IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 27, 2021, accepted April 9, 2021, date of publication April 13, 2021, date of current version April 20, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3072897

Object Detection-Tracking Algorithm for Unmanned Surface Vehicles Based on a Radar-Photoelectric System

QINGZE YU, BO WANG, AND YUMIN SU
Science and Technology on Underwater Vehicle Laboratory, Harbin Engineering University, Harbin 150001, China

Corresponding author: Bo Wang (wangboheu@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51709065, and in part by the Natural Science Foundation of Heilongjiang Province, China, under Grant JJ2020LH1535.

ABSTRACT Object tracking is an important basis for the autonomous navigation of unmanned surface vehicles. However, several problems still must be addressed for a wide applicating of object tracking in unmanned surface vehicles. First, if multiple objects of the same classification exist in the same field of view, then stable extraction of an object is difficult. Second, in an environment with a complex background and large changes in object shape, the tracking accuracy is low, and object tracking errors and tracking loss can easily occur. Third, much time is required to detect a high-resolution real-time video stream, not meeting the delay requirement of the photoelectric servo stable tracking. To resolve these problems, this paper proposes an object detection-tracking algorithm based on a radar-photoelectric system. The algorithm combines an object detection algorithm with an object tracking algorithm and involves the following steps. First, a first-frame object extraction algorithm is used to extract the tracking object from the first frame. Second, a region of interest (ROI)-prediction algorithm is used to predict ROIs and detect objects in these ROIs. This algorithm can effectively solve the above problems in marine tests. When multiple objects of the same classification exist in the same field of view, the algorithm can extract the radar-guided object stably. When faced with a complex background and a large change in object shape, the algorithm substantially improves the accuracy and robustness of object tracking. Compared with the conventional object detection algorithm, the time consumption of this algorithm is reduced by 25.8%.

INDEX TERMS Unmanned surface vehicle, radar-photoelectric system, first-frame object extraction algorithm, ROI-prediction algorithm, object detection algorithm, object tracking algorithm.

I. INTRODUCTION

An unmanned surface vehicle (USV) [1], [2] is a ship that can be operated by a monitoring center onshore and a program control device onboard. It is completely or intermittently controlled by the onboard computer. Compared with manned ships, unmanned ships are more suitable for high-risk, high-repetition and harsh environments. Based on their application field, unmanned ships can be divided into military and civil ships. In the military, USVs are mainly used in antisubmarine, coastal defense, mine detection [3], pirate attack and other tasks. For civil use, USVs are mainly used in hydrological monitoring, maritime search and rescue, seabed exploration [4] and other tasks. Fig. 1 shows the ''Tianxing-1'' USV used in this paper.

The associate editor coordinating the review of this manuscript and approving it for publication was Ehsan Asadi.

A. USV PERCEPTION SYSTEM

The USV perception system includes a perception computer, marine radar, a photoelectric device and other equipment. The perception computer includes a communication module, tracking module and decision module. The CPU model of the perception computer is i7-7700t, and the GPU model is 2080ti. The resolution of the photoelectric visible light sensor is 1920 × 1080, and the field of view is 1.97 to 40 degrees. The resolution of the infrared sensor is 640 × 512, and the field of view is 2 and 10 degrees. The range of the laser rangefinder is 20 m to 10 km with the error of less than 5 m, the tracking accuracy of the photoelectric servo is

FIGURE 1. "Tianxing-1" unmanned surface vehicle.

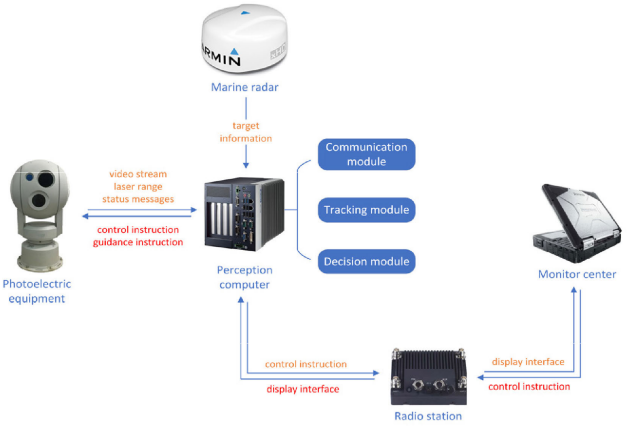

FIGURE 2. Perception system.

1 mrad, and it can generally navigate in the sea condition of grade 4. The azimuth accuracy of the navigation radar is 0.2 degrees, and the effective detection range is 40 m to 18 km. As shown in Fig. 2, first, marine radar detects the sea surface environment and transmits the detected object information (The object information includes the target's longitude and latitude, the direction angle of the target, the target's speed, the size of the target, and the course of the target) to the perception computer that transmits the processed object information to the photoelectric sensor. The photoelectric sensor points the USV toward the object according to the guidance of the perception computer. Then, the detection object enters the photoelectric field of view, and the photoelectric sensor transmits the optical data of the field of view to the perception computer. The perception computer recognizes and detects the object in the video and then transmits the first frame of the video data to the tracking algorithm to track the object in the video. Then, the tracking algorithm returns the pixel location deviation of the object and transmits it to the photoelectric servo motor. The servo tracking is controlled according to the pixel location deviation to realize real-time photoelectric object tracking.

B. MARINE TARGET DATASET

The targets tested in this paper mainly include buoy and ship. Prior to the algorithm test, we first set up the dataset, respectively, in different weather, different sea areas, different times, and different angles. We carried out multiple data collection and screened the data according to the specific data collection effect and the subsequent test environment analysis, to complete the establishment of the dataset.

Fig. 3 shows a partial display of the buoy dataset. These data samples include various characteristics such as small-scale target, large-scale target, target in fog, target in solid light, target in the fuzzy situation, target in complex background, target in occlusion.

Fig. 4 shows a partial display of the marine ship dataset. The data sample includes data with various characteristics such as small-scale target, large-scale target, target in fog, target in solid light, target in the fuzzy situation, target in

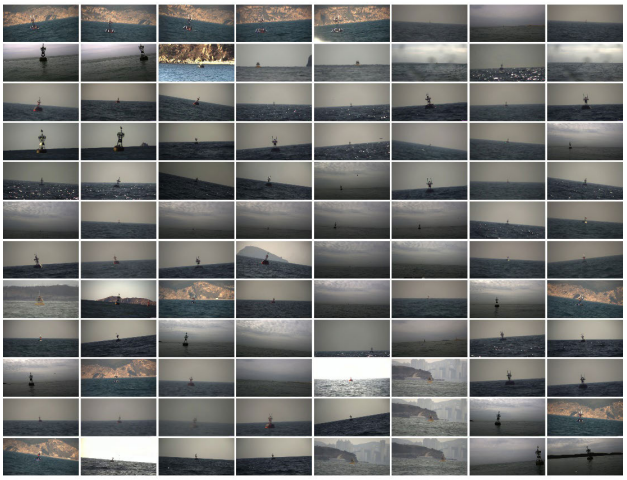

FIGURE 3. Partial display of buoy data.

complex background, target in static state, target in a dynamic state, target in high-speed navigation, target in bow direction and target in ship side direction, the target in the stern direction, the sample of the ship and buoy appearing at the same time.

C. TYPICAL OBJECT TRACKING ALGORITHM

Moving object tracking [5] is a challenging and popular topic research topic in USV research. With the rapid development of hardware equipment and artificial intelligence, object tracking performance has been greatly improved, but at the same time, the rapid development of USVs has led to higher requirements for object tracking algorithms. The complex environmental background, multi-object interference, and large changes in the object shape and object size still hinder the widespread application of object tracking algorithms.

