A Water Surface Moving Target Detection Based on Information Fusion Using Deep Learning

Wei Zhang¹, Chi-fu Yang¹, Feng Jiang²*, Xian-zhong Gao³, Kai Yang⁴

¹ School of Mechatronics Engineering, Harbin Institute of Technology, Harbin, Heilongjiang, 150000, China
² School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang, 150000, China
³ College of Aerospace Science and Engineering, National University of Defence Technology, Changsha, Hunan, 410073, China
⁴ Department of Aeronautical Mechanical Engineering, Army Aviation Academy

*Corresponding author’s e-mail: fjiang@hit.edu.cn

Abstract. An unmanned surface vehicle (usually referred to as USV) is an unmanned surface ship that has been developed in recent years. Surface unmanned boats can be used for a variety of civil and military missions, such as marine environmental monitoring, personnel search, anti-mine mines; island map mapping, offshore facility maintenance, and hydrographic surveys. Therefore, the research on unmanned surface vehicles has important theoretical significance and practical application value. The optical image information collected by the camera and the point cloud information collected by the laser radar are used for information fusion, and the optical image is processed by the deep neural network to obtain the target type, the detection confidence and the target position. At the same time, the point cloud information of the lidar image is used to obtain the distance and orientation information of the target. Finally, the two information is combined to obtain the type of surface target or obstacle, the confidence of the detection, and the distance and bearing information from the unmanned surface vehicles.

1. Introduction

When performing tasks or doing work on the ocean, you must first consider the unique workspace environment of the ocean. Compared with other working environments, the efficiency, cost performance and safety of personnel equipment of marine equipment are more important reference conditions. With the comprehensive development of science and technology, unmanned intelligent marine vehicles have developed rapidly and are widely used in civil and military fields. The unmanned intelligent marine vehicles previously defined by the US Department of Defense mainly consist of two major categories: surface unmanned boats and underwater unmanned aerial vehicles [1-4]. As a new type of surface mobile platform that can sail independently on the water surface, the Unmanned Surface Vehicle (USV) has the ability to perform tasks independently or completely independently. It can be seen that in the context of accelerating the construction of maritime powers and the development of marine equipment, with the comprehensive development of various technologies, surface unmanned boats will be widely used in civil and military fields and occupy an important position. The development of surface unmanned boat technology has received increasing attention. It is a prerequisite for water surface unmanned boats to perform various tasks.
2. Water surface moving target detection

Unmanned boats mainly undertake tasks such as intelligence gathering, surveillance and reconnaissance, mine clearance, anti-submarine, search and arrest, and hydrographic survey. According to the mission requirements, the unmanned boat should have the ability to detect and identify the surface target, that is, to obtain the position and motion information of the target. With the unique advantage of light vision, in the close-range detection area of the water surface, light visual perception is easier to obtain the surface target information than other means. With the development of related technologies, the surface motion detection technology based on information fusion has become more mature. According to the characteristics of unmanned surface vehicles, it can be equipped with various sensing devices such as cameras, laser radars, infrared sensors, and millimeter wave radars. Different sensors can extract different features for the same target, exert their respective advantages, and integrate detection information to improve the detection effect of moving targets. Therefore, target detection based on multi-sensor information fusion is the development trend [5-7].

Harbin Engineering University Wan Lei et al. proposed an automatic detection method for offshore targets based on coastline information for the detection of maritime targets in unconstrained coastal backgrounds, and obtained the target location [8]. Yan Xiaozheng of Nanyang Technological University proposed a real-time vision-based long-distance target detection and tracking algorithm for unmanned surface vehicles. High-resolution images (2736 × 2192) were used in the study to obtain a high-precision target distance [9]. Zhejiang University, Longgang, etc., based on the research and development of the visual perception system of unmanned surface vehicles, used the deep learning framework of cascaded principal component analysis network to study the surface vessel detection algorithm. The suspected target area is determined by saliency detection, and the detected target area is extracted using the PCANET model. The result is input into the support vector machine to obtain the final two-class result [10]. However, these research methods require coastal information assistance or lack of practical application on unmanned surface vehicles.

