

Confined Space Underwater Positioning Using Collaborative Robots

Xueliang Cheng*

Department of Electrical and Electronic Engineering
University of Manchester
Manchester, M13 9PL
xueliang.cheng@manchester.ac.uk

Kanzhong Yao

Department of Electrical and Electronic Engineering
University of Manchester
Manchester, M13 9PL
kanzhong.yao@manchester.ac.uk

Mellisa Sandison

Department of Electrical and Electronic Engineering
University of Manchester
Manchester, M13 9PL
melissa.sandison@manchester.ac.uk

Andrew West

Department of Electrical and Electronic Engineering
University of Manchester
Manchester, M13 9PL
andrew.west@manchester.ac.uk

Simon Watson

Department of Electrical and Electronic Engineering
University of Manchester
Manchester, M13 9PL
simon.watson@manchester.ac.uk

Ognjen Marjanovic

Department of Electrical and Electronic Engineering
University of Manchester
Manchester, M13 9PL
Ognjen.Marjanovic@manchester.ac.uk

Barry Lennox

Department of Electrical and Electronic Engineering
University of Manchester
Manchester, M13 9PL
barry.lennox@manchester.ac.uk

Keir Groves

Department of Electrical and Electronic Engineering
University of Manchester
Manchester, M13 9PL
keir.groves@manchester.ac.uk

STATEMENTS

Funding: This research was funded by Engineering and Physical Sciences Research Council (EPSRC) under grants: EP/P01366X/1, EP/W001128/1 and EP/V026941/1.

Author contributions: Keir Groves, Melissa Sandison and Xueliang Cheng contributed to hardware and software design and build. The experiment was designed by Keir Groves and Xueliang Cheng, and performed by Xueliang Cheng, Kanzhong Yao, Melissa Sandison and Keir Groves. Keir Groves, Andrew West and Xueliang Cheng conceived and formulated the CAP-CPnP and CAP-CD techniques. Kanzhong Yao, Ognjen Marjanovic and Simon Watson contributed and integrated the leader-follower algorithm. Xueliang Cheng performed all data analysis. Keir Groves, Barry Lennox, Ognjen Marjanovic and Simon Watson provided funding, supervised and guided the research. Xueliang Cheng and Keir Groves wrote the manuscript. All authors edited and advised on the manuscript.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: Videos and Dataset has been uploaded to:

*To whom correspondence should be addressed; Alternative e-mail: u98908xc@gmail.com

https://livemanchesterac-my.sharepoint.com/:f:/g/personal/xueliang_cheng_postgrad_manchester_ac_uk/Em_0ol2h60dLicXANMnqVfQB88aHH5fvvru5QpY3QzChOw?e=wTTHy8

The source code is hosted at the Github repository:

`git@github.com:Xueliang/CAP-system.git`

Abstract

Positioning of underwater robots in congested and enclosed spaces remains unsolved for field operations. Well existing field ready systems are generally more suited to use in large, open marine environments. In enclosed and congested environments, which are common in industrial settings, existing systems suffer from a mixture of issues, including: poor coverage, reliance on added infrastructure and the need for feature rich environments. Accurate and readily deployable positioning is a prerequisite for performing repeatable autonomous missions and therefore, until now, there has been a technological bottleneck in such environments. The Collaborative Aquatic Positioning system presented in this paper uses a mixture of collaborative robotics and sensor fusion to solve the problem. The proposed positioning system is deployed in a large water tank and repeatable autonomous missions are performed using the system's position measurement for real-time feedback. Experimental results show that the system can achieve a Euclidean distance Root Mean Square Error (RMSE) of 70 mm while operating in real-time. The system enables almost complete coverage of the body of water in large pools without requiring fixed infrastructure, lengthy calibration, or feature rich environments. The Collaborative Aquatic Positioning system builds upon recent advances in mobile robot sensing and a recently developed leader follower control system to provide a step-change in positioning capability for real-world, high-precision autonomous underwater navigation.

1 INTRODUCTION

Over the last decade, the field of underwater robotics has grown substantially. Today, the deployment of Remotely Operated Vehicles (ROVs) is both safe and routine, extending beyond offshore industries (Reach Robotics, 2023; Smith Jr et al., 2021) to include operations in spatially restricted aquatic environments. Presently, the absence of an adequate positioning system represents a significant technological bottleneck, hindering the introduction of higher levels of autonomy in the field use of underwater robots in industrial applications.

Thus far there has not been an underwater positioning system available, with sufficient accuracy, scalability and real world viability, to facilitate the use of autonomous robotics in typical industrial underwater settings. The goal of this ongoing research is to provide an underwater positioning system that can function in highly physically constrained environments (Negahdaripour and Firoozfam, 2006; Ayoola, 2019) with sufficient accuracy to facilitate repeatable and reliable autonomous robotic missions in real-world scenarios (FIS 360, 2021; Karlsen et al., 2021). To have greatest impact in real-world applications, the positioning system should; not require the addition of fixed infrastructure, such as cameras; it should cover large areas with minimal blind spots; and it should provide positional errors (Euclidean distance Root Mean Square Error, RMSE) less than approximately 100 mm, to enable navigation in cluttered environments.

1.1 Highly constrained underwater environments

There are numerous categories of highly physically constrained underwater environments that must be accessed on a regular basis for purposes such as inspection, maintenance, repair, or decommissioning. These environments are characterized by limited spatial conditions, which may include confinement by walls, narrow passageways, or areas that are densely populated with obstacles. Examples of such environments include nuclear fuel storage pools (Griffiths et al., 2016), liquid storage facilities (Duecker et al., 2019), flooded mines (Álvarez-Tuñón et al., 2018), ship hulls (Song and Cui, 2020), pipelines (Zhao et al., 2022), and off-shore wind turbine foundation tripods (Barari et al., 2021). In many cases, regular inspection of these environments is a regulatory necessity. For example, the Harmonized System of Survey and Certification (HSSC) Guidelines, Resolution A.1140 (International Maritime Organization, 2015), specify that the outside of a passenger ship's hull should be carried out twice in a five year period.

Traditionally, accessing highly constrained underwater environments requires human divers. However, working in these environments can be high-risk, tedious and expensive (Lin and Dong, 2023). This reality, alongside technological advancements and the need to reduce costs, has led to increased use of robotic vehicles to access such areas (Christ and Wernli Sr, 2013; FitzGerald et al., 2022; Brantner and Khatib, 2021). In addition, small robots can be used to operate in environments where human access would not be possible due to physical constraints, for example, access through pipes or openings that are too narrow for humans to access (Fackler, 2017).

1.2 Positioning for underwater robots

Typically robotic underwater vehicles are split into two categories, Remote Operated Vehicles (ROVs) and Autonomous Underwater Vehicles (AUVs). Whilst ROVs are remotely driven by a human operator, AUVs are autonomous vehicles that receive high-level commands from the operator, such as a list of waypoints that the robot must navigate to. AUVs typically necessitate a positioning system to facilitate autonomous navigation. Conversely, remotely operated vehicles (ROVs) do not inherently require such systems for operation; however, incorporating a positioning system can enhance their performance.

For remotely driven ROV missions, there are two main benefits associated with the provision of accurate positioning. First, the operator has an additional source of information, which can aid navigation. This assists the operator in driving through the environment and not losing track of the robot's location, which is a common issue (Loebis et al., 2004; Shabani et al., 2015; Li et al., 2021). Second, an accurate positioning system allows any data from the sensor payload to be geo-spatially located (geo-tagged), meaning that sensor readings can be repeatably mapped in the underwater environment and presented in human readable formats, such as a heat map format.

For AUVs, an accurate positioning system is essential (Ferreira et al., 2010; Palomeras et al., 2019; Hegrenas et al., 2008), with effective autonomous navigation relying on regular, accurate position updates. There are several well documented benefits to performing fully autonomous unmanned robotic missions, where the operator has minimal input. These include cost reduction, improved repeatability, increased survey frequency. Aside from fully autonomous systems, lower levels of autonomy, such as position and velocity control, which can provide smooth and accurate navigation in the presence of disturbances, also require accurate positioning information.

1.3 Accuracy requirements in constrained environments

Navigating underwater vehicles through highly constrained physical environments is challenging and requires precise movement. This is in contrast to operating in open oceans where robots generally move in free space and therefore the precision and accuracy requirements can be relaxed. The accuracy and precision of a robot's positioning and pose estimation is a limiting factor affecting the performance of any navigational control system, since a robot can't navigate with higher accuracy than its state estimation system.

While the accuracy requirements of a positioning system will vary depending on the mission and environment, it is useful to have quantitative targets, even if they can only be approximate. To gauge the accuracy requirements for

navigation, a representative example of a small underwater vehicle navigating through an opening that is 500 mm wide is considered. Assuming that the robot is 340 mm wide (Blue Robotics Inc, 2024), this would leave 80 mm either side to account for both position and control errors. In a recent challenge statement from the UK nuclear industry (FIS 360, 2021), an accuracy requirement of 50 mm was specified for revisiting the same position in a small, enclosed storage pond of 7 m x 7 m. Therefore, it is concluded that positioning accuracy in the 0-100 mm range would be acceptable for many common missions in constrained underwater environments. The accuracy is defined here as the RMSE euclidean distance between the estimated position and the actual position.

1.4 Infrastructure and coverage

To be useful in practical situations, a confined space positioning system should require minimal infrastructure and be capable of good coverage of the environment. Fixing infrastructure, such as installing underwater cameras, beacons or markers, in the environment is time consuming, expensive and generally requires lengthy calibration. Moreover, installing infrastructure is often not feasible in environments where access is highly restricted due to safety concerns. In such scenarios, deploying infrastructure becomes impractical. A pertinent example is in nuclear fuel ponds, where safety protocols severely limit the introduction of external equipment. It is also important that the system can operate over a high proportion of the environment, not suffer from blind spots, nor be confined to a local area. Systems that rely on fixed infrastructure often suffer from such problems because the fixed equipment has restricted field of view and range.

2 REVIEW OF UNDERWATER LOCALISATION SYSTEMS

The reason that positioning robots underwater remains particularly challenging relates mostly to the properties of water itself. Technologies that are commonly used in air, such as GPS and LiDAR, rely on electromagnetic frequency bands that are highly attenuated by water, rendering them largely unusable. There are some exceptions to this. For example, there is a relative reduction in attenuation of visible light frequencies (380-750 nm) and as a consequence, in aquatic environments, cameras are the most successfully used sensor that relies on propagation of electromagnetic waves. Aquatic applications that require information to be transferred over long distances typically rely on the use of acoustic signals, which are not significantly attenuated in water. However, acoustic signals are subject to relatively slow propagation, low frequency, and are impacted by multipath issues (Stojanovic and Preisig, 2009; Singer et al., 2009), these factors limit their accuracy and refresh rate when used for underwater positioning. In highly constrained environments, multipath problems are exacerbated, rendering many acoustic positioning systems unsuitable (Horri, 2020), particularly those that operate in lower frequency bands within which signals suffer less attenuation and therefore echoes dissipate slowly.

2.1 Acoustic positioning

The most widely used underwater positioning technology is based on acoustic triangulation and there are several standard system configurations available for its use in marine field robotics. The main difference between the different setups is the distance between acoustic transponders (termed the baseline) and whether transponders are mounted to the seabed or to a surface ship. For use in highly constrained environments, ultra short baseline (USBL) systems would be the most appropriate as they are designed for lower ranges and do not require transponders to be fixed to the infrastructure. However, typical accuracy of USBL systems is relatively low, 3-5% of the range (Sonardyne, 2023; ADVANCED NAVIGATION, 2022; Sonardyne, 2022), which equates to up to 0.5 m over 10 m, and so is insufficient for the previously mentioned applications, with the additional problem that refresh rates will be relatively slow.

