IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium | 978-1-7281-6374-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/IGARSS39084.2020.9323826

A HIGH RESOLUTION SAR SHIP SAMPLE DATABASE AND SHIP TYPE CLASSIFICATION

BAO Meng[1,2], MENG Jun min[1,2,], ZHANG Xi[1,2], LIU Gen wang[1,2]*

1. First Institute of Oceanography, Ministry of Natural Resources, Qingdao 266061, China
2. Ocean Telemetry Technology Innovation Center, Ministry of Natural Resources, Qingdao 266061, China

ABSTRACT

As the improving of the synthetic aperture radar (SAR) resolution and the increase in the amount of data acquisition, the ship type recognition has become an important research topic. In order to meet the precise identification for ship types, 101 SAR data and the Automatic Identification System (AIS) were used to build a SAR ship database. The database contains 5288 ship samples with different polarizations, incidence angle and resolutions, including more than 20 kinds of ship type such as cargo, container, oil tankers, and fishing boats. Furthermore, the influence of different polarization, incidence angle and heading on ship geometry parameters was analyzed. Moreover, a random forest (RF) classifier was used to carry out the ship type recognition experiment, and the classification accuracy reached more than 60%.

Index Terms**—**High resolution synthetic aperture radar (SAR), ship type, feature analysis, type classification

1. INTRODUCTION

Synthetic aperture radar (SAR) is of great significance in maritime surveillance. With the increasing number of ships around the world and the increasing threats at sea in recent years, ship detection can no longer meet the current application needs [1-2]. With the improvement of SAR resolution, vessel type identification has become an important research topic.

The classification of ships depends on a large number of ship samples. In order to achieve better identification result based on SAR, several SAR ship databases have been built in the past years. Chen et al. constructed a total of 250 ship samples including cargo ships, containers and oil tankers by using the TerraSAR-X data, and carried out classification applications [3]. Huang et al. constructed a data set named openSARship which contains 11346 ships by using 41 Sentinel-1 images and AIS, and it has certain potential in ship classification and Sentinel-1 evaluation [4]. Ma et al. built a data sets named MTCD by using high-resolution GF-3 image to conducted maritime target type identification, this database contains a total of 3200 samples in 8 categories, including boat, cage, cargo, container, tower platform, tanker and windmill [5]. Wang et al. constructed a ship database with 102 GF-3 and 108 Sentinel-1 SAR images, which contained 43819 ship chips and was mainly used for deep learning detection of ships [6]. These databases have their own characteristics and play important roles in SAR target detection and classification. However, the ship type labels for most of the databases are determined by visual interpretation, and lack of ground truth verification. Parameters such as length and width cannot be provided and can only be measured in SAR images. Otherwise, for databases with AIS information verification, such as openSARship, the resolution is not high enough.

In this paper, a high-resolution ship sample database with real information is built by SAR image and AIS data, which covers a variety of sensors (such as TerraSAR-X, RADARSAT-2, GF-3 etc.) with different polarization mode, resolution and incidence angle. Each sample in the database has real AIS information, and all ship attributes during SAR imaging are recorded. Furthermore, the geometric parameters of ships in the data set are fully analyzed, and the preliminary application of ship type identification is carried out, which provides a reference for the selection of ship classification algorithm and features.

2. DESCRIPTION OF SAR SHIP SAMPLE DATABASE

SAR and AIS are needed to construct the sample database. After obtaining SAR data, it is necessary to query the corresponding AIS data. The AIS information within 30 minutes was obtained taking the SAR imaging time was as the center as shown in Fig 1. Considering that the time of the two is not completely coincident and the positioning of the ship may be deviated, the interpolation processing of AIS is carried out to unify its time and space position with the SAR. Linear interpolation and extrapolation methods were usually used according to the AIS information time. Finally, the interpolated AIS data was superimposed on the SAR satellite image, and the ships in the SAR were labeled with AIS and their information was extracted. The process is shown in Figure 2.

