IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-2792-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/IGARSS46834.2022.9883323

HETEROGENEOUS IMAGE CHANGE DETECTION BASED ON TWO-STAGE JOINT FEATURE LEARNING

Te Han, Yuqi Tang, Yuzeng Chen

School of Geosciences and Info-Physics, Central South University, Changsha 410083, China
* Correspondence: yqtang@csu.edu.cn; Tel.: +86-139-7311-5356

Abstract

Heterogeneous image change detection, in contrast to homogeneous image change detection, has been a research hotspot due to the information complementary of different imaging mechanisms. However, the imaging difference leads to challenges on change detection by image comparison. To address the incomparability among heterogeneous images and improve the efficiency of heterogeneous image change detection, this paper proposes a novel heterogeneous image change detection method based two-stage joint feature learning. Assuming that the change is few and the image differences in unchanged areas between heterogeneous images are related to the imaging and environmental differences, it maps heterogeneous images into a similar feature space for comparison. Firstly, the bi-temporal similar feature maps with high similarity are extracted after joint feature learning of heterogeneous image. And the similar feature maps are used for joint feature learning optimized by a similarity measure in order to map them to an approximate feature space for comparison. Then the change map is obtained by segmenting the difference between the optimal feature maps. The experiments prove its superiority over existing methods on two heterogeneous image datasets (optical and synthetic aperture radar (SAR) images).

Index Terms**—**Change detection; heterogeneous image; joint feature learning

I. INTRODUCTION

The rapid development of remote sensing technique has provided different types of remote sensing data (e.g., multispectral, hyperspectral and synthetic aperture radar (SAR) images). How to effectively extract surface change information from massive amounts of remote sensing data has become one of the key topics in remote sensing application research. The purpose of remote sensing image change detection is to identify surface changes using remote sensing images covering the same region at different times[1], which is commonly used in land-use monitoring[2], and disaster assessment[3]. Remote sensing image change detection can be divided into homogeneous image change detection and heterogeneous image change detection based on the type of the data used.

The homogeneous image change detection utilized images obtained from the same satellite sensors, such as optical-optical images and SAR-SAR images. However, because of the constraints of sensor performance and temporal resolution, such techniques are prone to lack for appropriate data in practical applications. For example, optical images offer rich spectrum and textural features, while being vulnerable to imaging conditions; SAR sensors can image all day and any weather, but the images have significant speckle noise, making image interpretation challenging. As a result, in some observation tasks accompanied by severe weather, such as floods, earthquakes, and other natural calamities, collaborative observations by multi-sources images are required for timely changes detection.

In recent years, heterogeneous image change detection has received much attention. Liu *et al*. [4] proposed a homogeneous pixel transformation (HPT) based change detection method for heterogeneous image, which constructed the pixel transformation relationship between heterogeneous image by manually selecting samples and obtained the changed regions by comparing the original image and the transformed image. To enhance the automaticity of the algorithm, Sun *et al*.[5] proposed a nonlocal patch similarity-based graph (NPSG) to detect changed regions. To obtain the same spectral features from heterogeneous image, Niu *et al*. [6] used conditional generative adversarial network (CGAN) to map the images into the same feature space for comparison. Zhan *et al*. [7] proposed a feature learning framework based on logarithmic transformation (LTFL) to extract feature maps from heterogeneous image, and classified the feature maps to obtain the changed regions. However, these methods directly use the original heterogenous images for feature space mapping, which easily leads to inaccurate mapping. And the detection efficiency needs to be improved.

In this paper, a change detection method for SAR and optical images based on two-stage joint feature learning (TSJFL) is proposed. It assumes that the change is few and the image differences in unchanged areas between heterogeneous images are related to the imaging and environmental differences. A pair of similar feature maps of heterogeneous images are firstly extracted and selected by joint feature learning and similarity calculation, respec-

Authorized licensed use limited to: Wuhan University. Downloaded on March 16,2024 at 03:10:22 UTC from IEEE Xplore. Restrictions apply.

tively. And they are further conversed into an approximate feature space by joint feature learning optimized with similarity measure. As the imaging and environmental differences are suppressed by above two stages of joint feature learning, the optimal feature maps are compared and segmented to detect and surface changes in heterogeneous images. In the experiments, two sets of heterogeneous images are used to validate the proposed TSJFL method compared with four state-of-the-art ones.

