COARSE-TO-FINE UNSUPERVISED CHANGE DETECTION FOR REMOTE SENSING IMAGES VIA OBJECT-BASED MRF AND INCEPTION UNET

Xuan Hou[1], *Yunpeng Bai*[2], *Haonan Shi*[1], *Ying Li*[1,*]

[1]School of Computer Science, National Engineering Laboratory for Integrated Aero-Space-Ground -Ocean Big Data Application Technology, Shaanxi Provincial Key Laboratory of Speech Image Information Processing, Northwestern Polytechnical University, Xi'an 710129, China
[2] Department of Computer Science, Faculty of Business and Physical Sciences, Aberystwyth University, Aberystwyth SY23 3DB, UK

ABSTRACT

With the rapid development of various satellite sensor techniques, remote sensing imagery has been an important source of data in change detection applications. This paper aims to propose an unsupervised change detection method based on Object-based Markov Random Filed (OMRF) and Inception UNet (IUNet). Our method first utilizes a difference image (DI) obtained from two bi-temporal images as the initial feature, and proposes the OMRF algorithm based on homogeneous region to pre-classify the DI thus derive the coarse change map. The IUNet is then constructed to extract the points with high confidence from the coarse change map for training. Eventually, the trained model is fed to classify the original feature, then the final change map is obtained. Experimental results indicate that our method yields great detection results even without supervision.

Index Terms— change detection, object-based markov random filed, coarse-to-fine model, unsupervised learning

1. INTRODUCTION

Change Detection (CD) of remote sensing imagery is to analyze multi-temporal images at the same place, and then acquire the change information [1]. Indeed, remote sensing image CD has been extensively applied to address various issues, including: disaster assessment[2], land management[3], resource management[4],urban expansion research[5]. For CD task, label samples often require labeling by specific experts, while the workload is enormous. Therefore, unsupervised CD algorithm is of crucial practical significance. However, existing unsupervised CD algorithms has two limitations: (1) The performance of this algorithm is much inferior compared to the supervised learning algorithm. (2) Global information of the images haven't been fully utilized which will inhibit the accuracy of detection.

Markov Random Field (MRF) model has made gratifying achievements for unsupervised CD due to its capability to exploit spatial context information in the images[6]. Bruzzone et al.[7] first adopted MRF for remote sensing images CD. Benedek et al.[8] raised the conditional mixed Markov model for CD, which is a multi-layer MRF model combining the mixed Markov model and the conditional independent random field of signals. Gong et al.[6] proposed a Synthetic Aperture Radar (SAR) CD algorithm based on fuzzy clustering and the improved MRF energy function. Subudhi et al.[9] proposed a CD method by using fuzzy Gibbs Markov Random Field to model the spatial gray grey characteristics of DI. It effectively avoids the issue that traditional algorithms cannot obtain the global optimum. Convolutional Neural Network (CNN) is also populated in supervised CD [10, 11]. Zhang et al.[12] proposed a novel supervised CD approach based on Deep Siamese Convolutional Networks (DSCN) for direct convolution of the bi-temporal images of optical aerial images without pooling to ensure that the same size of the output and input. Then, the threshold is used to segment the distance map obtained from the output feature map to obtain the change map. Daudt et al.[13] proposed a Fully Convolutional siamese network for CD (FC-siam) and fused the features of two encoders with the decoder in the skip connection. Li et al.[14] used unsupervised spatial fuzzy clustering to produce false labels, and then proper labels were selected to train CNN (SFCCNN).

The contribution of this paper can be summarized in three points: firstly, we proposed the OMRF algorithm which exploits the global information and improved the stability of classification. Second, the idea of coarse-to-fine is introduced in unsupervised CD in our paper, which causes effective improvement of change map accuracy. Thirdly, experiments performed on four public remote sensing datasets demonstrate that our proposed method surpasses other unsupervised methods and even outperforms some supervised ones.

*Ying Li is the corresponding author. This work was supported in part by the National Natural Science Foundation of China under Grant 61871460; in part by the Shaanxi Provincial Key Research and Development Program under Grant 2020KW-003

2. THE PROPOSED METHOD

In this section, we elaborate our proposed method. As shown in Fig.1, our model mainly consists of two parts: coarse CD and refined CD. The coarse CD uses OMRF algorithm to analyze the DI and obtain the coarse change map. The refined CD is a IUNet which is trained by a coarse change map and DI, and then a progressive change map can be obtained by the trained IUNet.

