FUSION DETECTION OF CLOSED WATER IN MEDIUM-LOW RESOLUTION REMOTE SENSING IMAGERY

Yuanyong Ning, Yanan You, Jingyi Cao, Fang Liu, Qing Yan, Yue Zhang

School of Artificial Intelligence, Beijing University of Posts and Telecommunications

IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-0369-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IGARSS47720.2021.9553554

ABSTRACT

Aiming at the closed water detection in remote sensing imagery at medium-low resolution, this paper proposes a novel method for closed water detection based on fusion detection which conducts detection via informative fused images blended by Synthetic Aperture Radar (SAR) and optical images. Firstly, it utilizes SAR and optical image pairs containing the same closed water object to generate aligned image pairs according to latitude and longitude information. Next, generative adversarial network (GAN) is adopted to fuse two categories of images. At last, a target detection network driven by optical image samples is used to detect the closed water on the fused image. The experiment result on Sentinel-1&2 shows that the proposed method can effectively make up for the shortage of SAR image in closed water detection and improve the detection performance.

***Index Terms*—**Closed water detection, Image fusion, GAN

1. INTRODUCTION

Object detection in remote sensing image is a process to detect and locate the object with a certain class from the static image. With the development of deep learning, especially convolutional neural network (CNN) which is proved as an excellent architecture to extract and represent image features, many object detection methods have been developed based on CNN [1,2,3]. Optical image has spectral information that can be intuitively felt by human beings, which provides the possibility for researches to mark a reasonably large-scale dataset for the training of target detection network. However, it may suffer from the bad weather and season changes. Conversely, SAR with capacity to image all-day under most meteorological conditions [4], but it has certain limitations in the expression ability of ground objects. In particular, for medium-low resolution SAR images, targets are difficult to be labeled, resulting in unavailability in target detection task. The fusion of optical and SAR images can colorize the SAR images and provides an effective data source for fusion detection, and the application of fused images in the field of target detection has crucial scientific significance [5]. There are several literatures treating methods for object detection based on fusion method. [6] blended the infrared band and thermal infrared band of multi-spectral data, then the fusion results are applied to low-resolution ship detection, which greatly reduces false alarm rate.

In this paper, a new method for closed water detection based on fusion detection algorithm is presented and analyzed on Sentinel-1&2 experimental data sets. First of all, we propose an image registration methodology based on latitude and longitude information for the wide swath SAR and optical images. After that, pix2pixHD [7] network was used for image fusion. At last, we utilize Faster R-CNN framework, driven by optical image samples, to detect the closed water on the fused images. The result of experiment shows that the proposed algorithm can effectively detect the closed water targets.

2. THE PROPOSED ALGORITHM

The fusion detection algorithm is summarized in the given Fig.1. The following sections describe in detail the major steps of the algorithm.

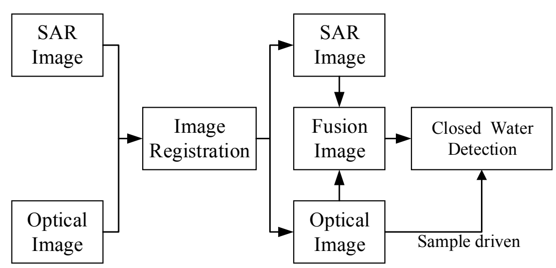


Fig.1. Flowchart of the proposed algorithm.

2.1. Image Registration

Before fusing optical and SAR images, image registration should be done with a higher priority. Image registration is a process of matching and superimposing images acquired at different time, different sensors, different viewing angles and different imaging conditions. Assuming that there is a reference image, $I_{SAR}(x, y)$, and a map to be registered, $I_{opt}(x, y)$, the transformation relationship between the two images can be simply expressed in the following equation.

$$I_{opt}(x, y) = I_{SAR}(f(x, y)) \tag{1}$$

Where f denotes the coordinate transformation of a two-dimensional space. At present, the feature-based image registration method is the most widely used method in remote sensing image registration. This method is not affected by the gray level, and the image is matched by extracting point feature, line feature or surface feature, with strong robustness and good matching effect. Considering the difference of imaging mechanism between optical image and SAR image, the same target on the optical image and SAR image may present a distinct feature description. Besides, owning to the specificity in image resolution (more than 10000), it is difficult to directly compute feature descriptors with the traditional methods like scale-invariant feature transform (SIFT). Therefore, this paper proposes a remote sensing imagery registration method based on latitude and longitude information. The process of multi-source remote sensing image registration algorithm is illustrated in the underneath figure.

