Towards Mapping of Underwater Structures by a Team of Autonomous Underwater Vehicles

Marios Xanthidis[1], Bharat Joshi[2], Monika Roznere[3], Weihan Wang[4], Nathaniel Burgdorfer[4], Alberto Quattrini Li[3], Philippos Mordohai[4], Srihari Nelakuditi[2], and Ioannis Rekleitis[2]

[1] SINTEF Ocean, Trondheim, Norway, 7010,
marios.xanthidis@sintef.no
[2] University of South Carolina, Columbia, SC, USA, 29208,
bjoshi@email.sc.edu, {yiannisr,srihari}@cse.sc.edu
[3] Dartmouth College, Hanover, NH, USA, 03755
{monika.roznere.gr, alberto.quattrini.li}@dartmouth.edu
[4] Stevens Institute of Technology, Hoboken, NJ, USA, 07030,
{wwang103,nburgdor,pmordoha}@stevens.edu

Abstract. In this paper, we discuss how to effectively map an underwater structure with a team of robots considering the specific challenges posed by the underwater environment. The overarching goal of this work is to produce high-definition, accurate, photorealistic representation of underwater structures. Due to the many limitations of vision underwater, operating at a distance from the structure results in degraded images that lack details, while operating close to the structure increases the accumulated uncertainty due to the limited viewing area which causes drifting. We propose a multi-robot mapping framework that utilizes two types of robots: proximal observers which map close to the structure and distal observers which provide localization for proximal observers and bird's-eye-view situational awareness. The paper presents the fundamental components and related current results from real shipwrecks and simulations necessary to enable the proposed framework, including robust state estimation, real-time 3D mapping, and active perception navigation strategies for the two types of robots. Then, the paper outlines interesting research directions and plans to have a completely integrated framework that allows robots to map in harsh environments.

Keywords: Underwater, Multi-Robot, Navigation, Mapping, and Localization

1 Introduction

Underwater structure mapping is an important capability applicable to multiple domains: marine archaeology, infrastructure maintenance, resource utilization, security, and environmental monitoring. The underwater environment is challenging and dangerous for humans in many ways, while robotic operations face additional challenges compared to the above-water ones. In particular, sensing and communications are restricted and planning is required in three dimensions based on limited information. Current approaches for underwater autonomous operations are limited to hovering at a distance [1–3], possibly resulting in occluded views and coarse models; operating autonomously near an underwater structure, thus obtaining high-resolution images, has been impossible so far.

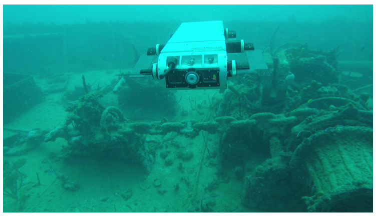


Fig. 1: Aqua2 Autonomous Underwater Vehicle exploring the top structure of the Stavronikita shipwreck, Barbados.

The overarching goal of this work is to create a 3D model of the underwater structure providing a high-resolution photo-realistic representation. To achieve this goal, we propose a framework that considers a team of robots operating in close cooperation. Some Autonomous Underwater Vehicles (AUVs), termed *proximal observers*, will be operating close to the underwater structure generating a dense vision-based 3D reconstruction of the observed surface; see Fig. 1 where an Aqua2 vehicle swims over the deck of a wreck. The rest of the robots, termed *distal observers*, will operate further out maintaining the global picture of the underwater structure and the pose of the proximal observers with respect to the structure.

In this paper, we discuss the fundamental components we have developed that contribute to the realization of the above framework. The distal observer monitors the relative pose of the proximal observer utilizing a Cooperative Localization (CL) [4, 5] scheme. The proximal observer maintains its current pose estimate even in the face of sensor failures. The visual data from the proximal observer are integrated into a 3D map, utilizing either real-time dense depth map fusion or a formulation of photometric stereo. Finally different motion strategies are employed by the distal and proximal observers to maximize target visibility. Experimental results for each component from deployments over a shipwreck are presented together with simulations in a realistic 3D robotic simulator (Gazebo [6]). These results highlight the potential of the proposed approach and provide insights on interesting research directions for mapping in harsh environments and on future plans for full system integration.

