Analysis and innovation prospect of intelligent ship visual tracking and recognition system

Shiguan Yu¹, Hongshuo Zhang², *, Shilong Mu³, Shizhong Liu¹, Haojie Ding¹

¹ College of transportation, Shanghai Maritime University, Shanghai, China
² Merchant Shipping College, Shanghai Maritime University, Shanghai, China
³ School of information and control engineering, China University of mining and technology, Xuzhou, China

*Corresponding author e-mail: 201810111164@stu.shmtu.edu.cn

Abstract. In recent years, due to the rise of the concept of intelligent ship and the development of the intelligent ship, the ship intelligent has become a global shipping trend of The Times, with the rapid development of maritime shipping industry, sea lanes and port ship intensity increasing, shipping safety risk increasingly improve, intelligent traffic regulation becomes increasingly important task, shipping information acquisition is more accurate. Ship tracking and identification system is not only the core of shipping safety supervision system, but also has an effective role in the collection of ship navigation information and intelligent supervision of navigation traffic. Computer vision is a technology developed rapidly in recent years, as the current research hotspot, with its access to large amount of information, and the advantages of fast response speed, strong anti-jamming capability compared with traditional of AIS and radar easily affected by obstacles, affected by the operator subjective behavior of communication system device, it can well realize the target ship recognition and tracking, to the modern channel intelligent regulation of ship sailing information collection, transportation, ship navigation safety are very important significance. This article mainly from the principle of development background significance, system characteristics, the visual technology, tracking algorithm for visual identification and tracking system to do a detailed description, on the basis of AIS and radar, based on traditional visual algorithm combined with some data and AIS system on personal improvement ideas on the application of the system innovation, and finally puts forward personal has been developed out of a visual tracking innovative device of the ship.

Keywords: Computer vision technology; Automatic identification; Visual orientation; Visual tracking.

1. Introduction
Nowadays, with the rapid development of economic globalization, the marine shipping industry is increasingly busy, and the number of all kinds of activities related to the sea is increasing rapidly. However, the increase in the number of ships has resulted in the situation of higher water traffic density. [1] Therefore, a comprehensive and effective intelligent ship tracking and identification system is
essential in order to achieve the purposes of safe operation of ships, effective channel information collection and accurate intelligent traffic monitoring. At present, the main monitoring means are automatic identification system (AIS) and shore based radar. However, there are some problems in AIS information, such as subjective error information and so on. [2] There are some problems in radar target, such as the lack of target information and low accuracy. In addition, the weather conditions, sea clutter and other obstacles can have a great impact on it, the subjective behavior of the operator is also crucial to the efficiency of tracking and recognition, and even there is a problem that it can not be used well in some dangerous situations.

With the development of technology, in view of the above problems of AIS and shore based radar, we are constantly looking for effective ways to improve. For example, in view of the lack of communication information, a large number of video cameras are arranged along the coast, [3] ports and rivers to play an auxiliary role in ship monitoring. However, due to the limitation of the installation location, orientation and distance of the marine video camera, the video quality is affected by it, and the utilization rate of video surveillance is generally low. In addition, it mainly depends on human visual observation. In view of the above problems, computer vision technology has a good solution, it can achieve most of the target information acquisition, and the perception of information, processing speed is fast, coupled with the strong resistance to interference obstacles, so the effect is good. The core of intelligent ship visual tracking and recognition system is computer vision technology. [4] On the basis of its application, different machine algorithms are used to effectively extract the ship's navigation trajectory, so as to achieve the high resolution of image features, and ultimately achieve the accuracy of ship target tracking and recognition, the stability of efficient detection, which is of great significance for ensuring the safe navigation at sea and improving the operation efficiency of maritime traffic It is of great significance to carry out maritime law enforcement, [5] efficient and intelligent ship monitoring and other maritime tasks.

