Article

Computer Vision-Based Position Estimation for an Autonomous Underwater Vehicle

Jacek Zalewski * and Stanisław Hożyń

Faculty of Mechanical and Electrical Engineering, Polish Naval Academy, 81-127 Gdynia, Poland; s.hozyn@amw.gdynia.pl
* Correspondence: j.zalewski@amw.gdynia.pl

Abstract: Autonomous Underwater Vehicles (AUVs) are currently one of the most intensively developing branches of marine technology. Their widespread use and versatility allow them to perform tasks that, until recently, required human resources. One problem in AUVs is inadequate navigation, which results in inaccurate positioning. Weaknesses in electronic equipment lead to errors in determining a vehicle's position during underwater missions, requiring periodic reduction of accumulated errors through the use of radio navigation systems (e.g., GNSS). However, these signals may be unavailable or deliberately distorted. Therefore, in this paper, we propose a new computer vision-based method for estimating the position of an AUV. Our method uses computer vision and deep learning techniques to generate the surroundings of the vehicle during temporary surfacing at the point where it is currently located. The next step is to compare this with the shoreline representation on the map, which is generated for a set of points that are in a specific vicinity of a point determined by dead reckoning. This method is primarily intended for low-cost vehicles without advanced navigation systems. Our results suggest that the proposed solution reduces the error in vehicle positioning to 30–60 m and can be used in incomplete shoreline representations. Further research will focus on the use of the proposed method in fully autonomous navigation systems.

Keywords: computer vision; deep learning; robotics; navigation

1. Introduction

The marine environment poses inherent risks to human exploration, characterized by various hazards that threaten human safety during deep-sea endeavors. The physiological limitations of the human body during prolonged exposure at significant depths require specialized life support equipment, making human presence on board a hazardous prospect [1]. Recognizing this imperative, the integration of Unmanned Underwater Vehicles (UUVs) has emerged as a key aspect of modern maritime technology. Due to their ubiquity and adaptability, UUVs have become indispensable tools, effectively replacing solutions that traditionally require human involvement. Eliminating the need for human presence aboard these vehicles is paramount to ensuring both safety and operational efficiency in the challenging marine environment. In addition to overcoming the constraints imposed by human physiological limitations, UUVs are distinguished by their ability to carry out prolonged missions, facilitating extended data collection without the constraints of human fatigue. The unparalleled endurance of unmanned vehicles enables them to operate continuously for extended periods of time, covering large areas and performing repetitive tasks with a level of precision that is unattainable by their human counterparts. This capability has not only revolutionized the scope of marine exploration but has also significantly improved the efficiency and effectiveness of marine data collection. As a result, the integration of unmanned vehicles is a cornerstone of modern maritime technology, introducing a new era of exploration and data collection capabilities in the challenging and dangerous marine environment [2].

Two common solutions have emerged in the field of underwater exploration, each addressing different operational requirements. The first solution involves the use of Remotely Operated Underwater Vehicles (ROUVs) attached to a Surface Control Unit (SCU) via a cable harness [3]. This configuration facilitates the transmission of power, control signals, and imagery between the ROUV and the SCU. ROUVs are highly adaptable and can integrate a wide range of sensors, cameras, and manipulators. This versatility enables them to perform various tasks from underwater inspection and maintenance of pipelines and cables to exploration, scientific research, and salvage operations.

In contrast, the second category includes Autonomous Underwater Vehicles (AUVs) [4]. AUVs, unlike ROUVs, operate autonomously and do not require continuous operator supervision. This autonomy gives them the ability to perform complex tasks independently. The development of AUV technology is strongly influenced by both civil and military applications. AUVs are becoming increasingly useful in scenarios where the use of manned vessels is either infeasible or impractical. In particular, in a military context, small AUVs equipped with specialized sensors constitute valuable tools for defensive and offensive operations in coastal zones [1]. The inherent autonomy of AUVs makes them well-suited for scenarios that require a high degree of flexibility, adaptability, and independence in underwater exploration and operations.

Accurate navigation, which is crucial for determining precise positions, is a major challenge for the effective deployment of Autonomous Underwater Vehicles (AUVs). One of the oldest methods of position determination, known for centuries, is dead reckoning [5]. This technique uses information about the ship's course, speed, and initial position to calculate its successive positions over time. However, the effectiveness of dead reckoning is hampered by the inherent imperfections of the measuring sensors used, compounded by the influence of external factors, leading to the accumulation of errors over time [6].

A critical aspect to consider is that the accuracy of the AUV's position decreases with increasing distance traveled. Consequently, there is an increased likelihood that the estimated position of the vehicle will deviate significantly from its actual position. Such discrepancies can pose significant challenges and potentially prevent the AUV from effectively performing its designated tasks. Consequently, there is an urgent need to develop and implement countermeasures to mitigate the impact of these error-prone phenomena on AUV navigation. Overcoming these challenges is critical to improving the reliability and effectiveness of AUV operations, particularly in scenarios where precise navigation is paramount.

Positioning errors can be reduced using radio navigation devices that mainly utilize satellites. Unfortunately, the high-frequency radio waves used for this purpose are very strongly suppressed by water. It is, therefore, necessary for the ship to be temporarily brought to the surface or for a special receiving antenna to be used. These devices are currently inexpensive and small, which means that they are fitted to most AUVs. The error is corrected to within a few meters due to the high accuracy of radio navigation systems. Once the error has been corrected, the vehicle will dive again. It will continue to navigate by dead reckoning.

In scenarios where the radio signal is either unavailable or deliberately jammed, making it difficult to reduce dead reckoning errors, an alternative positioning system is essential. Such a system should be able to provide accurate positioning using the onboard equipment of the AUV. The solution proposed in this paper is designed to enable the AUV to estimate its position by exploiting the surface environment without relying on radio navigation methods.

The proposed methodology involves the AUV temporarily surfacing to capture images of its surroundings. A process of semantic segmentation is then initiated, using a trained neural network to delineate and categorize objects within the captured images. This segmentation results in a comprehensive representation of the vehicle's environment. This representation is then transformed into a vector that encapsulates information about the height of the coastline (land) within a 360° area. The next stage involves a comparison of

this vector with the map stored in the AUV's memory, facilitating an estimate of the current position of the AUV.

By bypassing radio navigation and exploiting the inherent capabilities of onboard sensing and processing systems, this innovative solution has the potential to provide accurate AUV positioning even in radio-denied environments. The integration of image-based environmental analysis and neural network-based semantic segmentation underlines the adaptability and effectiveness of this approach for robust AUV navigation.

The solution was tested in real environmental conditions and compared with GNSS information. The research was carried out in the Gulf of Gdańsk and the Baltic Sea waters.

The contribution of the work is as follows:

1. Proposal of an algorithm for estimating the position of ships in coastal areas;
2. The algorithm is designed for GNSS and environments in which radio navigational signals are denied;
3. The algorithm's effectiveness has been verified under real-world conditions.

The paper is divided into six distinct sections, each of which contributes to the comprehensive exploration and presentation of the proposed solution. The initial section focuses on a review of relevant work in the field, providing a contextual background for the proposed solution. The subsequent section describes the architecture of the proposed solution, which includes an overview of the system's components, outlining how they interact and contribute to attaining the designated objectives. After that, authors explain the methodology used in the investigation and describe the step-by-step approach taken to investigate and validate the proposed solution. The Section 4 presents the outcomes of experiments conducted in a real marine environment. This includes a detailed analysis of the results, highlighting the performance and effectiveness of the proposed solution under real-world conditions. Future research identifies areas that could benefit from further exploration or refinement, providing a roadmap for ongoing investigations and improvements to the proposed solution. The final section encapsulates the paper's findings, summarizing the research's key insights, contributions, and implications. It links the various elements presented in the previous sections, offering a holistic perspective on the proposed solution and its significance.