The rest of the paper is structured as follows. Section II details the methodology of this study. Section III describes the experimental setup and results, and concludes the full paper in Section IV.

II. METHODOLOGY

The optical and SAR images covering the same geographical area acquired at various times are respectively given as $X=\{X(i,j,b), 1\le i\le H, 1\le j\le W, 1\le b\le B\}$ and $Y=\{Y(i,j,b), 1\le i\le H, 1\le j\le W, 1\le b\le B\}$, where H, W and B are the images' length, width and the number of bands, respectively. A logarithmic transformation of SAR images is implemented to obtain the same data distribution as optical images.

Between bi-temporal images, the unchanged region is much larger than changed one in general. The image difference in unchanged region between heterogeneous image is arised from various imaging mechanisms and condition. Therefore, this paper regards that the higher similarity between feature maps of heterogeneous image, the less affected by imaging mechanism and environment. And the feature maps optimized with similarity learning can be compared for change detection in an approximate feature space. Fig. 1 shows the flowchart of TSJFL. It consists of three major steps: 1) the first stage of joint feature learning selects similar feature maps; 2) the second stage of joint feature learning further maps the features to an approximate feature space with the highest similarity; and 3) change map is generated by comparing the optimal feature maps and clustering the difference map for change detection.

A. First stage of joint feature learning

Considering the task of detecting changes in heterogeneous image, it is required to map the heterogeneous image to an approximate feature space. The deep learning (DL) are being commonly employed in feature extraction tasks. It has a deep model structure and can transform the original data into a feature space, in which the data information is simpler to be recognized and predicted using a layer-by-layer feature representation. Therefore, multi-layer forward encoding (MFE) [8] is utilized as the feature extraction network, which has been proved effective and efficient. In MFE, each layer is an independent feature extraction unit. And the weights of its hidden layers are assigned randomly, which do not need to be fine-tuned. The output of the hidden layer is defined for each layer of MFE as follows:

$$\boldsymbol{H}_i = g(\boldsymbol{H}_{i-1} \times \boldsymbol{\beta}^*) \tag{1}$$

where, $\boldsymbol{H}_i$ denotes the output of layer i ($i\in[1,M]$), M represents the number of hidden layers, $\boldsymbol{\beta}^*$ means the optimized hidden layer weight, and $g(\bullet)$ is the activation function of the hidden layer.

$\boldsymbol{\beta}^*$ is optimized as follows:

$$\boldsymbol{\beta}^* = \arg\min\left\{\|\boldsymbol{H\beta}-\boldsymbol{X}\|^2+\|\boldsymbol{\beta}\|_{l1}\right\} \tag{2}$$

where, $\boldsymbol{X}=\{\boldsymbol{x}_i \mid \boldsymbol{x}_i\in\mathbf{R}^d, i=1,\ldots,N\}$ means the input data, N denotes the total number of input data. $\boldsymbol{\beta}$ is the randomly generated hidden layer weights for initialization, L represents the number of hidden layer nodes. MFE optimized the model using a fast iterative shrinkage thresholding algorithm (FISTA) [9].

We constructed the network net_1 with structure of $w\times w\times B\times 2$ -100-150-50-30 for the first stage of joint feature learning based on MFE, where w is the size of image patch to avoid the vulnerability to image noise.

The Maximum Mean Discrepancy (MMD) distance [10] is used to calculate the similarity between feature maps of heterogeneous images. A shorter MMD distance indicates that the bi-temporal feature maps are more similar.