2. Features of the system

2.1. Research background of Shipborne intelligent vision technology

Computer vision technology is a computer science rising in recent years. It refers to the machine vision technology that uses cameras and computers to recognize, track and measure targets instead of human eyes. It is a technology that simulates the imaging process of human vision through computers.[6-7] It can further process the images according to the collected information, and achieve accurate information detection, face recognition and target recognition It has many advantages, such as non-contact, fast access to a large amount of information, fast detection speed, strong resistance to external factors, etc. it can realize tracking and positioning at the same time, so it has been widely used in the field of industrial production.[8]

Visual tracking algorithm involves a wide range of applications. With the continuous development of its application, there are many different kinds of algorithms. According to different angles, visual tracking algorithms can be classified in different ways. [9] According to the number of tracking targets, it can be divided into single target tracking and multi-target tracking; according to the composition of tracking targets, it can be divided into rigid target tracking and non rigid target tracking; according to the certainty of target positioning output, it can be divided into deterministic target tracking and probabilistic target tracking. [10] The classification of different tracking algorithms is described in detail in Figure.1.
In the 21st century, with the rapid development of artificial intelligence and big data technology, computer vision application technology continues to mature and expand, intelligent navigation visual perception analysis has become an important research field. Under the background of intelligent development, the rapid development of image processing, pattern recognition, machine learning, computer vision, artificial intelligence and other fields of visual tracking technology promotes the development of vision based moving object analysis technology. Jiang Jingjing et al. [11] proposed a ship detection method based on improved vibe (visual background extractor) algorithm, Based on the traditional vibe algorithm, the improved Canny operator is fused, and the smooth filtering and adaptive threshold segmentation strategy are adopted to improve the accuracy of ship contour extraction. Li Shuangshuang [12] proposed to combine the codebook model background detection method with the improved visual saliency detection algorithm MSS (maximum symmetric surround) to complete the ship target detection. However, the traditional scheme has the problems of low recognition rate and poor real-time performance, so a fast and accurate ship target detection scheme is urgently needed.

At present, deep learning algorithm has a good effect in computer vision target recognition. Ross gershick et al. [13] proposed to combine region proposals with convolutional neural networks (CNNs), use large convolutional neural networks from bottom to top in candidate regions, and use the method of supervised pre training and task specific optimization to improve detection performance. On this basis, Ross gershick et al. [14-15] proposed a fast region based convolutional network method fast r-cnn and a faster region based convolutional network method fast r-cnn. Fast r-cnn designed a multi task loss function to unify the classification task and border regression into one framework. Fast r-cnn puts forward the regional proposal network (RPN) network replacement.Selective search algorithm makes use of neural network to generate candidate regions, so that neural network can learn more high-level, semantic, abstract features, and greatly improve the reliability of candidate regions. Liu et al. [16] proposed the single shot multibox detector (SSD) detection method, combined with fast r-cnn and regression method, in the final level of convolution, used anchor method to extract the candidate frames of the feature map on each scale, and judged the target type and position according to the candidate frames obtained by anchor on different scales.

Although the intelligent ship technology has just developed in China, it is not strong enough. At present, the "made in China 2025" has clearly put forward and started the top-level design and research work of intelligent ship, and gradually increase the capital and human investment in this area, in order to reach the advanced level in this field. What remains unchanged in the future is that the shipping market will expand with the economic globalization. In this regard, more powerful measures are needed to ensure maritime safety, further technology is needed to ensure maritime traffic efficiency, and more effective and accurate supervision is needed for maritime law enforcement. Therefore, the application of visual tracking technology is of great significance in the future era of intelligent navigation, And it
brings many development points for dynamic monitoring based on visual perception, which will effectively realize the combination of human and technology for further development. In this paper, based on the above research background of ship vision technology, based on the analysis of computer vision, combined with the development prospect of ship intelligence, through the experimental simulation analysis in different environments, combined with deep learning algorithm and binocular camera ranging method, innovative ship intelligent sensing technology is proposed. It is of great significance to achieve accurate identification and ranging between intelligent navigation ships in the future, and then help safe navigation and good traffic management.

2.2. The principle of ship identification and tracking
In marine intelligent traffic monitoring, ship tracking and recognition based on video is one of the most important tasks. Efficient ship tracking algorithm and recognition algorithm can not only effectively track the ships in the region of interest in different application scenarios, but also identify a variety of ship types. The traditional processing methods are based on the combination of AIS system and radar. [17] Now through the improvement of STMs algorithm, Kalman filter, particle filter, mean shift and other algorithms, some results have been achieved, but there is no good combination of tracking and recognition, so it can not be well applied to the Maritime Intelligent Traffic management. In this paper, a real-time ship tracking and recognition algorithm is constructed based on Darknet network and YOLOv3 algorithm to achieve the balance of speed and accuracy. In the proposed ship tracking and recognition model, the classification and recognition of ship feature extraction network (Darknet network) and YOLO3 algorithm are the key links to realize ship tracking and recognition. Therefore, this paper focuses on the feature extraction network and the learning process of YOLOv3 algorithm. The chart of tracking and identification is shown in Fig. 2.

