IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Research on unmanned surface vehicles environment perception based on the fusion of vision and lidar

Wei, Zhang[1], Feng, Jiang[1], Chi-fu,Yang[1], Zhi-peng, Wang[1],Tie-jun, Zhao[1], Member, IEEE

[1] Harbin Institute of Technology, Harbin, CO 150001 CHN

Corresponding author: (e-mail: fjiang@hit.edu.cn).

This work is partly funded by National Key Research and Development Program of China via grant 2018YFC0806802 and 2018YFC0832105.

ABSTRACT The research of unmanned surface vehicles （usually called USV） is becoming a research hotspot, and real-time and accurate environmental perception is the key technology for achieving autonomous navigation and completing tasks of USV. In this paper, aiming at the technical problems of environmental perception of the USV, from the perspective of sensor multi-information fusion, we build a complete USV perception system (hardware platform and software system). This paper mainly studies objects on the water detection and obstacle avoidance methods based on the information fusion of lidar and vision. In addition, we also studies the method based on morphology to process the reflection of the objects on the water, and the sea-sky detection based on support vector machine (SVM) to assist the objects on the water detection. Through the construction of the USV environment perception system, improving the ability of USV environment perception, and ensuring the safety of autonomous navigation of USV. Our proposed methods have been tested by simulation experiments and actual marine experiment as well as the Hawaii Unmanned Surface Vehicles Challenge, which proves the practicability and stability of our method in the actual environment.

INDEX TERMS Unmanned surface vehicle, Object detection, Sea surface, Autonomous vehicles, Collision avoidance

I. INTRODUCTION

With the comprehensive development of science and technology, unmanned intelligent marine vehicles are developing rapidly and are widely used in civil and military fields. Unmanned surface vehicles (USV) is an intelligent system that is unmanned, relies on remote control or completely autonomous way to navigate on the water surface. It can carry a variety of sensors, special equipment or weapons to perform anti-mine, anti-submarine, anti-ship, maritime security and electronic warfare tasks in the target sea area. As a new type of surface mobile platform that can navigate independently on the water surface, the USV has the ability to complete tasks partially or completely autonomously.

Compared with other conventional marine equipment, the USV has the characteristics of maintenance cost, low energy consumption, and long continuous operation time [1]. It can meet the long-term research tasks and engineering projects in a large area on the water surface. In addition, by carrying different functional modules, USV can replace people in complex and dangerous tasks, such as disaster and accident search and rescue, hydrological information monitoring and collection, marine biological information collection, regional sea chart and terrain drawing, marine weather forecast; adjacent sea defense Tasks; search, detection and demining of specific waters, anti-piracy, anti-terrorism tasks, etc. USV is an autonomous marine vehicle. Its biggest advantage is that it can operate in special complex and dangerous environments or when manned vessels are not suitable for work [2]. Compared with unmanned aerial vehicles, land robots and underwater robots, USV has the advantages of the widest observation space, low cost and the most stable system. The working space of the USV covers the water surface, low altitude and underwater, meeting the requirements of all-round three-dimensional perception of environmental information.

USV is restricted by the working environment when

performing tasks, and often encounter wind and waves, sea fog or high humidity, and other harsh sea conditions. At the same time, USV may encounter high-speed ships, etc. These requirements put forward higher requirements for the environmental awareness of USV [3]. Traditional target information perception technologies mainly include radar information perception, infrared perception, ultrasonic perception, visible light visual perception, and underwater acoustic information perception technology. The optical image obtained by visible light visual perception contains richer target information and background information, which can more conveniently obtain target information in the region through target region segmentation and detection algorithms. The lidar information can accurately obtain the three-dimensional position of the target information and accurately locate the target. Therefore, the fusion of visible light visual perception information and lidar information can use the rich information contained in visible light pictures and the three-dimensional positioning of radar point cloud information. Carrying out research on surface moving target detection and obstacle avoidance technology based on optical vision and lidar information fusion, which will help it complete tasks such as autonomous planning, autonomous collision avoidance and environmental monitoring, so as to avoid collisions between USV and surface targets. On the other hand, it can also ensure the accuracy of the information of the monitoring target, and improve the intelligence level of the USV itself and the ability to perform tasks.

II. RELATED WORKS

Environmental perception technologies based on USV mainly include sea-sky detection, object detection and tracking, and obstacle avoidance technologies. Compared with other unmanned robots, USV environmental perception technology is relatively weak, mainly due to limited conditions and short research time. In this part, we analyze and summarize the characteristics and shortcomings of existing technologies, and introduce the main contributions of our work.

Sea-sky detection. Sea-sky detection can provide very meaningful information. For example, the sea-sky line can be used to solve the calibration problem of a stereo camera; sea-sky detection is sometimes a key step in target detection, which can reduce the search space; sea-sky detection can provide good reference information for binocular ranging. At present, there are many methods about water boundary detection, which are mainly based on the principles of row mapping histogram, gradient transform, wavelet transform, Radon transform, maximum inter-class variance, texture features and Hough transform. Zhao Ningxia, Zhang Bing, Wei Junjie and Pei Lili detect the water boundary by using wavelet transform to the image obtained by the unmanned vehicle [4-6]. Usually, the threshold is used to segment the image, and then the specific position of the sea-sky line in the image is extracted by wavelet transform to avoid a large number of calculations caused by iteration. Wang Bo of Harbin Engineering University proposed a sea-sky line detection method based on gradient saliency [7]. The region growing method is used to realize the detection and identification of the sea antenna, and the accuracy and real-time performance of the method are excellent. At present, the main popular methods are to determine the position of sea antenna by using the assistance of region of interest (ROI). Xiao Zheng of Nanyang University of Technology selects the region of interest through conversion and cutting, and further processes the region of interest to get the location of the sea-sky [8]. In addition, sea-sky line detection is carried out by using the global sparsity of image blocks, but the real-time effect of this method is poor compared with the traditional method, and the practical application effect is not good.

Object detection. For the difficult problem of direct detection of water surface targets, some researchers have combined the specific characteristics of water surface images for target detection. Aiming at the problem of unmanned boat's detection of aquatic targets in a complex coastal background, Wan Lei proposed an automatic detection method of offshore targets based on coastline information [9]. On this basis, Zhang Tiedong et al. proposed a small and weak target detection method without the premise of detecting sea-sky horizons [39]. Chang Li proposed using the concept of saliency to obtain salient features [10]. The accuracy of the algorithm is 82%, and the time per frame is 0.268s. Li Chang et al. aimed at the complex environment faced by rapid detection of water surface moving targets, object uncertainty and viewing angle changes, and a fast water surface moving target detection method based on target surface characteristics [11]. In the later period, Li Chang combined target characteristics and saliency on the basis of two persons, eliminated false targets, obtained the accurate position of the target, and verified that the accuracy rate was over 80%. Wang Han developed a real-time obstacle detection system. The system can detect and locate multiple obstacles within a range of 30 to 300 meters on the sea surface [12]. The later improved algorithm uses high-definition images (2736×2192) to estimate the object distance to a higher accuracy. The rise of deep learning has enriched the detection methods of water surface moving targets. Long Gang et al. used the deep learning framework of the cascaded principal component analysis network to conduct research on the detection algorithm of sea ships, and the effective range of detection was 20 to 200 meters. Yang Jian et al. proposed a neural network-based monitoring and tracking system for surface targets [13]. The use of segmentation accurate detection results solves the problem of low positioning accuracy of current CNN-based detection methods. And using KF in the algorithm can track objects in multiple frames at the same time to improve efficiency.

To sum up, there are typical methods of object detection and sea-sky detection for USV. Sea-sky detection is mainly based on the principles of mapping histogram, gradient

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3057863, IEEE Access

transform, wavelet transform, Radon transform, maximum inter-class variance, texture features and Hough transform. The target detection algorithm is mainly based on the characteristics of water surface background or the method based on deep learning. However, these schemes still have the following shortcomings: (1) Most of the methods are not applied to USV, only the data collected by USV are used for simulation experiments [14-18]; (2) The existing methods have the problem of incompatibility between real-time and robustness, which affects the stability of USV environment perception performance; (3) The environmental perception system needs to have good signal transmission with the control system, power system, and communication system [19-23].

In order to overcome the shortcomings of the above methods, we propose a new method of target detection and obstacle avoidance based on the fusion of image information and lidar information. We apply the USV environment perception system to the USV instead of just using the collected data for simulation experiments. Moreover, our image processing method is based on deep learning, which has the characteristics of high accuracy and high robustness. Our computer is equipped with a GPU to ensure real-time performance. The USV platform we designed is an overall structure. The environmental perception system interacts with other systems, which truly improves the intelligence of the USV. It consists of five modules: surface image denoising module (SIDM), a water surface image quality evaluation index based on image three-channel information is proposed, which can accurately determine image quality. If the image quality is lower than the threshold, the image denoising process is performed. water reflection removal module (WRRM), sea-sky detection module (SSDM), surface target detection module (STDM) and obstacle avoidance module (OAM). The WRRM mainly uses morphological methods to distinguish surface targets and water surface reflections, and then removes the reflections to prevent the reflections from interfering with target detection; The SSDM mainly uses the support vector machine method to classify water and non-water areas, and the separating hyperplane of the two areas is the sea-sky line; In the STDM, we fuse image information and lidar information to detect and locate the target, and obtain the target's category, position and confidence; In the OAM, we mainly use lidar to record the position of obstacles and avoid obstacles.