This paper mainly introduces an unmanned surface vehicles target detection and identification method based on the fusion of monocular camera and lidar information. On-line detection and identification of water surface targets, that is, fusion of lidar information and camera information to detect, identify and locate targets within the sensing range. This puts forward higher requirements for the sensing part of the unmanned surface vehicles, and can independently perform maritime target detection, autonomous path planning and automatic completion of obstacle avoidance [11]. The on-line detection, identification and localization method of water surface targets combined with lidar and camera information can provide environmental perception for tasks of unmanned surface vehicles.

3. Camera and lidar joint calibration

The cost of the camera is low, and the researchers who use the camera to make the perceptual algorithm are the most, and the corresponding technology is relatively mature. However, the camera also has some shortcomings: first, the three-dimensional position information of the target cannot be accurately obtained; secondly, the imaging effect of the camera is greatly affected by the environment. The laser radar can make up for the shortcomings of the camera. On the one hand, the distance of the target is far away, and the three-dimensional information of the target can be accurately obtained. On the other hand, it is less affected by the external environment, and the data obtained by the detection is robust.

Depending on the characteristics of the camera and lidar, the two sensors can be used to detect surface targets. If the task is to identify the type of surface obstacles, it is easy to classify and identify the obstacles by relying on deep learning. However, the point cloud image processed by the lidar cannot accurately discriminate the target type. According to the radar point cloud information, the obstacle contour can be extracted, and only the obstacle can be roughly judged, but the target motion state information is acquired by using the laser radar. The fusion of lidar and camera is actually to take advantage of the lidar and camera to make up for their shortcomings. It is necessary to fuse the two sensor information at the bottom [12-16].
The image data captured by the camera is represented by a 3D lattice cloud captured by Lidar. The goal is to create a transformation matrix that maps 3D points to 2D points, namely:

$$\begin{bmatrix}
u \\
v \\
1 \\
1 \\
\end{bmatrix} = 
\begin{bmatrix}
f_x & 0 & u_0 \\
0 & f_y & v_0 \\
0 & 0 & 1 \\
\end{bmatrix} 
( R \ t ) \begin{bmatrix}
x \\
y \\
z \\
1 \\
\end{bmatrix} = M 
\begin{bmatrix}
x \\
y \\
z \\
1 \\
\end{bmatrix}$$

(1)

The matrix is the camera parameter, and is the XY axis direction scale factor (the effective focal length in the horizontal and vertical directions), and the center point of the image plane, also known as the principal point coordinates. For the rotation matrix, the translation vector.

$$\begin{align*}
u &= m_{11}x + m_{12}y + m_{12}z + m_{14} \\
v &= m_{21}x + m_{22}y + m_{23}z + m_{24} \\
v &= m_{31}x + m_{32}y + m_{33}z + m_{34} \\
v &= m_{41}x + m_{42}y + m_{43}z + m_{44} \end{align*}$$

(2)

The camera calibration determines the camera's camera parameters, ie the horizontal effective focal length, the vertical effective focal length, the center point of the image plane. Depending on the actual position of the camera and the laser beam, the rotation matrix and the motion vector can be obtained by measuring with a vernier caliper. Therefore, the conversion matrix can be obtained, and the three-dimensional information obtained from the laser radar can be converted to the two-dimensional information of the image obtained by the camera.

4. Camer and lidar information fusion

4.1. Faster-RCNN

The Faster R-CNN model uses a 4-step iterative training strategy: (1) first pre-training the RPN on ImageNet and finetuning on the PASCAL VOC dataset; (2) training the candidate region s generated using the trained PRN separately A Fast R-CNN model, this model is also pre-trained on ImageNet; (3) Initialize the RPN with the CNN model part (feature extractor) of Fast R-CNN, and then finetuning the remaining layers in the RPN, at this time Fast R - The feature extractor of the CNN and the RPN is shared; (4) The fixed feature extractor performs finetuning on the remaining layers of the Fast R-CNN. After several iterations, Fast R-CNN can be integrated with RPN to form a unified network.

