The Journal of Navigation (2021), **74**:6 1336–1352
doi:10.1017/S0373463321000503

RESEARCH ARTICLE

Identity recognition on waterways: a novel ship information tracking method based on multimodal data

Zishuo Huang,[1] Qinyou Hu,[1]* Qiang Mei,[1,2] Chun Yang,[1] and Zheng Wu[3]

[1] Merchant Marine College, Shanghai Maritime University, Shanghai, China.
[2] Navigation College, Jimei University, Xiamen, China.
[3] Department of Mathematics and Computer Science, Information Engineering University, Zhengzhou, China.
*Corresponding author. E-mail: qyhu@shmtu.edu.cn

Received: 1 April 2021; **Accepted:** 23 May 2021; **First published online:** 25 June 2021

Keywords: ship information tracking, multimodal data, object detection, object tracking

Abstract

Video monitoring is an important means of ship traffic supervision. In practice, regulators often need to use an electronic chart platform to determine basic information concerning ships passing through a video feed. To enrich the information in the surveillance video and to effectively use multimodal maritime data, this paper proposes a novel ship multi-object tracking technology based on improved single shot multibox detector (SSD) and DeepSORT algorithms. In addition, a night contrast enhancement algorithm is used to enhance the ship identification performance in night scenes and a multimodal data fusion algorithm is used to incorporate the ship automatic identification system (AIS) information into the video display. The experimental results indicate that the ship information tracking accuracies in the day and night scenes are 78·2% and 70·4%, respectively. Our method can effectively help regulators to quickly obtain ship information from a video feed and improve the supervision of a waterway.

1. Introduction

Inland waterways are important components of waterway transportation. A vessel traffic transit service (VTS) can effectively supervise the fixed waters of an inland waterway in real time, playing an important role in improving navigation efficiency and ensuring the navigation safety of vessels (Can, 2017); as such, it is a key research field in the shipping industry. As one important means of VTS supervision, maritime video surveillance primarily relies on humans to analyse and determine ship movement information in a video feed; however ship information obtained from such videos is extremely limited. A ship's automatic identification system (AIS) can collect a large number of spatial position data generated during ship navigation to provide intelligent ship management. The integration of AIS technology and ship VTS will provide a safe navigation system for maritime transportation (Kartika et al., 2018).

Augmented reality (AR) technology has been developing rapidly in recent years. This technology can be used to register virtual information (e.g., text, symbols and pictures) with objects in a real scene and, at the same time, achieve enhanced information tracking of real objects. As more and more intelligent algorithms are applied in maritime affairs, marine ship navigation and waterway supervision are gradually moving toward the field of AR. Researchers are committed to adding more navigation information to ship navigation videos to aid pilots. Frydenberg et al. (2018) developed a shipborne AR deployment scheme from the ship driving environment, human movements and illumination factors, and provided additional environment and navigation information for ship pilots by superimposing graphics and audio. Hugues et al. (2014) proposed an image analysis algorithm capable of detecting the horizon in

maritime scenes, creating an indirect visual function in an AR system for ships. Oh et al. (2016) analysed current shipboard equipment, such as radar and electronic charts, to design an AR user interface that integrates the visual view in front of a ship with virtual images and navigation information and presents it to the pilot via an independent computer screen.

With the continuous development of communication technology, the utilisation rate of AIS will be continuously improved in the future (Liu et al. 2020). Based on the massive parallel computing power of graphics processing units (GPUs), Huang et al. (2020) compacted the massive AIS tracks to realise visualisation. To obtain additional ship information, some researchers have tried to incorporate AIS information to enable visual support for AIS information. Lee et al. (2016) calculated the ship attitude by detecting an image ahead of the ship and then used image analyses and AIS data to detect the ship position and generate an enhanced display of that information. Based on AIS data, An et al. (2019) designed a visual analysis platform for global shipping routes that can display multiple aspects of the AIS data and generate a display of the entire sea route. Lukas et al. (2014) developed a multi-signal fusion monitoring method that combines electronic charts, radar and AIS information with the video feed, so that users can touch a screen to obtain all the ship information from the video.

Although researchers are increasingly incorporating AR technology in maritime applications. However, there are limited AR studies concerning inland waterway monitoring, which is often of higher application value in some important sections. This paper focuses on the special scenes of inland rivers and proposes innovations in improving the accuracy of ship detection, night image enhancement and maritime information fusion. In this paper, based on AIS data and combined with computer vision analysis methods, virtual information is registered with real images and an identification and tracking system for ship information suitable for inland waterway supervision is constructed, providing a convenient method for waterway supervisors to obtain ship information.

2. Related computer vision research

2.1. Object detection

Target detection is an important research direction of computer vision, which is used to lock the target position in a single-frame image. Research on ship detection began in 2002, code-named 'Spartan scout,' in the form of a project to self-control a pilotless ship. The ship was equipped with a ship detection system that could detect and track ship objects at sea (Rao et al., 2014). Later, Guang et al. (2011) proposed a method of region segmentation to extract the features of different segmentation regions; then, a machine-learning algorithm was used to detect the ships in each region of the image. Kim et al. (2010) used background estimation to enable ship detection and merged with AIS to implement relevant ship information matching. Liu (2010) detected ships in an inland waterway by separating the sky–water line and incorporating the average ship features in the region of interest.

In recent years, deep learning has been developing rapidly and is widely used in the field of object detection. Object detection algorithms are divided into two categories: object detection models based on classification, including region-based convolutional neural network (R-CNN) (Girshick et al., 2014), Fast R-CNN (Girshick, 2015) and Faster R-CNN (Ren et al., 2017), and detection models based on regression, including the YOLO (Redmon et al., 2016) and single shot multibox detector (SSD) (Liu et al., 2016) series.

Object detection methods based on deep learning have been widely used in various fields, including ship detection. Betti et al. (2020) constructed a YOLOv3 detector based on the Keras application programming interface to detect four types of ships: cargo ships, warships, oil tankers and tugboats. He et al. (2019) designed a method based on the combination of a Gabor filter and Faster R-CNN to effectively improve the accuracy of ship object detection in satellite images. Wang et al. (2017) proposed a ship detection method based on the SSD algorithm and transfer learning according to different image input sizes; two models, SSD-300 and SSD-500, were designed to test the detection performance. Guo et al. (2020) proposed a rotational Libra CNN method and introduced the concept of a balanced

https://doi.org/10.1017/S0373463321000503 Published online by Cambridge University Press

feature pyramid to improve the detection effect for different sizes of ships. At present, deep-learning algorithms are widely used in ship object detection. However, in real circumstances, the types of ships are diverse, their sizes are different and the water traffic environment is complex; in addition, it is necessary to consider recognition under night visual conditions, which poses a huge challenge to water object detection.

2.2. Object tracking

Object tracking is a key technology used to solve the real-time locking of moving objects in video feeds. Classic object-tracking algorithms include mean shift, particle filter and correlation filter. These methods mainly model the current object region and then find similar regions in the next image. Well-known correlation filtering algorithms include circulant structure tracking with kernels (CSK) (Henriques et al., 2012) and kernelised correlation filter (KCF) (Henriques et al., 2015), both of which determine the moving direction of an object by finding the relationship between two adjacent frames.