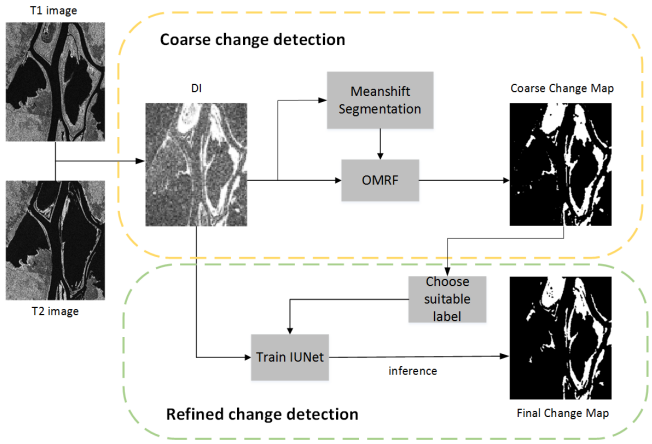

Fig. 1. Framework of the proposed method

2.1. Coarse change detection

The traditional MRF-based CD algorithms are under the assumption of global consistency. However, in the actual scenario, the image possesses its local characteristics. The inconsistency between the assumption and the actual situation affects the accuracy of the algorithm results. Therefore, a CD algorithm based on OMRF model is proposed to eliminate the inconsistency between them.

In our proposed method, the DI is regarded as the initial feature. The DI of the spectral image is $DI = |T1 - T2|$, where $T1$ and $T2$ represent remote sensing images of the same place at different times. The DI of SAR image is $DI = |\log T1 - \log T2|$. Structure $S = \{(i,j) \mid 1 \leq i \leq M, 1 \leq j \leq N\}$ is constructed for the image with size $M \times N$. According to MRF theory, the random field of the structure can be defined as $X = \{x_s | s \in S\}$,where $x_s = \Lambda$ represents the status of random filed. In S, the neighborhood structure centered on each pixel point s can be defined as $\delta = \{\delta(s) s \in S\}$, which satisfies:

$$\begin{aligned} &1)\delta(s) \subset S \\ &2)s \notin \delta(s) \\ &3)\forall s, r \in S, s \in \delta(r) \Leftrightarrow r \in \delta(s) \end{aligned} \tag{1}$$

The first-order potential function and second-order potential function are required in the modeling process of MRF. Let $D(i,j)$ be the observation field, corresponding to the pixel at image coordinate(i,j), and $C(i,j)$ be the label field, corresponding to real change label at image coordinate (i,j). The first-order potential function is as follows:

$$U_{data}(D(i,j)/C(i,j)) = \frac{\ln 2\pi\sigma_k^2}{2} + \frac{D(i,j) - \mu_k}{2\sigma_k^2} \tag{2}$$

where $U_{data}(D(i,j)/C(i,j))$ represents the relationship between $D(i,j)$ and $C(i,j)$, which refers to the corresponding relationship between observation field and label field.μ_k and σ_k are Gaussian model parameters estimated by the expectation–maximization (EM) algorithm.

In the CD task, if the first-order potential function is applied between two images (fields), then the second-order potential function will be applied to the neighborhood structure. Let $C_l(i,j)$ be the predicted value of the point (i,j), and the neighborhood of the point in the neighborhood structure is (g,h), the second-order potential function as follow

$$U_{con}(C_l(i,j)/C_l(g,h)) = \sum_N \beta \times \delta(C_l(i,j), C_l(g,h)) \tag{3}$$

where β is the adjustable weight coefficient and $\delta(\cdot)$is the Kronecker function.$N(i,j)$ refer to the neighborhood of pixel (i,j). Traditional MRF-based change detection algorithms generally apply 4-neighborhood or 8-neighborhood systems, which assumes that a pixel belongs to the same class as other pixels in its neighborhood. However, this assumption is based on the concept of global consistency and the second-order neighborhood system, which doesn't consider the anisotropy of the region boundary and cannot fully utilize the spatial relationship between pixels. Therefore, we propose an OMRF algorithm. Distinct from MRF, OMRF assumes that pixels in the homogeneous region (object) belong to the same class. For boundary pixels, OMRF does not calculate the contribution of their neighbors to other objects, which effectually minimizes the edge shrinking effect of the object. During the calculation, we use the Mean Shift algorithm to segment DI to obtain objects. Therefore, the second-order potential function is optimized as follows:

$$\begin{aligned} &U_{con}(C_l(i,j)/C_l(g,h)) \\ &= \textstyle\sum_{(g,h)\in SC_k} \beta \times \delta(C_l(i,j), C_l(g,h)) \end{aligned} \tag{4}$$

where SC is the object set obtained by the Mean Shift algorithm and SC_k is the set of all the coordinates in the kth object. OMRF and MRF shared the same solution for minimizing the energy function(as Eq.5), in this paper, the minimization energy function is solved by Iterated Conditional Mode (ICM)[15]

$$\begin{aligned} &U\left(\mathbf{D}, \mathbf{C}_l\right) = \textstyle\sum_{\forall(i,j)} U_{\text{data}}\left(D(i,j)/C_l(i,j)\right) \\ &+U_{\text{con}}\left(C_l(i,j)/\left\{C_l(g,h), (g,h) \in SC_k \cap (i,j) \in SC_k\right\}\right) \end{aligned} \tag{5}$$

