Development of a Low-Cost Payload for Littoral ROV Navigation

Aaron Marburg, Nicholas Guimond
University of Washington
Applied Physics Laboratory
Seattle, Washington, USA
Email: amarburg@uw.edu (Corresponding author)

Gene Mary Cheruvathur,
Celina Liao, Mingyue Shao,
Tegbir Lalli, and Patrick Heller
University of Washington
Electrical and Computer Engineering Department
Seattle, Washington, USA

Zachary Randell, Clyde McQueen
Seattle Aquarium
Seattle, Washington, USA
Email: {z.randell,c.mcqueen}@seattleaquarium.org

***Abstract*—The increasing availability of low-cost underwater robotic platforms has enabled and inspired new forms of community-based scientific inquiry and exploration. In particular, low-cost, hand-launched ROVs can provide greater access to littoral waters as a complement or supplement to scientific divers, allowing operations in more inclement conditions, providing access to communities not trained in science diving, and allowing rapid spot assessment at multiple sample points without requiring time for diver decompression. One impediment to broad adoption of such small vehicles is the relatively high cost of acoustic navigation sensors. This project assesses the utility of a stereo computer vision system running the ORB-SLAM3 visual SLAM algorithm as a substitute or complement to a doppler velocity log (DVL) in measuring vehicle altitude, velocity, and estimating vehicle position. Under the constrained conditions where the system is able to maintain visual tracking with the seafloor, the vision-based estimate of altitude and velocity is highly comparable to an acoustically derived value; and the vision solution provides localization capabilities comparable to IMU-DVL-based dead reckoning.**

I. Introduction

The increasing availability of low-cost underwater robotic platforms has enabled and inspired new forms of community-based scientific inquiry and exploration. Relative to divers, ROVs can remain at depth for longer periods (depending on power configuration), can operate in all water temperatures, and do not require lengthy decompression time during dives. Low-cost remotely-operated-vehicles (ROVs) can allow novel modes of operation including sampling in inclement weather, in high-risk locations, and providing access for users who are not trained for or not able to perform scientific diving. However, current ROVs, particularly modestly-equipped, low-cost models, have limited sensory, maneuvering, and manipulation capabilities and are not an immediate substitute for a trained scientific diver in many cases.

A mission profile of particular interest is photogrammetric imaging of relatively constrained extents of the seafloor. This form of data collection allows for, for example, localized assessment of flora and fauna abundance, and search for objects of interest or invasive species. For population-level assessment, in particular, thorough coverage of a broad area is not necessarily required and sparse sampling within a basin of interest can provide meaningful insights. Again, this plays to the strengths of the ROV which can rapidly cycle to depth at multiple sampling locations without requiring time for decompression.

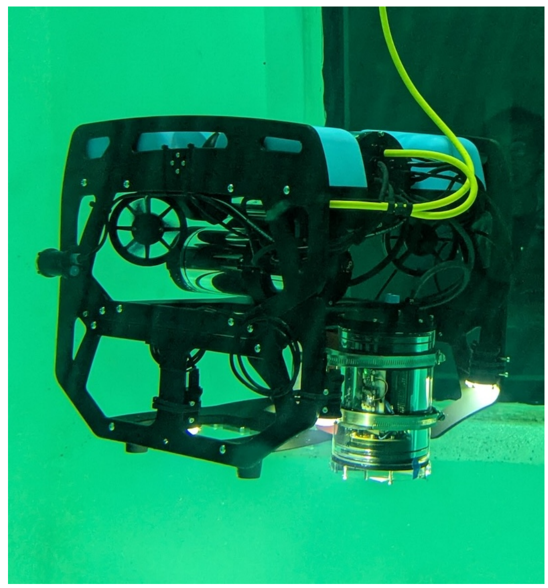


Fig. 1: Stern of BlueROV with integrated downward-looking camera system.