2 Related Work

Many approaches utilize visual and visual/inertial data for estimating the pose of a robot [7–10]. However, evaluations on a variety of underwater datasets have demonstrated the challenges of the underwater domain [11, 12]. Even when the integration of multiple sensors produces consistent estimations [13], operating around a 3D structure often results in loss of tracking when no unique visual features are present. An alternative to estimating the state of a moving robot is relative localization from another robot [4, 5, 14]. This is a challenging problem underwater due to limited visibility and potential occlusions.

Active perception, which was first introduced in the context of exploration [15–17], enables robots to actively map the environment. Due to the challenges of the underwater domain, there are only few active perception applications aiming to minimize uncertainty during coverage [18, 19] or explore interesting features [20] in simplistic environments with respect to obstacles. Advancements in deep learning have produced robotic systems that move freely while observing areas of interests [21–23] based on datasets collected by human operators. By construction, these systems are limited by the agility of the operator, the domain of the training set, and the excessive data needed for the production of such frameworks. Past works utilized well-established sampling-based techniques which provide strong guarantees [24] but are computationally expensive. On the other hand AquaVis [25], our previous work, proposed a lightweight real-time framework based on path-optimization that can navigate safely a complex 3D environment and at the same time observe multiple visual objectives using an arbitrary multi-sensor configuration.

Several active sensing approaches for 3D mapping require enumerating and simulating sensing from different discrete 6-D pose hypotheses [26–29] at high computational cost; other approaches are limited to 2-D slices of constant depth or height [30, 31] or they require the use of rough initial models [32, 33]; in addition, most operate on occupancy grids reducing the resolution of the reconstructed surfaces drastically. Exploration strategies [34, 35] that guide a vehicle towards frontier voxels without requiring sampling in pose space are closely related to our work, but they are limited to a single robot, and require a prior map. Multi-robot 3-D reconstruction methods have been presented [36], but robots are distributed in space to map independently without tight cooperation, and operate at distance to the target structure [37, 38].

3 Proximal-Distal Mapping Framework

Our proposed mapping framework relies on proximal and distal observers to overcome the inherent challenges of the underwater domain. Fig. 2 shows the full envisioned process. Here, we discuss each fundamental component that will enable the mapping by these two types of robots, highlighting the current results. For grounding our discussion, we refer to the specific underwater robots used, although the components and framework can be generalized.

The main target vehicle is the Aqua2 AUV [39]. Aqua2 utilizes the motion of six flippers, each one independently actuated by an electric motor, to swim. Aqua2 has 6 DOF, of which five are controllable: two directions of translation (forward/backward and upward/downward), along with roll, pitch and yaw. The robot's primary sensor is vision, more specifically three iDS USB 3.0 UEye cameras: two facing forward and one in the back. Aqua2 also has a pressure sensor and an IMU which are used for controlling the motions and can be utilized for visual-inertial state estimation [40, 41, 8, 42, 13].

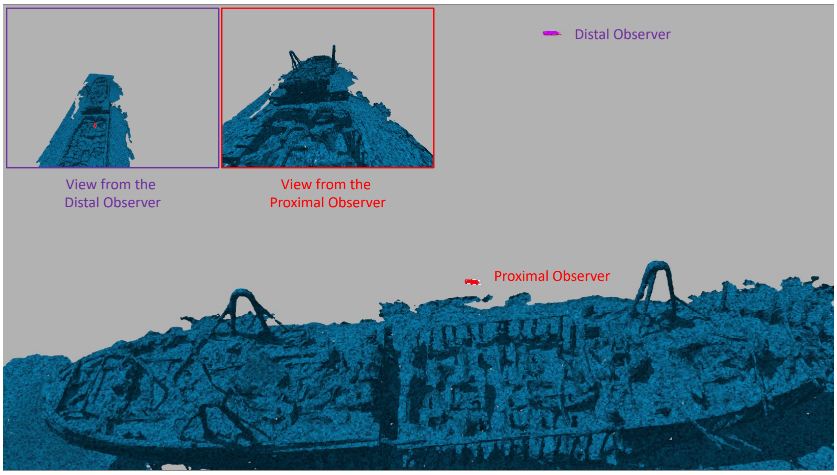

Distal Observer
View from the Distal Observer
View from the Proximal Observer
Proximal Observer

