A Hybrid Method for Inland Ship Recognition Using Marine Radar and Closed-Circuit Television

Authors: Xinglong Liu, Yicheng Li, Yong Wu, Zhiyuan Wang, Wei He, Zhixiong Li, Yicheng Li

Date: 2024-03-28T00:00:00+00:00

Abstract

Vessel recognition plays important role in ensuring navigation safety. However, existing methods are mainly based on a single sensor, such as automatic identification system (AIS), marine radar, closed-circuit television (CCTV), etc. To this end, this paper proposes a coarse-to-fine recognition method by fusing CCTV and marine radar, called multi-scale matching vessel recognition (MSM-VR). This method first proposes a novel calibration method that does not use any additional calibration target. The calibration is transformed to solve an N point registration model. Furthermore, marine radar image is used for coarse detection. A region of interest (ROI) area is computed for coarse detection results. Lastly, we design a novel convolutional neural network (CNN) called VesNet and transform the recognition into feature extraction. The VesNet is used to extract the vessel features. As a result, the MVM-VR method has been validated by using actual datasets collected along different waterways such as Nanjing waterway and Wuhan waterway, China, covering different times and weather conditions. Experimental results show that the MSM-VR method can adapt to different times, different weather conditions, and different waterways with good detection stability. The recognition accuracy is no less than 96%. Compared to other methods, the proposed method has high accuracy and great robustness.

Full Text

Preamble

A Hybrid Method for Inland Ship Recognition Using Marine Radar and Closed-Circuit Television

Xinglong Liu[1,2,3], Yicheng Li[4,*], Yong Wu[5], Zhiyuan Wang[1,2,3], Wei He[1,2,3] and Zhixiong Li[6]

[1] College of Physics and Electronic Information Engineering, Minjiang University, Fuzhou 350108, China
[2] Fujian Engineering Research Center of Safety Control for Ship Intelligent Navigation, Minjiang University, Fuzhou 350108, China
[3] Engineering Research Center of Fujian University for Marine Intelligent Ship Equipment, Minjiang University, Fuzhou 350108, China
[4] The Hubei Key Laboratory of Transportation Internet of Things, Wuhan University of Technology, Wuhan 430063, China
[5] National Engineering Research Center for Water Transport Safety, Wuhan University of Technology, Wuhan 430063, China
[6] Faculty of Mechanical Engineering, Opole University of Technology, 45758 Opole, Poland

- Correspondence: liyicheng070@ujs.edu.cn; Tel.: +86-0591-83761627

Abstract

Vessel recognition plays a critical role in ensuring navigation safety. However, existing methods predominantly rely on single sensors such as the Automatic Identification System (AIS), marine radar, or closed-circuit television (CCTV). To address this limitation, this paper proposes a coarse-to-fine recognition framework that fuses CCTV and marine radar data, termed Multi-Scale Matching Vessel Recognition (MSM-VR). The method introduces a novel calibration technique that requires no additional calibration targets, transforming the calibration problem into solving an N-point registration model. Furthermore, marine radar images are employed for coarse detection, and a Region of Interest (ROI) area is computed based on these coarse detection results. Finally, we design a novel Convolutional Neural Network (CNN) called VesNet that transforms recognition into a feature extraction task. The VesNet extracts vessel features, and the complete MVM-VR method is validated using actual datasets collected along different waterways, including the Nanjing and Wuhan waterways in China, under varying temporal and weather conditions. Experimental results demonstrate that the MSM-VR method adapts effectively to different times, weather conditions, and waterways while maintaining stable detection performance. The recognition accuracy achieves no less than 96%, and compared to alternative methods, the proposed approach exhibits higher accuracy and greater robustness.

Keywords: marine vessel recognition; multi-scale matching; convolutional neural network (CNN); radar-camera calibration; vessel net (VesNet)

1. Introduction

In recent years, the economic significance of inland waterway navigation has become increasingly prominent alongside China' s rapid economic development [1]. With growing vessel traffic, navigation safety has emerged as a critical concern, focusing on timely monitoring and accurate recognition of vessels. Various systems can be employed for this purpose, including the Automatic Identification System (AIS), marine radar, and closed-circuit television (CCTV). AIS serves as a navigation aid system for maritime safety and vessel-to-shore communication [2], enabling automatic exchange of critical information such as position, speed, course, vessel name, and call sign. This facilitates effective traffic information acquisition without radar detection and consequently reduces vessel collision accidents. However, many vessels operate without AIS in certain areas, particularly in China, despite mandatory equipment requirements. This necessitates the development of a hybrid method for inland ship recognition that integrates marine radar and CCTV.