Sonar based simultaneous localisation and mapping (SLAM) is another acoustic technique that is widely reported in the literature (Suresh et al., 2020; McConnell et al., 2022; Ling et al., 2023; Li et al., 2018; Westman et al., 2018; Teixeira et al., 2019). Sonar SLAM in an underwater environment is analogous to LiDAR based SLAM in a terrestrial environment (Oliveira et al., 2021). The technique has been reported to achieve positioning errors of 0.2 m over a

2.5 km trajectory, when used in combination with both IMU and DVL sensors (Ozog et al., 2016). Although this is an improvement it is still insufficient. More importantly that the quoted figures is the fact that the systems are unproven for applications in confined industrial aquatic environments.

2.2 Vision based positioning

The most accurate underwater positioning systems that are available use several cameras, fixed to the perimeter of an environment, to track an array of markers that are fixed to the robot. The commercial underwater motion capture system produced by Qualisys (Qualisys, 2022), for example, achieves sub-centimeter accuracy, low latency and fast refresh rates of 100 Hz. Despite this impressive performance, such systems are typically more suited to lab settings, as they have significant setup, calibration and infrastructure requirements, are highly sensitive to water clarity, as well as having limited volume coverage. Duecker et al. (Duecker et al., 2020) inverted this principle, using a single camera and many marker objects. They placed an array of 63 artificial markers around the perimeter of a tank and used a vehicle mounted camera, combined with AprilTag tracking, to estimate the pose of the camera, which is fixed on an underwater robot. Although cheaper and able to cover a greater proportion of the environment than a system with wall mounted cameras, placing and maintaining many markers at known locations in the environment is not a practicable solution.

Vision-based SLAM using onboard cameras is also a common solution, with at least one commercial product (Vaarst, 2023) and several examples in the literature (Zhou et al., 2022; Campos et al., 2021; Wang et al., 2021; Zhu et al., 2023). Vision-based SLAM depends heavily on recognizing and tracking salient environmental features and reliance on such features is a problem that has been reported when using aerial vehicles in GPS denied environments (Weiss et al., 2011; Chowdhary et al., 2013). As discussed above, vision penetration is reduced underwater, particularly in turbid waters, and causes image sharpness and visibility range to be reduced, which introduces significant challenges when using vision-based SLAM in aquatic applications. Studies have confirmed that the performance of vision-based SLAM used underwater is inferior to that in air due to the low contrast of underwater images. In most cases, features are difficult to extract and are highly dependent on environmental conditions (Zhao et al., 2020).

2.3 Contribution

In this work, a first-of-a-kind collaborative aquatic positioning (CAP) system, which aims to satisfy the requirements defined earlier, is proposed and evaluated experimentally in water tank that is a direct copy of a typical industrial liquid storage tank (see MOVIE 1). The fundamental concept behind the CAP system is inspired by the mother-ship model, used in open oceans, where a surface vehicle with it's own sensor suite is used to help localise a subsurface vehicle. The key difference being that the collaborating surface vehicle is highly mobile and able to move autonomously, staying above the subsurface vehicle. By combining information from sensors that are fixed to both the on the surface and underwater vehicles, and tracking a fiducial marker onboard the underwater vehicle, the position of the underwater vehicle can be determined.

Using a collaborative autonomous surface robot in this way has several benefits. First, the majority of the translation from the origin coordinate system to the underwater vehicle frame is performed in air using an accurate LiDAR based approach; only the direct translation down to the underwater vehicle is performed in water. This means that camera based underwater localisation, which is the fastest and most accurate underwater technique, is appropriate due to the relatively short distances involved. Second, coverage of the environment is almost complete because the camera follows the underwater vehicle and actively keeps it in the field of view. Third, the system does not require any fixed infrastructure, and calibration is as simple as choosing the location of the reference coordinate system origin on the water surface.

Two variants of the CAP system are proposed in this paper: CAP-CPnP, which uses camera based object tracking and a Perspective-n-Point (PnP) algorithm; and CAP-CD, which uses a novel formulation to combine camera based object tracking and a pressure sensor on the underwater vehicle. Both variants are evaluated experimentally during an autonomous underwater mission, with position data from the CAP system being produced in real-time and fed back to

the underwater vehicle (in real-time) to enable an autonomous (waypoint guided) mission.

3 MATERIALS AND METHODS

3.1 Overview

The two variants of the CAP system, namely CAP-CPnP and CAP-CD, share several components such as the estimation of the surface robot’s 6-DOF pose and use of the AprilTag to identify camera pixels that represent the corners of a fiducial marker. The key difference between the two systems is that CAP-CD does not require multiple pixels to be identified at known locations to enable use of a PnP algorithm (Lepetit et al., 2009). Instead, only a single pixel needs to be identified, which broadens the horizon of image processing techniques that can be applied. However, removing the PnP solver means that there is no longer a serial transform chain. Therefore additional sensing as well as a new mathematical formulation are required to enable full and direct calculation of the underwater robot’s pose.

3.2 Hardware architecture - robotic platforms and sensors

The hardware used in this study represents one possible physical incarnation of the positioning systems. As would be expected, the underlying mathematical formulations are agnostic the choice of sensing methods and robots and there are several possible configurations. However, detail has been given below to facilitate understanding of the systems.

The aquatic surface robot used in the proposed positioning system is MallARD (sMall Autonomous Robotic Duck) platform (Groves et al., 2019), which is shown in Figure 1. The dual pontoon configuration of MallARD ensures stability and also creates space at the robot’s centre for sensor payloads. To facilitate locomotion, MallARD is equipped with four bidirectional Blue Robotics T200 thrusters. The thrusters are in a 45-degree configuration relative to the x or y axis, which allows vectoring in the robot’s x and y axes and rotation about the robot’s z -axis. MallARD has an on-board computer and is Robotic Operating System (ROS) (Open Robotics, 2024) enabled. Motion commands are sent from ROS over a serial connection to a control unit (Pixhawk), which generates pulse-width modulation (PWM) signals that are sent to the electronic speed controllers (ESCs). The ESCs in turn provide a phased output to the four brushless motor powered thrusters.

For use in the CAP system, MallARD was modified to include a downward facing low light HD camera and an inertial measurement unit (IMU). The modified robot layout and dimensions are depicted in Figure 1 A and B respectively, while Figure 2 shows the electrical connections of the full system. The camera was included to provide the surface robot with a clear video stream directly beneath the robot to enable tracking of the underwater robot. This camera is a low-light HD USB Camera, which was mounted in a waterproof enclosure in the robot’s central payload area. The IMU was added to enable full pose estimation of MallARD relative to the fixed environment.

The underwater robot utilized in this study is a commercially available BlueROV2, which has been customised and made to run ROS. A fiducial marker, constructed using laser-cut acrylic sheet material, is fixed on top of the BlueROV2, as depicted in Figure 3. This marker is used to track the underwater robot in the field of view of the downward facing camera fixed to MallARD. It is important to note that Additionally, a pressure sensor, positioned at the BlueROV2’s rear and illustrated in Figure 3, has been integrated. This calibrated sensor provides accurate depth measurements (See Supplementary Material, Section 2).

3.3 Mathematical background - Homogeneous transforms

Given \mathbf{p}_X^1 that represents a point labelled $X \in \mathbb{R}^3$ in coordinate frame \mathcal{F}_1 , the coordinates of the same point can be represented in a different coordinate frame \mathcal{F}_0 , given the transform from \mathcal{F}_0 to \mathcal{F}_1 . This coordinate frame transform

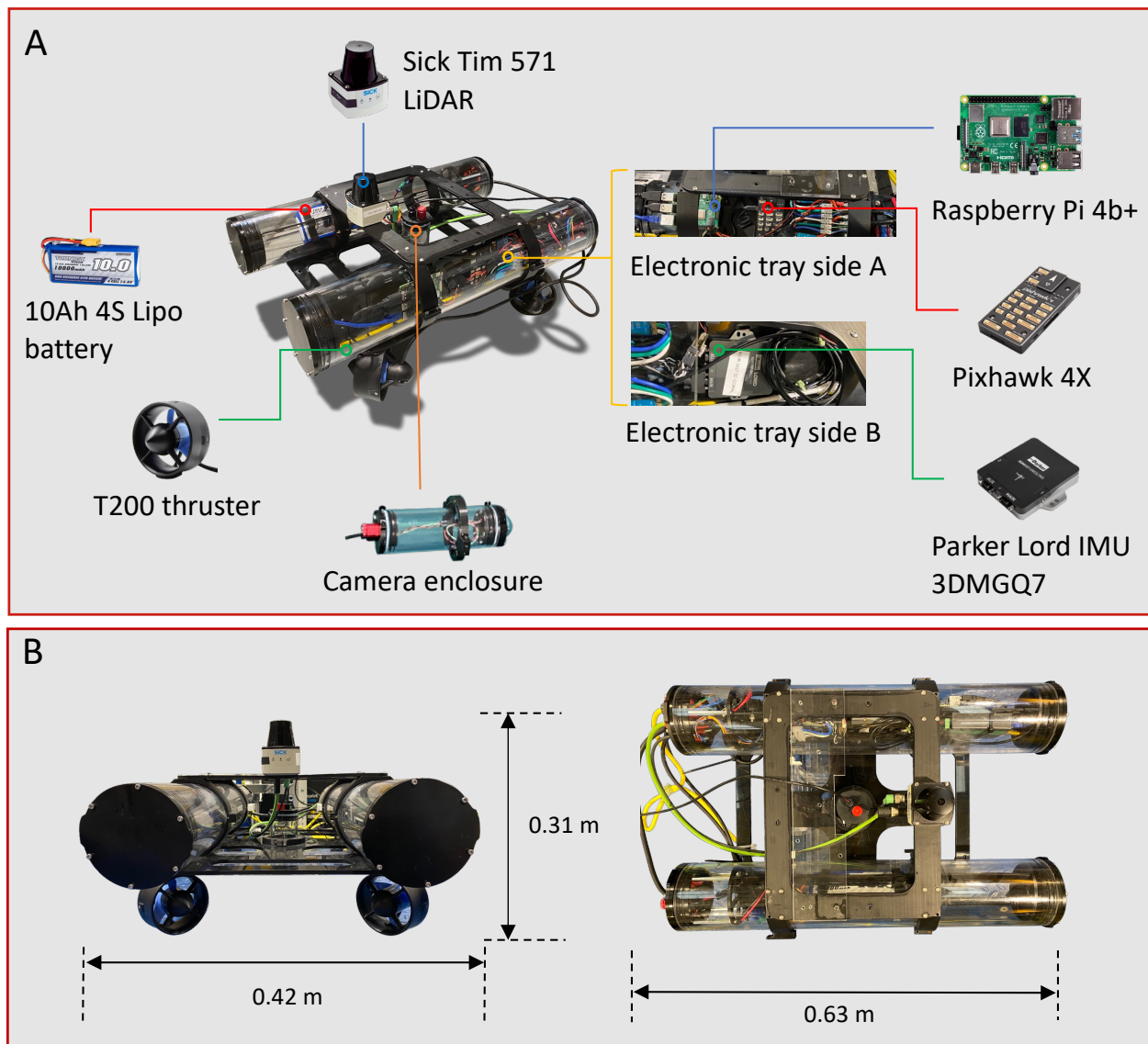


Figure 1: Hardware and system architecture. **A:** MallARD components and their layout. **B:** MallARD 003 platform dimensions