Fig. 2: Two AUVs exploring a wreck. Inserts present the view of each observer: the distal observer, in purple, keeps a large portion of the wreck and the proximal observer in view. The proximal observer, in red, has a close-up view of the wreck.

A lower-cost robot considered is the BlueROV2, a thruster-vectored robot that has an inexpensive sensor suite composed of: a monocular camera, an IMU, and a pressure sensor.

3.1 Robust State Estimation

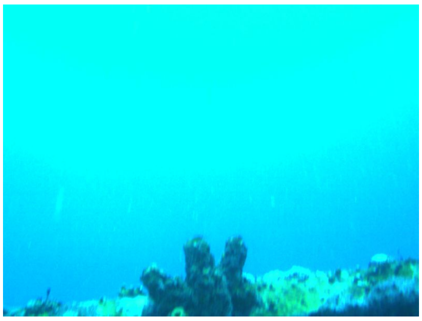


Fig. 3: The view of the Aqua2 AUV just before traveling over the starboard side of the Stavronikita wreck.

A major challenge underwater is robust robot state estimation, given the lack of global localization infrastructure. In addition to many underwater challenges such as lighting variations, limited visibility, and color absorption by distance, when AUVs operate around underwater structures they often encounter complete loss of visual tracking due to the field of view facing only open water, or a featureless surface such as a sandy bottom. Utilizing a dataset collected over a shipwreck by an Aqua2 AUV we report average time for loss of tracking on some of the most common VO/VIO software packages. As can be seen from Table 1, most of the packages get completely lost when the robot reaches the starboard side of the deck; see Fig. 3 for the view from the Aqua2 AUV about to travel over the railings at the starboard side of the deck; the image is dominated by blue water.

There are two major components that are necessary for proximal and distal observers to coordinate: 1) relative pose estimation so that the local observations can be mapped into a global reference frame; 2) robust single AUV pose estimation so that each robot can localize.

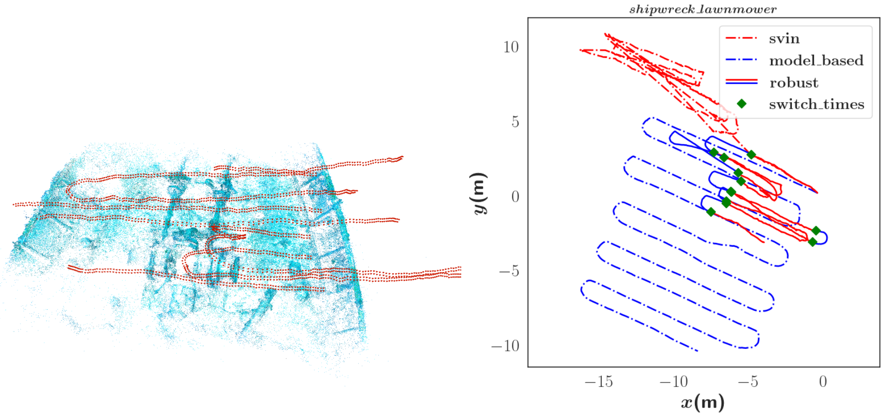
Table 1: Performance of popular open-source VIO packages on the wreck dataset. The root mean squared ATE compared to COLMAP trajectory after se3 alignment.