Marine radar technology has advanced rapidly for vessel detection, with high-sensitivity radar systems providing informative imagery of surroundings. Modern S-band marine radar can track a 0.5 m^2 object at distances up to 5 nautical miles. While maritime radar excels at long-range detection, it cannot identify vessel types. Conversely, CCTV provides high-quality images revealing vessel details [3,4]. Fusing these two modalities and leveraging their complementary advantages can improve both detection speed and recognition accuracy, as illustrated in Figure 1 [Figure 1: see original paper]. The overarching objective of this work is to fuse these two systems for fast and accurate vessel recognition. The nomenclature used throughout this paper is summarized in Table 1 .

---

2. Related Work and Contributions

Vessel recognition comprises two stages: vessel detection and vessel classification. Numerous marine radar-based methods have been proposed for these tasks. For instance, a pre-processing approach was developed to estimate the length of small, slow-moving marine targets for forward-scatter maritime radar. Unfortunately, most marine radar systems operate in low pulse repetition frequency mode, resulting in ambiguous Doppler signals and velocity measurements. Ma et al. proposed a generalized Bayesian inference method called the evidential reasoning (ER) rule for vessel recognition using radar images. This three-step method first computes likelihoods for velocity and direction from radar images, transforms these likelihoods into multiple evidence sources, and then trains weight coefficients using a nonlinear optimization model, with vessels recognized based on these coefficients [5]. A recent Bayesian network-based methodology extracted vessels from radar images [6], primarily applying inter-frame differences to account for vessel velocity, direction, and shape, while establishing a directed acyclic graph for recognition. This approach extracts targets from original radar images and verified records without imposing unrealistic assumptions about object states.

Recent studies have explored laser-based object recognition for vessel detection, as lasers provide more accurate distance information. For example, Misović et al. proposed a lock gate zone concept based on a laser monitoring system developed under modular principles with functional partitions based on pattern recognition [7]. However, laser detection range is relatively short (approximately 80–150 m), limiting applicability to special environments such as ports and wharves. Given that inland waterways often exceed 1 km in width, laser-based methods are unsuitable for such applications.

Cameras represent popular, cost-effective sensors for object recognition that provide rich target details. Typical maritime applications use CCTV [8], where image-based recognition focuses on feature extraction, converting images into multi-dimensional vectors to reduce data storage requirements. Scale-Invariant Feature Transform (SIFT), proposed in 2004, remains a popular descriptor known for extraction stability. Various recognition approaches employ SIFT, such as a bag-of-raw-features model for object recognition [9], despite its disadvantages of low detection efficiency and computational complexity. To address these limitations, Speeded-Up Robust Features (SURF) was proposed to simplify computational complexity while achieving results similar to SIFT [10].

With deep learning gaining popularity, Convolutional Neural Networks (CNN) have been widely utilized for feature extraction. For example, a novel CNN architecture was proposed for multi-label problems, demonstrating superior classification performance compared to alternatives [11]. An unsupervised learning method presented a semantic encoder [12], while another approach merged fragment pair descriptors during scene segmentation by computing distances between descriptor pairs [13]. However, due to the relatively short history of deep learning in this domain, few public studies address vessel recognition specifically. Solmaz et al. proposed a deep learning-based multi-task learning framework divided into five tasks (visual recognition, coarse-grained classification, fine-grained classification, coarse-grained retrieval, fine-grained retrieval, and verification), improving recognition performance through this framework [14]. Similarly, Voinov et al. proposed a CNN-based vessel recognition method with potential for near-real-time applications using optical satellite sensors primarily for marine applications [15]. Although these studies applied deep learning to vessel recognition, no network has been specifically designed for this purpose to date. Consequently, this paper proposes a neural network model tailored for inland vessels.