As with all underwater platforms, a key technical compromise inherent in many small ROVs is limited navigation and positioning performance due to constrained volume and power capacity available for the integration of additional sensors, the specialized expertise required for optimal operation of

This work was funded by the Office of Naval Research under grants N00014-21-1-2052 and N00014-22-1-2196

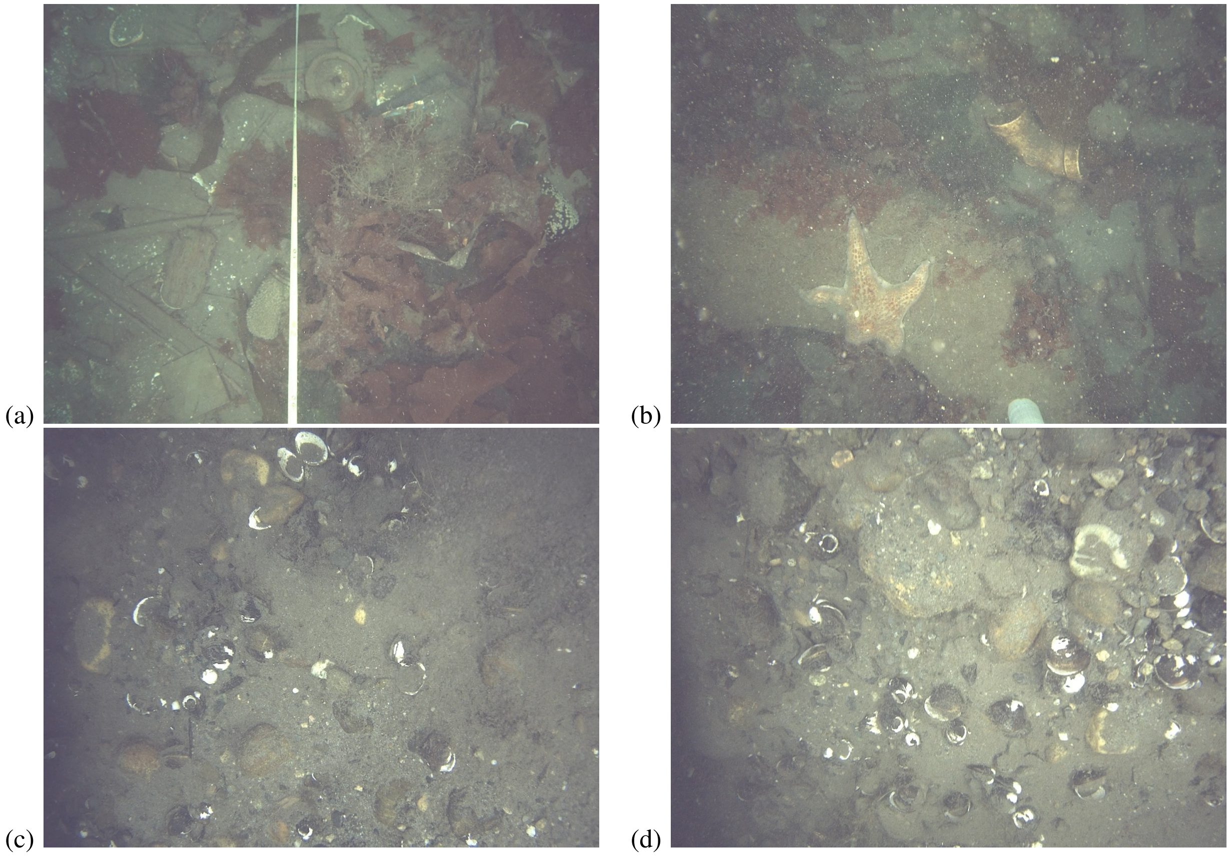


Fig. 2: (a) and (b) Sample imagery from mapping evaluation at Seattle Aquarium pier. (a) shows the 10m tape. (c) and (d) Sample imagery from velocity and altitude evaluation in Seattle Ship Canal. All images captured at 0.5-0.8m altitude.

complex navigation systems, and the cost of many subsea navigation sensors, which can exceed the cost of the core vehicle itself. The current marketplace of low-cost ROV navigation systems includes a diversity of inertial measurement units (IMUs), single-beam and mechanically scanned sonar/echo-sounders, doppler velocity logs (DVL), and long-, short- and ultra-short-baseline acoustic positioning systems (LBL, SBL, and USBL resp.) [1]. Depending on the resources available to an ROV operator, mission needs, and the operating environment, different combinations of sensors may be integrated to produce an estimate of vehicle global position and local motion, and to inform various vehicle control loops.