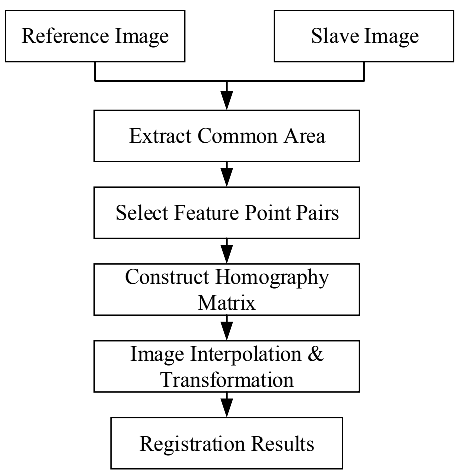


Fig. 2. Flowchart of image registration algorithm.

The Sentinel-1 satellite is equipped with a C-band SAR sensor with a resolution of 5 meters. The Sentinel-2 satellite carries a multi-spectral imager (MSI) that covers 13 spectral bands. The ground resolution of the bands is generally 10 meters. Therefore, Sentinel-1&2 images can be used as an alternative to detect closed water at medium-low resolution.

Geographic coordinate system and projection coordinate system were used in the image production of Sentinel-1&2. Based on the criterion that the same target has the same latitude and longitude information in different images, the coordinate transformation was used to convert geographic coordinates and projection coordinates into image coordinates. In addition, due to the different shooting angles of remote sensing images, the imaging content will undergo projection deformation, therefore, we construct the homography matrix according to the feature points to adapt to this change. Assuming that the coordinate before image transformation is (x, y), and after transformation is (x', y'), the projection transformation is shown in the following equation:

$$\begin{bmatrix} x_s' \\ y_s' \\ s \end{bmatrix} = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{2}$$

Where $x_s' = x'/s$, $y_s' = y'/s$, s is the scaling factor and the 3×3 matrix constituted by a_{ij} $(0 \le i, j \le 2)$ is the homography matrix. With transformation matrix available, the registration results can be obtained by setting the image interpolation mode and executing the image transformation. The registration results are shown as Fig.3.

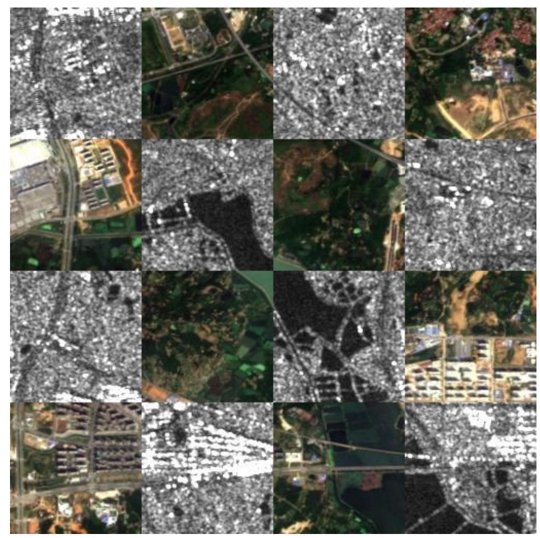


Fig. 3. The registration results based on latitude and longitude information.

2.2. Image Fusion

Recently, the rapid development of generative adversarial network makes image-to-image translation results more realistic. We choose pix2pixHD to fuse SAR image and optical image in this paper as it synthesizes images with more natural textures and details. Pix2pixHD framework uses a coarse-to-fine generator, a multi-scale discriminator architecture, and a robust adversarial objective function.