Current methods predominantly rely on single sensors (cameras, maritime radar, or lasers). Integrating these sensors to leverage their distinct advantages could improve recognition accuracy. Calibration represents the first step in data fusion, with most existing methods focusing on cameras or laser scanners. For instance, Zhang and Pless proposed a calibration method using a chessboard, which has become perhaps the most popular approach for laser-camera calibration, requiring at least five chessboard poses [16]. Subsequent work reduced the required scans from five to three [17,18]. Naikal et al. established a Perspective-n-Point (PnP) model to manually align images with LIDAR data [19], though invisible and large-error calibration results make correspondence difficult. A calibration method for CCTV and maritime radar based on checkerboard calibration has been proposed, but such methods are unsuitable for maritime sensors. Recent target-less calibration approaches utilize LRF reflectivity as a constraint to detect different objects, while edge-based calibration matches LIDAR and image edges to compute sensor relationships [20], estimating the degree of edge orientation alignment between LIDAR and images for LRF-camera calibration.

This paper proposes a hybrid method for inland ship recognition using marine radar and CCTV. The contributions are summarized as follows:

1. A novel coarse-to-fine recognition framework is proposed that fuses image and radar data for inland vessel recognition, representing the first such approach to our knowledge. Adding a detection step for coarse recognition before vessel classification narrows the search area and improves recognition accuracy.
2. A novel calibration method for CCTV and marine radar is presented that requires no chessboard or maritime-unrelated objects. By utilizing inland vessels as calibration targets, the calibration problem is transformed into N-point registration, enabling computation of calibration parameters between CCTV and marine radar.
3. Deep learning theory is incorporated into a novel network called VesNet for vessel recognition. Unlike other CNNs, VesNet is specifically designed for extracting vessel features. This eliminates the need to search for vessel datasets for training, as the recognition process employs traditional feature matching methods similar to SIFT, SURF, or ORB, reducing CNN training time. The proposed recognition method thus combines CNN with traditional feature matching.

The remainder of this paper is organized as follows: Section 3 details the MSM-VR method for vessel recognition, Section 4 presents experimental results, and Section 5 provides conclusions.

---

3. Inland Ship Recognition Framework Using CCTV and Marine Radar

A coarse-to-fine recognition method called Multi-Scale Matching Vessel Recognition (MSM-VR) is proposed based on the fusion of CCTV and marine radar. The multi-scale concept employs radar scale for detection and camera scale for recognition. Mathematical notation is summarized in Table 2 . The method comprises three steps: (1) a calibration method establishing the foundation for CCTV and maritime radar data fusion; (2) marine radar utilization for coarse detection with ROI area computation; and (3) a CNN-based vessel recognition network using CCTV images to detect vessels and determine their type and position.

The MSM-VR method is schematically illustrated in Figure 2 [Figure 2: see original paper].

3.1. Sensors Calibration

Calibration of CCTV and radar, as the first step in vessel recognition, primarily computes the relationship between radar and CCTV coordinate systems. Two three-dimensional coordinate systems are established (radar coordinate system, RCS, and CCTV coordinate system, CCS) to transform the calibration problem into solving two parameters: rotation matrix and translation vector, using the following equation:

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_R \\ Y_R \\ Z_R \\ 1 \end{bmatrix}$$

where $[X_C, Y_C, Z_C, 1]^T$ represents a point in CCS, R denotes the rotation matrix (size: 3×3), t denotes the column translation vector (size: 3×1), and $[X_R, Y_R, Z_R, 1]^T$ represents the same point in RCS.

Equation (1) indicates that calibration critically depends on finding corresponding points in RCS and CCS. Unlike other calibration methods, this approach utilizes vessels themselves, computing their centroid points as corresponding points. This ensures practicability without requiring additional calibration targets.

First, at least three vessels are manually selected from CCTV images, with corresponding vessels sequentially identified in radar data. Calibration is then divided into two steps: rotation and translation. RCS and CCS points are normalized for rotation using:

$$d = \frac{1}{\sqrt{X_R^2 + Y_R^2 + Z_R^2}} \begin{bmatrix} X_R \\ Y_R \\ Z_R \end{bmatrix}, \quad b = \frac{1}{\sqrt{X_C^2 + Y_C^2 + Z_C^2}} \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix}$$

Thus, RCS and CCS vessel points can be rewritten as n_1, n_2, n_3 and m_1, m_2, m_3, respectively. An N-point registration model is established to compute R:

$$[m_1 \quad m_2 \quad m_3 \quad ...] = R \cdot [n_1 \quad n_2 \quad n_3 \quad ...]$$