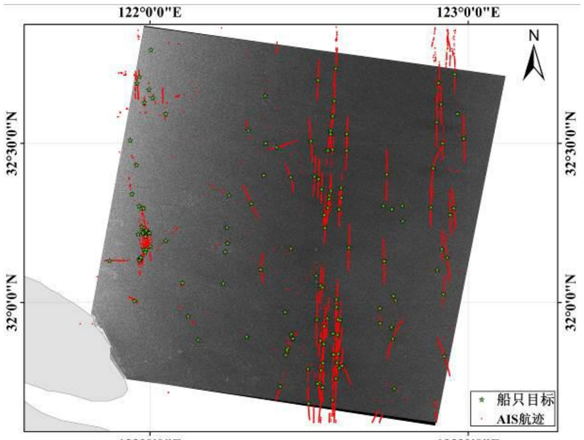

Fig.1 Schematic diagram of the superposition of AIS ship track and SAR satellite

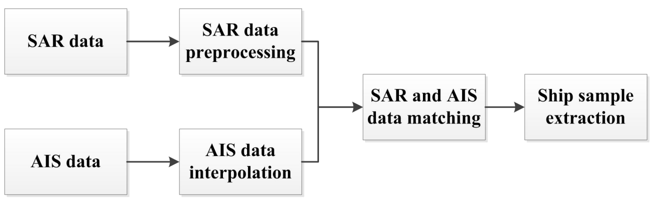

Fig.2 flowchart of SAR ship data set construction

A total of 101 SAR images were acquired in this paper, in which are 75 C-band images and 26 X-band images. The image covers four polarization and various resolutions, and the details of the data are shown in Table 1. All data have been geometric corrected and radiometric calibrated.

Table1 SAR data

Satellite	Polarization	Resolution （m）	Incidence angle (°)	Band	Number of data
RADARSAT-2	HH/VH/VV	3/5/25/50	22.1-43.3	C	27
GF-3	AHH	3/5/8/10/25	15.2-50.3		44
ASA	VV	150	20-34.7		1
Sentinel-1A	VH/VV	25	35.6-45.8		3
TerraSAR-X	HH/HV/VV	3/8	29.6-45.5	X	7
GNX	-	3	23.6-46.1		13
UAVSAR	AHH	10	-		6

A standard ship sample mainly contains three files: SAR image chip, attribute file and ship optical photo (Fig. 3). Among them, the SAR image (format is. tiff) is a geometrically corrected image. The properties file (format is .xlm) stores the original SAR image information such as imaging time, incident angle, polarization method, resolution, and ship attribute information such as Maritime Mobile Service Identify (MMSI), latitude, longitude, ship length, ship width, speed, heading, type, etc. The optical photo format is .jpg.

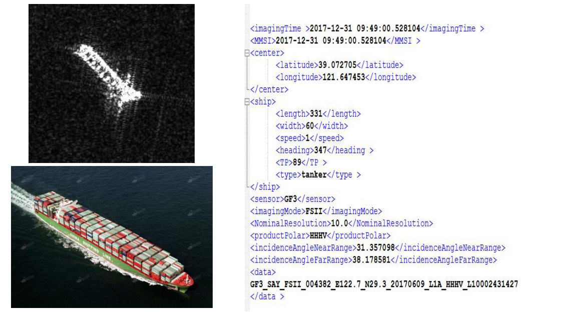

Fig.3 sample example

In all 5288 samples, there are 4047 ship samples with resolution better than 10 m, about 77% of the total samples. The dataset contains more than 20 types of ships such as cargo ships, tankers, container ships and fishing ships. Fig. 4 presents the statistics of the SAR ship dataset

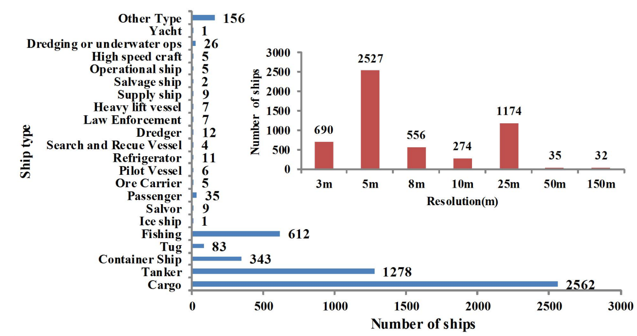

Fig.4 SAR ship sample database statistics

3. ANALYSIS OF SHIP GEOMETRIC CHARACTERISTICS AND TYPE CLASSIFICATION

3.1. Analysis of Ship Geometric Parameters

In this paper the difference between the ship length extracted from SAR and the real ship length is analyzed. The ship length from SAR data is extracted by visual interpretation. The real ship length is from the AIS record. In order to evaluate the accuracy of the ship length estimation, we define a coefficient: the relative ship length error ε,