The object tracking algorithms are classified according to the time sequence: the early classic tracking algorithm, the tracking algorithm based on the kernel correlation filter, and the tracking algorithm based on deep learning. Early object tracking algorithms have been mainly based on object

FIGURE 4. Partial display of marine ship data.

modeling or search. Object-based modeling involves modeling the appearance of an object and then finding that object in the following frame. The most commonly used method is feature matching. First, the object features are extracted, and then the most similar features are found in subsequent frames to determine the object location. The commonly used features include those derived from the scale-invariant feature transform (SIFT) feature [6], speeded up robust features (SURF) [7], and Harris corner detector. However, the real-time performance of this method is poor. With increased research, the prediction algorithm has been added to the object tracking field, and tracks an object near a predicted value to reduce the tracking range. The common prediction algorithms include the Kalman filter [8] and particle filter [9]. Another method to reduce the search range is the kernel method that uses the principle of the steepest descent method to iterate the object template in the direction of gradient descent until it reaches the optimal position; examples include the MeanShift [10] and continuously adaptive mean shift (CAMShift) algorithms [11]. However, the above algorithms contain fatal defects. First, the background information is not taken into account; second, the tracking algorithms are slow.

In recent years, object tracking has experienced much progress mainly due to the introduction of correlation filtering from the communication field into object tracking. Typical tracking algorithms based on correlation filtering include the minimum output sum of squared error (MOSSE) [12], circulant structure tracking with kernels (CSK) [13], kernel correlation filter (KCF) [14], background-aware correlation filter (BACF) [15] and scale-adaptive multiple-feature (SAMF) [16] approaches. Before the correlation filtering method is applied to object tracking, all tracking algorithms are processed in the time domain. The process of operation involves a complex matrix inverse operation that incurs a high computational cost and yields poor real-time performance. An object tracking method based on the correlation filter transfers the calculation to the frequency domain, and the cyclic matrix can be diagonalized in the frequency domain, substantially reducing the number of calculations and improving operation speed.

Recently, with the rapid development of deep learning [17], deep learning methods have been widely used in the field of object tracking. The hierarchical convolution feature (HCF) [18] approach integrates the depth feature domain correlation filtering algorithm because the features extracted from different layers of a convolution neural network [19] have different characteristics. The former image has high resolution, and the feature contains more location information; the latter contains more semantic information. The three-layer network is used to train the correlation filter, and the final response position of the object is obtained after weighting. In the continuous convolution operator tracker (C-COT) [20], a continuous convolution filter method is proposed. On the basis of the spatially regularized discriminative correlation filter (SRDCF), the SRDCFdecon approach [21] adopts a multilayer depth feature, and a continuous spatial domain difference conversion operation is used to determine that the input of the filter can have different resolution characteristics. The efficient convolution operator (ECO) [22] is an improved version of C-COT and considers the model size, sample size and update strategy. First, the filter with a small contribution is filtered out; then, the training samples are simplified to reduce the redundancy between adjacent samples; and finally, the model is updated every six frames.

1) MEANSHIFT

The MeanShift algorithm is a nonparametric estimation method. By calculating the similarity between the candidate template of the object position in the previous frame and the object position in the current frame, the template with a high similarity is selected to obtain the MeanShift vector of the object. The vector points to the current frame position of the object from the previous frame position. The mean shift vector is calculated by the iterative mean shift algorithm. The final algorithm locates the final position of the object in the current frame to achieve object tracking. Although the MeanShift algorithm achieves a good tracking effect, it cannot track the change in object scale adaptively and cannot track against a complex background.

2) KCF

Correlation filter theory is introduced into the KCF algorithm. An object tracking algorithm based on the correlation filter transfers the calculation to the frequency domain. The cyclic matrix can be diagonalized in the frequency domain, substantially reducing the computational cost and improves the operation speed. The cyclic matrix can increase the number of negative samples and improve the quality of classifier training. A Gaussian kernel is added to ridge regression to simplify the calculation. The introduction of the Gaussian kernel function can transform nonlinear problems into linear problems in high-dimensional space, making the algorithm more general.

3) ECO

The ECO algorithm is a filtering feature point tracking algorithm based on a convolution neural network spatial gradient calculation and histogram of oriented gradients (HOG) feature recognition technology that serves as an improvement of the C-COT. Based on the traditional discriminant correlation filter, Danelljan improved the time and space efficiency by improving the model size, training collective quantity and model updating method. Feature dimension reduction, sample space compression and a sparse update strategy are used to reduce the computational burden and improve the tracking efficiency of the algorithm. Although ECO greatly reduces the overfitting of the C-COT and reduces most of the unnecessary calculations, the ECO algorithm cannot benefit from the deeper CNN.

D. THE CONTENT OF THIS PAPER AND HIGHLIGHTS

To prevent the tracking loss of specific objects under multiple objects in USV object tracking based on marine radar guidance photoelectric mechanisms, and the poor tracking robustness and high object detection delay under complex backgrounds and large changes in the object morphology and object scale, an object detection-tracking algorithm based on a radar-photoelectric system is proposed. The algorithm is based on marine radar guiding photoelectric object pointing. After the photoelectric object is in place, the first-frame image is preprocessed, and then the object size, position and classification are obtained by the object detection algorithm. Then, the first-frame object extraction algorithm is used to filter the object in the first frame, the marine-radar-guided object is selected from the first-frame image, and then that object is fused with the object in the photoelectric image. The region of interest (ROI)-prediction algorithm is used to detect the object in each frame. According to the object position in the previous frames, the next pixel position of the object is predicted, the object is detected near the predicted position, and the local position of the object is returned to the global position. If the object is lost or the detection result does not match the radar guidance object, the predicted ROI interval is increased, and the object is detected and matched again. Thus, the object detection-tracking algorithm based on a radar-photoelectric system is realized. The algorithm makes full use of the depth information of the convolution neural network, improves the robustness of water object tracking, and substantially reduces the number of calculations compared with the traditional object detection algorithm. This algorithm can meet the accuracy requirement of object tracking and improve the computational efficiency. The main contributions of this paper are summarized as follows:

1) AN OBJECT DETECTION-TRACKING ALGORITHM BASED ON A RADAR-PHOTOELECTRIC SYSTEM IS PROPOSED

Compared with the KCF and ECO tracking algorithms, this algorithm can make full use of depth information. For a complex background environment, the robustness is greatly improved. Based on the results of object detection, the ROI-prediction algorithm can achieve stable tracking of an object. Compared with You Only Look Once, version 3 (YOLOv3) and other detection algorithms, it greatly improves the detection efficiency and reduces the time consumption by 25.8%.

2) A FIRST-FRAME OBJECT EXTRACTION ALGORITHM IS PROPOSED TO REALIZE THE FUSION OF MARINE RADAR GUIDANCE OBJECTS AND PHOTOELECTRIC TRACKING OBJECTS

Based on the marine-radar-guided photoelectric object tracking mechanism, the first-frame object extraction algorithm enables realization of the autonomous sensing of USVs. Compared with the tracking algorithm, the first-frame object extraction algorithm realizes the automatic extraction of tracking objects under multiple objects and fuses them with radar-guided objects. The object is matched by photoelectric laser positioning and marine radar positioning deviation, and photoelectric tracking direction angle deviation and marine radar guidance direction angle deviation.

3) AN ROI-PREDICTION ALGORITHM IS PROPOSED

According to the object location, a prediction algorithm is proposed to predict the position of the object in the vicinity of the object. The local position of the object is reset to the global position. If the object is lost or the detection result does not match the radar guidance object, the predicted ROI interval is amplified, detected and matched again. When it is enlarged to the maximum global area, if no object exists that meets the conditions, then the object is considered to be lost, and the marine radar objects undergo the photoelectric process again.

