A ship target dynamic recognition method based on image recognition and VR technology

Peng Fei¹,a, Xiumin Chu¹,b, Xiongfei Geng²,c, Jianan Luo²*  
¹School of Transport and Logistics Engineering, Wuhan University of Technology, Wuhan, China  
²China Waterborne Transport Research Institute, Beijing, China  
¹abelfei@hotmail.com  
°chuxm@whut.edu.cn  
³gengxf@wti.ac.cn  
*Corresponding author: luojianan@wti.ac.cn

Abstract—Effective identification of ship identity is an important means of water traffic management. Aiming at the technical limitations of AIS, radar and other means, a ship automatic recognition method based on machine learning is proposed, and a water traffic visualization system integrating multi-source recognition technology is realized based on VR technology. The test shows that the method and system can solve the technical problems of ship identification caused by data easy to be tampered with, difficult to distinguish massive targets and large obstacles, and provide an effective method for navigation support in key waters.

1. INTRODUCTION
With the rapid development of social economy and the deepening of water transportation, the number of ships is increasing, and the types are diversified. There are ship types such as oil tanker, bulk carrier, container ship and fishing boat, which makes the action coordination of ships difficult. At the same time, the ship is developing towards large-scale and high-speed, which not only increases the operation difficulty of the ship itself, but also puts forward higher requirements for the channel. It is necessary to provide a good traffic environment, special measures and services for its safe navigation. In addition, the increasing number of dangerous goods on board and the restriction of the controlled river section have led to the increase of navigation pressure of inland waterway, traffic congestion and hidden dangers of water traffic accidents, threatening shipping safety and river ecological environment, and restricting the normal operation of social economy [1]. Therefore, the intelligent tracking and monitoring of inland waterway, especially the navigation ships in the control river section, will help to improve the navigation command decision-making and improve the navigation status of the control river section.

Water traffic dynamic monitoring can find and coordinate maritime traffic targets in time. It is an important technical means to improve the efficiency and safety of ships entering and leaving the port, berthing and berthing, and navigation in the port channel. At present, VTS, AIS, radar, real-time monitoring video, GNSS positioning terminal, binding of mobile app and ship location, ship report, remote identification (RFID) are the main technical means of dynamic supervision of water traffic [2-3].

However, inland waterways face more technical obstacles, such as narrow waterways, many intersections, high ship density, many water navigation obstructions, large changes in channel water depth,
complex navigation environment, long and wide navigation area, etc. Therefore, it is difficult to apply image recognition technology to inland ship tracking and monitoring. To sum up, the existing data fusion methods can not fuse the image with ship positioning, resulting in the decline of ship tracking accuracy at the port and poor navigation efficiency in the port channel.

This study uses modern information collection and AR monitoring technology to monitor ship traffic conditions, rectify ship traffic order and assist ship navigation. Through these measures, the establishment of good traffic order and the reduction of ship collision, grounding, reef and other maritime accidents are of great significance to improve the navigation capacity of inland waterways, reduce ship traffic accidents, reduce pollution and improve the information level. [4] At the same time, it helps to strengthen the law enforcement of waterways and improve the efficiency of waterway supervision. The research on port digital service technology can promote the application of land traffic tracking and monitoring and other video tracking and monitoring fields.

2. SHIP TARGET LOCATION AND RECOGNITION

2.1. Ship target detection in HSV color space

The detection of ship targets through HSV color space is different from frame difference detection method. Frame difference method belongs to continuous multi frame dynamic target detection, while HSV color space detection belongs to single frame static target detection [5]. Because the color texture of water is obviously different from the ship target, the hue, saturation and lightness of water in HSV color space image are different from the ship target.

HSV space ship target detection principle: through research and analysis, most of the upper part of the hull of many ship targets are brightly colored, and the s component is obviously different from that of water, so the s component is an important means to detect the hull; However, the lower part of the hull is in contact with the water surface for a long time. Generally, the color is gray and the lightness value is relatively low, especially for small cargo ships. Therefore, there is a certain difference between the V component of the lower part of the hull and the water, so it is more effective to detect the lower part of the hull; Then combine s and V component detection to get a more complete target. In addition, due to the shadow of shore objects or ship hull affected by light, the effect of target detection will also be affected. Through the monotonous hue H component of shadow and fixed within a certain range, the influence of shadow can be removed by finding out the relevant range value through experiments.