The main contributions of this paper are summarized as follows:

(1)A morphological-based method for removing water surface reflections is proposed, which can avoid interference with target detection;

(2)Through data labeling and pre-training of SVM-based water and non-aquatic classifiers, sea antennas can be labeled by separating hyperplanes;

(3)Propose a new type of water surface target detection and obstacle avoidance scheme fusing image information and lidar information, which can be applied stably on USV.

III. METHODS

The research is based on "WAM-V-USV" as the experimental platform, which is a complex system platform that integrates many advanced subjects such as mechanical design, electronic communication, environmental perception and ship maneuverability. This part first introduces the architecture of the "WAM-V-USV" test platform, and then designs an environment perception system based on the hardware system and software system of the test platform. The following introduces the principles of the algorithm of the environment perception system: SIDM, WRRM, SSDM, STDM, OAM.

A. "WAM-V-USV" PLATFORM SYSTEM STRUCTURE

1) MAIN STRUCTURAL SYSTEM OF "WAM-V-USV"

"WAM-V-USV" is a small USV, the hardware frame is produced by Marine Advanced Research Company. We later added a power system, a sensing system, a communication system, and a control system, which can be remotely controlled and have a certain degree of autonomy for small USV. In order to further improve safety, we have additionally developed safety indicator lights and emergency brake valves. The appearance view of "WAM-V-USV" is shown in Figure 2 below. The carrier of "WAM-V-USV" is with an inflatable fender on the side. The power unit adopts the propeller propulsion method powered by lithium batteries. The total length is 3.91m and the height is 1.27m and the width is 2.44m.

FIGURE 2. **"WAM-V-USV" structure.**

The structure of the USV defines the interrelationship and function allocation among the various parts of a USV. "WAM-V-USV" is mainly composed of control system, environment perception system, communication system and safety system, power system, etc.:

(1) Control system: This part is mainly responsible for the control strategy of the "WAM-V-USV" test platform, mainly responsible for USV heading, speed control, and autonomous and remote control mode switching. The control platform is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3057863, IEEE Access

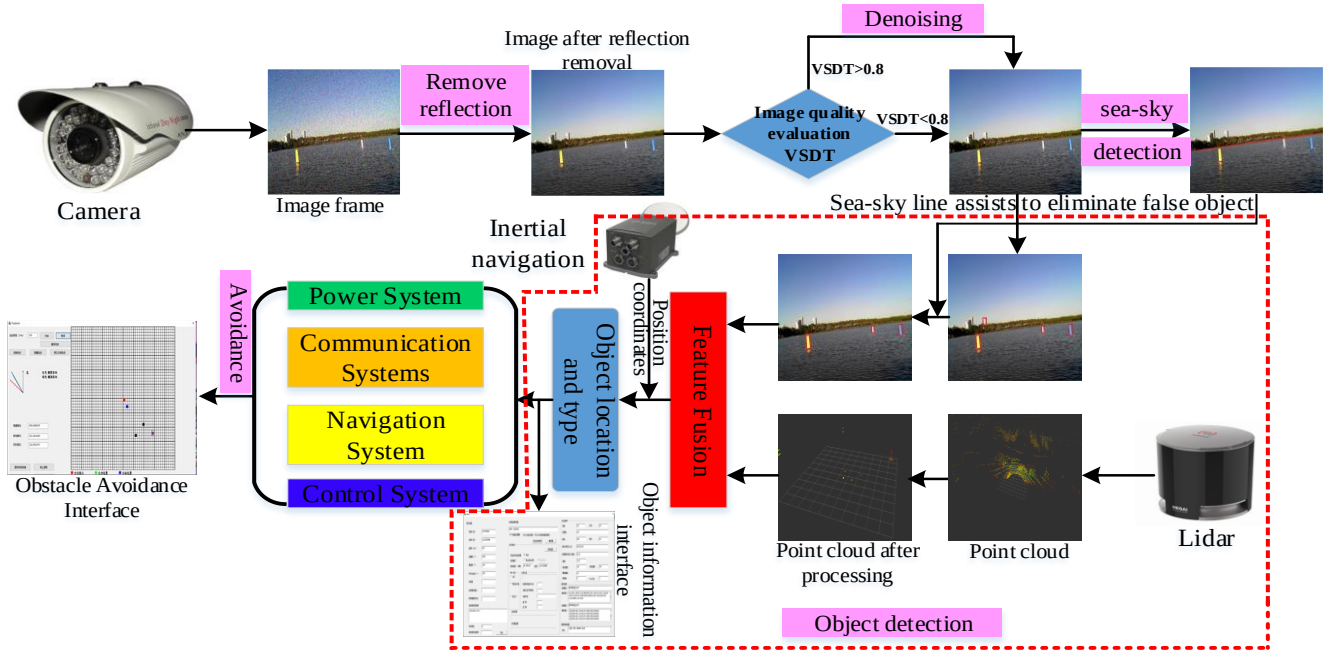


FIGURE 1. **USV environment perception system based on the fusion of vision and lidar. It mainly contains five modules: surface image denoising module (SIDM), water reflection removal module (WRRM), sea-sky detection module (SSDM), surface target detection module (STDM) and obstacle avoidance module (OAM).**

composed of a shore console and a shipboard computer.

(2) Environmental perception system: This part is mainly responsible for collecting the environmental perception data of the "WAM-V-USV" test platform, which consists of the corresponding image processing program, laser radar data processing program and camera, laser radar and processing industrial computer (image processing Board) hardware composition. This platform is equipped with an industrial camera (VCXG-25M.I) based on Gige protocol and Velodyne VLP16 lidar. The computing board is Jetson Tx2 from NVIDIA.

(3) Communication system: The shore console and the onboard computer transmit commands and send and receive data via radio.The internal information transmission of the onboard computer, such as image processing program and lidar processing program, exchange data through TCP/IP.

(4) Safety system: In order to ensure the safety of navigation, we have added a safety system, which mainly refers to the emergency brake valve and safety indicator. The emergency brake valve is located on the side of the boat.

(5) Power system: It mainly includes two lithium batteries and a pair of thrusters, which are responsible for providing control points and power electricity.

2) "WAM-V-USV" ENVIRONMENT PERCEPTION SYSTEM

According to the environmental information collection and processing process, the software architecture of the environment perception system of the "WAM-V-USV" can be divided into five parts. The specific structure diagram is shown in Figure 3.

(1) The information data collection layer mainly includes two parts: image data collected by cameras and point cloud information data collected by laser mines.

(2) Information data preprocessing layer, which mainly includes image de-dusting filtering and lidar point cloud filtering. The laser radar point cloud sampling range is limited to a range of 45 degrees in front of the USV, and then all scattered data points in the point cloud that do not meet the clustering requirements are removed, and finally the target is segmented.

(3) The information data processing layer mainly includes target detection on the preprocessed image to obtain the target type; target extraction on the preprocessed lidar point cloud data to obtain the target position.

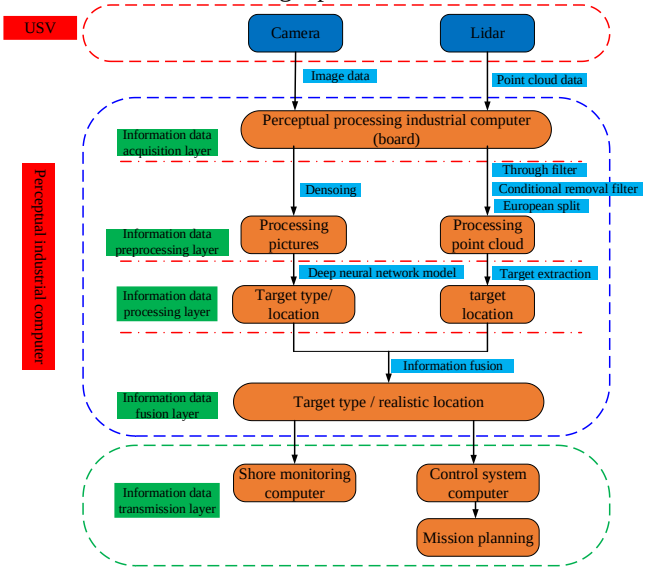


FIGURE 3. **"WAM-V-USV" environment-aware software system.**

(4) The information data fusion layer refers to data fusion between the results of image detection and the results of lidar

point cloud data processing to obtain the type and location of the final target.

(5) The information data transmission layer refers to the transfer of the results obtained by data fusion to the control system and the shore-side monitoring computer.

B. Water reflection removal module

1) THE CHARACTERISTICS OF THE REFLECTION OF THE SURFACE TARGET

The surface target forms a reflection on the surrounding water surface, which will interfere with the subsequent detection and tracking of the surface moving target. It can be analyzed that the reflection of the surface target has the following characteristics:

1. Irregularities. Because of the influence of water waves or ripples, the reflection of the target on the water surface is not exactly the same as the side view of the target, and there is irregular deformation.