Consequently, R is computed by:

$$R = \begin{bmatrix} m_1 & m_2 & m_3 & ... \end{bmatrix} \cdot \begin{bmatrix} n_1 & n_2 & n_3 & ... \end{bmatrix}^{-1}$$

Notably, the rank of $\begin{bmatrix} n_1 & n_2 & n_3 & ... \end{bmatrix}$ must be at least three. The translation vector t is solved as:

$$t = \begin{bmatrix} m_1 & m_2 & m_3 & ... \end{bmatrix} \cdot \begin{bmatrix} n_1 & n_2 & n_3 & ... \end{bmatrix}^{-1} \cdot \begin{bmatrix} d_1 - b_1 \\ d_2 - b_2 \\ d_3 - b_3 \end{bmatrix}$$

3.2. Coarse Detection

In the second step of coarse detection, radar images determine vessel presence in the waterway. Radar enables faster and more efficient vessel detection than CCTV due to its wider perspective. A frame difference method detects vessels using:

$$D_k(x, y) = |f_k(x, y) - f_{k-1}(x, y)|$$

where $f_k(x, y)$ represents the k-th frame image and $f_{k-1}(x, y)$ represents the $(k-1)$-th frame image, as shown in Figure 3 [Figure 3: see original paper]. The difference image is binarized as:

$$B_k(x, y) = \begin{cases} 0 \quad \text{Background}, \quad D_k(x, y) \leq \text{Threshold} \\ 1 \quad \text{Vessel Candidate}, \quad D_k(x, y) > \text{Threshold} \end{cases}$$

A threshold compares with difference image pixels during binarization. Pixels below the threshold are classified as background; otherwise, they are considered vessel candidates.

The binarized image undergoes further processing through image filtering and morphological operations to denoise targets and smooth boundaries. Another threshold is then applied based on candidate area comparison. Candidates with areas exceeding the threshold are confirmed as vessels; smaller regions are deemed false positives and removed. Results are shown in Figure 4 [Figure 4: see original paper]. Finally, detected vessels are mapped into CCTV images, with mass centers computed using an ROI radius of 50 m.

3.3. Fine Recognition

The ROI obtained from coarse detection enables deep learning to enhance vessel recognition accuracy. Recognition is divided into three steps: VesNet establishment, vessel feature extraction, and VesNet-based recognition.

(1) Establishment of VesNet: A deep learning network called VesNet is developed for vessel recognition. The AlexNet architecture, a typical CNN that won the 2012 ImageNet competition [21], serves as the foundation. AlexNet introduced novel techniques such as Local Response Normalization (LRN) and employs an eight-layer structure: five convolutional layers followed by three fully connected layers, as shown in Figure 5 [Figure 5: see original paper]. However, network depth can attenuate vessel detection performance in AlexNet. Therefore, considering vessel characteristics, the residual concept [22] is incorporated to create VesNet based on AlexNet.

Figure 6 [Figure 6: see original paper] illustrates the residual learning module added to VesNet, which builds deep networks through shallow network application and self-mapping. This ensures that training error does not exceed that of shallow networks as depth increases. The computation is:

$$q = F(p, \{W_i\}) + p$$

where p and q represent module input and output, respectively, and $F(p, \{W_i\})$ represents residual mapping. Note that p dimensions must be consistent with $F(p, \{W_i\})$. Otherwise, a linear projection is added to the shortcut connection:

$$q = F(p, \{W_i\}) + W_i \cdot p$$

The VesNet neural network model for vessel recognition follows two design rules: (1) network layers with identical output feature map sizes maintain the same number of filters and channels; (2) when output pattern size halves (via pooling), filter numbers double. Consequently, a 34-layer VesNet is designed, as structurally shown in Figure 7 [Figure 7: see original paper], where "3 conv, 64" denotes a convolutional layer with filter size 3 extracting vessel features (edges, corners, arcs) and 64 feature map channels; "avgpool" denotes average pooling that compresses vessel features, simplifying computational complexity while extracting principal features; and "fc" denotes fully connected layers that map upper-layer features via convolution to the sample label space defined during annotation. Specifically, input consists of vessel features from upper layers, while output is a 1000-dimensional vector where each element represents vessel type probability.