The pair of feature maps with the shortest MMD distance are obtained as the similar feature maps in the first stage of joint feature learning (Fig. 1).

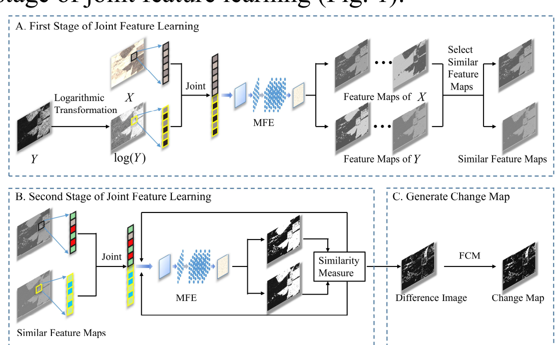


Fig. 1. Flowchart of the proposed TSJFL method.

B. Second stage of joint feature learning

In order to map images X and Y to an approximate feature space, the similar feature maps extracted from the first stage of joint feature learning are used as input of the second one. The network net_2 with structure of $w\times w\times B\times 2$ -100-150-50-1 based on MFE is constructed for this second stage. The main steps are as follows:

Authorized licensed use limited to: Wuhan University. Downloaded on March 16,2024 at 03:10:22 UTC from IEEE Xplore. Restrictions apply.

Step 1: Randomly initialize the weights of *net*$_2$, noted as $\boldsymbol{\beta}_0$;

Step 2: Learning update of *net*$_2$. The $w \times w$ image patch is used as the processing unit to input the similar feature maps into *net*$_2$ for joint feature learning. The initialization weight at the *t*th($t \geq 1$) iteration of the *net*$_2$ is calculated according to Eq. (1) and is denoted as $\boldsymbol{\beta}_{t-1}^{*}$. The feature maps of images X and Y obtained at the *t*th iteration are noted as $f_t(X)$ and $f_t(Y)$. The difference between $f_t(X)$ and $f_t(Y)$ is:

$$dif_t = MMD\left(f_t(X), f_t(Y)\right) \tag{4}$$

Step 3: Calculate the absolute difference Δ between the dif_t and dif_{t+1}. If $\Delta = \left|dif_{t+1} - dif_t\right| \leq \varepsilon$, stop iteration and output the optimal feature maps $f_{t+1}(X)$ and $f_{t+1}(Y)$. Otherwise, the iteration continues.

C. Change map generation

Through above two stages of joint feature learning, the optimal feature maps $f_{t+1}(X)$ and $f_{t+1}(Y)$ are in an approximate feature space unaffected from the various imaging mechanisms and environments. And the difference map of heterogeneous images can be obtained by direct compare between $f_{t+1}(X)$ and $f_{t+1}(Y)$. Consequently, the change map is obtained by binary segmentation of the difference map using the fuzzy c-means (FCM) algorithm[11] in this paper, which is a classical clustering algorithm.