Algorithm	Time to first track loss (in sec)	Recovery?	RMSE (in m)
OKVIS [7]	23.4	Partial	5.199
VINS-Fusion[43]	23.6	Partial	53.189
SVIn2[44]	23.4	Yes	1.438
Robust Switching Estimator	N/A	Yes	1.295

Relative Pose Estimation To locate robots in a common reference frame, we formulate a cooperative localization framework, where robots can estimate their relative pose. A major challenge underwater is the lack of ground truth, as setting up a motion capture system is prohibitively complicated. As a result, most learning based approaches face a shortage of training data. We employ a novel approach by Joshi et al. [45] utilizing a Generative Adversarial Network (GAN) [46] to train on simulated images, where the simulator provides the pose of the AUV, and then test on real images. Estimating a number of fixed points on the AUV in conjunction with the vehicle's geometry and a calibrated camera yields accurate estimates of the 3D pose of the observed AUV; for details please refer to [45].

Robust AUV Pose Estimation In order to address this common challenge, a novel estimator robust to VIO failures s outlined here. A model-based estimator is employed in conjunction with SVIn2 [44], an accurate VIO package in a robust switching estimator framework. The proposed estimator monitors VIO health based on the number of features tracked, their spatial distribution, feature quality, and their temporal continuity. When health deteriorates below a certain threshold, the model-based estimator is utilized, initialized at the last accurate pose of the VIO system. When VIO recovers and features are tracked again, the estimator switches back to SVIn2 which is itself initialized to the corresponding model-based estimator pose. The result is a sequence of segments (model-based and VIO) maintaining a consistent pose through the whole trajectory. In the event of loop closure, the corrections are propagated, through the pose graph, to the complete trajectory.

Results: Some of the most popular open-source packages were tested on a sequence collected over a wreck. The images from the sequence were fed to COLMAP [47, 48] and the estimated poses of the cameras are used as reference trajectory. The resulting reconstruction from COLMAP can be seen in Fig. 4(a). As can be seen in Table 1, the state-of-the art methods lost track when the AUV faced blue water with no visible structure. SVIn2 managed to recover due to loop closure, while OKVIS without loop closure drifted much further and VINS-fusion drifted even further, with respect to the COLMAP reference trajectory, at 53 meters. Our proposed switching estimator managed to keep track throughout the trajectory with the lowest RMSE error. Figure 4(a) presents the sparse reconstruction from COLMAP together with the estimated poses. When the AUV was facing blue water, the camera pose was not tracked. In contrast, the switching estimator managed to accurately track the trajectory throughout as can be seen in Fig. 4(b).



(a) (b)

Fig. 4: (a) Wreck reconstruction using COLMAP together with the estimated camera poses [47, 48]. (b) Trajectory estimation utilizing a switching estimator using SVIn2 [44] and a model based estimator.

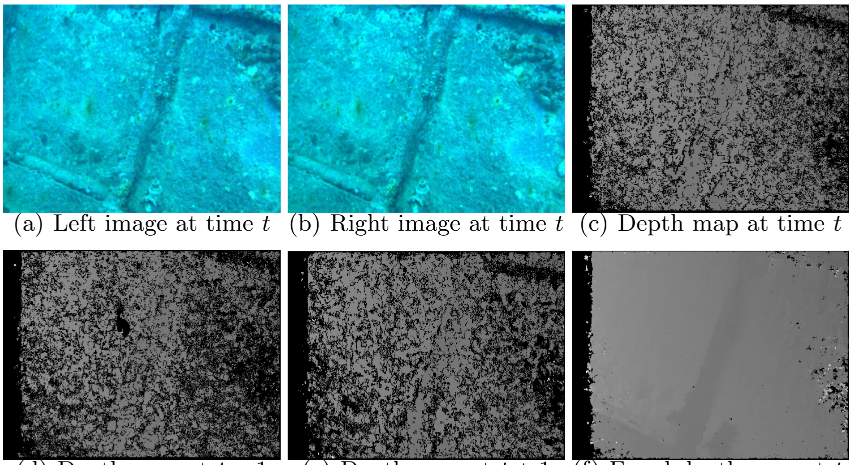
3.2 Photorealistic Reconstruction

With a robust state estimate, the robots can create a 3D map. Here, we discuss components that enable real-time dense 3D mapping within the proximal/observer framework: 1) using just a stereo camera; 2) using lights to have a more robust 3D map.