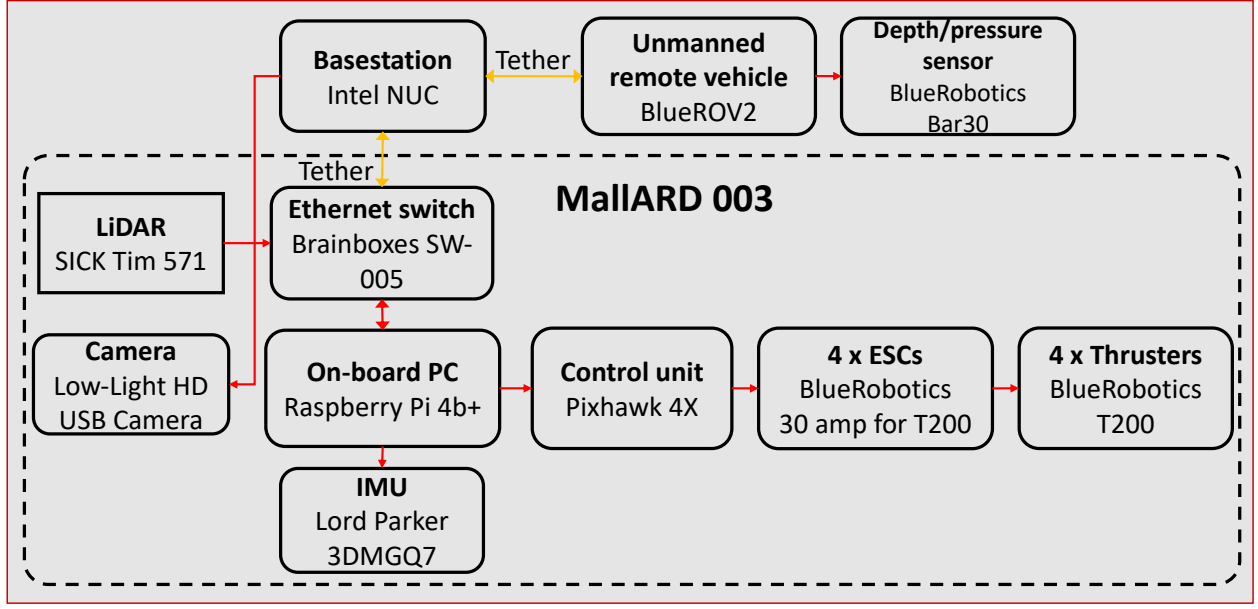


Figure 2: MallARD 003 electronic architecture.

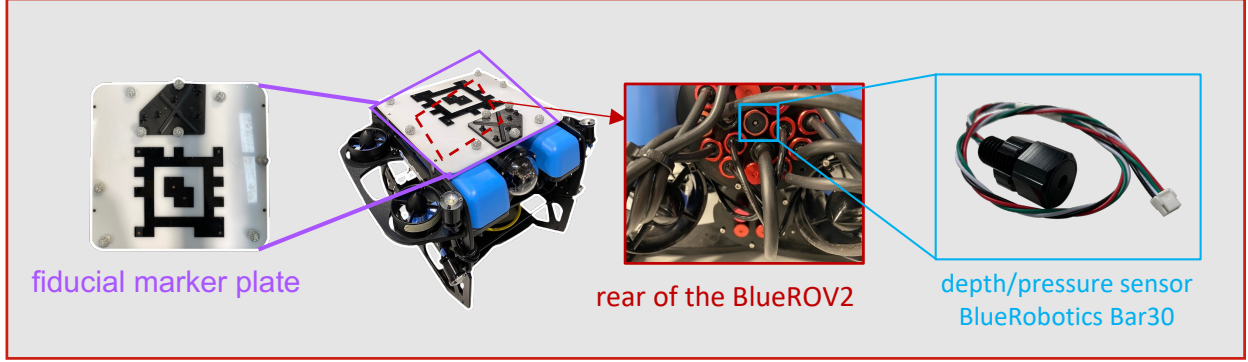


Figure 3: Customised BlueROV2 equipped with a depth/pressure sensor

can be expressed using a 4×4 homogeneous transform matrix \mathbf{H}_1^0 , which represents the pose of \mathcal{F}_1 with respect to \mathcal{F}_0

$$\begin{bmatrix} \mathbf{p}_X^0 \\ 1 \end{bmatrix} = \mathbf{H}_1^0 \begin{bmatrix} \mathbf{p}_X^1 \\ 1 \end{bmatrix}, \quad (1)$$

where

$$\mathbf{H}_1^0 = \begin{bmatrix} \mathbf{R}_1^0 & \mathbf{p}_1^0 \\ 0 & 1 \end{bmatrix} \in SE(3), \quad (2)$$

and $\mathbf{p}_1^0 \in \mathbb{R}^3$ is the translation from the origin of \mathcal{F}_0 to the origin of \mathcal{F}_1 , $\mathbf{R}_1^0 \in SO(3)$ is the rotation matrix from \mathcal{F}_0 to \mathcal{F}_1 . Homogeneous transforms can be formed into serial chains. For instance, if there is a third coordinate frame \mathcal{F}_2 and the transform from \mathcal{F}_1 to \mathcal{F}_2 is given by \mathbf{H}_2^1 , the transform from \mathcal{F}_0 to \mathcal{F}_2 can be obtained by right multiplication of the transform chain, in order from start frame to end frame:

$$\mathbf{H}_2^0 = \mathbf{H}_1^0 \mathbf{H}_2^1. \quad (3)$$

3.4 Coordinate frames

The coordinate systems involved in the design of each part of the CAP system are shown in Figure 4A. The full positioning system is composed of the following coordinate frames: world frame \mathcal{F}_W , MallARD baselink frame \mathcal{F}_B , IMU frame \mathcal{F}_I , camera frame (monocular) \mathcal{F}_C and marker frame \mathcal{F}_M . The world frame (\mathcal{F}_W) origin is assigned to a corner of the testing tank. MallARD baselink (\mathcal{F}_B) is the geometric centre of MallARD. The IMU frame (\mathcal{F}_I) is attached to the IMU and has a fixed transform from \mathcal{F}_B . The origin of the camera frame (\mathcal{F}_C) is located at the optical centre of the camera lens and also has a fixed transform from \mathcal{F}_B .

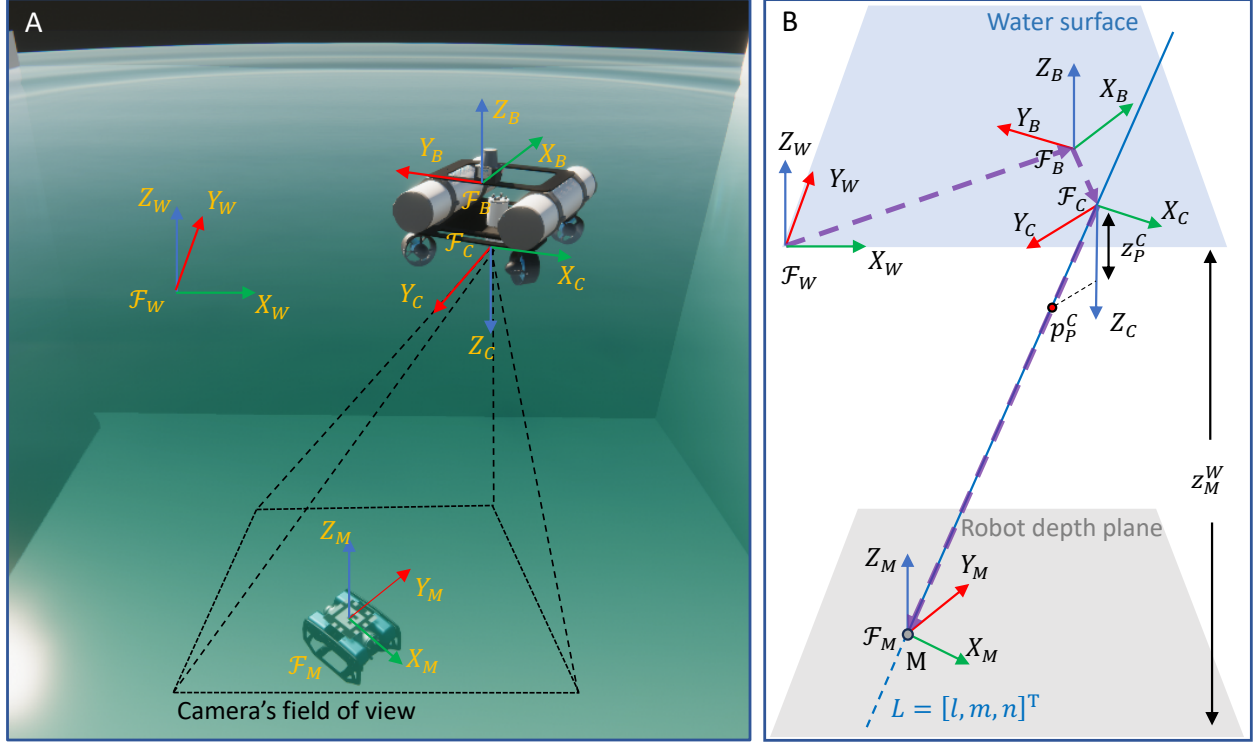


Figure 4: **A:** Overview of the proposed CAP system with coordinate systems. **B:** Coordinate frames, depth plane and Plücker line employed in the CAP-CD formulation.

3.5 CAP-CPnP formulation

The CAP system aims to determine \mathbf{p}_M^W : the position of the origin of the marker frame \mathcal{F}_M in the world frame \mathcal{F}_W . In the CAP-CPnP formulation, all elements of the serial transform chain can be determined independently; therefore, using Equations (1) and (3), \mathbf{p}_M^W can be calculated directly:

$$\begin{bmatrix} \mathbf{p}_M^W \\ 1 \end{bmatrix} = \mathbf{H}_B^W \mathbf{H}_C^B \begin{bmatrix} \mathbf{p}_M^W \\ 1 \end{bmatrix}. \quad (4)$$

The terms on the right hand side of Equation (4) can be determined as follows. From Equation (2), \mathbf{H}_B^W is composed of \mathbf{R}_B^W and \mathbf{p}_B^W . \mathbf{R}_B^W can be calculated using data from the IMU and 2D SLAM system on the surface vehicle, as detailed in Section 3.9. \mathbf{p}_B^W can be constructed from the x and y components of the 2D SLAM output (Section 3.8), while the z component is assumed to be a static value that represents the offset between the world frame and MallARD's body frame. Further detail regarding the method of obtaining \mathbf{R}_B^W and \mathbf{p}_B^W is given in the Section 3.8 and Supplementary Materials, Section 1.2 respectively. \mathbf{H}_C^B is a measured static transformation from the ASV body frame \mathcal{F}_B to the camera frame \mathcal{F}_C . \mathbf{p}_M^C is the position of the marker frame (\mathcal{F}_M) origin in the camera frame (\mathcal{F}_C) and is calculated

using a fiducial marker tracking technique, which is a two step process. First, the camera image is processed using AprilTag to detect the four pixel locations that relate to the four corners of a fiducial marker in the camera's image. Second, the four pixel locations, together with the marker's dimensions and the camera's intrinsic matrix are used to determine the 6 DOF pose of the fiducial marker in the camera frame, from which \mathbf{p}_M^C can be extracted. This is a standard problem known as Perspective-n-Point (PnP) and has several available solutions, for example Direct Linear Transformation (Hartley and Zisserman, 2003) and Efficient PnP (Lepetit et al., 2009). Further detail regarding fiducial tracking and PnP usage is provided in the Supplementary Materials, Section 2.