2. Light absorption. There is an error between the color of the reflection and the target, and the color of the reflection will be darker than that of the surface target.

3. Variety. The reflection of the surface target's own movement and the influence of waves is constantly changing.

According to the difference between the reflection of the surface target and itself, the surface target and the reflection can be quickly separated in the image.In order to remove irregular water reflections, opencv can be used to perform threshold segmentation, contour detection and other related processing on the image. The specific methods are as follows:

(1) First, convert the input image into a grayscale image according to formula1, and use the mean threshold function, formula 2 and 3 to convert the image into a binary image.

$$Gray_{ij} = \left(R_{ij} * 0.299 + G_{ij} * 0.587 + B_{ij} * 0.114\right) \quad (1)$$

$$m = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} Gray_{ij}}{a * b} \quad (2)$$

$$m_{ij}^{*} = \begin{cases} 0 & Gray_{ij} \le m \\ 255 & Gray > m \end{cases} \quad (3)$$

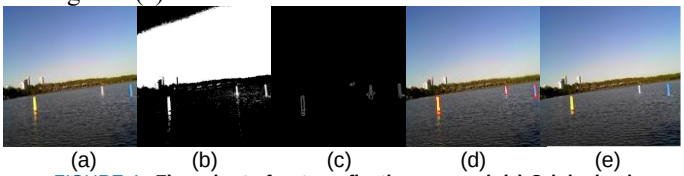
(2) We use the 9-square grid method for contour detection. If there is a difference in 8 pixels around a pixel, it is set as the background, otherwise it is the same as the contour, the calculation formula is shown in (4) and (5). Then use ContourArea and ArcLength functions in Opencv to remove large-scale white spots in the sky and ocean caused by strong light.

$$N_{ij} = \sum_{i=i}^{i=i\pm1}\sum_{j=j}^{j=j\pm1} m_{ij}^{*} \quad (4)$$

$$N_{ij}^{*} = \begin{cases} 255 & if \sum_{i=i}^{i=i\pm1}\sum_{j=j}^{j=j\pm1} m_{ij}^{*} = 0 or 2295 \\ 0 & others \end{cases} \quad (5)$$

(3) Perform the straight line detection operation on Figure 4(c) again, use the Hough Lines function to detect the approximate straight line of the contour, and remove the contour without a straight line, and then project the remaining contour points into the original image, as shown in Figure 4(d), the detection of cylindrical water surface targets has been completed at this time.

(4) First convert the color space in Figure 4(d) to HSV, and then use the point PolygonTest function to determine whether a certain point is within the contour. This function can be used to find all in Figure 4(d) Surrounding pixels outside the outline. Take out the pixel value and determine whether the value meets the range of colors such as yellow, white, and blue in the HSV color space. If it meets (that is, the reflection in the water), let the pixel be replaced by the pixel value of the surrounding sea water, complete The operation to remove the reflection. The final result is shown in Figure 4(e).



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

FIGURE 4. **Flow chart of water reflection removal. (a) Original color image; (b) Binary image; (c) Binary graph contour detection; (d) Straight line detection; (e) The final reflection image is removed.**

C. SURFACE IMAGE DENOISING MODULE

The images collected by the USV sometimes have more noise interference, which affects the image quality, which requires denoising processing to improve the image quality so as not to affect the subsequent detection tasks. However, the denoising task will waste computing resources and time, so we recommend to judge the image quality first. If the image quality is high enough, the denoising step is omitted.

1) IMAGE QUALITY EVALUATION PARAMETERS

Image quality can refer to the accuracy with which different imaging systems capture, process, store, compress, transmit and display the signals that form the image. Traditional methods of objectively evaluating image quality mainly include structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) [24,25]. PSNR is an objective standard for evaluating images. It is the most common objective evaluation method used to evaluate image quality. The higher the PSNR, the higher the image quality. SSIM is a more direct way to compare the difference in information structure between the test image and the reference image.

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left\|I(i,j) - K(i,j)\right\|^{2} \quad (6)$$

$$PSNR = 10 \times \log_{10}\left(\frac{\left(2^{n}-1\right)^{2}}{MSE}\right) \quad (7)$$

$$SSIM(x,y) = \frac{\left(2u_x u_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(u_x^2 + u_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \quad (8)$$

Where μ_x is the average of x, μ_y is the average of y, σ_x^2 is the variance of x, and σ_y^2 is the variance of y, σ_{xy} is the covariance of x and y. $c_1 = \left(k_1 L\right)^2$, $c_2 = \left(k_2 L\right)^2$ is a constant defined for stability.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3057863, IEEE Access

$k_1 = 0.01, k_2 = 003, L$ represents the floating range of the pixel.

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

FIGURE 5. **Original image and noise-added image.(a) Original image; (b) Noise is 0.04; (c) Noise is 0.08; (d) Noise is 0.16; (e) Noise is 0.20.**

We add salt and pepper noise to the original picture with an intensity of 0.02~0.2. It can be seen that as the intensity of the salt and pepper noise increases, the original picture becomes more and more blurred, and the picture quality becomes worse and worse. Through statistics, it is known that PSNR and SSIM are suitable for judging the quality of water surface images.

2) IMAGE QUALITY EVALUATION DESIGN

Image quality attributes mainly include: clarity, resolution, noise, contrast, color accuracy, and whether it is true or not. We, our parameters for image quality judgment are derived from the analysis of image channels. Each image has at least one color channel, and the number of color channels of an image is determined by its color mode. RGB images have 3 color channels, and the three color channels store different color information. Comparing the three-channel difference of two images, you can analyze the difference and quality of the two images. First, the original color image and the image with salt and pepper noise added are separated in three channels, and the standard deviation of the three channel matrices are calculated at the same time, and then the mean value of the standard deviation of the three channels is calculated. Finally, the variance of the standard deviation of the three-channel matrix is obtained, which is our evaluation parameter.

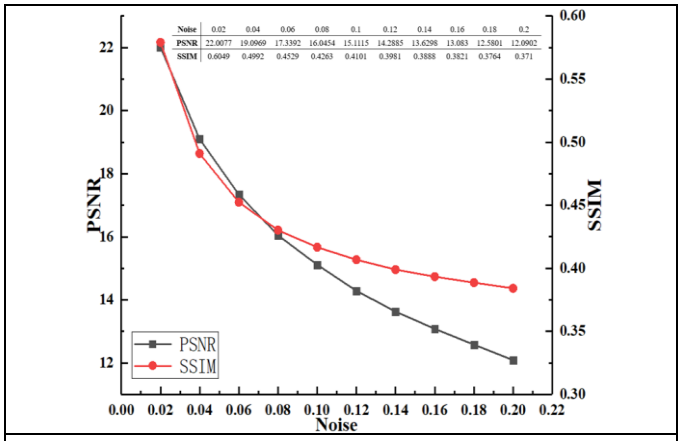
$$\sigma_1 = \sum_{i=1}^{m}\sum_{j=1}^{n}\sqrt{\frac{\left(x_{ij1} - x'_{ij1}\right)^2}{m*n}} \tag{9}$$

$$\sigma_2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\sqrt{\frac{\left(x_{ij2} - x'_{ij2}\right)^2}{m*n}} \tag{10}$$

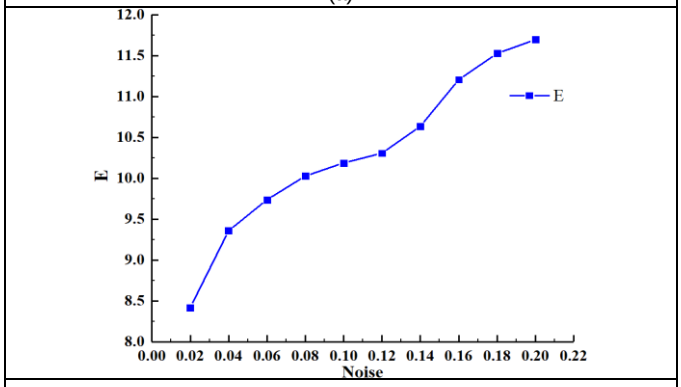
$$\sigma_3 = \sum_{i=1}^{m}\sum_{j=1}^{n}\sqrt{\frac{\left(x_{ij3} - x'_{ij3}\right)^2}{m*n}} \tag{11}$$

$$E = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}{3}} \tag{12}$$

In the above formula, m and n respectively indicate that the dimension of the image is $m*n*3$, x_{ij} indicates the pixel with coordinates (i, j) in the original image, and x'_{ij} indicates the pixel with coordinates (i, j) in the image after adding noise. That is, $\sigma_1, \sigma_2, \sigma_3$ are the standard deviation of the two images of the R,G,B channels. Finally, find the color space difference between the two images, which can also reflect the quality of the images. It can be seen that as the noise intensity increases, the picture quality becomes worse and worse, and the E value becomes larger and larger. It not only meets the comfort of human eyes, but also meets the accuracy of the calculation model, so this method of calculating image quality is effective.



