Proceedings of the 2021 IEEE
International Conference on Robotics and Biomimetics
December 27-31, 2021, Sanya, China

2021 IEEE International Conference on Robotics and Biomimetics (ROBIO) | 978-1-6654-0535-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/ROBIO54168.2021.9739449

# A Novel Unmanned Surface Vehicle with 2D-3D Fused Perception and Obstacle Avoidance Module

Zhe Chen [1], Tao Huang [1], Zhenfeng Xue [1,2], Zongzhi Zhu [3], Jinhong Xu [1] and Yong Liu [1,2]

***Abstract*— Unmanned surface vehicles (USVs) are important intelligent equipment that can accomplish various tasks on open area marine. During operation, environmental perception and obstacle avoidance is of vital significance to its autonomy. In this paper, we propose a novel USV equipped with fused perception and obstacle avoidance module that contains robust perception, localization and effective obstacle avoidance strategy. The new module is named Three-Dimensional Perception Module (PMTD), which utilizes camera and LiDAR to integrate multi-dimensional environmental information. It is able to detect, identify and track target objects in the process of autonomous travel. The localization precision achieves a centimeter-level with GPS and IMU devices. Meanwhile, the obstacle avoidance strategy allows the USV to efficiently keep away from static and dynamic floating objects in water areas. Through real-world experiments, we show that with the help of the proposed module, the USV can complete stable and autonomous operation and obstacles avoidance path planning even without any manual intervention. This indicates the strong ability of the module in autonomous driving for USVs.**

## I. INTRODUCTION

In conditions of complex water area and uncertain climate, artificial exploration or searching is commonly inefficient and even dangerous. As such, unmanned surface vehicles (USVs) are attracting much more attention in recent years [1], [8] due to the increasing demands on open-area marine for different tasks, such as environmental monitoring, good transportation, military reconnaissance, and *etc*. It faces more challenges than ground vehicles mainly because of the influence by water tides as well as high similarity of surrounding environment in water areas. Robust environmental perception and obstacle modules are required to ensure stable operation for USVs in real-world scenes.

With the development of various sensor, data transmission, and image processing technologies, there has been a continuous improvement in localization accuracy, especially for unmanned ground vehicles. In the ground-based unmanned systems, the target objects are usually described by a collection of 3D points, and then camera and LiDAR are used to detect and track small objects like pedestrians and vehicles [10]. For example, Chen *et al.* [4] propose a vehicle detection, tracking and classification system based on road edge objects. Zhang *et al.* [15] combine the data of stereo camera and LiDAR for vehicle detection. For this method, Random Hypersurface Model (RHM) [2] is employed to extract object shape parameters based on visual detection.

By far, there has been very few methods for environmental perception and obstacle avoidance in the field of USVs. Zhan *et al.* [14] propose an online learning method to detect in unknown water areas through Convolutional Neural Networks (CNNs). Wang *et al.* [13] further propose a multi-sensor-based urban waterway autonomous system to show the joint performance of localization, perception, planning and control. It is able to detect static and dynamic obstacles, but suffers from the detection accuracy with only one single laser.

Based on noisy stereo camera data, a real-time algorithm for detecting and tracking dynamic objects is proposed in [5]. The SSD [7] network with MobileNet [6] backbone is employed to perform object detection in the image. However, the accuracy and speed of above method is limited and not enough for real applications. In this paper, we replace it with YOLOv4 [3] for better performance.

This article proposes a complete perception and obstacle avoidance system for unmanned surface vehicles. Based on the fusion of laser and camera, we

[1]Zhe Chen, Tao Huang, Zhenfeng Xue, Jinhong Xu and Yong Liu are with College of Control Science and Technology, Zhejiang University, Hangzhou, China.

[2]Zhenfeng Xue and Yong Liu are with Research Center for Intelligent Perception and Control, Huzhou Institute of Zhejiang University, Huzhou, China. (Yong Liu is the corresponding author, yongliu@iipc.zju.edu.cn.)

[3]Zongzhi Zhu is with Zhejiang Guoli Xinan Technology Co., Ltd, China, zhu_zz1106@163.com.

raise a real-time method that identifies target, with the capacity of decimeter-level positioning and detection of obstacles within close range. The 3D perception module can effectively identify boats, vessels, buoys, and other objects, ensuring that unmanned ships generate reasonable obstacle avoidance strategies during autonomous navigation. The main contributions are listed as follows.

