Journal of
Marine Science and Engineering

MDPI

Article

Distance Estimation Approach for Maritime Traffic Surveillance Using Instance Segmentation

Miro Petković * and Igor Vujović

Faculty of Maritime Studies, University of Split, Ruđera Boškovića 37, 21000 Split, Croatia; ivujovic@pfst.hr
* Correspondence: mpetkovic@pfst.hr

Abstract: Maritime traffic monitoring systems are particularly important in Mediterranean ports, as they provide more comprehensive data collection compared to traditional systems such as the Automatic Identification System (AIS), which is not mandatory for all vessels. This paper improves the existing real-time maritime traffic monitoring systems by introducing a distance estimation algorithm for monocular cameras, which aims to provide high quality maritime traffic metadata collection for traffic density analysis. Two distance estimation methods based on a pinhole camera model are presented: the Vessel-Focused Distance Estimation (VFDE) and the novel Vessel Object-Focused Distance Estimation (VOFDE). While VFDE uses the predefined height of a vessel for distance estimation, VOFDE uses standardized dimensions of objects on the vessel, detected with a Convolutional Neural Network (CNN) for instance segmentation to enhance estimation accuracy. Our evaluation covers distances up to 414 m, which is significantly beyond the scope of previous studies. When compared to the distances measured with a precise instrument, VOFDE achieves a Percentage Deviation Index (PDI) of 1.34% to 9.45%. This advance holds significant potential for improving maritime surveillance with monocular cameras and is also applicable in other areas, such as low-cost maritime vehicles equipped with single cameras.

Keywords: maritime surveillance; distance estimation; pinhole camera model; instance segmentation

check for updates

Citation: Petković, M.; Vujović, I. Distance Estimation Approach for Maritime Traffic Surveillance Using Instance Segmentation. *J. Mar. Sci. Eng.* **2024**, *12*, 78. https://doi.org/10.3390/jmse12010078

Academic Editors: Jinfen Zhang and Ângelo Palos Teixeira

Received: 14 November 2023
Revised: 22 December 2023
Accepted: 28 December 2023
Published: 28 December 2023

1. Introduction

Analysis of maritime traffic density plays an important role in efficiently managing port operations and ensuring safe navigation. Conventionally, this analysis relies on data obtained from radar-based systems, the Automatic Identification System (AIS), and human observation—each method has its own challenges and limitations. As a result, it is difficult to distinguish between different types of vessels such as passenger ships, fishing vessels, recreational boats, etc. [1]. This lack of specificity hinders a comprehensive analysis of maritime traffic density and complicates the decision-making processes of port authorities. On the other hand, AIS, a transponder-based system, provides data on a vessel's identity, type, position, course, and additional safety-related information. In previous research, it was demonstrated that exclusive reliance on AIS data leads to an incomplete representation of maritime traffic in the Mediterranean, primarily due to the high number of vessels operating without AIS [2]. Results showed that automated maritime video surveillance, employing Convolutional Neural Networks (CNN) for detailed vessel classification, can capture 386% more traffic than the AIS alone. This highlights the significant potential of automated maritime video surveillance systems to enhance maritime security and traffic monitoring. These systems can overcome the shortcomings of traditional methods by providing more detailed, reliable, and comprehensive maritime surveillance.

The existing real-time maritime traffic counting system based on a neural network, presented in [2], is used to monitor incoming and outgoing traffic in the port of Split, Croatia. This research aims to improve this system by introducing distance estimation between the camera and the vessels, an important component of advanced monitoring systems for analyzing maritime traffic density. Distance estimation methods from images or videos, i.e., the extraction of 3D information from 2D sources, is an area that has been intensively researched over the last decade. Its applicability extends across various fields, including autonomous driving [3], traffic safety [4], animal ecology [5], assistive technologies for the visually impaired [6], etc.

As shown in Figure 1, the existing counting system uses a single static long-range video camera to monitor traffic. The captured video stream is then processed by an object detection module that identifies and classifies vessels, which are then tracked and quantitatively counted. Vessels are expected to pass through the monitored zone at distances ranging from 150 m to 500 m from the camera. To implement distance estimation, a monocular approach using the pinhole camera model in conjunction with triangle similarity is investigated. This technique requires knowledge of the height (or width) of the observed vessel, as well as a reference height (or width) for each class of vessels to enable accurate distance estimation. It is important to acknowledge that the pinhole camera model can incur estimation errors of up to 20%, as noted in [5–7].

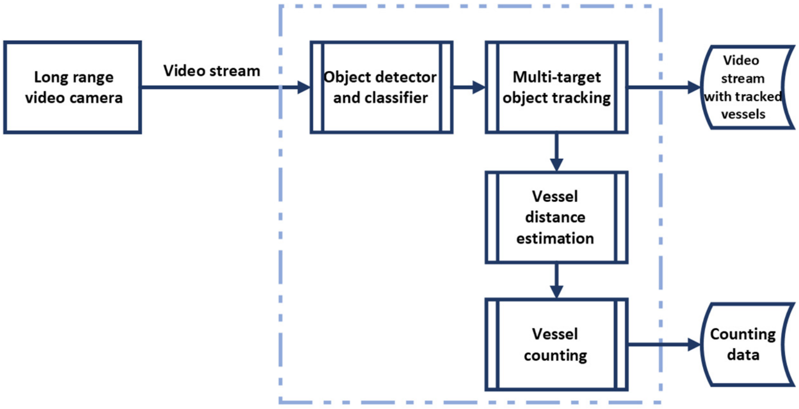

Long range video camera
Video stream
Object detector and classifier
Multi-target object tracking
Video stream with tracked vessels
Vessel distance estimation
Vessel counting
Counting data

Figure 1. Illustration of the real-time maritime traffic counting system.