(a)

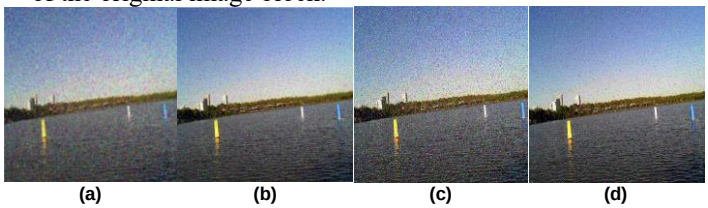


(b)

FIGURE 6. **PSNR,SSIM and E vs noise intensity curve.**

3) IMAGE DENOISING ALGORITHM

When the image quality evaluation parameter $E > 10.5$ (Because when the $PSNR \le 15$, it is recognized that there is a large loss in image quality, and the corresponding noise is 0.1. at this time, E is 10.5.), we believe that the image quality has unacceptable distortion, and it is necessary to remove noise. We choose the best denoising algorithm-BM3D algorithm, which finds two-dimensional image blocks that are similar to the reference block through similarity determination, combines similar blocks into three-dimensional groups, performs collaborative filtering processing on the three-dimensional groups, and aggregates the processing results to the position of the original image block.



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

FIGURE 7. **BM3D algorithm denoising comparison.(A) is an image with a noise of 0.2,PSNR=12.0902,SSIM=0.371,E=11.76;(B) is the image after denoising, PSNR=18.473,SSIM=0.4697,E=9.72 (C) is an image with a noise of 0.12, PSNR=14.2285,SSIM=0.3981,E=10.32 (D) is the image after denoising, PSNR=20.4725,SSIM=0.5183,E=9.41.**

D. SEA-SKY DETECTION MODULE

Generally, the detection, recognition, positioning and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3057863, IEEE Access

tracking of targets in water surface images require sea-sky line to play an auxiliary role, so the extraction of sea-sky line information is of great significance. For example, a sea-sky line can be used to solve the problem of stereo camera calibration; water boundary detection is sometimes a key step in surface target detection, which can narrow the search space. At present, there have been many detection methods for sea-sky line, which are mainly based on the principles of row mapping histogram, gradient transform, wavelet transform, Radon transform, and maximum inter-class variance.

SVM is a two-class classifier. It is trained according to the sample training set, and a relatively optimal hyperplane is obtained through training (as shown in Figure 2-7). This hyperplane can perfectly separate the two types of data, and the distance between the two types of data separated at the same time is the largest among all planes. Support vector machines have the advantages of strong adaptability, optimization, substantial theory, and low training cost [26]. The sea-sky we need is the hyperplane that separates the sky area and the water surface in the water surface image, so the SVM idea can be used for sea-sky detection.

We set the training set as: $S=\{(x_1,y_1)...(x_i,y_i)\}$. Among them, $x_i \in R^n, y_i \in \{-1,1\}$, n is the dimensionality of each training point, that is, the number of features. The goal of the support vector machine is to construct a decision function $f(x)=(\bar{w}\bullet x)+\bar{b}$ that can correctly divide the training data into two parts as much as possible. The final classification hyperplane of the support vector machine is:

$$(\bar{w}\bullet x)+\bar{b}=0 \tag{13}$$

$$\begin{cases}((\bar{w}\bullet x)+\bar{b})\geq 0, y_i=+1\\ ((\bar{w}\bullet x)+\bar{b})\leq 0, y_i=-1\end{cases} i,..1, \tag{14}$$

Among them: $\bar{w}$ is the normal vector of the hyperplane, which can be obtained:

$$y_i((\bar{w}\bullet x)+b\geq 0, i=1,..1 \tag{15}$$

There is ε for a finite number of samples, such that:

$$y_i((\bar{w}\bullet x)+b\geq \varepsilon, i=1,..1 \tag{16}$$

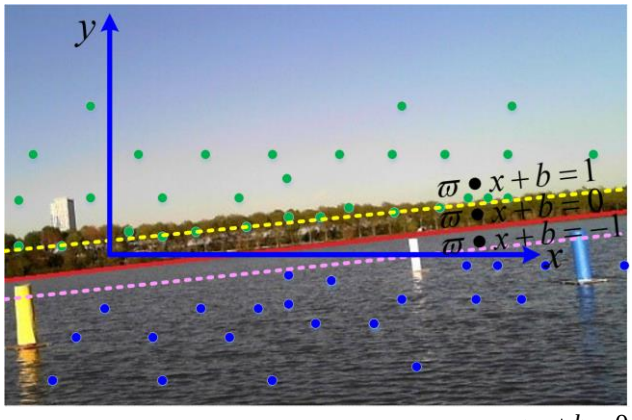
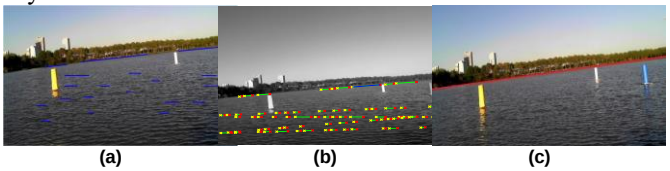


FIGURE 8. **Sea-sky line (Hyperplane, that is the red line $\varpi \bullet x+b=0$) detection method based on SVM.**

We set $w=\frac{\bar{w}}{\varepsilon}, b=\frac{\bar{b}}{\varepsilon}$, we can get: $y_i(\bar{w}\bullet x_i)+\bar{b}\geq 1, \forall_i$.

In order to get the best classification results, we will choose to train such a hyperplane, which can accurately separate the training sample data, and the closest distance to the hyperplane in the two types of training data is the largest. This hyperplane is the optimal hyperplane. This maximum distance is recorded as $\rho(w,b)$, $(w\bullet x)+b=\pm 1$ is called the support hyperplane.

The existing sea-sky detection methods mainly include Hough detection, and based on gradient saliency. we have compared these two methods with our method. It can be seen that the detection method based on gradient saliency detects many interference lines, and detects some waves on the water surface as straight lines, and the real sea-sky cannot be detected correctly. We have performed the longest line segment detection based on the Hough line detection, as shown in the blue line segment in Figure (b), because the sea-sky is the longest line segment in the image. It can be seen that this method also produces many interference line segments, and it cannot accurately detect the sea-sky. Our SVM-based sea-sky detection method can accurately detect the sea-sky, but the disadvantage is that a large amount of data needs to be calibrated in advance, which is not required by the first two methods. The next improvement method is to improve the SVM into multiple classifications, that is, the image is divided into three parts: sky, shore and water surface, which is more useful for sea-sky detection.



(a) (b) (c)

FIGURE 9. **Sea-sky detection comparison experiment.(a) Sea-sky detection based on gradient significance;(b) Sea-sky detection based on Hough;(c) Our sea-sky detection based on SVM.**

E. SEA-SKY DETECTION MODULE

In order to achieve high-quality perception in a complex environment, a variety of sensors must be used to achieve a more comprehensive perception by fusing a variety of different sensor data. In the perception layer of USV systems, the most commonly used sensors are cameras and lidars. Lidar can obtain high-precision depth information. However, only sparse point clouds with limited resolution can be obtained. The camera can obtain high-resolution color and texture information from the environment, but cannot obtain high-precision depth information. Therefore, cameras and lidars are complementary at the data level. In order to achieve high-quality sensor fusion, the calibration of external parameters of the camera and lidar is a very important part. When the precise coordinate system transformation relationship between the camera and the lidar is obtained, can the camera image data and the lidar point cloud data be accurately matched, and data can be merged at various levels.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3057863, IEEE Access

1) JOINT CALIBRATION OF CAMERA AND LIDAR

There are four coordinate systems in the camera model, world coordinate system, camera coordinate system, image physical coordinate system and pixel coordinate system. The world coordinate system, on which the spatial position of the camera and the object to be measured can be described. The position of the world coordinate system can be freely determined according to the actual situation [27]. The camera coordinate system's origin is at the center of the lens, the x and y axes are parallel to the opposite sides of the phase, and the z axis is the lens optical axis, perpendicular to the image plane.

