PAPER

Maritime Target Detection Based on Electronic Image Stabilization Technology of Shipborne Camera

Xiongfei SHAN[†], **Mingyang PAN**[†a)], **Depeng ZHAO**[†], **Deqiang WANG**[†], **Feng-Jang HWANG**[††], ***Nonmembers***, ***and*** **Chi-Hua CHEN**[†††], ***Member***

SUMMARY During the detection of maritime targets, the jitter of the shipborne camera usually causes the video instability and the false or missed detection of targets. Aimed at tackling this problem, a novel algorithm for maritime target detection based on the electronic image stabilization technology is proposed in this study. The algorithm mainly includes three models, namely the points line model (PLM), the points classification model (PCM), and the image classification model (ICM). The feature points (FPs) are firstly classified by the PLM, and stable videos as well as target contours are obtained by the PCM. Then the smallest bounding rectangles of the target contours generated as the candidate bounding boxes (bboxes) are sent to the ICM for classification. In the experiments, the ICM, which is constructed based on the convolutional neural network (CNN), is trained and its effectiveness is verified. Our experimental results demonstrate that the proposed algorithm outperformed the benchmark models in all the common metrics including the mean square error (MSE), peak signal to noise ratio (PSNR), structural similarity index (SSIM), and mean average precision (mAP) by at least −47.87%, 8.66%, 6.94%, and 5.75%, respectively. The proposed algorithm is superior to the state-of-the-art techniques in both the image stabilization and target ship detection, which provides reliable technical support for the visual development of unmanned ships.

key words: *shipborne camera; electronic image stabilization; points line model; feature points; convolutional neural network*

1. Introduction

In recent years, artificial intelligence [1]–[3] develops rapidly, and it is widely used in fields such as unmanned driving and computer vision, shipborne cameras are gradually applied to unmanned ships with the characteristics of intuitiveness, reliability, rich information and cost-effective performance. Among smart ship projects around the world, almost all projects adopt cameras as important sensors of situational awareness to provide reliable data source for intelligent decision-making, such as Maritime Unmanned Navigation through Intelligence in Networks (MUNIN) [4], Rolls-Royce's Advanced Autonomous Waterborne Applications (AAWA) [5], and DNV GL's concept ship "The Revolt" [6].

As an important task of ship situational awareness technology, maritime target detection has the following characteristics. First, the shipborne camera generally bumps and sways along with the ship's navigation due to the influence of the wind and waves, which causes a six-degree-of-freedom low-frequency sway, at the same time, mechanical vibrations of the ship's main engine causes high-frequency shaking [7]. Both affect the stability of videos, which in turn leads to relatively large irregular movements of the background pixels and foreground pixels. Second, due to the characteristics of the complex maritime environment, the ships' posture, lighting and other conditions change, even the same target ships may have different Intersection over Union (IoU) scores in consecutive frames. With a lower threshold, ships in consecutive frames can be detected, but a large number of false positive samples generate, resulting in a higher false detection rate. With a higher threshold, it generates many false negative samples, resulting in a higher missed detection rate [8].

Target detection algorithms can be divided into two categories, namely target detection based on background modeling and target detection based on foreground modeling [9]. The former models the image background and segments the moving targets through background subtraction technology, but due to the influence of the shipborne camera's own motion, this type of algorithm is difficult to accurately model the moving background, which leads to the failure of target ship detection. The latter mainly extracts appearance features of the targets through convolutional neural networks (CNN) and generates the bounding boxes (bboxes) and confidence scores. However, this type of algorithm ignores the spatio-temporal relations between consecutive frames, and the detection speed is slow due to the repeated feature calculation of a large number of overlapping candidate bboxes.

In the maritime video sequences, the inter-frame motion of pixels includes the background pixel motion and foreground pixel motion. The background pixel motion is mainly formed by the motion of the camera, while the foreground pixel motion includes the motion of the target ship itself in addition to the motion of the camera. It can be seen that accurate classification of background pixels and foreground pixels in the sea-sky image is crucial for video image stabilization and target detection tasks. However, existing studies generally consider these two technologies separately. The image stabilization algorithms do not consider the effective detection of the targets after acquiring stable videos, and the target detection technologies handle stable video by default. This is inconsistent with the actual detection of maritime

Manuscript received July 9, 2020.

[†]The ~~author~~ are with the Navigation College, Dalian Maritime University, China.

[††]~~The author~~ is with the School of Mathematical and Physical Sciences, Transport Research Centre, University of Technology Sydney, Australia.

[†††]~~The author~~ is with the College of Mathematics and Computer Science, Fuzhou University, China.

a) E-mail: E-mail: panmingyang@dlmu.edu.cn

DOI: 10.1587/transinf.E0.D.1

targets.

This study proposes a novel maritime target detection algorithm, which effectively integrates electronic image stabilization technology and maritime target detection technology to achieve the target ship detection closer to actual maritime conditions. The algorithm flowchart is show in Figure 1. The blue region, the green region, and the yellow region represent the points line model (PLM), the points classification model (PCM), and the image classification model (ICM), respectively. The PLM mainly processes the feature points (FPs) and sea-sky lines (SSLs) in two consecutive frames. The corner region and edge region are judged by the feature value of the corner detection. In the edge region, the SSLs are extracted, the region of interest (ROI) is determined, and the SSL motions in two consecutive frames are estimated. In the corner region, the optical flows are clustered through the points-line fusion to obtain background FPs (BFPs) and foreground FPs (FFPs). The PCM mainly processes background BFPs and FFPs. For the BFPs, the stable frames are obtained through the motion estimation, filter smoothing and motion compensation of the image stabilization algorithm, and then the difference images of the stable frames are calculated. For the FFPs, the transformation matrices before and after the image stabilization are used to transform the FFPs of two consecutive frames into the stable frames. Therefore, the candidate bboxes are obtained through the PCM. Finally, the ICM classifies the candidate bboxes and displays the classification results and position information in the stable video sequences.

The remainder of the study is organized as follows. Section 2 discusses the literature review of electronic image stabilization and target detection technologies. The construction of the PLM, the PCM and the ICM are presented in Section 3, Section 4 and Section 5, respectively. Section 6 shows the experimental results in practice. The contributions of this study are summarized and the future work is suggested in Section 7.

2. Literature Reviews

Considering the rapid development of electronic image stabilization and target detection technologies, this study proposes a main technical framework based on the comprehensive analysis of relevant research.

2.1 Electronic image stabilization technology

Electronic image stabilization technology mainly includes three parts: motion estimation, motion filtering, and motion compensation. Motion estimation is the premise of electronic image stabilization technology. Accurate global motion estimation directly determines the performance of image stabilization. The global motion estimation algorithms are widely used based on image block matching methods [10]–[12] and FPs matching methods [13]–[15]. In general, due to the invariance of translation, rotation, and lighting of FPs, they are more effective than the image block matching method, but the above algorithms are only applicable to the cases where there are no significant moving targets in the video. To cope with this problem, Grundmann et al. [16] used a grid-based segmentation technology, which applied the Random Sample Consensus (RANSAC) algorithm [17] to motion estimation in each grid. This algorithm improved the estimation accuracy, but it still did not distinguish the BFPs from the FFPs accurately . Kim et al. [18] proposed an FPs classification algorithm that used a projection transformation to transform the FPs from one frame to next frame. By repeatedly calculating the minimize distances between FPs, FPs are divided into BFPs and FFPs. Based on the foregoing, Chen and Hu et al. [19]–[23] further explored the FP classification method, then, they achieved accurate classification of BFPs and FFPs through clustering, homography transformation, and pole geometry. Ling and Zhao et al. [24]–[26] extended the classification of FPs to the classification of optical flow trajectories and their trajectory derivatives. At the same time, they proposed a feedback mechanism to further strengthen the judgment of foreground optical flows. It had a significant effect when the targets occupied a large area of the image. The above algorithms achieved good results, but more custom thresholds were introduced in the calculation process. In addition, Liu et al. [27]–[29] proposed a concept of pixel profile instead of optical flow trajectories for video stabilization, but such algorithms were also based on the FPs detection. The performance was poor when there were large foreground targets at close range. In terms of maritime image stabilization, Cao [[30] presented a dual image stabilization technology based on the combination of mechanical image stabilization and electronic image stabilization. Liu [31] developed a two-level electronic image stabilization algorithm based on vibration measurement and SSL detection. These algorithms used SSLs to estimate inter-frame image motion. However, they only considered the translation motion in the vertical direction and the rotation motion but fail to consider the translation motion in the horizontal direction.