Real-time Dense 3D Mapping Dense surface reconstruction relies on stereo matching across the left and right camera of the proximal observer and on fusing multiple depth maps to achieve improved robustness and accuracy. To achieve real-time performance, we decompose dense surface estimation in stereo matching and depth map fusion modules with constant computational complexity.

The core of our dense 3D reconstruction pipeline is a binocular stereo matching module which estimates depth for the pixels of the left image from a rectified stereo pair of images. We are able to process image pairs at several frames per second on the CPU using a publicly available multi-threaded implementation.[5]

[5] https://github.com/kbatsos/Real-Time-Stereo


(a) Left image at time t (b) Right image at time t (c) Depth map at time t

(d) Depth map at $t-1$ (e) Depth map at $t+1$ (f) Fused depth map at t

Fig. 5: (a)-(b): input images. (c)-(e): input depth maps for fusion. (f): fused depth map.

Stereo matching operates by assigning a cost or score to each possible disparity[6] that can be assigned to a given pixel of the reference image, typically the left. (We will use cost in the remainder without loss of generality.) Cost is computed in square windows centered around the pixels under consideration.

All matching costs are stored in a *cost volume* with dimensions equal to the width and height of the images and the number of disparity candidates for every pixel. (The maximum disparity corresponds to the minimum depth of interest, while minimum disparity can be set to 0.) The cost volume can be optimized to impose piece-wise smoothness and extract accurate depth maps via the widely used Semi-Global Matching algorithm (SGM) [49]. Here, we integrate the rSGM implementation of Spangenberg et al. [50] into the stereo matching code. Finally, disparity is converted to depth using the known baseline and focal length of the cameras. Sub-pixel precision is obtained by fitting a parabola to the minimum cost and its two neighboring values [51]. To support the subsequent depth map fusion module we associate a confidence value to each depth estimate. To this end, we adopt the PKRN measure [52], which is the ratio of the second smallest over the smallest cost for a given pixel after SGM optimization. An example of a pair of input images and the resulting depth map can be seen in Fig. 5.

Depth maps estimated by the stereo matching module suffer from artifacts due to lack of texture, occlusion, and motion blur. Assuming that errors do not persist over multiple frames, we propose to improve the depth maps by fusing them.

[6] Disparity is defined as the difference between the horizontal coordinates of two potentially corresponding pixels in the same epipolar line (scanline) in the left and right image. Disparity is inversely proportional to depth.

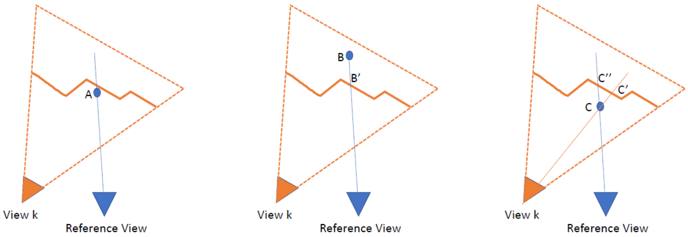


Fig. 6: Constraints used in depth map fusion. Points A, B and C are depth candidates either estimated for the reference view directly or rendered to it from other views. The solid orange polyline is the cross section of the surface estimated by view k. Left: point A is supported by the orange surface. Middle: point B is occluded by B' which is in front of B in the ray of the reference view. Right: point C violates the free space of C' on the ray of view k.

The principle behind depth map fusion is that, as long as individual overlapping depth maps produce either relatively consistent 3D estimates for the same part of the scene or uncorrelated noisy estimates, measuring the consensus and conflicts among depth estimates allows us to improve the accuracy of the correct estimates and to reject outliers. To achieve real-time, scalable 3D reconstruction, our approach operates in sliding window fashion, keeping a small number of recent depth maps in memory at a given time. This decomposition allows the pipeline to operate at constant speed regardless of the size of the scenes and the number of frames that have been collected. At each time step, the middle depth map in the sliding window is used as reference and the remaining depth maps are rendered onto it along with the corresponding confidence maps.