In the coordinate system conversion, the world coordinate system (x_w, y_w, z_w) and the camera coordinate system (x_c, y_c, z_c) can be described by the following formula, where R is the rotation matrix and t is the translation matrix

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{17}$$

Considering that the unit length of the pixel and the unit length of the image coordinate system are scaled by α times and β times respectively in the x direction and the y direction, and the pixel coordinate system takes the upper left corner of the image as the origin, and the image coordinate system has offsets of c_x and c_y in the x direction and y direction, respectively. The relationship between the pixel coordinates and the image coordinates is:

$$\mu = \alpha x + c_x, \nu = \beta y + c_y \tag{18}$$

In summary, the relationship between arbitrary world coordinates and pixel coordinates can be obtained:

$$z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \left(\begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \right) \tag{19}$$

FIGURE 10.**Baumer camera(VCXG-25M.)**

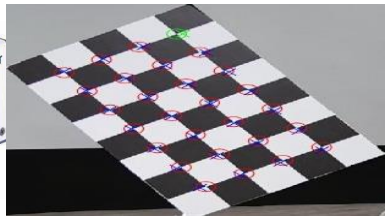


FIGURE 11. **Opencv camera calibration tool.**

According to formula (19), divided by the scale factor Zc, we establish a mapping relationship from world coordinates to pixel planes. For the convenience of calculation, we assume that the plane where the object point is located in the world coordinate system passes through the far point of the world coordinate system and is perpendicular to the Z axis, so that $Zw = 0$. Define the homography matrix as:

$$H = s \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \left(\begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ 0 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \right) = s \begin{bmatrix} R_1 & R_2 & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ 1 \end{bmatrix} \tag{20}$$

Among them, s is the scale factor, R_1, R_2 are the corresponding column vectors in the rotation matrix.

$$\begin{aligned} h_1 &= sKR_1, R_1 = \lambda K^{-1} h_1 \\ h_2 &= sKR_2, R_2 = \lambda K^{-1} h_2 \\ h_3 &= sKt, t = \lambda K^{-1} h_3, \lambda = s^{-1} \end{aligned} \tag{21}$$

According to the property that the length of the rotation vector does not change after rotation, we can get:

$$\|R_1\| = \|R_2\| = 1 \tag{22}$$

$$R_1^T R_1 = R_2^T R_2 \tag{23}$$

$$h_1^T K^{-1^T} K^{-1} h_1 = h_2^T K^{-1^T} K^{-1} h_2 \tag{24}$$

$$B = \left(K^{-1}\right)^T K^{-1} = \begin{bmatrix} \frac{1}{f_x^2} & 0 & -\frac{c_x}{f_x^2} \\ 0 & \frac{1}{f_y^2} & -\frac{c_y}{f_y^2} \\ -\frac{c_x^2}{f_x^2} & -\frac{c_y^2}{f_x^2} & \frac{c_x^2}{f_x^2} + \frac{c_y^2}{f_y^2} + 1 \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} \tag{25}$$

The formula (24) can be simplified to:

$$h_1^T B h_2 = h_2^T B h_2 \tag{26}$$

Since B is a symmetric matrix, so the formula (25) is:

$$h_i^T B h_j = v_{ij}^T b \tag{27}$$

$$v_{ij} = \left[h_{i1}h_{j1}, h_{i1}h_{j2} + h_{i2}h_{j1}, h_{i2}h_{j2}, h_{i3}h_{j1} + h_{i1}h_{j3}, h_{i3}h_{j2} + h_{i2}h_{j3}, h_{i3}h_{j3} \right] \tag{28}$$

Therefore, the formula (27) can be written as:

$$\begin{bmatrix} v_{12}^T \\ (v_{11} - v_{22})^T \end{bmatrix} b = 0 \tag{29}$$

When we provide a certain amount of data, equation (29) has a solution to b, and then the camera internal parameter K can be obtained. After the internal parameters are obtained, the external parameters can be further solved by formula (24).

The image data captured by the camera is represented by a 3-dimensional lattice cloud captured by Lidar. The goal is to create a transformation matrix that maps 3D points to 2D points, namely:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{30}$$

TABLE I

INTERNAL PARAMETER VALUES FOR CALIBRATION OF CAMERA

Camera	f_x	f_y	c_x	c_y
	1289.193642	1289.865271	654.435167	568.754916

TABLE II

THE DISTORTION COEFFICIENT OF THE CAMERA

Camera	p_1	p_2	p_3	p_4	p_5
	-0.10652	0.10764	-0.00862	0.00573	0.17954

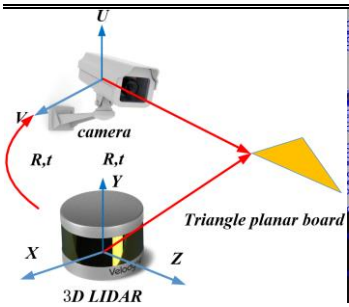


FIGURE 12.**Schematic diagram of the joint calibration of camera and lidar.**

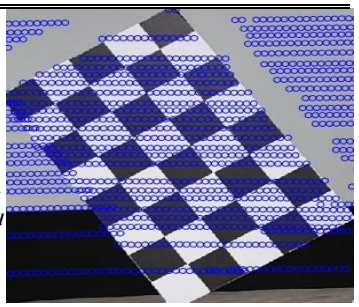


FIGURE 13.**Lidar point cloud and calibration board image matching.**

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.

TABLE III

THE ROTATION MATRIX AND TRANSLATION MATRIX VALUES OBTAINED BY THE JOINT CALIBRATION OF THE CAMERA AND THE LIDAR

R			*t*		
-0.0651	1.5833	-0.0104			
0.3586	-0.0162	-1.2761	-0.0658	0.0198	0.1935
-1.2489	-0.0561	-0.4582			

2)TARGET DETECTION BASED ON INFORMATION FUSION

YOLOv3 is the third version of the YOLO (You Only Look Once) series of target detection algorithms. Compared with the previous algorithms, especially for small targets, the accuracy has been significantly improved. Our surface targets have many small targets, so this algorithm is also suitable for the application environment of USV. The main improvements of YOLO3 are: adjust the network structure; use multi-scale features for object detection; use Logistic to replace softmax for object classification [28].

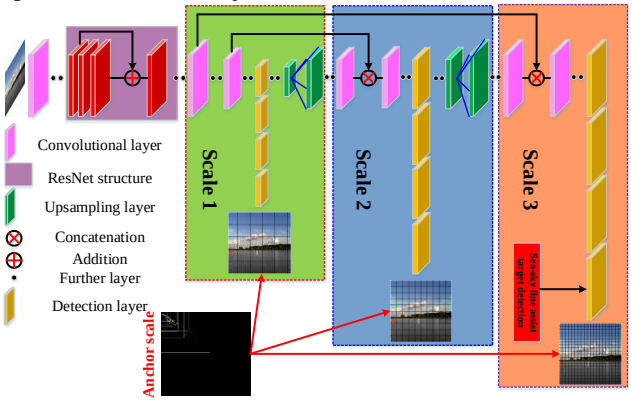


FIGURE 14. **YOLOV3 target detection used in our paper. On the basis of the original YOLOV3, we improve the size of anchor and introduce sea-sky assistance, both of which improve the accuracy of target detection.**

The basic idea of YOLO V3 algorithm can be divided into two parts:

(1)A series of candidate regions are generated on the picture according to certain rules, and then the candidate regions are labeled according to the positional relationship between these candidate regions and the real frame of the object on the picture. Those candidate regions that are sufficiently close to the ground truth box will be marked as positive samples, and the position of the ground truth box will be used as the position target of the positive sample. Those candidate regions that deviate greatly from the true frame will be marked as negative samples, and negative samples do not need to predict the position or category.

(2)Use convolutional neural network to extract image features and predict the location and category of candidate regions. In this way, each prediction box can be regarded as a sample, and the label value is obtained according to the position and category of the real box relative to it. The network model predicts its location and category, and compares the network prediction value with the label value to establish a loss function.

Datasets. We produce16 types of surface targets, including red, green, blue, black, white, and yellow surface buoys, red, green, and blue triangle signs, red, green, and blue circle signs, and red, green, and blue cross signs. Surface floats and so on. Firstly, image data is collected, and video data is collected using the camera carried by the USV. Pass the video data through FFMPEG to obtain pictures that meet the requirements, the number of pictures is 8000, and the picture size is 1920*1080; then use Labelimg to label the collected 8000 pictures. And according to the difficulty of the target, calibrate the difficult nature, and generate the data set in VOC format.

Data enhancement. In order to improve the generalization ability and the robustness of the model, so we increase the amount of training data. We change the brightness of the image and adjust it to 0.5, 0.8 and 1.2 times of the original image, as shown in the following figure (b), (c) and (d); rotate the image, change the angle of the target in the image, respectively, 5, -5, 10, -10, the scores are (e), (f),, (g) and (h); mirror operation, symmetrical operation of the position of the target in the image, as shown in figure (I);change the size of the target in the image and zoom in, as shown in figure (m);change the contrast and increase the contrast of the image, as shown in figure (n);change the hue of the image and adjust the saturation of the image, as shown in (o). Therefore, the data set has increased from the original 8000 images to the current 96000 images.

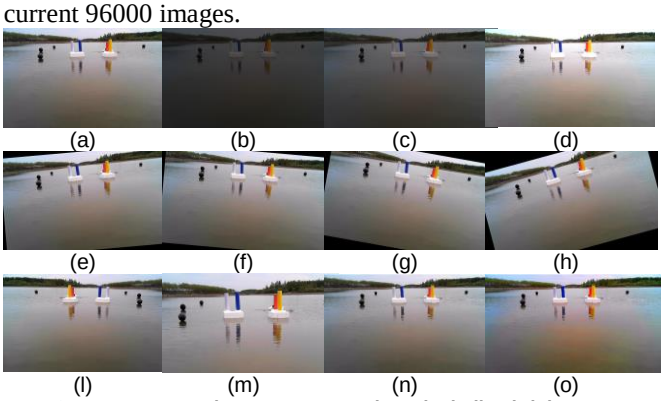


FIGURE 15. **Data enhancement operations, including brightness, rotation, mirroring, scale, contrast, hue changes, etc.**

Generate new anchors. The size of the target in our surface environment is different from the VOC and COCO data sets, so if we continue to use the default anchor size in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3057863, IEEE Access

the algorithm for target detection, there will be errors, so we recalculate the anchor size for the target in the USV specific environment, the calculation method adopts the K-means method. The specific calculation steps are as follows:

Step 1: We need to extract all bounding box coordinates, extract all rectangular boxes of all pictures, and put them together.