- We propose a novel USV that has better autonomous perception and obstacle avoidance capabilities in water environments.
- We design a 2D-3D fused perception module (PMTD) based on camera and laser, showing high accuracy in positioning, obstacle detection and target recognition.
- We have conducted real-world experiments to verify the effectiveness of the whole system.

## II. Hardware Overview

As a common practice, the proposed unmanned surface vehicle contains a set of modules, including the V-shaped hull structure, positioning and communication, obstacle detection, core control and computing unit, power system, and *etc.*


Fig. 1. An overview of the unmanned surface vehicle

In order to ensure its autonomy, many devices are equipped on the new USV, such as the controller, IMU, LiDAR, GPS. An Intel I7 quad-core processor with 8G memory is employed as the core computing unit. The high-performance processor enables to reduce the impact of communication delays and insufficient computing power on the overall performance. The USV also uses a series of high-precision sensors to gather positioning information. Among them, an IMU (Xsens) device is installed parallel to the surface, and its direction is consistent with the forward. The IMU provides high-frequency triaxial acceleration and triaxial angular velocity.And the GPS device provides complete location information. For long-range detection and better positioning, A 3D LiDAR (RoboSense, RSlidar) is installed on the upper end of the hull. What's more, the USV contains a novel perception and obstacle avoidance module that is specifically designed by us, named three-dimensional perception module (PMTD). Within this module, a 3D LiDAR (Livox Horizon) and a standard monocular camera are installed. The 2D image data by the monocular camera is fused with the 3D point cloud data by the LiDAR, which is an essential source of sensor information for the close-sensing module.

According to different propulsion methods, the USV can be divided into tail push type and double push type. In our USV, a semi-fixed single propeller outboard engine is adopted. The propeller is fixed in the vertical direction relative to the boat. Here, the common way to generate various sizes and directions of propulsion are two-folds. First is through modulating the received speed signal into the corresponding PWM. Second is to receive the angle signal by the linkage connected with the steering mechanism and the vector rudder to drive the propeller in the horizontal direction to the corresponding angle. Compared with standard single-propeller-single-rudder and single-propeller-double-rudder model, that fix the propeller and transform the stern flow direction by controlling the orientation of the stern rudder, the dynamic model of the system is closer to the Ackermann trolley model on the ground.

## III. Perception and Autonomy System

The system flow diagram of USV is shown in Fig. 2. In general, an autonomous navigation task from upper-level human interface is issued to the computing unit. During the execution of the whole mission, the positioning module determines the USV's pose in real-time. And the perception module recognizes the targets and inputs the 3D information of the obstacle into the path planning module. Based on the acquired pose, speed, and the obstacle information, the path planning module calculate a safe and reliable obstacle avoidance trajectory.

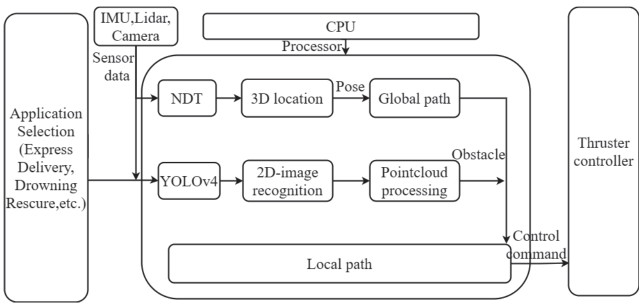



Fig. 2. Illustration of the USV system framework

### A. 2D-3D fused object perception

The 3D point cloud data obtained by LiDAR contains distance and intensity information. In contrast, the 2D image data obtained by camera is able to extract color and texture information. In summary, LiDAR has good robustness in 3D information but is difficult to identify objects in the transformation of adjacent frames accurately, and the camera can make up for this shortcoming.

Many works have confirmed that the sparse laser point cloud has well effect on marking the target in the case of time-synchronized images [12]. Camera and radar are complementary on target recognition in different aspects. Therefore, in this part, the 3D laser point cloud is merged with the 2D image data to be applied in target recognition and obstacle detection.

LiDAR's perception is divided into obstacle detection and target recognition. The obstacle detection with low-level perception usually extracts the length, width, and other specification information of the object, which meets the definition requirements of obstacles. The high-level perception, based on obstacle detection, fuses image recognition results to define the target point cloud data in 3D space.