In this paper, we adopt visibility-based fusion from our previous work [53, 54]. The setting here is more challenging since the input depth maps are estimated from two images only, and are thus more susceptible to occlusion. The input for computing a fused depth map for a given *reference view* is a set of N_f depth maps and the corresponding confidence maps. The fusion process begins by rendering the depth and confidence maps to the reference view yielding a new set of N_f depth and confidence maps from the perspective of the reference view.

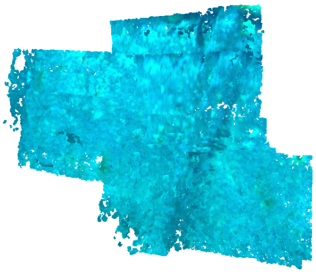


Fig. 7: Shipwreck's partial point cloud from the stereo dense 3D mapping pipeline.

At the end of the rendering stage, we have at most N_f depth candidates per pixel of the reference view as depths may project out of bounds. For each depth candidate d_j, we accumulate *support* and *visibility violations*. Support comes from other depth candidates for the same pixel that are within a small distance of d_j. d_j is then replaced by the confidence-weighted average of the supporting depths. The confidence of the fused depth estimate d_j is set equal to the sum of the supporting confidences. See Fig. 6 (left).

Visibility violations are of two types: occlusions and free space violations. An *occlusion* occurs when d_j appears behind a rendered depth map from view k, D_k^r on the ray of the reference view, as in Fig. 6 (middle), while a *free space violation* occurs when d_j appears in front of an input depth map D_l on the ray of view l, as in Fig. 6 (right). For each detected violation, we penalize the confidence of d_j by subtracting the confidence of the conflicting depth estimate.

At the end, we assign to each pixel the depth candidate with the maximum fused confidence, as long as it is positive. (We can also apply larger thresholds on the confidence to remove noisy depths.) Because processing is independent per pixel, it can be performed in parallel, with the most computationally expensive step being rendering to the original depth maps to detect free space violations. An example is shown in Fig. 5, while a point cloud made of a several fused depth maps is shown in Fig. 7.

Photometric Stereo Mapping Below 20-30 m deep in the water column, the sun's rays diminish significantly. To sufficiently illuminate the scene, AUVs are commonly equipped with independently controllable lights. Interestingly, what the AUVs perceive through their camera-imagery while their lights are on/off can provide information on the 3D structure and albedo of the scene.

The problem of estimating the visible scene given images illuminated by light sources is called the photometric stereo (PS) problem [55, 56]. The main principle of the PS algorithm is that a surface point's albedo and normal can be recovered by modeling the changes in that surface point's reflectance under various lighting source orientations. From previous work [57, 58], four images and their light correspondences ensure that surface points are illuminated sufficiently and uniquely. This configuration can be a very simple model to solve with the assumption that the camera never moves, the light orientations are known, and the surface material of the object in focus is also known. However, this assumption does not hold since the AUV is traveling underwater. We expand the PS algorithm to address the in-water light behavior as well as the AUV non-stationarity characteristic, that the camera is never still. Our complete model allows the AUV, with at least one camera, to estimate high-resolution 3D models of the scene by flicking on four different lights in sequence and capturing their respective images.

When light travels through water, it is continuously attenuated over distance as it collides with different particles, characterized by the uniqueness of the waterbody's properties. Therefore, the final image captured by the camera is composed of the direct signal – the light that traveled from a light source, interacted with the scene, and reflected back to the image sensor – and backscatter – the light that traveled from a light source and reflected back from particles not part of the scene. We refer to our prior work [59] for a deeper explanation of the image formation model. The attenuation parameters in the image formation model can be calibrated prior to deployment with a color chart or a black and white marker.

How the light reflects from the scene depends on the object's material. Most underwater PS works assume that the underwater scene is comprised of Lambertian-type surfaces. Simply, the amount of light reflected from the surface is independent of the viewing direction. It is only dependent on the incoming light's intensity and the angle between the light's direction and the surface normal.