II. FIRST-FRAME OBJECT EXTRACTION ALGORITHM

A. OBJECT TRACKING MECHANISM OF UNMANNED SURFACE VEHICLE

The real-time tracking and positioning of an object by a sensing system is an important part of the sea object tracking task of USVs. This real-time object tracking and positioning process is mainly divided into the following three parts: first, the marine radar detects the surface object and guides the photoelectric equipment toward the object; second, the perception computer identifies and detects the object sequence images returned by the photoelectric system to return the object pixel location deviation; and third, according to the pixel location deviation, the photoelectric servo realizes the tracking and laser positioning of the object. The USV object tracking mechanism is shown in Fig. 5.

The return accuracy of the direction angle of marine radar is high, and in this test, it is 0.2 degrees. Even if the photoelectric system is adjusted to the minimum field of view (the minimum field of view of the photoelectric device used in this test is 1.97 degrees), the radar can stably guide the photoelectric object pointing. After photoelectric pointing toward the object is accomplished, the object guided by radar appears in the photoelectric field of view. If only one object

FIGURE 5. USV tracking mechanism.

of this classification is in the field of view, the pixel location deviation of the object can be obtained by the object recognition and detection algorithm in the perception computer to realize stable tracking of the object by the photoelectric servo. However, in actual task execution, multiple objects of the same classification often appear in the same field of view after the object is identified. In this case, the object detection algorithm cannot accurately return the pixel location deviation of the radar-guided object, causing problems such as incorrect object tracking and servo oscillation.

To solve this problem, this paper proposes a first-frame object extraction algorithm, which can realize the filtering of nonguided objects in the same field of view under the guidance of marine radar and realize the fusion of radar-guided objects and photoelectric visible light objects. The proposed algorithm improves the robustness of object tracking and enables the sensing system to realize the stable return of object pixel location deviation in the first photoelectric image in a complex environment.

B. PRINCIPLE OF THE FIRST-FRAME OBJECT EXTRACTION ALGORITHM

The implementation of the first-frame extraction algorithm mainly includes the following four parts: first, elimination of the outliers in the first frame of the photoelectric image; second, filtering the object according to the horizontal pixel location deviation; third, eliminating the left-right oscillation when the pixel location deviation is close; and fourth, applying the pixel area filtering method.

First, according to the characteristics of radar guidance, when the USV and object are relatively stationary, the maximum deviation of the direction angle is α degrees. Then, according to the relative position and relative speed of the USV and the object, we can calculate the object pixel position with an optoelectronic field of view of β and a video resolution of $P_x \times P_y$ after radar guidance. Based on this, the fault tolerance value ε is added to ensure the stable filtering of outliers.

$$A_0 = \frac{\alpha}{\beta} \cdot P_x \tag{1}$$

$$A_1 = \frac{P_x}{2} - (A_0 + \varepsilon) \tag{2}$$

$$A_2 = \frac{P_x}{2} + (A_0 + \varepsilon) \tag{3}$$

where A_0 is the maximum deviation of the actual pixel position of the object in the horizontal direction after radar-guided photoelectric object pointing. Filtering removes objects less than A_1 or greater than A_2 in the optoelectronic field of view. When the pixel position of the object behind the radar object is calculated, each object is weighted according to the guidance characteristics of the radar and the pixel location deviation in the horizontal direction of each object.

$$d_i = \left| x_i - \frac{P_x}{2} \right| \tag{4}$$

$$p_i = \sum_{j=1}^{n} sign(d_i - d_j) \tag{5}$$

$$q_i = \begin{cases} \dfrac{n+1}{2} + sign(p_i) \cdot \left| p_i - p_{\frac{n+1}{2}} \right|, & n \text{ is odd} \\ \dfrac{n}{2} + sign(p_i) \cdot \dfrac{\left| p_i - p_{\frac{n}{2}} \right|}{2}, & n \text{ is even} \end{cases} \tag{6}$$

where i is the serial number of each object in the original data, x_i is the pixel coordinates of the object in the horizontal direction, and d_i is the distance between the object and the image center in the horizontal direction. q_i is the new serial number of each object.

When two objects with the same classification appear in the same field of view and the absolute pixel location deviation values of the two objects are close to each other, optoelectronic servo tracking oscillation occurs during the actual task execution. When the difference between the two frames is less than x, the pixel location deviation of this frame is calculated with the pixel location deviation of the previous frame to ensure that the calculated results of the two frames are the same in order to solve the problem of photoelectric servo oscillation.

$$\begin{cases} X_k = x_k - \dfrac{P_x}{2} \\ Y_k = y_k - \dfrac{P_y}{2} \end{cases} \tag{7}$$

$$S_k = sign(X_k \cdot X_{k-1}) \tag{8}$$

where X_k is the object pixel location deviation in the horizontal direction in the current frame and Y_k is the pixel location deviation in the vertical direction of the object in the current frame. S_k is the oscillation flag of two adjacent frames. If S_k is 0 or 1, then it is in the non-oscillatory state; if S_k is -1, then it is in the oscillation state.

We can obtain the approximate distance from the object via detection of the surrounding environment by radar. At the same time, we can classify the object according to the anisotropy of the object echo, the echo size and the object behavior characteristics. The pixel size of the object can be estimated according to the classification of its distance. Each

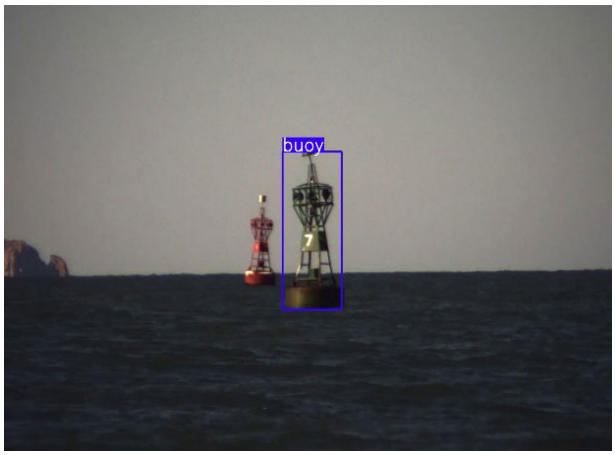

FIGURE 6. Effect of the first-frame object extraction algorithm In the case of multiple buoys (simple background).

pixel is used as the object weight to estimate the object size.

$$s_i = w_i \cdot h_i \tag{9}$$

$$\|\mathbf{s}\|_\infty = \max_{1 \le i \le m} |s_i| \tag{10}$$

where w_i is the object width, h_i is the object height, and s_i is the object pixel area. m is the number of objects retained after filtering by previous filtering algorithms. In the reserved objects, the object with the largest pixel area is selected as the final guidance object input to the photoelectric servo.

C. TEST RESULTS OF THE FIRST-FRAME OBJECT EXTRACTION ALGORITHM

First, the first-frame object extraction algorithm is tested against a simple background. As shown in Fig. 6, two objects with the same classification exist in the same field of view. After marine radar detection, the perception decision-making program automatically selects the nearest object to the radar object to guide the object. In this field of view, the green buoy is the object guided by the radar, and the red buoy is successfully filtered out by the first-frame object extraction algorithm.

Against a complex background, the first-frame object extraction algorithm is tested in the case of multiple buoys. As shown in Fig. 7, two objects of the same classification exist in the same field of view. In addition, the background behind the buoy is more complex. The red buoy is again filtered out by the first-frame object extraction algorithm.