Considering the large amount of point cloud data and the noise, it is necessary to pre-process the point cloud data. The detection module adopts the voxel grid filtering method to down-sample the original data, which can smooth the point cloud density, and reduce the scale of data. What's more, due to the large viewing angle of LiDAR, it is easy to collect obstacle point cloud data with low correlation. Therefore, the ROI of the original laser point cloud is determined by the area scanned through the 2D image. In ground vehicles, considering the influence of the ground point cloud by clustering method, they generally eliminate ground segmentation.However, in consideration of the characteristic that the laser can penetrate the water, the operation of ground segmentation can be ignored and directly eliminate the outliers from the data by range filtering.

According to the determined ROI area and the preprocessed point cloud data, the data points are divided into clusters by the Euclidean Clustering method. Each cluster represents a single obstacle. Then the bounding box method of the smallest bounding box is employed to fit each cluster. From each bounding box, the physical information of the obstacle is easily to be obtained. It is also possible to estimate the movement state by correlating the obstacle data of multiple consecutive frames and calculate the movement information such as the speed and direction of the corresponding obstacle.

Based on LiDAR obstacle detection, the camera's image data is merged to have a better target recognition effect.For image recognition, the YOLOv4 network is adopted as the detection model. YOLOv4 [3] is a well-known convolutional neural network (CNN) designed for object detection. Compared with other detectors, it has better speed (FPS) and accuracy (mAP). Under the framework of previous YOLO versions, YOLOv4 combines other suitable detection structures. For example, the head part of YOLOv3 is retained, but the backbone network is modified to CSPDarknet53, and spatial pyramid pooling (SPP) is adopted to expand the receptive field, and the path aggregation module is employed. These strategies optimize the overall detection model from data processing, backbone network, network training, activation function, loss function, and other aspects, balancing speed and accuracy requirements.

Finally, target recognition is performed on the 2D image collected by the camera through YOLOv4. And the laser point cloud provided by the LiDAR is projected onto a 2D plane. The target data points of the image recognition are correlated to obtain the 3D data corresponding to the target object.

### B. Precise positioning and path planning

*1) Precise positioning:* The positioning module adopts the LiDAR 3D data above the hull in static and dynamic conditions through Normal Distributions Transformation (NDT) [9] matching algorithm.

The advantage of the algorithm is that the matching effect is not related to the initial value of the input algorithm. Even if the difference between the hypothetical value and the actual value is significant, NDT can correct the error shortly, accurately.Firstly, the algorithm divides the spatial range of the point cloud data in the reference frame into cell grids or voxels of fixed size. The PDF of each cell can be regarded as an approximate representation of the surface.In other words, it can be interpreted as the generation of surface points in a cell. The eigenvectors and eigenvalues of the covariance matrix can be employed to represent the surface information of cells, such as orientation, flatness.

$$\begin{cases} \vec{\mu} & = \frac{1}{m} \sum_{k=1}^{m} \vec{y}_k \\ \Sigma & = \frac{1}{m-1} \sum_{k=1}^{m} \left(\vec{y}_k - \vec{\mu}\right)\left(\vec{y}_k - \vec{\mu}\right)^{\mathrm{T}}, \end{cases} \tag{1}$$

where $\vec{\boldsymbol{y}}_{k=1,\ldots,m}$ denotes the positions of all reference scan points included in one cell.

When using NDT to register the current scanning data points, the main purpose is to find the pose of the current scanning frame relative to the reference frame. The fundamental implementation is through maximizing the possibility of each scanning point on the reference scanning plane. For example, given a set of current lidar scanning data points $X = \{\vec{x}_1, \ldots, \vec{x}_n\}$, a pose parameter to be optimized $\vec{p}$, and a transform function $T(\vec{p}, \vec{x})$ to transform the point cloud data to the reference coordinate system (usually the previous frame). Calculate the normal probability distribution function of the current point falls in the cell of the reference frame as follows:

$$p(\vec{x}) = \frac{1}{(2\pi)^{D/2}\sqrt{|\Sigma|}} \exp\left(-\frac{(\vec{x}-\vec{\mu})^{\mathrm{T}}\Sigma^{-1}(\vec{x}-\vec{\mu})}{2}\right), \tag{2}$$

where $\vec{\boldsymbol{\mu}}$ and $\boldsymbol{\Sigma}$ denote the mean vector and covariance matrix of the reference scan surface points within the presumptive cell where $\vec{x}$ lies.

