Accurate detection method of moving target based on sequential SAR images with multiple azimuth squint angles

Shiqi Yang¹, *, Yang Liu², Peili Xi², Chunsheng Li¹, Wei Yang¹, Hongcheng Zeng¹
¹School of Electronics and Information Engineering, BeiHang University, Beijing, China
²Shanghai Institute of Satellite Engineering, Shanghai 201109, Shanghai, China

*Corresponding author e-mail: qianyx612@nbt.edu.cn

Abstract. In this paper, a novel moving target detection method for sequential Synthetic Aperture Radar (SAR) images with different azimuth-squint angles is proposed. In sequential SAR images, due to the movement of the target, the imaging position of moving targets among different frames differs. The method proposed in this paper uses this kind of motion characteristics to achieve the detection of moving targets in multi-frame SAR images. This algorithm can be divided into two parts: block-level detection and pixel-level detection. Block-level detection is achieved by stacked denoising autoencoders to extract the high-dimensional features of the moving target. Pixel-level detection consists of Local Binary Similarity Patterns (LBSP) coding as well as grayscale background subtraction. Pixel-level detection only needs to consider the pixels of foreground image pieces which contain moving targets. This method can not only increase the detection speed, but also suppress the false alarm problem caused by clutter. Experiments are carried out for verifying the validation of the method and the comparison are made between the proposed method and the traditional Constant False Alarm Rate (CFAR) algorithm.

1. Introduction
With the development of azimuth multi-angle SAR imaging mode, SAR system has the ability of obtain a series of sequential images in a short time interval, in which moving targets such as ships will appear in different positions [1].

Currently, the traditional Constant False Alarm Rate (CFAR) algorithm mainly calculates the threshold based on the false alarm rate, and separates the ship target from the background clutter in sliding window to achieve the purpose of pixel-level detection. This method is to consider each pixel in the sliding window of the scene image, and then determine whether the pixel belongs to the target. As a result, CFAR algorithm often brings high time cost when processing sequential SAR images, and it is difficult to make full use of the related information between multiple frames of images. The SAR target detection algorithms based on deep learning include Faster-RCNN [3] algorithm, YOLO [4] algorithm, SSD[5] method, etc. The detection result of the deep learning method is the anchor frame information of the target in the image, without pixel-level detection. Also, it is difficult for both CFAR and deep learning methods to use the target’s motion information between multiple SAR image sequences.
In this paper, a novel moving target detection method for sequential SAR images with different azimuth-squint angles is proposed. This paper is organized as follows. Section 2 introduces the complete detection process as well as the detecting mechanism. Section 3 addresses the simulation data and processing results. Conclusions are drawn in Section 4.

2. Method
This paper proposes a moving target detection method based on the SDAE-LBSP framework. The method is designed to detect moving targets in sequential SAR images with multiple azimuth-squint angles in short time interval. The method can be divided into block-level detection and pixel-level detection. In the period of block-level detection, it will extract high-dimensional features of the sequential SAR images through the deep learning method so as to select the foreground image pieces which contain targets. For the detected foreground image pieces, grayscale and local binary similarity pattern (LBSP) are used to determine whether the pixels in the foreground image pieces belong to the background or the target. Figure 1 shows the complete detecting process.

2.1. Block level detection
Block level detection is based on deep learning methods, i.e., stacked denoising autoencoders (SDAE). During detection, the sequential graph is divided into small image pieces, and the cosine distance of the high-dimensional features extracted from both the image piece of the sequential graph and the image piece corresponding to the background frame is calculated, so as to determine whether the image piece in the sequential graph belongs to the background or contains moving targets.

2.1.1. Stacked Denoising Autoencoders. Deep learning methods can be divided into supervised learning and unsupervised learning. In general, supervised methods require labels corresponding to input signals. Label production often requires a lot of manpower and time costs. As an unsupervised neural network model, SDAE training does not need to build labels.

Autoencoders are often used for dimensionality reduction or feature learning [6]. The trained autoencoders are able to copy the input to the output. As is shown in Figure 2, the autoencoder network can be divided into two parts: an encoder that extracts features \( h_i = f(W_{1}h_{i-1} + b) \) and a decoder that generates reconstruction \( y_i = f(W'_{1}h_i + b') \). In general, the autoencoder needs to add some constraints so that it does not simply learn to copy the input to the output. The denoised autoencoder adds noise to the network input, and it is trained to restore noisy input to noise-free signals. The method of adding noise is to randomly discard several neurons. Then the input signals with noise are encoded and decoded in order. By minimizing the reconstruction error, the distance between the decoded result and the original signal without noise is minimized.