(2) Vessel Feature Extraction: The MSM-VR method realizes vessel recognition by establishing VesNet to extract features for all vessel types. According to Changjiang Maritime Safety Administration regulations, inland vessels are categorized into three types: container, cargo, and dangerous vessels [23,24]. However, extensive experiments reveal a unique vessel type—empty vessels—that warrants separate classification based on distinctive characteristics and waterway management requirements, as shown in Figure 8 [Figure 8: see original paper].

Figure 9 [Figure 9: see original paper] shows features extracted by VesNet for these four vessel types, each represented as a 1$\times$1000 dimensional vector. The distinct features enable differentiation between vessel types.

(3) VesNet-Based Recognition: After extracting features for four vessel types, each type’ s features are trained to establish a database offline. When marine radar detects a vessel, its features are extracted online and matched against the database. Since extracted features are row vectors (size: 1$×$1000), cosine similarity between query features and database features is computed:

$$\mathrm{Co}(F_p, \mathrm{model}_{k,l}) = \frac{F_p \cdot \mathrm{model}}{\|F_p\| \times \|\mathrm{model}\|} = \frac{\sum_{j=1}^{1000} \omega_j \tau_j}{\sqrt{\sum_{j=1}^{1000} \omega_j^2} \times \sqrt{\sum_{j=1}^{1000} \tau_j^2}}$$

where $F_p = (\omega_1, \omega_2, \dots, \omega_{1000})$ and model $= (\tau_1, \tau_2, \dots, \tau_{1000})$. The maximum similarity is 1, with higher values indicating greater feature similarity. The vessel type with peak similarity across four databases is selected as the query vessel’ s classification.

3.4. MSM-VR Algorithm

The MSM-VR algorithm is summarized as follows: (1) Select up to three vessels in both marine radar and CCTV, compute their mass points in each sensor, and calibrate sensors using the N-point registration model (Equations (1)-(10)). (2) Apply frame difference to radar data for coarse detection (Section 3.2, Figures 3 and 4), determine ROI areas, and map results to CCTV using calibration parameters. (3) Establish VesNet to extract features for four vessel types, training them to create a database (Figure 9) while extracting online features from ROI areas (Figure 7). (4) Match features against the database, compute cosine similarities, select the maximum, and determine vessel type (Equation (15)).

---

4. Experimental Results

Experiments using actual data evaluate the proposed method. The experimental site is the Yangtze River, China’ s longest river and the world’ s third-longest, serving as the golden waterway connecting eastern and western China. With the world’ s largest cargo volume, it holds strategic significance for China’ s economic development and social progress. Wuhan and Nanjing, located in the middle and downstream reaches, respectively, were selected to comprehensively evaluate MSM-VR under different navigable environments.

4.1. Case 1: Experiment in Nanjing

As Jiangsu Province’ s capital and a central city in the Yangtze River’ s downstream region, Nanjing hosted experimental setup at Banqiao ferry with co-located CCTV and marine radar. Sensors were calibrated initially for coarse-to-fine recognition using MSM-VR. Experiments spanned over one week (8:00-17:00 daily), detecting more than 400 vessels under varying weather conditions (sunny and rainy days).

Figure 10 [Figure 10: see original paper] shows recognition results in Nanjing, where blue, red, and yellow frames represent cargo, empty, and container vessels, respectively. Notably, no dangerous vessels appeared during experiments. Vessels are recognized upon entering CCTV surveillance areas, and the method successfully recognizes multiple simultaneous vessels.

Comparative evaluation against methods from References [3,5] using 400 test images shows the MSM-VR method recognized 387 images, achieving 96.75% accuracy with approximately 0.50% false positives, as multiple vessels occasionally occlude each other. Figure 11 [Figure 11: see original paper] illustrates an error case where a small vessel behind an empty vessel remained unrecognized. False negatives approximate 2.75% due to small vessels being too distant from the camera or having colors similar to the background. Table 3 indicates MSM-VR outperforms comparison methods achieving approximately 90.50% and 91.25% accuracy, respectively.