Step 2: Data processing to obtain the width and height data of all the training data bounding boxes. The training data given is often the 4 coordinates of its bounding box, but we need to convert the coordinate data to the width and height of the box. The calculation method is as follows:

$$\begin{aligned} w &= anchor_{width} \times input_{width} / downsamples \\ h &= anchor_{height} \times input_{height} / downsamples \end{aligned} \tag{31}$$

Length = abscissa of lower right corner-abscissa of upper left corner, width = ordinate of lower right corner-ordinate of upper left corner.

Step 3: Initialize k anchor boxes, and randomly select k values from all the bounding boxes as the initial values of the k anchor boxes.

Step 4: Calculate the IOU value of each bounding box and each anchor box.

The original official default anchor size is (10, 13 16, 30 33, 23 30, 61 62, 45 59, 119 116, 90 156, 198 373, 326). After K-means calculation [29], set the input network image size to 416*416, get new anchors, the size is (12.2742 11.9937, 29.3959 26.8536, 43.8952 41.6611, 53.2567 64.8536, 72.9625 53.5407, 77.5597 86.2044, 126.3383 85.6498, 104.5958 119.1917, 224.8133 184.2880)

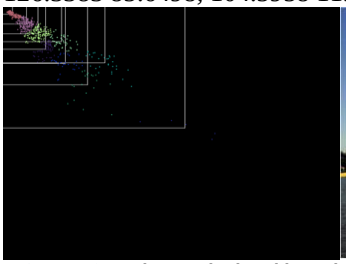


FIGURE 16. **Anchors calculated based on K-means.**

FIGURE 17. **False targets that may be detected in target detection.**

Sea-sky line assists in eliminating false positive targets. Water surface target detection is different from natural image target detection in that the targets are distributed on the water surface, that is, below the sea-sky line. This is the prior experience of surface target detection, which is why the sea-sky detection is performed before surface target detection. Therefore, many false targets can be removed based on this prior experience, as follows:

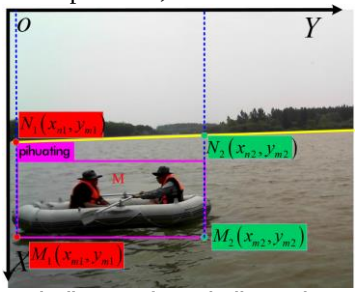


FIGURE 18. **Schematic diagram of sea-sky line assists target detection to eliminate false targets.**

Since we know the position of the sea-sky line according to the SSDM, we can get the two-point coordinates of the two possible maximum X coordinates x_{m1}, x_{m2} according to the target detection bounding box (such as the position of the red point and the green point on the detection frame), and also get the two points $N_1(x_{n1}, y_{m1}), N_1(x_{n1}, y_{m1})$ on the sea-sky line corresponding to the Y coordinate value, and some of the surface targets must exist below the sea-sky line, that is, they exist, so the judgment formula is as follows:

$$M = \begin{cases} True & x_{n1} \le x_{m1} \ or \quad x_{n2} \le x_{m2} \\ False & others \end{cases} \tag{32}$$

Training. Train on our data set and set different parameters for training. When the input network image size is 416*416, batch=32, subdivisions=8, the learning rate is set to 0.001, the anchor point is the official default size, and the number of training times is set to 70,000. The loss curve is as follows: (1)Classic YOLOV3;(2) use K-means to generate new anchors instead of default anchors;(3)use sea-sky to help eliminate false targets;(4)combine K-means to generate new anchors and sea-sky to assist with false targets. The model trained by the four methods calculates the mean average precision (MAP)on the test set, and the original YOLOV3 is 0.834, the YOLOV3+anchors is 0.846, the YOLOV3+sea-sky is 0.843 and the YOLOV3+anchors+sea-sky is 0.849. Therefore, it can be seen that the two auxiliary target detection methods are effective, and the maximum increase is 1.5%.

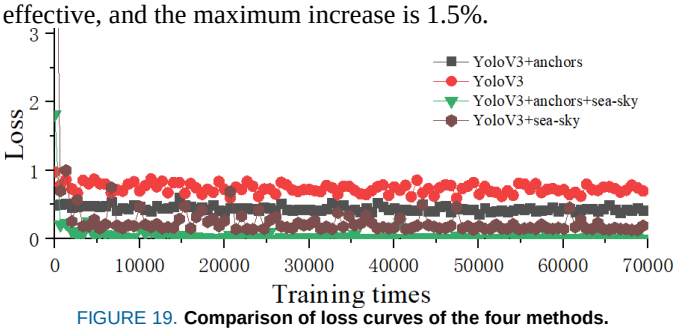


FIGURE 19. **Comparison of loss curves of the four methods.**

TABLE IV
YOLOV3 AND YOLOV3+ANCHORS, YOLOV3+SEA-SKY AND YOLOV3+ANCHORS+SEA-SKY FOUR METHODS MAP COMPARISON.

Methods	YOLOV3	YOLOV3+A	YOLOV3+S	YOLOV3+A+S
MAP	0.834	0.846	0.843	0.849

Lidar data processing. The point cloud image of the lidar is processed to obtain the point cloud cluster information of the surface target. The specific segmentation target steps are as follows:

(1) The lidar uses 16 threads. After actual water tests, the test distance is set to 60m as the longest distance.

(2) Use a straight-pass filter to limit the sampling range to 45 degrees on the port and starboard sides of the USV, reduce the number of data points that need to be processed and reduce the amount of calculation to achieve real-time requirements, which is beneficial to the angle of view of 90 degrees with the camera.

(3) Then use Conditional Removal filter to delete all scattered data points in the input point cloud that do not meet the clustering requirements. Due to the particularity of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3057863, IEEE Access

the water surface environment and the impact of shore objects when the USV is near the shore, the point cloud density distribution obtained by lidar scanning is uneven, and the point cloud data under the water surface environment is more sparse. And there are scattered interference points. This will cause interference to point cloud feature recognition and point cloud clustering, resulting in missed or false target detection. The Conditional Removal filter analyzes the neighborhood of each point.

(4) Calculate the Euclidean distance between all neighboring points of a certain point, and identify the points with too few neighboring points as outlier interference points and remove them from the data set.

(5) This removes scattered data points on the water surface. Then input the filtered point cloud data into Kd-Tree to simplify the calculation. Kd-Tree is a high-dimensional index tree data structure developed from BST. It can save the time required for clustering and meet the requirements of real-time detection.

(6) Then use Euclidean clustering to segment the target object. The feature points are extracted and the 3D points in the depth point cloud image are projected to the 2D image coordinate system.

(7) By searching the neighboring points in Kd-Tree, based on the distance threshold from a certain point to other neighboring points, segment the filtered point cloud to obtain the target. At the same time get the bounding boxes of the segmented target, and add the label B_j

Image and lidar information fusion. Image information and lidar point cloud information are fused. Multi-sensor information fusion includes the unity of space and time. The GPS frequency on the USV is 6, the camera frequency is 20, and the lidar frequency is 15. We take the same image frame and point cloud data in the time dimension.

In order to ensure the uniformity of lidar data and image data in space, we use the following methods, which mainly includes the following steps:

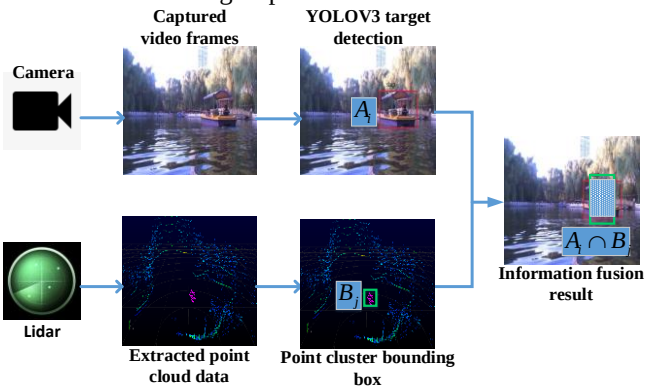


FIGURE 20. **Image and point cloud two-dimensional bounding boxes fusion process.**

(1) Project the obtained target point cloud cluster and the bounding boxes surrounding it onto the two-dimensional image plane to obtain the two-dimensional bounding boxes of the point cloud;

(2) Then calculate the two-dimensional bounding boxes and image bounding boxes A_i IOU of all point clouds,

(3) Obtain the two-dimensional bounding boxes B_j of the point cloud with the maximum value of the bounding boxes A_i IOU of the image target detection frame;

(4) Obtain the A_i and B_j obtained in the above steps, and merge the information of the two, that is, the target type of the target detection frame A_i , and the B_j point cloud detection frame includes the distance and orientation of the target from the USV.