The pix2pixHD generator is composed of global generator, G_1, and local enhancer network, G_2, and the generator's structure is depicted in Fig.4. Both generative networks include a convolutional front-end, a set of residual blocks, a transposed convolutional back-end. In this paper, the global generator network processes at a resolution of 256×256, and the local enhancer network produces an image with a resolution that is $4\times$ the output size of the previous one ($2\times$ along each image dimension), so the final synthesized image resolution via the generator $G = \{G_1, G_2\}$ is 512×512.

Pix2pixHD uses a multi-scale discriminator, which means that 3 discriminators have an identical network structure but operate at different image scales. The

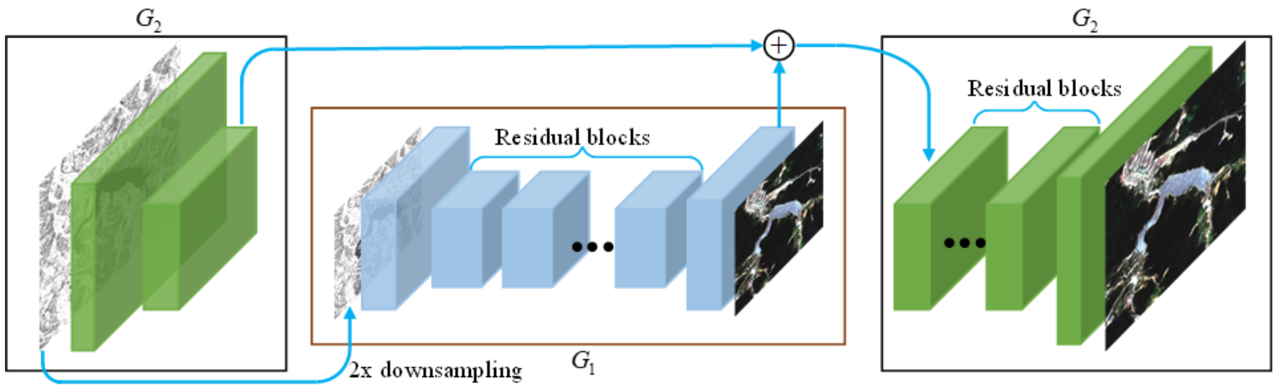


Fig. 4. The Generator structure of pix2pixHD

author downsamples the real and synthesized high-resolution images by a factor of 2 and 4 to create an image pyramid of 3 scales. Then the discriminators D_1, D_2 and D_3 were trained to distinguish between real and synthesized images at 3 different scales, respectively. The full objective used in pix2pixHD framework combines both GAN loss and feature matching loss as:

$$\min_G((\max_{D_1,D_2,D_3} \sum_{k=1,2,3} L_{GAN}(G,D_k) + \lambda \sum_{k=1,2,3} L_{FM}(G,D_k))) \tag{3}$$

Where λ is weight and L_{FM} is the feature matching loss.

2.3. Closed water detection

Faster R-CNN network is a typical deep learning target detection model. Compared with previous networks, Faster R-CNN first unified region proposal network and classification into the same network framework, thus achieving a significant improvement in detection speed and accuracy. The later deep learning network, such as YOLO, SSD and other networks, basically followed the core idea of Faster R-CNN. Although the improved model improved in detection speed, almost no model could significantly outperform the Faster R-CNN in detection accuracy. Therefore, this paper utilizes Faster R-CNN to detect the closed water.

3. EXPERIMENTS AND RESULTS

In experiments, we apply Sentinel-1&2 dataset constructed in this paper to test the performance of the proposed fusion detection algorithm. After image registration, we annotate the closed water on the optical images, and map the annotation results to the SAR image, then we have two different kinds of closed water detection data sets. The experimental dataset consists of 3677 aligned optical and SAR image pairs with a size of 512×512 pixels. The same Faster R-CNN network structure is used in SAR image and optical image detection. The algorithm's performance is measured by detection rate (DR) and false alarm rate (FAR). DR is the ratio of the correct number of closed water detected by algorithm to the sum of real closed water number and FAR is the ratio of false alarm number to the sum of detection boxes. In all experiments, we set the weight λ=10 in image fusion, the confidence threshold is 0.8 and the intersection over union (IOU) threshold is 0.5 in the closed water detection. The results of the experiments conducted are shown in Table 1.