2. Related Work

In principle, dead reckoning is a seemingly straightforward method of determining a vehicle's position. Knowing the initial position, speed, and direction, one should theoretically be able to determine the vehicle's position with pinpoint accuracy at any given time. Unfortunately, the practical application of dead reckoning is fraught with challenges due to the inherent inaccuracies of measurement sensors and the ever-present influence of external environmental factors, which introduce errors into the input data critical to calculating the vehicle's current position. The primary dead reckoning system used in AUVs typically revolves around the use of an Inertial Measurement Unit (IMU) [7–9]. This IMU consists of three rotating axes: an accelerometer, a gyroscope, and a magnetometer. This hardware configuration is complemented by specialized software that processes the raw signals from these devices to estimate the vehicle's key navigation parameters. Unfortunately, the effectiveness of this system is hampered by significant discrepancies in the accurate determination of navigation parameters due to internal noise and increased susceptibility to external interference. Consequently, this discrepancy culminates in a gradual accumulation of errors over time, making the dead reckoning approach unsustainable for sustained and accurate vehicle performance.

In study [10], the authors conduct a comparative analysis of professional-grade, high-rate, and low-cost IMU devices based on microelectromechanical systems (MEMS). The results show the good performance of these devices in determining pitch and roll, although with an observable increase in azimuth error over time. These challenges are compounded by the complex task of determining the correct speed of the AUV, particularly when dealing with the influence of sea currents, which the IMU fails to detect [11]. These

collective complexities underscore the intricacies associated with implementing dead reckoning in AUV navigation systems, requiring a nuanced consideration of sensor limitations and environmental factors to mitigate the accumulation of errors over time. The use of optical gyroscopes and a Doppler log will reduce the occurrence of this phenomenon. Their tactical performance is 10 times better than that of MEMS, according to the analysis presented in [12]. They can be successfully integrated into all types of unmanned vehicles due to their moderate price and compact size. Despite enormous technological progress and increasing accuracy of dead reckoning devices, error accumulation is still a problem. Additional systems are needed.

In waters where it is possible to build a suitable infrastructure, Long Baseline (LBL), Short Baseline (SBL), or Ultra-Short Baseline (USBL) acoustic reference positioning systems are used [13–15]. In [16], the authors present a solution that enables the estimation of the vehicle's position in the horizontal plane based on two known position acoustic transponders located in water. The distance from them is measured based on the time of the total signal propagation to and from the transponder. Results were obtained in which the maximum error in determining the AUV position did not exceed 10 m, and in most cases, it was about 2–3 m. Unfortunately, these systems can only be used in locations where there is a suitable infrastructure or surface vessel.

The source of signals can also constitute a surface vessel equipped with a precise DGPS positioning system. The system proposed by the authors of [17] uses a combination of an external USBL acoustic system with a Doppler velocity logo and a depth sensor. Such a combination of sensors enables a precise estimation of the AUV position in three planes. The system guarantees excellent results but requires the constant presence of a reference station near the vehicle.

The mitigation of navigation errors becomes an imperative consideration, especially during periods when the autonomous vehicle temporarily surfaces. Currently, the predominant approach involves the use of GNSS in the majority of unmanned vehicles [18–20]. This system generally provides exceptional accuracy, often to within a few meters. However, a notable drawback of this solution is its dependence on the continuous reception of satellite signals. The temporary disappearance or deliberate interruption of these signals poses a significant risk to the integrity of the AUV navigation system, resulting in inaccuracies that could lead to dangerous situations. It is important to emphasize the vulnerability of GNSS-based systems to signal interruption, which can occur for a variety of reasons, including atmospheric conditions or deliberate jamming. Such interruptions can lead to errors in the AUV navigation system, compromising the reliability of its position information. This inherent vulnerability is critical to ensuring the safety and effectiveness of unmanned vehicles operating in dynamic and potentially hostile environments. Furthermore, it is pertinent to highlight a fundamental limitation of GNSS systems—their inoperability below the water's surface. The GPS signal, which is an integral part of GNSS, is obstructed by water, making it unavailable for underwater navigation purposes. This underscores the need for alternative navigation methods and solutions that can be used during underwater operations and further highlights the multiple challenges associated with achieving accurate and robust navigation for autonomous underwater vehicles.

Currently, a commonly used and intensively developed technique of precise navigation is Simultaneous Localization and Mapping (SLAM) [21,22]. Based on the readings from the sensors, the displacement of characteristic points in relation to the vehicle is determined. Sonar [23,24], radar [25], LIDAR [26], acoustic sensors [27], and cameras [28] are used as sensors collecting information about the environment. On the basis of the collected information, a map is created and updated, against which the vehicle's position is simultaneously estimated. Data processing is done by a computing unit that uses advanced algorithms [29] with Bayesian [30] and Kalman [31] filters. However, SLAM requires constant monitoring of the vehicle's surroundings.

In [32], the authors used this technique combined with LIDAR to obtain information on the coastline and determine the surface vehicle's position in a seaport. The applied

technique made it possible to estimate vehicle motion parameters without using elements for dead reckoning. Unfortunately, publicly available laser devices have a range of up to several hundred meters and require waters with low undulation.

Similar imaging is possible with the use of radar systems. However, because of the large size of these devices, it is difficult to install them on small autonomous vehicles. In [33], the authors tested the radar used in autonomous cars in the marine environment. Unfortunately, the achieved range of the system (approximately 100 m) means that it can only affect collision avoidance. A similar situation occurred in [34], where the authors checked the capabilities of the micro-radar mounted on a UAV. The use of navigation radars with a range of several kilometers requires a significant increase in the AUV volume and the space for a sufficiently large antenna.

Vision systems are a common and relatively cheap tool enabling the registration of the vehicle's surroundings [35,36]. Information about objects in the near and distant surroundings can be obtained using the installed visible light or infrared camera. Due to the low cost and small size of the image recording devices, they are often used in all unmanned vehicles. Due to this, the operator can avoid obstacles and identify objects in the vicinity. A very popular solution is to use a double camera system (stereovision) [37,38], which allows the extraction of 3D information. In this approach, two horizontally shifted cameras provide information about relative depth, thanks to which we can determine the vehicle's position in relation to individual characteristic objects [39]. However, this solution requires placing the cameras at a certain distance from each other, which significantly increases the space occupied by the system.

The other solutions use a single camera. Low weight and dimensions make it possible to use for any vehicle, with the main disadvantage being the difficulty in determining the distance from the object. In [40], the author presents a method based on the results of horizontal angle measurements; however, the indication of characteristic points in an image with a known position requires analysis by a human. An AUV camera is also widely used for relative pose estimation and docking with self-similar landmarks [41] or with lights mounted around the entrance [42]. It can also be used for obstacle detection. The advantages of optical detectors in this field and tracking objects in maritime environments are described in [43].

The creation of an accurate autonomous positioning system based on one camera requires the use of appropriate algorithms to obtain as much information as possible from the input data [44–46]. One of the more significant methods is image segmentation. This is a computer vision technique that divides an image into segments and that has found application in many aspects of image processing. In [47], the authors proposed adaptive semantic segmentation for visual perception of water scenes. Adaptive filtering and progressive segmentation for shoreline detection are shown in [48].

The many advantages in navigation encouraged the authors of this paper to apply the above computer vision methods to solve the problem of AUV position estimation in surface position.

