constructed feature spaced of traditional DI-based method [3].

Recently, the self-similarity property is exploited to project the pre-event (post-event) image to the post-event (pre-event) image and then obtain a pixelwise DI for heterogeneous CD. In [14], the self-similarity property is used to complete the fractal projection, which contains a fractal encoding step and a fractal projection/decoding step. In [3], the self-similarity is used to complete self-expression, which learns a patch similarity graph matrix (PSGM) for each image and then measures the change level by multiplying one image with the learned PSGM from the other image and calculating the difference. In this letter, the self-similarity is further used to construct graphs representing the local structures for each image and establish the relationship between heterogeneous images, which is similar to the patch similarly graph based method in [15]. However, instead of using a fixed graph as in [15], the proposed method adaptively learns a distance induced probabilistic graph, which is more robust. This local structure consistency exploits the fact that the heterogeneous images share the same structure information for the same ground object, which is imaging modality-invariant. Different from the previous fractal projection and PSGM based methods, the proposed ALSC based method neither reconstructs the image nor transforms the image to the domain of the other image. It only focuses on the changes of local structure. At the same time, it also takes into account the prior statistics of heterogeneous images that are not used in fractal projection [14] and PSGM [3], which makes the CD results more accurate. The main contributions of this letter are summarized...
as follows.

1) A heterogeneous CD method based on adaptive local structure consistency is proposed, which measures the change level between the pre- and post-event images by calculating how much the local structure of one image still conforms to that of the other image.

2) The leakage of heterogeneous data in the DI generation is avoided by the graph projection, which calculates the structure difference within the same image domain.

3) The ALSC based method is completely unsupervised, and it exploits the inherent structure property of any satellite image that appeals quite imaging modality-invariant, so it has a strong flexibility to deal with a variety of heterogeneous image processing tasks.

II. ALSC BASED HETEROGENEOUS CD METHOD

We consider two coregistered heterogeneous images, \( X = \{ x(m, n, c) \mid 1 \leq m \leq M, 1 \leq n \leq N, 1 \leq c \leq C_X \} \) lining in the domain \( X \) and \( Y = \{ y(m, n, c) \mid 1 \leq m \leq M, 1 \leq n \leq N, 1 \leq c \leq C_Y \} \) lining in the domain \( Y \), which are acquired on the same geographical area by two different sensors at two different times, respectively. Here, \( M, N \) and \( C_X, C_Y \) represent the height, width and channel number of two images, respectively.

Using the self-similarity property, a small patch in the image can always find some similar patches within the image. This relationship between the target patch and its similar patches (K-nearest neighbors) can be regarded as the local structure of this target patch, which is quite well preserved across the different types of imaging modality [15]. As shown in Fig. 1, in the SAR image \( X \), the local structure of target patch \( X_i \) is represented by the relationship between \( X_i \) and its similar patches \( X_j \). If the area represented by these patches has not changed in the event, the local structure of target patch \( X_i \) can be preserved by the patch \( Y_i \) in the optical image \( Y \), which means that the patch \( Y_i \) is also very similar to the patches \( Y_j \). On the contrary, if the area represented by \( X_i \) has changed in the event, the local structure is no longer preserved by \( Y_i \), showing that \( Y_i \) and \( Y_j \) are very different. Therefore, we can see that the ALSC based heterogeneous CD method mainly needs to solve two problems: how to construct local structure and how to measure structural difference. The proposed method consists of four steps: 1) construct adaptive local structure; 2) calculate the structure difference; 3) generate the difference image; 4) obtain the binary change map.

A. Adaptive local structure

For a square target patch, \( X_i = \{ (m_i + \vartheta_m, n_i + \vartheta_n, c) \mid \vartheta_m, \vartheta_n \in [-p, p], 1 \leq c \leq C_X \} \) is a subset of \( X \), centered on position \( (m_i, n_i) \), in order to capture the local structure of \( X_i \), we construct a graph by adaptively assigning a probability \( S^X_{i,j} \) for \( X_j \) as the neighborhood of \( X_i \), and the probability is treated as the similarity between two patches, where \( X_j \) is another patch in a \( \omega \times \omega \) search window \( W \) centered on \( (m_i, n_i) \). Here, we use a search step size \( \Delta_s > 1 \) in \( W \) to make the distance between adjacent patches equal to \( \Delta_s \), which can reduce the number of neighbors and avoid redundancy.

