A Stereovision-based Navigation System for Autonomous Vessels

Ir. M Hepworth*, Ir. V Garofano, Dr. Y Pang, Dr. V Reppa

Department of Maritime and Transport Technology, Delft University of Technology
*Corresponding author. Email: m.hepworth-1@tudelft.nl

Synopsis

The autonomous inland vessel has the potential to enable a hinterland modality shift to the waterway, however economically-conscious solutions to the technological challenges are lacking. This paper explores the potential role of stereovision technology as a primary sensor set within the application of autonomous inland vessels. The design of a multi-camera navigation system for mid-range perception and localisation tasks is presented with a view to providing sufficient data for collision avoidance tasks. The performance of these sensors in the application is evaluated with experimental testing using small-scale vessels in an indoor tank, demonstrating their capabilities and limitations for consideration in ongoing research.

Keywords: Autonomous surface vehicle, Robot Navigation, Programming and Vision, Neural networks Multi sensor systems, Perception and sensing.

1 Introduction

Urban freight transportation (UFT) requires smaller vessels, specifically those under 300 tonnes for navigating in restricted waterway environments. However, economic viability eludes the smaller inland vessel. In the Netherlands for example, the cumulative number of small inland vessels has been gradually declining since the turn of the century (Sys and Vanelslander, 2011). As existing vessels reach their end-of-life, the preference for replacement leans towards the economies of scale offered by larger ships. The impact of losing these small vessels is felt most in urban areas where the freight transportation must shift modalities to the road network. This is a highly undesirable trend given the less sustainable nature of road transportation and the ever-growing concerns about congestion.

Crew costs dominate the operational expenditure of small vessels, drawing key attention when attempting to revive economic viability. Autonomy is considered to hold the potential to eliminate a large proportion of crew costs. However with profitability being prominent, the initial investment and running costs of an autonomous vessel must also be considered during technological research and development. Technologies that can enable the automation of these vessels without infringing on economical objectives are consequently of critical interest to the field of inland autonomy. Yet this factor is seldom encountered as a major research incentive.

The navigation system of an autonomous vessel requires sensors to perceive the environment and localise the vessel's own position. Radar and GPS/GNSS technology play primary roles in both autonomous applications and conventional vessels, providing perception and localisation data respectively. In Table 1, typical technologies encountered in autonomous vessel research are presented. The selection of these sensors is influenced by the requirements of the application environment, with the majority of research to date simply focusing on acquiring sufficient data to enable autonomous navigation. With the exception of a low-cost radar solution (Schuster et al., 2014), little consideration has been made to the economical branch of reasoning during sensor selection.

Table 2 presents the main sensor options along with a comparison of the crucial technological characteristics. Radar is the primary choice for perception tasks in autonomous vessels due to its proven, dependable reputation from decades of being used as a navigation aid. Inland vessels operate in denser environments, experiencing closer encounters with other vessels and static obstacles than sea-going vessels. Consequently achieving autonomy in this setting requires extra provision for near-field perception. LiDAR is typically the go to solution as it provides dense 3D point clouds of the nearby surroundings at a reasonable frequency, with only a linear range error. The cost of LiDAR however is considered one of the main limitations, especially where 3D sensor is required. The Puck™ from Velodyne Lidar© for example is one of the most affordable options yet still costs over $4000 USD (Hu et al., 2021) and unlike more expensive solid-state sensors can also be susceptible to mechanical failures.

Imaging sensors offer an alternative means of collecting depth data through stereovision, with devices such as the Intel© RealSense™ starting from around $200 USD (Léger et al., 2021). Although a single stereovision device

Authors' Biographies

Ir. Matt Hepworth is a Research Engineer at the Researchlab Autonomous Shipping at Delft University of Technology. He is specialised in the application of computer vision and artificial intelligence to situational awareness tasks in the field of maritime autonomy.

Ir. Vittorio Garofano is a Research Engineer at the Researchlab Autonomous Shipping at Delft University of Technology. His work focuses on the development of high-level control strategies for single and multi-vessel coordination.

Dr. Yusong Pang is an Assistant Professor in the Department of Maritime and Transport Technology of Delft University of Technology. His research interests fall into the automation, monitoring and intelligent control for large scale transport technology and logistics.

Dr. Vasso Reppa is an Assistant Professor in the Department of Maritime and Transport Technology of Delft University of Technology. Her current research interests include multi-level fault diagnosis and fault tolerant control, cooperative control, (adaptive) learning, observer-based estimation, and applications of autonomous systems in (waterborne) transport and smart buildings.

Citation	Technology
Peeters et al. (2020)	GPS, IMU, LiDAR, Imaging
Wang et al. (2020)	IMU, LiDAR, Imaging
Song et al. (2019a)	Radar
Shi et al. (2019)	Radar, AIS
Song et al. (2019b)	AIS, Radar
Sun et al. (2019)	AIS, Radar
Fan et al. (2019)	GPS, Radar, AIS, Sonar
Zhuang et al. (2016)	Radar
Lyu and Yin (2019)	AIS, Radar, GPS
Eriksen et al. (2018)	Radar
Villa et al. (2020)	GPS, LiDAR
Faggioni et al. (2022)	LiDAR, Imaging
Villa et al. (2020)	GPS, Radar, AIS, LiDAR, Imaging

Table 1: Navigation technologies utilised in maritime autonomy.

Sensor	Range	Data Frequency	Cost	Limitations
Radar	High	Low	Mid	Surface reflections
LiDAR	Mid	Mid	High	Vertical noise & weather sensitive
Imaging	Low-Mid	High	Low	Non-linear range error & light sensitive
Sonar	Low-Mid	Mid	Mid	Near-surface noise

Table 2: Comparison of sensor technologies.

cannot provide a 360° horizontal coverage like Radar and LiDAR, a degree of vertical coverage comes at no extra cost. Data from imaging sensors can also be processed using a wealth of computer vision techniques which are highly enabling to automation. Range is however limited as stereovision is subject to a quadratic depth error and to benefit fully from the sensors, good light and visibility are required. Sonar sensors are occasionally utilised by autonomous surface vehicles however the technology does not provide a great source for perception due to surface noise.