Against a simple background, the first-frame object extraction algorithm is tested in the case of multiple ships. As shown in Fig. 8, two objects with the same classification exist in the same field of view. After marine radar detection, the perception decision-making program automatically selects the nearest object to the radar object to guide the object. In this field of view, the ship in the middle is the object guided by the radar, and the other ship is successfully filtered out by the first-frame object extraction algorithm.

Against a complex background, the first-frame object extraction algorithm is tested in the case of multiple ships.

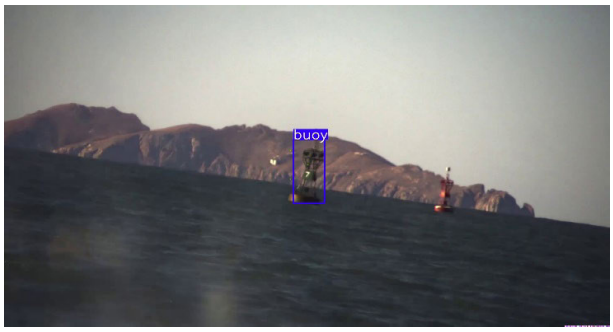

FIGURE 7. Effect of the first-frame object extraction algorithm In the case of multiple buoys (complex background).

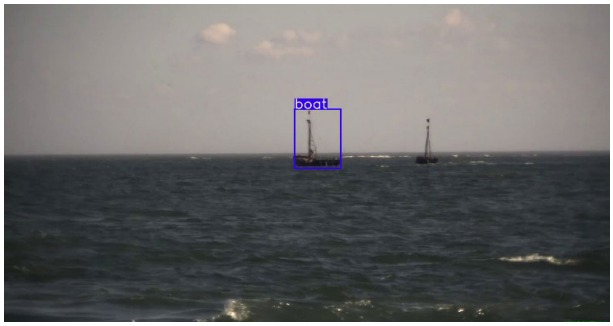

FIGURE 8. Effect of the first-frame object extraction algorithm In the case of multiple ships (complex background).

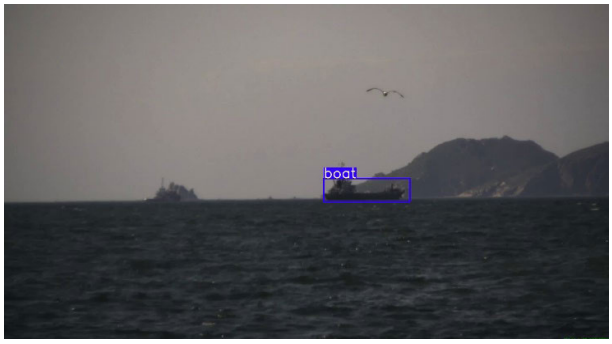

FIGURE 9. Effect of the first-frame object extraction algorithm In the case of multiple ships (complex background).

As shown in Fig. 9, two objects with the same classification exist in the same field of view. In addition, the background behind the ship is more complex. Using the first-frame object extraction algorithm, the other ships are again filtered out successfully.

In the test, ten different sea test environments are selected to test the first-frame target extraction algorithm in the case of multiple buoys and multiple ships. The results are shown in Fig. 10.

III. ROI-PREDICTION ALGORITHM

A. VALUE OF THE ROI-PREDICTION ALGORITHM

For the USV object tracking task, the time consumed during object detection and recognition is a very important index.

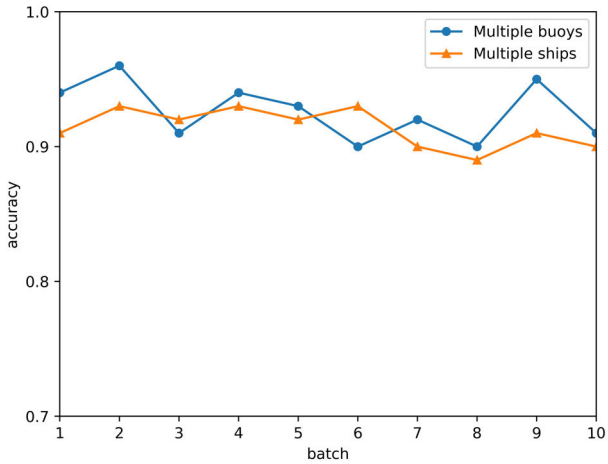

FIGURE 10. Accuracy of the first-frame object extraction algorithm in different environments.

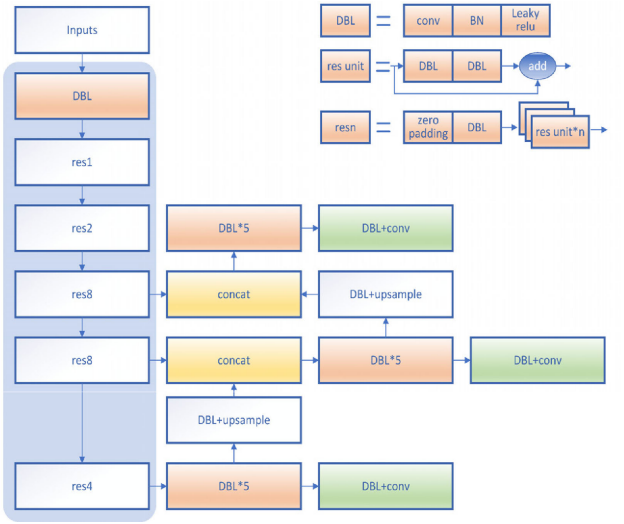

FIGURE 11. Network structure of object detection algorithm.

The traditional object detection and recognition algorithm detects and recognizes all of the images in each frame. Processing all of the images in each frame is a highly time-consuming task. If the image enhancement algorithm is added to the front end or other image processing algorithms are added to the back end, the time required increases. Therefore, the delay may not meet the requirements of the photoelectric servo tracking delay. In fact, little change occurs in the position of the object pixel between two adjacent frames, and the approximate position of the object in the next frame can be predicted by the Kalman filter algorithm. If the calculation time of the predicted object area image processing and object detection and recognition is reduced, a better object tracking effect can be achieved. At the same time, occurrences of object false detections can be greatly minimized.

Compared with the traditional depth difference algorithm, the prediction algorithm can solve the problem of object tracking with large differences in depth and shape. In addition, when the object is against a complex background, the ROI-prediction algorithm can provide a better tracking effect.

B. THE PRINCIPLE OF THE ROI-PREDICTION ALGORITHM

Realization of the ROI-prediction algorithm mainly involves the following three components: first, the object area prediction algorithm; second, the local object detection algorithm; and third, object loss detection.

When the target motion model and sensor observation model are linear functions and the noise model is Gaussian white noise, the Kalman filter is the optimal solution for the tracking problem. First, according to the position of the object pixel in the previous frame or several frames, the Kalman filter algorithm is used to predict the position of the object in the next frame. The object global position and size in the first frame are extracted, and the global coordinates of the object are used as input. The new object position is predicted by the following state prediction equation and state update equation.

State prediction equation:

$$\begin{cases} \hat{x}_t^- = F\hat{x}_{t-1}^- + Bu_{t-1} \\ P_t^- = FP_{t-1}^- F^T + Q \end{cases} \tag{11}$$

State update equation:

$$\begin{cases} K_t = P_t^- H^T (HP_t^- H^T + R)^{-1} \\ \hat{x}_t = \hat{x}_t^- + K(z_t - H\hat{x}_t^-) \\ P_t = (I - K_t H)P_t^- \end{cases} \tag{12}$$

Then, according to the size of the object in the last frame, the prediction area is appropriately enlarged to improve the probability of the existence of a real object at the predicted object position. At the same time, properly enlarging the detection area does not cause a large increase in the time consumed. The depth features of the enlarged object prediction area are extracted, and object detection and recognition are carried out. The network structure is shown in Fig. 11. After the object is successfully detected, the new local coordinates and new size of the object are extracted. Then, the local coordinates of the object are converted to global coordinates, and the global coordinates and size of the object are taken as new inputs for the iterative predicted the object position in the next frame.