Initially, the Vessel-Focused Distance Estimation (VFDE) method was developed using pinhole camera model. This method estimates the distance directly by focusing on the vessel as a whole, using the vessel's height as a parameter for the distance calculation. VFDE integrates the YOLOv4 object detector [8], a component already integrated into the existing counting system. While the preliminary results of VFDE were promising, it was found that significant height variability within certain vessel classes led to increased estimation errors. These errors depend on the discrepancy between the reference and actual heights of the vessels [5,7].

To address this challenge, a new approach, the Vessel Object-Focused Distance Estimation (VOFDE), is proposed. This method uses the standardized dimensions of objects commonly found on the decks and sides of vessels, such as safety equipment and recovery devices, for distance estimation. These objects are identified using an instance segmentation CNN that accurately delineates their shapes and sizes, thus increasing the accuracy of the distance estimation process. The main scientific contribution can be summarized as follows:

- Proposal of the VOFDE method, a monocular camera distance estimation, to reduce estimation errors for vessel classes with significant height variability;
- Validation of the VFDE and VOFDE approaches by comparison with actual distance measurements and distances derived from AIS in real-world applications.

This paper is organized as follows: Section 2 presents Related Work, Section 3 presents Methodology, while Section 4 presents Results. Finally, a discussion and conclusions are given.

2. Related Work

Camera-based distance estimation methods can be divided into two main categories: the use of a monocular camera and a stereo vision [9]. Stereo vision emulates how human eyes perceive depth, employing two cameras placed at a certain distance apart. The major advantage of these systems is their ability to provide rich, accurate, and detailed depth information in real-time. However, stereo vision systems are more complex and expensive due to the need for two cameras, and they require careful calibration to ensure the two cameras are properly aligned [7]. Monocular vision systems, on the other hand, use only one camera to capture images or videos. Since there is only one input, these systems do not natively provide depth information. Instead, they rely on other cues to estimate depth, such as object size, perspective, and shadows. Moreover, some articles proposed adding RFID [10] as a complement to the surveillance camera for distance estimation, or additional sensors such as a LiDAR sensor [11] or Kinect sensor [12]. Given that our system employs a single camera for capturing maritime traffic and adding a secondary sensor is not possible, our research narrows its focus to monocular camera distance estimation methods.

The paper [13] proposes a monocular vision-based method for estimating face-to-camera distance using a single camera. The approach includes three steps: extraction and location of feature region, calculation of the pixel area of the characteristic triangle, and construction of a measurement formula (derived from the pinhole camera model). According to the experimental analysis, the proposed method shows over 95% accuracy with a processing time of about 230 ms. Authors in [14] evaluated three distance estimation methods for road obstacles. The methods are based on the pinhole camera model: one uses the geometry of similar triangles, another utilizes the cross-ratio of a set of collinear points, and the last relies on camera matrix calibration. Results suggest that the triangle similarity method is more suitable for distance estimation at wide angles. For accurate distance estimation to the object (from a vehicle) using a single camera, the paper [15] presented the technique that treats the camera as an ideal pinhole model. This technique employs the principle of triangle similarity for distance estimation after determining the camera's focal length in pixels. Object detection is achieved through image processing techniques, and the known width of the object, along with the calculated focal length, is used by the algorithm to calculate the distance after identification. Experimental results showed satisfactory accuracy at short range distances for objects of varying widths, while it is noted that the technique may not perform as well with unknown or variable width objects. While in [7], an experiment is conducted to calculate human distance using a single camera, pinhole camera model, and triangle similarity concept to estimate the distance. The results reveal that for shorter distances, an estimation error ranges from less than 10% [5,7] up to 17% [6], while the error rate increases to 20% [7] for longer distances. It is noted that the error depends on the difference between the reference height and the actual height of the target [5,7].

Prior to distance estimation, there were two main approaches for object detection in the literature: traditional machine learning techniques [10,13,15], and neural network methods [3,6,7,16–18]. The results indicate that the accuracy of distance estimation is closely related to the quality of the detection results. Furthermore, Ref. [18] argues for future improvements by using more sophisticated CNN models that include object segmentation. The goal of segmentation is to make detailed predictions by assigning labels to each pixel of an image, and classifying each pixel based on the object or region it belongs to [19]. Instance segmentation, which combines the tasks of object detection and segmentation [20], has evolved into a technique in its own right. It is characterized by its ability to identify and separate individual objects of the same class, which enables detailed examination at the pixel level [19,21].

3. Methodology

This section presents the methodology, beginning with an overview of the pinhole camera model, followed by detailed descriptions of the VFDE and proposed VOFDE methods.

3.1. Pinhole Camera Model

The object detector provides essential 2D information about the detected target, including its x and y coordinates (in pixels) and the width and height of the target (in pixels). This information serves as the basis for distance estimation using the pinhole camera model, applying the concept of triangle similarity. Distance estimation based on the pinhole camera model relies on the relationships among key variables, as illustrated in Figure 2.

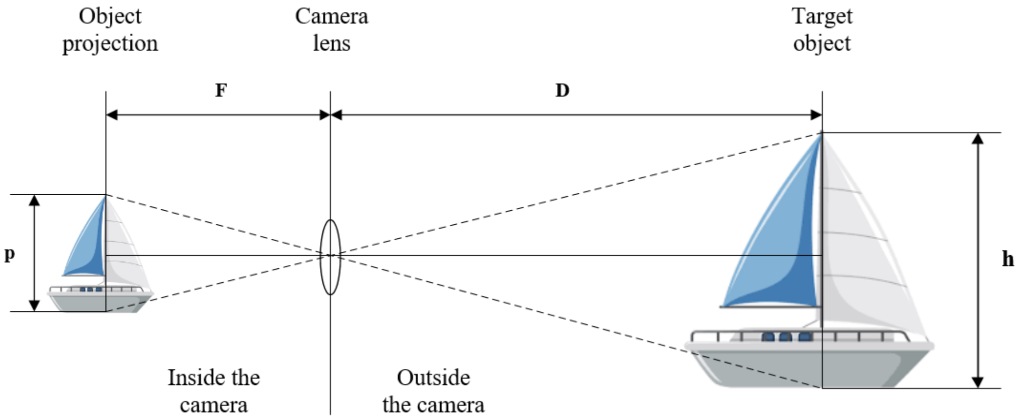


Figure 2. Distance estimation based on the pinhole camera model and triangle similarity concept.