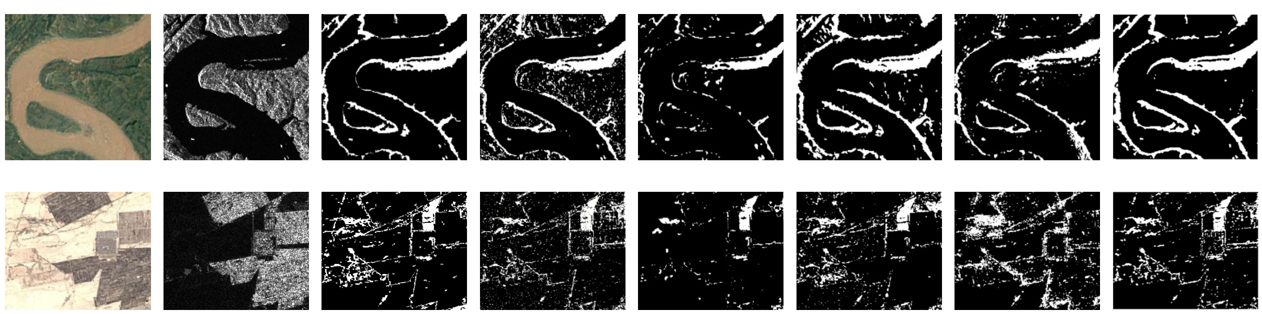
III. Experiment

A. Experiment settings

To validate the proposed TSJFL method, two sets of heterogeneous images acquired by optical and SAR sensors are used. Yangtze River dataset shown in the first row of Fig. 2(a) and Fig. 2(b). The size of each image is 600 x 600 pixels with the resolution of 10 m, which is acquired in December 1999 and November 2017, respectively. There are land and water in this area, and the changes are in water. As shown in the last row of Fig. 2(a) and Fig. 2(b) is the Guazhou dataset, which is obtained in December 2014 and July 2018, respectively. Both images are 1310 × 888 pixels in size at a resolution of 5 m. It is used to monitor the change of vegetation cover in Guazhou. Fig. 2(c) and Fig. 3(c) respectively shows the reference image of each dataset obtained by visual interpretation with ancillary data.

To quantitatively assess the change detection method proposed in this paper, the accuracy evaluation is performed using the *overall error* (OE), *overall accuracy* (OA), *Kappa coefficient* (KC) and run time/s. The Kernel parameter σ of MMD distance, window size w and tolerance value ε is respectively set to 6, 5 and 0.01, in which w can be selected according the image resolution and change scale, σ and ε can be regarded as constant between Google Earth and Sentinel 1 images.

The existing methods HPT[4], CGAN[6], LTFL[7] and NPSG[5] are used for comparison methods to validate the proposed method in this paper. All algorithms were written in MATLAB language and tested running environment: AMD Ryzen 7 CPU at 3.89 Hz, 64 GB RAM, Windows 10 (64 bit), MATLAB 2020a.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 2. Change maps generated by different methods on Guazhou dataset. (a) and (b) Images acquired from different sensors. HPT. (c) Reference map. (d) HPT. (e) CGAN. (f) LTFL. (g) NPSG. (h) TSJFL.

B. Result analysis

Fig. 2 shows the change detection results of the comparison methods and proposed TSJFL on the Yangtze River dataset. Visually, the noise of HPT, LTFL and NPSG is more serious, which producing many false detections, and CGAN misses some changed regions. Through two-stage joint feature learning, the result of proposed TSJFL is the closest to the reference image since it can map heterogeneous images to an appropriate feature space for comparison, excluding "pseudo-change" caused by imaging mechanism and environmental differences.

The results on the Guazhou dataset are shown in Fig. 2. By visual comparison, HPT, LTFL and NPSG have some noise and false detections, while CGAN is more robust to image noise but misses some changes. The result of the

Authorized licensed use limited to: Wuhan University. Downloaded on March 16,2024 at 03:10:22 UTC from IEEE Xplore. Restrictions apply.

proposed TSJFL is the closest to the reference image with the least noise. It verifies the TSJFL can not only detect water changes, but also improve the accuracy and completeness of change detection in more complex vegetation areas by deeply excavating the changes in heterogeneous images on the basis of excluding the "pseudo-change" caused by imaging mechanism and environmental differences.

Table I and II show the results of quantitative analysis among the proposed TSJFL and comparison methods on two datasets, respectively. The TSJFL has the lowest OE and the highest OA and KC among all methods for both datasets, which indicates that it can accurately detect changed regions by mapping heterogeneous images to an approximate feature space. In terms of computational efficiency, the proposed TSJFL improves at least 76.25% over comparison methods. It is proved to be effective and efficient for change detection in optical and SAR images.