Camera tracking systems have inherent noise due to their derivation from (necessarily) pixelated camera images. Noise is exacerbated when 3-DOF translations are derived using Perspective-n-Point (PnP) methods (Lensgraf et al., 2021). This is primarily due to the low sensitivity of PnP solutions to depth variations (z_M^C). Changes in depth cause relatively small changes to the image and, in turn, have a lesser effect on changing the pixels that are identified as corners of objects, causing low relative sensitivity and ambiguity. Noise and ambiguity in z_M^C mostly affects z_M^W , due to the fact that their associated axes are generally well aligned, however, noise in z_M^C also translates onto x_M^W and y_M^W when the ASV pitches and rolls. To overcome the shortcomings of CAP-CPnP, CAP-CD is proposed, which does not require the use of a PnP solver and, instead, incorporates a depth sensor onboard the underwater robot.

3.6 CAP-CD formulation

The CAP-CD formulation does not use a PnP solver (or equivalent) and, therefore, \mathbf{p}_M^C is undefined, breaking the transform chain. However, with the inclusion of a pressure sensor on the underwater robot, that can be calibrated to measure water depth, it is possible to directly calculate \mathbf{p}_M^W .

Figure 4B gives a graphical representation of the method. Consider a Plücker line that passes through the origins of \mathcal{F}_C and \mathcal{F}_M , and a horizontal plane defined by the depth sensor measurement z_M^W . By finding the intersection between the Plücker line and the horizontal plane, \mathbf{p}_M^W can be calculated.

The Plücker line is defined by two points in the world frame \mathcal{F}_W . The first point \mathbf{p}_C^W is the origin of \mathcal{F}_C which can be found using

$$\begin{bmatrix} \mathbf{p}_C^W \\ 1 \end{bmatrix} = \mathbf{H}_B^W \begin{bmatrix} \mathbf{p}_C^B \\ 1 \end{bmatrix}. \quad (5)$$

Since the origin of \mathcal{F}_M is unknown, another point on the line must be found for the line can be defined. To find this second point, AprilTag is used to identify the camera pixel locations that represent the corners of the fiducial marker; these are averaged to give the pixel location of the centre of the marker: u_p and v_p . The camera's intrinsic matrix is then used to identify the components x_P^C and y_P^C of a projected point \mathbf{p}_P^C . According to the definition of the intrinsic matrix, $z_P^C = 1$ for all cases. Therefore \mathbf{p}_P^C can be identified as follows:

$$\mathbf{p}_P^C = \begin{bmatrix} x_P^C \\ y_P^C \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} u_{centre} \\ v_{centre} \\ 1 \end{bmatrix}, \quad (6)$$

where the 3×3 matrix is the camera's intrinsic matrix. \mathbf{p}_P^C lies on the the Plücker line that passes through the origin of \mathcal{F}_C and \mathcal{F}_M . However, for the Plücker line to be defined in \mathcal{F}_W the point must be transformed into \mathcal{F}_W , the world frame:

$$\begin{bmatrix} \mathbf{p}_P^W \\ 1 \end{bmatrix} = \mathbf{H}_B^W \mathbf{H}_C^B \begin{bmatrix} \mathbf{p}_P^C \\ 1 \end{bmatrix}. \quad (7)$$

Now, given \mathbf{p}_C^W and \mathbf{p}_P^W the Plücker line can be defined in the world frame. In general, the equation of a line with direction vector $\mathbf{l} = [l, m, n]^\top$ that passes through the point $[x_1, y_1, z_1]^\top$ is given by the formula

$$\frac{x - x_1}{l} = \frac{y - y_1}{m} = \frac{z - z_1}{n} = k, \quad (8)$$

where k ranges over all real numbers and represents the position on the line. By defining

$$[l, m, n]^\top = [x_C^W - x_P^W, y_C^W - y_P^W, z_C^W - z_P^W]^\top \quad (9)$$

the Plücker line can be expressed as:

$$x = x_P^W + (x_C^W - x_P^W) k, \quad (10)$$

$$y = y_P^W + (y_C^W - y_P^W) k, \quad (11)$$

$$z = z_P^W + (z_C^W - z_P^W) k. \quad (12)$$

Since z_M^W can be found directly from the calibrated pressure sensor measurement, the value required for k which effectively identifies the intersection between the horizontal plane given by z_M^W and the Plücker can be computed by:

$$k = \frac{-z_M^W - z_P^W}{z_C^W - z_P^W}. \quad (13)$$

Therefore, the tag's unknown coordinates x_M^W and y_M^W can be found by substituting k back into Equation (10) and (11).

3.7 Autonomous following

For the CAP system to function, the marker on the underwater robot must be within the field of view of the surface vehicle's downward facing camera. To achieve this a range of control techniques could be applied and in this work visual servoing was implemented (Chaumette, 2004; Yao et al., 2023).

Initially, four target feature points that represent the desired position of the corners of the fiducial marker are defined on the projected image plane. The marker tracking system (AprilTag) then continuously detects these four points and compares them with the corresponding target projection points. The aim of the visual servoing system is to minimize the difference between the desired and tracked positions. This difference is translated into how the surface robot should move to ensure that the detected points match (or fall within an acceptable range of deviation) the target feature points, thereby enabling the surface robot to automatically follow the underwater robot.

3.8 MallARD's 2D SLAM system

MallARD is equipped with a waterproof 2D LiDAR, enabling planar localisation relative to the pool walls. Because the 2D LiDAR operates in a plane which is parallel to the water surface, the SLAM system provides positions x_B^W and y_B^W as well as yaw angle ψ_B^W on that plane. Because in this application there is no reliable source of odometry, the choice of 2D slam algorithms is limited. In the current work, a customised version of Hector mapping is used as the SLAM framework as this does not require any odometry (Kohlbrecher et al., 2011b). The customisations made allow the map to be locked, preventing corruption in longer trials, and allows the position and rotation to be output relative to a fixed location in the pool. By default the SLAM output is with reference to the start location, which is not practical or repeatable.

Upon immersion and activation in the aquatic environment, MallARD transmits LiDAR scans to the SLAM software. To build a 2D map of the pool's walls, MallARD is manually navigated around the pool using a joystick. Once the mapping phase reaches completion, the map is fixed and MallARD's autonomous following mode is initiated. During this phase, the primary function of the SLAM algorithm is localisation, given that the spatial map undergoes no changes, as depicted in Figure 5(A).

3.9 MallARD's rotation relative to the world frame \mathbf{R}_B^W

As MallARD navigates through water, it undergoes roll and pitch due to resultant hydrodynamic forces and small waves on the water surface (Figure 5 B(i) and B(ii)). For both CAP-CD and CAP-CPnP, it is necessary to know

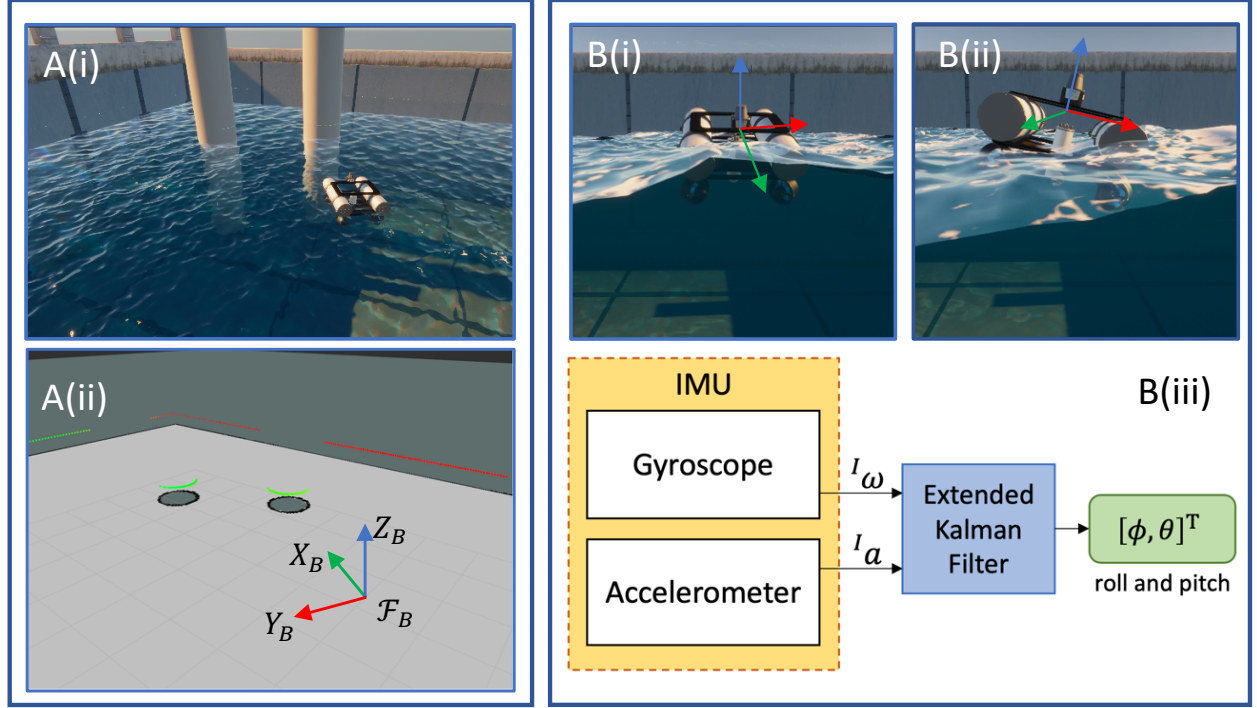


Figure 5: Simulated scenarios to aid description of self-localisation and tilting analysis of the surface robot: **A(i)**: Mallard in a simulated aquatic environment, and **A(ii)**: The SLAM system visualisation including the multi coloured dots of the laser scan, the black lines of the map and a body frame pose estimate. **B(i)** and **B(ii)**: Tilting of the surface robot due to surface waves, and **B(iii)**: Schematic diagram of Extended Kalman Filter inputs and outputs.

Mallard's body frame rotation relative to the world coordinate system \mathbf{R}_B^W . For mathematical convenience \mathbf{R}_B^W is calculated using Euler angles in the Z-Y-X sequence and then converted to a rotation matrix. While the issue of gimbal lock is a known problem when using Euler angles, it is unlikely to occur in this case because rotations about the y-axis generally remain within approximately 10 degrees of zero.

To calculate \mathbf{R}_B^W in the Z-Y-X Euler angle form, the rotation about z-axis is decoupled from the rotations about y-axis and X-axis. Although Mallard's IMU has a built in 3-axis compass, the compass-provided measurements are unreliable due to the magnetic fields generated by metallic structures and Mallard's own electronic equipment and motors. Therefore, the magnetometer cannot provide a lock for the yaw measurement (rotation about Z-axis in the Euler sequence). However, yaw can be acquired through LiDAR-based SLAM and this is used as the z-axis component of the sequence. This approach is valid because z is the first rotation in the sequence and is therefore about the z -axis of \mathcal{F}_W . Mallard's roll and pitch (y-axis and x-axis rotations in the Euler sequence) must now be computed relative to the stabilised body frame (Kohlbrecher et al., 2011a), which is a version of the body frame without any roll or pitch.