Usually, a smaller distance between \( X_i \) and \( X_j \) should be assigned a large probability \( S^X_{i,j} \). To achieve this goal, we can solve the following problem

\[
\min_{S^X_{i,j}} \sum_{j=1}^{N_i} dist^X_{i,j} S^X_{i,j} + \gamma (S^X_{i,j})^2 \quad s.t. \quad 0 \leq S^X_{i,j} \leq 1, \quad \sum_{j=1}^{N_i} S^X_{i,j} = 1
\]

(1)

where \( dist^X_{i,j} \) represents a distance measure of two patches \( X_i \) and \( X_j \), \( N_i \) is the number of all potential neighbors of \( X_i \) in the search window \( W \), and \( \gamma > 0 \) is the trade-off parameter. The second term of the objective function in (1) is a regularization, which is used to avoid the trivial solution, i.e., only the nearest patch can be the neighbor of \( X_i \) with probability 1 (when \( \gamma = 0 \)). If we only focus on the second term (when \( \gamma \to \infty \)), the optimal solution of (1) is that all the patches in \( W \) can be the neighbors of \( X_i \) with the same probability \( 1/N_i \).

The patch distance \( dist^X_{i,j} \) should be computed based on the prior statistics of the image \( X \). For two patches \( X_i \) and \( X_j \), we vectorized them and denote each element as \( x_i(q) \) and \( x_j(q) \), respectively. 1) For the optical image, assuming the additive white Gaussian noise (AWGN) model, the traditional Euclidean distance is usually used

\[
dist^X_{i,j} = \| X_i - X_j \|^2_F
\]

(2)

2) For the SAR image, assuming the multiplicative speckle noise modeled by a Gamma distribution, the following distance criterion [16] can be used

\[
dist^X_{i,j} = \sum_{q=1}^{(2p+1)C_X} \log \left( \frac{x_i(q) + x_j(q)}{2 \sqrt{x_i(q) x_j(q)}} \right)
\]

(3)

It should be noted that the distance calculation function is not fixed, which should be determined according to the prior statistical information of the image.

Denote \( S^X \in \mathbb{R}^{N_i} \) and \( dist^X \in \mathbb{R}^{N_i} \) as column vectors composed of \( S^X_{i,j} \) and \( dist^X_{i,j} \), respectively, and denote \( I \) as a column vector with all the elements being one, the problem (1) can be reformulated as

\[
\min_{S^X} \sum_{1 \leq i \leq 1, 0 \leq S^X_{i,j} \leq 1} \left( S^X_{i,j} - \frac{1}{2\gamma} \right)^2
\]

(4)
and the number of nearest neighbors (that is, \(i; i = 2\) and backward and the backward structure difference in ascending order as can also be generated in \(i\) and \(i\) is only connected with its \(\text{dist}\) to the sets \(i\) and \(i\), we need to compare them to calculate, we give a summary of \(i = 1\) and \(i\) can be repre-

C. Generate the difference image

Once the forward structure difference \(f^X\) is calculated, it is assigned to all the pixels of patch \(i\). After we calculate the forward structure differences for all the overlapping patches of the two images, for each pixel \(j, j = 1, \cdots, MN\) in the forward DI there is a set \(f^Y\) of structure difference that covers the pixel \(j\). Therefore, the final forward change level of this specific pixel \(j\) is the mean of this set as

\[
DI_j^Y = \frac{1}{|F_j^Y|} \sum_{i \in F_j^Y} f_j^Y
\]

Then we can obtain the forward difference image \(DI_j^Y\). At the same time, the backward \(DI_j^X\) can also be generated in a similar way. We can obtain the final \(DI_{\text{final}}\) by fusing the forward \(DI_j^Y\) and backward \(DI_j^X\) as

\[
DI_{\text{final}} = DI_j^X + DI_j^Y
\]

Given the set \(\Lambda\) of all target patches spaced by a patch step size \(\Delta_p \in [1, p]\), Algorithm 2 listed in Table II summarizes the procedure of generating the ALSC based DI. With this \(\Delta_p\), the amount of target patches can be reduced by a factor of \(\Delta_p\), thereby speeding up the algorithm.

D. Obtain the binary change map

After the fused DI is obtained, the CD can be completed as an binary segmentation problem. Then, we can use the thresholding method (such as the Otsu threshold method [18]) or the clustering method (such as the PCAKM [19] that using principal component analysis to extract the features and employing the \(k\)-means with \(k = 2\)) to obtain the binary CM.