HSV space ship target detection process: since the measurement scene is in the water area, the ship targets on the water shall be separated. In order to eliminate the influence of the shore, a detection area shall be set in the water area. Since the water area in the detection area is generally larger than the ship target area, the maximum value of the statistical histogram obtained for the s and V components in the detection area is the water background reference value t of the component T(s) and T(v), and then detect the s and V components at the virtual coil position of the current frame detection area, which is consistent with the reference values T(s) and T(v). If the absolute value of the difference exceeds a certain threshold, the pixel is judged as 1, otherwise it is 0; then the absolute value diagrams of the two components are calculated or fused together to judge that pixel 1 is the target and pixel 0 is the water background; finally, the shadow is removed through the h-component value of the virtual coil position. When the h-component value of the pixel with target 1 is within the shadow fixed value range, the pixel is reset to 0.
2.2. Virtual coil target capture

Several problems to be solved in ship target data acquisition and capture: First, capture when the ship target appears in the best position in the camera. Second, ensure that the capture of ship targets in the image is complete and the screen ratio meets the standard requirements. Third, the captured ship target shall be marked without repeated capture.

2.2.1. Setting of virtual detection coil and detection area: on the video image of the fixed gun, the virtual coil is set in the center of the water area, perpendicular to the navigation direction of the ship, and is composed of three rows of coil blocks side by side. Each coil block is adjacent to 5 to 8 coil blocks. The detection area is set to facilitate the use of HSV space ship target detection method. The detection area is set in the water surface area where ships other than the shore often appear.

2.2.2. Virtual coil block detection: each coil block, as an independent calculation unit, calculates whether the change of each pixel of the coil block exceeds the set threshold through the frame difference method or HSV color space detection method, and then counts the ratio of the number of pixels exceeding the threshold of the whole coil block to the total pixels of the coil block. If it is greater than 50%, it is judged that the target is detected by the virtual coil block and marked as 1. If it is less than or equal to 50%, it is marked as 0.

2.2.3. Target detection judgment: the rule for judging whether the virtual coil detects the target, and neighborhood retrieval is carried out for each coil block marked 1. If the number of neighborhood coil blocks marked 1 of a coil block marked 1 is greater than or equal to 3, it is judged that the virtual coil detects the ship target. If the number of adjacent coil blocks marked 1 of all coil blocks marked 1 is less than 3, it is judged that no ship target is detected.

2.2.4. Start capture: when the ship target triggers the virtual coil, the capture machine will capture the complete target through focusing, moving and other operations at the virtual coil position.

2.2.5. Waiting for the capture to be completed: waiting for the capture to be completed is a process of constantly reading video images. When the ship target is found to start the capture, the ship target will continue to enter the virtual coil until the stern leaves the virtual coil. This process is a process of waiting for the capture to be completed. When the ship continues to enter the virtual coil until it leaves, the virtual coil block will continue to detect the ship target until there is no ship target, indicating that it is waiting for the capture to be completed.
2.3. Ship location and segmentation
The core task of ship target location and segmentation is to obtain the position information and contour information of all ship targets in a given image, so as to accurately separate the image of each ship target from the original image. The practical approach is to detect all ship targets in the image and frame the object with a rectangular frame. This is also a basic task in computer vision - target detection. This is a basic core module for us to build the whole system. Only by accurately framing all ship targets can we build an efficient ship identification and monitoring system. Target detection algorithms are roughly divided into traditional algorithms and algorithms based on deep learning.

Another category of target detection algorithms based on deep learning is non-region-based recommendation. The representative of this kind of algorithm is SSD algorithm [6]. The design idea of this kind of algorithm is end-to-end. The generation of candidate regions, feature extraction, location offset prediction and category prediction are encapsulated in the same network, which greatly improves the overall performance of the algorithm. SSD algorithm is one of the detection networks with the best real-time performance and the highest detection accuracy. Therefore, the construction of this system will be based on SSD algorithm.

In the training phase, the SSD algorithm framework is composed of input, feature extraction, default box generation, matching, prediction network and loss function, as shown in Figure 3 below. It is introduced as follows:

2.3.1. Input: the input information required by the SSD algorithm framework includes the training picture and the position information of all targets in the picture (i.e. the coordinates of the real frame) and category information (i.e. category number);

2.3.2. Feature extraction: it is composed of vgg-net (visual geometry group) and other neural network architectures with excellent performance to extract picture features and output feature maps extracted from different convolution layers. The selected convolution layer also becomes the feature extraction layer;

2.3.3. Default box generation and matching: according to the preset parameters, a certain number of default boxes with different aspect ratios are generated on different feature extraction layers for target box selection. After the default box is generated, it will be matched with the real box in the input information according to a certain matching strategy, and the matching result information will be sent to the prediction network;

2.3.4. The prediction network is composed of a group of convolution layers. After synthesizing the characteristic map and matching information, the target category prediction and position offset prediction are given;

2.3.5. The loss function layer will evaluate the gap between the predicted value and the real value, and the optimization of the loss function will adjust the weight of the network in back propagation, so that
the network can better detect the target and fit the location information and category information of the target.