Compared to general motion environments, water traffic environments are more complex, including the partial or complete occlusion of adjacent ships, attitude and illumination changes, sudden scale changes and motion blur (Chen et al., 2014; Li and Zhu, 2014; Li et al., 2018). Chen et al. (2017) used the difference in the grey peak to detect the position of a ship and tracked ships based on the mean-shift algorithm. Dong et al. (2019) designed an object-tracking algorithm based on an improved KCF, in which a Kalman filter module was added to predict the position of the ship tracking object in the next frame, and the concept of object-tracking critical probability was used to evaluate whether the object tracking was abnormal. However, in maritime video feeds, because of occlusion, ships are often not detected. Cheng et al. (2019) improved the KCF algorithm by adding a scale-transform frame and tracking-effect detection method to solve the object occlusion problem or tracking failures caused by object detection. Chen et al. (2020) proposed an enhanced ship-tracking framework based on KCF and a curve-fitting algorithm and used a data anomaly detection and correction program to correct the positions of occluded ships.

With the development of deep-learning algorithms, such algorithms have been applied to object tracking. This method trains a CNN model to predict the object region in the next frame (Dorai et al., 2017; Pang et al., 2017). Recently, such algorithms have appeared in maritime applications, where researchers (Shan et al., 2020) have used a modified Siamese network combined with multi-region proposal networks to build a tracking pipeline to track maritime ships. ResNet-50 with a feature pyramid network structure was used as the CNN of the Siamese detection subnetwork.

However, research concerning multi-object ships is relatively sparse. Vivone et al. (2015) used a prior knowledge-based multi-object tracking method, prior information given by the ship channel, and its related motion model as the basic components of a variable structure interactive multi-model process and used the joint probabilistic data association rules to deal with false and missed detections. Xiao and Gang (2011) employed the continuously adaptive mean shift (camshaft) multi-ship tracking algorithm and multi-feature adaptive fusion based on colour, shape and texture to improve the robustness of the model.

3. Methodology

This paper uses a multi-object ship tracking mechanism based on the combination of a detector and a tracker and, using a data fusion algorithm, achieves AIS AR. The steps are as follows. (1) We enhance the night images to increase the contrast between the ship objects and the water surface background. (2) We use the SSD model to build a detector and introduce a self-attention mechanism to improve the model to enhance the accuracy of object detection. (3) We use the DeepSORT algorithm to build a tracker, according to the current tracking results, to predict the next position of the object and obtain the best match. (4) We establish a mapping between the ship positioning points and the world coordinate

https://doi.org/10.1017/S0373463321000503 Published online by Cambridge University Press

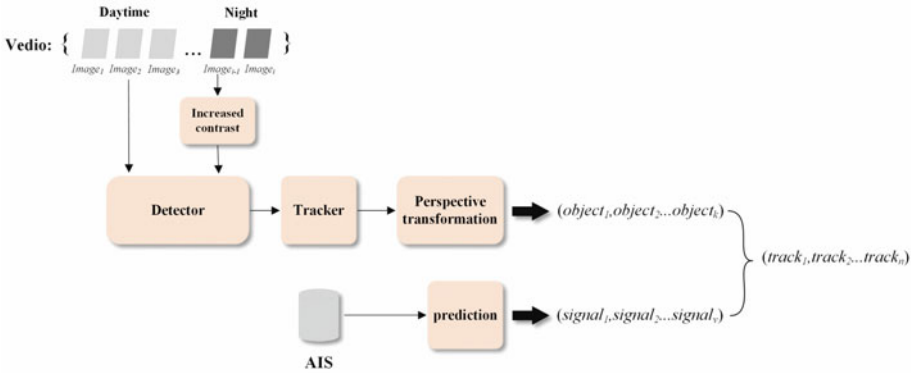

Figure 1. *Schematic of the algorithm discussed in this paper.*

system. (5) We track and predict the AIS signal via a distance calculation matching an AIS signal with a ship object. The structure of the algorithm discussed in this paper is shown in Figure 1.

3.1. Increased contrast at night

Computer vision began to dabble in maritime video-surveillance and finds many both civilian and military applications (Grimaldi et al., 2015). Therefore, it is necessary to improve the image quality before further analysis. Yang et al. (2019) estimated the initial brightness through Max-RGB and then refined it with the WLS filter. Based on the well-constructed brightness, the enhanced image at a low light level was obtained. In the inland waterway, the problem of night image detection mainly came from the low contrast between the ship target and background.

In this paper, we chose the RGB colour space based on the visible light image. The differences in red (R), green (G) and blue (B) in three dimensions cause the image to present different colour representations; however, it can be difficult to express the deep image characteristics. By constructing the pixel matrices hue (H), saturation (S) and brightness (V), the image can be easily quantified according to its colour, saturation and brightness. For image P, as shown in Figure 2(a), the corresponding S and V grey matrices are calculated as

$$\begin{cases} S_{ij} = 1 - \dfrac{3\min(R_{ij}, G_{ij}, B_{ij})}{R_{ij} + G_{ij} + B_{ij}} \\ V_{ij} = \dfrac{1}{\sqrt{3}}(R_{ij} + G_{ij} + B_{ij}) \end{cases} \tag{3.1}$$

where R_{ij}, G_{ij}, and B_{ij} represent the three colour weights for the pixel coordinates (i, j) after normalisation.

A generic ship tends to show high saturation and low brightness in night images; accordingly, the pixels in the ship region have higher S and lower V. The surface region in the background, conversely, exhibits lower S and higher V, e.g., $S_{\text{boat}} > S_{\text{river}}$ and $V_{\text{boat}} < V_{\text{river}}$. At the same time, because of the influence of the illumination on both sides of the water surface, the pixels have high gradient characteristics; therefore, the Sobel operator can be used to extract the gradient image of $P : L$, as shown in Figure 2(b), and then transform it into the grey pixel matrix L', as shown in Figure 2(c). Then,

https://doi.org/10.1017/S0373463321000503 Published online by Cambridge University Press