2.2 Maritime target detection technology

Target detection experiences two historical periods in the development, namely traditional target detection (before 2014) and target detection based on deep learning (after 2014) [32]. Traditional target detection is mainly based on the characteristics of artificial design, and three milestone detectors are formed. In 2001, Viola and Jones [33] proposed the VJ detector, which achieved the real-time detection of human faces for the first time without any constraints. In 2006, Dalal and Triggs [34] proposed the histogram of oriented gradients (HOG) detector, which made significant improvements to the scale-invariant feature transform SIFT algorithm. In 2008, Felzenszwalb [35] proposed the deformable part based model (DPM), which was derived from the HOG detector. It was the peak of traditional target detection algorithms. Target detection based on deep learning mainly includes two types of detection models, namely the two-stage detection

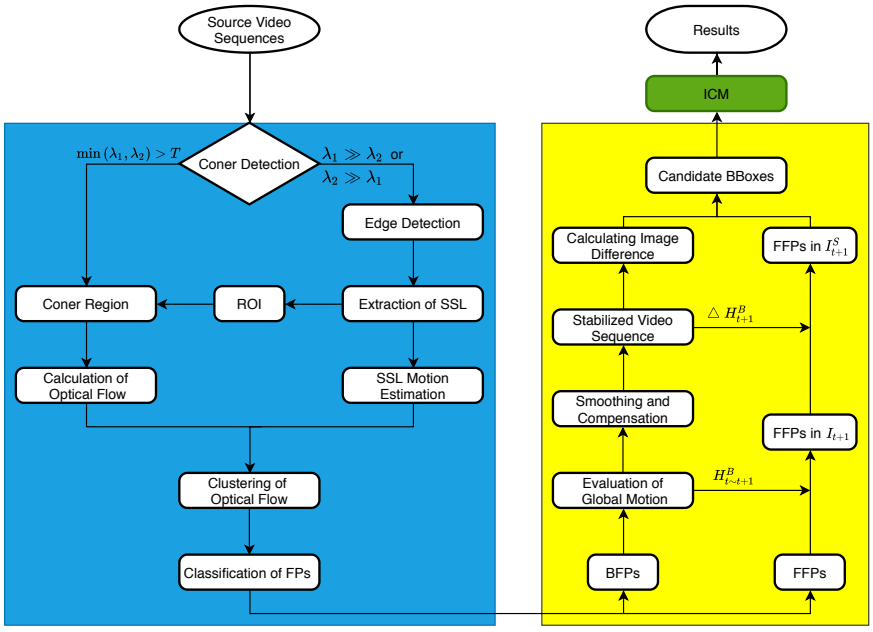


Fig. 1 Algorithm flowchart, where λ_1 and λ_2 represent the feature values of the corner detection, ΔH^B_{t+1} and $H^B_{t\sim t+1}$ represent the transformation matrices, and I_{t+1} and I^S_{t+1} respectively represent the t-th frame and the $(t+1)$-th stable frames.

(a) (b)

Fig. 2 FPs detection results with $N = 50$. (a) Corner detection in the entire image. (b) Corner detection in ROI.

model and the single-stage detection model. The former is represented by the region-based CNN (R-CNN) series [36]–[38], and the latter is represented by You Only Look Once (YOLO) [39] and the single shot multibox detector (SSD) [40]. Then, with the continuous development of target detection, the detection model achieves the balance between accuracy and speed gradually, through the new feature extraction network [41], accurate region proposal network (RPN) [42], complete classification method of ROI [43], sample post-processing technology [44] and mutual reference between various models [45]–[47].

3. Construction of the PLM

In the sea-sky image, the FPs are usually used for motion estimation because of their rich local information and large intensity changes in various directions. However, due to the complexity and variability of the maritime environment, it is difficult to be distinguished accurately between the BFPs and FFPs, which is not conducive to the realization of video stabilization. As SSLs are important background features in maritime video sequences, analyzing the SSLs' motion characteristics of two adjacent frames can perform vertical and rotative compensation. Nevertheless, it cannot achieve horizontal compensation. Thus, the PLM is constructed for accurate classification of BFPs and FFPs.

The PLM uses the corner detection algorithm proposed by Harris et al [48]. It uses the first-order partial derivatives to describe the intensity change of the pixel in any direction, i.e. the gray change caused by the image moving a small

displacement (u, v) of the image in any direction, as shown in Eq. (1) below, where $w(x, y)$ is a window function.

$$E(u,v) = \sum_{(x,y)} w(x,y)[I(x+u, y+v) - I(x,y)]^2 \quad (1)$$

In order to better evaluate the detection results, this study uses the Shi-Tomasi scoring function, as shown in Eq. (2) below, where λ_1, λ_2 are the eigenvalues.

$$R = \min(\lambda_1, \lambda_2) \quad (2)$$

- When λ_1 and λ_2 are both small, then $|R|$ is also small and this region is a flat region.
- When $\lambda_1 \gg \lambda_2$ or $\lambda_2 \gg \lambda_1$, then $R < 0$ and this region is an edge region.
- When both λ_1 and λ_2 are large, then R is also large, $\min(\lambda_1, \lambda_2)$ is larger than the threshold T, and this region is a corner region.

For the corner region, the threshold value T used to generate the corner points is a global threshold value, which will cause a large deviation in the region with a strong image texture. In the sea-sky image, the texture information of the sky region is weak, and it is difficult to generate stable corner points. The sea region is rich in texture information due to noise interference such as wave and sea clutter, which will generate a large number of corner points, resulting in a large deviation of the global threshold. The texture information of the region near the SSL is relatively rich, and the features of the target ships are obvious. So the region near the SSL is suitable for corner detection to obtain stable FPs as ROI.

For the edge region, the non-maximum suppression and lag thresholds are used to obtain accurate edges. Then, the longest straight line in the edge image is extracted by the Hough transformation to extract the SSL. Considering that SSL is usually a straight line that runs through the entire image and generally has a certain oblique angle, this study uses a rectangle to describe the region of the SSL. In order to reduce the error, N pixels are added both above and below the rectangular region as ROI. The FP detection results with $N = 50$ are shown in Figure 2.

The SSL is simple in description and has a stronger anti-noise capability than the FPs. It can estimate the global motion vector, which is of great significance to the stabilization of maritime videos. In order to facilitate the analysis of SSL motion characteristics, this study simplifies the image motion model of two consecutive frames into a rigid transformation model, i.e. only considering translation and rotation motion. The rigid transformation model is shown in Eq. (3) below:

$$E = \begin{pmatrix} cos\theta & -sin\theta & d_x \\ sin\theta & cos\theta & d_y \\ 0 & 0 & 1 \end{pmatrix} \quad (3)$$

where d_x and d_y represent translation motions along the x axis and y axis, respectively, and θ represents the rotation motion around the optical axis of the camera. This model neither distorts the original shape of the object nor changes the size of the object.

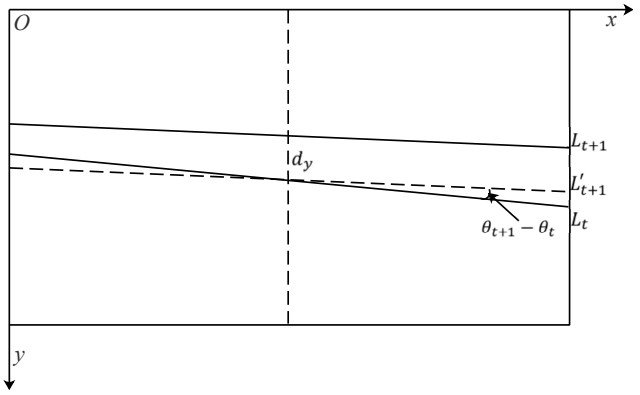


Fig. 3 SSL inter-frame motion diagram.