Figure 3. Framework and flow diagram of SSD algorithm training stage

SSD algorithm regards location prediction and category prediction as regression problems at the same time, so the loss function can combine location loss and category loss. Therefore, the form of loss function is multitasking. The loss function is an important driving force of neural network, because its optimization determines how the network learns the content. Therefore, SSD network adopts "smooth" L1 loss function is used as position loss function, and softmax loss function is used as confidence loss function, and then through a coefficient factor α Combine the two. The mathematical expression of the loss function in SSD algorithm is shown in formula (1).

$$L(x, c, l, g) = \frac{1}{N}(L_{conf}(x, c) + \alpha L_{loc}(x, l, g))$$

(1)

Where n represents the number of matched candidate boxes. If n is equal to 0, is consideredloc=0. α It is a weight coefficient used to balance the losses of the two parts and avoid excessive proportion of the losses of one part. The expression of position loss function is shown in formulas (2) to (6).

$$L_{loc}(x, l, g) = \sum_{i \in Pos} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k \text{smooth}_{L1}(l_i^m - \hat{g}_{ij}^m)$$

(2)

$$\hat{g}_{ij}^{cx} = \frac{(g_{ij}^{cx} - d_{ij}^{cx})}{d_{ij}^{w}}$$

(3)

$$\hat{g}_{ij}^{cy} = \frac{(g_{ij}^{cy} - d_{ij}^{cy})}{d_{ij}^{w}}$$

(4)

$$\hat{g}_{ij}^{w} = \log\left(\frac{g_{ij}^{w}}{d_{ij}^{w}}\right)$$

(5)

$$\hat{g}_{ij}^{h} = \log\left(\frac{g_{ij}^{h}}{d_{ij}^{h}}\right)$$

(6)

Where, $x_{ij}^k$ indicates the matching condition. When it is 1, it means that the $i_{th}$ default box matches the $j_{th}$ real box, and the real category is k, otherwise it is 0. The position loss function adopts the position regression function smooth L1 loss, where $l$ represents the predicted coordinate offset, $\hat{g}_{ij}$ represents the coordinate offset corresponding to the real box on the matching of the current default box. $d_i$ stands for the default box. The mathematical expression of the confidence loss function is as follows:

$$L_{conf}(x, c) = -\sum_{i \in Pos} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i^0)$$

(7)

where $\hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)}$

(8)

Where, $C$ represents confidence, and the meaning of other variables is the same as above.
3. SHIP FEATURE EXTRACTION AND MATCHING METHOD

3.1. Ship feature extraction
In this study, ORC (Oriented FAST and Rotated BRIEF) method is used to extract ship propagation features. ORB is a fast feature point extraction and description algorithm. The following are the steps of ORB feature extraction and corresponding feature matching [7].

ORB uses fast (features from accelerated segment test) algorithm to detect feature points. Taking the point to be detected as the center, the gray value is compared with the gray value of 16 adjacent pixel points. When compared with other pixels, if the gray difference is greater than a set threshold, it is considered that the point to be detected is different from the adjacent point. If 12 consecutive points are different from the point to be detected, the point to be detected is considered as a corner point, that is, a feature point.

In order to avoid taking too much time to calculate most non feature points, this method first calculates whether the gray values of 1, 5, 9 and 13 points are the same. Because if the point to be detected is a feature point, at least three of the four points are the same. Therefore, by comparing the above four points, the detection speed of feature points can be accelerated.

ORB uses BREF algorithm to calculate the descriptor of feature points. The core idea of BREF algorithm is to select n point pairs around feature point P with specific rules and combine the comparison results of these n point pairs as descriptors. Find the point pair around the feature points and obtain the 0 or 1 description of the point pair by comparing the gray value of the point pair. The same method is used for other point pairs to finally generate orb descriptors.

Specifically, the feature description area is framed with the feature point as the center and a certain length as the radius. Select 512 pairs of points and number them, P1 (a, b)...PN (a, b)... P512 (a, b). The descriptor of feature points is obtained by combining the calculated attribute results of multiple points. The characteristic descriptor of orb can be considered as a string of 0 and 1.

Since the descriptor of ORB feature points is a string of 0 and 1 binary strings, the matching of feature points can be transformed into string matching, which greatly saves storage and matching time. The matching of orb features is actually to calculate the Hamming distance between two feature vectors, that is, the number of different characters in the same position. It is defined as follows:

$$d(f^1, f^2) = \sum_{i=1}^{d} \text{XOR}(f^1_i, f^2_i)$$  \hspace{1cm} (9)