The stacked denoising autoencoder is composed of multiple denoising autoencoders. The output of the previous encoder is the input of the next encoder. Each denoising autoencoder is trained separately.
The outer encoder is trained first, and then the coding results of the previous encoder are used to train the next encoder. After the training is completed, each encoder is extracted to form a feedforward neural network, which is used to abstract high-dimensional features of sequential SAR image pieces. Figure 3 shows the structure of 5-layer stacked denoising autoencoders.

2.1.2. Training and detecting. The network structure proposed in this paper consists of 5 encoders. The sizes of encoders are 1024, 2560, 1024, 512, 256 respectively. During training, a 28-size sampling window is first used to sample one single image. Then we expand the samples into 784-dimensional vectors, and send 100 samples as a batch to the network for training. The training data used in this experiment comes from 5192 real ship images generated by the SSDD data set after data expansion, each with a resolution of 280×280.

The establishment of the background model is based on the following assumption: For an image sequence, the pixels in each frame belong to the background for the most of the time. Therefore, the establishment of the background frame is through the median method, that is to say, the grayscale of each pixel in the SAR image sequences is sorted in different frames, and the median value among the sequences is picked out to form a new image as the background frame. The generated background frame $B$ and sequential image $I$ are divided into small image pieces respectively. The trained SDAE feedforward network is able to calculate the eigenvectors of image pieces at the corresponding positions in the background frame $B$ and the original sequential image $I$. Then the cosine similarity is calculated to get the distance between the two image pieces, which is used to determine whether the image piece in the sequential image $I$ contains a moving target. The greater the cosine similarity, the more similar the two image slices, i.e., the image pieces in the sequential image $I$ is more likely not to contain target. The cosine similarity is calculated as follows:

$$\cos \theta = \frac{c_i \cdot c_j}{\|c_i\| \|c_j\|}$$

2.2. Pixel level detection

Due to the influence of the clutter in the scene, if only the gray value is used to detect the pixels of moving targets, it may cause false alarms as well as missed detection. Therefore, the gray background subtraction result is combined with the LBSP. LBSP coding takes into account local similarity and has the ability to suppress noise. Therefore, it can be used as a supplement to the gray value background subtraction method to make the target detection result more precise[7].

The result of LBSP encoding of pixels is a 16-bit binary sequence. Its core idea is to compare the similarity between the central pixel and the surrounding 16 adjacent pixels. Figure 4 shows the 16-bit LBSP encoding template. If the adjacent pixel and the central pixel have high similarity, the binary value of the corresponding position of the encoding result is set to 1. In the detection, the LBSP codes of the
foreground image slices in the sequential image and the corresponding image slices in the background model are respectively calculated. The Hamming distance between the binary codes of the above image slices determines whether the pixel belongs to the moving target.

\[
\text{Figure 4. Template of 16-bit LBSP encoding.}
\]

The calculation formula of LBSP code can be written as:

\[
d_p(v_p - v_s) = \begin{cases} 
1, & |v_p - v_s| \leq T \\
0, & \text{else}
\end{cases}
\]  

\[
\text{LBSP}(x, y) = \sum_{p=0}^{15} d_p(v_p - v_s) \cdot 2^p
\]  

In the detection, the LBSP encoding results of the foreground image pieces in the sequential image and the corresponding image pieces in the background model are calculated respectively. As is shown in (4), according to the Hamming distance between the two binary codes, it is judged whether the pixel belongs to the moving target.

\[
\text{FG}_{\text{LBSP}}(x, y) = \begin{cases} 
1, & D(S_p(x, y), S_b(x, y)) > T_{\text{LBSP}} \\
0, & D(S_p(x, y), S_b(x, y)) \leq T_{\text{LBSP}}
\end{cases}
\]  

The final pixel-level detection is the result of LBSP encoding and grayscale background subtraction. Based on the above block-level detection, pixel-level detection only needs to compare the pixels of foreground image pieces in both sequential frame and background frame. This processing can not only increase the detection speed, but also suppress the false alarm problem caused by clutter.

3. Experiment

In order to verify the effectiveness of the proposed algorithm, experiments are carried out in this section. The experiments consist of detecting moving ship targets in sequential SAR images with different azimuth squint angles and evaluating the detection results.

\[
\begin{array}{ccc}
\text{(a) original image} & \text{(b) block-level detection} & \text{(c) pixel-level detection} \\
\text{(frame 2)} & \text{(frame 2)} & \text{(frame 2)} \\
\text{(d) original image} & \text{(e) block-level detection} & \text{(f) pixel-level detection} \\
\text{(frame 12)} & \text{(frame 12)} & \text{(frame 12)} \\
\end{array}
\]  

\[
\text{Figure 5. Original images and detecting results.}
\]
The moving target echo and clutter simulation are separately carried out and superimposed frame by frame according to the multiple imaging parameters. The data acquisition mode for each frame is the stripe mode and CS imaging algorithm is applied to generate images from the echo.