2.3. Ship feature extraction network
The innovative idea of this paper is based on the deep learning framework of Darknet network structure for feature extraction learning. The connection idea of residual network is adopted to ensure that the network can converge even though the network structure is very deep. With the deepening of the network structure, each layer can extract different feature expression, generate more feature combination learning, get the object feature graph, and initialize the subsequent detection and tracking model. Using full convolution structure and small convolution kernel operation, the calculation of parameters is greatly reduced, and the speed and accuracy are improved. Through the convolution layer operation of the network structure, the output data of the previous deep network layer is studied to extract highly abstract ship features. From the ship feature map output from the convolution layer, the deep-seated features with semantic information in the ship feature map are learned and extracted, and the learning results of the secondary feature map are ignored. Therefore, the maximum pooling algorithm is used to pool the ship map output from the convolution layer. In order to improve the speed of feature learning, the activity of local neurons is introduced to create a competition mechanism, and the LRN layer is added. The larger the response ratio, the larger the pair value, and inhibit other neurons with smaller feedback, which enhance the generalization ability of the model.
2.4. **Yolov3 structure in Darnet-53**

YOLO series is an excellent network model in the field of target detection, just like SSD, r-cnn, faster, r-cnn and other models. YOLOv3 network is a single-stage target detection method, which aims to find out the coordinate position and category of the target object in a given picture. The first is feature extraction. The input of yolov3 is generally 416x416x3 original image matrix, n coordinate labels and N classification labels, where n represents the number of target objects in each image. After the feature extraction network, we can get three different sizes of feature images: 13x13x255, 26x26x255, 52x52x255. Then there is the YOLO layer, which is responsible for the prediction or training of the obtained feature map. [18] During the training, the coordinate loss and classification loss are calculated respectively. This paper is mainly for ship classification and recognition. The feature fusion layer of traditional yolov3 algorithm consists of three feature maps of different scales, N / 32 × n/32,n/16 × N / 8 (input image size is n) × n) To detect large, medium and small targets. In view of the small proportion of long-distance ships and short-distance small ships in the view.

2.5. **Binocular positioning**

Use two cameras to position. For a feature point on an object, the image of the object is captured by two cameras fixed at different positions, and the coordinates of the point on the image plane of the two cameras are obtained respectively. As long as we know the exact relative position of the two cameras, we can get the coordinates of the feature point in the coordinate system of a fixed camera by geometric method, that is, we can determine the position of the feature point.

In this paper, the traditional ranging algorithm is to get the binocular camera matrix, distortion vector, left and right rotation matrix and translation vector through stereo correction. Then the global stereo matching algorithm is used to match the same features of the left and right camera views to get the disparity map. Finally, the distance information of the target object can be calculated through the triangular similarity principle. The dual camera model is shown in Fig.3.

![Binocular camera model](image)

**Fig. 3** binocular camera model

Combined with figure 3 and trigonometric similarity principle, the following formula can be obtained

\[
X = z \times \frac{x_l}{d} \\
Z = b \times \frac{f}{d}
\]

(1)
(2)

Where: \(d = x_l - x_r\), \(x_l\) and \(x_r\) are the abscissa of \(x_l\) and \(x_r\), \(b\) is the baseline distance between the left and right cameras, and \(f\) is the focal length of the camera, all of which are in mm.
3. Overview of innovative principles of ship visual perception system based on deep learning

3.1. Improvement of ship identification

In this paper, the mesoscale detection module of yolov3 is improved, the original three scale detection is expanded to four scale detection, and the size of feature map is \( n / 64 \times N / 64 \), so as to improve the accuracy of small target detection. In this paper, the improved yolov3 multi-scale detection module uses darknet-53 network model as convolution layer for 32 times down sampling, which is divided into input layer, convolution layer, feature fusion layer and output layer. In this paper, the input image size is 416 \( \times 416 \), layer 0-74 is the convolution layer, which is composed of residual module and convolution module to realize down sampling between residual modules. Sampling is carried out under convolution with step size of 2, so as to construct deep network and propose target features. Layer 75-105 is the feature fusion layer, which is divided into three scales, 13 \( \times 13 \), 26 \( \times 26 \), 52 \( \times 52 \), each scale features interaction, complete the pyramid feature fusion. Finally, the scale feature map output from the feature fusion layer is classified accurately and the position is regressed.