The PS objective function is referred to as photometric consistency [58]:

$$o(\mathbf{n}, Z) = \frac{\sum_{i=1}^{N_I}(|I_i - I_i'(\mathbf{n}, Z)|)}{N_I} \tag{1}$$

As the name dictates, the goal is to minimize the difference between the predicted I' and observed I pixel values of a scene point in the set of N_I images, by estimating the scene point's surface normal $\mathbf{n}$ and distance from camera Z.

As the AUV is non-stationary, the cameras' and collocated lights' changing orientations need to be considered accordingly and fed into the PS model. If the AUV has only one camera, the PS framework requires additional pose estimation information that could be estimated using a SLAM system (e.g., Monocular ORB-SLAM [60]). In our case with Aqua2 equipped with two cameras, pose can be estimated as presented above in Section 3.1. The estimated depth maps provided by the stereo cameras can be utilized as initial guesses for the PS model – thus, the PS model can be used to improve the overall 3D scene modelling capabilities. By solving for the unknowns in the non-stationary PS model, one can derive the 3D scene or the distances of the scene points from the camera.

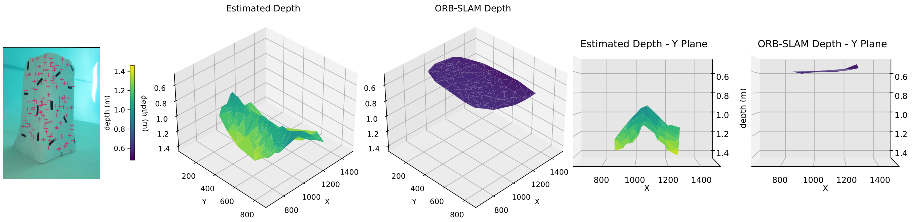


Fig. 8: Non-stationary photometric stereo model results given AUV with one camera and pose estimation and initial depth map from ORB-SLAM [60].

Results: Simple tests were performed with an AUV (BlueROV2) integrated with a single camera and four lights. Pose estimates and initial depth maps are provided via ORB-SLAM [60]. Fig. 8 shows the non-stationary photometric stereo modeling results of a synthetic rock viewed at its edge. Eight images were captured, four of them with lights on and four corresponding lights off. As ambient light was present, the images with lights off are integral for subtracting the ambient illumination within the PS model. These results show the capability of using lights to improve the 3D modeling of scenes when initial depth maps and pose estimations are available.

3.3 AUV Navigation

With localization and mapping techniques, the next goal is to effectively coordinate the proximal and distal observers for a fully integrated system. The two AUVs start at a location near the wreck, with the proximal observer in the front and the distal observer behind. Initially, the two robots convoy together [61] approaching a target structure such as a wreck; the proximal leads and the distal follows keeping the proximal inside its field of view. When the two robots have reached the starting positions, they start mapping the target structure in a collaborative way by assuming two different behaviors: the proximal observer focuses on covering the target structure in close proximity collecting high resolution observations, while the distal observer focuses on tracking the proximal observer from a larger distance maintaining a more informative and macroscopic perspective of the general structure and aiding the localization of the proximal robot simultaneously.

Multi-robot operations in the proposed scheme could expand from reactive coordination for each robot independently by assuming absence or very limited explicit communication, to high-level deliberative collaboration with increased communication capabilities. Addressing the specifics of the communication (acoustic or light-based) is an interesting future research direction; here, we discuss how the coordinated navigation between proximal and distal observers can be adapted according to different levels of communication.

Proximal Observer Navigation As a base case, we assume no communication from the distal to the proximal observer, thus the proximal observer will greedily attempt to cover the entirety of the target structure. Several exploration or mapping strategies could be deployed where the proximal observer actively decides which areas to visit with an informed global planner, but for the sake of simplicity in order to showcase the fundamental concept of the proximal observer, the robot will follow a predefined lawnmower pattern in close proximity to the structure. The proximal observer will have to operate in close proximity to highly unstructured environments through a complex terrain, therefore, instead of blindly reaching local goals, it has to avoid obstacles. For this purpose, it can utilize AquaNav [62], a state-of-the-art real-time vision-based underwater navigation framework developed in our previous work that enabled underwater robots to navigate safely through challenging 3D environments. Though, since the main purpose of the proposed framework is to map and not just navigate challenging terrains, to maximize observations of the target structure, AquaVis [25], an extension of AquaNav that performs active perception on automatically extracted visual objectives, is employed instead.