III. EXPERIMENTAL RESULTS

Table I IMPLEMENTATION STEPS OF ALGORITHM 1.

| Input: Distance vector \(dist^X_i\) and the number of nearest neighbors \(K\). | Output: The similarity vector \(S^X_i\). |
|---|---|
| 1. Sorting \(dist^X_i\) in ascending order as \(dist^X_i; i = 1, \cdots, \text{dist}^X_{(N+1)}\). |  |
| 2. Calculating the \(S^X_i\) with \(i; i = 1, 2, \cdots, N\) as the \(i\)-th smallest element of \(dist^X_i\), we give a summary of this efficient solution as shown in Algorithm 1 of Table I, and more details can be found in [17]. |  |

From Algorithm 1, we can find that the regularization parameter \(\gamma\) is replaced by the number of neighbors \(K\), which are equivalent when we set \(\gamma\) to be

\[
\gamma = \frac{K}{2} \cdot \frac{dist^X_i; i = 1, \cdots, N}{2} \cdot \frac{dist^X_i; i = 1, \cdots, K}{2} \cdot \frac{dist^X_i; i = 1, \cdots, K}{2}
\]

In this way, the search of parameter \(\gamma\) can be better handled by searching the neighborhood size \(K\), which is more intuitive (it has explicit meaning) and easy to tune (it is an integer).

B. Calculate the structure difference

As the local structure of patches \(X_i\) and \(Y_i\) can be represented by \(S^X_i\) and \(S^Y_i\), we need to compare them to calculate the structure difference. However, because \(S^X_i\) and \(S^Y_i\) are generated based on different distance vectors \(dist^X_i\) and \(dist^Y_i\), which are calculated on different domains \(X\) and \(Y\), directly comparing \(S^X_i\) and \(S^Y_i\) (such as \(\|S^X_i - S^Y_i\|_2\)) will cause the leakage of heterogeneous data. To avoid the leakage, we first project the graph \(S^X_i\) (or \(S^Y_i\)) of image domain \(X\) (or \(Y\)) to the other image domain \(Y\) (or \(X\)), and then compare the difference in the same image domain to obtain the structure difference. Therefore, we can compute the forward structure difference \(f^Y\) and the backward structure difference \(f^X\) as

\[
f^Y_i = \sum_{h=1}^{h=K} dist^Y_{(h)} S^X_{(h)}; f^X_i = \sum_{h=1}^{h=K} dist^X_{(h)} S^Y_{(h)}
\]

(6)

The intuitive explanation of \(f^Y_i\) in (6) is: for the patch \(X_j\) with \(j = i\), which is the \(h\)-th nearest neighbor of target patch \(X_i\), the \(dist^Y_{(h)}\) is the patch difference between the projected patch \(Y_j\) (that is \(Y_j\)) and the \(h\)-th nearest neighbor of patch \(Y_i\) (that is \(Y_j\)), and the \(S^X_{(h)}\) is the weight of this patch difference. From this, we can find that the patch difference is calculated in the same image domain, which avoids the leakage of heterogeneous data.

Algorithm 2. ALSC based DI generation

| Input: Heterogeneous images \(X, Y\), parameters \(p, w, \Delta_p, \Delta_s\) and \(K\). | Output: The similarity vector \(S^X_i\). |
|---|---|
| 1. Calculate the structure difference for all target patches \(X_i, Y_i, i \in \Lambda\) do | Compute the distances \(dist^X_i\) and \(dist^Y_i\) with different criteria. Compute the similarities \(S^X_i\) and \(S^Y_i\) by using Algorithm 1. Compute the structure difference \(f^X_i\) and \(f^Y_i\). Add \(f^X_i\) and \(f^Y_i\) to the sets \(F^X_i\) and \(F^Y_i\), respectively. |
| end for |  |

2. Compute the forward and backward DIs for all pixels \(j, j = 1, \cdots, MN\) do |

Compute the forward \(D^Y_i\) and backward \(D^X_i\).

end for

3. Fuse the forward and backward DIs

\[
D^Y_{\text{final}} = D^Y_i + D^Y_i
\]