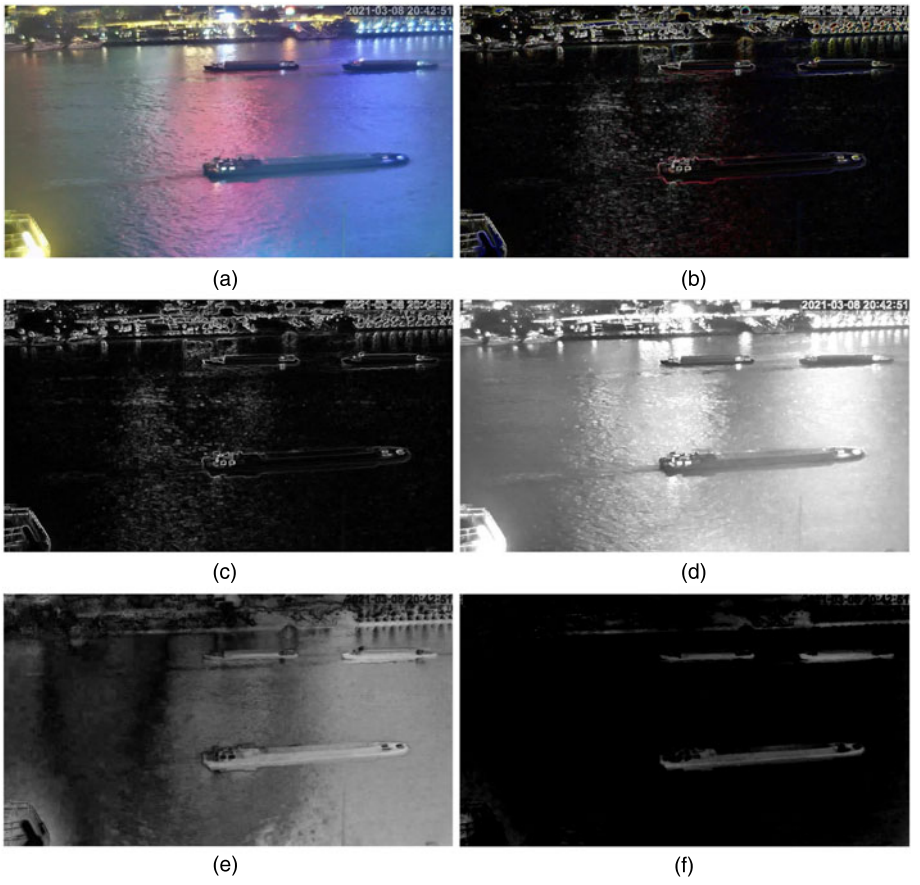

Figure 2. *Image enhancement process. (a) Original image. (b) Gradient characteristic image. (c) Greyscale image of gradient characteristic image. (d) Saturation image. (e) Brightness image. (f) Grey image after enhancement processing.*

the contrast of the water region pixels in S and V can be enhanced:

$$\begin{cases} S_{ij} = \dfrac{1}{\exp\left(\frac{L'_{ij}}{P_{\max}}\right)} S_{ij} \\ V_{ij} = \min\left(V_{ij} \times \exp\left(\dfrac{L'_{ij}}{P_{\max}}\right), \quad P_{\max}\right) \end{cases} \tag{3.2}$$

where $P_{\max} = 255$. The processed greyscale images of S and V are shown in Figure 2(d) and (e), respectively. It is obvious from the figure that the ship region is more obvious in the saturation image S than in the brightness image V. The ship region can be enhanced by subtracting the two images:

$$T_{ij} = S_{ij} - \min(S_{ij},\ V_{ij} - \varepsilon). \tag{3.3}$$

where T_{ij} = grey image after enhancement processing and the parameter $\varepsilon = 40\varepsilon$. The grey image T is shown in Figure 2(f). Now that the ship region is clearly expressed, the contrast enhancement calculation

https://doi.org/10.1017/S0373463321000503 Published online by Cambridge University Press

for the object region can be carried out using

$$P_{ij} = \begin{cases} \min\left(P_{\max} \times \left(\frac{P_{ij}}{P_{\max}}\right)^{\beta},\ P_{\max}\right) & T_{ij} > 0 \\ T_{ij} & T_{ij} = 0 \end{cases} \tag{3.4}$$

where P_{ij} = pixel after contrast enhancement and β = controllable parameter, where $\beta = 1.15$. Each pixel in T is compared with 0, and the ship region is segmented from T. Then, the RGB component of the corresponding position in the original image P is exponentially calculated. The processed image P is shown in Figure 2(f). The ship object in the figure shows higher contrast compared with the background region.

3.2. Self-attention mechanism

The self-attention mechanism is a recently proposed improved machine-learning method that enables the model to fully consider the relationship between the features of the samples in training and enhances the feature learning of key regions. This mechanism has been widely used in natural language processing and image learning (Cao et al., 2020). Because of the small size of the convolution kernel, CNNs only capture regional features with the same size and cannot effectively reflect the relationships between pixels in different regions. In this paper, we improve the network structure of SSD, based on the self-attention mechanism, so that it can analyse the correlations of all the pixels in the different scale feature maps, making the correlation of the extracted features stronger.

First, we input the feature map x with a structure $c \times w \times h$, use a 1×1 convolution kernel to convolute twice, and compress the channel number according to $c = c \div 8$ and obtain matrices F and G with structures $c' \times w \times h$, which are then transformed into $c' \times n$, where $n = w \times h$. The third convolution operation is performed using a 1×1 convolution kernel without channel number compression, and a matrix H with a $c \times w \times h$ structure is obtained, which is then transformed into $c \times n$. The matrices F and G are used to calculate the attention matrix

$$e_{i,j} = \frac{\exp(s_{ij})}{\sum_{i=1}^{N} \exp(s_{ij})}, \quad \text{where } s_{ij} = f(X_j)^T g(X_i), \tag{3.5}$$

where $e_{i,j}$ = element values of the attention matrix E, which has dimensions $n \times n$; s_{ij} = relative weight of the i position to the j position; N = number of elements in the characteristic graph matrix; and $f(x_i) = W_f x_i$, $g(x_j) = W_g x_j$, where W_f and W_g represent 1×1 convolutions.

We continue to multiply the matrices H and E to obtain a matrix with dimensions $c \times n$ as the self-attention map:

$$q_{i,j} = y(h(X_i)e(X_j)), \tag{3.6}$$

where $q_{i,j}$ = elements in the matrix Q and y = learning parameter, which is used to represent the dependence of the network on the relationship between the long-distance regions. $h(x_j) = W_h x_i$, where W_h is a 1×1 convolution.

The matrix Q is expanded into a characteristic graph with a $c \times w \times h$ structure as the output result of the self-attention module, as shown in Figure 3. The input structure is a characteristic graph with a $c \times w \times h$ structure. Because the structure of the entire process characteristic graph remains unchanged, the self-attention module can be directly added to the convolution network without changing its basic structure.

https://doi.org/10.1017/S0373463321000503 Published online by Cambridge University Press

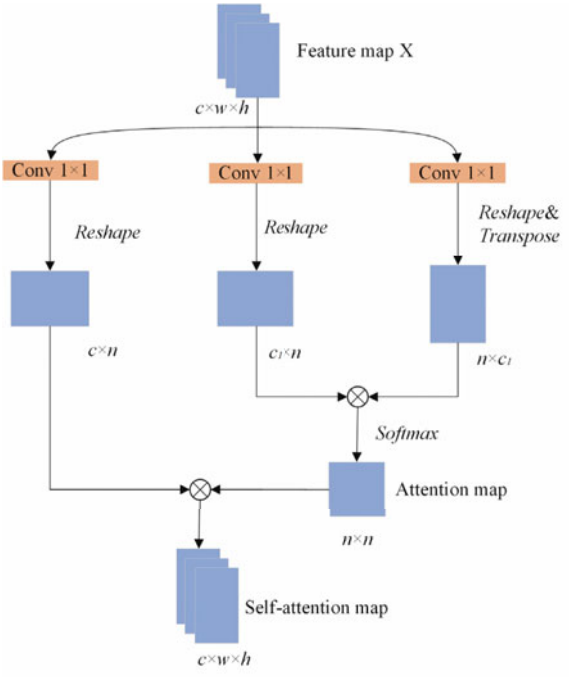

Figure 3. *Calculation method for a self-attention characteristic graph.*

3.3. The detector

The basic network of SSD was improved compared with VGG16. The full connection layer of the original VGG16 was deleted, and the depth of the convolution layer was increased. In addition, four convolution layers were added after the network to extract the network features. In the process of the network moving forward, a series of prediction scores of bounding box sets and categories are generated and then the final detection results are determined via nonmaximum suppression (NMS). As the depth of the network increases, the size of the feature map decreases and the NMS layer collects the object information at each scale; therefore, the default box generated by SSD is multiscale.