Performance under varying conditions is summarized in Table 4 . For two time periods (8:00–12:00 and 12:00–17:00) and two weather conditions (rainy and sunny days), 200 images were selected per period and sunny day, with 100 images per rainy day due to lower occurrence probability. Accuracies reach 96.00% and 97.50% across time periods, and 96.00% and 97.00% for rainy and sunny conditions, respectively. These results demonstrate MSM-VR' s improved accuracy and robustness, particularly on rainy days.

Precision-recall curves evaluate vessel type recognition performance. Figure 12 [Figure 12: see original paper] shows no dangerous vessels in the Nanjing test area, with recalls reaching 59%, 56%, and 48.5% for cargo, empty, and container vessels, respectively, at peak precision.

4.2. Case 2: Experiments in Wuhan

As Hubei Province' s capital and a central city in the Yangtze River' s middle stream, Wuhan hosted experimental setup at the Wuhan Yangtze River Bridge for approximately one month. Fewer vessels appeared in the middle stream test area compared to downstream (approximately 800 vessels total). Testing covered 8:00–17:00 daily, including sunny and rainy days.

Comparative results in Table 5 show MSM-VR processing 800 images with 97.63% accuracy—higher than Nanjing due to fewer simultaneous vessels. False positives dropped to zero, and MSM-VR outperformed comparison methods achieving 91.88% and 90.38% accuracy, respectively.

Table 6 presents results for two weather conditions and time periods. Four hundred images were selected per condition except rainy days (300 images). Accuracies reach 98.00% and 97.25% across time periods, with 96.33% accuracy on rainy days—slightly lower than sunny days but demonstrating robustness.

Precision-recall curves in Figure 13 [Figure 13: see original paper] confirm dangerous vessels in Wuhan Waterway, with recalls of 68%, 63%, 48%, and 48.5%

for dangerous, cargo, container, and empty vessels, respectively, at 100% precision. These stable results across test sites validate method robustness.

Table 7 summarizes combined results from Wuhan and Nanjing experiments.

---

5. Conclusions

This study fuses CCTV and marine radar, presenting a novel calibration method (MSM-VR) that establishes an N-point registration model using corresponding vessels as calibration targets, eliminating the need for additional objects. Marine radar performs coarse detection via frame differences, outputting ROI areas mapped to CCTV. A novel CNN called VesNet is designed for vessel recognition, transforming the task into accurate and efficient feature extraction through an added residual learning module. These components form the complete coarse-to-fine vessel recognition framework.

Experimental results demonstrate robust performance across different test sites and conditions, with recognition accuracies exceeding 96% except on rainy days where accuracy decreases moderately. The method achieves high accuracy with strong robustness, applicable to inland river waterways, particularly the Yangtze River.

Limitations include suitability only for narrow rivers due to camera detection range constraints. Future work will focus on rainy condition detection, integration of sensor networks for small boat detection, and fusion of AIS data to enhance recognition precision.

Author Contributions: Conceptualization, W.H. and Z.L.; methodology, Y.L. and X.L.; software, Y.W. and Y.L.; writing–review and editing, X.L. and Z.W.; funding acquisition, X.L, W.H. and Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fujian Province Natural Science Foundation (No: 2021J011028, 2020J01860) and University-industry cooperation program of Department of Science and Technology of Fujian Province (No. 2019H6018) and The Key Project of Science and Technology of Wuhan (No. 201701021010132) and Fund of National Engineering Research Center for Water Transport Safety (No. A2021001) and Fund of Hubei Key Laboratory of Transportation Internet of Things (No. WHUTIOT-2019002). The research leading to these results has received funding from the Norwegian Financial Mechanism 2014-2021 under Project Contract No 2020/37/K/ST8/02748.

Institutional Review Board Statement: Not applicable.
Informed Consent Statement: Not applicable.
Data Availability Statement: Not applicable.
Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wei, Z.; Xie, X.; Lv, W. Self-adaption vessel traffic behaviour recognition algorithm based on multi-attribute trajectory characteristics. *Ocean. Eng.* 2020, 198, 106995. [CrossRef]

2. Zhao, L.; Shi, G. A trajectory clustering method based on Douglas-Peucker compression and density for marine traffic pattern recognition. *Ocean. Eng.* 2018, 172, 456–467. [CrossRef]

3. Chen, Z.; Xue, J.; Wu, C.; Qin, L.; Liu, L.; Cheng, X. Classification of vessel motion pattern in inland waterways based on Automatic Identification System. *Ocean. Eng.* 2018, 161, 69–76. [CrossRef]