4.2. Test data and test

According to the loss curve and computing power, the number of trainings is determined to be 70,000 times. The test was performed with 50% of the test data set, and the AP of each target was obtained. It can be seen that the MAP is higher than the MAP of the official VOC data set, which proves that the training effect is better. Finally, the training and test flow diagrams are generated.

4.3. Image target detection and lidar point cloud fusion

Because the lidar point cloud is very fast for target detection, the average is 0.1s, and the image is detected slowly by Faster-RCNN. In order to maintain the consistency of the spatial dimensions of the two, a historical data matching algorithm is proposed. Assume that the camera captures the target image, and the lidar collects the point cloud information for target detection, and converts the point cloud information into planar two-dimensional information to extract the point cluster circle. After the moment, the image processing algorithm detects the type, confidence and coordinates of the target in the image. The algorithm stores the two-dimensional information converted from the point cloud information at all times during the period, and matches the detected result of the image target with the two-dimensional information during the period to obtain an optimal matching result. The entire integration mainly includes the following steps:

1. Projecting the clusters according to the clustering onto the two-dimensional camera plane, using the mean to calculate the point cluster center point O1..In, set the point cluster center circle for radius;
(2) Perform classification judgment based on the calculated degree of overlap. If degree of overlap is greater than 0.5, the image and the target of the lidar detection are considered to be the same target;

(3) If the degree of overlap is less than 0.5, then the image does not detect the target in the area or the laser radar does not detect the target;

(4) If the image detects the target and the laser radar does not detect the target, and the unmanned surface vehicles is commanded to advance in the direction according to the direction in which the laser radar detects the target. Step flow chart.

![Figure 1. Various target detection accuracy](image1.png)

![Figure 2. Target recognition and positioning method based on information fusion](image2.png)
5. Camer and lidar information fusion
The single target and multiple targets were tested separately, and the obtained image and radar point cloud are shown in Figure 3. The transmission interface of the information fusion and the control interface of the shore control box are made by MFC. The lower left corner is the result of the displayed target detection. The actual test results of the obstacle avoidance test prove that the fusion is effective. After the actual application of the Robotx competition in Hawaii in 2018, compared with the domestic and foreign university teams, the accurate performance of this method is higher than that of most teams for the surface target detection performance of unmanned surface vehicles.

![Figure 3. Target information fusion experiment on the river](image)

In the Songhua River waters, the target detection image and the point cloud information fusion test were carried out. We randomly selected 400 frames of multi-target detection and 100 frames of single-target detection. There are 18 types of targets to be detected, and the number of frames is divided into five categories according to the number of occurrences within the detection range: single target, double target, three target, four target and five target. The statistics of these detection frames are divided into two categories: correct detection of one target and no detected target; dual target is divided into correct detection of one target, two targets and all detection errors; Divided into six categories. The data of the information fusion detection results are as follows:

Table 1. Fusion detection result of multi-target detection frame

| Number of target | 1  | 2  | 3  | 4  | 5  | 6  |
|------------------|----|----|----|----|----|----|
| Number of frame  | 100| 78 | 120| 134| 68 |    |
| Number of correctly detected | 0  | 1  | 0  | 2  | 1  | 0  |
| Correctly detected frames | 2  | 98 | 3  | 7  | 68 | 2  |
| Proportion(%)     | 2  | 98 | 4  | 9  | 87 | 2  |

6. Conclusion
In this paper, considering the time difference between image detection and point cloud processing, an information fusion algorithm for optimal matching of historical information is proposed. The problem of the difference between the point cloud information processing time and the image processing time dimension is solved, and the actual experiment achieves a high precision effect. After the field test, the proposed method can accurately detect the water surface moving target. It mainly studies the principle of camera calibration and the method of joint calibration of camera and laser radar. Provides sufficient parameters for joint calibration of the camera and lidar. After testing in the Songhua River field, it is proved that the information fusion algorithm based on this paper has a good detection effect and meets the real-time requirements.
Acknowledgments
This work is partly funded by National Key Research and Development Program of China via grant 2018YFC0806802 and 2018YFC0832105.