\[
\mathbb{F} = \begin{cases} \mathbb{C} & \text{if } b \cap c \\ \mathbb{D} & \text{otherwise} \end{cases}
\]

Where: \( b \) is the label box, \( c \) is the cluster center box, and \( d \) is the distance from the label box center to the cluster center. The smaller the distance between the annotation box and the cluster center, the higher the detection accuracy. So the maximum intersection and union ratio of the corresponding anchor box is obtained by calculation.

![Diagram of Improved yolov3 Network](image_url)
3.2. Improvement of ship positioning and navigation

The effect of disparity map directly affects the ranging effect. However, in the case of complex background, small proportion of ships in the view, long distance and unobvious features, the traditional method can not get effective disparity map. Therefore, this paper proposes a local matching method based on ship detection and recognition, so as to obtain effective parallax, calculate distance and locate. After the ship is detected and identified by yolov3, the center position of the ship boundary box detected by the left and right cameras is sorted. According to the principle that the same target has the same ordinate value in the plane coordinate of the two cameras view, the similarity matching algorithm is used to match the corresponding target, and then the disparity of the same target in the two cameras view is calculated by the feature matching algorithm. Finally, the ship position is calculated by formula (1) and (2).

3.3. Data and image fusion location

AIS data and image contact calculation process: obtain the position of the ship and the target ship in the AIS data, calculate the azimuth of the line between the ship and the target ship, at the same time check the bow data, camera angle, etc., calculate the orientation of the target ship relative to the ship on the plane, and then combine the position of the image on the computer. The AIS data is matched with the ship, and the image is marked with MMSI number, ship name information, navigation status information, position information, speed and course information. The received AIS data will be saved in the database for ship trajectory analysis. The calculation flow of AIS and image fusion is shown in fig.5.

![Flow chart of AIS and image fusion](image)

3.4. Result analysis

Further, the position data supplemented by AIS and radar data are matched, AIS position and radar position are selected according to the rectangular range of statistics, and certain screening conditions are set up to select the target ship meeting the specified conditions to enter the rectangular screening list. Based on the list of related ships, the target ship identification algorithm combines matching as the main method, and the AIS data is used to match radar data as the auxiliary. If there are multiple radar targets that can match in the matching area, the shortest distance comparison of position data is made. Under short distance, the Euclidean distance between AIS and multiple matching targets is calculated according to plane rectangular coordinate system in short distance, and the matching is achieved with the minimum range radar target. When the location match is successful, the target matching is determined to be successful. If AIS does not have matched radar target, it can output AIS target data only, and combine it with shooting video to prompt. By combining the radar target with AIS target location and target vessel identification algorithm, the blind area and false echo of radar detection and the blind area and deviation of AIS equipment signal are effectively compensated, and the accuracy and reliability of target
position information are improved, and the matching rate is improved. [19] Object recognition of intelligent vision system based on improved algorithm is shown in Fig.6.

![Object recognition of intelligent vision system based on improved algorithm](image)

**Fig. 6** Object recognition of intelligent vision system based on improved algorithm

The figure shows the recognition of objects by the intelligent vision system based on the improved Yolo algorithm, which is more accurate for the recognition of the hull itself, and higher for the recognition of sea conditions and dangerous things. After learning and research, it is concluded that the recognition accuracy of objects on water is more than 90%. Its main innovation is the introduction of improved machine learning algorithm, combined with binocular visible light and infrared camera function, to achieve all-weather target positioning, object recognition, detection and tracking. By fusing machine vision data with radar and AIS data, route planning is realized, which has the ability of machine vision navigation.