$$\varepsilon = |L_{AIS} - L_{SAR}|/L_{AIS}\ .$$

Where L_{AIS} is the ship length from AIS record, and L_{SAR} is the ship length estimated from SAR data.

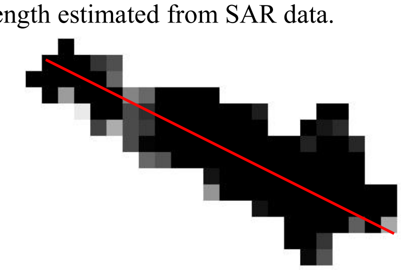

Fig.5 Schematic diagram of ship measurement length extraction

3.1.1. Polarization

Fig. 6 shows the relationship between L_{AIS} and L_{SAR} in HH, VH/HV, and VV polarization, respectively. Among all three polarizations, VV polarization has the best ship length detection accuracy, and the mean value of relative ship length error ε for all high-resolution samples is 0.20. So the ship length extracted by the VV polarization is more consistent with the real ship length. In contrast, the worst ship length detection accuracy is by HH polarization. The mean value of ε is 0.34.

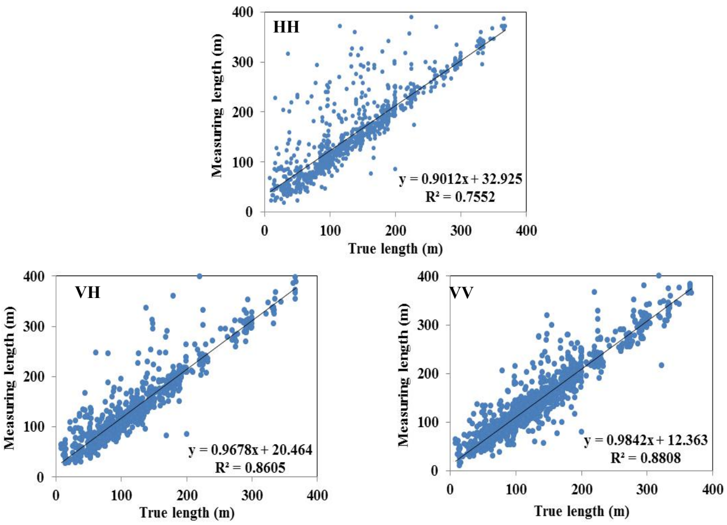

Fig.6 Comparison of the true length and measured length in different polarization

3.1.2. Incidence angle

Fig. 7 shows the mean value μ and standard deviation σ of relative ship length error ε under the different incident angle. of the HH, VH, and VV polarization modes. The three colors (blue, red and green) in the Fig. 7 represent HH, VH, and VV respectively.

We divide the ship database into four intervals by incidence angles. Fig. 7 reveals that the ship length extraction capability of SAR is not significantly affected by the variation of incidence angle. Meanwhile, we can find the standard deviation of relative ship length error ε is higher in the range of incidence angle of 30°-40° in cross-polarization. This may be caused by due to the VH polarization being susceptible to noise and sea state.

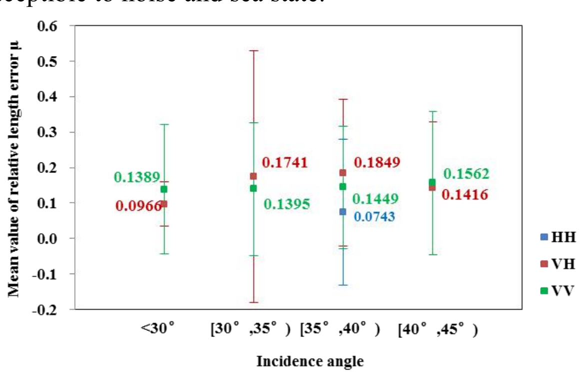

Fig.7 Relative length error statistics in different Incidence angle

3.2. Type identification performance analysis

The purpose of building a SAR ship sample database is to better serve the ship type recognition. Therefore, this section uses the ship sample database preliminary evaluates the performance of ship type recognition.