A sensor fusion technique with GPS and digital compass data has shown significant improvements over raw sensor information (Liu et al., 2022) for self-localisation and an inland application has also taken the fusion approach by utilising LiDAR and cameras to further precision for an urban waterway environment (Wang et al., 2020). The inland systems (Wang et al., 2020) (Peeters et al., 2020) in Table 1 both utilise stereovision however the technology takes a back seat role to LiDAR. In the automotive sector, LiDAR was also once the primary sensor choice, however its cost presented a barrier to consumer accessibility and studies proved the capability of vision technology (Wang et al., 2019). A stereovision-based navigation system also has the potential to provide an apt solution for autonomous inland vessels, especially when utilised for collision avoidance (Hepworth et al., 2021).

Rule five of the international collision regulations states that a vessel must maintain a proper look-out at all times so to make a full appraisal of the situation and the risk of collision (IMO, 1972). Whilst a human will never be able to maintain a look-out in all directions surrounding the vessel, at all times, their vision does not have a fixed orientation, providing flexibility in field of view. When perceiving the environment using stereovision devices, such flexibility is not achievable in the absence of a rotating mount, such as those used by LiDAR and Radar sensors. Nonetheless, the application of a multi-device setup can overcome this obstacle and imaging sensors are not susceptible to the same mechanical failures attributed with a rotating sensor. Furthermore, the application of a multi-device setup provides an enhanced sensor fault tolerance as there is some physical redundancy should one sensor fail.

This paper contributes a multi-device vision solution for autonomous vessels, utilising stereovision technology and artificial intelligence to automate situational-awareness and self-localisation. This economically-conscious navigation system is particularly well suited to application on the inland waterway, helping work towards the restoration of the small inland vessel fleet and a sustainable modality shift.

The paper is organised as follows. Section 2 details the stereovision-based navigation system including its application to a small-scale test vessel. Section 3 then covers a mapping procedure to post-process the navigation data into a useful format for collision avoidance. Section 4 presents the results of experimental testing in an indoor tank environment and in the Section 5, conclusions are drawn and recommendations for future research are presented.

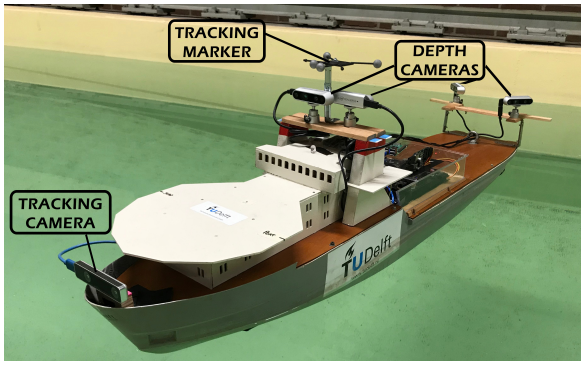

Figure 1: Model-scale test vessel utilised as the rig for experimental testing.

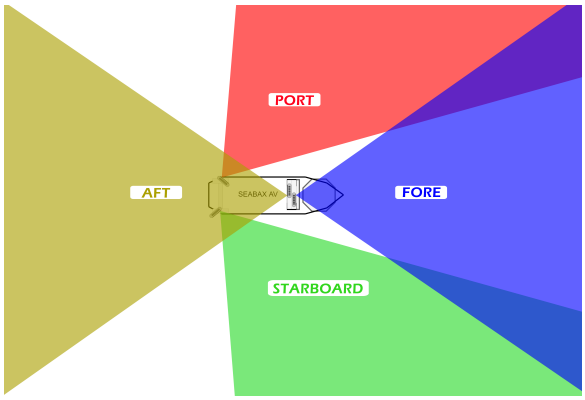

Figure 2: Perception Sensors - Field of View

2 Stereovision-based Navigation System

Supplying sufficient environmental awareness for a vessel to autonomously avoid collisions makes the role of a navigation system critical. The stereovision-based solution presented in this section consequently applies its focus to enabling interactions with dynamic obstacles on the waterways. A fully operational navigation system of an autonomous vessel must have the capacity to perform perception and localisation under any condition of visibility to comply with collision regulations. This paper narrows its focus to clear, well lit conditions for initial development. This section is split into two main topics. The first sub-section details the perception tasks, which involves the detection of obstacles and consequently determination of obstacle position to gain environmental awareness. The second topic then covers the localisation tasks which concern the determination of the autonomous vessel's current position.

2.1 Perception

The proposed camera configuration can be seen in Fig. 1 with the coverage illustrated in Fig.2. A total of four Intel© RealSense™ D435i depth devices are utilised covering a forward weighted coverage around the vessel, focusing on the critical interaction region for collision avoidance. The two blind spots that can be seen at the aft of the vessel only cause a momentary lapse in perception tracking deemed permissible at this stage of testing. However a comprehensive lookout could be achieved by the use of devices with wider angle lenses.

To minimise computational load for the multi-camera setup, a resolution of 480x240 at 15Hz was chosen for this application. Due to the correlation between the distance and the depth error in a stereovision setup, it is important to consider the range within which the device is designed to operate. In this case, the manufacturer recommends use cases with a range of up to ten metres, which is sufficient for the scale vessel (1:75) in question. For larger scale applications, the specific requirements would need to be considered, with the individual sensor specifications and baseline influencing the operating range.

Each of the depth cameras has two infrared sensors that allow for a depth frame to be extracted from the disparity map of the overlapping sensor coverage. An additional RGB sensor is also integrated into the device, which in this application is utilised for obstacle detection, with the resultant region of interest (ROI) being used to extract relevant position data from the point cloud data. As the colour and depth frames are not physically aligned in the device, alignment must be conducted during post-processing. Sub-sampling is applied to the depth frame directly prior to this alignment by decimating the depth frame through a factor of two, reducing subsequent computation by a factor of four. It further has the benefit of removing dead depth pixels, which occur in instances where the built-in stereo-matching algorithm cannot extract a depth value of acceptable accuracy.