As shown in Figure 4, based on the SSD network, the self-attention module is added after the 4th, 7th, 8th, 9th, 10th and 11th layers. The output characteristic diagram of the original layer is taken as the input of the self-attention module, and the corresponding feature map is the output. With the deepening of the network, the scale of the generated feature map decreases, and different scale feature maps produce different default boxes. NMS is used again to fuse the self-attention feature maps of the different receptive fields, and the final detection result is calculated.

3.4. The tracker

This component of the algorithm must continuously track all the ship objects detected by the detector. Multiple-object tracking (MOT) is used to track and extract the trajectories of multiple objects of interest in a video sequence. As opposed to single-object tracking, MOT must identify different objects in each frame. For new objects, it must generate new trajectories, and for objects that have left the field of view of the camera, it must terminate the tracking of their trajectories.

The traditional multi-object tracking algorithm is based on correlation filtering, with its representative algorithm being a KCF-based multi-object tracking algorithm (Henriques et al., 2015) that uses multithreading to track multiple single objects. The SORT (Bewley et al., 2016) algorithm proposed in 2016 regards multi-object tracking as a data association problem. SORT is based on accurate object

https://doi.org/10.1017/S0373463321000503 Published online by Cambridge University Press

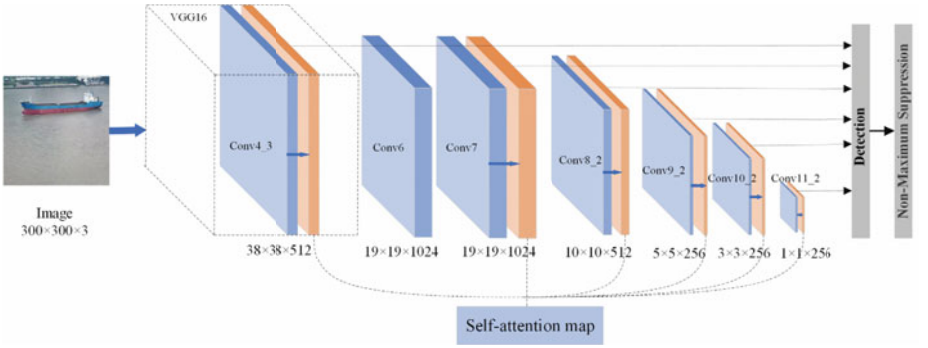

Figure 4. *Improved SSD network.*

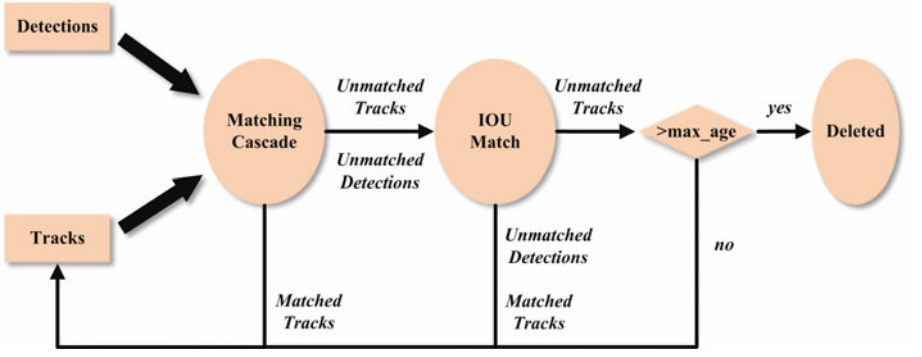

Figure 5. *DeepSORT algorithm processes.*

detection, according to the location information in the detection box, and uses Kalman filtering and the Hungarian algorithm to match the tracking objects of the front and back two frames of the image. However, because SORT ignores the surface characteristics of the object being detected, it is easy to lose an object.

DeepSORT (Wojke et al., 2017) has been improved, combining object motion and surface feature information and using the Mahalanobis and cosine distances as a measure of the similarity between the motion and depth features within the detection box, while at the same time using the Hungarian algorithm to cascade match the predicted track with the detections in the current frame, giving priority to the objects that have not been lost. Then, use the minimum IOU threshold to filter the low confidence matching to reduce error matching, and set a threshold to remove trackers that loop too many times. The algorithm structure is shown in Figure 5.

3.5. Multimodal data fusion

(1) Perspective transformation

Perspective transformation is the process of projecting an image onto a new view plane by establishing a projection matrix. The projection process is the product of the coordinate vector and the transformation matrix, as

$$[x' \quad y' \quad w'] = [u \quad v \quad w] \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \tag{3.7}$$

https://doi.org/10.1017/S0373463321000503 Published online by Cambridge University Press

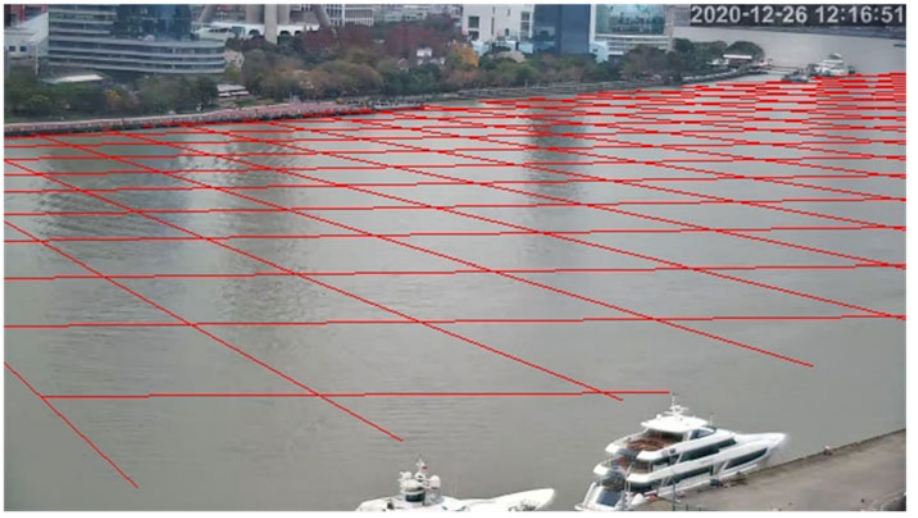

Figure 6. *Perspective effect of a virtual grid in world coordinates on an image.*

where u and v = coordinates of the original plane; the parameter $w = 1$; the perspective matrix contains the zoom, rotation and translation operations of the plane; and $a_{33} = 1$. The transformed coordinates are represented as (u', v'), where $u' = x'/w'$ and $v' = y'/w'$. Entering this into Equation (3.7), we obtain Equation (3.8). According to this formula, the parameters in the perspective matrix can be calculated by simply identifying four sets of key points:

$$\begin{cases} u' = \dfrac{a_{11} \times u + a_{21} \times v + a_{31} \times 1}{a_{13} \times u + a_{23} \times v + 1 \times 1} \\ v' = \dfrac{a_{12} \times u + a_{22} \times v + a_{32} \times 1}{a_{13} \times u + a_{23} \times v + 1 \times 1} \end{cases}. \tag{3.8}$$