Here 1548 high-resolution ship samples are selected. The selected ship types are cargo ship, container and oil tanker. The number of these three ship types are 848, 260 and 440, respectively. Random forest (RF) classifier is used to carry out the ship type recognition experiment.

Feature characteristics are the basis and key for ship classification and recognition. Geometric feature, scattering feature and polarization information are the most three importance features for ship target classification in SAR image [7]. In this paper five kinds of feature information are extracted, including geometry, radar intensity, texture and scattering features. They are described as follows:

$f_1 = \{G_i\}_{i=1}^{8}$ is the geometry features of ship sample, including length, width, area, the ratio between length to width, perimeter, shape complexity and centroid.

$f_2 = \{I_i\}_{i=1}^{3}$ is the radar intensity features of ship sample, including mean value, standard deviation, and the ratio between standard deviation and mean value.

$f_3 = \{H_i\}_{i=1}^{7}$ is the Hu moment invariants features, this feature has the invariability of rotation, translation and scale [8].

$f_4 = \{T_i\}_{i=1}$ is the texture features, expressing the complexity and irregularity of objects.

$f_5 = \{S_i\}_{i=1}^{8}$ is the ship target scattering features, including scattering center energy, number of scattering centers, and the scattering density characteristics of bow, middle and stern, left, main and right sides of the ship as described in [9]. Scattering density characteristics can be calculated by follow

$$\delta_i = \left(M_i/A_i\right) / \max_{1\le i\le p}\left(M_i/A_i\right); \quad (1)$$

$$\delta_j = \left(M_j/A_j\right) / \max_{1\le j\le q}\left(M_j/A_j\right)。 \quad (2)$$

Where M_i and M_j are the sum of the backscatter intensities of the section elements perpendicular and parallel to the principal axis respectively. A_i and A_j are unit area perpendicular and parallel to principal axis. p and q are the number of units in the direction of vertical and parallel principal axes.

Based on the above these feature types, a total of 31 feature combinations are generated, as shown in Table 2

Tab.2 Ship features

feature	combination	feature	combination	feature	combination
F_1	$\{f_1\}$	F_{12}	$\{f_2, f_5\}$	F_{23}	$\{f_2, f_3, f_5\}$
F_2	$\{f_2\}$	F_{13}	$\{f_3, f_4\}$	F_{24}	$\{f_2, f_4, f_5\}$

F_3	$\{f_3\}$	F_{14}	$\{f_3, f_5\}$	F_{25}	$\{f_3, f_4, f_5\}$
F_4	$\{f_4\}$	F_{15}	$\{f_4, f_5\}$	F_{26}	$\{f_1, f_2, f_3, f_4\}$
F_5	$\{f_5\}$	F_{16}	$\{f_1, f_2, f_3\}$	F_{27}	$\{f_1, f_2, f_3, f_5\}$
F_6	$\{f_1, f_2\}$	F_{17}	$\{f_1, f_2, f_4\}$	F_{28}	$\{f_1, f_2, f_4, f_5\}$
F_7	$\{f_1, f_3\}$	F_{18}	$\{f_1, f_2, f_5\}$	F_{29}	$\{f_1, f_3, f_4, f_5\}$
F_8	$\{f_1, f_4\}$	F_{19}	$\{f_1, f_3, f_4\}$	F_{30}	$\{f_2, f_3, f_4, f_5\}$
F_9	$\{f_1, f_5\}$	F_{20}	$\{f_1, f_3, f_5\}$	F_{31}	$\{f_1, f_2, f_3, f_4, f_5\}$
F_{10}	$\{f_2, f_3\}$	F_{21}	$\{f_1, f_4, f_5\}$		
F_{11}	$\{f_2, f_4\}$	F_{22}	$\{f_2, f_3, f_4\}$		

70% of ship samples for classifier training and 30% for testing. The number of trees in the RF classifier is set to 150. Fig. 8 shows the classification results of the ship samples. From the figure, we can find that the best classification accuracy achieved by using the combined features $F_{24}=\{f_2, f_4, f_5\}$. The accuracy reaches 61.85%, and the classification accuracy was better than the single-class features and other feature combinations.