3. Architecture Overview

The solution outlined in this paper represents an innovative approach to determining the X,Y coordinates of an AUV through the integration of advanced computer vision and artificial intelligence methods. This advanced methodology is mainly applied in AUV navigating areas without radio navigation signals. During a temporary surface, the system captures images of the environment using a camera mounted on the vehicle. The semantic segmentation is then performed using a trained neural network, the purpose of which is to obtain information about the height of the land in a given bearing. The resulting output provides a comprehensive representation of the vehicle's environment, facilitating comparative analysis with a corresponding map-derived representation. The complex architecture of the proposed system is illustrated by the components shown in Figure 1:
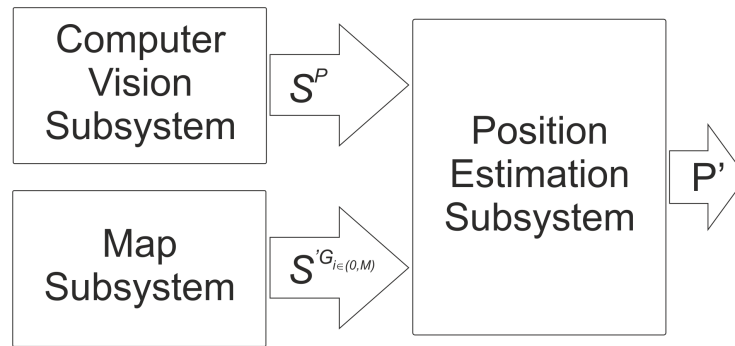
Figure 1. Block diagram of the proposed system.

Where:
S^P—the representation of the surroundings of the vehicle at point P
P—the point where the vehicle is actually located
S'^{G_i}—map representations of the points G_i, $i \in (0, M)$
M—the set of random points G_i
G_i—point on the map where the map representation S'^{G_i} is created
P'—estimated AUV position.

3.1. Computer Vision Subsystem

The task of the subsystem is to determine the visual representation of the vehicle surrounding S^P in point P, where the actual location of the vehicle is. A camera mounted on the vehicle is used for this purpose. The vehicle or camera is rotated $360°$ to provide full coverage of the surroundings. Actual camera angle information A_k^O is read from on-board systems and appended to each kth image, $k = 1 \ldots N$. Knowing this parameter and the horizontal angle of view α_h, the bearing (angle) can be determined for each image pixel column (Figure 2).
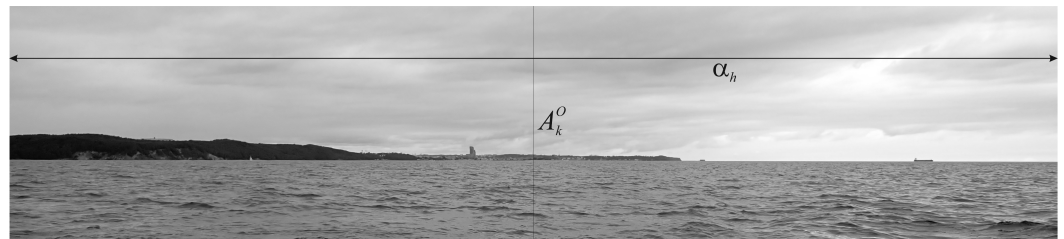

Figure 2. Actual camera angle A_k^O and angle of view α_h.

The recorded images are then segmented to determine where the land is located. This is done using a trained convolutional neural network (Figure 3).

Figure 3. Segmented image (ground truth): sky (white), land (black), sea (gray).

The information from all segmented N images is then used to create a representation of the vehicle's surroundings S^P. Formally, the representation $S^P = <S_0^P, S_1^P, \ldots, S_l^P, \ldots, S_t^P>$

(Figure 4) is the height of the land as seen from the point P, where S_l^P is the height of land in the direction $A_l^P = lR^\circ$, $l = 0 \ldots t$, $t = \frac{360}{R} - 1$, and R is the resolution of S^P.

Figure 4. The representation of the vehicle's surroundings S^P.

S^P is determined from the N images captured by the camera, of which the kth image is represented by (Figure 5):

$$S_k^O = < S_{k,0}^O, S_{k,1}^O, \ldots, S_{k,i}^O, \ldots, S_{k,(w-1)}^O > \tag{1}$$

where:

$S_{k,i}^O$—is height of land in kth image and ith image column;
A_k^O—is azimuth (angle) for kth image;
$A_{k,i}^O$—is azimuth (angle) for kth image and ith image column;
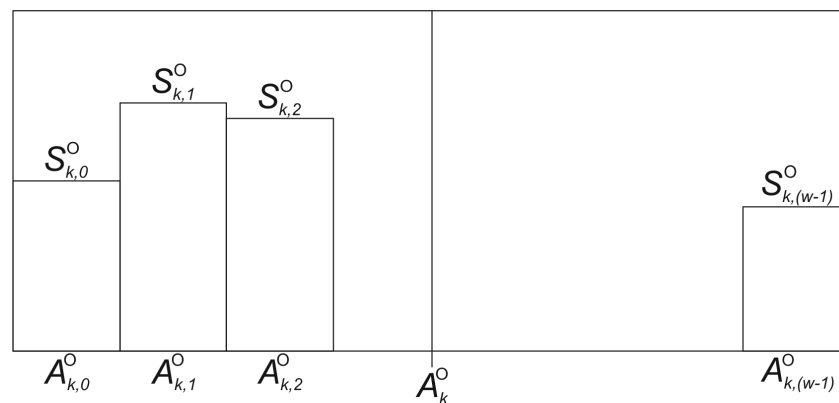w—determines the width of the image.

Figure 5. Land representation on the image S_k^O for the kth image.

Finally, S^P is calculated as follows:

$$S_l^P = \frac{S_{k^*,j^*}^O}{S_{max}^O} \tag{2}$$

where:

$S_{max}^O = \max_{k,j} S_{k,j}^O$
$k^* = \operatorname{argmin}_k |A_l^P - A_k^O|$—is the image that was selected to calculate S_l^P
$j^* = \operatorname{argmin}_j |A_l^P - A_{k^*,j}^O|$—is the column of k^* image, which was selected to calculate S_l^P.

3.2. Map Subsystem

The primary objective of the subsystem is to generate a map representation S'^G corresponding to any given point G_i in the area covered by the terrain model. This representation is formulated by integrating a stored topographic map, a digital terrain model, or, in particular, a digital surface model. Of these options, the digital surface model emerges as the most accurate representation of the terrain, as it not only encapsulates the topographical features, but also incorporates elevation information relating to objects on the ground, such as trees and buildings, thereby increasing the overall accuracy of the representation. The construction process S'^{G_i} requires the application of appropriate transformations, taking

into account the Earth's curvature and the objects' mutual occlusion, as shown in Figure 6. This complex process ensures a faithful representation of the terrain that is consistent with the topography of the real world. A critical consideration in this endeavor is the phenomenon of mutual occlusion, where objects occlude each other as the observer moves away from them. The size of the visible portion (h_{vis}) of an object is mainly a function of the elevation of the camera and the distance from the object, as described by Equation (3). This mathematical relationship emphasizes the dynamic nature of the visible part of an object as a function of both camera placement and spatial proximity. The systematic consideration of these factors is essential for the accurate and comprehensive generation of the map representation S', which reflects the complex interplay between terrain features and observation parameters.
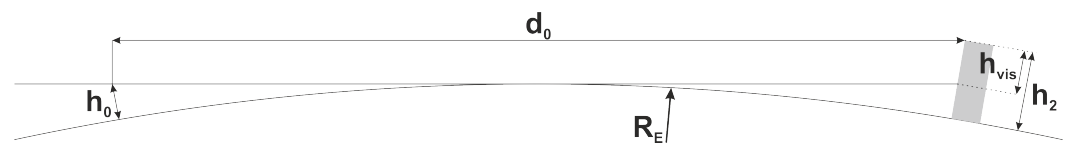

Figure 6. Height of an object beyond the horizon [49].

$$h_{vis} = h_2 - \left(\sqrt{(d_0 - \sqrt{h_0(h_0 + 2R_E)})^2 + R_E^2} - R_E\right) \tag{3}$$

where:

h_0—height of observer (camera)
h_{vis}—height of an object visible above horizon
h_2—height of an object
d_0—distance from the observer to object
R_E—radius of earth.