During reoperation, the angle between the object and the bow direction in the geodetic coordinate system is calculated according to the center point of the object. If the angle between the object and the radar exceeds the maximum error angle or the object is not detected in the detection process, the object is regarded as lost.

C. TEST RESULTS OF THE ROI-PREDICTION ALGORITHM

As shown in Figs. 12 and 13, the ROI-prediction algorithm is used to track a buoy. As shown in the figure, the initial

FIGURE 12. First-frame object extraction algorithm for buoy extraction.

FIGURE 13. Test results for a buoy based on ROI-prediction algorithm.

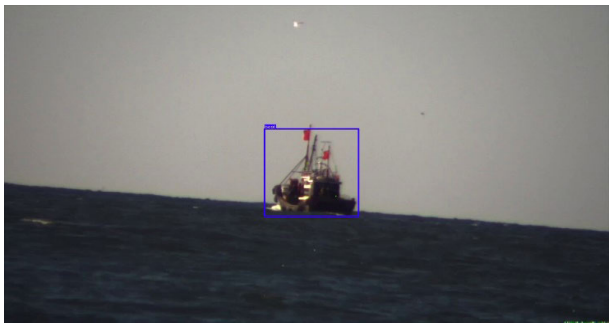

FIGURE 14. First-frame object extraction algorithm for ship extraction.

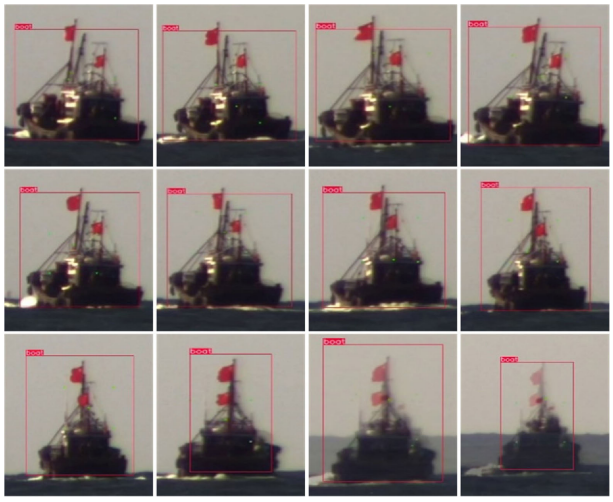

FIGURE 15. Test results for a ship based on the ROI-prediction algorithm.

position of the object is extracted through the first-frame object extraction algorithm, and then the position of the buoy at the next moment is predicted by the ROI-prediction algorithm. The object is detected near the prediction area, the new local position of the object is successfully extracted, and then the local position is converted to the global position so that the next frame position of the buoy can be predicted again until the object is lost.

Even when the buoy is interleaved with several ships, the ROI-prediction algorithm can still realize stable tracking of the buoy.

As shown in Figs. 14 and 15, the ROI-prediction algorithm is used to test the object tracking of ships at sea. In the figures, the initial object position is extracted through the first-frame object extraction algorithm, and then the ship position at the next moment is predicted by the ROI-prediction algorithm. The object is detected near the prediction area, and the new local position of the object is successfully extracted. Then, the local position is converted to the global position so that the next frame position of the buoy is predicted again until the object label is missing.

In the actual test, the ROI-prediction algorithm greatly reduces the time consumption of object detection and improves the robustness of object tracking.

IV. OBJECT DETECTION-TRACKING ALGORITHM FOR UNMANNED SURFACE VEHICLES BASED ON RADAR-PHOTOELECTRIC SYSTEMS

A. THE VALUE OF THE USV OBJECT DETECTION-TRACKING ALGORITHM BASED ON A RADAR-PHOTOELECTRIC SYSTEM

Compared with the KCF and ECO tracking algorithms, our algorithm can fully utilize the depth information. For a complex background environment, the robustness is substantially improved. Based on the object detection results, the ROI-prediction algorithm can achieve stable object tracking. Compared with YOLOv3 [23], [24] and other detection algorithms [25], [26], it greatly improves the detection efficiency and reduces the time consumption by 25.8%.

The implementation of the object tracking algorithm based on object detection mainly involves two algorithms: the first-frame extraction algorithm and the ROI-prediction algorithm. The specific process is shown in Fig. 16.

First, marine radar scans the area and then guides the photoelectric system to point toward the object. Next, the first-frame object extraction algorithm is executed to extract the pixel coordinates and size of the radar-guided object. Then, the pixel coordinates of the object are taken as the input, and the pixel position of the object in the next frame is

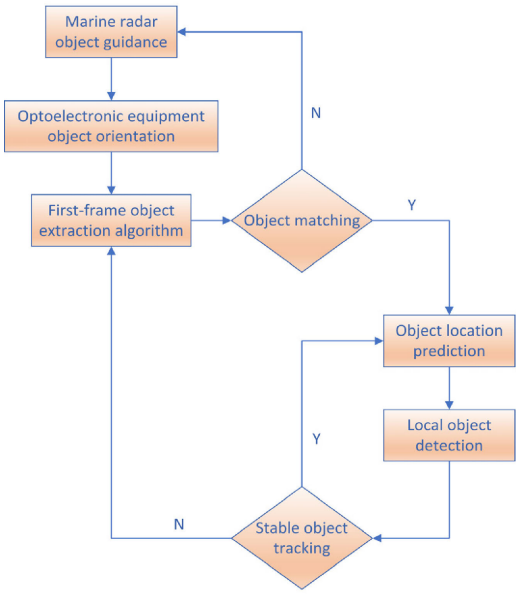

FIGURE 16. Flow chart of the object detection-tracking algorithm.

predicted by the ROI-prediction algorithm. Then, the local pixel coordinates of the object are converted into global pixel coordinates that are used as the new input to predict the next frame region. When object loss occurs after the object detection algorithm is executed, the first-frame object extraction algorithm is executed again to carry out new global object detection. If no matching object is found after the first-frame object extraction algorithm is executed, the marine radar carries out a photoelectric object search again.

B. ADVANTAGES OF THE USV OBJECT DETECTION-TRACKING ALGORITHM BASED ON A RADAR-PHOTOELECTRIC SYSTEM AGAINST A COMPLEX BACKGROUND

Compared with KCF and other object tracking algorithms, the object detection-tracking algorithm in this paper has higher stability. As shown in Fig. 17, it has better performance in the actual test process when the background is complex and the object is obscured by fog.

When the environmental background is complex and the color of the background is very similar to that of the buoy, the algorithm in this paper can still achieve stable object tracking.

Moreover, this paper tests the algorithm in cloudy and dark weather. As shown in Fig. 18, the unmanned surface ship approaches the buoy from 2 km away. Despite the cloudy weather, it can still achieve stable tracking of the target during the navigation. As shown in Fig. 19, when the unmanned surface ship is in dark weather, the perception computer automatically switches to the infrared sensor, and the target tracking can still be realized in dark weather by detecting the infrared image.