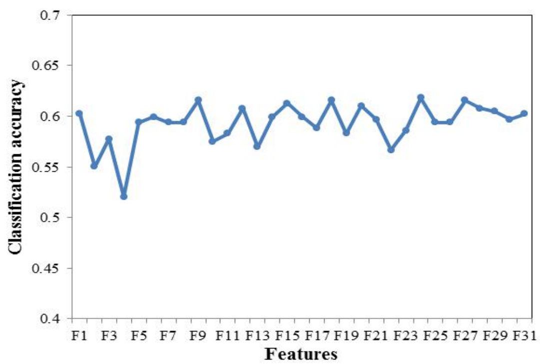

Fig.8 Ship classification results (range from 0 to1)

4. CONCLUSION

This article uses 101 scene SAR images and AIS data to create a ship samples database. The database contains 5288 SAR ship samples in which 4047 (about 77%) ship samples with 10 m higher resolution, and includes more than 20 ship types. Based on the analysis of the ship's geometric parameters, the results show that the relative error of the ship length extracted by VV polarization is the smallest and closer to the true ship length which recoded by AIS. For vessel type recognition, based on the random forest (RF) classifier, 31 types of ship characteristics are combined to compare the accuracy of marine vessel type recognition. The results show that the random forest (RF) classifier with $F_{24}=\{f_2, f_4, f_5\}$ feature combines has the best classification accuracy 61.85%.

Our research team will continue to expand the number of ship sample databases in the future. The database introduced in this article will be officially released on the day of the IEEE IGARSS 2020 conference.

ACKNOWLEDGEMENT: This research was supported by the National Key R&D Program of China (No. 2017YFC1405204), the National Natural Science Foundation of China (No. 61971455)

REFERENCES

[1] Velotto D, Bentes C, Tings B and Lehner S, "First Comparison of Sentinel-1 and TerraSAR-X Data in the Framework of Maritime Targets Detection: South Italy Case," IEEE J. Oceanic Eng., 41(4): 993-1006, Jan. 2016.

[2] Schwegmann C P, Kleynhans W and Salmon B P, "Synthetic Aperture Radar Ship Detection Using Haar-Like Features," IEEE Geosci. Remote Sens. Lett, 14(2): 154-158, Nov. 2017.

[3] Chen W T, Ji K f., Xing X. W, and Zou H. X., "Ship recognition in high resolution SAR imagery based on feature selection," in Proc. CVRS, pp. 301–305, 2012.

[4] Huang L Q, Liu B, Li B Y and Guo W W, "OpenSARShip: A Dataset Dedicated to Sentinel-1 Ship Interpretation," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 2017, 11(1): 195 - 208.

[5] Ma M Y, Chen J, Liu W and Yang W, "Ship Classification and Detection Based on CNN Using GF-3 SAR Images," Remote Sens., vol. 10, no. 2043, Dec. 2018.

[6] Wang Y, Wang C and Zhang H, "Automatic Ship Detection Based on RetinaNet Using Multi-Resolution Gaofen-3 Imagery," Remote Sens., 11(5): 1-14, 2019.

[7] Lang H T and Wu S W, "Ship Classification in Moderate-Resolution SAR Image by Naive Geometric Features-Combined Multiple Kernel Learning," IEEE Geosci. Remote Sens. Lett., 14(10): 1765-1769, Sep. 2017.

[8] Chen W T, Ji K F, Xing X W Zou H X and Sun H, "Ship Recognition in High Resolution SAR Imagery Based on Feature Selection," International Conference on Computer Vision in Remote Sensing, Xiamen, China, Dec. 2012.

[9] Margarit G and Tabasco A, "Ship Classification in Single-Pol SAR Images Based on Fuzzy Logic," IEEE Geosci. Remote Sens., 49(8): 3129-3138, Aug. 2011.