In addition, there are cases where the apparent height of an object in the foreground is less than that of a more distant object, but the former still blocks the view of anything behind it. This scenario is particularly relevant in the context of map creation, where the goal is to capture information about objects that occupy the largest number of pixels in the image, prioritizing visibility over physical height. Figure 7 presents a visibility analysis carried out along a specified line of approximately 4 km, accompanied by its corresponding cross section or elevation profile. This analytical display illustrates the nuanced interplay between foreground and background objects and highlights the critical nature of prioritizing pixel coverage in the display process. The intricacies of visibility, as delineated in the cross section, underscore the need for a sophisticated approach to map representation that goes beyond mere physical height considerations to ensure a more comprehensive and contextually relevant representation of the observed terrain.

Formally, a map representation for point G_i looks as follows:

$$S'^{G_i} = <S_0'^{G_i}, S_1'^{G_i}, \ldots, S_t'^{G_i}>, S_k'^{G} = h_{vis,w}$$

$$k = 0, 1, \ldots, t \tag{4}$$

$$w = \begin{cases} \text{argmax}_i \frac{h_{vis,i}}{x_i} & \iff \exists_{i \in B_k} h_{vis,i} > 0 \\ 0 \text{ otherwise} \end{cases} \tag{5}$$

where:

B_k—a numbered sequence of pixels in direction kR from point G_i
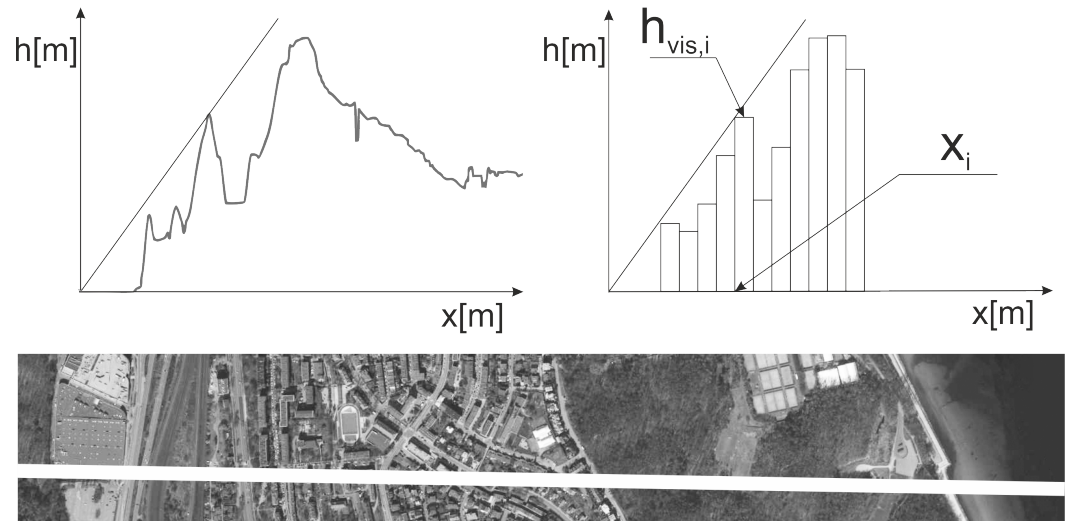x_i—distance from point G_i to the ith pixel.

Figure 7. Example determination of h_{vis}: **bottom**—white line is vision for a given h_{vis}; **top left**—continuous height profile, **top right**—discrete height profile; where: $x[m]$—distance from point G_i, $h[m]$—height of an object.

3.3. Position Estimation Subsystem

The task of the subsystem is to determine the position estimated at point P. This is done by comparing the S^P representation with the S'^G map representations for the set M of random points G_i located in a circle with radius r and center at point Q, which is the estimated position of the submersible vehicle determined by the dead reckoning navigation system. The position of point P corresponds to the position of point G_i, whose representation S'^{G_i} is more similar to the representation of S^P with respect to the selected measure of similarity T. Formally, the estimated position P' for the point P is defined as follows:

$$P' = G_k \tag{6}$$

$$k = \operatorname*{argmax}_{i \in 0...|M|} T(S^P, S'^{G_i}) \tag{7}$$

where:

T is a measure of the similarity of representations S^P and S'^G.

4. Experiments

4.1. Dataset

To validate the methodology proposed in this paper, the authors conducted testing using real-world image data. The research began with the assembly of an image database. In order to compute the map representation, S'^G, a prerequisite is the use of a neural network capable of delineating the location of land within an image through semantic segmentation. To facilitate this, a corpus of 3000 images was collected, covering a range of weather and lighting conditions, as shown in Figure 8. These images were captured in various locations, including different areas of the Baltic Sea and inland waters, to ensure a diverse and representative dataset for robust testing.

A crucial phase of the experiments involved image labeling. The authors initially used an algorithm, as previously proposed in [48], for an automated labeling process; the inherent imperfections of the algorithm required manual improvement for some images. Consequently, each image destined for neural network training was segmented into distinct segments, namely sea (I_1), land (I_2), and sky (I_3). This labor-intensive labeling process ensures the accuracy and reliability of the training dataset and lays the groundwork for the subsequent stages of verification of the proposed method.

Figure 8. Sample images from the training set.

In addition, to increase the robustness and diversity of the dataset, all the recorded images, along with their carefully labeled counterparts, underwent an augmentation process. Several methods available in the Pyplot library were used for this purpose, including horizontal and vertical shift, horizontal and vertical flip, random brightness adjustment, rotation, and zoom. The augmentation process significantly expanded the dataset used to train the neural networks, contributing to the models' adaptability to different environmental conditions and scenarios. The dataset was then split into training and validation sets at a ratio of 9:1. Based on the findings of an extensive literature review, four widely accepted neural network architectures [50,51] were selected for experimentation: FCN-32, Segnet, Unet, and PSPNet.

In the next phase, the selected neural networks were trained using the TensorFlow and Keras libraries. The training and validation accuracies of all selected models are shown in Table 1. The best results were achieved by Segnet. To evaluate the effectiveness of the trained models, an additional test set of 150 labeled images taken in different regions of the Baltic Sea was used. Each image of the test set was carefully compared to the ground truth, as shown in Figure 9. This comparative analysis facilitated the quantification of correctly identified pixels, which served as a metric for evaluating the segmentation accuracy of the neural networks. After careful evaluation, the Segnet architecture emerged as the most effective, achieving the highest accuracy in image segmentation. As a result, Segnet was selected as the designated image segmentation neural network for the subsequent stages of the research, demonstrating its effectiveness for the intended application.

Table 1. The training and validation accuracies.

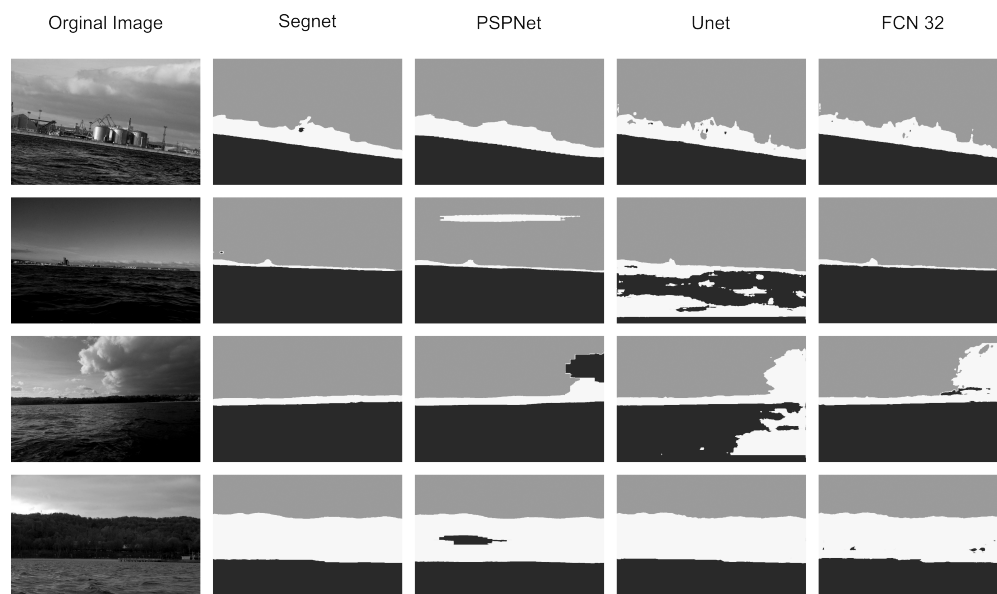| Model	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
FCN-32	1.0985	0.3398	1.0984	0.3431
PSPNet	0.1685	0.9480	0.6678	0.8120
Unet	0.1355	0.9595	0.3672	0.8847
Segnet	0.0614	0.9802	0.0833	0.9694

Figure 9. Comparison of the effectiveness of the tested neural networks.