The tilting EKF (Welch and Bishop, 1995) is used to find the second two rotations in the sequence. As shown in Figure 5 B(iii) the EKF takes two vector inputs, which are 3-axis angular rate $\boldsymbol{\omega} = [\omega_x, \omega_y, \omega_z]^T$ measured by the gyroscope and 3-axis acceleration $\mathbf{a} = [a_x, a_y, a_z]^T$ measured by the accelerometer, and outputs the rotations, which are pitch(θ) and roll(ϕ). In this research, the motion of the USV does not exhibit prolonged substantial accelerations (other than gravitational acceleration) for an extended period of time. Therefore, it is assumed that the acceleration vector is identical to the gravity vector. Using the full Euler angle Z-Y-X sequence \mathbf{R}_B^W can be transformed to a rotation matrix using the following equation:

$$\mathbf{R}_B^W = \begin{bmatrix} c\psi c\theta & c\psi s\theta s\phi - s\psi c\phi & c\psi s\theta c\phi + s\psi s\phi \\ s\psi c\theta & s\psi s\theta s\phi + c\psi c\phi & s\psi s\theta c\phi - c\psi s\phi \\ -s\theta & c\theta s\phi & c\theta c\phi \end{bmatrix} \quad (14)$$

4 RESULTS

4.1 Experiment setup

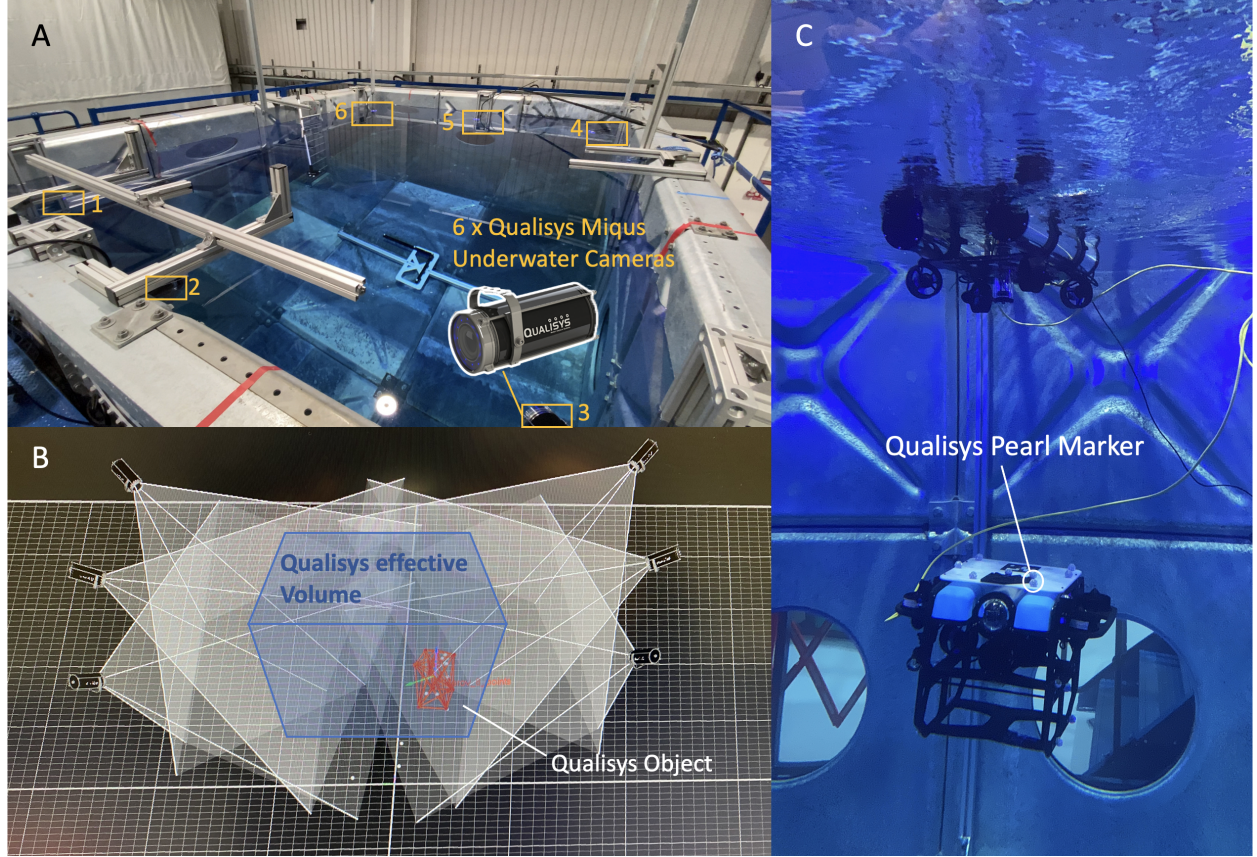


Figure 6: Experimental field and setup. **A:** Overview of the experimental tank. **B:** Qualisys system setup and effective volume **C:** MallARD and BlueROV2 deployed in the experimental pond. BlueROV2 is mounted with pearl marker for Qualisys system tracking.

The CAP system was evaluated using data collected in the $4.8 \times 3.6 \times 2.0$ m (length, width, depth) indoor test tank shown in Figure 6A. The positioning accuracy of the system was validated using a high-accuracy, 6 camera Qualisys Miquis M5 underwater motion tracking system (Qualisys, 2024). The submerged Qualisys cameras, as shown in Figure 6A, were fixed to the walls of the tank. Due to field of view limitations, the Qualisys system could not cover the entire tank's volume, and as a consequence the experiments were conducted in a smaller region of the tank, as illustrated in Figure 6B. Qualisys tracking markers were placed on the BlueROV2 and on a customised marker plate that was used to allow the Qualisys object frame to be accurately located on to the BlueROV2. Figure 6C shows the BlueROV2 with Qualisys pearl markers attached. When calibrated, the accuracy of the Qualisys system over the effective volume was 1 mm with regard to position and 0.1° for rotation (Qualisys, 2024) (Qualisys calibration is detailed in Supplementary Materials, Section 5).

BlueROV2, MallARD, and the basestation all utilize ROS, enabling real-time data sharing and synchronization to a

single clock, specifically the clock of the basestation. All sensor data was generated and processed in real-time and recorded on the basestation. Real-time pose data from the Qualisys system were also bridged into the ROS system and recorded on the basestation.

4.2 Experimental validation in the test tank

The two positioning systems described in Section 3 were evaluated in this work, namely CAP-CD and CAP-CPnP. To evaluate the performance of CAP system, the underwater robot was programmed to move autonomously along three pre-programmed trajectories. These trajectories were: square, random and lawnmower pattern. To comprehensively assess the impact of depth variations on the system, the depth of the underwater robot was varied by up to 1 m as it followed the pre-programmed trajectories. Concurrently, the surface robot (MallARD) autonomously followed the underwater robot, ensuring that the fiducial marker, fixed on the underwater robot, remained within the field of view of the downward-facing camera on the surface robot. The validation experiment collected three sets of data for each of the square, random, and lawnmower patterns with relatively large depth variations (up to 1 m), with each dataset lasting for 120 seconds. Furthermore, the datasets for each of the square, random, and lawnmower patterns, were collected with minor depth fluctuations (up to approximately 0.3 m).

Figure 7A displays overlaid snapshots of the real-time positioning trajectory of the underwater robot over a duration of 21 seconds. Figure 7B shows the positioning performance of CAP-CD and CAP-CPnP, corresponding to the results in Figure 7A. These overlaid snapshots and trajectories are presented for clarity and conciseness. Beyond the fixed-depth square trajectory illustrated, the performance of CAP-CD and CAP-CPnP under various more complex trajectories is detailed in Figure 9 and Table 1. Figures 8A(i), B(i), and C(i) illustrate the translation of the underwater robot in the world-fixed frame, estimated using both CAP methods in comparison with the ground truth, for the X , Y , and Z axes respectively. In Figure 8A(i) and A(ii), it is evident that the positioning error of CAP-CD increased during changes in the underwater robot's direction of motion, both in the X and y axes, presented as fluctuations within the graphs. This phenomenon occurred because CAP-CD assumed the depth measured by the depth sensor was at the centre of rotation of the underwater robot. However, the sensor was actually located towards the rear of the robot and changed when the robot's motion caused the robot to tilt. In contrast, the ground truth, which tracked the centre of the robot, was minimally impacted by this tilt. The depth sensor, positioned at the BlueROV2's rear, registered significant depth alterations due to the BlueROV2's inclination, which lead to observable positioning fluctuations along the X - Y plane. Similarly, since the AprilTag was located towards the rear of the robot rather than at the centre, CAP-CPnP experienced the same issue.

In terms of the positioning results from the x , y , and z axes, the Euclidian RMSE of CAP-CD was slightly lower than that of CAP-CPnP, shown in 8A(ii)(iii), B(ii)(iii) and C(ii)(iii). The Euclidean RMSE of CAP-CD over the 120 s was 70.2 mm, while that of CAP-CPnP was slightly higher at 100.3 mm. Table 1 shows the full breakdown of the results for each dataset. The results indicate that across a variety of trajectories, the accuracy of CAP-CD surpasses that of CAP-CPnP in the x , y , and z axes, respectively. Consequently, the Euclidean RMSE of CAP-CD is lower than that of CAP-CPnP. The Euclidean RMSE for CAP-CD is concentrated between 90 mm to 130 mm, with the highest value reaching 123.4 mm.

To demonstrate consistency between the results from different experiments, results of the CAP system operating on an underwater robot in square, lawnmower, and random patterns (accompanied by variations in depth) are shown in the Figure 9. More details and plots of the results for each dataset can be found in Supplementary Materials, Section 3 and MOVIE 3, 4, 5 and 6.

4.3 CAP system in waters of varying turbidity

The accuracy of positioning in waters with varying levels of turbidity is an important metric for underwater positioning systems. To address this, a further set of experiments were designed whereby a fiducial marker was laminated and fixed in the middle of the underwater test volume, and the water's turbidity altered by adding talcum powder. During the experiments, the water's turbidity was adjusted to 0.12 NTU (Nephelometric Turbidity Units), 2.58 NTU, and