2.2. Refined change detection

CD results obtained by OMRF algorithm have reached high accuracy to a certain extent. Nevertheless, there are still a certain number of misclassification points and noise in the coarse change map. Recently, due to the extensive applications of CNN in image processing and its promising performance, we deployed CNN to tune the coarse change map.

2.2.1. Training samples generation

we require to select several reliable samples (the probability of pixels being correctly classified is higher) as the training samples to train IUNet as the coarse CD results obtained by OMRF are not exactly accurate. It is assumed that P_{ij} represents the pixel at (i, j) on the DI, and M_{ij} represents the label of P_{ij} in the OMRF coarse change map. N_{ij}^k represents the label corresponding to the kth neighbor in eight field pixels of the center pixel. There are three types of labels in the coarse change map, as shown in Fig.2. Fig.2(a) denotes to the correctly classified sample points within the changed or unchanged region. Fig.2(b) represents misclassified labels in the changed or unchanged region. Fig.2(c) represents the labels at the changed edge, which is unknown whether these are correctly classified. When selecting reliable samples as the training data, the initial consideration is to avoid the situation as shown in Fig.2(b). Removed samples should meet the following conditions:

$$\begin{aligned} Q_{D_{ij}} &= \sum L_{ij} \geq 7 \\ &\begin{cases} L_{ij} = 1, \text{ when } N_{ij}^k \neq M_{ij} \\ L_{ij} = 0, \text{ when } N_{ij}^k = M_{ij} \end{cases} \end{aligned} \tag{6}$$

where Q represents the number of pixels different from the center pixel label.A centered pixel is considered a misclassified point if its label M_{ij} differs from seven or more labels in that field.

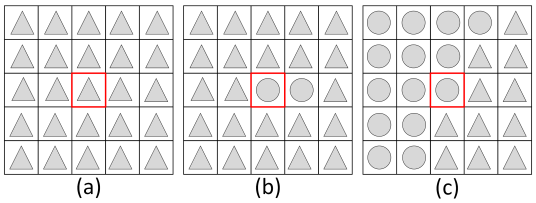

Fig. 2. The situation of label in coarse change map

2.2.2. Network architecture and training

In deep learning, The conditions of the task should be considered initially. From the objective point of view, CD is a binary classification task. From the perspective of network input, CD is a pixel-level classification.Therefore, we design Inception UNet (IUNet) to detect changes on remote sensing images(as shown in Fig.3). UNet was originally used for medical image segmentation[16]. One of its characteristics is the symmetrical structure of the left and right branches,which consists of multiple blocks inside its encoder-decoder structure. The left branch is a contracting path, which is composed of two 3×3 convolutional layers and a 2×2 max-pooling layer (down-sampling) in each block to achieve feature extraction. The right branch comprises two 3×3 transposed convolutional layers and an up-sampling layer using bilinear upsampling in each block to realize the upsampling operation. We add Inception block[17] to UNet, which enables the extraction of multi-scale features, making the network sensitive to change regions of different sizes and effectively improving the performance of the model. In addition, Cross Entropy loss is employed during the entire training.

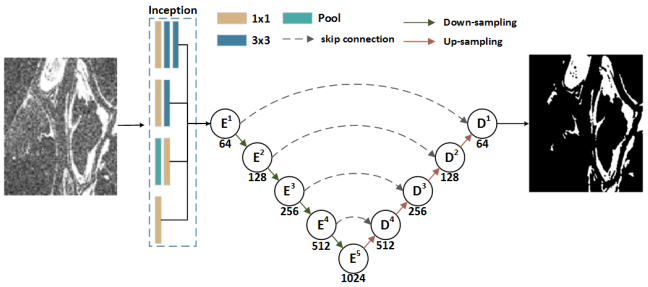

Fig. 3. The network architecture of proposed CNN

2.3. Coarse-to-fine Model

We propose to combine OMRF and IUNet to obtain a coarse-to-fine CD model (as shown in Fig.1), the detailed steps are as follows:

Algorithm 1: coarse-to-fine CD.