Table I. Quantitative analysis of different methods on Yangtze River dataset

Methods	OE	OA(%)	KC	run time/s
HPT	29,849	91.71	0.6854	177.19
CGAN	31,338	91.30	0.5551	860.47
LTFL	28,646	92.04	0.7119	2596.56
NPSG	38,065	89.43	0.5394	655.89
TSJFL	**18238**	**94.93**	**0.8134**	**42.07**

Table II. Quantitative analysis of different methods on Guazhou dataset

Methods	OE	OA(%)	KC	run time/s
HPT	84,566	91.35	0.6039	628.32
CGAN	87,948	91.00	0.4490	2295.29
LTFL	77,804	92.04	0.6206	4358.68
NPSG	144,226	85.25	0.3831	1858.14
TSJFL	**71847**	**92.65**	**0.6383**	**112.21**

IV. CONCLUSION

In order to solve the problem of inaccurate and inefficient feature space mapping in heterogeneous image change detection, a two-stage joint feature learning (TSJFL) based method is proposed in this paper. It regards that the higher similarity between feature maps of heterogeneous image, the less affected by imaging mechanism and environment. And the feature maps optimized with similarity learning can be compared for change detection in an approximate feature space. In the first stage of joint feature learning, it extracts the similar feature maps of the heterogenous images, which are easier to be further mapped to a more approximate feature space for comparison in the second stage. Consequently, the change map is obtained by binary segmentation of the difference map. Comparing four methods with two sets of optical and SAR images in the experiments, the proposed TSJFL improves the accuracy and computational efficiency respectively by more than 2.85% and 76.25% in terms of KC and run time, which demonstrate its effectiveness on change detection in heterogeneous images.

V. ACKNOWLEDGMENT

The authors would like to thank the IEEE Geoscience and Remote Sensing Society and the IGARSS 2022 Organizing Committee. This research was funded by National Natural Science Foundation of China under Grant, 41971313; Key Research and Development Program of Hunan Province, 2019SK2112.

VI. REFERENCE

[1] D. Lu, P. Mausel, E. Brondízio, and E. Moran, “Change detection techniques,” *Int. J. Remote Sens.*, 2004, 25(12):p.2365-2407.

[2] Mas and J.-F., “Monitoring land-cover changes: A comparison of change detection techniques,” *Int. J. Remote Sens.*, vol. 20, no. 1, pp. 139–152, 1999.

[3] D. Al-Khudhairy, I. Caravaggi, and S. Giada, “Structural Damage Assessments from Ikonos Data Using Change Detection, Object-Oriented Segmentation, and Classification Techniques,” *Photogramm. Eng. Remote Sens.*, vol. 71, no. 7, p. págs. 825-838, 2005.

[4] Z. Liu, G. Li, G. Mercier, Y. He, and Q. Pan, “Change Detection in Heterogenous Remote Sensing Images via Homogeneous Pixel Transformation,” *IEEE Trans. Image Process.*, 2018, 27(4):1822.

[5] Y. Sun, L. Lei, X. Li, H. Sun, and G. Kuang, “Nonlocal patch similarity based heterogeneous remote sensing change detection,” *Pattern Recognit.*, vol. 109, p. 107598, 2020.

[6] X. Niu, M. Gong, T. Zhan, and Y. Yang, “A Conditional Adversarial Network for Change Detection in Heterogeneous Images,” *IEEE Geosci. Remote Sens. Lett.*, 2019. 2018, PP(1):1-5.

[7] T. Zhan, M. Gong, X. Jiang, and S. Li, “Log-based transformation feature learning for change detection in heterogeneous images,” *IEEE Geosci. Remote Sens. Lett.*, 2018, PP:1-5.

[8] J. Tang, C. Deng, and G. Bin Huang, “Extreme Learning Machine for Multilayer Perceptron,” *IEEE Trans. Neural Networks Learn. Syst.*, 2016, 809-821..

[9] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[10] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, “A kernel two-sample test,” *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 723–773, 2012.

[11] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. 1981. 4, 95-154.

Authorized licensed use limited to: Wuhan University. Downloaded on March 16,2024 at 03:10:22 UTC from IEEE Xplore. Restrictions apply.