This study marks the SSLs of two consecutive frames in the same coordinate system, as shown in Figure 3. The lines L_t and L_{t+1} represent the SSLs at the t-th and $(t+1)$-th frames, respectively, and their respective angles are θ_t and θ_{t+1}. The parameter d_y represents the difference between the midpoints of the lines L_t and L_{t+1} in the y axis, and $(\theta_{t+1} - \theta_t)$ represents the angle difference between the lines L_{t+1} and L_t.

From Figure 3, it can be seen that there is a rotation and translation motion of SSL position relationship from the t frame to $(t+1)$ frame, i.e., L_t rotates $(\theta_{t+1} - \theta_t)$ and moves along the y axis by d_y to L_{t+1}. When calculating the motion vector between frames using SSLs, it is impossible to calculate the translation in the x axis direction (i.e. $d_x = 0$), and only d_y and θ can be calculated. Therefore, the rigid transformation model can be described as Eq. (4) below:

$$E = \begin{pmatrix} cos(\theta_{t+1} - \theta_t) & -sin(\theta_{t+1} - \theta_t) & 0 \\ sin(\theta_{t+1} - \theta_t) & cos(\theta_{t+1} - \theta_t) & d_y \\ 0 & 0 & 1 \end{pmatrix} \quad (4)$$

In ROI, the optical flow method is used to obtain the matching FPs in two consecutive frames. This study defines the FPs in the t-th frame by P_t and the matching FPs in the $(t+1)$-th frame by P_{t+1}. The optical flow vectors in two consecutive frames can be expressed as $V_{t\sim t+1}$, such as Eq. (5). This study defines $V_{t\sim t+1}$ with two elements of optical flow angle δ and length ρ, as shown in Eq. (6), where $(x_{t+1}, y_{t+1}) \in P_{t+1}$ and $(x_t, y_t) \in P_t$. The distribution of $V_{t\sim t+1}$ in the Cartesian coordinate system is shown in Figure 4(a), where the x axis represents δ and the y axis represents ρ. $V_{t\sim t+1}$ can be divided into background optical flows (defined as $V^B_{t\sim t+1}$) and foreground optical flows (defined as $V^F_{t\sim t+1}$). In the rigid model $V^B_{t\sim t+1}$ is affected by d_x, d_y, and θ, while $V^F_{t\sim t+1}$ is affected by its own motion in addition to the above factors. In this study, the rigid model E obtained by the SSL motion is used to transform the FPs of the t-th frame, and

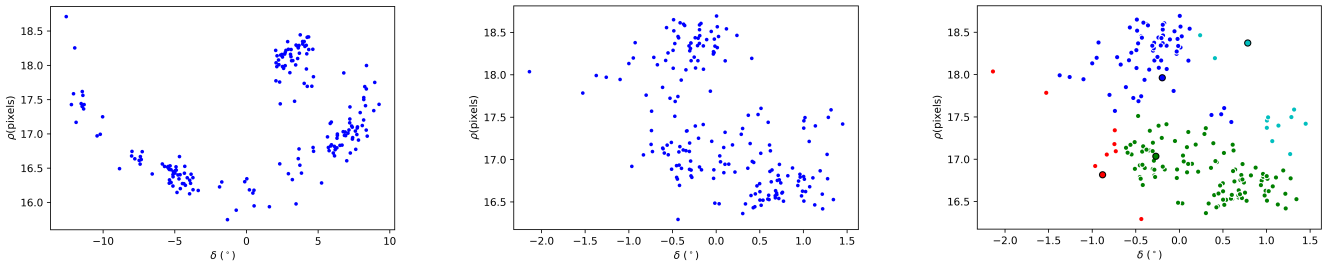


Fig. 4 Optical flow vectors distribution for two consecutive frames in the Cartesian coordinate system. (a) Initial optical flow vectors distribution. (b) Optical flow vectors distribution after SSL motion model transformation. (c) Distribution of clustering results in the Cartesian coordinate system.

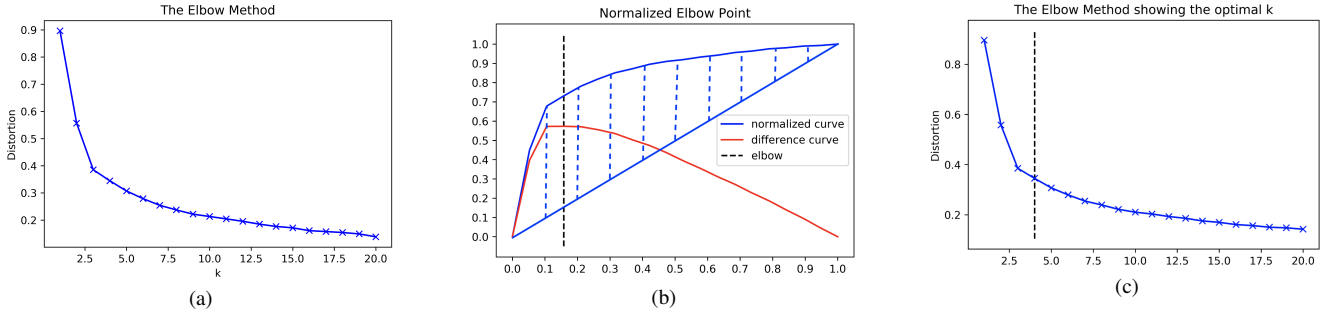


Fig. 5 Optimization of the elbow method. (a) The original elbow method. (b) Optimization algorithm. (c) Define the optimal clusters ($k = 4$)

Fig. 6 Distribution of FPs, $V^B_{t\sim t+1}$ and $V^F_{t\sim t+1}$ from two consecutive frames.

then the new optical flow vectors ($V^E_{t\sim t+1}$) are obtained by using Eq. (7). The distribution of $V^E_{t\sim t+1}$ in the Cartesian coordinate system is shown in Figure 4(b). By using this method, it can eliminate the effects of d_y and θ, the difference between $V^B_{t\sim t+1}$ and $V^F_{t\sim t+1}$ is more obvious, which is convenient for their further classification.

$$V_{t\sim t+1} = P_{t+1} - P_t \tag{5}$$

$$\begin{cases} V_{t\sim t+1} = (\delta, \rho) \\ \delta = tan^{-1}(y_{t+1} - y_t)/(x_{t+1} - x_t) \\ \rho = \sqrt{(x_{t+1} - x_t)^2 + (y_{t+1} - y_t)^2} \end{cases} \tag{6}$$

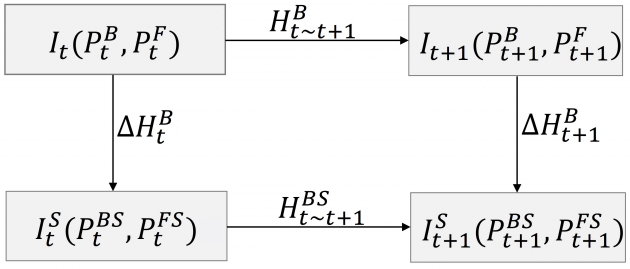
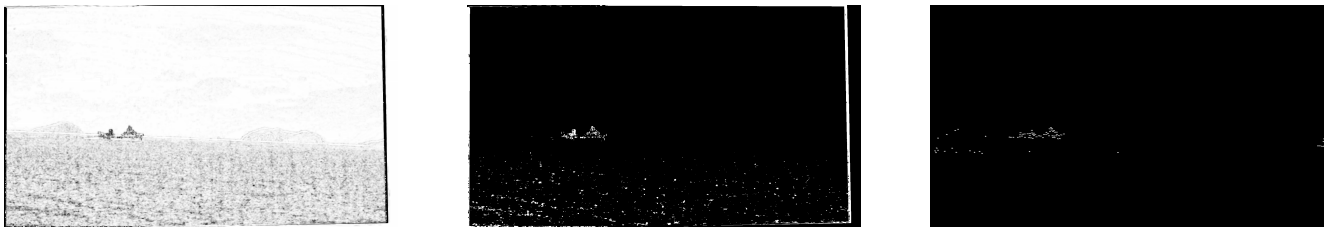


Fig. 7 FPs classification processing model diagram.