The specific process of the moving ship detection method based on the SDAE-LBSP framework is as follows: First of all, the high-dimensional abstract features of the image slices are extracted through the trained SDAE feedforward neural network, which is used to select the foreground image pieces containing the target. In addition, Combining the gray value background subtraction method and local binary similarity pattern method in order to determine whether the pixels in the foreground image piece belong to the background or the target, so as to complete the pixel-level target detection.

The processing result of block detection based on SDAE is shown in Figure 5. In the scene image, if the ship target appears at the edge of the image block to be detected, it is likely to cause the missed detection of the image block, causing the pixel-level detection result to appear jagged edges, and resulting in the loss of detailed information in the detection result. Therefore, to alleviate this problem, the detected foreground image block is expanded in four directions: up, down, left, and right. Then the expanded image pieces are used for pixel-level detection. This processing step can avoid jagged edges to a certain extent and increase the detection accuracy but it means that the number of candidate pixels for fine detection increases, and it takes more time to detect targets. It can be seen from the experimental results that after the features are extracted by the stacked denoising autoencoder, the foreground image pieces containing targets can be screened out. The block detection results are used for LBSP for pixel-level detection, and the binary image of the ship target can be obtained finally.

Figure 6 shows the pixel-level detection results after the expansion of the coarse detection image. It can be seen that the detection results basically retain the contour information of the moving ship.

| Numble | Frame 1 | Frame 2 | Frame 3 | Frame 4 | Frame 5 |
|--------|---------|---------|---------|---------|---------|
| Sp     | 0.99698 | 0.997200| 0.995534| 0.997037| 0.996738|
| Re     | 0.803674| 0.765576| 0.812230| 0.770093| 0.782137|
| FPR    | 0.00302 | 0.002800| 0.004466| 0.002963| 0.003262|
| FNR    | 0.00399 | 0.005241| 0.00373 | 0.004781| 0.004689|
| PWC    | 0.006874| 0.007877| 0.008034| 0.007595| 0.007790|
| Precision | 0.844051| 0.859683| 0.783069| 0.844089| 0.837885|
| F-Measure | 0.823368| 0.809905| 0.797383| 0.805395| 0.809052|
In the part of evaluating detection result, binary image obtained by the CFAR detection algorithm is used as the true value. The numbers of pixels of True Positive, True Negative, False Positive and False Negative in the detection result are calculated. The detection effect is evaluated by seven indicators: Re (Recall), Sp (Specificity), FPR (False Positive Rate), FNR (False Negative Rate), PWC (Percentage of Wrong Classification), Precision and F-Measure[8]. The evaluation results are shown in Table 1.

It can be seen from the table that the overall accuracy of the detection is relatively high, and the error rate stays low. The detection effect is relatively stable among the multiple sequential images. Since only the pixels in the foreground image are counted, compared with the CFAR algorithm, the detection results in this paper effectively suppress false alarms caused by clutter.

4. Conclusion

In this paper, a novel moving target detection method for sequential SAR images with different azimuth-squint angles is proposed. Through the trained SDAE, the high-dimensional features of the moving target can be extracted, so as to select the image pieces where the target is located in the scene. Moreover, the influence of clutter and noise is depressed by LBSP coding and gray-scale background subtraction, and the target pixel-level detection is realized. Detecting results of the simulation data and the its comparison with traditional CFAR algorithm validate the effectiveness of the proposed method.

Acknowledgments

This work was supported by Shanghai Aerospace Science and Technology Innovation Fund SAST2019-038.

References

[1] A. Damini, B. Balaji, C. Parry, et al. A VideoSAR mode for the X-band wideband experimental airborne radar[C]. Algorithms for Synthetic Aperture Radar Imagery XVII, Orlando, 2010, 76990E.
[2] J. Mittermayer, S. Wollstadt, P. Prats, et al. Bidirectional SAR imaging mode[J]. IEEE Transaction on Geoscience and Remote Sensing, 2013, 51(1):601-614.
[3] S. Q. Ren, K. M. He, R. Girshick, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
[4] J. Redmon, A. Farhadi. YOLOv3: an incremental improvement[EB/OL]. https://arxiv.org/abs/1804.02767.
[5] W. Liu, D. Anguelov, D. Erhan, et al. SSD: single shot MultiBox detector.[EB/OL]. https://arxiv.org/abs/1512.02325.
[6] P. Vincent, H. Larochelle, I. Lajoie, et al. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion[J]. Journal of Machine Learning Research, 2010, 11, 3371-3408.
[7] G. A. Bilodeau, J. P. Jodoin, N. Saunier. Change Detection in Feature Space Using Local Binary Similarity Patterns[C]. IEEE International Conference on Computer and Robot Vision, 2013: 106-112.
[8] Y. Wang, P. M. Jodoin, F. Porikli, et al. CDnet 2014: An Expanded Change Detection Benchmark Dataset[C]. IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014:387-394.