Let D represent the distance from the camera to the target, h as the actual height (or width) of the target, p as the height (or width) of the object in the image (in pixels), and F as the camera's focal length. In the pinhole model (Figure 1), these variables form two similar triangles, giving rise to the equation:

$$\frac{F}{p} = \frac{D}{h}, \tag{1}$$

The calibration process, which establishes the focal length F as a constant, is detailed in the following section. If we assume that the actual height (or width) h of the target is known, while the height (or width) p of the target in an image frame is provided by the object detector, the formula for calculating the distance estimation D is derived from Equation (1).

3.2. Vessel-Focused Distance Estimation

Preliminarily, the Vessel-Focused Distance Estimation (VFDE) method was developed to estimate the distance directly to the detected vessel using the pinhole camera model. This model requires the specification of a reference height (or width) for each object class that is detected and classified by YOLOv4, a component of the real-time maritime counting system. An important consideration is the movement of vessels entering the port. This movement is usually either parallel to the camera or at a certain angle, which is influenced by the positioning and viewing angle of the camera. When estimating distance, the use of the vessel's width can lead to estimation errors if the vessel approaches at an angle, as the actual observed width can deviate considerably from the set reference width. To mitigate this problem, the VFDE method uses the height of the vessel as a reference variable for distance estimation.

Furthermore, certain height factors of the vessels must be taken into account. As seen in Figure 3, the overall height of a vessel, referred to as Keel to Masthead (KTM), includes the draft, i.e., the vertical distance between the waterline and the deepest point of the vessel. Since only the part of the vessel that is above the waterline is captured by the camera, the air draft (AD), i.e., the vertical distance between the waterline and the highest point of the vessel, is used as the reference height.

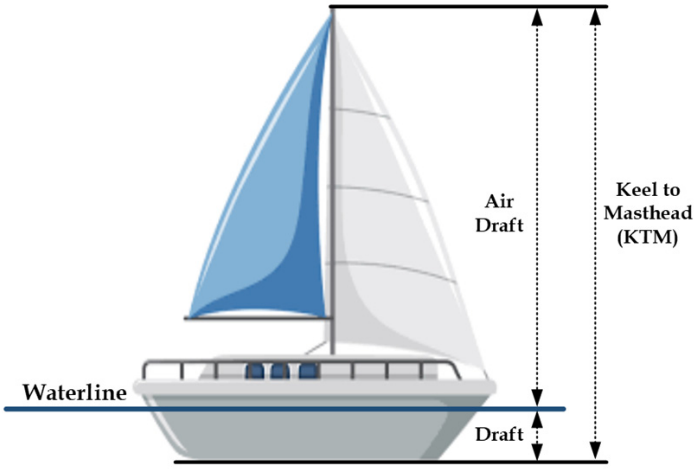


Figure 3. Vessel vertical distances.

3.3. Vessel Object-Focused Distance Estimation

To reduce the distance estimation error for vessel classes with significant height variability, we propose the Vessel Object-Focused Distance Estimation (VOFDE) method, shown in Figure 4. Similar to the VFDE, VOFDE is based on the pinhole camera model. Initially, the object detector YOLOv4 detects and classifies the vessel. If the height of the vessel class is defined (indicating no or minimal height variability), the VFDE method is applied. On the other hand, if the height of the vessel class is undefined (indicating significant height variability), the VOFDE method is employed. The VOFDE method uses the standardized dimensions of objects located on the deck and sides of vessels, such as safety equipment (life raft container, lifeboat, etc.), recovery devices, fenders, etc., to estimate the distance. The YOLOv5 instance segmentation CNN [22] is used both for detecting the objects and for determining their dimensions, which enables an accurate distance estimation to the object and, consequently, to the vessel.

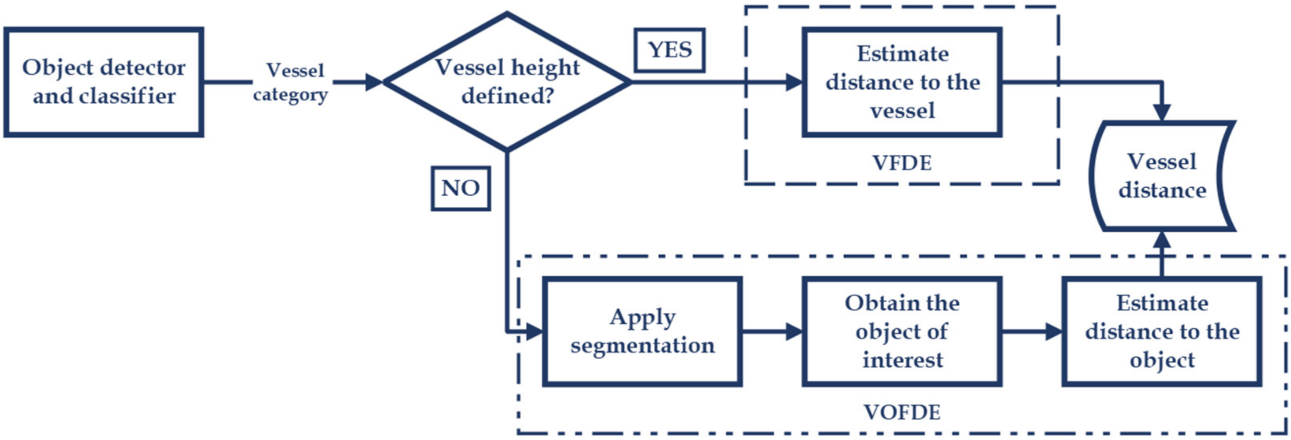

Object detector and classifier
Vessel category
Vessel height defined?
YES
Estimate distance to the vessel
VFDE
Vessel distance
NO
Apply segmentation
Obtain the object of interest
Estimate distance to the object
VOFDE

Figure 4. Flowchart of the proposed distance estimation.