3.2. Ship feature matching
Performing the above processing on each ship target feature point to obtain a feature point set of an effective ship target image. Using FLANN matching algorithm, the first feature point set kp1 and the second feature point set kp2 are extracted from any two effective ship target images respectively [8]; According to the first feature point set kp1 and the second feature point set kp2, the corresponding first feature description set des1 and the second feature description set des2 are obtained by using the SIFT feature extraction method or surf feature extraction method. Then using the feature matcher FLANN, the first feature description set des1 and the second feature set des2 are matched to obtain multiple sets of feature matching point pairs, and the Euclidean distance between each set of feature matching point pairs is calculated.

$$A = (x_1, x_2, ..., x_n)$$
$$B = (y_1, y_2, ..., y_n)$$

$$AB = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$  \hspace{1cm} (10)

Where, A and B are two different feature matching points in a set of feature matching point pairs, $x_n$ and $y_n$ is the coordinate of the feature point corresponding to a feature matching point and the coordinate of the feature point corresponding to the B feature matching point in the nth group of feature matching point pairs, and $ab$ is the Euclidean distance.
Select a set of feature matching point pairs with the smallest distance as the minimum distance and record it as the standard value. A set of feature matching point pairs with a minimum distance of less than 3 times are used as the matching degree. If the matching degree is less than the preset matching degree threshold, the matching feature corresponding to the matching degree is stored in the ship database as the nearest matching feature for subsequent search. If the matching degree is greater than or equal to the preset matching degree threshold, the matching degree is used for feature matching in the ship database to obtain the nearest neighbor matching feature. According to the nearest matching feature, obtain the corresponding electronic chart in the ship database; Using the ship automatic identification system, retrieve the identity of the target ship, determine the identity information of the target ship and calibrate it, compare the location information of the ship automatic identification system, obtain the AIS location information of the target ship, and then obtain the ship data;

Among them, the identity information of the target ship includes ship call sign, mobile identification code, course, speed, Captain, ship width, draft, port of destination, cargo type and sailing time. The ship data includes AIS data of each target ship and a set of feature matching point pairs corresponding to the nearest matching feature. Further, we coded the above method and applied it to the Lancang River AR [9-10] ship monitoring system in Yunnan Province, China, as shown in Figure 4.

Figure 4. Demonstration of ship monitoring system in Yunnan Province

4. CONCLUSION
This study proposes an automatic ship identification method based on machine learning, which can automatically capture and identify ship objects through video system, and quickly identify ship identity and location by integrating ship data such as AIS and VTS. By further integrating AR application scenarios, the corresponding ship digital information can be superimposed on the real visual scene, and the information beyond vision can be superimposed to enrich the information perception of the channel. In the application scenario of ship dynamic monitoring in key waters, the use of AR technology can make managers more simply and intuitively observe the navigation environment information of the ship and the ship navigation in the port.

ACKNOWLEDGMENT
This work was supported by Green Intelligent Inland Ship Innovation Programme.

REFERENCES
[1] XU Wuxiong, CHU Xiumin, LEI Jinyu, Parallel-Control-Based Navigation Risk Management System for Busy Inland Waters, 3rd ed., vol. 2. Oxford: Clarendon, 2018, 41(01), pp. 64-69.
[2] Gangyan Xu, Chun Hsien Chen, Fan Li, Xuan Qiu. AIS data analytics for adaptive rotating shift in vessel traffic service[J]. Industrial Management & Data Systems, 2020, 120(4).
[3] Sun Yang, Peng Xinya, Ding Zexuan, Zhao Jiansen. An Approach to Ship Behavior Prediction Based on AIS and RNN Optimization Model[J]. International Journal of Transportation Engineering and Technology, 2020, 6(1).
[4] Bai X, Chen Z, Zhang Y, et al. Infrared ship target segmentation based on spatial information improved FCM[J]. IEEE transactions on cybernetics, 2015, 46(12): 3259-3271.
[5] Huang Z K, Liu D H. Segmentation of color image using EM algorithm in HSV color space[C]//2007 International Conference on Information Acquisition. IEEE, 2007: 316-319.

[6] Chen Q, Huang N, Zhou J, et al. An SSD Algorithm Based on Vehicle Counting Method[C]//2018 37th Chinese Control Conference (CCC). IEEE, 2018: 7673-7677.

[7] Shu C W, Xiao X Z. ORB-oriented mismatching feature points elimination[C]//2018 IEEE International Conference on Progress in Informatics and Computing (PIC). IEEE, 2018: 246-249.

[8] Gupta A, Chaudhary D K, Choudhury T. Stock prediction using functional link artificial neural network (FLANN)[C]//2017 3rd international conference on Computational Intelligence and Networks (CINE). IEEE, 2017: 10-16.

[9] Lvov M S, Popova H V. Simulation technologies of virtual reality usage in the training of future ship navigators[J]. 2019.

[10] Wu H, Yang J, Chen C, et al. Research of Virtual Ship Fire-fighting Training System Based on Virtual Reality Technique[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2019, 677(4): 042100.