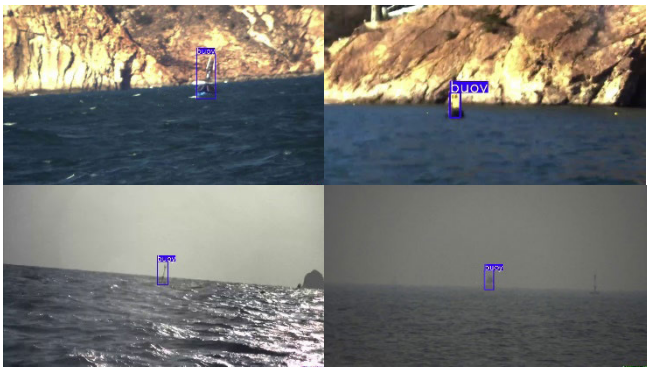

FIGURE 17. Object tracking effect against a complex background and under fog.

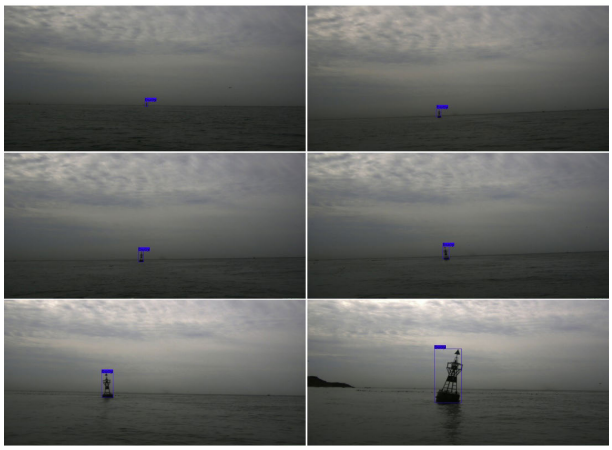

FIGURE 18. Object tracking effect in cloudy weather.

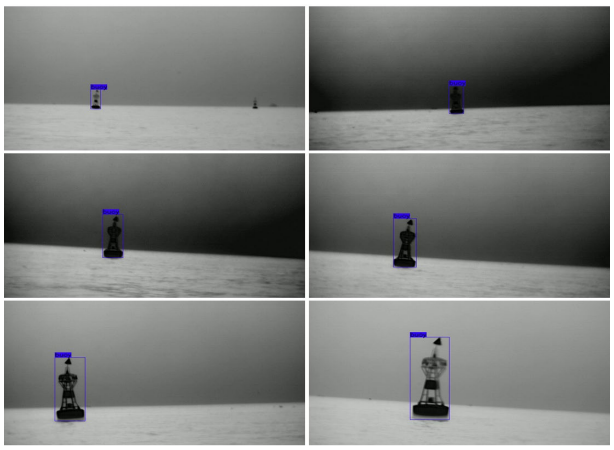

FIGURE 19. Object tracking effect in dark weather.

During the navigation of unmanned surface vehicles, the mast will block the lens and the wave on the deck, leading to a significant time interruption between the image sequences. In view of these situations, this paper has carried out the relevant tests. When the image sequence is interrupted, the photoelectric system will not detect the target.

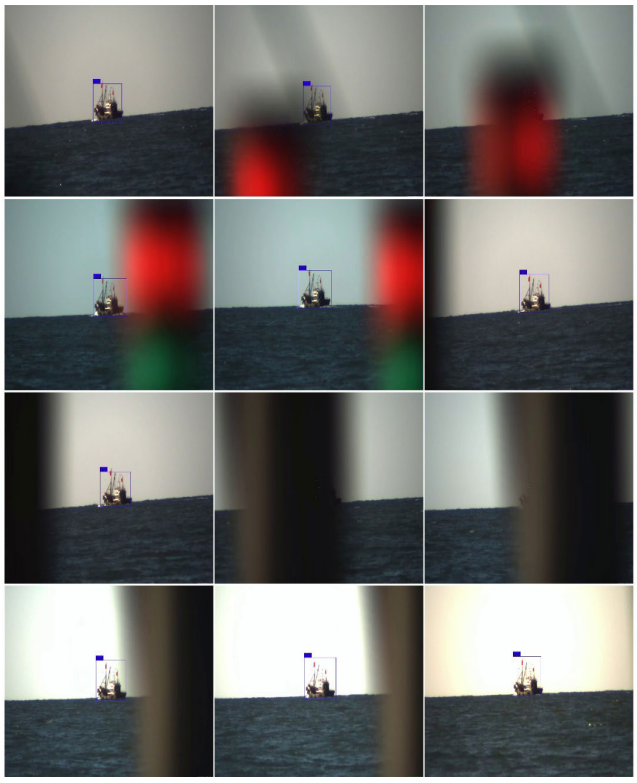

FIGURE 20. Object tracking effect in the case of mast blocking the lens.

FIGURE 21. Object tracking effect in the case of seawater blocking the lens.

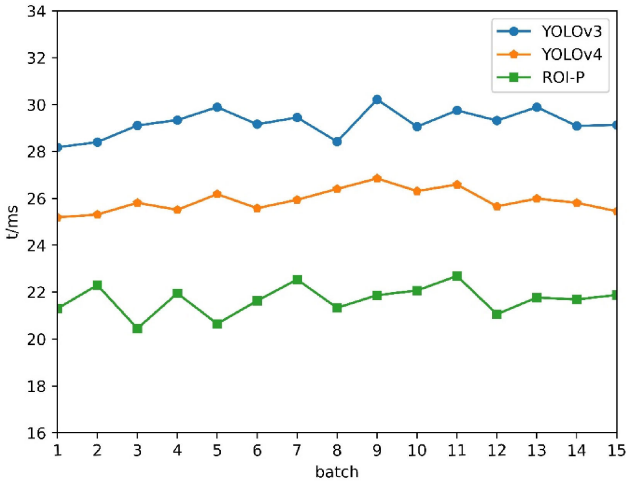

FIGURE 22. Comparison of the time consumption between the ROI-prediction algorithm and YOLOv3 and YOLOv4 in the buoy tracking test.

At this time, the navigation radar target information will be called again, and the angle tracking of the target will be carried out. When the target appears again in the image, the tracking mechanism will immediately change to the tracking according to the pixel location deviation.

As shown in Fig. 20, when the mast blocks the lens, the target ship disappears in the image for a while. Simultaneously, the angle tracking of the target through the navigation radar information can still achieve the target ship tracking. When the mast no longer blocks the target ship, the tracking mechanism will immediately change to tracking according to the pixel location deviation to achieve higher precision target tracking.

As shown in Fig. 21, when there is a wave on the deck, after the splashing water blocks the lens, the target ship will disappear in the image for a while. At this time, angle tracking can still be achieved by using navigation radar information. When the navigation is stable, and there is no wave on the deck, the tracking mechanism will immediately switch to tracking according to the pixel location deviation to achieve higher accuracy dynamic target tracking.

The USV object detection-tracking algorithm based on a radar-photoelectric system predicts the pixel position of the object in the next frame and only detects the object in the prediction area. Compared with the traditional object detection algorithm, by predicting the object area and reducing the detection area, the detection efficiency is largely improved. Under the same experimental environment, the algorithm's time consumption is reduced by 25.8% compared with that of the YOLOv3 algorithm. Compared with the yolov4 algorithm, the time consumption of the algorithm is reduced by 16.4%. Some of the data are compared in Figs. 22 and 23.

V. SEA TRIALS

A. THE PRINCIPLE OF THE ROI-PREDICTION ALGORITHM

A sea test is conducted to examine the dynamic object tracking effect of the USV object detection-tracking algorithm based on a radar-photoelectric system in an actual marine environment.