4. Results

In this section, we detail the practical aspects of our paper, covering hardware settings, software implementation, camera calibration, and the validation of methods with corresponding results.

4.1. Hardware Settings

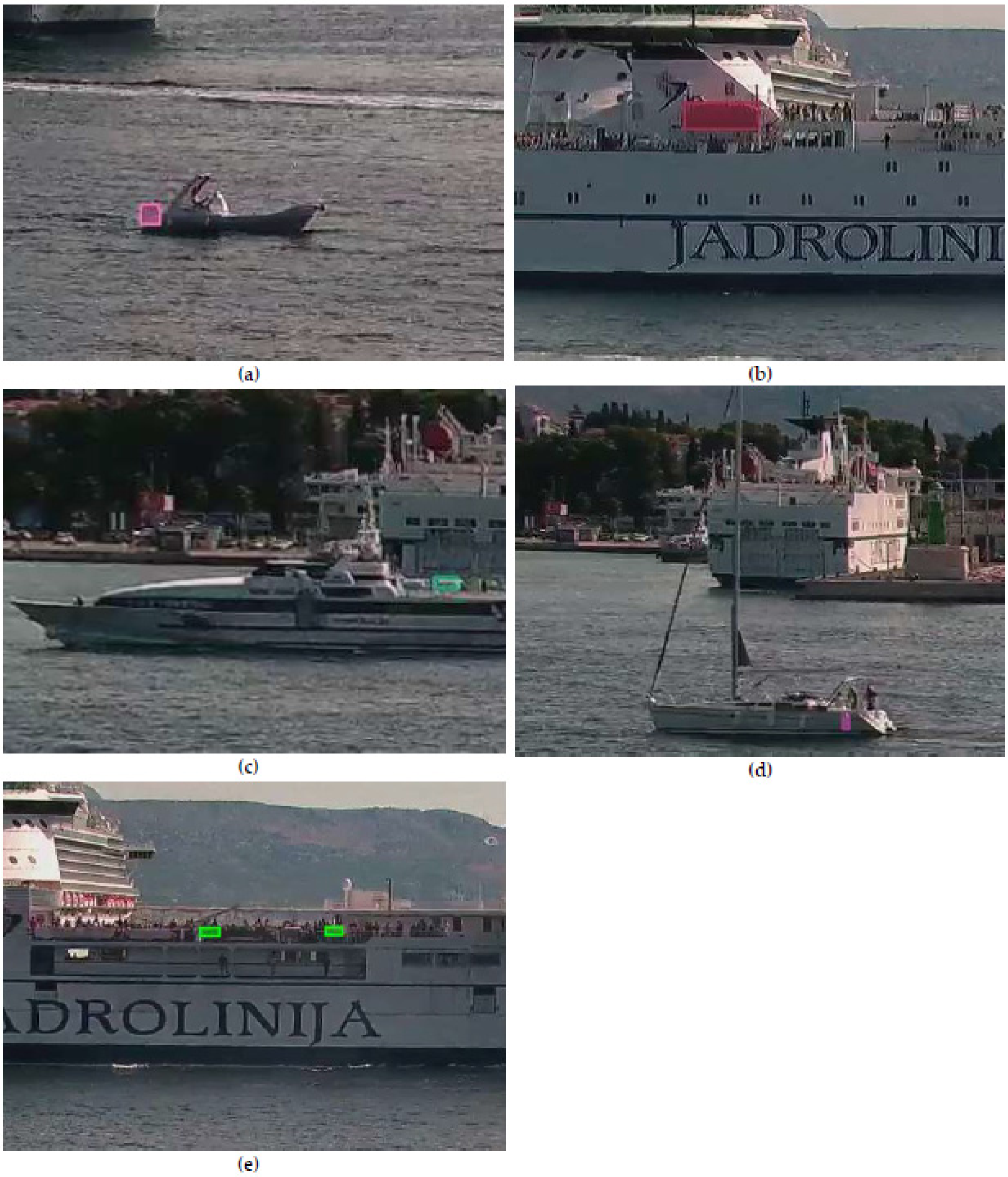
For monitoring all incoming and outgoing maritime traffic at the entrance to the Port of Split, a long-range surveillance Dahua DH-TPC-PT8620A-T camera (Dahua, Hangzhou, China) [23] is employed. The camera is positioned on a building overlooking the port entrance, located at 43°30′04″ N latitude and 16°25′48″ E longitude, with an elevation of 9 m above sea level. This positioning provides an optimal vantage point for capturing maritime traffic. The camera delivers a video stream in H.264 compression, a resolution of 1920 × 1080, and a frame rate of 25 frames per second (FPS). The computational resources used for this project include an Intel Core i7-9700K processor (Intel, Santa Clara, CA, USA), 32 GB of DDR4 RAM, and a Nvidia GeForce RTX 2080 with 8 GB RAM (Nvidia, Sanata Clara, CA, USA). Additionally, the Stonex R15 Total Station [24], a high-precision instrument, was employed to provide accurate reference distance measurements necessary for the validation and calibration of our distance estimation methods.