The potential of artificial intelligence in maritime navigation is well founded, both for object detection (Prasad et al., 2017) (Shao et al., 2018) (Kim et al., 2018) (Faggioni et al., 2022) and for trajectory prediction Tang et al. (2022). The application of a Convolutional Neural Network (CNN) for object detection is further explored in this implementation. The MobileNet Single Shot Detector v2 framework is utilised, with a model being custom trained for this application from the COCO 2017 pre-trained backbone. At this stage of research the model is only trained for one object, a model scale tugboat. A total of 1200 manually annotated images of the model-scale tugboat were used to train the final model.

Running inference on the trained model creates a set of obstacle bounds (O_B) for each object detected within each image. These bounds describe are rectangular coordinates within the frame as defined in Eq. 1, where x and y describe the horizontal and vertical pixel position respectively. These coordinates, together with the classification of the object (C_{type}), are the key outputs of the neural network and are illustrated by the annotated box in Fig. 3.

Figure 3: Example Obstacle Detection Output - *The resultant bounds and classification of the detected obstacle are annotated on the input image.*

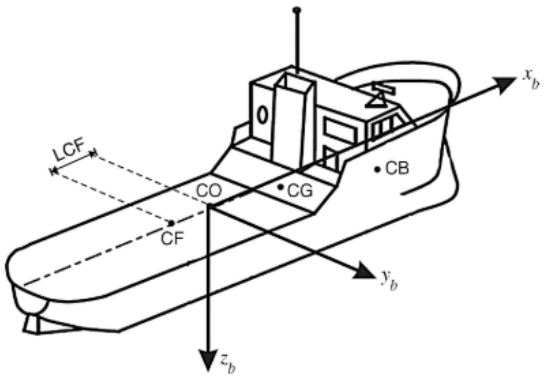

Figure 4: Body Reference Frame (Fossen, 2011)

As the model is only trained to recognise one vessel type at this stage, all detections will consequently carry the classification of 'tugboat'.

$$O_B = [x_{min}, y_{min}, x_{max}, y_{max}] \tag{1}$$

The obstacles detected are subsequently analysed to determine their actual position within 3D space and due to the collision avoidance application, this position is defined by the Closest Obstacle Point (COP). Whilst possible to extract all three dimensions, the surface vessels only operate within two dimensions of freedom therefore this point is defined only as a two dimensional coordinate. The extraction of the third dimension from the stereovision data could be useful for navigation purposes when determining the height of overhead obstacles, such as bridges. However, such information could better be determined using an algorithm specifically designed for this task, leaving the scope of this project to focus on enabling avoidance around obstacles.

The minimum depth value is extracted from the ROI of each detected obstacle as defined in Eq. 2, where z is the depth value at each pixel. The known depth value is then scaled to a distance (X) in metres and the lateral position (Y) of the obstacle can also be extracted from the device using built-in geometrical optics, taking the centroid of the obstacle bounds as reference. The position of the contact vessel is at this stage defined in coordinates relative to the device (D) with which the obstacle has been detected, as defined in Eq. 3. These device coordinates shall later be translated in the mapping section, to merge all four devices into one coordinate system.

$$z_{min} = min(z_{x_i y_j}),\ x_i = x_{min}, ..., x_{max},\ y_j = y_{min}, ..., y_{max} \tag{2}$$

$$COP = (X_D, Y_D) \tag{3}$$

2.2 Localisation

To localise the Autonomous Vessel, a dedicated tracking device is utilised which makes use of two fish eye lenses and a built-in inertial measurement unit to conduct stereo V-SLAM (Visual- Simultaneous Localisation and Mapping). The device selected for this application also belongs to the Intel© RealSense™ range, with the T265 camera fulfilling this role.

The tracking camera is mounted to the bow of the vessel as seen in Fig. 1. The position at the bow of the vessel was selected to minimise vision occlusions and IMU noise due to vibrations. The SLAM algorithm that runs onboard the T265 uses a Visual-Intertial-Odometry (VIO) technique and provides pose directly to the vessel. The two fish-eye sensors allow for a large field of view to be covered and a calibrated scale of the environment. The further fusion of visual feature recognition with high frequency IMU data allows for the VIO algorithm to further track movements over shorter time steps, enhancing the localisation accuracy. Allowing for initial acceleration from a stationary position to be measured where vision alone would struggle.

The device directly provides position and orientation, therefore the only post-processing tasks concern translation to the desired coordinate frame and units. The local coordinate system is translated from the device coordinate axes to correspond with the body reference frame and the extraction of Euler angles from the quaternion orientation data. The position of the vessel is only required within two body-frame dimensions, x_b and y_b as defined in Fig. 4 and the Euler angle of main interest is the yaw angle (ψ) which provides a means to derive vessel heading.

3 Mapping Strategy

Mapping represents the final sub-task in the navigation system and is responsible for post-processing the perception and localisation outputs into meaningful data for collision avoidance. Post-processing involves two main procedures. Firstly, the translation of perception results to map positions into a single homogeneous reference frame, and secondly the fusion of localisation and perception data to map the position of the AV and obstacles in the global coordinate system. Obstacles are hereon referred to as Contact Vessels (CV), due to the collision avoidance scope.