4.2. Representation of the Surroundings

In the next step, in order to verify the accuracy of the proposed method, 2046 images were registered at 100 measurement points *P* on the Baltic Sea and in Gdańsk Bay (Figure 10). Information on the position and azimuth of the camera's optical axis was attached to each shot.
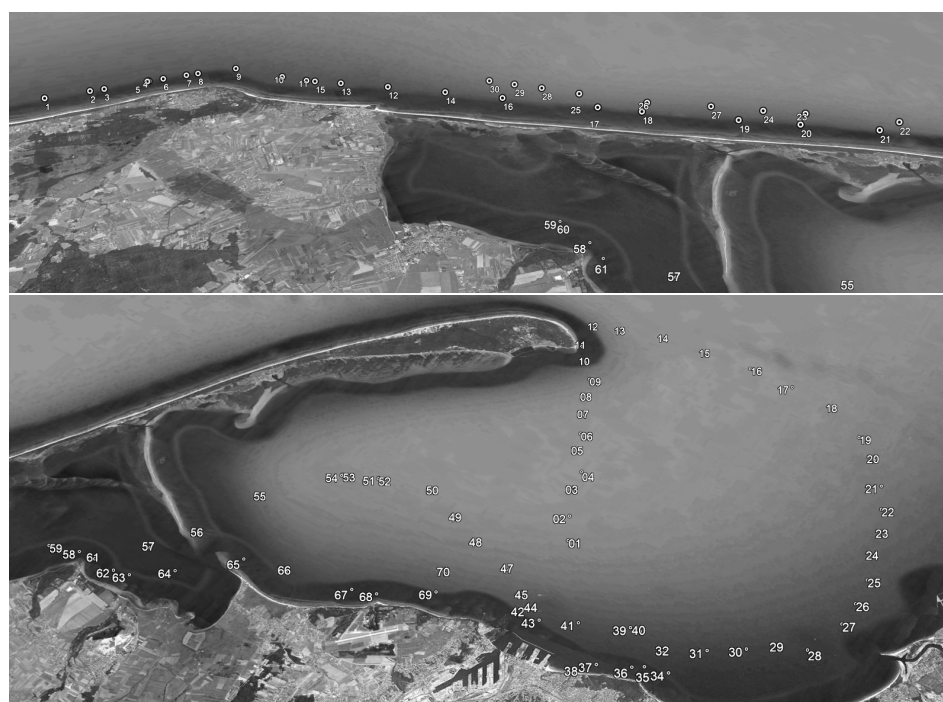

Figure 10. Measurement points on the open sea (**top**) and in the bay (**bottom**).

An important parameter influencing the nature of the image is the height of the lens above the water level. In order to simulate the camera of a surfaced underwater vehicle, all images were taken at a constant height of 1 m above sea level. An Olympus TG-6 camera equipped with a GNSS receiver and a digital compass was used to record the images. During the recording, the focal length of the lens was set to a value equivalent to 25 mm for a full frame matrix, thus obtaining an $\alpha_h = 71.52°$. Therefore, full 360° imaging could

be obtained by registering a minimum of $N = 6$ images. In practice, this number ranged from 8 to 16, which made it possible to counteract the distortion. The resolution of the representation S^P was set to $R = 0.1°$, so $S^P = < S_0^P, S_1^P, \ldots, S_{3600}^P >$.

The 2046 images recorded under the conditions specified above were used to create a representation S^P for each measurement point P.

4.3. Map of the Surroundings

The generation of an M set of points requires reference to the map of the area where the AUV performs its tasks. The more precise the map, the more accurate the representations of the points in the set M. A Digital Surface Model (DSM) was used as the map, on the basis of which the comparison points were generated. It is currently the most accurate representation of the Earth's surface along with the objects located on its surface. DSM is available as text files and contains the coordinates (X, Y, Z) of points, with a resolution of 0.5 m for urban areas and 1 m for other areas. The heights are generated with an accuracy of 0.15 m and, in some areas, up to 0.1 m. The data obtained cover the Polish area of the Gulf of Gdańsk, together with the Hel Peninsula, which fully coincides with the area where the research was carried out. However, the drawback of this representation is the size of the files, which is about 40 GB for the area shown in Figure 11. Therefore, it was necessary to adapt the data derived from DSM to the proposed method. To this end, numeric data were converted to a bitmap. Knowing that the maximum height is less than 200 m above sea level, a compressed DSM took the form of an 8-bit grayscale image. The graphic file in tiff format had a resolution of 39,135 × 27,268 pixels and covered an area of approximately 85 × 55 km, due to which an accuracy of about 2 m/pixel was obtained. This made it possible to reduce the image size to 1.2 GB, which significantly influenced the speed of generating points.

Figure 11. DSM after converting to bitmap (**left**: Gdynia city, **right**—Gdańsk Bay).

4.4. A Measure of the Similarity of Representations

The estimation of the AUV position at point P requires a comparison of the S^P representation with the S'^{G_i} map representations. Two common methods were used to determine the measure of similarity $T(S^P, S'^{G_i})$: the inverse of Euclidean distance (Equation (8)) and the Pearson correlation coefficient (Equation (9)).

$$T^E(S, S') = \frac{1}{1 + d(S, S')}$$

$$T^E(S, S') = \frac{1}{1 + \sqrt{\sum_{n=1}^{t}((S_i^P - S_i'^G)^2)}} \tag{8}$$

$$T^C(S, S') = d(S, S')$$

$$T^C(S, S') = \frac{\sum_{i=1}^{n}(S_i^P - \overline{S^P})(S_i'^G - \overline{S'^G})}{\sqrt{\sum_{i=1}^{n}(S_i^P - \overline{S^P})^2 \sum_{i=1}^{n}(S_i'^G - \overline{S'^G})^2}} \tag{9}$$

For each measurement point P, the algorithm generated the set M, consisting of 500 random points G_i located in a circle with radius r = 1 km and center at point P. For each point G_i, a map representation S'^{G_i} was generated using compressed DSM, and then the measure of similarity $T(S^P, S'^{G_i})$ was computed using Equations (8) and (9). In the final step, the estimated position P' for each point P was calculated according to Equation (6).

5. Experimental Results

In order to verify the effectiveness of the proposed algorithm, the errors (max, min, and average) in the calculation of the estimated position P' for each point P were applied. The errors are given in Table 2.

As it turned out, errors in the estimation of the position P for Equation (9) were 102.40 m (max), 8.93 m (min), and 64.60 m (avg). For Equation (8), the results were significantly worse and were 189.25 m (max), 10.43 m (min), and 91.6 m (avg). Pearson's correlation, as a measure of the linear relationship between variables, showed greater sensitivity to underlying patterns. This is particularly important for data where the relationships between variables are not strictly linear and may show non-linear trends. Unlike Euclidean distance, which is sensitive to the size and scale of the variables, Pearson's correlation takes into account relative differences, providing a more robust measure of similarity. The prevalence of Pearson correlation in our experiments suggests that it is more appropriate for our dataset.

Table 2. Comparing Euclidean distances and correlations.