Input: Two bi-temporal remote sensing images T1,T2.
Output: Change map of two bi-temporal images.
1:The DI is obtained from the original bi-temporal images.
2:OMRF algorithm was used to obtain a coarse change map.
3:Select suitable labels to mark the samples from the coarse change map.
4:IUNet model was trained with labeled samples.
5:The trained CNN model is used to predict the samples and to generate the final change map.

3. EXPERIMENTS AND RESULTS

3.1. Data and Criterion

Four open bi-temporal remote sensing datasets are adopted to validate the effectiveness of our proposed approach, named Lakemead, Feltwell, Ottawa, and Alaska respectively(partially shown in Fig.4). Lakemead and Alaska are spectral

Table 1. Quantitative results (%) of our method and other benchmark methods on four datasets

method	alaska				feltwell				lakemead				ottawa			
	pre	recall	PCC	kappa	pre	recall	PCC	kappa	pre	recall	PCC	kappa	pre	recall	PCC	kappa
MRF	91.97	99.77	99.39	95.38	82.20	**99.03**	99.40	89.53	90.81	99.20	97.83	93.45	98.49	77.51	96.26	84.61
OMRF	94.88	**99.87**	99.62	97.11	87.40	98.22	99.57	92.27	89.59	98.89	97.47	92.42	87.48	94.81	97.03	89.23
SFCCNN	95.64	98.95	99.62	97.06	91.83	98.59	99.73	94.95	93.07	97.46	98.04	93.98	93.67	96.45	98.41	94.09
FC-Siam	93.45	98.09	99.40	95.39	96.78	97.06	99.83	96.83	85.66	98.71	96.43	89.46	90.48	97.54	97.99	92.67
DSCN	98.38	94.93	99.55	96.38	**97.13**	96.73	**99.84**	**96.84**	**93.33**	97.46	**98.09**	**94.15**	91.71	**97.93**	98.27	93.69
our	**98.76**	97.55	**99.75**	**98.02**	95.96	97.03	99.81	96.39	90.79	**99.31**	97.84	93.49	**98.41**	94.67	**98.92**	**95.86**

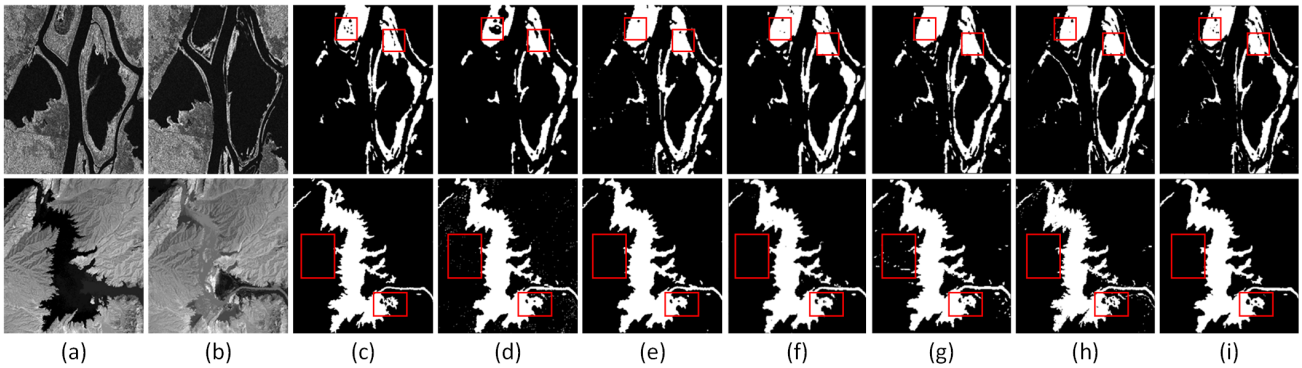

Fig. 4. Qualitative comparison between our method and benchmark SOTA algorithms on ottawa (first row) and lakemead (second row). (a) T1 image. (b) T2 image. (c) Ground truth map. Change maps predicted by means of (d) MRF, (e)OMRF, (f)SFCCNN, (g) FC-Siam, (h)DSCN (i) our.

remote sensing datasets (experiments use their first channel) with sizes of 316×351,475×449, Feltwell and Ottawa are SAR images with sizes of 470×335, 290×335. Evaluation metrics include Percentage Correct Classification (PCC), kappa coefficient (kappa), precision (pre), and recall.