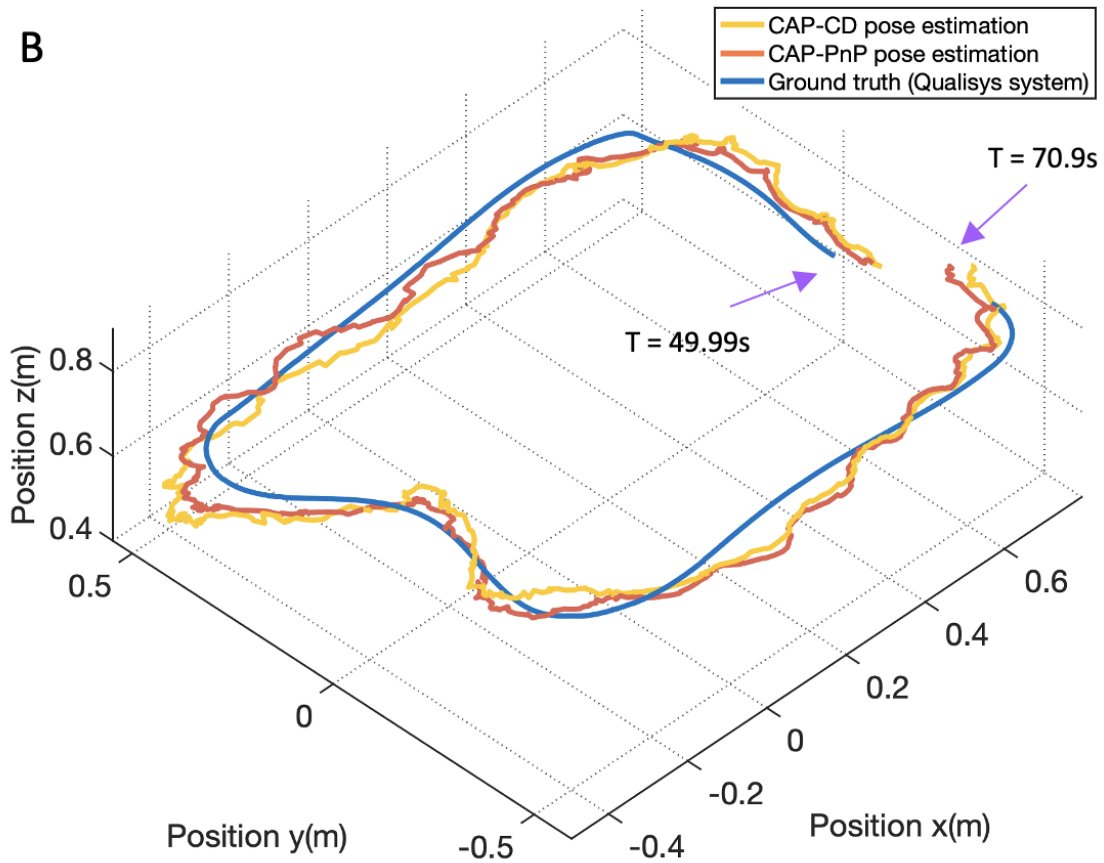
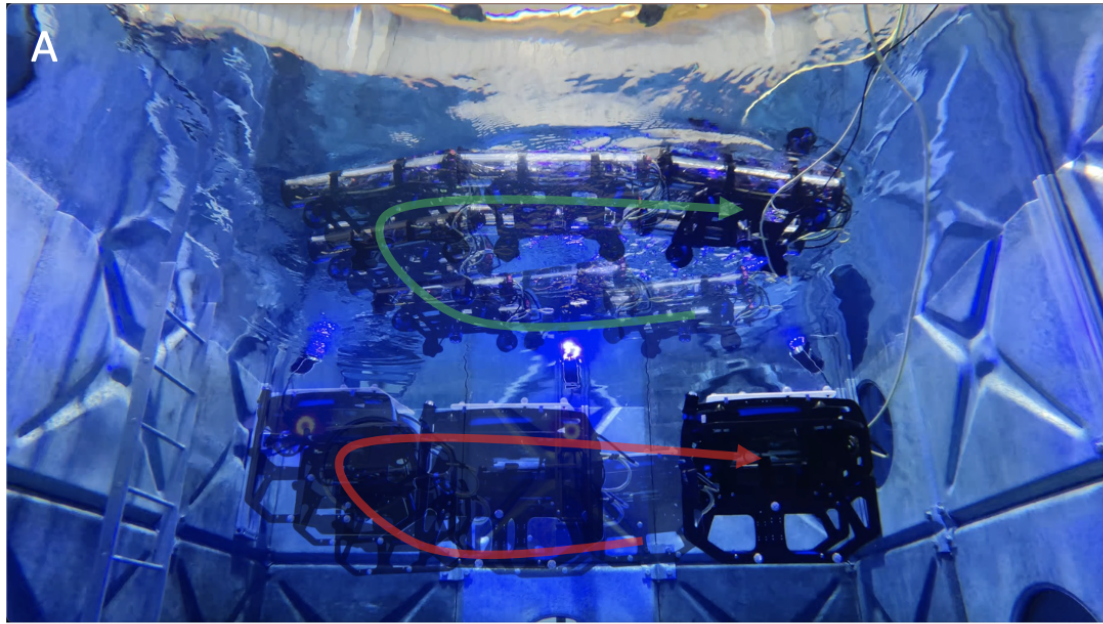


Figure 7: Overview of CAP system during testing. **A:** Overlaid snapshots of CAP system; the trajectories of MallARD and BlueROV2 are shown as the green and red path. **B:** 3D trajectory of BlueROV2 estimated by CAP-CD and CAP-CPnP against ground truth overtime respectively. BlueROV2 and MallARD operating in autonomous mode. The BlueROV2 is programmed to move in a specific trajectory. Meanwhile, the two robots of the CAP system operate in a leader-follower configuration (BlueROV2 as leader and MallARD as follower).

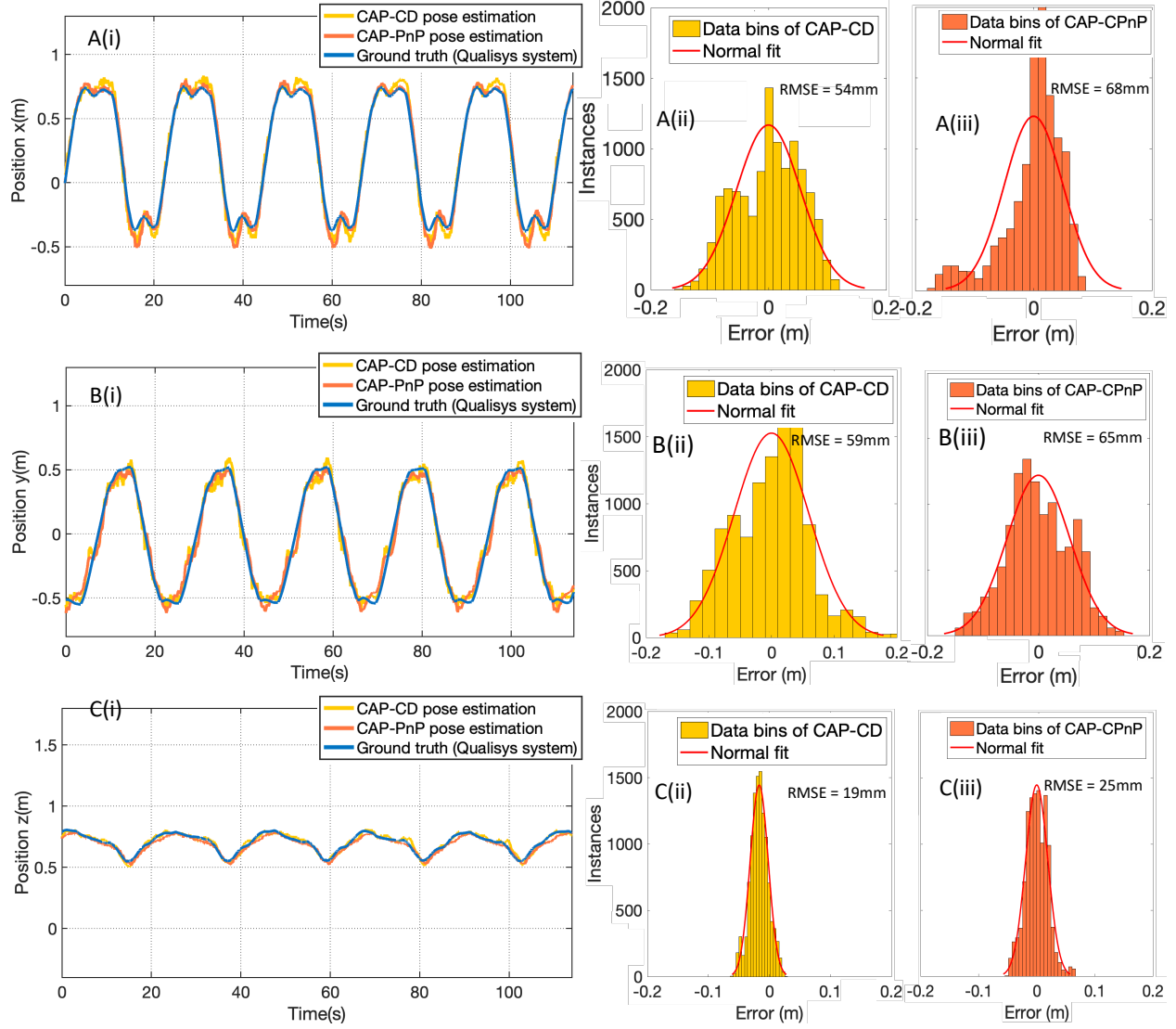


Figure 8: **The positioning results of the CAP system from the X, Y, and Z axes.** A(i), B(i) and C(i) The positioning of BlueROV2 by CAP-CD and CAP-CPnP in comparison with the ground truth along the X, Y, and Z axes, respectively. A(ii), B(ii) and C(ii) The error histograms of CAP-CD on the X, Y, and Z axes. A(iii), B(iii) and C(iii) The error histograms of CAP-CPnP on the X, Y, and Z axes.

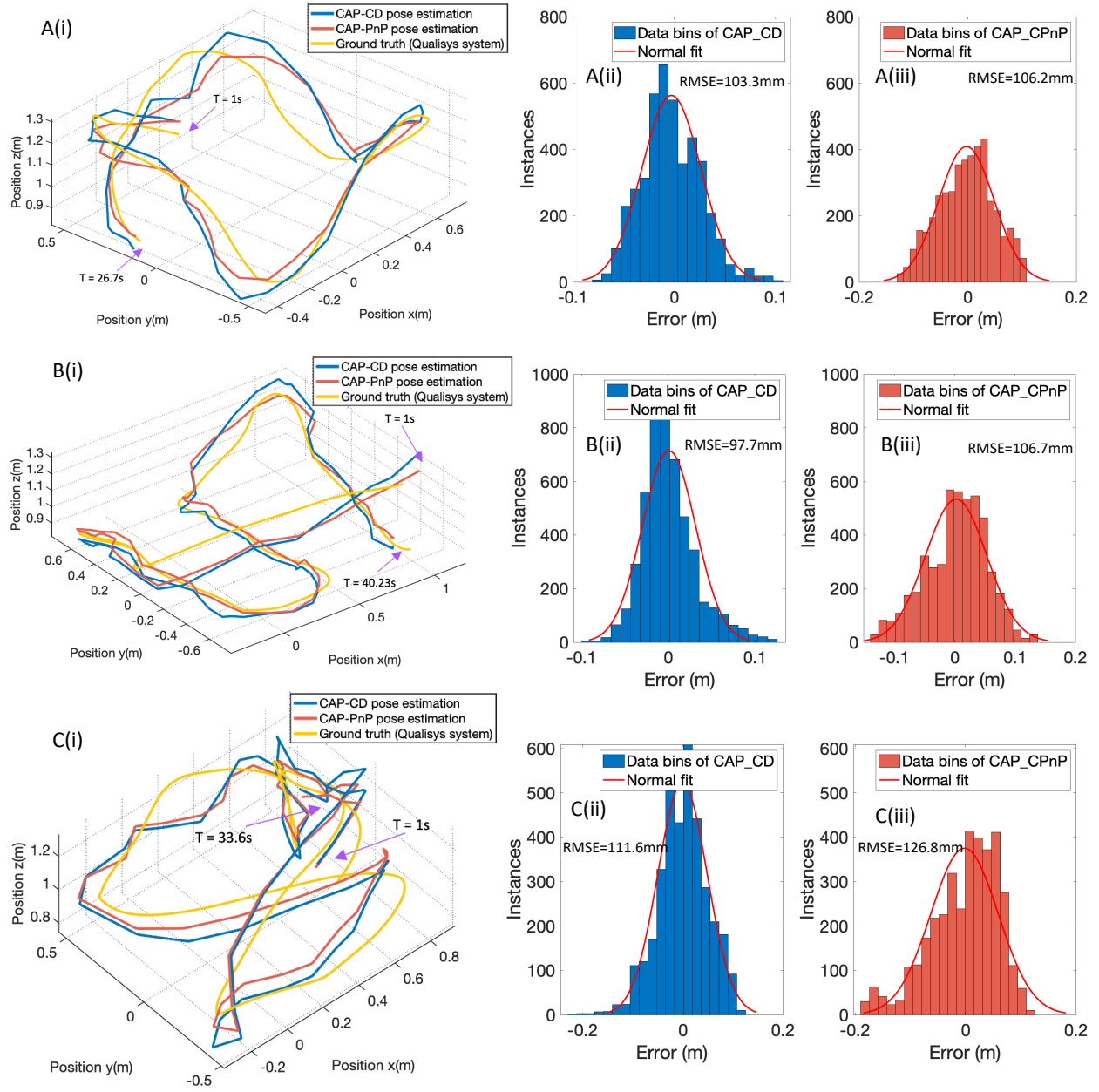


Figure 9: **The positioning results of the CAP system while the underwater robot moves in different patterns.** Figures A(i), B(i), and C(i) respectively show the trajectory plots of the CAP system positioning underwater robots operating in square, lawnmower, and random patterns. Figures A(ii), B(ii), and C(ii) present the error histograms of CAP-CD for Euclidean distance. Figures A(iii), B(iii), and C(iii) display the error histograms of CAP-CPnP for Euclidean distance.