4.2. Software Implementation

The real-time maritime traffic counting system, as detailed in [2], is developed in C++ and incorporates a YOLOv4 CNN for vessel detection. This detector plays a crucial role in the VFDE method. The YOLOv4 model was trained with the Split Port Ship Classification Dataset (SPSCD) [1], which contains 19,337 images with a resolution of 1920 × 1080. This dataset includes a total of 27,849 labelled vessels divided into 12 different classes, ensuring a comprehensive coverage of vessel types. Then, the reference heights were then defined for each of the 12 vessel classes. For example, the reference height for the vessel class 'Sailboat' was determined by examining the AD height of vessels of different lengths (between 32 and 47 feet) from several manufacturers [25,26]. The average AD height, measuring 17.48 m, was then calculated and set as the reference height for this particular vessel class. For the vessel classes 'Ferry' and 'Large Ferry', the national ferry company Jadrolinija [27] provided the AD heights of all its vessels. The average AD height is then calculated for each of these vessel classes and set as the reference height.

The VOFDE method, developed in C++, uses YOLOv5: 7.0, a CNN optimized for real-time instance segmentation. First, a database of objects located on the deck or sides of the vessels is created, ranging from safety equipment (life raft container, lifeboat, etc.) to recovery devices, fenders, and outboard engine caps. The YOLOv5 model is then trained with the pre-trained weights yolo5l-seg.pt for 300 epochs, with a custom dataset consisting of 300 images of the above objects with a resolution of 1920 × 1080. Figure 5 shows examples of the vessels with the detected objects. The input size of the image for YOLOv5 is set to 960 × 960 pixels, with default training parameters such as a learning rate of 0.01, a momentum of 0.937, a weight decay of 0.0005, and the activated mosaic data augmentation. The reference heights are then defined for each object class. For example, for the vessel classes 'Highspeed craft', 'Ferry', and 'Large Ferry', we distinguish two common container types for throw-overboard life rafts: a smaller container for life rafts with a capacity of less than 100 persons [28], and a larger container for life rafts with a capacity of more than 100 persons [29]. The average container size is calculated for each class, resulting in reference heights of 0.73 m and 1.2 m, respectively. When investigating vessel classes such as 'Sailboat' or smaller vessels such as 'Speedboat', various objects with relatively standardized dimensions are identified. For example, a fender hanging on the side of the vessel can be used to estimate the distance. They are divided into two classes: 'Small fender' with a reference height of 0.6 m, and 'Large fender' with a reference height of 1.2 m.

It is important to note that for the distance estimation in a particular frame, the results are averaged over the 10 previous frames. In addition, in the VOFDE method, if multiple objects are detected in a frame, the distance estimate to the vessel is calculated by averaging the results of all detected objects over the 10 previous frames.



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

Figure 5. Example images of the vessels (enlarged for better viewing) with detected objects are as follows: (**a**) Speedboat 1 with 'Motor cap'; (**b**) Large Ferry 1 with 'Free-fall Lifeboat'; (**c**) Highspeed craft with 'Large Life raft container'; (**d**) Sailboat with 'Small fender' vessels; (**e**) Ferry 2 with 'Small Life raft container'.

4.3. Camera Calibration

The calibration process is an important step of distance estimation based on pinhole camera mode. It begins by calculating the focal length (F, in pixels), which is an important variable. To do this, an object with a known real height is placed within the camera's field of view. At the same time, the actual distance between the camera and the object is determined using a precise distance measuring station. The object's height within the image (measured in pixels) is then determined. These variables are input into Equation (1) to calculate the focal length. This calibration process is conducted three times at different distances from the camera. The results, as shown in Table 1, are then averaged to determine the reference focal length for the system, as recommended in [5,15].

Table 1. Focal lengths at various distances.

Real Measured Distance (m)	Focal Length (px)
168.9	4011.5
350	4170.8
489.9	4347.3

4.4. Validation

To validate the proposed VOFDE method, multiple video sequences of vessels entering or leaving the port of Split were recorded. At the same time, the actual distance from the camera to each recorded vessel was measured using a precise distance measuring station positioned near the camera. Additionally, for vessels with active AIS, the data is used to calculate the distance between the vessel and the camera. Figure 6 presents a map of the entrance to the port of Split with the camera location.