3.2. Implementation

Regarding all compared methods, we adhere to their original papers and tune the parameters continuously until optimal and stable performance. In case of compared algorithms based on supervised learning, training and test samples are randomly selected from the whole dataset, with a proportion of 20% and 80%. All models evaluated are trained with Adam optimizer with an initial learning rate of 5e-3 and the entire model is trained for 100 epochs. The proposed model is based on PyTorch, implemented on the NVIDIA TITAN Xp GPU.

3.3. Performance comparison

Reference methods include OMRF, MRF [6], SFCCNN [14],FC-Siam [13] and DSCN [12], which are described in introduction section. Notably, FC-Siam and DSCN are supervised methods. The results on the four datasets are shown in Table 1, which indicates that the OMRF method has significantly improved the performance of CD compared with the MRF method. In addition, the utilization of IUNet results in a significant effect,with a robust suppression of noise and false detection in the change maps obtained by OMRF. The misclassification in the unchanged area can be corrected by IUNet adjustment as well. Compared with SFCCNN which is based on coarse-to-fine, our method has some improvement in each evaluation index. The qualitative results of the methods are shown in the Fig(4).

4. CONCLUSION

In this paper, we propose an unsupervised learning method based on coarse-to-fine model for CD on remote sensing imagery. The main concept is to generate a coarse change map through OMRF, selecting pixels with high confidence as the labeled samples for training IUNet, Our OMRF algorithm achieves high detection accuracy even as an unpervisied learning method. Further, IUNet is used to realize the refinement of coarse change map, enables the results attain or even exceed the performance of partially supervised deep model detection. Experimental results demonstrate the effectiveness of our approach.

5. REFERENCES

[1] Ashbindu Singh, "Review article digital change detection techniques using remotely-sensed data," *International journal of remote sensing*, vol. 10, no. 6, pp. 989–1003, 1989.

[2] Suming Jin, Limin Yang, Patrick Danielson, Collin Homer, Joyce Fry, and George Xian, "A comprehensive change detection method for updating the national land cover database to circa 2011," *Remote Sensing of Environment*, vol. 132, pp. 159–175, 2013.

[3] Jie Chen, Ziyang Yuan, Jian Peng, Li Chen, Haozhe Huang, Jiawei Zhu, Yu Liu, and Haifeng Li, "Dasnet: Dual attentive fully convolutional siamese networks for change detection in high-resolution satellite images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1194–1206, 2021.

[4] Dominik Brunner, Guido Lemoine, and Lorenzo Bruzzone, "Earthquake damage assessment of buildings using vhr optical and sar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2403–2420, 2010.

[5] Xuan Hou, Yunpeng Bai, Ying Li, Changjing Shang, and Qiang Shen, "High-resolution triplet network with dynamic multiscale feature for change detection on satellite images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 177, pp. 103–115, 2021.

[6] Maoguo Gong, Linzhi Su, Meng Jia, and Weisheng Chen, "Fuzzy clustering with a modified mrf energy function for change detection in synthetic aperture radar images," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 1, pp. 98–109, 2013.

[7] Lorenzo Bruzzone and Diego F Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Transactions on Geoscience and Remote sensing*, vol. 38, no. 3, pp. 1171–1182, 2000.

[8] Csaba Benedek and Tamás Szirányi, "Change detection in optical aerial images by a multilayer conditional mixed markov model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 10, pp. 3416–3430, 2009.

[9] Badri Narayan Subudhi, Francesca Bovolo, Ashish Ghosh, and Lorenzo Bruzzone, "Spatio-contextual fuzzy clustering with markov random field model for change detection in remotely sensed images," *Optics & Laser Technology*, vol. 57, pp. 284–292, 2014.

[10] Chenxiao Zhang, Peng Yue, Deodato Tapete, Liangcun Jiang, Boyi Shangguan, Li Huang, and Guangchao Liu, "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 183–200, 2020.

[11] Hongruixuan Chen, Chen Wu, Bo Du, Liangpei Zhang, and Le Wang, "Change detection in multisource vhr images via deep siamese convolutional multiple-layers recurrent neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 4, pp. 2848–2864, 2020.

[12] Yang Zhan, Kun Fu, Menglong Yan, Xian Sun, Hongqi Wang, and Xiaosong Qiu, "Change detection based on deep siamese convolutional network for optical aerial images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1845–1849, 2017.

[13] Rodrigo Caye Daudt, Bertr Le Saux, and Alexandre Boulch, "Fully convolutional siamese networks for change detection," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 4063–4067.

[14] Yangyang Li, Cheng Peng, Yanqiao Chen, Licheng Jiao, Linhao Zhou, and Ronghua Shang, "A deep learning method for change detection in synthetic aperture radar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5751–5763, 2019.

[15] Julian Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 48, no. 3, pp. 259–279, 1986.

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[17] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.