Table 1: Comparative performance of CAP-CD versus CAP-CPnP across various datasets and trajectory types.

Trajectory	Dataset No.	Method	X RMSE (mm)	Y RMSE (mm)	Z RMSE (mm)	Euclidean RMSE (mm)
Square	1	CAP-CD	66.2	61.6	24.1	103.3
		CAP-CPnP	55.6	58.8	61.7	106.2
	2	CAP-CD	67.7	49.1	63.0	99.3
		CAP-CPnP	76.9	45.3	82.2	102.4
	3	CAP-CD	65.0	51.7	52.4	99.2
		CAP-CPnP	59.2	49.9	71.8	92.5
Lawnmower	4	CAP-CD	55.2	76.9	29.0	98.7
		CAP-CPnP	76.9	83.2	65.8	106.7
	5	CAP-CD	58.2	73.7	30.8	87.3
		CAP-CPnP	46.8	71.5	62.2	94.5
	6	CAP-CD	56.5	68.5	36.0	85.5
		CAP-CPnP	45.5	65.2	62.0	92.1
Random	7	CAP-CD	104.4	89.2	45.3	116.1
		CAP-CPnP	100.1	89.1	75.2	126.8
	8	CAP-CD	80.4	88.6	48.3	117.7
		CAP-CPnP	74.9	85.2	68.5	122.7
	9	CAP-CD	87.4	78.9	59.4	123.4
		CAP-CPnP	83.4	81.4	80.2	134.2

3.74 NTU. The YSI ProDSS water quality meter(Van Walt, 2024) was used to measure turbidity and this probe was calibrated using deionized water. At each turbidity level, three sets of experiments were conducted at different depths: 0.9 m, 1.4 m, and 1.9 m. Simultaneously, at each depth, the surface robot was programmed to move along three trajectories: a square (with the surface robot performing a 90-degree turn at each corner of the square), a hexagon, and a lawnmower pattern (see MOVIE 2). The outputs of the CAP system and all sensors were recorded throughout these tests.

In these tests, the camera could clearly detect the AprilTag at turbidity levels ranging from 0.12 NTU to 2.58 NTU, as shown in Figure 10. The ability to detect the underwater fiducial marker correlates to two factors: turbidity and the distance between the camera and the target tag. The confidence in tag detection is quantified by the decision margin. As might be expected, the decision margin demonstrates a negative correlation with both water turbidity and the distance between camera and tag, as shown in Figure 10B(i), B(ii) and B(iii). As turbidity increased to 3.74 NTU, a significant reduction in the decision margin was observed. With turbidity at 3.74 NTU and the distance set to 1.9 meters, the efficacy of tag detection diminished considerably, although there were sporadic instances of successful detection under these circumstances.

5 DISCUSSION

The positioning of robots in constrained underwater environments introduces a range of challenges that are typically not encountered in open marine environments. Perhaps the most significant of these is that in a constrained environment, an underwater robot may require far greater positional accuracy than would be required in a marine environment. Whilst it is possible to design underwater positioning systems for environments where infrastructure, such as cameras or markers, can be added, or where there are a rich set of features, this research has presented a collaborative system

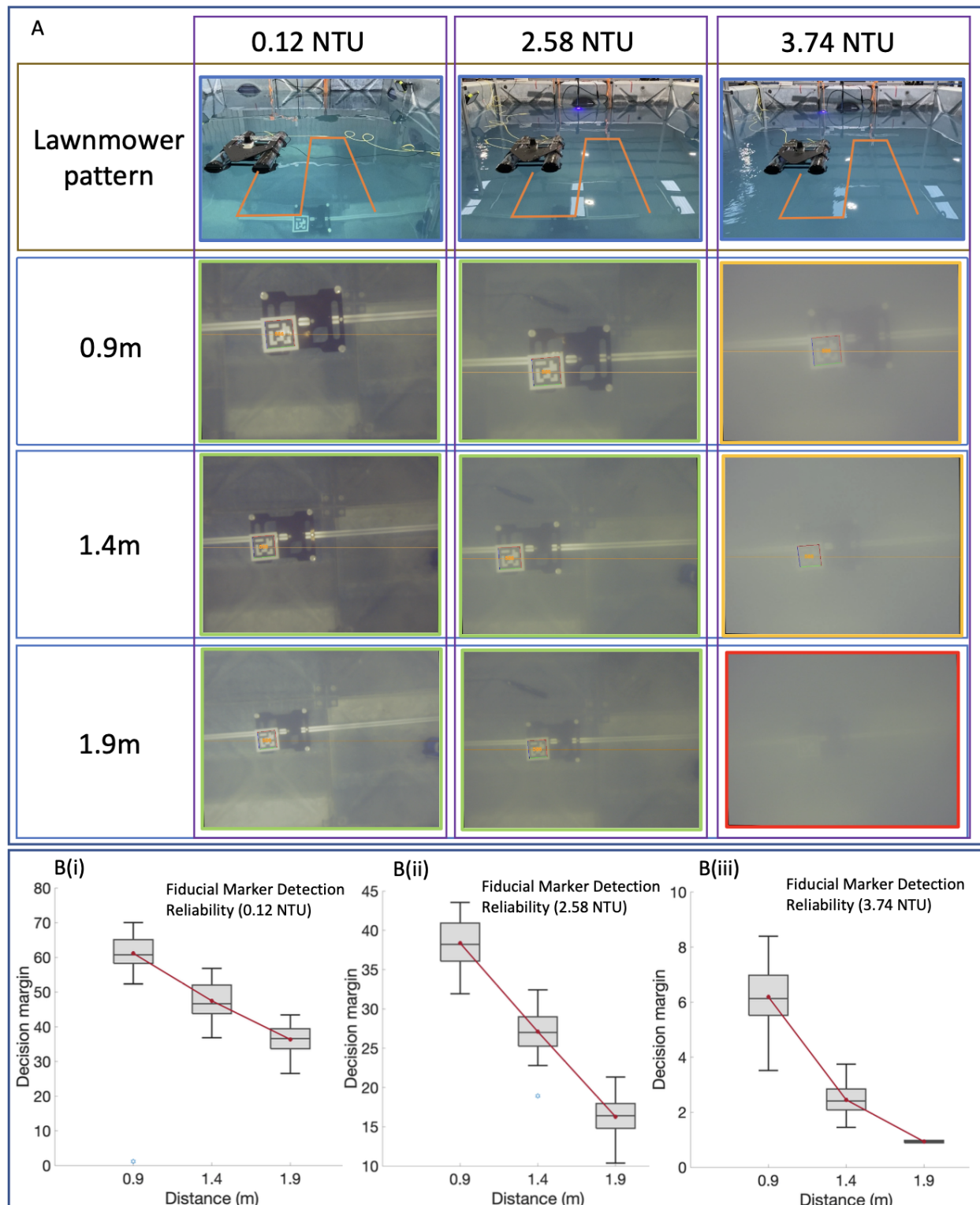


Figure 10: **CAP system in waters of varying turbidity.**(A) In situations where an underwater AprilTag remains stationary while the surface robot moves, and under varying depths as well as different turbidity levels of water, the downward-facing camera's recognition performance of the fiducial marker. **B(i)**, **B(ii)** and **B(iii)** are the box plots of the decision margins detected by fiducial detection under three different levels of turbidity.

that provides underwater positioning in confined or otherwise constrained underwater environments, without the need for infrastructure, system calibration or the environment to be feature rich. This capability marks a significant step towards enabling repeatable and reliable autonomous robotic missions in such challenging conditions. Comparative analysis indicates that the system's performance aligns well with the set objectives, offering a viable solution for the precise navigation and positioning that is required for successful underwater explorations and tasks.

An important point to note regarding the differences between the two CAP formulations is that, in addition to improved accuracy, the CAP-CD formulation is more flexible. The PnP element of the CAP-CPnP formulation relies on four corners of a geometric tag of known dimensions being identified. This is somewhat limiting, in that a fiducial marker tracking system must generally be employed. On the other hand, the CAP-CD system only requires a single point in a projected plane. This means that the system is open for use with other tracking systems; for instance, those based on deep learning, such as YOLO (Redmon et al., 2016) or fast RCNN (Girshick, 2015) which could track the robot without the need for a fiducial marker. This would have the added benefit of not requiring the underwater vehicle to be locked in roll and pitch, to maintain visibility of the fiducial marker.

5.1 Limitations of the study

A current limitation is that since the underwater portion of the proposed method is vision-based, it inherits some common issues of optical positioning; for instance, ambient lighting and turbidity. The experimental setup for the turbidity test, conducted within a water tank with a depth of 2.4 m limits the applicability of the results obtained. To fully assess the capabilities of the CAP system in locating the robot, it would be necessary to conduct measurements in environments that are deeper than the current experimental setup. However, the turbidity study demonstrates that tag detection has a reasonable degree of tolerance to turbid water.

Although the use of the positioning system as feedback to enable autonomous underwater missions has confirmed that the CAP system can operate successfully in real-time, this does not fully address the issue of temporal synchronization among multiple sensors. Instead the current system relies on using the most recent update from each sensor.

The current system provides the Cartesian position of the underwater robot in 3-DOF. However, the system could be expanded to cover 3-DOF rotation with relative ease by re-using components that were developed to determine the surface vehicle's 3-DOF rotation.

An important practical limitation of the setup is that it requires an open expanse of water surface above the underwater environment to allow deployment of the surface vehicle. In addition, the current system, with only one collaborating robot, can only operate to a depth where the fiducial marker can be tracked. Besides, as the self-localisation of surface robots relies on LiDAR-based SLAM, it necessitates that the environment above the water surface contains features within the range of LiDAR.

5.2 Possible extensions

The CAP positioning system has been designed to be applicable for a broad range of restricted underwater environments, hence the consideration of water turbidity and insufficient ambient lighting conditions. However, to expand its capabilities further, future work will involve using acoustic sensors, such as a short range multibeam sonar, to locate the underwater robot rather than the optical cameras. It is anticipated that this approach should enable the positioning system to have improved capability in highly turbid environments. To extend the capabilities of the proposed system further, it is feasible that multiple underwater vehicles could collaborate, enabling one underwater vehicle, in light of sight of the surface vehicle, to position a second underwater vehicle, not within light of sight of the surface vehicle. This could allow the range, in terms of depth, to be extended, and the system to be used for navigation of highly constrained environments where there is limited, or no line of sight between the submersible robot and surface of the water. For CAP-CD, currently, the plane in which the underwater robot, as defined by the depth sensor, resides does not take into account the pitch and roll rotations of the underwater robot. Once rotation occurs, the plane of the robot's depth will not be equivalent to the plane in which the depth sensor is located. To address this, the pose of the

underwater robot can be obtained through an IMU, thereby acquiring accurate depth information. Finally, the use of deep learning to process sonar images, as well as camera images will be investigated.