The objective function of obtaining the optimal pose parameter $\vec{\boldsymbol{p}}$ of all points is defined as

$$s(\vec{p}) = \sum_{k=1}^{n} p\left(T\left(\vec{p}, \vec{x}_k\right)\right), \tag{3}$$

where the optimal pose parameter $\vec{\boldsymbol{p}}$ needs to be calculated iteratively until the convergence condition is reached.

The initial position of USV is obtained based on the pre-obtained water area map and NDT algorithm. What's more, when a USV performs point-to-point obstacle avoidance planning according to the set task, the rotation and translation vectors between frames are registered using the NDT algorithm. Based on the initial calculation, the accurate position of USV can be acquired iteratively in real-time.

*2) Path planning:* As a typical heuristic search algorithm, A* is widely used in many problems such as path planning and graph traversal because of its fast search speed. A* is mainly used to find the optimal solution from the initial state S to the termination state G in the state space. The algorithm's input is the initial state and the termination state, while the output is the action sequence number of the target in each state transition. Due to the guidance of the heuristic function, the A* algorithm will transfer to the state at a low cost. Therefore, compared with other search algorithms, the A* algorithm usually has better performance in search efficiency.

In the state space of path planning, for a state N, g(N), h(N), and f(N) need to be provided to the A* algorithm. In detail, g(N) is expressed as the optimal cost from the initial state S to the current state N. h(N) is expressed as the estimated cost from state N to termination state N. f(N) = g(N) + h(N), which represents the estimated cost from the starting state S to the end state G through the current state N. A* uses the list OPEN to store all the current states to be detected and sort them in ascending order according to the f value of the states. Each time, the state at the top of open is extracted for extension, and the extented states are added to OPEN. The extracted state is placed in the list CLOSE for marking. The algorithm terminates When the state extracted from OPEN is the target state.

## IV. Experiment and Results

### A. Sensor external parameter calibration

Although multiple sensors are in the same system in the multi-sensor fusion problem, they work in their coordinate systems. In the final fusion positioning result, only the coordinate system of one sensor can be selected as the reference coordinate system. What's more, due to hardware triggers, initial delays, clock synchronization errors, *etc.*, the time stamp of each sensor is not synchronized [11]. Considering the influence of various errors aforementioned, it is significant for the positioning system to perform high-precision space-time calibration of multiple sensors.

The joint calibration of lidar and camera is a spatial calibration between multi-sensor fusion. Its primary purpose is to obtain the pose transformation matrix.

A checkerboard method is adopted to calculate the external calibration between radar and camera. First extract the corner points of the camera image, and then select the region of interest (ROI) within a specific range of the checkerboard point cloud. The plane from the ROI area should be extracted to obtain the corners of the point cloud. According to the lidar corner points, camera corner points, and camera internal parameters, the precise pose transformation matrix between the radar and the camera can be obtained through the algorithm.This method requires relatively high requirements for the calibration board. In particular, it requires that the reflection intensity of the laser in the white and black areas of the checkerboard is significantly different. At the same time, the distance between the radar and the camera needs to be close.

Through the obtained external parameters between the radar and the camera, the 3D laser point cloud is projected to the 2D image and color the point cloud corresponding to camera data. If the external parameters are accurate enough, the object will show an apparent edge contour in the colored three-dimensional point cloud, as shown in Fig.3.

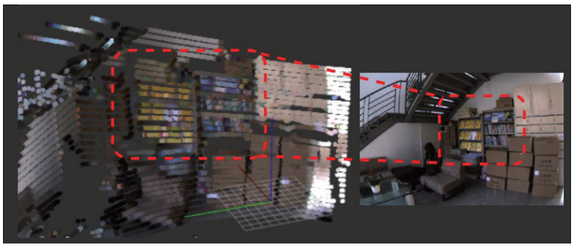


Fig. 3. Result of external parameter calibration.

In addition, our core arithmetic processing module supports external hardware time synchronization, and the synchronization accuracy is better than $1\mu$s.