Table 1. Detection results in different datasets.

Category	Real targets	DR	FAR
SAR	155	53.55%	17.82%
Optical	155	87.44%	9.34%
Fusion	155	89.68%	10.32%

From the Table 1, we can see that the optical image is more suitable for the detection of closed water than the SAR image. This is because the target of SAR image is vulnerable to speckle noise. Generally, the fusion image has the same detection performance as the optical image in closed water detection task.

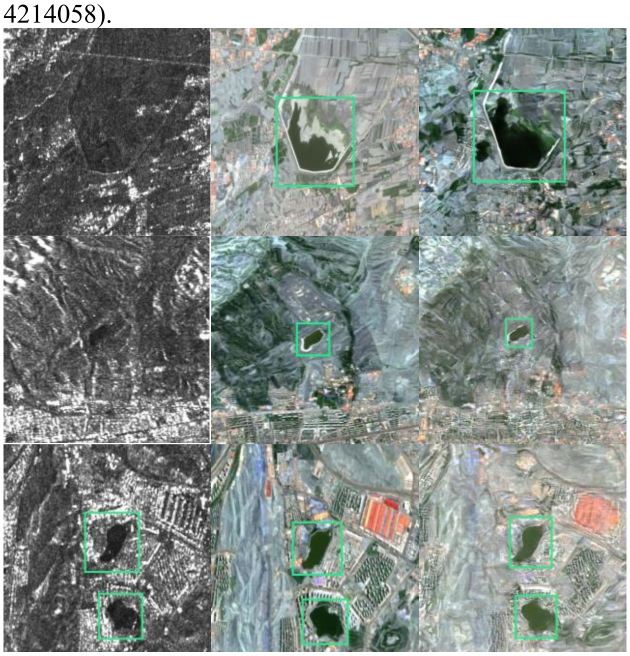
Fig.5 shows some of the detection results. As we can see, with the same confidence, more targets are not detected in SAR images compared with optical and fused images. In addition, Faster R-CNN has nearly the same detection performance in optical and fusion images.

4. CONCLUSION

In this paper, we present a fusion detection method for closed water target on Sentinel-1&2 data. The experiment shows that the method can effectively detect closed water in fusion images. It proved that the obtained fused image combines the features of optical image and SAR image, and the object detection network driven by optical images can be used to closed water detection for fused image, which is an important supplement for SAR image target detection. In future work, we will extend this method to detect more targets, especially those that are difficult to detect in medium-low resolution SAR images.

5. ACKNOWLEDGMENT

Sentinel-1&2 data used in this research was downloaded from https://scihub.copernicus.eu. This work is Supported by Beijing Natural Science Foundation, China (Grant No. 4214058).



(a) SAR (b) Optical (c) Fusion

Fig. 5. An example of detection results.

6. REFERENCES

[1] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." IEEE transactions on pattern analysis and machine intelligence 39.6 (2016): 1137-1149.

[2] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "YOLOv4: Optimal Speed and Accuracy of Object Detection." arXiv preprint arXiv:2004.10934 (2020).

[3] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.

[4] Franceschetti, Giorgio, and Riccardo Lanari. Synthetic aperture radar processing. CRC press, 1999.

[5] Irwin, Katherine, et al. "Fusion of SAR, optical imagery and airborne LiDAR for surface water detection." Remote Sensing 9.9 (2017): 890.

[6] Liu, Yong, et al. "Fusion detection of ship targets in low resolution multi-spectral images." 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2016.

[7] Wang, Ting-Chun, et al. "High-resolution image synthesis and semantic manipulation with conditional gans." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.