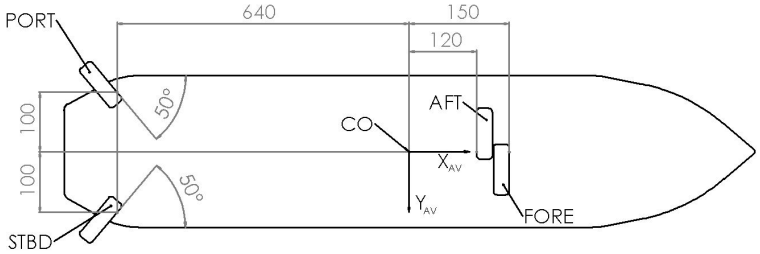

Figure 5: Multi-device Configuration for Perception. *All linear dimensions in millimetres.*

$$\begin{pmatrix} X_{AV_D} \\ Y_{AV_D} \end{pmatrix} = \begin{pmatrix} X_o \\ Y_o \end{pmatrix} + \begin{pmatrix} cos\gamma & sin\gamma \\ -sin\gamma & cos\gamma \end{pmatrix} \begin{pmatrix} X_D \\ Y_D \end{pmatrix} \tag{4}$$

Device	$X_o(m)$	$Y_o(m)$	$\gamma(^\circ)$
Fore	0.15	0	0
Aft	0.12	0	180
Port	-0.64	-0.1	50
Starboard	-0.64	0.1	-50

Table 3: Device orientation and position offsets.

Each CV position attained by the perception system is translated from the individual depth device coordinate system into a consolidated AV domain. The origin of the AV domain, is located at the geometric centre of the vessel as indicated by point CO in Fig.4, with the body axes orientation also matching that of the 2D AV coordinate system. The relative camera is translated by applying the displacement from the origin point and accounting for the configured angle through axes rotation. The displacement and angle for each camera can be found in Table 3. The result of translation is a set of CV coordinates within the AV domain (X_{AV_D}, Y_{AV_D}). One particularly useful attribute for collision avoidance tasks which can be derived from this is the relative bearing of the CV as seen in Eq. 6. This describes the angle between the AV's forward direction and the CV position as can be visualised in Fig. 6 (a).

$$\hat{\beta} = atan2(Y_{AV_D}, X_{AV_D}) \tag{5}$$

$$\beta = \begin{cases} \hat{\beta}, & \hat{\beta} \geq 0 \\ 2\pi + \hat{\beta}, & \hat{\beta} < 0 \end{cases} \tag{6}$$

With the perception output coalesced into a singular coordinate frame, any replications that exist in the overlapping field-of-view regions are removed and each CV is associated with an identification number for tracking over time. Tracking is achieved using a Global Nearest Neighbour approach whereby the coordinates of each CV detected in the current frame are compared with all CV positions in the previous frame and associated by minimising Euclidean distance.

Global mapping is conducted by utilising the localisation data. The global coordinate system requires a base map to be provided by the user that details the area of navigation, the starting position of the AV and the start heading (φ_0). The AV position (X_{av}, Y_{av}) can thus be defined by summing the start position with the position data from the localisation output. By monitoring the position over time, the AV speed (u_{av}) can be determined, following a constant time step of one second. The relative heading value is calculated through summing the initial offset with the yaw value (ψ) as defined in Eq. 7 and then applying the cases in Eq. 8.

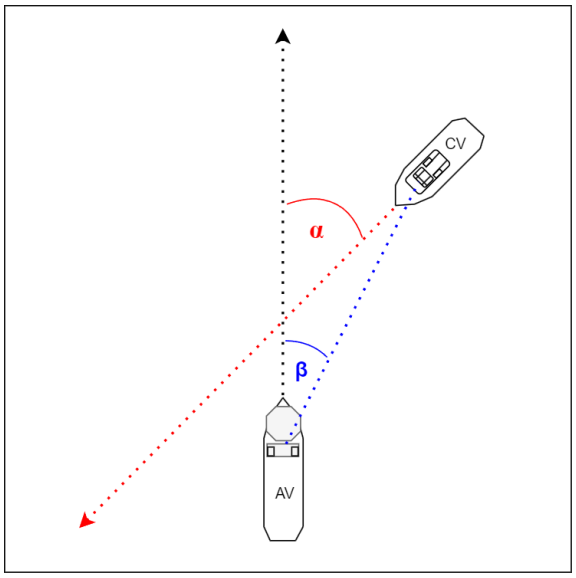

Figure 11(a): Bearing (β) and Approach Angle (α) of the CV in the AV domain.

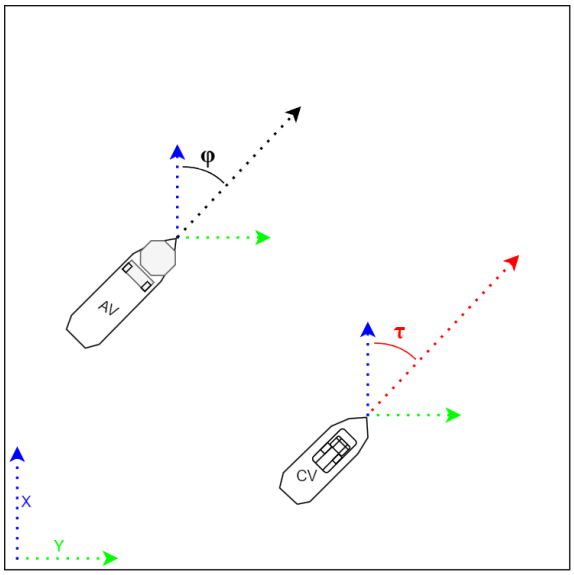

Figure 11(b): Heading (φ) of the AV and Track angle (τ) of the CV in the global domain.