References
[1] J. Qiu, Z. Tian, C. Du, Q. Zuo, S. Su and B. Fang. A Survey on Access Control in the Age of Internet of Things. IEEE Internet of Things Journal. 2020. DOI: 10.1109/JIOT.2020.2969326.
[2] M. Li, Y. Sun, H. Lu, S. Maharjan and Z. Tian. Deep Reinforcement Learning for Partially Observable Data Poisoning Attack in Crowdsensing Systems. IEEE Internet of Things Journal. 2020. DOI: 10.1109/JIOT.2019.2962914.
[3] Z. Tian, C. Luo, J. Qiu, X. Du and M. Guizani. A Distributed Deep Learning System for Web Attack Detection on Edge Devices. IEEE Transactions on Industrial Informatics. 2020. Vol 16(3): 1963-1971. DOI: 10.1109/TII.2019.2938778.
[4] Z. Tian, X. Gao, S. Su and J. Qiu. Vcache: A Novel Reputation Framework for Identifying Denial of Traffic Service in Internet of Connected Vehicles. IEEE Internet of Things Journal. 2020. DOI: 10.1109/JIOT.2019.2951620.
[5] Q. Tan, Y. Gao, J. Shi, X. Wang, B. Fang and Z. Tian, Toward a Comprehensive Insight Into the Eclipse Attacks of Tor Hidden Services, IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1584-1593, April 2019.
[6] Hermann D, Galeazzi R, Andersen J C, et al. Smart Sensor Based Obstacle Detection for High-Speed Unmanned Surface Vehicle [C]// Ifac Conference on Manoeuvring and Control of Marine Craft, 2015.
[7] Zhang X, Wang H, Cheng W. Vessel detection and classification fusing radar and vision data [C]// Seventh International Conference on Information Science and Technology. IEEE, 2017: 474-479.
[8] Wan Lei, Zeng Wenjing, Zhang Tedong, et al. Rapid extraction of near-shore targets[J]. Journal of Harbin Engineering University, 2012, 33(9): 1158-1163.
[9] Shih B S, Mou X, Mou W, et al. Vision-based navigation of an unmanned surface vehicle with object detection and tracking abilities [J]. Machine Vision and Applications, 2017.
[10] Long Gang, Ren Jianqiang, Gong Xiaojin. The Algorithm of Sea Surface Vessel Detection Based on PCANet[J]. Journal of Hangzhou Dianzi University, 2017(2).
[11] Li C, Cao Z, Xiao Y, et al. Fast object detection from unmanned surface vehicles via objectness and saliency [C]// Chinese Automation Congress. IEEE, 2016: 500-505.
[12] Ji Qiang. Research on Several Algorithms for Removing Redundant Data from Space Remote Sensing Images [D]. Wuhan University, 2013.
[13] Ren Guiwen. Research on Dual Camera Calibration Method Based on OpenCV[J]. Science Technology and Engineering, 2016, 16(03):211-214.
[14] Wang Han, Yin Hong, Research and implementation of binocular camera calibration based on OpenCV[J]. Fujian Computer. 2018, 34(08): 121-122+129.
[15] Gu C, Wang G, Li Y, et al. A Hybrid Radar-Camera Sensing System With Phase Compensation for Random Body Movement Cancellation in Doppler Vital Sign Detection [J]. IEEE Transactions on Microwave Theory & Techniques, 2013, 61 (12): 4678-4688.
[16] Zyczkowski M, Palka N, Trzcinski T, et al. Integrated radar-camera security system: experimental results [J]. Proceedings of SPIE - The International Society for Optical Engineering, 2011, 8021 (80211U).