F. OBSTACLE AVOIDANCE MODULE

In the marine environment with unknown dynamics, USV will inevitably encounter obstacles during autonomous navigation. USV can identify and avoid obstacles based on data collected by lidar and cameras. According to STDM, get the type of surface target, the distance and direction angle relative to USV. According to the electronic compass and inertial navigation, we can collect the USV's latitude and longitude position information and heading angle in real time. According to the distance and direction angle of the obstacle relative to the USV, we use the longitude and latitude calculation formula to calculate the longitude and latitude coordinates of the obstacle. Therefore, we save the types of obstacles, latitude and longitude coordinates in the perception map we constructed.

Suppose the positions of USV U and obstacle A are as shown in the figure below. The USV's own coordinate system $X_bO_bY_b$ is the right-handed coordinate system, and the geodetic coordinate system $X_eO_eY_e$ is the true north and true east coordinate system. According to the inertial navigation, the longitude and latitude of the USV is $D_U(J_U,W_U)$, and according to the formula (36), the longitude and latitude of the obstacle $D_A(J_A,W_A)$ can be obtained.

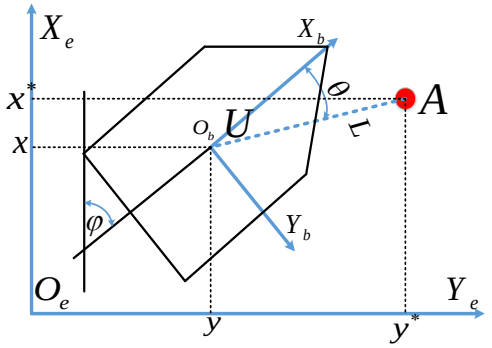


FIGURE 21. **USV, obstacles and geodetic coordinate system.**

$$J_A = J_U + \frac{\left[L * \sin(\theta * pi/180)\right]}{\left[111 * \cos(W_U * pi/180)\right]}$$
$$W_A = W_U + \frac{\left[L * \cos(\theta * pi/180)\right]}{111} \quad (34)$$

We stipulate that the distance between USV U and obstacle A is L, in km . $(J_U,W_U),(J_A,W_A)$ are all angles,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3057863, IEEE Access

the direction angle θ is the angle, and clockwise in the north direction is positive. At the same longitude, the latitude differs every other degree by 111 km; at the same latitude, the longitude differs every other degree by $111*\cos(latitude\ of\ the\ point)km$; the latitude distance difference between two points at the same longitude is $L*\cos(\theta*pi/180)$; the longitude distance difference between two points at the same latitude is $L*\sin(\theta*pi/180)$. So the offset in longitude is $\left[L*\sin(\theta*pi/180)\right]/\left[111*\cos(W_b*pi/180)\right]$; the offset degree in latitude is $\left[L*\cos(\theta*pi/180)\right]/111$.

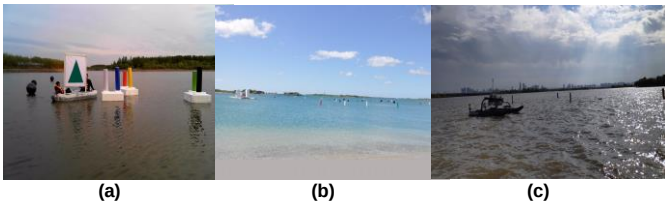
Based on the above information, we can construct a USV environment perception map to mark the types and real-time locations of targets within the perception range. Based on the above information, we can construct a USV environment perception map to mark the types and real-time locations of targets within the perception range. We add the distance, direction angle and target type of all targets relative to USV to a matrix M. According to formula (36), a USV environmental perception map (UEPM) matrix N can be calculated.

$$M*D_U=\begin{bmatrix}(\theta_1,L_1,C_1)\\(\theta_2,L_2,C_2)\\\vdots\\(\theta_n,L_n,C_n)\end{bmatrix}\left[(J_u,W_u)\right]=N=\begin{bmatrix}(J_1,W_1,C_1)\\(J_2,W_2,C_2)\\\vdots\\(J_n,W_n,C_n)\end{bmatrix} \quad (35)$$

After we have obtained the latitude and longitude coordinates and types of all targets above, we set different safety collision radii (3m for the sphere, 5m for the buoy, 6m for the kayak, etc.) according to the type, and finally, we use the A* algorithm to plan the path for collision avoidance.

IV. EXPERIMENT

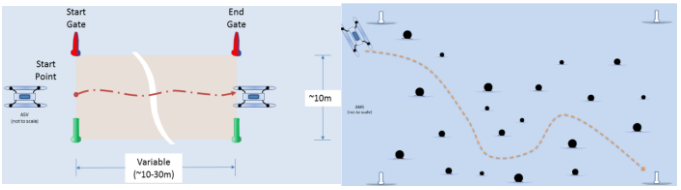
To test our algorithm, we conducted tests in Songhua River, China and Hawaii, USA. After completing the algorithm verification and the establishment of the USV platform, "WAM-V-USV" conducted an 80-day field test on the Songhua River in China and a 20-day Unmanned Surface Vehicles Challenge in the Hawaiian waters of the United States. We have produced 16 surface targets. The specific model is shown in Figure 24(a) below. According to the experimental arrangement, the target detection and tracking, intelligent navigation test and autonomous obstacle avoidance test were carried out.



(a) (b) (c)

FIGURE 22. **FUSV marine test.(a) Model on Songhua River;(b) Hawaii competition venues;(c) USV is on an obstacle avoidance mission.**

We set up the following tasks to test the target detection and obstacle avoidance performance of USV: USV recognizes the door and passes it automatically; USV passes through the obstacle area; USV surrounds different types of goal posts. The schematic diagram of the three tasks are shown blow:



(a) (b)

FIGURE 23. **Schematic diagram of two tasks. (a).USV recognizes the door and passes it; (b).USV automatically passes through the obstacle area.**

The first task is the task of passing the door, as shown in the figure (a) above, the first task is the task of passing the door, as shown in the figure (a) above, specifically, it is to identify and locate the first red and green post, after USV enters the first gate, then identify and locate the second goalpost, and go through, after successfully stepping out of the second door, the test is considered successful. Our strategy is to detect the red and green pillars of the first door, obtain the latitude and longitude coordinates $A(J_A,W_A)$ and $B(J_B,W_B)$ of the two pillars according to UEPM, and set the first target point of USV as the midpoint $E(J_E,W_E)$ between A and B, , after reaching the first target point E, continue to move forward. Then after detecting the red goalpost and the green goalpost of the second goalpost, the latitude and longitude coordinates of the two are $C(J_C,W_C)$ and $D(J_D,W_D)$,set the second target point of USV as the midpoint $F(J_F,W_F)$ between C and D.

In the experiment, we set the latitude and longitude position of the dock as the origin O,The picture above shows our shore console interface, we marked the location of the identification and location on the map, and carried out the door test according
to the above strategy. The red and green in the picture are the goal posts, the blue point is the target point, and the real-time position of the USV is marked by the model.

The second task is through the obstacle area. We identify and mark all the detected and located obstacles, mark them in our UEPM, and use the A* algorithm for path planning.

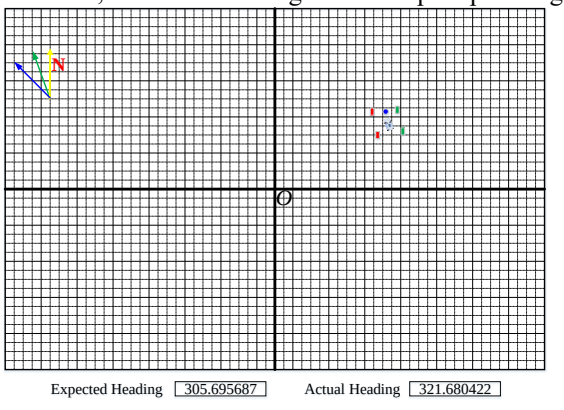


FIGURE 24. **In the task of passing the door, the position of the USV, the door and the target point are divided into alternate models, and the red and green signs and blue points are marked.**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3057863, IEEE Access

FIGURE 25. **Through the obstacle area task, spherical obstacles are marked with black dots, and USV is marked with models.**

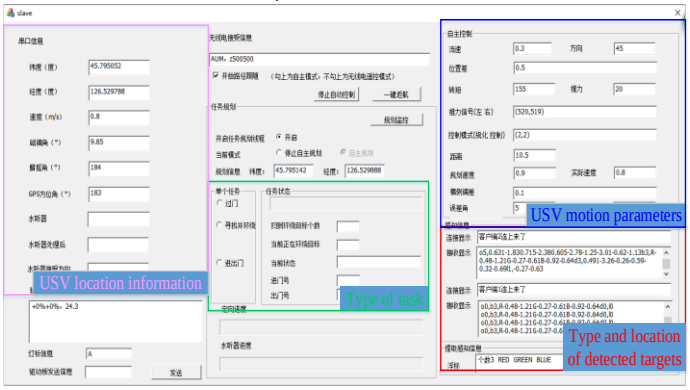


FIGURE 26. **The main page of the shore control box. Mainly include USV position information, posture information, task information, movement parameters and environmental perception information, etc.**