For the seafloor survey mission, the ROV has two navigation needs: (1) to understand its global position with sufficient precision to geolocate any collected data; and (2) to understand its relationship to the seafloor, including altitude and speed over ground, as well as local relative position to estimate data coverage and overlap. Arguably, these two navigation needs are lightly decoupled, with an acceptable global positioning accuracy on the order of 2–5% of water depth, a value consistent with SBL or USBL tracking, while operations near the seafloor require precise (cm-scale) local understanding of vehicle pose to avoid collisions and spatially correlate collected data.

The current best option for local navigation is a combination of IMU and DVL, potentially complemented by an external acoustic navigation system [1]. This pairing of IMU for attitude and DVL for seafloor-relative velocity and altitude provides an enhanced dead reckoning solution of vehicle track, which may be tied to a global coordinate frame through integration with an acoustic tracking solution.

While effective, DVL-based navigation has some disadvantages. Lower-cost models can suffer from processing latency, difficulties when operating close to uneven seafloor, and minimum operating ranges very close to desired survey altitude ($\sim$1m). Above all, even "low-cost" DVLs can cost on the same order of magnitude as the base ROV itself, increasing the barriers of entry to use for community science.

This project develops a prototype low-cost, computer-vision based navigation appliance for seafloor survey. The goal is not to use the camera system as the primary science sensor, as other higher-resolution, higher-image-quality camera systems were already integrated into the ROV workflow for data collection. Instead, the system is designed as a supplemental data source dedicated to vehicle localization. The overall project objective is to assess the feasibility of optical tracking of soft substrates and seafloor, including regions of extensive kelp growth, as a source of (a) vehicle altitude; (b) vehicle speed over ground; and (c) vehicle position within a local frame (mapping).

This paper describes the system hardware and software design, and presents preliminary results from system testing. Overall, the priority is to assess the technical feasibility of the concept and develop a baseline for future project development. The system is designed as an add-on to the BlueROV platform and to interact with the existing BlueROV piloting and navigation systems, rather than as a replacement for existing control algorithms. This project was completed by a student team within the University of Washington ENGINE Senior Capstone program, with support from the Coastal Climate Resilience program at the Seattle Aquarium.

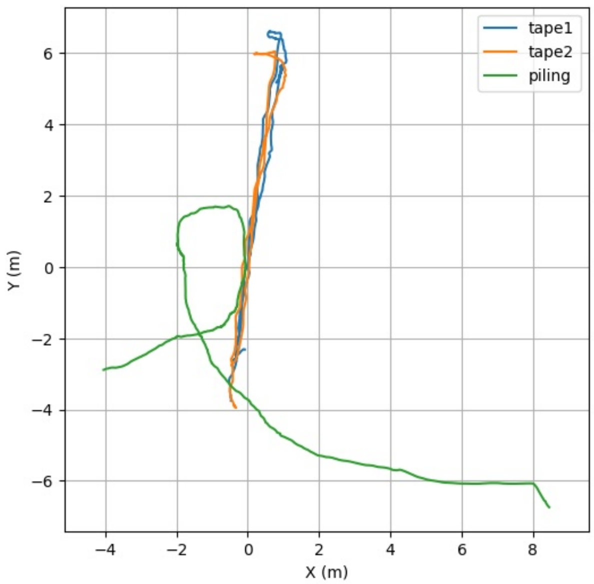


Fig. 3: Mission trajectories as reconstructed by COLMAP. Note the survey origin is arbitrary based on the order of processing within COLMAP.

II. Related Works

While fundamentally relying on ORB-SLAM3, this work is inspired by the long history of offline and realtime algorithms for improving nadir-looking photogrammetry and photo survey, both in the air and underwater [2]–[5]. The proposed approach is driven by the increasing capabilities of modern, factor-graph-based SLAM approaches, including ORB-SLAM3, and their application to a broad range of environments [6]–[8]. This system in particular is strongly inspired by AQUA-SLAM [9] which extends ORB-SLAM3 for the marine environment by adding custom factors for integration of DVL velocity estimates – their approach also performs online DVL calibration and DVL-IMU alignment, while leaving the ORB-SLAM3 visual frontend largely intact. As it performs DVL-inertial-visual fusion their approach is more sophisticated, and suitable for applications where both vision and a DVL are available.