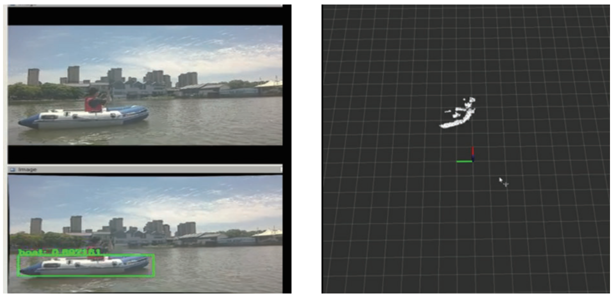
### B. *Real-world experiment*

Fig. 4. The actual water scene of experimental test.

Fig. 5 shows the ability of our perception system to recognize hull targets. The result of framing the target object in the two-dimensional image obtained from the camera is shown in the Fig. 5(a). What's more, the Fig. 5(b) shows that according to the fusion of the camera and lidar, the three-dimensional data points corresponding to the target object are entirely extracted, and the object's external contour can be extracted seen well.

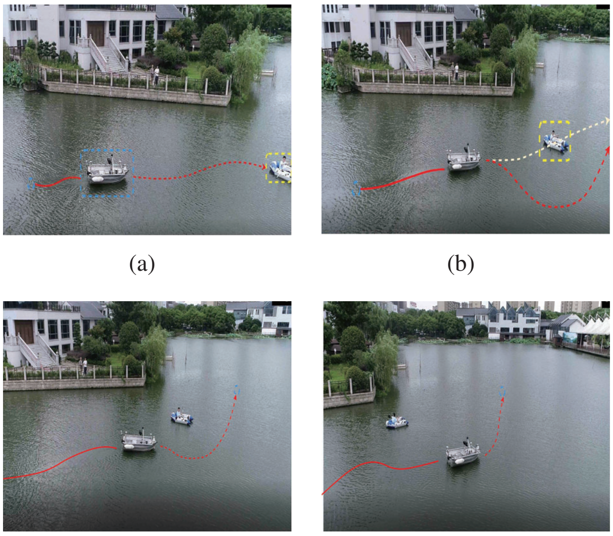
The experimental results of applying the data of surrounding target objects obtained by the sensing system to generate a practical obstacle avoidance path are shown in Fig. 6.

A small sport kayak is defined as a dynamic obstacle for testing obstacle avoidance performance in our obstacle avoidance planning experiment. The Fig. 6(a) presents that the upper layer generates a point-to-point autonomous navigation task based on the accurate



(a) (b)

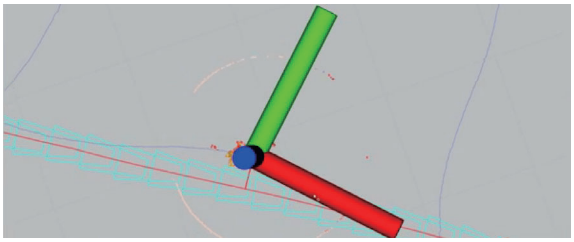
Fig. 5. (a) is the recognition result of the hull target in the two-dimensional image. (b) is the three-dimensional data point of the hull extracted from the laser point cloud according to the target recognition result.



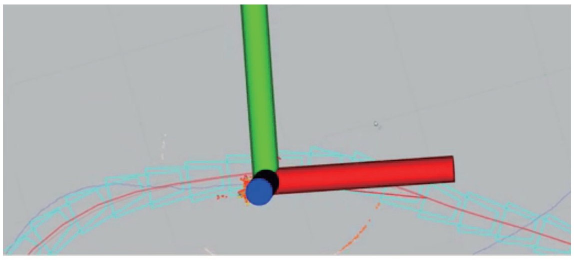
(a) (b) (c) (d)

Fig. 6. (a) is the initial trajectory generated according to the target point. (b) is the target obstacle detected. (c) is to generate a new obstacle avoidance trajectory. (d) is to execute the new trajectory to reach the target point.

position of the unmanned ship. If the perception system does not detect obstacles within the sensing range, the system directly generates a smooth optimal path. In reality, the planning module sends control instructions to the underlying power module to drive the unmanned ship to follow the planned trajectory. The process of regenerating the obstacle avoidance trajectory is shown in Fig. 6(b) - Fig. 6(d). When an obstacle enters the sensing range, the USV system frames the target in the two-dimensional image of the camera. It then extracts the data points corresponding to the kayak from the three-dimensional laser point cloud data. The obstacle avoidance planning system regenerates a new safe obstacle avoidance trajectory based on the obstacles detected in real-time.