Error	Min [m]	Max [m]	Average [m]
Euclidean	10.43	189.25	91.62
Correlation	8.93	102.40	64.60

After the analysis of the spatial distribution of the points G_i, it turned out that the points with the highest similarity measure $T^C(S, S')$ were generally located in the closest neighborhood to the point P. The distribution of the correlation coefficient for four sample points P is shown in Figure 12 (left column). The authors thus decided to estimate P' based on the five G_i points with the highest $T^C(S, S')$ according to the following equation:

$$P' = \frac{\sum_{i=1}^{L} T^C(S^P, S'^{G_i}) G_i}{\sum_{i=1}^{L} T^C(S^P, S'^{G_i})} \tag{10}$$

where:

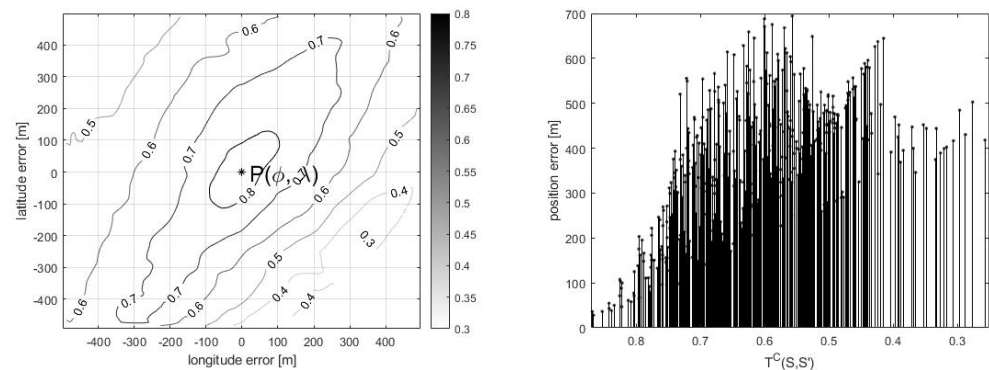L—a set of points G_i with the highest value of T^C.
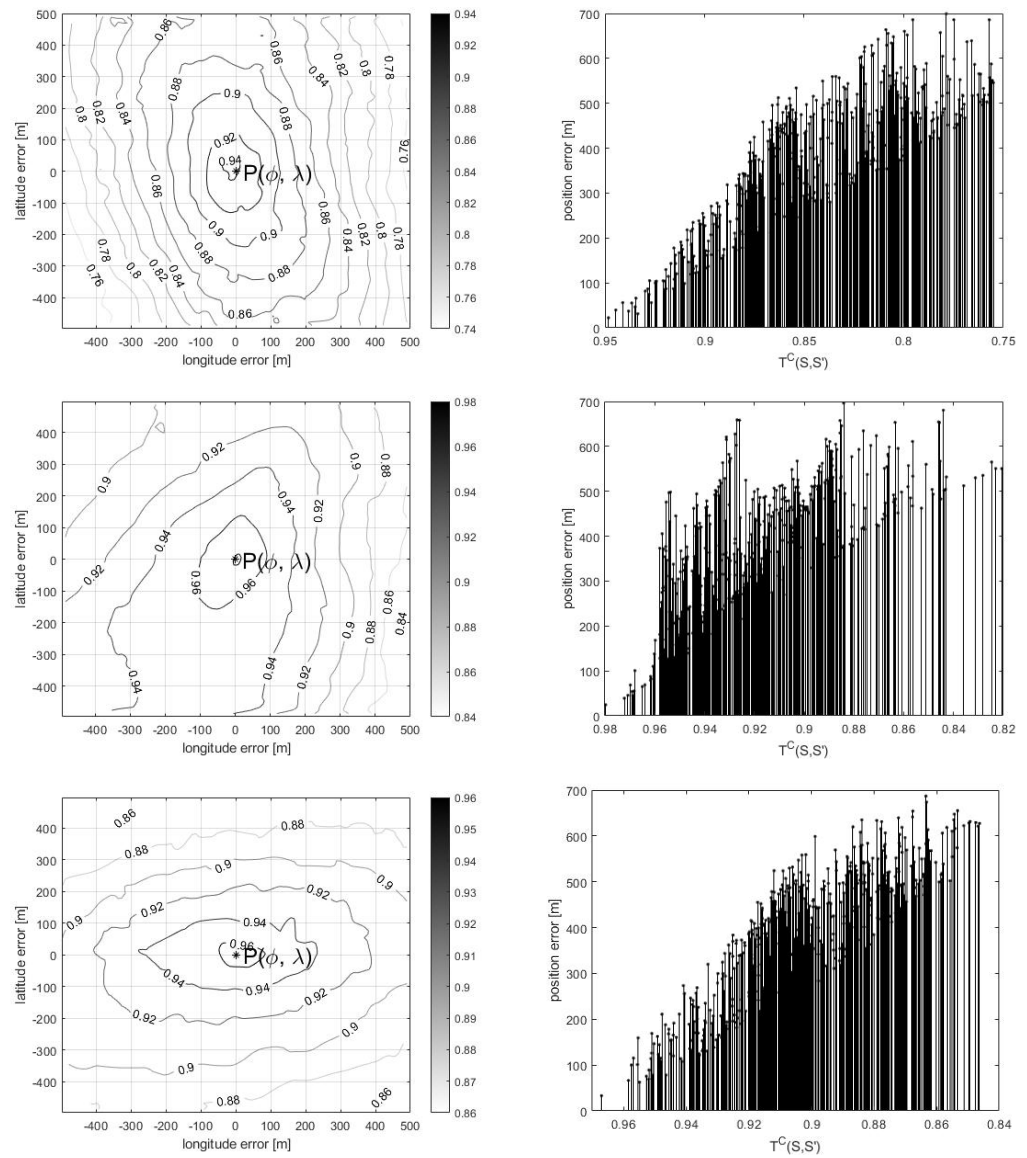

Figure 12. *Cont.*

Figure 12. Distribution (**left**—spatial, **right**—quantitative) of the correlation coefficient for four sample points P.

As it turned out, this procedure resulted in a significant reduction in the estimation error, as shown in Table 3.

Table 3. Comparing 1 and 5 points with the highest similarity measure.

Error	Min [m]	Max [m]	Average [m]
1 point	8.93	102.40	64.60
5 points	4.37	64.59	34.78

During the registration of the images, there were situations where the sea conditions (waves) or the appearance of surface vessels in the field of view prevented the correct generation of the representation S^P. Therefore, the authors decided to verify the effectiveness of the incomplete environment representation algorithm S^P. To this end, 10%, 30%, or 50% (in reality, it is very unlikely for the disruptions to exceed 30%) of items, ranging from the nth to the mth, were removed from $S^P = <S_0^P, S_1^P, \ldots, S_n^P, \ldots, S_m^P, \ldots, S_t^P>$. The resulting representation S_r^P after the removal procedure is given by Equation (11). The verification

of the algorithm for incomplete representation S_r^P was carried out for 20 random points P, and the results are presented in Table 4.

$$S_r^P = S^P \backslash < S_n^P, S_{n+1}^P, \dots, S_m^P > \tag{11}$$

where:

$m \in \{n + 360{:}10\%, n + 1080{:}30\%, n + 1800{:}50\%\}$
$n \in \{180, 720, 1800\}.$

Table 4. Comparing representations of incomplete surroundings.

Removed Part	n	Min [m]	Max [m]	Average [m]
10%	180	5.21	54.19	40.87
10%	720	7.87	74.02	29.08
10%	1800	4.17	62.94	47.78
30%	180	8.03	81.09	46.26
30%	720	14.01	101.29	69.01
30%	1800	10.65	57.94	36.28
50%	180	42.39	179.03	130.97
50%	720	38.67	237.14	158.30
50%	1800	49.49	213.59	143.00

The obtained results show that incomplete display around the vehicle does not make it impossible to estimate vehicle position. For disruptions equal to 10% and 30% of original S^P, the highest errors amounted to 47.78 m/69.01 m (average), 74.02 m/101.29 m (max), and 7.87 m/8.03 m (min), which is an even better result than for the undisturbed representation S^P. However, it should be noted that the results given in Table 3 are calculated for only 20 measurement points compared to the previous case in which all 100 points were considered. The noticeable deterioration of the results is only visible for 50% of the disturbances. In this case, the highest errors amounted to 158.30 m (average), 237.14 m (max), and 49.49 m (min).