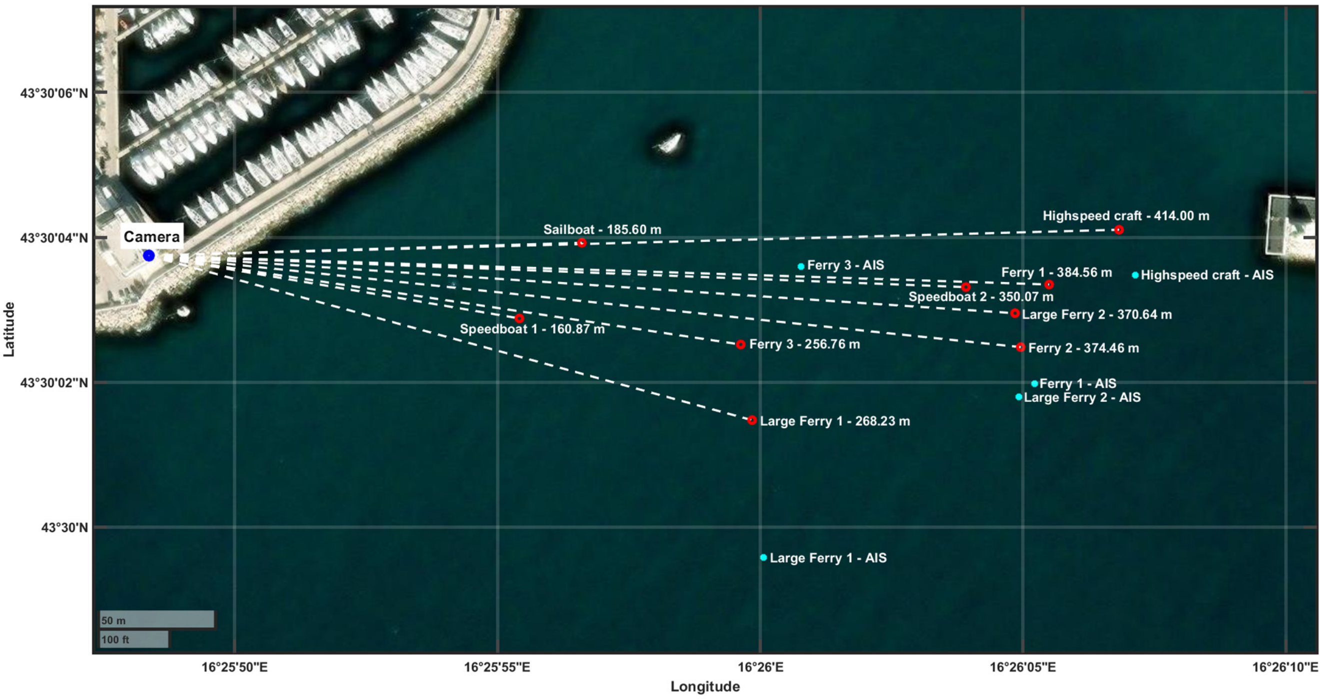


Figure 6. Entrance to the port of Split with the camera location.

On the map, the approximate positions where the distances to the vessels were measured are indicated with red circles. Each circle is labelled with the vessel class, the vessel number, and the measured distance. The cyan circles represent the AIS positions of the vessels that had the system active at the time.

The accuracy of the VFDE and the proposed VOFDE method were evaluated by comparing their results with the measured distances by using the Percentage Deviation Index (PDI) measure. To calculate the PDI, the measured distance (d_m) was used as a reference to express the PDI as follows:

$$\mathrm{PDI}[\%] = \frac{|d_e - d_m|}{d_m} \times 100, \tag{2}$$

where d_e denotes the estimated distance using the VFDE and VOFDE methods. The obtained distance estimates of each method and the PDI results are summarized in Table 2.

Table 2. Comparison of VFDE and VOFDE with the measured distance.

Vessel Class	Vessel Number	Measured Distance (m)	VFDE (m)	PDI_{VFDE} (%)	Proposed VOFDE (m)	$\text{PDI}_{\text{VOFDE}}$ (%)
Sailboat	1	185.60	208.62	12.40%	179.25	3.42%
Speedboat	1	168.87	190.47	12.79%	180.39	6.82%
Speedboat	2	350.07	377.89	7.95%	354.67	1.31%
Highspeed craft	1	414.00	367.65	11.20%	453.14	9.45%
Ferry	1	384.56	387.44	0.75%	369.87	3.82%
Ferry	2	374.46	431.91	15.34%	402.49	7.49%
Ferry	3	256.76	352.85	37.42%	279.49	8.85%
Large Ferry	1	268.23	218.14	18.67%	290.09	8.15%
Large Ferry	2	370.64	499.03	34.64%	401.60	8.35%

On average, VFDE achieved PDI of 16.80%, while the proposed VOFDE achieved PDI of 6.41%.

To further validate both VFDE and VOFDE methods, AIS data is utilized. The GPS coordinates for each vessel are determined from the AIS system at the same time as the precise distance measurement took place. The distance between the AIS vessel coordinates and the camera coordinates is then calculated. The accuracy of the VFDE and the proposed VOFDE method were evaluated by comparing their results with the AIS derived distance using the PDI measure. The results of the two methods and the PDI results are shown in Table 3.

Table 3. Comparison of VFDE and VOFDE with AIS distances.

Vessel Class	Vessel Number	AIS Distance (m)	VFDE (m)	PDI_{VFDE} (%)	Proposed VOFDE (m)	$\text{PDI}_{\text{VOFDE}}$ (%)
Highspeed Craft	1	420.80	367.65	12.63%	453.14	7.69%
Ferry	1	381.60	387.44	1.53%	369.87	3.07%
Ferry	3	278.20	352.85	26.83%	279.49	0.46%
Large Ferry	1	292.20	218.14	25.35%	290.09	0.72%
Large Ferry	2	375.80	499.03	32.79%	401.60	6.87%

5. Discussion

This paper addresses the need for reliable distance estimation between a monocular camera and the object of interest in real-time maritime video surveillance. Accurate distance estimation is important to collect valuable data for future scientific research in the dynamic maritime environment, characterized by a wide variety of vessel sizes and types. In this paper, two approaches for distance estimation using the pinhole camera model are evaluated in a real-world application in a maritime environment. Existing studies evaluated the pinhole camera model across various distances; distances up to 5 m are evaluated in [6,7,13,15,18], up to 15.5 m in [6], up to 30 m in [14], and up to 96 m in [5]. In contrast, our research extends this validation to distances up to 414 m. Additionally, the literature review suggests that the expected estimation error of the pinhole camera model ranges from under 10% [5,7], to 17% [6], and up to 20% [7].

A particular challenge arises from the significant size variability within certain classes of vessels, such as the ‘Sailboat’ class, which can vary in length from 24 to over 52 feet, affecting their height. This variability poses a challenge for distance estimation using the pinhole camera model and the triangle similarity approach, as it requires the establishment of a reference height (or width) for each vessel class to enable accurate distance estimation. Significant deviations in object dimensions from the reference height can lead to a significant increase in the distance estimation error.

Table 2 illustrates these challenges by comparing the VFDE estimates to distances measured with an accurate instrument. The PDI results in this approach ranged from 0.75% to 37.42%. In contrast, the proposed VOFDE method achieved a lower PDI, ranging from 1.31% to 9.45%. Notably, while the VFDE estimation achieved the lowest PDI of 0.75% in one instance, the VOFDE method generally performed better.