(a)



(b)

Fig. 7. (a) is the case of executing according to a predetermined straight line. (b) is the case of executing in accordance with the predetermined turning trajectory. Among them, the red line represents the preset trajectory, while the purple line represents the trajectory executed by the actual unmanned ship.

In addition, due to the limitations of the hull's dynamics, it is difficult for the power system to control the hull to move laterally directly. The experimental results also show that when the hull is moving in a straight line, there is a particular gap between the actual motion trajectory and the planned trajectory due to the interference of the water flow and the insufficient control of the lateral motion(Fig. 7). However, it can also be seen from the results that our planning module has good real-time performance and can control the hull to return to the set trajectory in time. Our experiments prove that the proposed unmanned surface vehicle system can efficiently detect obstacles and recognize target objects. Furthermore, when static or dynamic obstacles appear around the unmanned ship, which causes a risk of collision, our obstacle avoidance planning module can also generate a new safe trajectory in real-time.

## V. Conclusion and Future work

In this paper, we propose a novel USV equipped with high-performance 2D-3D fused perception and obstacle avoidance module. The new module is named Three-Dimensional Perception Module (PMTD). The perception system on USV faces challenges such as water vapor blocking the sensor's view, and inevitable ship shaking. PMTD can handle these challenges in USV's autonomy, so that USV can perceive, detect and avoid static and dynamic obstacles autonomously in the real-world water areas.

## References


[1] Azzeri, M., Adnan, F., Zain, M.M.: Review of course keeping control system for unmanned surface vehicle. Jurnal Teknologi **74**(5) (2015)

[2] Baum, M., Hanebeck, U.D.: Shape tracking of extended objects and group targets with star-convex rhms. In: 14th International Conference on Information Fusion, pp. 1–8. IEEE (2011)

[3] Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)

[4] Chen, Z., Ellis, T., Velastin, S.A.: Vehicle detection, tracking and classification in urban traffic. In: 2012 15th International IEEE Conference on Intelligent Transportation Systems, pp. 951–956. IEEE (2012)

[5] Eppenberger, T., Cesari, G., Dymczyk, M., Siegwart, R., Dubé, R.: Leveraging stereo-camera data for real-time dynamic obstacle detection and tracking. arXiv preprint arXiv:2007.10743 (2020)

[6] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)

[7] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., Berg, A.: Ssd: Single shot multibox detector,‖ in european conference on computer vision (eccv) (2016)

[8] Liu, Z., Zhang, Y., Yu, X., Yuan, C.: Unmanned surface vehicles: An overview of developments and challenges. Annual Reviews in Control **41**, 71–93 (2016). DOI https://doi.org/10.1016/j.arcontrol.2016.04.018

[9] Magnusson, M., Andreasson, H., Nuchter, A., Lilienthal, A.J.: Appearance-based loop detection from 3d laser data using the normal distributions transform. In: 2009 IEEE International Conference on Robotics and Automation, pp. 23–28. IEEE (2009)

[10] Meissner, D., Reuter, S., Dietmayer, K.: Real-time detection and tracking of pedestrians at intersections using a network of laserscanners. In: 2012 IEEE Intelligent Vehicles Symposium, pp. 630–635. IEEE (2012)

[11] Nilsson, J.O., Handel, P.: Time synchronization and temporal ordering of asynchronous sensor measurements of a multi-sensor navigation system. In: Proceedings of IEEE/ION PLANS 2010, pp. 897–902 (2010)

[12] Thompson, D.J.: Maritime object detection, tracking, and classification using lidar and vision-based sensor fusion (2017)

[13] Wang, W., Gheneti, B., Mateos, L.A., Duarte, F., Ratti, C., Rus, D.: Roboat: An autonomous surface vehicle for urban waterways. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 6340–6347. IEEE (2019)

[14] Zhan, W., Xiao, C., Wen, Y., Zhou, C., Yuan, H., Xiu, S., Zhang, Y., Zou, X., Liu, X., Li, Q.: Autonomous visual perception for unmanned surface vehicle navigation in an unknown environment. Sensors **19**(10), 2216 (2019)

[15] Zhang, F., Clarke, D., Knoll, A.: Vehicle detection based on lidar and camera fusion. In: 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp. 1620–1625. IEEE (2014)