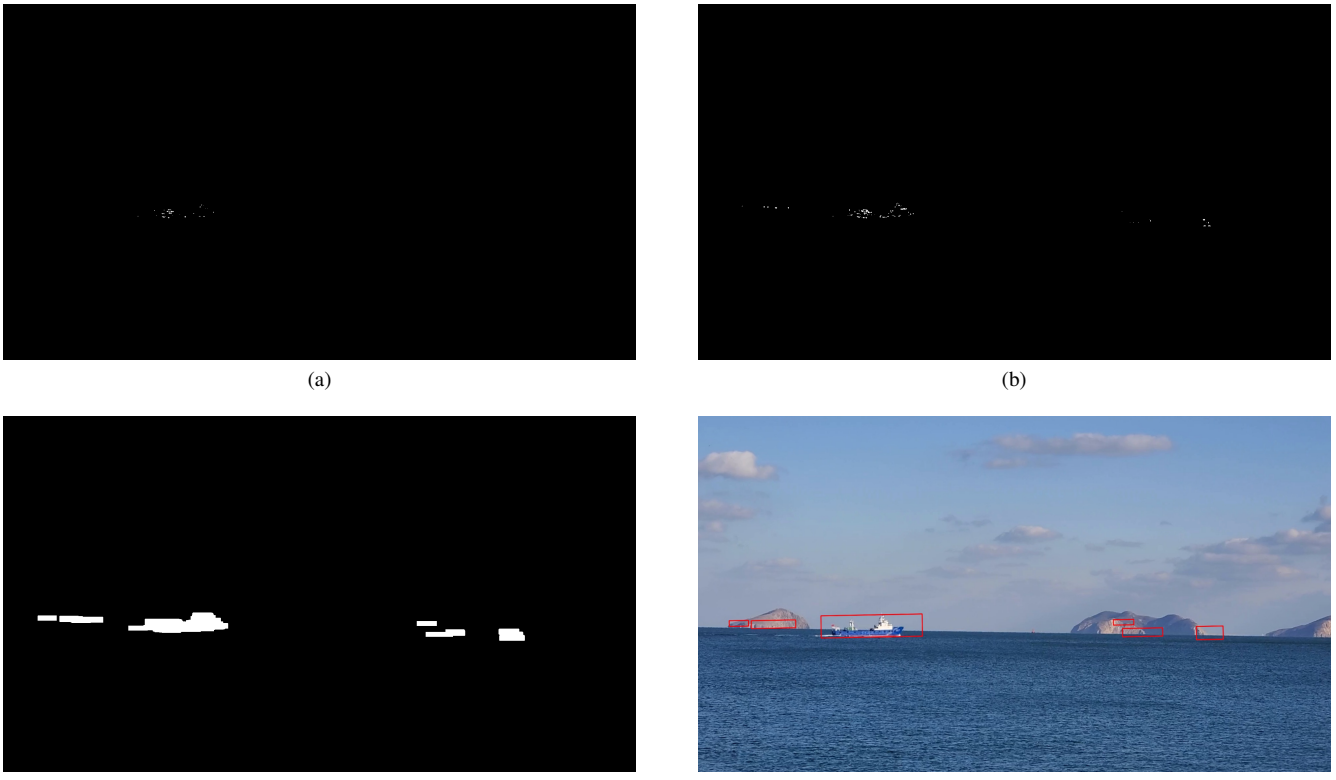
$$V^E_{t\sim t+1} = P_{t+1} - E * P_t \tag{7}$$

Then the K-means algorithm is used to cluster $V^E_{t\sim t+1}$, and the number of clusters is defined as k. The choice of the value of k directly determines the quality of the clustering results, and the elbow method is adopted, as shown in Figure 5(a). The value k is taken from 2 to 20, and the ordinate value represents the cost function of the clustering, which is expressed by the sum of the degree of distortion of the categories. In order to automatically obtain the k value corresponding to the elbow position, the optimization algorithm proposed by Ville et al. [49] is used to rotate and normalize the blue curve in Figure 5(a) to the position in Figure 5(b). This study connects the endpoints of the blue curve to make a straight line, find the difference between the ordinate values of the blue curve and the straight line. The difference is



(a) (b) (c)

Fig. 8 FPs classification processing results. (a) Difference image of two consecutive stable frames. (b) Binary image of the difference image ($\varepsilon = 8$). (c) FFPs processing results.



(a) (b)

(c) (d)

Fig. 9 Subsequent processing results. (a) Single frame processing results. (b) Three consecutive frames processing results. (c) Morphological processing results. (d) Target detection results.

represented by the red curve. So, this study can obtain the vertex abscissa values of the red curve and transform it to the initial coordinate system, as shown in Figure 5(c).

The clustering results obtained by the above method are shown in Figure 4(c). The k clusters are defined by $C = \{C_1, C_2, ..., C_k\}$, and the numbers of optical flows in the k respective clusters are defined by $n = \{n_1, n_2, ..., n_k\}$. According to the clustering results, $V^B_{t\sim t+1}$ and $V^F_{t\sim t+1}$ are classified according to the following principles:

For the number of optical flows n_i in any cluster C_i,

- if $n_i \leq \tau$, C_i is considered to be a noise cluster and needs to be deleted;
- if $n_i > \tau$ and $n_i = max(n)$, C_i is considered as the background cluster;
- the remaining clusters are considered as the foreground clusters.

According to the classification results of $V^E_{t\sim t+1}$, the FPs in the t-th frame and $(t+1)$-th frame are displayed in the t-th frame with red and green dots, respectively, as shown in Figure 6, where the blue lines represent $V^B_{t\sim t+1}$, and the yellow lines represent $V^F_{t\sim t+1}$.

4. Construction of the PCM

After classifying FPs by the PLM, the BFPs and FFPs are separately processed. The specific FP processing strategy is

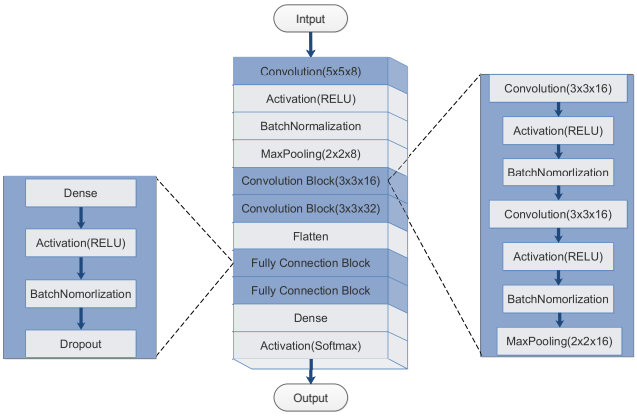


Fig. 10 The network structure diagram of the ICM.

shown in Figure 7, where I_t and I_{t+1} represent respectively the t-th frame and $(t+1)$-th frame of the original image sequences, P_t^B and P_{t+1}^B (also, P_t^F and P_{t+1}^F) represent the BFPs (also, FFPs) from two consecutive frames of the PLM, and $H_{t\sim t+1}^B$ represents the transformation matrix from I_t to I_{t+1}. Besides, I_t^S and I_{t+1}^S stand for the t-th frame and $(t+1)$-th frame after image stabilization, where P_t^{BS} and P_t^{FS} are the BFPs and FFPs, respectively, in I_t^S (also, P_{t+1}^{BS} and P_{t+1}^{FS} represent the BFPs and FFPs, respectively, in I_{t+1}^S). Also, $H_{t\sim t+1}^{BS}$ represents the transformation matrix from I_t^S to I_{t+1}^S, and ΔH_t^B (also, ΔH_{t+1}^B) refers to the transformation matrix from the original image sequence to the stable image sequence for the t-th frame (also, the $(t+1)$-th frame).