Other recent approaches have attempted to improve underwater visual SLAM through the integration of external modalities. The recent SVIN2 framework [10] fuses inertial, visual, pressure, and a mechanically scanned profiling sonar estimate vehicle trajectories, including loop closure. Wang et al. [11] were then able to use the navigation solution to produce dense realtime 3D maps of the environment.

III. Hardware Design

The system consists of an NVidia Jetson Orin Nano development board housed in a BlueRobotics 4-inch diameter acrylic housing. The housing also contains two Vision Components global shutter cameras based on the Sony IMX296 sensor with 2.7mm focal length wide-angle lenses. The cameras are hardware synchronized by digital I/O on the Jetson board and look out through a standard 12.7mm-thick BlueRobotics flat acrylic endcap with a 5.5cm stereo baseline. In these trials the cameras were operated at 10Hz. The housing also includes a small companion board which samples environmental and leak sensors, and a VectorNav VN-100 IMU which is sampled at 100Hz. A single subsea cable (ethernet plus power) connects the camera housing to the main BlueROV control vessel, which contains limited modifications from stock, notably the inclusion of a BotBlox Gigablox ethernet switch to provide connectivity to the Jetson development board.

The 4-inch housing is mounted in the stern of the ROV, pointed downward on the standard BlueRobotics payload sled (Figure 1) with the stereo baseline abeam relative to the ROV's frame.

IV. Software Design

The Jetson Nano development board runs the Jetpack operating system provided by NVidia, and the ROS2 "Humble" middleware layer [12], including a custom stereo camera driver for the stereo Vision Components cameras.

The intrinsic and extrinsic calibration of the stereo camera system, as well as the camera-IMU alignment, were estimated using Kalibr [13].

A. ORB-SLAM3 Integration

Multiple visual processing approaches were evaluated, with preliminary results from the ORB-SLAM3 ([14]) visual mapping package selected as the most promising for further development. Testing used a modified ROS2 wrapper for ORB-SLAM3 which adds key introspection capabilities, as well as the capacity to process data both in realtime and in an offline post-processed mode [15].

Mapping is evaluated using the optimized odometric output from ORB-SLAM3. The local origin of the SLAM map is set to the first keyframe, necessitating rigid body alignment for comparison with other trajectories. Body velocities are calculated from finite differences from ORB-SLAM3's realtime odometric output, while altitude is calculated from the ROV-relative positions of tracked map points. Rather than using raw ORB features, tracked map points are used as they have passed geometric consistency tests, eliminating false positives from drifting marine particulates or moving features on the seafloor. A center-weighted average of vertical distance to each feature is taken to simulate a "beam pattern" which concentrates on points near the center of the camera field of view.

V. TESTING AND EVALUATION

The system was evaluated in two field trials. The first occurred in Puget Sound adjacent to the Seattle Aquarium and evaluated mapping performance in realistic ocean survey conditions. Unfortunately, a DVL was not available at the time of this test and no ground truth of vehicle altitude or velocity was collected.

A subsequent test occurred in the freshwater Seattle Ship Canal with a Waterlinked A50 DVL mounted on the ROV. The ROV allows collection of ground truth altitude and velocity information, however this data is significantly more turbid and contains fewer seafloor features reducing its value in evaluating mapping performance.

A. *Mapping evaluation*

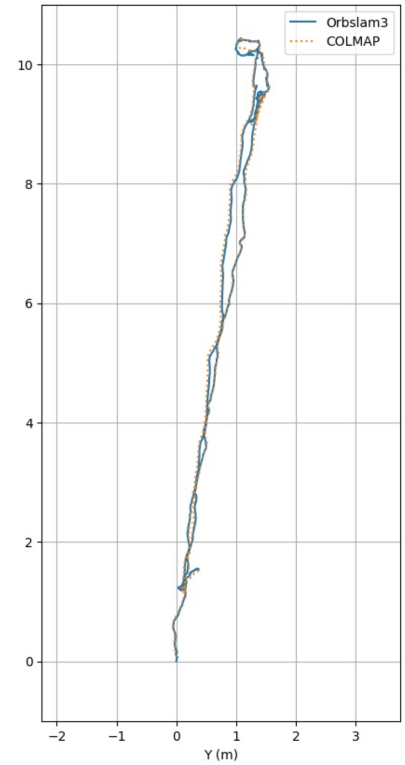
The ROV and mapping package were deployed adjacent to the Seattle Aquarium in ∼10m of water. This urban, littoral site features significant seafloor debris and offers a highly textured seafloor (Figure 2(a)-(b)). In preparation for the test, a 10m tape was placed on the seafloor extending from a pier piling. The ROV collected two out-and-back transects along the tape (*tape1* and *tape2*) as well a less constrained survey of a collapsed piling nearby (*piling*). The two tape missions are naturally overlapping, while the *piling* contains limited intersections with itself or with the *tape* missions.