4. Tai, J.C.; Tseng, S.T.; Lin, C.P.; Song, K.T. Real-time Image Tracking for Automatic Traffic Monitoring and Enforcement Application. *Image Vis. Comput.* 2004, 22, 485–501. [CrossRef]

5. Ma, F.; Chen, Y.W.; Yan, X.P.; Chu, X.M.; Wang, J. Target recognition for coastal surveillance based on radar images and generalised Bayesian inference. *IET Intell. Transp. Syst.* 2017, 12, 103–112. [CrossRef]

6. Ma, F.; Chen, Y.W.; Yan, X.P.; Chu, X.M.; Wang, J. A novel marine radar targets extraction approach based on sequential images and Bayesian Network. *Ocean. Eng.* 2016, 120, 64–77. [CrossRef]

7. Misović, D.S.; Milić, S.D.; Đurović, Ž.M. Vessel Detection Algorithm Used in a Laser Monitoring System of the Lock Gate Zone. *IEEE Trans. Intell. Transp. Syst.* 2015, 17, 430–440. [CrossRef]

8. Qiao, D.; Liu, G.; Lv, T.; Li, W.; Zhang, J. Marine Vision-Based Situational Awareness Using Discriminative Deep Learning: A Survey. *J. Mar. Sci. Eng.* 2021, 9, 397. [CrossRef]

9. Zhang, H. BoRF, Loop-closure detection with scale invariant visual features. In *Proceedings of the 2011 IEEE International Conference on Robotics and Automation*, Shanghai, China, 9–13 May 2011; pp. 3125–3130.

10. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* 2008, 110, 346–359. [CrossRef]

11. Zhang, X.; Zhou, F.; Lin, Y.; Zhang, S. Embedding label structures for fine-grained feature representation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016; pp. 1114–1123.

12. Kodirov, E.; Xiang, T.; Gong, S. Semantic autoencoder for zero-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 21–23 June 2017; pp. 3174–3183.

13. Teney, D.; Brown, M.; Kit, D.; Hall, P. Learning similarity metrics for dynamic scene segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 7–12 June 2015; pp. 2084–2093.

14. Solmaz, B.; Gundogdu, E.; Yucesoy, V.; Koç, A.; Alatan, A.A. Fine-grained recognition of maritime vessels and land vehicles by deep feature embedding. *IET Comput. Vis.* 2018, 12, 1121–1132. [CrossRef]

15. Voinov, S.; Krause, D.; Schwarz, E. Towards Automated Vessel Detection and Type Recognition from VHR Optical Satellite Images. In *Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, Spain, 22–27 July 2018; pp. 4823–4826.

16. Zhang, Q.; Pless, R. Extrinsic calibration of a camera and laser range finder (improves camera calibration). In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS' 04)*, 28 September–2 October 2004; pp. 2301–2306.

17. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* 2000, 22, 1330–1334. [CrossRef]

18. Vasconcelos, F.; Barreto, J.; Nunes, U. A minimal solution for the extrinsic calibration of a camera and a laser-rangefinder. *IEEE Trans. Pattern Anal. Mach. Intell.* 2012, 34, 2097–2107. [CrossRef] [PubMed]

19. Naikal, N.; Kua, J.; Zakhor, A. Image Augmented Laser Scan Matching for Indoor Localization; Technical Report. No. UCB/EECS-2009-35; University of California Berkeley: Berkeley, CA, USA, 2009.

20. Taylor, Z.; Nieto, J.; Johnson, D. Automatic calibration of multimodal sensor systems using a gradient orientation measure. In *Proceedings of the IEEE International Conference on Intelligent Robots & Systems (IROS)*, Tokyo, Japan, 3–7 November 2013; pp. 21.

21. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* 2012, 25, 1097–1105. [CrossRef]

22. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image

recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

23. Wu, Z.; Zhu, J. *Marine Traffic Engineering*; Dalian Maritime University Press: Dalian, China, 2004. (In Chinese)

24. Changjiang River Administration of Navigational Affairs, MOT. *Report on Changjiang River Shipping Development*; China Communications Publishing & Media Management: Wuhan, China, 2011. (In Chinese)

Note: Figure translations are in progress. See original paper for figures.