For the BFPs, this study calculates the motion trajectory of the original image sequence, obtains the transformation matrix and the stable image sequence through the mean filtering and motion compensation algorithms. In the stable image sequence, P_t^{BS} and P_{t+1}^{BS} can be obtained by Eq. (8) and (9), respectively. Then, this study uses the RANSAC method to calculate P_t^{BS} and P_{t+1}^{BS} for obtaining the transformation matrix $H_{t\sim t+1}^{BS}$. The difference image, which is shown in Figure 8(a), can be obtained by Eq. (10), where ε is the threshold. It is automatically obtained by the Otsu method. The inverse binary thresholding is used to process the difference image, as shown in Figure 8(b). For the FFPs, this study first transforms P_t^F from I_t to I_{t+1} by $H_{t\sim t+1}^B$ and then uses Eq. (11) to transform $(H_{t\sim t+1}^B * P_t^F)$ and P_{t+1}^F together from I_{t+1} to I_{t+1}^S, as shown in Figure 8(c).

$$P_t^{BS} = \Delta H_t^B * P_t^B \tag{8}$$

$$P_{t+1}^{BS} = \Delta H_{t+1}^B * P_{t+1}^B \tag{9}$$

$$I_{t\sim t+1}^{BS} = \begin{cases} 255, & if \quad |I_{t+1}^S - H_{t\sim t+1}^{BS} * I_t^S| > \varepsilon \\ 0, & otherwise \end{cases} \tag{10}$$

$$P_{t\sim t+1}^{BS} = \Delta H_{t+1}^B * (H_{t\sim t+1}^B * P_t^F + P_{t+1}^F) \tag{11}$$

In the subsequent processing stage, $I_{t\sim t+1}^{DS}$ and P_{t+1}^{BS} are fused by the "AND" operator to obtain the moving target contours. In order to obtain the quality contours, the results of three consecutive frames are fused and then the target contours are enhanced by morphological processing. Then the target positions are obtained using the smallest bounding rectangles. The processing results are shown in Figure 9.

5. Construction of the ICM

The above algorithm realizes rough detection of the positions of maritime targets. However, when using the PCM, some BFPs may be misclassified as FFPs. It results in the candidate bboxes containing ships and backgrounds, which causes false detection. Therefore, all the candidate bboxes are sent to the ICM for image classification, eliminating background bboxes and retaining ship bboxes. The network structure of the ICM is shown in Figure 10.

The ICM uses a convolutional (Conv) layer with a 5×5 filter to learn the larger features, such as the shape and the color of the ships, and then learns the detail features by two identical blocks. In each block, it contains two Conv layers, two activation layers, two batch normalization (BN) layers, and a pooling layer. In the first block, each Conv contains 16 filters and in the second block, each Conv contains 32 filters. The activation layer uses a linear activation function (RELU), and the pooling layer uses max pooling (MaxPool). Two fully connection (FC) blocks are followed, and each of them contains Flatten, Dense, RELU, BN and Dropout. In order to reduce the effect of overfitting, the Dropout are all set to 0.5.

6. Experimental Results and Discussion

This study mainly contains three datasets. The training set and the validation set were used to verify the performance of the ICM. In order to eliminate background interference, first, this study extracted every firth frame of each video from the Singapore Maritime Dataset (SMD) [50], then, used the YOLOv3 to pre-process the images and crops the bboxes. Through manual annotation, 12,000 training set images and 4000 verification set images were obtained. The type of ships includes ferry, vessel/ship, speed boat, boat, sail boat and others. The testing set includes testing set-1 and testing set-2, among which testing set-1 was used to verify the image stabilization algorithm, and testing set-2 was used to verify the performance of ship detection. Since the image stabilization algorithm is not driven by data, the testing set-1 only contained eight videos, where four videos came from onboard videos of the SMD taken by shipborne camera pitching and the others came from dynamic videos taken by shipborne camera rolling in the real ship. Testing set-2 contained seven videos on the basis of testing set-1, also from the onboard videos of the SMD. Each video was captured 200 consecutive frames. There were a total of 1600 frames in testing set-1, and a total of 3000 frames in testing set-2. The testing set-1 is shown in Figure 11.

In order to better analyze the effect of image stabi-

Fig. 11 The testing set-1, where V1, V2, V3 and V4 are from the SMD and V5, V6, V7 and V8 are from real ship shooting.

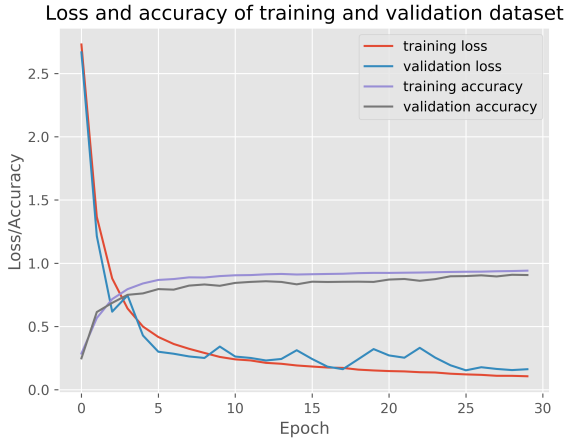


Fig. 12 Loss and accuracy on training and validation datasets.

lization, the common measures such as mean square error (MSE), peak signal to noise ratio (PSNR), and structural similarity index (SSIM) are introduced as the evaluation metrics of the algorithm. Besides, in order to accurately analyze the effect of target detection, the speed and mean average precision (mAP) are also set as the evaluation metrics of the algorithm.

6.1 Verification of the ICM

During training, Adma was selected as the optimizer, the epoch, batch size, learning rate and decay were set to 30, 64, 0.001 and 6.67×10^{-5}, respectively. To increase variability of the training set, the input images were augmented with random rotation, zoom, shift, shear, and flip settings. The loss function curve and accuracy curve obtained on the training set and validation set are shown in Figure 12. It can be seen that in the first 5 epochs, the loss function curve decreases sharply, and the accuracy curve increases rapidly. After 5 epochs, the training loss curve and the training accuracy curve are flattened, while validation loss curve and the validation accuracy curve have some fluctuations, mainly because in each epoch, the images used for verification have a certain randomness. After 30 epochs, the training loss value is reduced to 0.1063, and the training accuracy is 94.14%. On the validation set, the validation loss is 0.1621, validation accuracy is 91.67%. Good results were obtained on both the training and validation dataset, which verifies the effectiveness of the ICM.

6.2 Comparative Analysis of image Stabilization

In order to verify the image stabilization effects of proposed algorithm, testing set-1 was used for the comparison experiments. In the setting of benchmark algorithms, the popular YouTube stabilizer [13] and MeshFlow [21] were employed as Algorithm 1 and Algorithm 2, respectively. The results of the three algorithms can be obtained by calculating the average of the image stabilization metrics, as shown in Table 1.

In Table 1, the MSE, PSNR, and SSIM of Algorithm 1 after image stabilization were respectively reduced by approximately 45.61%–70.91%, increased by approximately 12.09%–24.56%, and increased by approximately 8.30%–19.65%. The MSE, PSNR, and SSIM of Algorithm 2 after image stabilization were respectively reduced by approximately 53.97%–79.41%, increased by approximately 18.50%–30.47%, and increased by approximately 7.27%–23.06%. After the proposed algorithm stabilized the video, the MSE was reduced by approximately 70.35%–90.93%, the PSNR was increased by approximately 21.01%–45.28%, and the SSIM was increased by approximately 13.00%–32.33%. Our proposed algorithm outperformed Algorithm 1 in the MSE, PSNR, and SSIM by −55.42%, 12.30%, 8.8%, respectively, while it outperformed Algorithm 2 in the MSE, PSNR, and SSIM by 47.87%, 8.66%, 6.94%, respectively. It can be seen that the image stabilization results of the proposed algorithm were better than those of the benchmark algorithms.