Figure 6: Visualisation of Key Angles from the Mapping Procedure

$$\hat{\varphi} = \varphi_0 + \psi \tag{7}$$

$$\varphi = \begin{cases} \hat{\varphi}\,, & 0 \leq \hat{\varphi} \leq 2\pi \\ 2\pi + \hat{\varphi}\,, & \hat{\varphi} < 0 \\ \hat{\varphi} - 2\pi, & \hat{\varphi} > 2\pi \end{cases} \tag{8}$$

Defining the global coordinates of each obstacle (X_{cv}, Y_{cv}) is accomplished through the translation of obstacle coordinates in the AV domain using the current AV position and orientation. The velocity vector between the CV position at the previous time step (one second) and the current position can be referred to as the relative track, with the vessel speed (u_{cv}) being the magnitude and the Track Angle (τ) being the direction. This Track Angle (τ) is defined in Eq. 13 and can be seen as an estimation of the CV heading as visualised in Fig. 6 (b). To assist in collision avoidance procedure it is also useful to know the angle at which a dynamic obstacle is moving relative to the AV itself. This directional measure of obstacle dynamics is referred to as the approach angle (α) which is defined in Eq. 15 and can be seen in Fig. 6 (a).

$$\begin{pmatrix} X_{cv} \\ Y_{cv} \end{pmatrix} = \begin{pmatrix} X_{av} \\ Y_{av} \end{pmatrix} + \begin{pmatrix} cos\psi & sin\psi \\ -sin\psi & cos\psi \end{pmatrix} \begin{pmatrix} X_{AV_D} \\ Y_{AV_D} \end{pmatrix} \tag{9}$$

$$\begin{pmatrix} \Delta X_{cv} \\ \Delta Y_{cv} \end{pmatrix} = \begin{pmatrix} X_{cv}(t) \\ Y_{cv}(t) \end{pmatrix} - \begin{pmatrix} X_{cv}(t-1) \\ Y_{cv}(t-1) \end{pmatrix} \tag{10}$$

$$u_{cv} = \sqrt{\Delta X_{cv}^2 + \Delta Y_{cv}^2} \tag{11}$$

$$\hat{\tau} = (atan2(\Delta Y_{cv}, \Delta X_{cv}) + \varphi_0) \tag{12}$$

$$\tau = \begin{cases} \hat{\tau}\,, & 0 \leq \hat{\tau} \leq 2\pi \\ 2\pi + \hat{\tau}\,, & \hat{\tau} < 0 \\ \hat{\tau} - 2\pi, & \hat{\tau} > 2\pi \end{cases} \tag{13}$$

$$\hat{\alpha} = \tau - \varphi \tag{14}$$

$$\alpha = \begin{cases} \hat{\alpha}, & \hat{\alpha} \geq 0 \\ \hat{\alpha} + 2\pi, & \hat{\alpha} < 0 \end{cases} \tag{15}$$

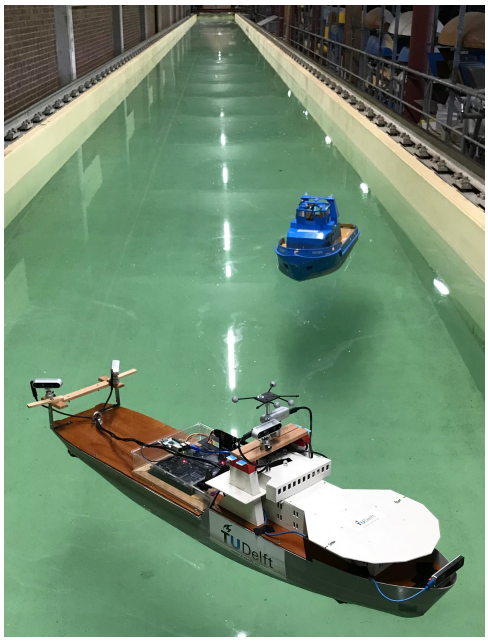

Figure 7: Experimental Setup - *Front: Autonomous Vessel(AV), rear: Contact Vessel (CV)*

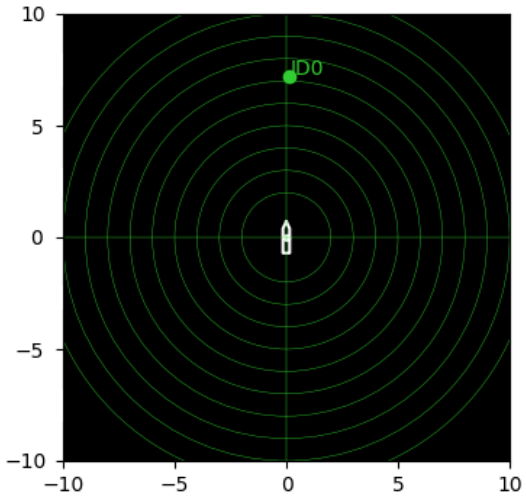

Figure 8: Pseudo-radar plot.

4 Experimental Results

The performance of the developed navigation system is evaluated hereon with data gathered through experimental testing. The test environment and vessels used in experimental testing can be seen in Fig. 7 with the vessel in the front acting as the autonomous vessel and the vessel behind as the contact vessel. The test environment is further equipped with an OptiTrack© 3D tracking system that is used to extract ground truth vessel pose data from tracking markers.

The performance of the mapping procedure is not the explicit focus of experimental testing in this work, rather it is a post-processing task for collision avoidance. Instead the primary data which it relies on from the perception and localisation tasks is assessed in detail. Nonetheless, the mapping procedure was reviewed for functionality during experimentation. In Figure 8, a pseudo-radar plot can be seen which provides real-time visualisation of the mapping output within the AV domain.

Evaluations of the perception and localisation systems are made independently in the following sub-sections so to highlight the specific performance of the stereovision sensor solutions under each task. The perception and navigation tasks were run onboard the AV's NVIDIA Jetson TX2 during all testing.