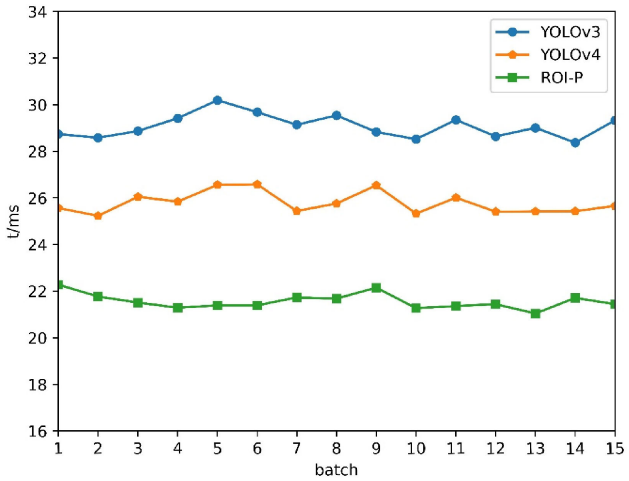

FIGURE 23. Comparison of the time consumption between the ROI-prediction algorithm and YOLOv3 and YOLOv4 in the ship tracking test.

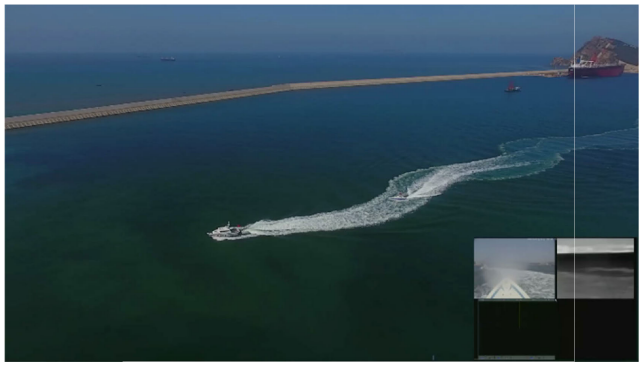

FIGURE 24. Dynamic object tracking of "Tianxing-1".

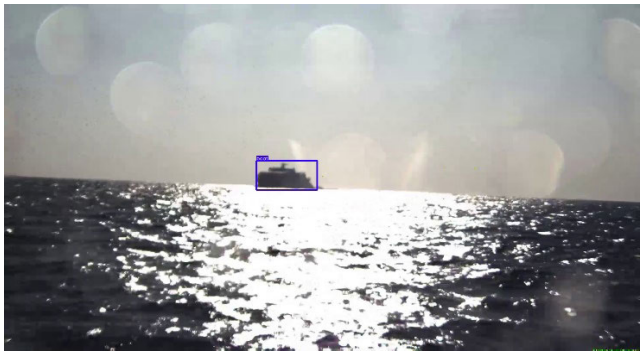

FIGURE 25. First-frame object extraction algorithm for ship extraction.

Fig. 24 shows a surface view of Tianxing-1 as it tracks a dynamic object.

The object ship sails at a speed of 25 knots. Tianxing-1 starts to sail when the object ship is at a distance of 8 km away. When it is 500 m away from the object ship, it starts evasion maneuvers. After sailing to the rear of the object ship, Tianxing-1 achieves stable tracking of the object ship at a distance of 400-500 m from the object ship.

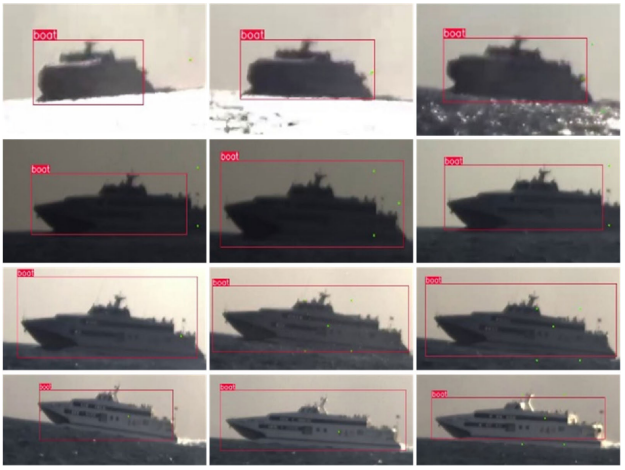

FIGURE 26. Test results for a ship based on the ROI-prediction algorithm.

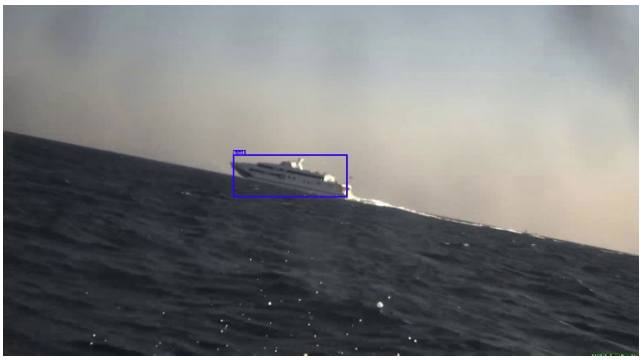

FIGURE 27. First-frame object extraction algorithm for ship extraction.

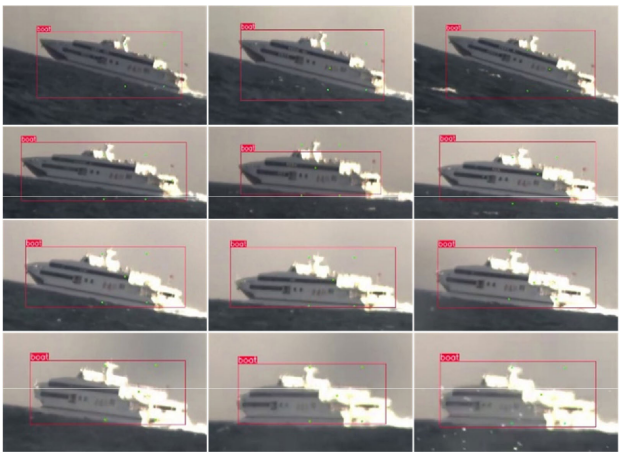

FIGURE 28. Test results for a ship based on the ROI-prediction algorithm.

B. PROCESS OF THE MARINE TEST

After the marine radar detects the object ship, it guides the photoelectric device to point toward the object. As shown in Fig. 25, the object identified by the radar in the field of view is extracted through the first-frame object extraction algorithm. Then, the pixel coordinates of the object are input into the ROI-prediction algorithm to predict the next frame position

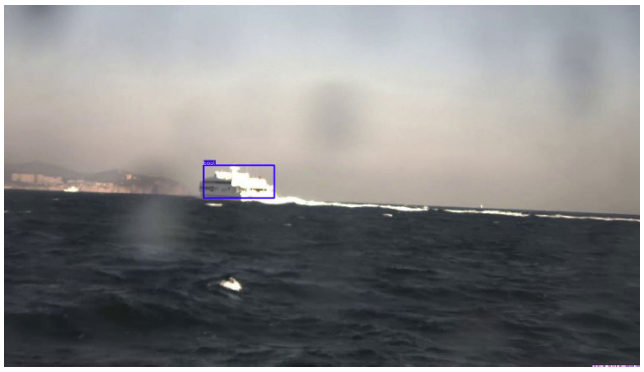

FIGURE 29. First-frame object extraction algorithm for ship extraction.

FIGURE 30. Test results for a ship based on the ROI-prediction algorithm.

FIGURE 31. First-frame object extraction algorithm for ship extraction.

of the object, and then the local object detection algorithm is implemented near the prediction area. Then, the local coordinates of the object are transformed into global coordinates, and the object position is predicted again. As shown in Fig. 26, the ROI-prediction algorithm achieves the object tracking effect after extracting the position information of the object ship with its size in the first frame.