In order to visually display the image stabilization re-

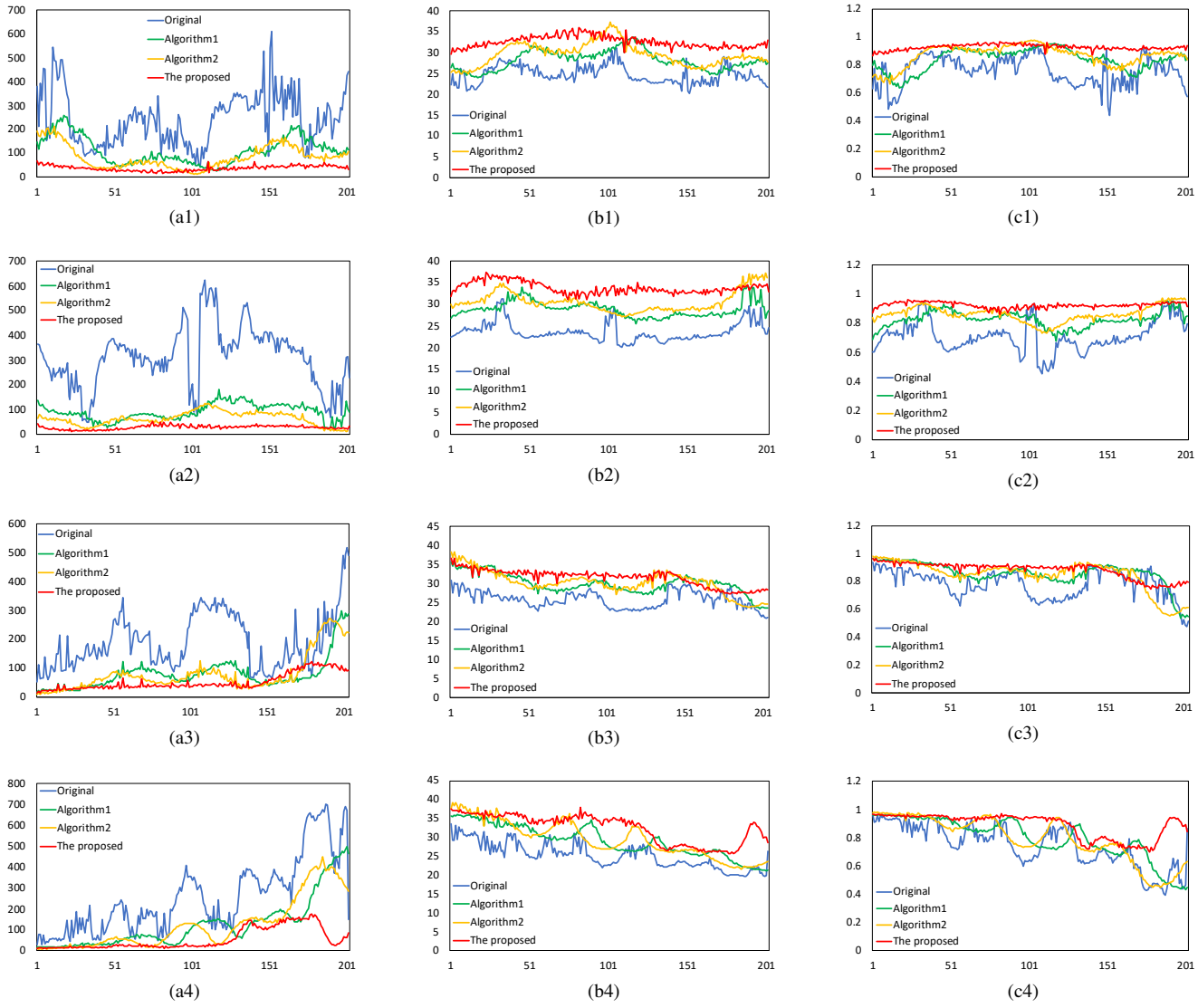
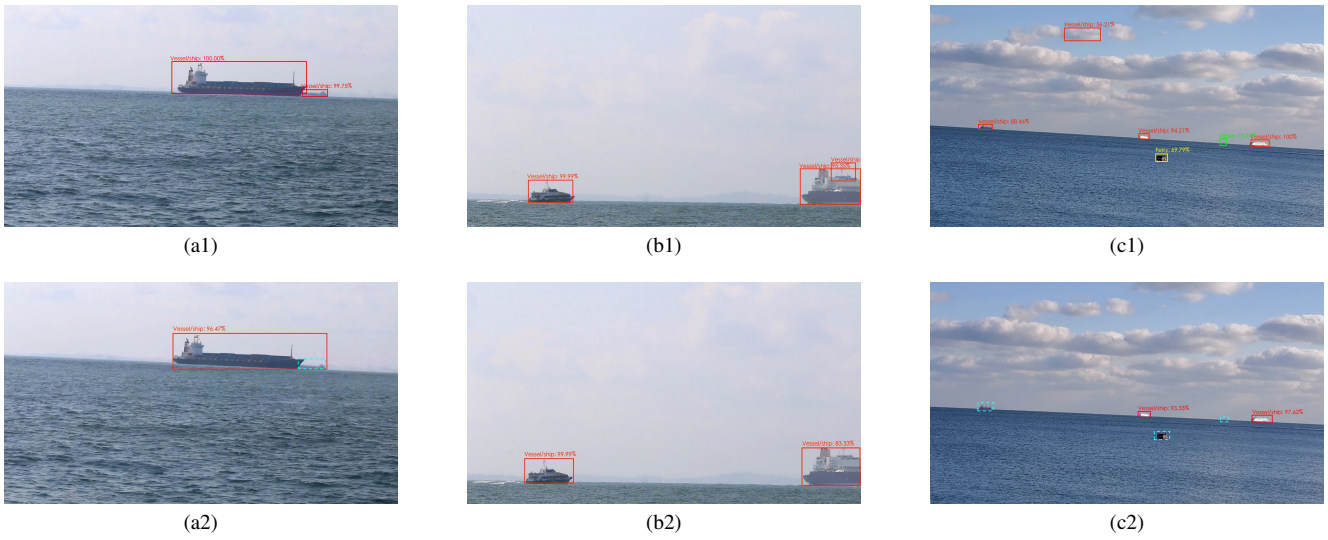


Fig. 13 Comparative analysis of video stabilization results. (a1) - (c1) Contrasts of MSE, PSNR and SSIM in V3. (a2) - (c2) Contrasts of MSE, PSNR and SSIM inV4. (a3) - (c3) Contrasts of MSE, PSNR and SSIM in V7. (a4) - (c4) Contrasts of MSE, PSNR and SSIM in V8.

Table 1 Average results of metrics with different algorithms on testing set-1.

Algorithms	Metrics	V1	V2	V3	V4	V5	V6	V7	V8	Average
Original	MSE	245.16	321.25	243.80	311.41	146.23	195.31	213.65	295.46	246.53
	PSNR(dB)	24.85	23.17	24.76	23.64	26.42	26.89	24.56	24.03	24.79
	SSIM(%)	74.23	73.14	75.75	70.93	78.97	74.15	71.24	69.35	73.47
Algorithm1	MSE	79.91	100.46	112.69	90.58	68.47	106.23	115.32	100.36	96.75
	PSNR(dB)	29.97	28.86	28.20	28.90	30.21	30.14	28.35	29.46	29.26
	SSIM(%)	86.71	79.21	83.70	82.63	87.31	82.45	85.24	81.22	83.56
Algorithm2	MSE	79.74	97.20	84.23	64.13	65.32	86.54	98.35	86.34	82.73
	PSNR(dB)	30.35	30.23	29.64	30.57	30.65	30.87	29.21	30.42	30.24
	SSIM(%)	84.85	78.46	86.87	86.40	87.45	86.56	84.32	85.34	85.03
The proposed	MSE)	51.43	68.45	37.50	28.25	43.36	34.14	49.24	32.67	43.13
	PSNR(dB)	31.60	32.31	32.60	33.82	31.97	33.65	32.01	34.91	32.86
	SSIM(%)	88.34	91.98	92.63	92.44	89.24	92.35	88.65	91.77	90.93

sults, V3, V4, V7 and V8 were taken as examples to draw the MSE, PSNR and SSIM curves after image stabilization, as shown in Figure 13. The blue curve represents the original image sequence. The green, yellow, and red curves represent



(a1) (b1) (c1)

(a2) (b2) (c2)

Fig. 14 Comparative analysis of some failed detection frames. (The blue dotted boxes represent the missed detection targets.) (a1) - (c1) Detection results of Algorithm3 in frame 0072 of V2, frame 0012 of V3 and frame 0007 of V6. (a2) - (c2) Detection results of the proposed algorithm in frame 0072 of V2, frame 0012 of V3 and frame 0007 of V6.

Table 2 The comparison results of speed and mAP on testing set-2.

Metrics	Algo1	Algo2	Algo3	The proposed
Speed(ms)	58.72	96.53	35.47	63.58
mAP(%)	72.21	81.34	84.49	90.03

the image stabilization results of Algorithm 1, Algorithm 2, and the proposed algorithm, respectively. It can be seen that all the algorithms significantly reduced the MSE, as well as increased the PSNR and SSIM of the images, compared with the original video. It indicates that the three image stabilization algorithms can all reduce the difference between two consecutive frames and achieve good effects of video stabilization, while the proposed algorithm is markedly superior to Algorithm 1 or Algorithm 2.