To develop a ground truth for vehicle motion, combined imagery from all three missions was post-processed in COLMAP [16], [17] to produce a synoptic reconstruction of the study area. An overview of the reconstructed trajectories is shown in Figure 3.

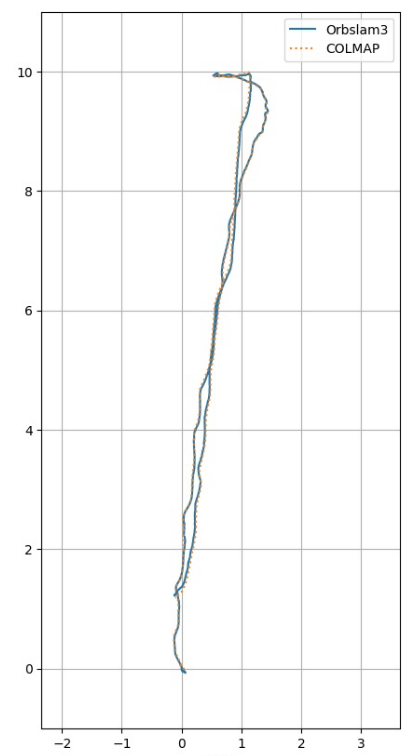
Each trajectory was processed individually in a new ORB-SLAM3 session. As the COLMAP and ORB-SLAM3 trajectories are derived from the same input image stream, the results are inherently time synchronized and can be compared directly using the *evo* toolkit [18]. As neither COLMAP nor ORB-SLAM3 are provided with global references, each exists within its own local frame; *evo* estimates a rigid body transformation between the two local frame before comparing trajectories.

For each trajectory, we consider two metrics, absolute pose error (APE) and relative pose error (RPE) in both translation and rotation, summarized in Table I. Tracking success gives fraction of the total mission duration where ORBSLAM successfully localizes relative to the local map. Note on the *piling* mission, the algorithm is unable to maintain tracking throughout the duration of the mission and as such the overall APE is not meaningful; however the reconstruction is successful for approximately the first half of the mission (followed by extended stretches where the algorithm is unable to track the seafloor), and the APE of that subset can be evaluated individually. The COLMAP- and ORB-SLAM3-derived trajectories for the *tape1* and *tape2* missions are shown in Figure 4.

Per Table I, the ORB-SLAM3 trajectory estimate over the $\sim 21-25$m-long *tape* missions was approx 0.6% of distance traveled, a value competitive with existing IMU-DVL dead reckoning solutions. This is assisted in this case by the



(a) Mission *tape1*



(b) Mission *tape2*

Fig. 4: Comparison of COLMAP and ORB-SLAM3 trajectories for both out-and-back surveys of 10M tape on seafloor.

Mission	Traj. length (m)	Tracking success	Translation (m)		Rotation (rad)	
			APE	RPE	APE	RPE
tape1	25.79	100%	0.150	0.004	0.146	0.005
tape2	21.59	100%	0.126	0.003	0.106	0.004
piling overall	24.774	82.8%	–	0.019	–	0.011
piling subset	13.22	100%	0.179	0.004	0.152	0.004

TABLE I: Summary of mapping performance from testing at Seattle Aquarium.