In this paper, the principle of perspective transformation is used to project the image coordinate system onto the longitude and latitude coordinate system. The process of transforming the pixel coordinate to longitude and latitude coordinates is called the forward perspective, while the opposite is called the reverse perspective. If we reverse perspective the plane of the river water, the effect is shown in Figure 6, which demonstrates the reverse perspective effect of a square virtual grid with a side length of 0·00045 in the longitude and latitude coordinate system.

(1) Ship position estimation

The previous steps transform any pixel coordinate into longitude and latitude coordinates on the river surface in an image. However, the longitude and latitude data in AIS are confirmed by the shipborne GPS, which is often close to the bow or stern of the ship; therefore, there may be a gap of tens of meters between the two sets of coordinates. This paper takes the ship centre as the estimated positioning point and marks the pixel coordinates of the ship centre position corresponding to the river surface in the video.

This is shown in Figure 7, where the point O on the long side of the detection box is the selected water surface projection point. A ray is traced along the angle θ to the right boundary and intersects with the boundary box at the point F, where θ is obtained by the DeepSORT tracker, such that the line segment OF can approximately represent the waterline of the ship. The position of B in world coordinates is

$$P_B^{'} = \frac{1}{2} \cdot (\text{Transform}(P_O) + \text{Transform}(P_F)), \tag{3.9}$$

https://doi.org/10.1017/S0373463321000503 Published online by Cambridge University Press

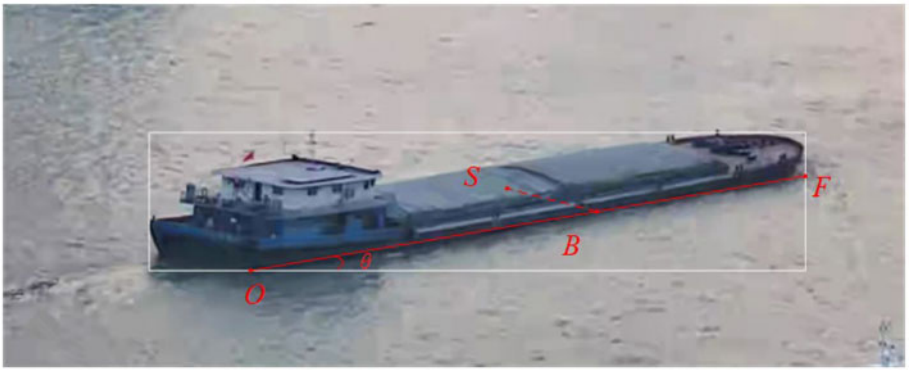

Figure 7. *Estimation method for the ship GPS positioning points.*

where P is used to represent the pixel coordinates of each point and Transform () represent the forward perspectives. From point B along the vertical travel direction, a ray is traced into the detection box to the ship positioning point S. The length of this line segment in world coordinates, l'_{BS}, is determined by the length/width ratio of the ship, such that

$$l'_{BS} = \frac{1}{2} \cdot \delta \cdot \text{Distance}(\text{Transform}(P_O), \text{Transform}(P_F)), \tag{3.10}$$

where δ represents the length/width ratio, which for bulk carriers, tankers and river ships is approximately 1 : 6 and for yachts, cruise ships and tugboats is approximately 1 : 3; Distance () is used to calculate the Euclidean distance between two points. On this basis, the position S in the world coordinate system P'_S can be obtained as

$$P'_S = P'_B + (l'_{BS} \cdot \sin(\beta), \quad l'_{BS} \cdot \cos(\beta)), \tag{3.11}$$

where β = azimuth of point S with respect to point B in world coordinates. Then, we can obtain the pixel coordinates of the point S:

$$P_S = \text{Transform}'(P'_S), \tag{3.12}$$

where Transform′() = reverse perspective.

(1) AIS signal prediction and data fusion

Next, we demarcate the object region within the longitude and latitude coordinate system. We receive the dynamic AIS information in the object region and extract the Maritime Mobile Service Identity (MMSI) number, longitude and latitude, direction and speed information. We track the latest signal points of the ship according to the MMSI number, where $D_{mmsi,t} = \{\text{lon}, \text{lat}, \text{sog}, \text{cog}\}$. The AIS signal is updated only every few seconds and is out of sync with the video signal; therefore, it is necessary to track the ship and continuously predict its position at every moment $A_{mmsi,t'} = \{\text{lon}', \text{lat}'\}$ according to $D_{mmsi,t}$, such that

$$\text{lon}' = \text{lon} + (t' - t) \cdot \text{sog} \cdot \sin(\text{cog} \cdot \pi/180), \tag{3.13}$$

$$\text{lat}' = \text{lat} + (t' - t) \cdot \text{sog} \cdot \cos(\text{cog} \cdot \pi/180), \tag{3.14}$$

Therefore, the position of AIS objects at each moment are expressed as P_A, the position of video ship objects after the positive perspective transformation are expressed at the same moment as P'_S, two points constitute the vector $R_1 = \overrightarrow{P'_S P_K}$. Their forward directions are represented by the vector

https://doi.org/10.1017/S0373463321000503 Published online by Cambridge University Press

Table 1. *Comparison of the accuracy of the test set detection results (%).*

	mAP	River ships	Bulk carriers	Cruise ships	Tankers	Tugboats	Yachts
SSD	82·83	86·43	82·48	89·93	71·26	90·82	76·03
SSD+	87·53	91·54	89·60	97·78	73·58	95·61	77·08

$R_2 = P_S^{'} + \overrightarrow{P_O^{'} P_F^{'}}$, and we calculate the distance between the two as

$$d = \text{Normalisation}(P_S^{'},\ P_{A2}) - \left| \frac{R_1 \cdot R_2}{R_1 R_2} \right| + 1, \tag{3.15}$$

where Normalisation() represents a normalised function, the first half of the formula represents the Euclidean distance between two points, and the second half represents the cosine similarity between the vectors. The range of values of distance is $[0, 2]$.

Using the near-matching mechanism, information pairs with smaller distances can be matched first.