6.3 Comparative Analysis of Target Ship Detection

In order to analyze the performance of the proposed ship detection algorithm, this study used SSD-resnet50-fpn and faster-rcnn-resnet50 in Tensorflow object detection API as Algorithm 1 (Algo1) and Algorithm 2 (Algo2), respectively. Also, YOLOv3 was set as Algorithm 3 (Algo3). All of them were fine-tuned in the SMD, and run on the PC equipped with Intel® Xeon(R) CPU E5-2620 v4 and NVIDIA GeForce GTX 1080. The comparison results with the proposed algorithm are shown in Table 2.

From the comparison results of the four algorithms, the proposed algorithm generated the best performance in testing set-2, with the mAP value of 90.03%. However, in terms of running speed, the tuned YOLOv3 was the fastest. The average processing speed of the proposed algorithm was 64 ms, which is only slightly faster than Algo2. This is mainly because the proposed algorithm obtained fewer candidate bboxes, but the process only used the CPU calculations and failed to use the GPUs for acceleration. It is shown that the proposed algorithm can be applied on a wider platform without relying on GPU. The following are some examples to analyze the failed frame detection of the proposed algorithm and Algorithm3, as shown in Figure 14.

The comparison between Figures 14(a1) and 14(a2) illustrates that the proposed algorithm had a poor performance on partially occluded targets as it is easy to cause missed detection for occluded small targets, such as the blue dotted boxes in Figure 14(a2). The comparison between Figures 14(b1) and 14(b2) shows that the candidate bboxes obtained by the PLM of the proposed algorithm can produce the overall outline of the targets well and avoid false detection caused by local features. The comparison between Figures 14(c1) and 14(c2) implies that by the processing of the ROI the proposed algorithm can effectively suppress the interference of cloud layer and prevent false detection targets. However, for the static or non-obvious motion targets, such as the blue dotted boxes in Figure 14(c2), the PLM model mistakes them as the background of the image due to the obscurity of optical flow motion, resulting in the missed detection of targets.

7. Conclusions and Future Work

This study proposes a novel algorithm for ship detection based on electronic image stabilization technology. Firstly, the FPs are accurately classified into the BFPs and FFPs through the PLM. Then the BFPs are used for image stabilization and difference images, and the FFPs together with difference images are used to obtain the candidate bboxes through the PCM. Finally, the ICM is used to show the classification results and candidate bboxes in the stable image sequence, so as to realize the detection and recogni-

tion of the maritime targets. The experiment results have shown that the ICM constructed in this study has competitive overall performance, which can provide high accuracy and effectively reduce the false detection of maritime targets. Compared with the YouTube stabilizer and MeshFlow algorithms, the proposed algorithm has achieved superior image stabilization results. The proposed algorithm has also outperformed traditional maritime target detection algorithms, which provides a novel and promising research idea for maritime target detection. However, the proposed algorithm still has some drawbacks in the extraction of candidate bboxes, which might cause missed detection of targets. Therefore, one of the future research directions could be investigating a robust candidate bboxes extraction model for maritime target detection by combining various characteristics of the sea-sky image.

Acknowledgments

This research was financially supported by National Natural Science Foundation of China (Nos. 61772102 and 61906043) and the Fundamental Research Funds for the Central Universities (No. 3132019400).

References

[1] R. Lan, L. Sun, Z. B. Liu, H. M. Lu, C. Pang and X. N. Luo, MADNet: A Fast and Lightweight Network for Single-Image Super Resolution, IEEE Transactions on Cybernetics, pp. 1-11, 2020.

[2] H. M. Lu, Y. J. Li, M. Chen, H. Kim and S. Serikawa, Brain Intelligence: Go beyond Artificial Intelligence, Mobile Networks and Applications, vol. 23, pp. 368–375, 2018.

[3] W. P. Lu, Y. T. Zhang, S. J. Wang, H. Huang, Q. Liu and S. Luo, Concept Representation by Learning Explicit and Implicit Concept Couplings IEEE Intelligent Systems, pp. 1–1, 2020.

[4] Maritime Unmanned Navigation through Intelligence in Networks, Final Report Summary - MUNIN, EU Publications Office [Online], Available: https://cordis.europa.eu/project/id/314286/reporting, Accessed on: February 21, 2020.

[5] A.Garcia-Dominguez, Mobile applications, cloud and big data on ships and shore stations for increased safety on marine traffic, a smart ship project, In Proceedings of 2015 IEEE International Conference on Industrial Technology (ICIT), Seville, Spain, March 17-19 2015.

[6] J. Pandey, and K. Hasegawa, Autonomous navigation of catamaran surface vessel, In Proceedings of 2017 IEEE Underwater Technology (UT), Busan, South Korea, February 21-24, 2017.

[7] J. T. Qiu, A study of electronic image stabilization algorithms and visual tracking algorithm, PhD thesis, Xidian University, Xian, China, 2013.

[8] V. Eiselein, E. Bochinski and T. Sikora, Assessing post-detection filters for a generic pedestrian detector in a tracking-by-detection scheme, In Proceedings of 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, Italy, August 29-September 1, 2017.

[9] H. P. Yin, B. Chen, Y. Chai and Z. D. Liu, Vision-based Object Detection and Tracking: A Review, ACTA AUTOMATICA SINICA, vol. 42, no. 10, pp. 1466-1489, 2016.

[10] J. P. Hsiao, C. C. Hsu, T. C. Shih, P. L. Hsu, S. S. Yeh and B. C. Wang, The real-time video stabilization for the rescue robot, In Proceedings of the ICROS-SICE International Joint Conference 2009, Fukuoka, Japan, August 18-21, 2009.

[11] H. Shen, Q. Pan, Y. Cheng and Y. Yu, Fast video stabilization algorithm for UAV, In Proceedings of 2009 IEEE International Conference on Intelligent Computing and Intelligent Systems, Shanghai, China, November 20–22, 2009.

[12] K. X. Liu, J. Qian and R. K. Yang, Block matching algorithm based on RANSAC algorithm, In Proceedings of 2010 International Conference on Image Analysis and Signal Processing, Xiamen, China, April 9-11, 2010.

[13] K. Y. Lee, Y. Y. Chuang, B. Y. Chen and M. Ouhyoung, Video stabilization using robust feature trajectories, In Proceedings of 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, September 29-October 02, 2009.

[14] C. Wang, J. H. Kim, K. Y. Byun, J. Ni and S. J. Ko, Robust digital image stabilization using the Kalman filter, IEEE Transactions on Consumer Electronics, vol. 55, no. 1, pp.6-14, 2019.

[15] S. Battiato, G. Gallo, G. Puglisi and S. Scellato, SIFT features tracking for video stabilization, In Proceedings of the 14th International Conference on Image Analysis and Processing (ICIAP 2007), Modena, Italy, September 10-14, 2007.

[16] M. Grundmann, V. Kwatra and I. Essa, Auto-directed video stabilization with robust L1 optimal camera paths, In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, June 20-25, 2011.

[17] M. Nicolescu and G. Medioni, A voting-based computational framework for visual motion analysis and interpretation, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp. 739-758, 2005.

[18] S.K. Kim, S.J. Kang, T.S. Wang and S.J. Ko, Feature point classification based global motion estimation for video stabilization, IEEE Transactions on Consumer Electronics, vol. 59, no. 1, pp. 267-272, 2013.

[19] C.H. Chen, T.Y. Chen, W.C. Hu and M.Y. Peng, Video stabilization for fast moving camera based on feature point classification, In Proceedings of 2015 Third International Conference on Robot, Vision and Signal Processing (RVSP). Kaohsiung, Taiwan, November 18-20, 2015.

[20] W.C. Hu, C.H. Chen, Y.J. Su and T.H. Chang, Feature-based real-time video stabilization for vehicle video recorder system, Multimedia Tools and Applications, vol. 77, pp. 5107-5127, 2018.

[21] W.C. Hu, C.H. Chen, T.Y. Chen, M.Y. Peng and Y.J. Su, Real-time video stabilization for fast-moving vehicle cameras, Multimedia Tools and Applications, vol. 77, pp. 1237-1260, 2018.