References

- ADVANCED NAVIGATION (2022). Subsonus USBL/INS. Retrieved January 12, 2024, from <https://docs.advancednavigation.com/subsonus/Introduction.htm>.
- Álvarez-Tuñón, O., Rodríguez, Á., Jardón, A., and Balaguer, C. (2018). Underwater robot navigation for maintenance and inspection of flooded mine shafts. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1482–1487. IEEE.
- Ayoola, O. S. (2019). *In-Situ Monitoring of the Legacy Ponds and Silos at Sellafield*. PhD thesis, The University of Manchester, United Kingdom.
- Barari, A., Glittrup, K., Christiansen, L. R., Ibsen, L. B., and Choo, Y. W. (2021). Tripod suction caisson foundations for offshore wind energy and their monotonic and cyclic responses in silty sand: Numerical predictions for centrifuge model tests. *Soil Dynamics and Earthquake Engineering*, 149:106813.
- Blue Robotics Inc (2024). Bluerov2. "Retrieved January 12, 2024, from <https://bluerobotics.com/store/rov/bluerov2/>.
- Brantner, G. and Khatib, O. (2021). Controlling ocean one: Human–robot collaboration for deep-sea manipulation. *Journal of Field Robotics*, 38(1):28–51.
- Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M., and Tardós, J. D. (2021). Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE Transactions on Robotics*, 37(6):1874–1890.
- Chaumette, F. (2004). Image moments: a general and useful set of features for visual servoing. *IEEE Transactions on Robotics*, 20(4):713–723.
- Chowdhary, G., Johnson, E. N., Magree, D., Wu, A., and Shein, A. (2013). Gps-denied indoor and outdoor monocular vision aided navigation and control of unmanned aircraft. *Journal of field robotics*, 30(3):415–438.
- Christ, R. D. and Wernli Sr, R. L. (2013). *The ROV manual: A User Guide for Remotely Operated Vehicles*. Butterworth-Heinemann, Oxford.
- Duecker, D. A., Bauschmann, N., Hansen, T., Kreuzer, E., and Seifried, R. (2020). Towards micro robot hydrobat-ics: Vision-based guidance, navigation, and control for agile underwater vehicles in confined environments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1819–1826. IEEE.
- Duecker, D. A., Geist, A. R., Kreuzer, E., and Solowjow, E. (2019). Learning environmental field exploration with computationally constrained underwater robots: Gaussian processes meet stochastic optimal control. *Sensors*, 19(9):2094.
- Fackler, M. (2017). Six years after fukushima, robots finally find reactors' melted uranium fuel. Retrieved January 15, 2024, from <https://www.nytimes.com/2017/11/19/science/japan-fukushima-nuclear-meltdown-fuel.html?smid=pl-share&r=0>.
- Ferreira, B., Matos, A., and Cruz, N. (2010). Single beacon navigation: Localization and control of the mares auv. In *OCEANS*, pages 1–9. IEEE.
- FIS 360 (2021). CHALLENGE: ROV positioning in featureless pond environments. Retrieved January 12, 2024, from https://www.gamechangers.technology/challenge/ROV_positioning_in_featureless_pond_environments/.
- FitzGerald, L., Weir, A., Duraibabu, D. B., Omerdic, E., Dooly, G., and Toal, D. (2022). Robotic ship hull inspection for damage repair. In *OCEANS*, pages 1–5. IEEE.
- Girshick, R. (2015). Fast r-cnn. In *IEEE international conference on computer vision*, pages 1440–1448. IEEE.

- Griffiths, A., Dikarev, A., Green, P. R., Lennox, B., Poteau, X., and Watson, S. (2016). Avexis—aqua vehicle explorer for in-situ sensing. *IEEE Robotics and Automation Letters*, 1(1):282–287.
- Groves, K., West, A., Gornicki, K., Watson, S., Carrasco, J., and Lennox, B. (2019). Mallard: An autonomous aquatic surface vehicle for inspection and monitoring of wet nuclear storage facilities. *Robotics*, 8(2):47.
- Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge Univ. Press, Cambridge.
- Hegrenas, O., Berglund, E., and Hallingstad, O. (2008). Model-aided inertial navigation for underwater vehicles. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1069–1076. IEEE.
- Horri, A. (2020). *Underwater Localization in a Confined Space Using Acoustic Positioning and Machine Learning*. PhD thesis, University of Windsor, Canada.
- International Maritime Organization (2015). Harmonized System of Survey and Certification (HSSC) Guidelines, Resolution A.1140 (31). Retrieved January 14, 2024, from <https://www.imo.org/en/OurWork/IIIS/Pages/Harmonized>
- Karlsen, H. O., Amundsen, H. B., Caharija, W., and Ludvigsen, M. (2021). Autonomous aquaculture: Implementation of an autonomous mission control system for unmanned underwater vehicle operations. In *OCEANS*, pages 1–10. IEEE.
- Kohlbrecher, S., Meyer, J., von Stryk, O., and Klingauf, U. (2011a). A flexible and scalable slam system with full 3d motion estimation. In *IEEE International Symposium on Safety, Security and Rescue Robotics*. IEEE.
- Kohlbrecher, S., Von Stryk, O., Meyer, J., and Klingauf, U. (2011b). A flexible and scalable slam system with full 3d motion estimation. In *IEEE international symposium on safety, security, and rescue robotics*, pages 155–160. IEEE.
- Lensgraf, S., Sniffen, A., Zitzewitz, Z., Honnold, E., Jain, J., Wang, W., Li, A., and Balkcom, D. (2021). Droplet: Towards autonomous underwater assembly of modular structures. In *Proceedings of Robotics: Science and Systems*.
- Lepetit, V., Moreno-Noguer, F., and Fua, P. (2009). Epnp: An accurate $O(n)$ solution to the pnp problem. *International journal of computer vision*, 81:155–166.
- Li, J., Du, J., and Chen, C. P. (2021). Command-filtered robust adaptive nn control with the prescribed performance for the 3-d trajectory tracking of underactuated auvs. *IEEE Transactions on Neural Networks and Learning Systems*, 33(11):6545–6557.
- Li, J., Kaess, M., Eustice, R. M., and Johnson-Roberson, M. (2018). Pose-graph slam using forward-looking sonar. *IEEE Robotics and Automation Letters*, 3(3):2330–2337.
- Lin, B. and Dong, X. (2023). Ship hull inspection: A survey. *Ocean Engineering*, 289:116281.
- Ling, Y., Li, Y., Ma, T., Cong, Z., Xu, S., and Li, Z. (2023). Active bathymetric slam for autonomous underwater exploration. *Applied Ocean Research*, 130:103439.
- Loebis, D., Sutton, R., Chudley, J., and Naeem, W. (2004). Adaptive tuning of a kalman filter via fuzzy logic for an intelligent auv navigation system. *Control Engineering Practice*, 12(12):1531–1539.
- McConnell, J., Chen, F., and Englot, B. (2022). Overhead image factors for underwater sonar-based slam. *IEEE Robotics and Automation Letters*, 7(2):4901–4908.
- Negahdaripour, S. and Firoozfam, P. (2006). An ROV stereovision system for ship-hull inspection. *IEEE Journal of Oceanic Engineering*, 31(3):551–564.
- Oliveira, A. J., Ferreira, B. M., and Cruz, N. A. (2021). Feature-based underwater localization using imaging sonar in confined environments. In *OCEANS*, pages 1–7. IEEE.
- Open Robotics (2024). ROS (Robot Operating System). Retrieved January 5, 2024, from <https://www.ros.org/>.

- Ozog, P., Carlevaris-Bianco, N., Kim, A., and Eustice, R. M. (2016). Long-term mapping techniques for ship hull inspection and surveillance using an autonomous underwater vehicle. *Journal of Field Robotics*, 33(3):265–289.
- Palomeras, N., Hurtós, N., Vidal, E., and Carreras, M. (2019). Autonomous exploration of complex underwater environments using a probabilistic next-best-view planner. *IEEE Robotics and Automation Letters*, 4(2):1619–1625.
- Qualisys (2022). Miquis. Retrieved January 8, 2024, from <https://www.qualisys.com/cameras/miquis/>.
- Qualisys (2024). Marine vessels and structures engineering with motion capture technology. Retrieved January 12, 2024, from <https://www.qualisys.com/engineering/marine-vessels-and-structures/>.
- Reach Robotics (2023). Offshore energy. Retrieved January 3, 2024, from <https://reachrobotics.com/offshore-energy/>.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *IEEE conference on computer vision and pattern recognition*, pages 779–788. IEEE.
- Shabani, M., Gholami, A., and Davari, N. (2015). Asynchronous direct kalman filtering approach for underwater integrated navigation system. *Nonlinear Dynamics*, 80:71–85.
- Singer, A. C., Nelson, J. K., and Kozat, S. S. (2009). Signal processing for underwater acoustic communications. *IEEE Communications Magazine*, 47(1):90–96.
- Smith Jr, K., Sherman, A., McGill, P., Henthorn, R., Ferreira, J., Connolly, T., and Huffard, C. (2021). Abyssal benthic rover, an autonomous vehicle for long-term monitoring of deep-ocean processes. *Science Robotics*, 6(60):eabl4925.
- Sonardyne (2022). AvTrak 6 nano. Retrieved January 16, 2024, from <https://www.sonardyne.com/products/avtrak-6-nano-auv-swarm-tracking-and-communications>.
- Sonardyne (2023). Micro-ranger 2 USBL. Retrieved January 20, 2024 <https://www.sonardyne.com/products/micro-ranger-2-shallow-water-usbl-system/>.
- Song, C. and Cui, W. (2020). Review of underwater ship hull cleaning technologies. *Journal of Marine Science and Application*, 19(3):415–429.
- Stojanovic, M. and Preisig, J. (2009). Underwater acoustic communication channels: Propagation models and statistical characterization. *IEEE communications magazine*, 47(1):84–89.
- Suresh, S., Sodhi, P., Mangelson, J. G., Wettergreen, D., and Kaess, M. (2020). Active slam using 3d submap saliency for underwater volumetric exploration. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 3132–3138. IEEE.
- Teixeira, P. V., Fourie, D., Kaess, M., and Leonard, J. J. (2019). Dense, sonar-based reconstruction of underwater scenes. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8060–8066. IEEE.
- Vaarst (2023). SubSLAM X2. Retrieved January 16, 2024 <https://vaarst.com/perception/subslam-x2/>.
- Van Walt (2024). YSI ProDSS. Retrieved January 20, 2024 <https://www.vanwalt.com/equipment/ysi-pro-dss/>.
- Wang, Z., Wu, S., Xie, W., Chen, M., and Prisacariu, V. A. (2021). Nerf-: Neural radiance fields without known camera parameters. <https://arxiv.org/abs/2102.07064>.
- Weiss, S., Scaramuzza, D., and Siegwart, R. (2011). Monocular-slam-based navigation for autonomous micro helicopters in gps-denied environments. *Journal of Field Robotics*, 28(6):854–874.
- Welch, G. and Bishop, G. (1995). An introduction to the kalman filter. Chapel Hill, NC, USA.
- Westman, E., Hinduja, A., and Kaess, M. (2018). Feature-based slam for imaging sonar with under-constrained landmarks. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 3629–3636. IEEE.

- Yao, K., Bauschmann, N., Alff, T. L., Cheah, W., Duecker, D. A., Groves, K., Marjanovic, O., and Watson, S. (2023). Image-based visual servoing switchable leader-follower control of heterogeneous multi-agent underwater robot system. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 5200–5206. IEEE.
- Zhao, C., Thies, P. R., and Johannning, L. (2022). Offshore inspection mission modelling for an asv/rov system. *Ocean Engineering*, 259:111899.
- Zhao, H., Zheng, R., Liu, M., and Zhang, S. (2020). Detecting loop closure using enhanced image for underwater vins-mono. In *OCEANS*, pages 1–6. IEEE.
- Zhou, X., Wen, X., Wang, Z., Gao, Y., Li, H., Wang, Q., Yang, T., Lu, H., Cao, Y., Xu, C., et al. (2022). Swarm of micro flying robots in the wild. *Science Robotics*, 7(66):eabm5954.
- Zhu, Z., Peng, S., Larsson, V., Cui, Z., Oswald, M. R., Geiger, A., and Pollefeys, M. (2023). Nicer-slam: Neural implicit scene encoding for rgb slam. <https://arxiv.org/abs/2302.03594>.