6. Future Research

The exploration of alternative terrain models is a promising avenue for further research. While DSM provides high accuracy in object elevation information, they are limited to specific geographic regions. In cases where DSM data are not available, alternative models as topographic maps or satellite imagery could be used. However, this will require adaptations to the existing algorithm to enable it to integrate and utilize different map information from different sources.

A second area of research is the development of an algorithm tailored to regions with recognizable landmarks in the vicinity of the vehicle. For this purpose, a terrestrial navigation approach can be used by training neural networks to classify objects such as forests, beaches, lighthouses, buoys, etc., and then using an algorithm to locate these objects on a map. This approach is promising in environments where identifiable landmarks can serve as navigational reference points.

The third area of interest revolves around exploring the size of objects in the image. By identifying objects of known size and estimating their distance based on the disparity between their actual size and their size in pixels, a methodology for distance estimation can be established. Integrating this distance information with object classification data further paves the way for an effective terrestrial navigation approach.

A fourth and equally important area for future research is the development of algorithms capable of detecting image clutter such as waves, vessels, and other potential sources of interference. The identification and mitigation of such clutter is critical to ensuring the accuracy of vehicle position calculations, particularly in challenging marine environments where external factors can impede accurate navigation.

7. Conclusions

This paper presents a new algorithm for estimating the position of an Autonomous Underwater Vehicle (AUV) when it operates on the surface under the condition that the shore is visible. The algorithm's core methodology involves comparing the real-time representation of the vehicle's environment, captured by a camera, and the map representation derived from a DSM.

The main results of the tests carried out, as reported in the paper, are as follows:

1. Convolutional neural networks accurately extract land features from marine imagery. This highlights the effectiveness of advanced machine learning techniques in the field of maritime image analysis.
2. The proposed algorithm reduces the error associated with dead reckoning navigation systems to 30–60 m. This indicates the algorithm's effectiveness in improving the accuracy of AUV navigation, particularly in challenging maritime environments.
3. The resilience of the algorithm is highlighted by its ability to operate effectively even in scenarios where land representation is incomplete. This adaptability is crucial for real-world applications where environmental conditions may limit the availability of comprehensive map data.
4. The proposed algorithm proves to be a robust means for developing a fully autonomous AUV navigation system. It shows particular promise in environments where access to GNSS signals is limited, positioning it as a viable solution for GNSS-denied scenarios. This capability opens up avenues for autonomous task performance in challenging maritime conditions, contributing to the advancement of AUV technology.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, [J.Z.], upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AUV	Autonomous Underwater Vehicle
DGPS	Differential Global Positioning System
DSM	Digital Surface Model
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
IMU	Inertial Measurement Unit
MEMS	Microelectromechanical System
LBL	Long Baseline
LIDAR	Light Detection and Ranging
ROUV	Remotely Operated Underwater Vehicle
SBL	Short Baseline
SCU	Surface Control Unit
SLAM	Simultaneous Localization and Mapping
UAV	Unmanned Aerial Vehicle
USBL	Ultra Short Baseline
UUV	Unmanned Underwater Vehicle

References

1. Sahoo, A.; Dwivedy, S.K.; Robi, P. Advancements in the field of autonomous underwater vehicle. *Ocean Eng.* **2019**, *181*, 145–160. [CrossRef]
2. Bovio, E.; Cecchi, D.; Baralli, F. Autonomous underwater vehicles for scientific and naval operations. *Annu. Rev. Control* **2006**, *30*, 117–130. [CrossRef]
3. Joochim, C.; Phadungthin, R.; Srikitsuwan, S. Design and development of a Remotely Operated Underwater Vehicle. In Proceedings of the 2015 16th International Conference on Research and Education in Mechatronics (REM), Bochum, Germany, 18–20 November 2015; pp. 148–153. [CrossRef]
4. He, Y.; Wang, D.B.; Ali, Z.A. A review of different designs and control models of remotely operated underwater vehicle. *Meas. Control* **2020**, *53*, 1561–1570. [CrossRef]
5. Maurelli, F.; Krupiński, S.; Xiang, X.; Petillot, Y. *AUV Localisation: A Review of Passive and Active Techniques*; Springer: Berlin/Heidelberg, Germany, 2021. [CrossRef]
6. Xie, Y.x.; Liu, J.; Hu, C.q.; Cui, J.h.; Xu, H. AUV Dead-Reckoning Navigation Based on Neural Network Using a Single Accelerometer. In Proceedings of the 11th ACM International Conference on Underwater Networks & Systems, Shanghai, China, 24–26 October 2016. [CrossRef]
7. Brossard, M.; Barrau, A.; Bonnabel, S. AI-IMU Dead-Reckoning. *IEEE Trans. Intell. Veh.* **2020**, *5*, 585–595. [CrossRef]
8. Rogne, R.H.; Bryne, T.H.; Fossen, T.I.; Johansen, T.A. MEMS-based Inertial Navigation on Dynamically Positioned Ships: Dead Reckoning. *IFAC-PapersOnLine* **2016**, *49*, 139–146. [CrossRef]
9. Chu, Z.; Zhu, D.; Sun, B.; Nie, J.; Xue, D. Design of a dead reckoning based motion control system for small autonomous underwater vehicle. In Proceedings of the 2015 IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE), Halifax, NS, Canada, 3–6 May 2015; pp. 728–733. [CrossRef]
10. De Agostino, M.; Manzino, A.M.; Piras, M. Performances comparison of different MEMS-based IMUs. In Proceedings of the IEEE/ION Position, Location and Navigation Symposium, Indian Wells, CA, USA, 4–6 May 2010; pp. 187–201. [CrossRef]
11. Vitale, G.; D'Alessandro, A.; Costanza, A.; Fagiolini, A. Low-cost underwater navigation systems by multi-pressure measurements and AHRS data. In Proceedings of the OCEANS 2017, Aberdeen, UK, 19–22 June 2017; pp. 1–5. [CrossRef]
12. Yoon, S.; Park, U.; Rhim, J.; Yang, S.S. Tactical grade MEMS vibrating ring gyroscope with high shock reliability. *Microelectron. Eng.* **2015**, *142*, 22–29. [CrossRef]
13. Vickery, K. Acoustic positioning systems. A practical overview of current systems. In Proceedings of the 1998 Workshop on Autonomous Underwater Vehicles (Cat. No.98CH36290), Cambridge, MA, USA, 21 August 1998; pp. 5–17. [CrossRef]
14. Thomson, D.J.M.; Dosso, S.E.; Barclay, D.R. Modeling AUV Localization Error in a Long Baseline Acoustic Positioning System. *IEEE J. Ocean. Eng.* **2018**, *43*, 955–968. [CrossRef]
15. Zhang, J.; Han, Y.; Zheng, C.; Sun, D. Underwater target localization using long baseline positioning system. *Appl. Acoust.* **2016**, *111*, 129–134. [CrossRef]
16. Matos, A.; Cruz, N.; Martins, A.; Lobo Pereira, F. Development and implementation of a low-cost LBL navigation system for an AUV. In Proceedings of the Oceans '99. MTS/IEEE. Riding the Crest into the 21st Century. Conference and Exhibition. Conference Proceedings (IEEE Cat. No.99CH37008), Seattle, WA, USA, 13–16 September 1999; Volume 2, pp. 774–779. [CrossRef]
17. Mandt, M.; Gade, K.; Jalving, B. Integrating DGPS-USBL position measurements with inertial navigation in the HUGIN 3000 AUV. In Proceedings of the 8th Saint Petersburg International Conference on Integrated Navigation Systems, Saint Petersburg, Russia, 28–30 May 2001; pp. 28–30.
18. Hegarty, C.J.; Chatre, E. Evolution of the Global Navigation SatelliteSystem (GNSS). *Proc. IEEE* **2008**, *96*, 1902–1917. [CrossRef]
19. Yang, Y.; Gao, W.; Guo, S.; Mao, Y.; Yang, Y. Introduction to BeiDou-3 navigation satellite system. *Navigation* **2019**, *66*, 7–18. [CrossRef]
20. Giorgi, G.; Teunissen, P.J.G.; Gourlay, T.P. Instantaneous Global Navigation Satellite System (GNSS)-Based Attitude Determination for Maritime Applications. *IEEE J. Ocean. Eng.* **2012**, *37*, 348–362. [CrossRef]
21. Marchel, L.; Naus, K.; Specht, M. Optimisation of the Position of Navigational Aids for the Purposes of SLAM technology for Accuracy of Vessel Positioning. *J. Navig.* **2020**, *73*, 282–295. [CrossRef]
22. Karkus, P.; Cai, S.; Hsu, D. Differentiable SLAM-Net: Learning Particle SLAM for Visual Navigation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 2815–2825.
23. Fallon, M.F.; Folkesson, J.; McClelland, H.; Leonard, J.J. Relocating Underwater Features Autonomously Using Sonar-Based SLAM. *IEEE J. Ocean. Eng.* **2013**, *38*, 500–513. [CrossRef]
24. Siantidis, K. Side scan sonar based onboard SLAM system for autonomous underwater vehicles. In Proceedings of the 2016 IEEE/OES Autonomous Underwater Vehicles (AUV), Tokyo, Japan, 6–9 November 2016; pp. 195–200. [CrossRef]
25. Han, J.; Cho, Y.; Kim, J. Coastal SLAM with Marine Radar for USV Operation in GPS-Restricted Situations. *IEEE J. Ocean. Eng.* **2019**, *44*, 300–309. [CrossRef]
26. Ueland, E.S.; Skjetne, R.; Dahl, A.R. Marine Autonomous Exploration Using a Lidar and SLAM, Volume 6: Ocean Space Utilization. In Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering, Trondheim, Norway, 25–30 June 2017. [CrossRef]
27. Hidalgo, F.; Braunl, T. Review of underwater SLAM techniques. In Proceedings of the 2015 6th International Conference on Automation, Robotics and Applications (ICARA), Queenstown, New Zealand, 17–19 February 2015; pp. 306–311. [CrossRef]