[22] W.C. Hu, C.H. Chen, C.M. Chen and T.Y. Chen, Effective moving object detection from videos captured by a moving camera, Advances in Intelligent Systems and Computing, vol. 297, pp.343-353, 2014.

[23] W.C. Hu, C.H. Chen, T.Y. Chen, D.Y. Huang and Z.C. Wu, Moving object detection and tracking from video captured by moving camera, Journal of Visual Communication and Image Representation, vol. 30, pp. 164-180, 2015.

[24] Q. Ling, S. Deng, F. Li, Q. Huang and X. Li, A feedback-based robust video stabilization method for traffic videos, IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 3, pp. 561-572, 2018.

[25] M. Zhao, S. Deng and Q. Ling, A fast traffic video stabilization method based on trajectory derivatives, IEEE Access, vol. 7, pp. 13422-13432, 2019.

[26] M. Zhao and Q. Ling, A robust traffic video stabilization method assisted by foreground feature trajectories, IEEE Access, vol. 7, pp. 42921-42933, 2019.

[27] S. Liu, L. Yuan, P. Tan and J. Sun, SteadyFlow: spatially smooth optical flow for video stabilization, In Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA June 23-28, 2014.

[28] S.C. Liu, P. Tan, L. Yuan, J. Sun and B. Zeng, MeshFlow: minimum latency online video stabilization, Lecture Notes in Computer Science, vol. 9910, pp. 800-815, 2016.

[29] Z. Ren, J. Li, S. Liu and B. Zeng, Meshflow video denoising, In Proceedings of 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, September 17-20, 2017.

[30] H. Cao, Research on video-image stabilization and moving-object tracking in ship monitoring and control, M.S. thesis, Marine Engineering College, Dalian Maritime University, Dalian, China, 2008.
[31] W. Liu, Research on the method of electronic image stabilization for shipborne mobile video, M.S. thesis, Dalian Maritime University, Dalian, China, 2017.
[32] Z. X. Zou, Z. W. Shi, Y. H. Guo and J. P. Ye, Object Detection in 20 Years: A Survey, Computer Vision and Pattern Recognition (cs.CV), May 2019.
[33] P. Viola and M. J. Jones, Rapid object detection using a boosted cascade of simple features, Computer Vision and Pattern Recognition (CVPR), In Proceedings of the 2001 IEEE Computer Society Conference on, vol. 1, pp. I–I, 2001.
[34] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, Computer Vision and Pattern Recognition (CVPR), In Proceedings of the 2005 IEEE Computer Society Conference on, vol. 1, pp. 886–893, 2005.
[35] P. Felzenszwalb, D. McAllester, and D. Ramanan, A discriminatively trained, multiscale, deformable part model, Computer Vision and Pattern Recognition (CVPR), In Proceedings of the 2008 IEEE Conference on. pp. 1–8, 2008.
[36] R. Girshick, J. Donahue, T. Darrell and J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, June 23-28, 2014.
[37] R. Girshick, Fast R-CNN. In Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, December 7-13, 2015.
[38] S. Q. Ren, K. M. He, R. Girshick and J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, vol. 39, no. 6, pp. 1137-1149, 2017.
[39] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, June 27-30, 2016.
[40] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu and A. C. Berg, SSD: Single Shot MultiBox Detector. In Proceedings of 14th European Conference on Computer Vision (ECCV), Amsterdam, NETHERLANDS, October 8-16, 2016.
[41] Z. W. Cai, Q. F. Fan, R. S. Feris, and N. Vasconcelos, A Unified Multi-Scale Deep Convolutional Neural Network for Fast Object Detection. In Proceedings of 14th European Conference on Computer Vision (ECCV), Amsterdam, NETHERLANDS, October 8-16, 2016.
[42] T. Y. Lin, P. Dollar, R. Girshick, K. M. He, B. Hariharan and S. Belongie, Feature Pyramid Networks for Object Detection. In Proceedings of 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, July 21-26, 2017.
[43] K. M. He, G. Gkioxari, P. Dollar and R. Girshick, Mask R-CNN. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, vol. 42, no. 2, pp. 386-397, 2018.
[44] N.Bodla, B. Singh, R. Chellappa and L. S. Davis, Soft-NMS - Improving Object Detection with One Line of Code. In Proceedings of 16th IEEE International Conference on Computer Vision (ICCV), Venice, ITALY, October 22-29, 2017.
[45] J. Redmon, and A. Farhadi, YOLO9000: Better, Faster, Stronger. In Proceedings of 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, July 21-26, 2017.
[46] J. Redmon and A. Farhadi, YOLOv3: An Incremental Improvement. Computer Vision and Pattern Recognition (cs.CV), April 2018.
[47] A. Bochkovskiy, C. Y. Wang and H. Y. M. Liao, YOLOv4: Optimal Speed and Accuracy of Object Detection, Computer Vision and Pattern Recognition (cs.CV), April 2020.
[48] C. Harris and M. Stephens, A Combine Corner and Edge Detector, In Proceedings. of Fourth Alvey Vision Conference, pp.147–151, 1988.
[49] V. Satopaa, J. Albrecht, D. Irwin and B. Raghavan, Finding a "Kneedle" in a Haystack: detecting Knee points in system behavior, In Proceedings of 2011 31st International Conference on Distributed Computing Systems Workshops, Minneapolis, MN, USA, June 20-24, 2011.
[50] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabaly and C. Quek, Video Processing from Electro-optical Sensors for Object Detection and Tracking in Maritime Environment: A Survey. IEEE Transactions on Intelligent Transportation Systems (IEEE), vol 18, no. 8, pp.1993-2016, 2017.

Xiongfei Shan received the master's degree in traffic information engineering and control from the Dalian Maritime University, Dalian, China, in 2013. From 2015 to the present, he studied for a Ph.D. degree at Dalian Maritime University. His research interests include artificial intelligence and computer vision.

Mingyang Pan received the Ph.D. degree in traffic information engineering and control from the Dalian Maritime University, Dalian, China, in 2004. His research interests include artificial intelligence and computer vision. He is a professor with the navigation college, Dalian Maritime University. He is the director of the technical institute of navigation, Dalian Maritime University. His research activities are in the fields of electronic chart display and information system (ECDIS), digital waterway system and intelligent waterway transportation system.

Depeng Zhao is a professor and doctoral supervisor at Dalian Maritime University in China. He has published more than 30 papers in academic journals at home and abroad and has successively completed many scientific research projects, such as "electronic chart satellite navigation system", "electronic chart display and information system", "China Ship report system", etc. Among them, "Electronic chart and its application system" won the second prize of national science and technology progress (2007), "Research on maritime search and Rescue Decision Support System" won the two prize of Liaoning province for scientific and technological progress (2000), and "electronic chart display and information system" won the first prize of Dalian technology development (1994).

Deqiang Wang is a professor at Dalian Maritime University in China. He has published over 80 academic articles and patents. His recent research interests include the graph theory, machine learning and deep learning, big data, and intelligent transportation systems.

F.J. Hwang is the Senior Lecturer (Level C, Associate Professorship equivalency in the North American system), the Leading PI of the Industrial Optimization Group, and the Program Director (Maths/Stats) at the Faculty of Science; Transport Research Centre, University of Technology Sydney. F.J. has published in the leading journals, including Journal of Scheduling, Annals of Operations Research, Journal of the Operational Research Society, Computers & Operations Research, etc. He has been serving as the Guest Editor of several Q1/Q2 SCI journals as well as the General/Session Chair of several international conferences/workshops, and invited to give more than 30 research talks and conference presentations. His research interests center around data-driven optimization, supply chain and logistics optimization, and computational intelligence.

Chi-Hua Chen is currently a distinguished professor at Fuzhou University and a chair professor at Dalian Maritime University in China. He has published over 300 academic articles and patents. His contributions have been published in IEEE Internet of Things Journal, IEEE Access, IEICE Transactions, WWW'20, SIGIR 2020, etc. He serves as an associate editor for several international journals (e.g. IEICE Transactions on Information and Systems, IEEE Access, etc.). His recent research interests include the Internet of things, machine learning and deep learning, big data, ~~cellular~~ networks, and intelligent transportation systems.