Deviations between 0.78% and 8.20% were found between the measured distance and the AIS derived distance. This deviation is attributed to the different locations of the measuring points. According to the IMO guidelines for AIS installation [30], the GNSS antenna should have a 360° clear view of the horizon, which is usually the case at the highest point of the vessel (on the masts, superstructure, flybridge, etc.). Therefore, there is a difference between the GPS coordinates obtained from the AIS and the measurement point of the precise instrument, as they were taken at random points on the vessel.

Furthermore, a comparative analysis of the two methods was conducted against the distances derived from the AIS position for each vessel. As shown in Table 3, the PDI for the VFDE method ranged from 1.53% to 32.79%, while the PDI for the proposed VOFDE method was consistently lower, ranging from 0.46% to 7.69%. It is important to note that when using AIS-derived distances, compensation for the vessel's beam is necessary. This adjustment accounts for the GPS coordinate typically being at the vessel's midpoint, whereas monocular camera distance estimations are generally to the vessel's edge.

An important conclusion can be drawn from this with regard to measurement uncertainty: errors in distance estimation also indirectly indicate possible discrepancies between the actual height of an object and its reference height. In addition, inaccuracies in object detection can contribute to these distance estimation errors. This is evident when the bounding box (or object segmentation) inaccurately represents the size of the detected object, and is either larger or smaller than the actual object.

This paper demonstrates the reliability of distance estimation algorithms using a monocular camera, particularly validating the pinhole camera model with triangle similarity for maritime video surveillance systems. This validation is crucial in scenarios where additional hardware for distance measurement or a second camera for stereo vision is not feasible. The lower error rates achieved with the proposed VOFDE method are valuable for research applications in maritime traffic density analysis. More accurate data collection through intelligent video surveillance is important for research focused on understanding and analyzing maritime traffic, and contributes to progress in this field. Furthermore, the impact of this research goes beyond maritime surveillance. The VOFDE method can also be used in underwater or surface low-cost vehicles equipped with a single camera, providing a cost-effective and practical solution for distance and position estimation. This adaptability of the VOFDE method to different maritime contexts emphasizes its versatility and broad applicability in the field of maritime observation and analysis.

In future studies, a range of CNNs for instance segmentation, such as YOLOv8, will be investigated to improve the accuracy of distance estimation in maritime surveillance systems. An in-depth study on error propagation is also planned, focusing particularly on how deviations of heights from the reference height affect the accuracy.

6. Conclusions

This paper enhances a real-time maritime traffic counting system by introducing an algorithm for distance estimation between vessels and a camera. Initially, the system used the Vessel-Focused Distance Estimation (VFDE) method based on the pinhole camera model and a predefined reference height for each vessel class. A novel Vessel Object-Focused Distance Estimation (VOFDE) method is developed to address the challenge of height variability in some vessel classes. VOFDE employs standardized dimensions of objects on vessels, identified through a Convolutional Neural Network (CNN) for instance segmentation, to improve distance estimation accuracy.

The significant contribution of this research lies in advancing distance estimation using the pinhole camera model and monocular cameras. Our real-world evaluations, covering distances up to 414 m and comparing results with actual measurements and AIS data, demonstrate that VOFDE achieves a Percentage Deviation Index (PDI) ranging from 1.31% to 9.45%. This improvement in accuracy is important for maritime traffic density analysis and has potential applications in GPS-denied environments and low-cost maritime vehicles equipped with single cameras.

Author Contributions: Conceptualization, M.P.; methodology, M.P. and I.V.; software, M.P.; validation, M.P. and I.V.; formal analysis, M.P. and I.V.; resources, M.P.; data curation, M.P.; writing—original draft preparation, M.P. and I.V.; writing—review and editing, I.V.; visualization, M.P.; supervision, I.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding, but the equipment used is obtained by the project "Functional integration of the University of Split, PMF/PFST/KTF through the development of scientific and research infrastructure in the three faculty (3F) building" grant number: KK.01.1.1.02.0018.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Video sequences with real distances, used for the validation of distance estimation, can be downloaded from: https://labs.pfst.hr/maritime-dataset/ (accessed on 10 November 2023) and used for scientific purposes only.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Petković, M.; Vujović, I.; Lušić, Z.; Šoda, J. Image Dataset for Neural Network Performance Estimation with Application to Maritime Ports. *J. Mar. Sci. Eng.* **2023**, *11*, 578. [CrossRef]
2. Petković, M.; Vujović, I.; Kaštelan, N.; Šoda, J. Every Vessel Counts: Neural Network Based Maritime Traffic Counting System. *Sensors* **2023**, *23*, 6777. [CrossRef] [PubMed]
3. Arabi, S.; Sharma, A.; Reyes, M.; Hamann, C.; Peek-Asa, C. Farm Vehicle following Distance Estimation Using Deep Learning and Monocular Camera Images. *Sensors* **2022**, *22*, 2736. [CrossRef] [PubMed]
4. Liu, L.C.; Fang, C.Y.; Chen, S.W. A Novel Distance Estimation Method Leading a Forward Collision Avoidance Assist System for Vehicles on Highways. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 937–949. [CrossRef]
5. Leorna, S.; Brinkman, T.; Fullman, T. Estimating Animal Size or Distance in Camera Trap Images: Photogrammetry Using the Pinhole Camera Model. *Methods Ecol. Evol.* **2022**, *13*, 1707–1718. [CrossRef]
6. Chou, K.S.; Wong, T.L.; Wong, K.L.; Shen, L.; Aguiari, D.; Tse, R.; Tang, S.K.; Pau, G. A Lightweight Robust Distance Estimation Method for Navigation Aiding in Unsupervised Environment Using Monocular Camera. *Appl. Sci.* **2023**, *13*, 11038. [CrossRef]
7. Saputra, D.E.; Senjaya, A.S.M.; Ivander, J.; Chandra, A.W. Experiment on Distance Measurement Using Single Camera. In Proceedings of the ICOIACT 2021—4th International Conference on Information and Communications Technology: The Role of AI in Health and Social Revolution in Turbulence Era, Yogyakarta, Indonesia, 30–31 August 2021; pp. 80–85. [CrossRef]
8. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
9. Aicardi, I.; Chiabrando, F.; Maria Lingua, A.; Noardo, F. Recent Trends in Cultural Heritage 3D Survey: The Photogrammetric Computer Vision Approach. *J. Cult. Herit.* **2018**, *32*, 257–266. [CrossRef]
10. Duan, C.; Rao, X.; Yang, L.; Liu, Y. Fusing RFID and Computer Vision for Fine-Grained Object Tracking. In Proceedings of the IEEE INFOCOM 2017—IEEE Conference on Computer Communications, Atlanta, GA, USA, 1–4 May 2017; pp. 1–9.
11. Heimberger, M.; Horgan, J.; Hughes, C.; McDonald, J.; Yogamani, S. Computer Vision in Automated Parking Systems: Design, Implementation and Challenges. *Image Vis. Comput.* **2021**, *68*, 88–101. [CrossRef]
12. Eric, N.; Jang, J.W. Kinect Depth Sensor for Computer Vision Applications in Autonomous Vehicles. In Proceedings of the International Conference on Ubiquitous and Future Networks, ICUFN, Milan, Italy, 4–7 July 2017; pp. 531–535. [CrossRef]
13. Dong, X.; Zhang, F.; Shi, P. A Novel Approach for Face to Camera Distance Estimation by Monocular Vision. *Int. J. Innov. Comput. Inf. Control* **2014**, *10*, 659–669.
14. Nienaber, S.; Kroon, R.S.; Booysen, M.J. A Comparison of Low-Cost Monocular Vision Techniques for Pothole Distance Estimation. In Proceedings of the 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, South Africa, 7–10 December 2015; pp. 419–426.
15. Megalingam, R.K.; Shriram, V.; Likhith, B.; Rajesh, G.; Ghanta, S. Monocular Distance Estimation Using Pinhole Camera Approximation to Avoid Vehicle Crash and Back-over Accidents. In Proceedings of the 10th International Conference on Intelligent Systems and Control (ISCO 2016), Coimbatore, India, 7–8 January 2016. [CrossRef]