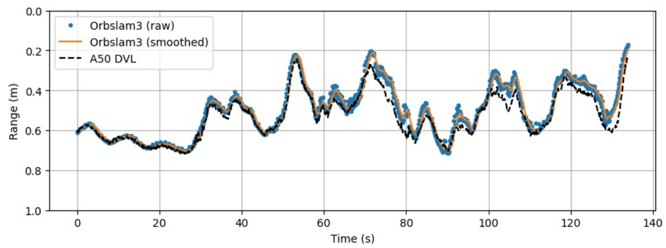
relatively clear water and the continuous, low altitude which allowed sustained visual tracking. The *piling* mission shows the hazards of not meeting those conditions as ORB-SLAM3 is unable to maintain tracking – nor to relocalize on the existing map, as the path cross itself only once – and as such the vehicle has no absolute position estimate in the second half of the mission. However, the relatively consistent RPE demonstrates that the instantaneous motion estimates remain accurate.

B. Range and velocity evaluation

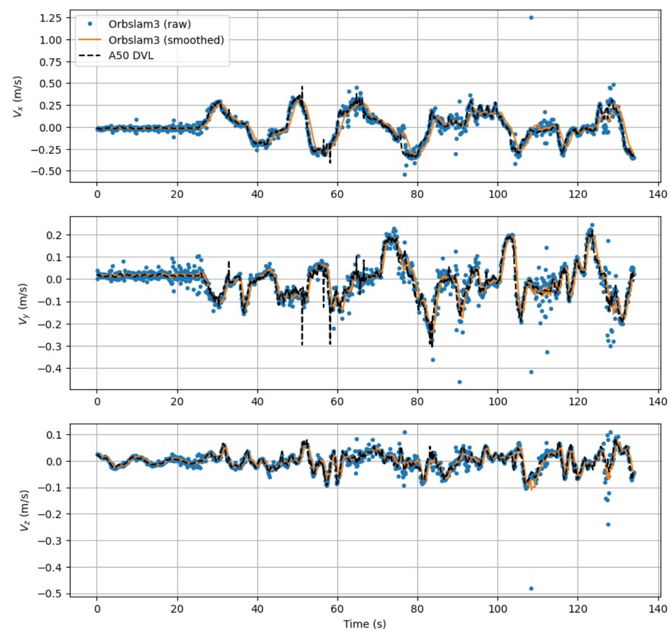
Range and velocity estimation testing occurred in the Seattle Ship Canal with a Waterlinked A50 DVL mounted on the ROV. The DVL was adjusted to the speed of sound in fresh water based on ambient water temperature. This test site was significantly more turbid than the Seattle Aquarium site, limiting effective ranges for visual testing but also providing ample opportunity for evaluating the robustness of ORB-SLAM3's visual odometry and map point tracking implementations. Evaluation focused on a $\sim$2 minute subset when the ROV was consistently 0.5-0.8m from the seafloor, maximizing visibility.

Comparison of DVL and visual estimates of vehicle range to bottom are shown in Figure 5a. For comparison, the ORB-SLAM3 estimate is smoothed by a 10-sample (1-sec) moving window average. Neither DVL nor ORB-SLAM3 ranges are compensated for vehicle pitch and roll.

Similarly, the ROV body velocities estimated by the DVL and ORB-SLAM3 are presented in Figure 5b. Again, both raw and smoothed velocities are presented; for comparison purposes, the ORB-SLAM3 velocities have been transformed from the camera frame to ROV body frame (X fore, Y starboard, Z down).



(a) Comparison of range to seafloor.



(b) Comparison of ROV body velocity.

Fig. 5: Comparison of Waterlinked A50 DVL and ORB-SLAM3 estimates of range to seafloor and ROV velocities.

VI. Conclusion

This project assessed the feasibility of using a low-cost embedded stereo system running a published visual processing algorithms to perform altitude, velocity, and position estimation as a potential substitute – or complement – for a DVL as a source of navigation information when performing near-bottom surveys.

As shown above, the system is capable of making the required measurements with high accuracy, although with the strong caveat that the system is only effective when the seafloor is visible. Though seemingly a disqualifying restriction, during visual seafloor surveys, such visibility is a prerequisite for mission success. Moreover, compared to a DVL the vision system can provide a dense estimate of vehicle altitude, allowing, for example, identification of rocks or cobbles which can be avoided, but do not require a significant change in altitude, and critically it offers the possibility of re-localization or "loop closure" to improve long-term position estimates, a capability which DVL-based dead reckoning cannot achieve.