4. Results

4.1. The experiment platform

The experimental scene selected in this paper is the Bund section of the Huangpu River, Yangtze River channel, which is in the central region of Shanghai, shown in Figure 8. During the day, this area accommodates many bulk carriers and river ships, which differ in distance and size. At night, the traffic volume increases and there are more cruise ships and yachts on the water surface. Under the illumination of the lights on both sides of the channel, the water surface is colourful and provides a variety of visual conditions for passing ships.

We used a long focal distance network camera to receive a video signal with a resolution of 2550 × 1440. The AIS signal receiving base station was set at the same location, and the video and AIS signals were transmitted to the remote experimental platform via the network at the same time.

Using the computer vision method for ship detection and tracking, this experiment employed an original, independently established dataset and a monitoring probe located on the river bank. Images of the channel were collected from different shooting angles, under different weather conditions, and for different time periods throughout the day, resulting in a total of 2,700 images, including daytime, night-time, rainy weather and foggy weather, in which 6,330 ships were marked. The collected image resolution was adjusted to 300 × 300, and the ship dataset was divided into six categories: river ships, bulk carriers, cruise ships, tankers, tugboats and yachts. We manually distinguished the night images, processed using the method in Section 3.1, and fed the processed images into the improved SSD–DeepSORT frames for training. We used Python programming software, and the training environment consisted of an Ubantu operating system, PyTorch 1.0 framework, TensorFlow 1.14 frame, GTX 1050ti (2 blocks) GPU, and 10 G of memory, using the NVIDIA CUDA9.0 acceleration toolbox.

4.2. Experiment and analysis

As can be seen from Table 1, the average detection accuracy of the improved SSD model in the six ship categories is 87·53, which is 4·7 higher than the original SSD model. The accuracy for the river ships, which are the most frequent type of ship, increased by 5·11; the accuracy for the cruise ships improved the most, by 7·85; and the accuracy increase for the yachts was smaller, only 1·05. Results show that the detector proposed in this paper has a higher detection performance than that of the original method.

We used continuous frames to test the effect of tracing and to draw the ship trajectories. Figure 9 shows the object-tracking test results. According to the objects detected in Figure 9(a),

https://doi.org/10.1017/S0373463321000503 Published online by Cambridge University Press

Figure 8. *Map representation of the experimental area.*

Even in rainy weather, some small-sized ship targets can still be tracked effectively. Figure 9(b) shows that the detection and tracking algorithm can still performs well in foggy weather. In Figure 9(c), the ships and the water surface region are dim and the degree of discrimination is not large. However, the trajectories of the objects indicate that the algorithm can solve the problem of ship object tracking at night. Figure 9(d) indicates that the algorithm can also achieve good detection and tracking effects for crowded ships.

After capturing the objects in the video, the angle ranging from 0° to 35° between the line of two central pixels one second apart and the horizontal line is figured out. Then according to the method described in Section 3.5, the real scene was modelled and coordinate matches between the video and AIS objects were made. The result is shown in Figure 10.

To verify the effect of the algorithm on the information enhancement at different times, we selected a video with a length of 1 h, selected a frame with an interval of 3·5 s to test and compiled statistics of

https://doi.org/10.1017/S0373463321000503 Published online by Cambridge University Press

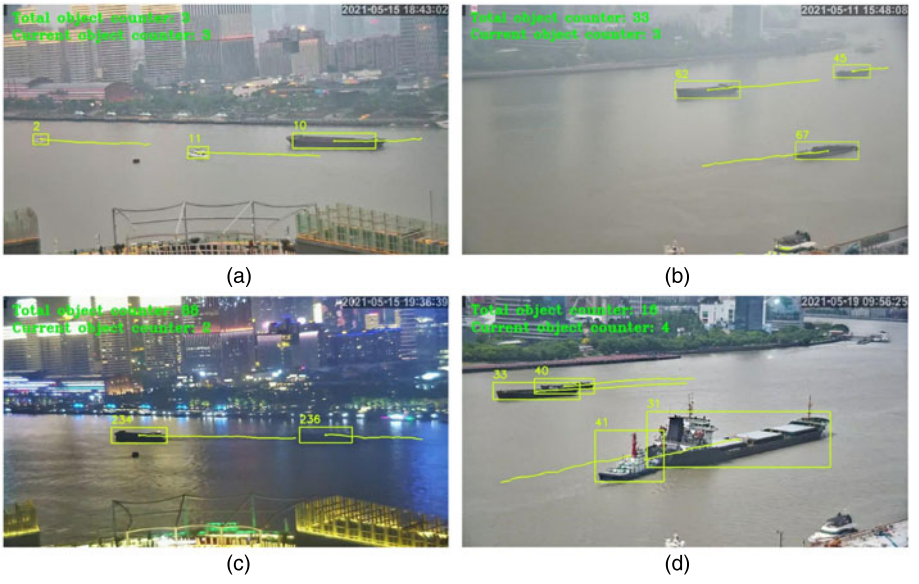

Figure 9. *Object tracking test results. (a) test in rainy weather. (b) test in foggy weather. (c) Test at night. (d) Test of crowded ships.*

the relevant indicators in each frame. These indicators included the total number of ships, the number of tracked ships, the number of successful matches with AIS information and the number of correct matches with AIS information. Then, all the indicators were summed over 12 time periods. Because the detection difficulty for ships in day and night scenes is different, and the types of ships on the waterway during these periods are different, 1 h each of day and night scenes was selected; the statistical results for the day and night scenes are shown in Figure 11(a) and (b), respectively.

5. Discussion

As can be seen in Figure 11, the gap between the total number of ships and the number of tracked ships is large, indicating that, in some cases, some ships fail to form a tracking trajectory because they are distant and small, resulting in object tracking interruption or failure to track. In most time periods, the gap between the number of tracked ships and the number of successful matches is significant, indicating that, even if the ship has a greater probability of successfully matching the AIS signal under the premise of successful tracking, in some cases, because of the gap between the AIS forecast position and the actual position of the ship, the matching conditions cannot be met. The gap between the number of successful matches and the number of correct matches is generally small, indicating that the AIS information for the successful matches has a high accuracy.

An analysis of Table 2 reveals that, given a period of 1 h, the total number of ships at night is 2,319, reflecting an increase of 525 compared with the same period during the day. This suggests that the traffic flow on the waterway is significantly greater at night than during the day. According to the summation calculation results for each indicator in Table 2, the ratios of the number of tracked ships to the total number of ships during the day and at night are 87·0% and 83·4%, respectively, which indicates that it is easier to lose or not track a ship at night. This is because there are more smaller yachts at night and the self-lighting of these ships makes it difficult to separate them from the reflected light on the water surface, leading to object leakage. The ratios of the number of successful matches to the number of tracked ships during the day and night were 93·0% and 89·5%, respectively; the difference between the two ratios is obvious. This is because the error in the ship signal position prediction at night makes it more difficult to meet the matching conditions. The ratios of the number of correct matches to the

https://doi.org/10.1017/S0373463321000503 Published online by Cambridge University Press

(a)

(b)

Figure 10. *AIS information visualisation. (a) During the day. (b) At night.*

number of successful matches during the day and night were 96·6% and 94·4%, respectively, indicating that the correct rate of matched information is lower at night than during the day because the ships are denser at night, resulting in interference between the ships, which are closer to each other. The overall accuracy of the algorithm is represented by the proportion of the number of correct matches to the total number of ships; the accuracies during the day and night were calculated to be 78·2% and 70·4%, respectively. The algorithm can visualise the AIS information for most ships in the surveillance video and that the display effect is obviously better during the day than during the night.