4. Conclusion

The training of ship radar and AIS route is carried out by genetic algorithm, and the video collected is combined. However, the automatic generation function of route can be realized. The route can be displayed on VTS or radar chart or displayed in the diagram designed by software. As shown in Figure 1 below, Arctic map shows the route map of polar ice area automatically planned in the system interface. At the same time, it can also be displayed in the video, as shown in Figure 2, which can show the navigation situation of the ship more intuitively and give guidance to the navigation of the future ships. [20]

The visual tracking and identification of ships has always been an important research field in maritime navigation. It is the most important in the future intelligent navigation era. This paper mainly summarizes the principle and personal innovation of the visual tracking and recognition function of intelligent ships, and describes the principle of recognition and tracking positioning, visual perception technology, the function characteristics of tracking and identifying positioning, and the visual tracking and positioning Based on the overview of the general ship vision tracking and identification system, the paper analyzes the personal innovation of visual tracking and positioning based on Algorithm and AIS. It explains the problems of the system, such as the system itself, the environment, [21] the artificial defects and the shortcomings of the current visual system finally, innovative technical ideas that individuals have been exploring in this area are specially added. It is mainly based on the known yoloy3 algorithm and darket network, and then extends the direction, and explores the future significance of the intelligent vision tracking system. [22] Intelligent ship vision tracking and identification system will play an important role in the future intelligent navigation era, which will directly affect the safety of maritime traffic, traffic efficiency, intelligent monitoring of ships, maritime law enforcement and other important maritime operations.

Acknowledgements

National Key Research and Development Program of China (2020YFE0201200).
References

[1] Hu Houli, Wei Wei, Hu Mona. The principle and application of deep learning algorithm [j]. Information technology, 2015 (2): 175-177.

[2] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[J]. 2015:779-788.

[3] Li Shuanghuang. Research on detection algorithm of moving ship in River video monitoring [d]. Dalian Maritime University, 2018.

[4] J. Zhao, H. Liu, Y. Feng, S. Yuan and W. Cai, "BE-SIFT: A More Brief and Efficient SIFT Image Matching Algorithm for Computer Vision," 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, Liverpool, 2015, pp. 568-574.

[5] FANG J, ZHOU Y, YU Y, et al. Fine-grained vehicle model recognition using a coarse-to-fine convolutional neural network architecture [J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(7): 1782 -1792.

[6] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1440-1448.

[7] Pan Meilian. Application of computer vision technology in ship automatic identification system [u]. Marine science and technology, 2019 (2): 76-78.

[8] Hua xufen, Zhao Yong. Visual tracking control in ship automatic identification system [u]. Ship science and technology, 2019 (2): 76-78.

[9] Zhang Xinjiang. Research on the method of ship bearing only target tracking based on computer vision [u]. Ship science and technology, 2020 (4): 34-36.

[10] Li Xiaofei. Research and application of visual based ship tracking and 3D positioning technology [u]. Guangdong University of technology, 2014:0-83.

[11] Jiang Jingjing, an Bowen. Ship detection method based on vibe algorithm in Low Altitude Aerial Video [J]. Microcomputer and application, 2017, 36 (10): 44-47.

[12] Li Shuanghuang. Research on moving ship detection algorithm in inland river video surveillance [D]. Dalian Maritime University, 2018.

[13] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 580-587.

[14] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1440-1448.

[15] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017.

[16] Lu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[J]. 2016.

[17] Li Xiaofei, songyann, xuronghua, Chen Jun. Ship tracking and positioning based on binocular vision [tp]. Journal of Nanjing University of Information Engineering (NATURAL SCIENCE EDITION), 2015, 51 (5): 799-804.

[18] Jin Zhi, Liangshan, caofenping. Visual servo tracking method of ships in Yangtze River control reach based on automatic identification system [u]. Computer application, 2011 (12): 3414-3417+3421.

[19] Huang Yuxin. Video path ship detection and tracking based on OpenCV [j]. Ship science and technology, 2017 (16): 28-30.

[20] Zhao Weixin, Cheng Chengqi. CAMSHIFT motion tracking algorithm based on Kalman filter [j]. Journal of Peking University (NATURAL SCIENCE EDITION), 2015, 51 (5): 799-804.

[21] Zhong Li, Jin Zhi, Liangshan. Design of intelligent visual tracking control system for control
reach [u]. Traffic technology, 2013 (4): 137-139+142.

[22] Liu Hongtao, research on the deployment and coordination technology of visual sensors in intelligent ship sensing system [u]. Dalian Maritime University, 2020:0-52.

[23] Tengfei, Liu Qing, Guo Jianming, et al. Inland river ship tracking under TLD framework [j]. Journal of Applied Science, 2014,32 (1): 105-110.