Given the set \(\Lambda\) of all target patches spaced by a patch step size \(\Delta_p \in [1, p]\), Algorithm 2 listed in Table II summarizes the procedure of generating the ALSC based DI. With this \(\Delta_p\), the amount of target patches can be reduced by a factor of \(\Delta_p\), thereby speeding up the algorithm.
Lake Mulargia, on the island of Sardinia, Italy. (2) Shuguang
data set: a pair of SAR/optical images. The 593 × 921 × 1 SAR
image is acquired by Radarsat-2 and the 593 × 921 × 3 optical
image is obtained from Google Earth. The ground truth shows
the changes of land use in Shuguang Village, Dongying City,
China. (3) Wuhan data set: a pair of SAR/optical images. The
495 × 503 × 1 SAR image is acquired by Radarsat-2 and the
495 × 503 × 3 optical image is obtained from Google Earth.
The ground truth shows the changes of new buildings and
roads in Wuhan City, China. (4) California data set: a pair of
multispectral/SAR images. The 875 × 500 × 11 multispectral
image is acquired by Landsat-8 and the 875 × 500 × 3 SAR
image is acquired by Sentinel-1A, which is recorded
in polarisations VV and VH, and augmented with the ratio
between the two intensities as the third channel. The ground
truth shows a flood in California, USA, which is provided by
Luppino et al. [13].

B. Experimental results and analysis

As listed in Table II, the parameters of ALSC based method
are the patch size \( p \), the search window size \( \omega \), the patch step
size \( \Delta_p \), the search step size \( \Delta_s \), and the number of neighbors
\( K \). For all the experimental results, we fix \( \omega = 75p \), \( \Delta_p = p \),
\( \Delta_s = 2p + 1 \), \( K = 35 \), and vary the \( p \) from 1 to 4 to select
the best one for each data set. Specifically, we set \( p = 1 \) for
Wuhan, \( p = 2 \) for Sardinia and California, and \( p = 3 \) for
Shuguang.

Fig. 2(d) shows the ALSC generated DI of different data
sets. We can find that the proposed ALSC can well establish
the relationship between heterogeneous images, which can
highlight the changes in the DI. Fig. 3 plots the empirical
receiver operating characteristics (ROC) curves of ALSC gen-
erated DI, and shows that the quality of these DI is very high,
which gain a large area under the ROC curves as 0.970, 0.979,
0.957 and 0.925 for Sardinia, Shuguang, Wuhan and California
data sets, respectively.

With these ALSC based DI, the binary CM can be obtained
by using the Otsu thresholding (denoted as ALSC-O for short)
and PCAKM (denoted as ALSC-P for short), as shown in
Figs. 2(e)-(f). Four measures are considered as the evaluation
criteria of CM: the false positives (FP) rate, the false negatives
(FN) rate, the overall accuracy (OA), and the Kappa coefficient
Performance on different heterogeneous data sets. Our method, ALSC based CD, clearly showed that the ALSC based CD method can achieve effective leakage of heterogeneous data. Experimental results clearly demonstrated that the ALSC based CD method can be implemented on the same image domain by graph projection, which can avoid the leakage of heterogeneous data. This is achieved by combining local invariance with adaptive neighbors, and then comparing the structure difference in the input image, and then comparing the structure difference in the target patch, this local invariance will no longer be maintained.

Therefore, the ALSC based CD method can be implemented in two steps. It first constructs adaptive local structure for each input image, and then compare the structure difference in the same image domain by graph projection, which can avoid the leakage of heterogeneous data. Experimental results clearly show that the ALSC based CD method can achieve effective performance on different heterogeneous data sets.

In order to further evaluate the performance of ALSC, some representative and state-of-the-art (SOTA) methods including PSGM [3], LTFL [8], M3CD [11], AM-IR [13], fractal projection and Markovian segmentation-based method (FP-MS) [14], nonlocal patch similarly graph-based method (NPSG) [15], reliable mixed-norm-based heterogeneous CD method (RMN) [20], and anomaly feature learning based deep sparse residual model (AFL-DSR) [21], are selected for comparison as summarized in Table IV. We can see that ALSC can achieve better or quite competitive performance by comparing with these SOTA methods, and ALSC shows the ability to gain consistent good results on different data sets. The average accuracy rates obtained on four heterogeneous data set of the ALSC with Otsu thresholding and PCAKM are 95.46% and 94.66%, respectively.

### IV. Conclusion

In this letter, we mainly address the problem of change detection in heterogeneous remote sensing. To find out a relationship between heterogeneous images and make them comparable, we explore the ALSC between heterogeneous images, which exploits the inherent structure property of images and appeals quite imaging modality-invariant. The core idea of the ALSC is that the local structure of each target patch in the pre-event (post-event) image will keep invariance in the post-event (pre-event) image when there is no change occurred; on the contrary, once a change occurred within this target patch, this local invariance will no longer be maintained. Therefore, the ALSC based CD method can be implemented in two steps. It first constructs adaptive local structure for each input image, and then compare the structure difference in the same image domain by graph projection, which can avoid the leakage of heterogeneous data. Experimental results clearly show that the ALSC based CD method can achieve effective performance on different heterogeneous data sets.