Tested in various environments, including rainy weather, cloudy weather and other bad weather, randomly selected 500 test pictures, the test results are as follows:

(a) (b)

FIGURE 27. **Our USV perform tasks in Unmanned Surface Vehicles Challenge in Hawaii. (a). USV detects and locates the post to complete the task; (b). USV autonomously passes through obstacle areas.**

TABLE V
FRAME DETECTION ACCURACY BASED ON INFORMATION FUSION

Weather conditions	rainy day	cloudy day	sunny day
Randomly selected frames	126	145	229
Accuracy	86.9%	90.2%	92.8%
Average accuracy		90.84%	
Only viison	89.9%	91.7%	93.2%

V. CONCLUSIONS

Aiming at the technical problems of USV environment perception, this paper builds a complete USV perception system from the perspective of multi-sensor information fusion. Because the water surface image is different from the natural image, we have verified through experiments that the PSNR and SSIM quality evaluation parameters are suitable for the water surface image. In addition, according to the multi-channel characteristics of the water surface image, we define an image quality evaluation parameter E, and then calculate it according to the value of E determine whether to use BM3D noise reduction operation. We also studied the method based on morphology to deal with the reflection of the object on the water surface, and studied the sea and sky detection based on SVM, and the experiment proved that it is better than the traditional method. And it assists the detection of surface targets. According to the characteristics of the surface targets, we eliminate false positive targets and improve the performance of target detection. Through the joint calibration of the camera and the lidar, we fused the processed point cloud and image information, and proposed a multi-modal data fusion method. Based on this, the environment perception map was constructed to assist the USV to complete Obstacle avoidance task. After testing in Songhua River, China and participating in the U.S. Unmanned Boat Challenge in Hawaii, it is proved that the information fusion algorithm based on this paper has good detection effect and obstacle avoidance function, and can meet the real-time and practicality and stability in the actual environment.

ACKNOWLEDGMENT

This work is partly funded by National Key Research and Development Program of China via grant 2018YFC0806802 and 2018YFC0832105.

References

[1]J. D. Koch, M. D. Smith, and R. Telang, “Camcording and film piracy in asia-pacific economic cooperation economies,” International Intellectual Property Institute, pp. 7–8, 2011.
[2] M. E. Shiffler and L. W. Loy, “Unmanned sea surface vehicle having a personal watercraft hull form,” Feb. 3 1998, uS Patent 5,713,293.
[3] Z. Liu, Y. Zhang, X. Yu, and C. Yuan, “Unmanned surface vehicles: An overview of developments and challenges,” Annual Reviews in Control, vol. 41, pp. 71–93, 2016.
[4]R.-j. Yan, S. Pang, H.-b. Sun, and Y -j. Pang, “Development and missions of unmanned surface vehicle,” Journal of Marine Science and Application, vol. 9, no. 4, pp. 451–457, 2010.
[5] M. Caccia, M. Bibuli, R. Bono, and G. Bruzzone, “Basic navigation, guidance and control of an unmanned surface vehicle,” Autonomous Robots, vol. 25, no. 4, pp. 349–365, 2008.
[6]M. Bibuli, G. Bruzzone, M. Caccia, and L. Lapierre, “Path-following algorithms and experiments for an unmanned surface vehicle,” Journal of Field Robotics, vol. 26, no. 8, pp. 669–688, 2009.
[7]W. Zhang, C.-f. Yang, F. Jiang, X.-z. Gao, and K. Yang, “A water surface moving target detection based on information fusion using deep learning,” in Journal of Physics: Conference Series, vol. 1606, no. 1. IOP Publishing, 2020, p. 012020.
[8] J. Larson, M. Bruch, R. Halterman, J. Rogers, and R. Webster, “ Advances in autonomous obstacle avoidance for unmanned surface vehicles,” SPACE AND NA V AL W ARFARE SYSTEMS CENTER SAN DIEGO CA, Tech. Rep., 2007.
[9]X. Mou and H. Wang, “Wide-baseline stereo-based obstacle mapping for unmanned surface vehicles,” Sensors, vol. 18, no. 4, p. 1085, 2018.
[10]D. Hermann, R. Galeazzi, J. C. Andersen, and M. Blanke, “Smart sensor based obstacle detection for high-speed unmanned surface vehicle,” IF AC-PapersOnLine, vol. 48, no. 16, pp. 190–197, 2015.
[11]X. Zhang, H. Wang, and W. Cheng, “Vessel detection and classification fusing radar and vision data,” in 2017 Seventh International Conference on Information Science and Technology (ICIST). IEEE, 2017, pp. 474–479.
[12]Q. Liu, X. Xie, and P. Fan, “A fast method for obtaining the region of interest of coastal infrared ship,” in 2016 8th International Confer-ence on

Intelligent Human-Machine Systems and Cybernetics (IHMSC),vol.2. IEEE, 2016, pp. 236–238.

[13] B.-S. Shin, X. Mou, W. Mou, and H. Wang, "Vision-based navigation of an unmanned surface vehicle with object detection and tracking abilities," Machine Vision and Applications, vol. 29, no. 1, pp. 95–112, 2018.

[14] N. Wang, B. Li, Q. Xu, and Y. Wang, "Automatic ship detection in optical remote sensing images based on anomaly detection and spp-pcanet," Remote Sensing, vol. 11, no. 1, p. 47, 2019.

[15] C. Li, Z. Cao, Y. Xiao, and Z. Fang, "Fast object detection from unmanned surface vehicles via objectness and saliency," in 2015 Chinese Automation Congress (CAC). IEEE, 2015, pp. 500–505.

[16] X. Wang, A. Shrivastava, and A. Gupta, "A-fast-rcnn: Hard positive generation via adversary for object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2606–2615.

[17] L. Chen, Y. Liang, and K. Wang, "Inspection of rail surface defect based on machine vision system," in The 2nd International Conference on Information Science and Engineering. IEEE, 2010, pp. 3793–3796.

[18] C. Persello and L. Bruzzone, "Active learning for domain adaptation in the supervised classification of remote sensing images," IEEE Transactions on Geoscience and Remote Sensing, vol. 50, no. 11, pp. 4468–4483, 2012.

[19] R. Guiwen, "Research on dual camera calibration method based on opencv [j]," Science Technology and Engineering, vol. 16, no. 03, pp. 211–214, 2016.

[20] J. Xiangkui and J. Xu, "Design and implementation of camera calibration system based on opencv and matlab," Computer & Digital Engineering, no. 8, p. 34, 2015.

[21] C. Gu, G. Wang, Y. Li, T. Inoue, and C. Li, "A hybrid radar-camera sensing system with phase compensation for random body movement cancellation in doppler vital sign detection," IEEE Transactions on Microwave Theory and Techniques, vol. 61, no. 12, pp. 4678–4688, 2013.

[22] M. Zyczkowski, N. Palka, T. Trzcinski, R. Dulski, M. Kastek, and P. Trzaskawka, "Integrated radar-camera security system: experimental results," in Radar Sensor Technology XV, vol. 8021. International Society for Optics and Photonics, 2011, p. 80211U.

[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances in neural information processing systems, 2015, pp. 91–99.

[24] A. Salvador, X. Giró-i Nieto, F. Marqués, and S. Satoh, "Faster r-cnn features for instance search," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2016, pp. 9–16.

[25] K. Shi, H. Bao, and N. Ma, "Forward vehicle detection based on incremental learning and fast r-cnn," in 2017 13th International Conference on Computational Intelligence and Security (CIS). IEEE, 2017, pp. 73–76.

[26] J. Jiao, Y. Zhang, H. Sun, X. Yang, X. Gao, W. Hong, K. Fu, and X. Sun, "A densely connected end-to-end neural network for multiscale and multiscene sar ship detection," IEEE Access, vol. 6, pp. 20 881–20 892, 2018.

[27] X. Chen and A. Gupta, "An implementation of faster rcnn with study for region sampling," arXiv preprint arXiv:1702.02138, 2017.

[28] H. Wang, X. Mou, W. Mou, S. Y uan, S. Ulun, S. Yang, and B.-S. Shin, "Vision based long range object detection and tracking for unmanned surface vehicle," in 2015 IEEE 7th International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM). IEEE, 2015, pp. 101–105.

[29] W. Zhang, X.-z. Gao, C.-f. Yang, F. Jiang, and Z.-y. Chen, "A object detection and tracking method for security in intelligence of unmanned surface vehicles," Journal of Ambient Intelligence and Humanized Computing, pp. 1–13, 2020.

[30] J. Yang, Y .Xiao, Z. Fang, N. Zhang, L. Wang, and T. Li, "An object detection and tracking system for unmanned surface vehicles," in Target and Background Signatures III, vol. 10432. International Society for Optics and Photonics, 2017, p. 104320R.