4.1 Perception

The perception system is built within a ROS package and designed to run on the onboard Jetson TX2. The output is published over the ROS network and was found to be capable of reliably supplying data at a frequency of 5Hz. The robustness of obstacle detection and the accuracy of positioning are selected as key criteria for the quantitative analysis. The precision/recall KPIs defined in Eq. 16 and 17, and the position error KPI defined Eq. 18 are used for this purpose.

$$Precision = \frac{correct\ predictions}{all\ predictions} \tag{16}$$

$$Recall = \frac{correct\ detections}{all\ ground\ truths} \tag{17}$$

$$RMSE_{pos} = \sqrt{\frac{\sum_{i=1}^{N}(\hat{o}_i - o_i)^2}{N}} \tag{18}$$

where: o_i = measured position, $\hat{o}_i$ = ground truth position, N = number of samples

$$ATE_{pos} = \sqrt{\frac{\sum_{i=1}^{N}(\hat{p}_i - p_i)^2}{N}} \tag{19}$$

$$ATE_{rot} = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}} \tag{20}$$

where: p_i, y_i = measured values, $\hat{p}_i$, $\hat{y}_i$ = ground truth values, N = number of samples

CNN model inference was tested on the NVIDIA Jetson TX2 mounted onboard the vessel, with frames being processed from all four camera streams. Running inference on a TensorRT engine model was found to reduce latency fivefold over a standard Tensorflow format. This enables obstacle detection to be processed at up to 10Hz and does so without exhibiting a reduction in average precision. The obstacle detector demonstrated this high precision with increasing recall, achieving a total precision of 97.4% and recall of 94.6%. However as the neural network was both trained and tested with one type of obstacle in one environment, the significance of this result is limited to the specific experimental environment.

The accuracy of obstacle positioning is evaluated using the accepted ground truth from the OptiTrack camera system. The ground truth position is aligned with the centre-point (CO in Fig. 4) of the CV, whereas the perception system provides the closest point (Eq. 3) to the AV. Data was collected within the ten metre range covered by the perception system, with a total of 280 data points being gathered across three tests with varied vessel positioning. A root mean square position error of 0.31 metres is obtained from the experimental data, as can be seen in Figure 9. Whilst this appears high with respect to the vessel dimensions, it is valuable to remember that the the COP approach of the perception system carries an inherent safety factor. Correcting for the minimum causal deviation of this safety factor (half vessel beam), already reduces this error to 0.16m. This evaluation therefore delivers insight to the performance of the specific perception approach, rather than the stereovision sensor itself.

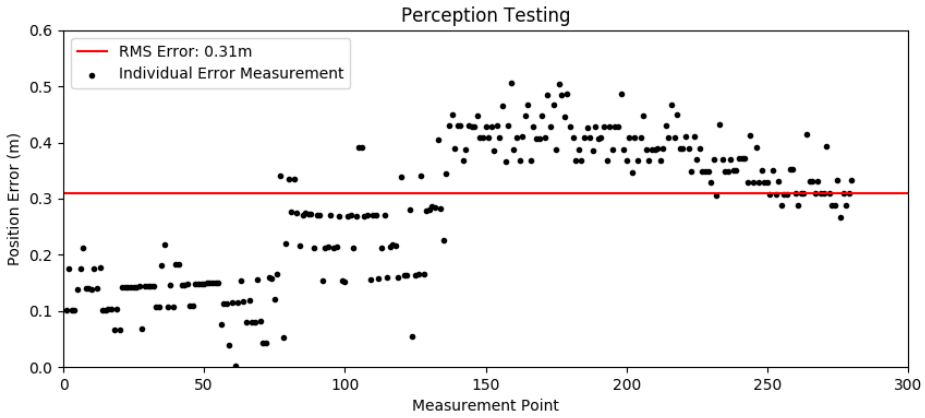

Figure 9: Perception System Position Error

4.2 Localisation

The localisation system is also built within a ROS package and tested on the Jetson TX2. A consistent stream frequency of 10Hz was attainable. As such, should an ongoing process require a higher rate of localisation data, this could be attained directly from the localisation system, bi-passing the mapping procedure which publishes data at half the rate.

Experimental testing to evaluate localisation performance was conducted using two path types, a slalom/zigzag style path and a straight path as can be seen in Figure 10. These two path types aim to review performance under typical sailing conditions, i.e. a straight path between waypoints, but also when conducting obstacle avoidance where sharp turns are required. The two KPIs selected are the position error and heading error, which are both defined by the root mean square deviance between the acquired sensor value and the accepted ground truth value (Eq. 19 and 20). These errors were again calculated by taking the pose data from the Optitrack system to be the ground truth values.

Both the straight paths and slalom paths exhibited a correlation between the magnitude of error and the path length, a behaviour that is commonly faced by localisation systems that utilise a Visual-SLAM approach. This behaviour is frequently characterised by the sensor drift whereby position error accumulates over the length of a path. As well as being visually evident in the path plots, this behaviour can be further observed in the boxplots in Figure 11. When the sampling length of the waterway is reduced, the position and heading errors also reduce significantly. The average error across the two path types were found to be 0.129m and 5.3° for the position and

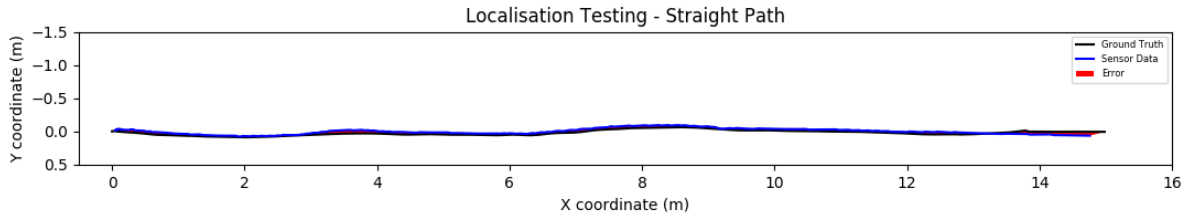

Figure 10(a): Straight Path

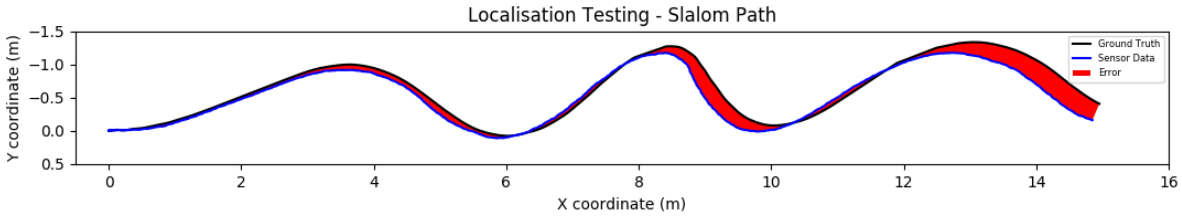

Figure 10(b): Slalom Path

Figure 10: Localisation System Testing - Performance following Straight and Slalom Paths.