28. Jung, J.; Lee, Y.; Kim, D.; Lee, D.; Myung, H.; Choi, H.T. AUV SLAM using forward/downward looking cameras and artificial landmarks. In Proceedings of the 2017 IEEE Underwater Technology (UT), Busan, Republic of Korea, 21–24 February 2017; pp. 1–3. [CrossRef]

29. Taketomi, T.; Uchiyama, H.; Ikeda, S. Visual SLAM algorithms: A survey from 2010 to 2016. *IPSJ Trans. Comput. Vis. Appl.* **2017**, *9*, 16. [CrossRef]

30. Blanco, J.L.; FernÁndez-Madrigal, J.A.; GonzÁlez, J. Toward a Unified Bayesian Approach to Hybrid Metric–Topological SLAM. *IEEE Trans. Robot.* **2008**, *24*, 259–270. [CrossRef]

31. Huang, S.; Dissanayake, G. Convergence and Consistency Analysis for Extended Kalman Filter Based SLAM. *IEEE Trans. Robot.* **2007**, *23*, 1036–1049. [CrossRef]

32. Naus, K.; Marchel, L. Use of a Weighted ICP Algorithm to Precisely Determine USV Movement Parameters. *Appl. Sci.* **2019**, *9*, 3530. [CrossRef]

33. Stateczny, A.; Kazimierski, W.; Burdziakowski, P.; Motyl, W.; Wisniewska, M. Shore Construction Detection by Automotive Radar for the Needs of Autonomous Surface Vehicle Navigation. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 80. [CrossRef]

34. Scannapieco, A.; Renga, A.; Fasano, G.; Moccia, A. Ultralight radar for small and micro-uav navigation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *42*, 333. [CrossRef]

35. Wolf, M.T.; Assad, C.; Kuwata, Y.; Howard, A.; Aghazarian, H.; Zhu, D.; Lu, T.; Trebi-Ollennu, A.; Huntsberger, T. 360-degree visual detection and target tracking on an autonomous surface vehicle. *J. Field Robot.* **2010**, *27*, 819–833. [CrossRef]

36. Stateczny, A.; Gierski, W. The concept of anti-collision system of autonomous surface vehicle. *E3S Web Conf.* **2018**, *63*, 00012. [CrossRef]

37. Huntsberger, T.; Aghazarian, H.; Howard, A.; Trotz, D.C. Stereo vision–based navigation for autonomous surface vessels. *J. Field Robot.* **2011**, *28*, 3–18. [CrossRef]

38. Hożyń, S.; Zak, B. Stereo Vision System for Vision-Based Control of Inspection-Class ROVs. *Remote Sens.* **2021**, *13*, 5075. [CrossRef]

39. Sinisterra, A.J.; Dhanak, M.R.; Von Ellenrieder, K. Stereovision-based target tracking system for USV operations. *Ocean Eng.* **2017**, *133*, 197–214. [CrossRef]

40. Naus, K. Accuracy in fixing ship's positions by CCD camera survey of horizontal angles. *Geomat. Environ. Eng.* **2011**, *5*, 47–61.

41. Negre, A.; Pradalier, C.; Dunbabin, M. Robust vision-based underwater homing using self-similar landmarks. *J. Field Robot.* **2008**, *25*, 360–377. [CrossRef]

42. Park, J.Y.; huan Jun, B.; mook Lee, P.; Oh, J. Experiments on vision guided docking of an autonomous underwater vehicle using one camera. *Ocean Eng.* **2009**, *36*, 48–61. [CrossRef]

43. Prasad, D.K.; Rajan, D.; Rachmawati, L.; Rajabally, E.; Quek, C. Video Processing From Electro-Optical Sensors for Object Detection and Tracking in a Maritime Environment: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 1993–2016. [CrossRef]

44. Lu, Y.; Song, D. Visual Navigation Using Heterogeneous Landmarks and Unsupervised Geometric Constraints. *IEEE Trans. Robot.* **2015**, *31*, 736–749. [CrossRef]

45. Mu, X.; He, B.; Zhang, X.; Yan, T.; Chen, X.; Dong, R. Visual Navigation Features Selection Algorithm Based on Instance Segmentation in Dynamic Environment. *IEEE Access* **2020**, *8*, 465–473. [CrossRef]

46. Moskalenko, V.; Moskalenko, A.; Korobov, A.; Semashko, V. The Model and Training Algorithm of Compact Drone Autonomous Visual Navigation System. *Data* **2019**, *4*, 4. [CrossRef]

47. Zhan, W.; Xiao, C.; Wen, Y.; Zhou, C.; Yuan, H.; Xiu, S.; Zou, X.; Xie, C.; Li, Q. Adaptive Semantic Segmentation for Unmanned Surface Vehicle Navigation. *Electronics* **2020**, *9*, 213. [CrossRef]

48. Hozyn, S.; Zalewski, J. Shoreline Detection and Land Segmentation for Autonomous Surface Vehicle Navigation with the Use of an Optical System. *Sensors* **2020**, *20*, 2799. [CrossRef] [PubMed]

49. Commons, W. File: Calculating How Much of a Distant Object Is Visible above the Horizon.jpg—Wikimedia Commons, the Free Media Repository. 2022. Available online: https://commons.wikimedia.org/wiki/File:Calculating_How_Much_of_a_Distant_Object_is_Visible_Above_the_Horizon.jpg (accessed on 26 July 2022).

50. Sultana, F.; Sufian, A.; Dutta, P. Evolution of Image Segmentation using Deep Convolutional Neural Network: A Survey. *Knowl. Based Syst.* **2020**, *201–202*, 106062. [CrossRef]

51. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image Segmentation Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 74–80. [CrossRef]