Of course, when the project budget and ROV payload allows, the ultimate solution would integrate DVL, IMU, and visual information (as in [9]) for local navigation, potentially complemented with acoustic ranging for global localization.

The described results are preliminary and rely heavily on the unaltered ORB-SLAM3 package; future work will include both evaluation of alternative vision processing components, including other SLAM packages, integration of more robust filtering and failure detection in the presence of poor imaging conditions (due to e.g., turbidity), and continued testing.

REFERENCES

[1] Bingbing Zhang, Daxiong Ji, Shuo Liu, Xinke Zhu, and Wen Xu. Autonomous underwater vehicle navigation: A review. *Ocean Engineering*, 273:113861, 2023.

[2] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustment—a modern synthesis. In *International workshop on vision algorithms*, pages 298–372. Springer, 1999.

[3] Hanumant Singh, Jonathan Howland, and Oscar Pizarro. Advances in large-area photomosaicking underwater. *IEEE Journal of Oceanic Engineering*, 29(3):872–886, 2004.

[4] Nuno Ricardo Gracias, Sjoerd Van Der Zwaan, Alexandre Bernardino, and José Santos-Victor. Mosaic-based navigation for autonomous underwater vehicles. *IEEE journal of oceanic engineering*, 28(4):609–624, 2004.

[5] Ryan M Eustice, Hanumant Singh, John J Leonard, and Matthew R Walter. Visually mapping the rms titanic: Conservative covariance estimates for slam information filters. *The international journal of robotics research*, 25(12):1223–1242, 2006.

[6] Zijun Sha, Syuuhei Shiro, Kazuki Shibamiya, Kazuhiro Shintani, and Kazuhiro Tanaka. Improving orb-slam3 performance in textureless environments: Mapping and localization in hospital corridors. In *2025 IEEE Conference on Artificial Intelligence (CAI)*, pages 723–726. IEEE, 2025.

[7] Qihang Chen, Chengxiang Xie, Sutthiphong Srigrarom, and Boo Cheong Khoo. 3d reconstruction of indoor environment using rgb-d camera based on orb slam3. In *Proceedings of the 2024 10th International Conference on Robotics and Artificial Intelligence*, pages 24–30, 2024.

[8] Qiuyu Zang, Kehua Zhang, Ling Wang, and Lintong Wu. An adaptive orb-slam3 system for outdoor dynamic environments. *Sensors*, 23(3):1359, 2023.

[9] Shida Xu, Kaicheng Zhang, and Sen Wang. Aqua-slam: Tightly-coupled underwater acoustic-visual-inertial slam with sensor calibration. *IEEE Transactions on Robotics*, 2025.

[10] Sharmin Rahman, Alberto Quattrini Li, and Ioannis Rekleitis. Svin2: A multi-sensor fusion-based underwater slam system. *The International Journal of Robotics Research*, 41(11-12):1022–1042, 2022.

[11] Weihan Wang, Bharat Joshi, Nathaniel Burgdorfer, Konstantinos Batsos, Alberto Quattrini Li, Philippos Mordohai, and Ioannis Rekleitis. Real-time dense 3d mapping of underwater environments. *arXiv preprint arXiv:2304.02704*, 2023.

[12] Steven Macenski, Tully Foote, Brian Gerkey, Chris Lalancette, and William Woodall. Robot operating system 2: Design, architecture, and uses in the wild. *Science Robotics*, 7(66):eabm6074, 2022.

[13] Paul Furgale, Joern Rehder, and Roland Siegwart. Unified temporal and spatial calibration for multi-sensor systems. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1280–1286. IEEE, 2013.

[14] Carlos Campos, Richard Elvira, Juan J Gómez Rodríguez, José MM Montiel, and Juan D Tardós. Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE transactions on robotics*, 37(6):1874–1890, 2021.

[15] Aaron Marburg. https://gitlab.com/apl-ocean-engineering/orbslam3_odometry_ros2, 2025.

[16] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016.

[17] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[18] Michael Grupp. evo: Python package for the evaluation of odometry and slam. https://github.com/MichaelGrupp/evo, 2017.