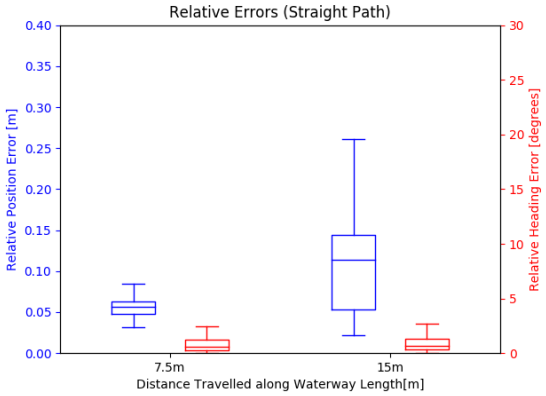

Figure 11(a): Straight Path Boxplot

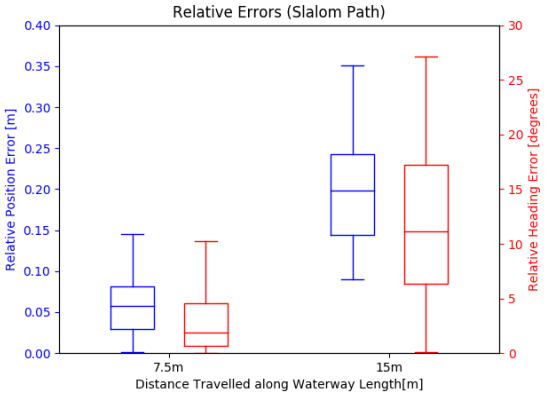

Figure 11(b): Slalom Path Boxplot

Figure 11: Boxplot representation of relative position and heading errors for divided sub-trajectory lengths. Boxplot colour corresponds to y-axis of the same colour. Specifically. blue: position error and red: heading error.

heading errors respectively along the full path lengths. With a reduced sampling length of 7.5m, the position and heading errors also fall to 0.068m and 2.6° respectively.

It is furthermore clear from the results that the localisation performance along a straight path is far superior with this technology when compared to path with slalom manoeuvres. The operational distance and anticipated vessel behaviour should therefore be considered when utilising V-SLAM based technology for autonomous vessel applications, where possible utilising additional sensor data to supplement. These evaluations were made under normal operational cases, however the localisation system was further observed on multiple occasions to be susceptible to the impact of the kidnapped robot problem, hindering the overall system performance.

5 Conclusions and Further Research

A multi-device stereovision based-navigation system has been proposed to achieve low-mid range perception and localisation onboard autonomous inland vessels. An additional mapping procedure has further been developed to post-process the perception and localisation data into desirable attributes for collision avoidance.

The perception system introduced enables an autonomous vessel to maintain a constant lookout in areas of collision risk. The utilisation of a convolutional neural network to conduct object detection has demonstrated its suitability to the autonomous inland vessel application. With the model maintaining high precision with increasing recall at impressive inference rates. A refined approach was taken towards the processing of depth data which avoids the computationally expensive manipulation of dense point clouds.

A dedicated stereovision tracking camera, utilising a Visual Inertial Odometry approach was also integrated to

provide pose data for autonomous vessel localisation. The system proved capable of providing accurate tracking data within a short range however, the localisation performance was seen to reduce as the vessel progressed down the path length. The system was further observed on multiple occasions to be susceptible to the impact of the kidnapped robot problem, limiting its reliability as an independent solution.

A clear progression for future work involves focus on collision avoidance using the developed navigation system, allowing the outputs of the mapping procedure to be tested for suitability. The development of an Inland Maritime Dataset for CNN training would enable research to apply focus to model optimisation, perhaps creating a similar initiative to the KITTI benchmark suite (Geiger et al., 2012). Finally, the developed navigation system would benefit from the addition of radar and GPS sensors for gathering complementary data for fusion with the stereovision sensors. This sensor fusion approach could alleviate the reliability issues of the localisation system and the range of the perception system whilst remaining economically conscious. Future research also aims to overcome shortfalls in the system during low light conditions and/or poor conditions of visibility. Furthermore, during scale up from model-scale to full-scale, consideration must be made to the selection of suitable imaging sensors, lenses and baseline when configuring the stereovision setup.

6 Acknowledgement

The research presented in this paper was supported by the Researchlab Autonomous Shipping (RAS) and funded by the European Union and the European Regional Development Fund, as part of the Interreg North Sea Region project AVATAR. This research is also supported by the EFRO REACT-EU Op-Zuid Project "Fieldlab Autonomous Sailing Technology (FAST)" (no. 4119).

References

Eriksen, B.O.H., Wilthil, E.F., Flten, A.L., Brekke, E.F., Breivik, M., 2018. Radar-based maritime collision avoidance using dynamic window. IEEE Aerospace Conference Proceedings 2018-March, 1–9. doi:10.1109/AERO.2018.8396666.

Faggioni, N., Leonardi, N., Ponzini, F., Sebastiani, L., Martelli, Michele", e.J., Fagiolini, A., Vasik, P., Turi, M., Bruzzone, A., Pickl, S., Neumann, V., Stodola, P., 2022. Obstacle detection in real and synthetic harbour scenarios , 26–38.

Fan, Y., Sun, X., Wang, G., 2019. An autonomous dynamic collision avoidance control method for unmanned surface vehicle in unknown ocean environment. International Journal of Advanced Robotic Systems 16, 1–11. doi:10.1177/1729881419831581.

Fossen, T.I., 2011. Handbook of Marine Craft Hydrodynamics and Motion Control. doi:10.1002/9781119994138.

Geiger, A., Lenz, P., Urtasun, R., 2012. Are we ready for autonomous driving? the kitti vision benchmark suite, in: Conference on Computer Vision and Pattern Recognition (CVPR).