16. Ahmed, I.; Ahmad, M.; Rodrigues, J.J.P.C.; Jeon, G.; Din, S. A Deep Learning-Based Social Distance Monitoring Framework for COVID-19. *Sustain. Cities Soc.* **2021**, *65*, 102571. [CrossRef] [PubMed]
17. Xu, X.; Chen, X.; Wu, B.; Yip, T.L. An Overview of Robust Maritime Situation Awareness Methods. In Proceedings of the 6th International Conference on Transportation Information and Safety: New Infrastructure Construction for Better Transportation (ICTIS 2021), Wuhan, China, 22–24 October 2021; pp. 1010–1014.
18. Li, H.; Qiu, J.; Yu, K.; Yan, K.; Li, Q.; Yang, Y.; Chang, R. Fast Safety Distance Warning Framework for Proximity Detection Based on Oriented Object Detection and Pinhole Model. *Measurement* **2023**, *209*, 112509. [CrossRef]
19. Hafiz, A.M.; Bhat, G.M. A Survey on Instance Segmentation: State of the Art. *Int. J. Multimed. Inf. Retr.* **2020**, *9*, 171–189. [CrossRef]
20. Sharma, R.; Saqib, M.; Lin, C.T.; Blumenstein, M. A Survey on Object Instance Segmentation. *SN Comput. Sci.* **2022**, *3*, 1–23. [CrossRef]
21. Jung, S.; Heo, H.; Park, S.; Jung, S.U.; Lee, K. Benchmarking Deep Learning Models for Instance Segmentation. *Appl. Sci.* **2022**, *12*, 8856. [CrossRef]
22. Jocher, G.; Chaurasia, A.; Stoken, A.; Borovec, J.; NanoCode012; Kwon, Y.; Michael, K.; Xie, X.; Fang, J.; Imyhxy; et al. Ultralytics/Yolov5: V7.0—YOLOv5 SOTA Realtime Instance Segmentation. Zenodo. 2022. Available online: https://zenodo.org/records/7347926 (accessed on 5 May 2023).
23. Dahua TPC-PT8620A-TB. Available online: https://www.dahuasecurity.com/products/productDetail/26237 (accessed on 11 February 2023).
24. R15 Total Station—Stonex. Available online: https://www.stonex.it/project/r15-total-station/#data%20sheet (accessed on 14 October 2023).
25. Sun Odyssey. Jeanneau Boats. Range of Boats Jeanneau. Available online: https://www.jeanneau.com/en-us/boats/sailboat/2-sun-odyssey (accessed on 3 November 2023).
26. Beneteau Oceanis—31- to 60-Foot Cruisers. Beneteau. Available online: https://www.beneteau.com/sailing-yachts/oceanis (accessed on 3 November 2023).
27. Jadrolinija. Available online: https://www.jadrolinija.hr/ (accessed on 3 September 2023).
28. Liferaft—VIKING, 35DK+, Throw Overboard (35 Pers.). Available online: https://www.viking-life.com/shop/liferafts-and-accessories/liferafts/throw-overboard/liferaft-viking-35dkplus-throw-overboard-35-pers/ (accessed on 4 October 2023).
29. Liferaft—VIKING, 150DKS, Throw Overboard, Self-Righting, (153 Pers.). Available online: https://www.viking-life.com/shop/liferafts-and-accessories/liferafts/throw-overboard-self-righting/liferaft-viking-150dks-throw-overboard-self-righting-153-pers/ (accessed on 4 October 2023).
30. Annex—Guidelines for the Installation of a Shipborne Automatic Identification System (AIS). Available online: https://www.imorules.com/SNCIRC_227_ANN.html (accessed on 10 December 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.