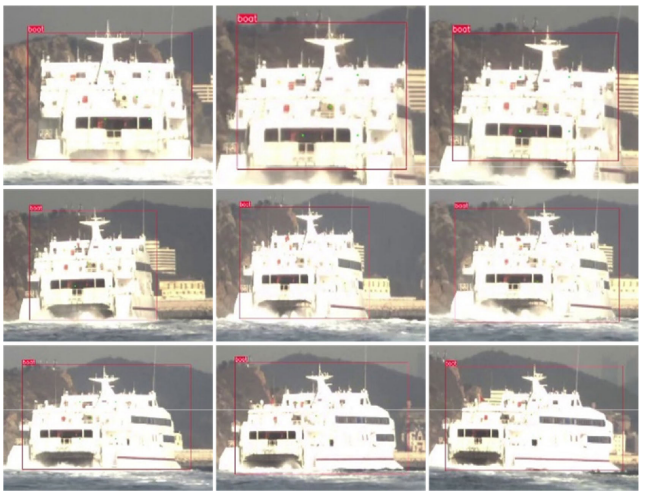

FIGURE 32. Test results for a ship based on the ROI-prediction algorithm.

Due to a wave on the deck, the object is lost in the field of view. Then, the marine radar reattempts the photoelectric object pointing and re-executes the first-frame object extraction algorithm to extract the new pixel position of the object ship and track it, as shown in Figs. 27 and 28.

At this point, the evasion action of the object ship is nearly complete, and then the tracking ship circles to the rear of the object ship for constant speed tracking; see Figs. 29 and 30.

At this time, ''Tianxing-1'' circles to the rear of the object ship and keeps a distance of 400-500 m from the object ship for stable tracking; see Figs. 31 and 32.

VI. CONCLUSION

An object detection-tracking algorithm based on a radar-photoelectric system is proposed to avoid the tracking loss of specific objects among multiple objects in the object tracking of USVs that leads to poor tracking robustness and high object detection delay when the object shape and scale substantially change against a complex background. The first-frame object extraction algorithm fuses marine radar guidance objects and photoelectric tracking objects. When multiple objects of the same classification appear in the same field of view, an object guided by radar can be extracted stably. When the object is against a complex background, the algorithm can also achieve stable object tracking, and the robustness is greatly improved. Due to the proposed ROI-prediction algorithm, the detection area of each frame is reduced, the stability of object tracking is improved, and the detection efficiency is substantially increased. Compared with the YOLOv3 detection algorithm, the time consumption is reduced by 25.8%. Compared with the yolov4 detection algorithm, the time consumption is reduced by 16.4%. In a sea test of dynamic objects, the stability of object tracking is notably improved.

REFERENCES

[1] Z. Liu, Y. Zhang, X. Yu, and C. Yuan, ''Unmanned surface vehicles: An overview of developments and challenges,'' *Annu. Rev. Control*, vol. 41, pp. 71–93, Jan. 2016.

[2] M. Breivik, V. E. Hovstein, and T. I. Fossen, "Straight-line target tracking for unmanned surface vehicles," *Model., Identificat. Control A, Norwegian Res. Bull.*, vol. 29, no. 4, pp. 131–149, 2008.
[3] G. Bruzzone, G. Bruzzone, M. Bibuli, and M. Caccia, "Autonomous mine hunting mission for the Charlie USV," in *Proc. OCEANS IEEE Spain*, Jun. 2011, pp. 1–6.
[4] Y. Ohta, H. Yoshida, S. Ishibashi, M. Sugesawa, F. H. Fan, and K. Tanaka, "Seabed resource exploration performed by AUV 'Yumeiruka,'" in *Proc. Oceans MTS/IEEE Monterey*, Sep. 2016, pp. 1–4.
[5] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surveys, Rev.*, vol. 38, no. 4, 2006, Art. no. 13.
[6] J.-Y. Choi, K.-S. Sung, and Y.-K. Yang, "Multiple vehicles detection and tracking based on scale-invariant feature transform," in *Proc. IEEE Intell. Transp. Syst. Conf.*, vols. 1–2, Sep. 2007, pp. 349–354.
[7] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Computer Vision—ECCV 2006*, vol. 3951. 2006, pp. 404–417.
[8] S. Y. Chen, "Kalman filter for robot vision: A survey," *IEEE Trans. Ind. Electron.*, vol. 59, no. 11, pp. 4409–4420, Nov. 2012.
[9] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "An adaptive color-based particle filter," *Image Vis. Comput.*, vol. 21, no. 1, pp. 99–110, Jan. 2003.
[10] G. Tian, R.-M. Hu, Z.-Y. Wang, and L. Zhu, "Object tracking algorithm based on meanshift algorithm combining with motion vector analysis," in *Proc. 1st Int. Workshop Educ. Technol. Comput. Sci.*, vol. 1, 2009, pp. 987–990.
[11] H. Chu, S. Ye, Q. Guo, X. Liu, and Ieee, "Object tracking algorithm based on Camshift algorithm combinating with difference in frame," in *Proc. IEEE Int. Conf. Automat. Logistics*, vols. 1–6, Aug. 2007, pp. 51–55.
[12] D. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2544–2550.
[13] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Computer Vision—ECCV*, vol. 7575. 2012, pp. 702–715.
[14] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.
[15] H. K. Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1144–1152.
[16] Z. Liu, Z. Lian, and Y. Li, "A novel adaptive kernel correlation filter tracker with multiple feature integration," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 8926, Sep. 2017, pp. 254–265.
[17] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
[18] A. Bibi, M. Mueller, and B. Ghanem, "Target response adaptation for correlation filter tracking," in *Computer Vision—ECCV*, vol. 9910. 2016, pp. 419–433.
[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
[20] M. Danelljan, A. Robinson, F. S. Khan, and M. Felsberg, "Beyond correlation filters: Learning continuous convolution operators for visual tracking," in *Proc. Eur. Conf. Comput. Vis.*, vol. 9909, 2016, pp. 472–488.
[21] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1430–1438.
[22] S. Yun, J. Choi, Y. Yoo, K. Yun, and J. Y. Choi, "Action-decision networks for visual tracking with deep reinforcement learning," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1349–1358.
[23] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. New York, NY, USA: IEEE, Jun. 2016, pp. 779–788.
[24] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. New York, NY, USA: IEEE, Jul. 2017, pp. 6517–6525.
[25] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*. New York, NY, USA: IEEE, Dec. 2015, pp. 1440–1448.
[26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

QINGZE YU received the B.Eng. degree in naval architecture and ocean engineering from Dalian Maritime University, Dalian, China, in 2015. He is currently pursuing the Ph.D. degree with Harbin Engineering University, Harbin, China. His research interests include intelligent marine vehicles, intelligent computing in marine environments, and marine environment comprehending.

BO WANG received the Ph.D. degree in design and manufacture of marine structures and vessels from Harbin Engineering University, Harbin, China, in 2017. He is currently an Associate Professor with the College of Shipbuilding Engineering, Harbin Engineering University. His research interests include autonomous marine vehicles, marine environment comprehending, and intelligent computing in marine environments. He serves as a Vice Chairman for the *Journal of Shipbuilding Engineering*.

YUMIN SU received the B.Eng. degree in naval architecture and ocean engineering and the M.S. degree in design and manufacture of marine structures from Harbin Engineering University, Harbin, China, in 1982 and 1984, respectively, and the Ph.D. degree in civil engineering from Yokohama National University, Yokohama, Japan, in 1999. His research interests include intelligent marine vehicles and bionics.