Hepworth, M., Garofano, V., Reppa, V., Negenborn, R.R., 2021. Collision Avoidance for Autonomous Inland Vessels using Stereovision. URL: https://repository.tudelft.nl/islandora/object/uuid:5e00f536-fa99-4116-8f87-cfb01a525fcb?collection=education.

Hu, T., Sun, X., Su, Y., Guan, H., Sun, Q., Kelly, M., Guo, Q., 2021. Development and performance evaluation of a very low-cost uav-lidar system for forestry applications. Remote Sensing 13. URL: https://www.mdpi.com/2072-4292/13/1/77.

IMO, 1972. Convention on the International Regulations for Preventing Collisions at Sea.

Kim, K., Hong, S., Choi, B., Kim, E., 2018. Probabilistic ship detection and classification using deep learning. Applied Sciences 8, 936. URL: http://dx.doi.org/10.3390/app8060936, doi:10.3390/app8060936.

Léger, É., Gueziri, H.E., Collins, D.L., Popa, T., Kersten-Oertel, M., 2021. Evaluation of low-cost hardware alternatives for 3d freehand ultrasound reconstruction in image-guided neurosurgery , 106–115.

Liu, W., Liu, Y., Bucknall, R., 2022. Filtering based multi-sensor data fusion algorithm for a reliable unmanned surface vehicle navigation. Journal of Marine Engineering & Technology 0, 1–17. URL: https://doi.org/10.1080/20464177.2022.2031558, doi:10.1080/20464177.2022.2031558, arXiv:https://doi.org/10.1080/20464177.2022.2031558.

Lyu, H., Yin, Y., 2019. COLREGS-Constrained Real-Time Path Planning for Autonomous Ships Using Modified Artificial Potential Fields. Journal of Navigation 72, 588–608. doi:10.1017/S0373463318000796.

Peeters, G., Kotzé, M., Afzal, M.R., Catoor, T., Van Baelen, S., Geenen, P., Vanierschot, M., Boonen, R., Slaets, P., 2020. An unmanned inland cargo vessel: Design, build, and experiments. Ocean Engineering 201. doi:10.1016/j.oceaneng.2020.107056.

Prasad, D.K., Rajan, D., Rachmawati, L., Rajabally, E., Quek, C., 2017. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. IEEE Transactions on Intelligent Transportation Systems 18, 1993–2016. doi:10.1109/TITS.2016.2634580.

Schuster, M., Blaich, M., Reuter, J., 2014. Collision avoidance for vessels using a low-cost radar sensor. volume 19. IFAC. URL: http://dx.doi.org/10.3182/20140824-6-ZA-1003.01872, doi:10.3182/20140824-6-za-1003.01872.

Shao, Z., Wu, W., Wang, Z., Du, W., Li, C., 2018. Seaships: A large-scale precisely annotated dataset for ship detection. IEEE Transactions on Multimedia 20, 2593–2604. doi:10.1109/TMM.2018.2865686.

Shi, B., Su, Y., Wang, C., Wan, L., Luo, Y., 2019. Study on intelligent collision avoidance and recovery path planning system for the waterjet-propelled unmanned surface vehicle. Ocean Engineering 182, 489–498. URL: https://doi.org/10.1016/j.oceaneng.2019.04.076, doi:10.1016/j.oceaneng.2019.04.076.

Song, L., Chen, H., Xiong, W., Dong, Z., Mao, P., Xiang, Z., Hu, K., 2019a. Method of Emergency Collision Avoidance for Unmanned Surface Vehicle (USV) Based on Motion Ability Database. Polish Maritime Research 26, 55–67. doi:10.2478/pomr-2019-0025.

Song, L., Chen, Z., Dong, Z., Xiang, Z., Mao, Y., Su, Y., Hu, K., 2019b. Collision avoidance planning for unmanned surface vehicle based on eccentric expansion. International Journal of Advanced Robotic Systems 16, 1–9. doi:10.1177/1729881419851945.

Sun, X., Wang, G., Fan, Y., Mu, D., Qiu, B., 2019. Fast Collision Avoidance Method Based on Velocity Resolution for Unmanned Surface Vehicle. Proceedings of the 31st Chinese Control and Decision Conference, CCDC 2019 , 4822–4827doi:10.1109/CCDC.2019.8832993.

Sys, C., Vanelslander, T., 2011. Future challenges for inland navigation : a scientific appraisal of the consequences of possible strategic and economic developments up to 2030. First ed.

Tang, H., Yin, Y., Shen, H., 2022. A model for vessel trajectory prediction based on long short-term memory neural network. Journal of Marine Engineering & Technology 21, 136–145. URL: https://doi.org/10.1080/20464177.2019.1665258, doi:10.1080/20464177.2019.1665258, arXiv:https://doi.org/10.1080/20464177.2019.1665258.

Villa, J., Aaltonen, J., Koskinen, K.T., 2020. Path-following with lidar-based obstacle avoidance of an unmanned surface vehicle in harbor conditions. IEEE/ASME Transactions on Mechatronics 25, 1812–1820. doi:10.1109/TMECH.2020.2997970.

Wang, W., Gheneti, B., Mateos, L.A., Duarte, F., Ratti, C., Rus, D., 2020. Roboat: An Autonomous Surface Vehicle for Urban Waterways , 6340–6347doi:10.1109/iros40897.2019.8968131.

Wang, Y., Chao, W.L., Garg, D., Hariharan, B., Campbell, M., Weinberger, K.Q., 2019. Pseudo-lidar from visual depth estimation: Bridging the gap in 3D object detection for autonomous driving. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2019-June, 8437–8445. doi:10.1109/CVPR.2019.00864, arXiv:1812.07179.

Zhuang, J., Zhang, L., qi Zhao, S., Cao, J., Wang, B., bing Sun, H., 2016. Radar-based collision avoidance for unmanned surface vehicles. China Ocean Engineering 30, 867–883. doi:10.1007/s13344-016-0056-0.