As shown in Figure 10, the information displayed in the upper part of the ship detection box includes the MMSI number, type, speed and course, where the type is derived from the detector and the rest of the information is obtained from the AIS dynamic information and is constantly updated. However, since the missed detection and occlusion of the ship will cause the temporary disappearance of the tracking target, the cascade matching algorithm of DeepSORT can retain the missing track information for a

https://doi.org/10.1017/S0373463321000503 Published online by Cambridge University Press

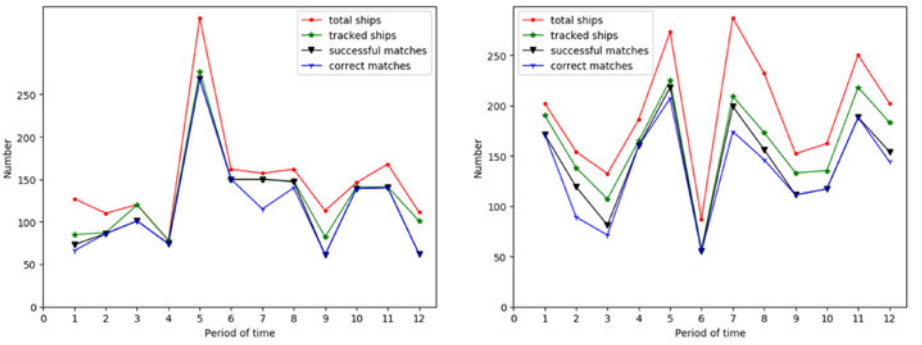

Figure 11. *Test results for each index during different periods. (a) 1 h during the day. (b) 1 h during the night.*

Table 2. *Test results for each index for 1 h.*

		Result	
Order	Index	Day	Night
I_1	All vessels	1794	2319
I_2	Tracked ships	1560	1933
I_3	Successful matches	1451	1729
I_4	Correct matches	1402	1633
I_5	I_2/I_1 (%)	87·0	83·4
I_6	I_3/I_2 (%)	93·0	89·5
I_7	I_4/I_3 (%)	96·6	94·4
I_8	Accuracy: I_4/I_1 (%)	78·2	70·4

period of time, which requires the combination of the video frame rate and the ship moving speed to set the optimal *max_age*. These are basic pieces of information concerning a ship that can help an observer quickly understand the driving status of the ship and can be used to search for the ship data and voyage information according to the ship MMSI number. Our method can enhance the display of the basic ship information in a channel video. It can also be adapted to similar scenes, and further developments can be applied to various channel video surveillance platforms to achieve the intelligent supervision of waterways.

6. Conclusions

VTSs are key to ensuring the efficient navigation of ships. This paper makes full use of channel surveillance video and AIS data to design a ship information tracking scheme. We standardise virtual information with actual ships and superimpose a visual expression of AIS information on video images. The algorithm was tested on day and night scenes, and the results show that the algorithm can be effective in real scenes. This study proposes a constructive plan for an intelligent upgrade for river shipping supervision.

This paper proposes a solution to the problem that the contrast between a ship object and background is not strong in night images, which is helpful for detecting ordinary bulk ships, river vessels, tankers and tugboats because of their special appearance in HSV images, but not for ship objects with higher brightness; therefore, this method has local applicability in night scenes. In this paper, a popular object

https://doi.org/10.1017/S0373463321000503 Published online by Cambridge University Press

detection algorithm was selected and improved upon, and the advantages of the improved algorithm were verified via comparative tests. Tracking tests were performed on both day and night scenes. The application of computer vision principles and algorithms illustrates that artificial intelligence has broad prospects for maritime applications. Based on the completion of image object positioning, this paper designed a signal-fusion calculation method based on the perspective transformation, simulating the positioning point of a ship, and predicting the AIS signal position that can be applied to a variety of overlooking monitoring scenes. To achieve coordinate matching between the virtual information and a ship object by combining the Euclidean distance and the cosine similarity between the multimodal data.

The establishment of the ship information tracking model in this paper needs to be adjusted according to different video images. The target detection based on SSD can adapt to the change of the camera's focal length in a certain range, but the camera has different perspectives to collect the video of the waterway; thus, it is necessary to consider the multiangle features of the ship when building the ship data set. Under different focal lengths and angles of view, different coordinates of reference points are also needed to establish the perspective transformation model, to ensure the basic conditions of interrelation between objects.

Financial statement. This research was funded by the National Natural Science Foundation of China (Grant No. 71804059), Social Development Major Project of Shanghai Municipal Science and Technology Commission (grant number 18DZ1206300).

References

An, J., Qiao, T., Yang, X., Hong, H. and Bai, X. (2019). Design of a Visual Analysis Platform for Sea Route Based on AIS Data. *2nd International Conference on Artificial Intelligence and Big Data (ICAIBD)*. Chengdu: IEEE.

Betti, A., Michelozzi, B., Bracci, A. and Masini, A. (2020). Real-Time Object Detection in Maritime Scenarios Based on YOLOv3 Model. *the 9th International Symposium on Optronics in Defence and Security*. Paris: 3AF.

Bewley, A., Ge, Z., Ott, L., Ramos, F. and Upcroft, B. (2016) Simple Online and Realtime Tracking. IEEE *International Conference on Image Processing (ICIP)*, Phoenix: IEEE.

Can, X. (2017). Research and simulation of information fusion technology for inland river AIS and VTS. *Ship Science and Technology*, **39**(22), 49–51.

Cao, J., Chen, Q., Guo, J. and Shi, R. (2020). Attention-guided Context Feature Pyramid Network for Object Detection. *Computer Vision and Pattern Recognition*. Online: IEEE.

Chen, D., Yuan, Z., Wu, Y., Zhang, G. and Zheng, N. (2014). Constructing Adaptive Complex Cells for Robust Visual Tracking. IEEE *International Conference on Computer Vision*, Sydney: IEEE.

Chen, Z., Li, B., Tian, L. F. and Chao, D. (2017). Automatic Detection and Tracking of Ship Based on Mean Shift in Corrected Video Sequences. *The 2nd International Conference on Image, Vision and Computing (ICIVC)*, Chengdu: IEEE

Chen, X., Xu, X., Yang, Y., Wu, H., Tang, J. and Zhao, J. (2020). Augmented ship tracking under occlusion conditions from maritime surveillance videos. *IEEE Access*, **8**, 42884–42897.

Cheng, Z., Zhilin, L. and University, H. E. (2019). Application of improved kernel correlation filtering algorithm in small ship dynamic object tracking. *Applied Science and Technology*, **46**(01), 36–42.

Dong, C., Zheng, B., Li, B., Tian, L. F. and Liu, W. (2019). Shiptarget tracking with improved kernelized correlation filters. *Optics and Precision Engineering*, **27**(4), 911–921.

Dorai, Y., Chausse, F., Gazzah, S. and Amara, N. E. B. (2017). Multi Object Tracking by Linking Tracklets with aConvolutional Neural Network. *International Conference on Computer Vision Theory and Applications*, Porto: IEEE.

Frydenberg, S., Nordby, K. and Eikenes, J. O. (2018). Exploring designs of augmented reality systems for ship bridges in Arctic waters. *Human Factors*. London: RINA.

Girshick, R. (2015) Fast R-CNN. *International Conference on Computer Vision (ICCV)*. Santiago: IEEE.

Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. IEEE *Conference on Computer Vision and Pattern Recognition (CVPR)*. Columbus: IEEE.

Grimaldi, M., Bechar, I., Lelore, T., Guis, V. and Bouchara, F. (2015). An Unsupervised Approach to Automatic Object Extraction From A Maritime Video Scene. *4th International Conference on Image Processing Theory, Tools and Applications (IPTA)*. Paris: IEEE.

Guang, Y., Qichao, L. and Feng, G. (2011). A Novel Ship Detection Method Based on Sea State Analysis From Optical Imagery. *Sixth International Conference on Image and Graphics*. Hefei: IEEE.

Guo, H., Yang, X., Wang, N., Song, B. and Gao, X. (2020). A rotational libra R-CNN method for ship detection. *IEEE Transactions on Geoence and Remote Sensing*, **58**(8), 5772–5781.

He, L., Yi, S., Mu, X. and Zhang, L. (2019). Ship Detection Method Based on Gabor Filter and Fast RCNN Model in Satellite Images of Sea. *the 3rd International Conference on Computer Science and Application Engineering(CSAE)*, **111**, 1–7.

Henriques, J. F., Rui, C., Martins, P. and Batista, J. (2012). Exploiting the Circulant Structure of Tracking-by-Detection with Kernels. *The 12th European Conference on Computer Vision (ECCV)*, Berlin: Springer.

Henriques, J. F., Caseiro, R., Martins, P. and Batista, J. (2015). High-Speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**(3), 583–596.

Huang, Y., Li, Y., Zhang, Z. and Liu, W. (2020). GPU-Accelerated Compression and visualization of large-scale vessel trajectories in maritime IoT industries. *IEEE Internet of Things Journal*, **7**(11), 10794–10812.

Hugues, O., Cieutat, J. M. and Guitton, P. (2014). Real-time infinite horizon tracking with data fusion for augmented reality in a maritime operations context. *Virtual Reality*, **18**(2), 129–138.

Kartika, I. V., Siswandari, N. A and Puspitorini, O. (2018). Application of Genetic Algorithm for Placement of AIS (Automatic Identification System) Base Station. *Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)*. Batu: IEEE.

Kim, H. T., Park, J. S. and Yu, Y. S. (2010). Ship detection using background estimation of video and AIS informations. *Journal of the Korea Institute of Information and Communication Engineering*, **14**(12), 2636–2641.

Lee, J. M., Lee, K. H. and Nam, B. (2016). Study on Image-Based Ship Detection for AR Navigation. *6th International Conference on* IT *Convergence and Security (ICITCS)*. Prague: IEEE.

Li, Y. and Zhu, J. (2014). A scale adaptive kernel correlation filter tracker with feature integration. *Lecture Notes in Computer Science*, **8926**, 254–265.

Li, F., Tian, C., Zuo, W., Zhang, L. and Yang, M. H. (2018). Learning Spatial-Temporal Regularized Correlation Filters for Visual Tracking. IEEE/CVF *Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City: IEEE.

Liu, J. (2010). Moving ship detection and tracking from infrared image for collision avoidance of ships. *Opto-Electronic Engineering*, **37**(9), 8–13.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y. and Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. *European Conference on Computer Vision (ECCV)*. Amsterdam: Springer.

Liu, W., Nie, J., Garg, S., Xiong, Z. and Hossain, M. S. (2020). Data-Driven trajectory quality improvement for promoting intelligent vessel traffic services in 6G-enabled maritime IoT systems. *IEEE Internet of Things Journal*, **8**(7), 5374–5385.

Lukas, U., Vahl, M. and Mesing, B. (2014). Maritime Applications of Augmented Reality- Experiences and Challenges. *Virtual, Augmented and Mixed Reality. Applications of Virtual and Augmented Reality - 6th International Conference*. Berlin: Springer.

Oh, J., Park, S. and Kwon, O. S. (2016). Advanced navigation aids system based on augmented reality. *International Journal of E Navigation & Maritime Economy*, **5**(C), 21–31.

Pang, S., Coz, J. J. D., Yu, Z., Luaces, O. and Diez, J. (2017). Deep learning to frame objects for visual target tracking. *Engineering Applications of Artificial Intelligence*, **65**(oct.), 406–420.

Rao, A., Wang, H., Hu, Z. C. and Mullane, J. (2014). A Gaussian Particle Filter Based Factorised Solution to the Simultaneous Localization and Mapping Problem. *Advanced Robotics and Its Social Impacts*. Tokyo: IEEE.

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. IEEE *Conference on Computer Vision and Pattern Recognition (*CVPR*)*. Las Vegas: IEEE

Ren, S., He, K., Girshick, R. and Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**(6), 1137–1149.

Shan, Y., Zhou, X., Liu, S., Zhang, Y. and Huang, K. (2020). Siamfpn: A deep learning method for accurate and real-time maritime ship tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, doi:10.1109/TCSVT.2020.2978194

Vivone, G., Braca, P. and Horstmann, J. (2015). Knowledge-Based multiobject ship tracking for HF surface wave radar systems. *IEEE Transactions on Geoence and Remote Sensing*, **53**(7), 3931–3949.

Wang, Y., Wang, C., Hong, Z., Cheng, Z. and Fu, Q. (2017). Combing Single Shot Multibox Detector with Transfer Learning for Ship Detection Using Chinese Gaofen-3 Images. *Progress in Electromagnetics Research Symposium - Fall (PIERS - FALL)*. Singapore: IEEE.

Wojke, N., Bewley, A. and Paulus, D. (2017). Simple Online and Realtime Tracking with A Deep Association Metric. IEEE *International Conference on Image Processing (ICIP)*, Beijing: IEEE.

Xiao, Y. and Gang, X. (2011). Camshift ship tracking algorithm based on multi-feature adaptive fusion. *Opto- Electronic Engineering*, **38**(5), 52–58.

Yang, M., Nie, X. and Liu, R. W. (2019). Coarse-to-Fine Luminance Estimation for Low-Light Image Enhancement in Maritime Video Surveillance. IEEE *Intelligent Transportation Systems Conference (ITSC)*. Auckland: IEEE.

Cite this article: Huang Z, Hu Q, Mei Q, Yang C, Wu Z (2021). Identity recognition on waterways: a novel ship information tracking method based on multimodal data. *The Journal of Navigation* **74**: 6, 1336–1352. https://doi.org/10.1017/S0373463321000503