Distal Observer Navigation The distal observer should move in a way that will keep a distance from both the target structure and the proximal observer while following the latter. To follow the proximal observer, the distal observer in absence of communication could utilize motion predictors. On the other hand, assuming limited communication, the proximal observer could publish the trajectory it plans to follow, allowing for more informed decisions. Then, the distal observer could employ the strong capabilities of AquaVis to track the proximal observer by processing as visual objectives the future positions of the target robot, while also avoiding collisions with other potential objects. At this stage, AquaVis was modified to consider only the expected position of the proximal observer at each corresponding state and produce a solution that ensures visibility of the target robot at all times, assuming no occlusions. Finally, for simplicity the distal observer will be moving towards global waypoints in a similar pattern as the proximal observer. More deliberative and informed policies will be investigated in the future.

Results: Fig. 9 presents the trajectory of the distal (purple) and proximal (red) observers as they cover the deck of a USS YP-389 shipwreck [63] model (187.5m × 63.6m × 29.7m) simulated in gazebo with seabed at depth 39.7m. The dynamics of the AUVs together with the selection of viewpoints result in a variety of poses such that visual objectives (the wreck for the proximal, and the proximal AUV for the distal) are kept in the center of the field of view. The proximal AUV utilizes a lawnmower pattern as the starting point for AquaVis [25] while the distal observer uses the states of the planned trajectory of the proximal as the objectives. The simulation takes about 10 minutes with both robots trying to maintain a forward velocity of 0.5m/s.

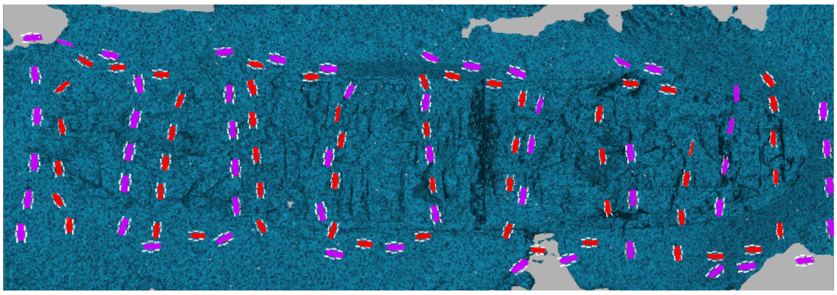


Fig. 9: Two AUVs exploring a wreck. Multiple snapshots combined to illustrate the two trajectories. The distal observer, in purple, keeps a large portion of the wreck and the proximal observer in view hovering above. The proximal observer, in red, utilizes a lawnmower pattern to cover the top of the wreck.

4 Conclusions, Lessons Learned, and Future Directions

While each component is necessary to achieve photorealistic mapping of underwater structures, the results presented here provide insights for the integration and research questions that are interesting to pursue:

- What is the trade-off between real-time 3D reconstructions and accurate ones? The underwater robots need enough information to navigate safely around the underwater structures; at the same time the more details added will contribute to denser reconstructions. We will encode hybrid hierarchical representations that can be used from the different components.

- How to optimize the AUV navigation given the number of conflicting optimization criteria? Examples of criteria include fine-grained coverage of the structure to ensure accurate reconstruction; minimization of overlaps; ensuring that the robots' state estimates do not diverge. We will explore Pareto-optimal decision framework.
- How to optimize communication between robots? Here, we did not explicitly discuss the communication bandwidth available; in general, communication is very limited underwater – with a bandwidth in the order of kb/s. It is important to identify efficient data representation of the 3-D reconstruction and on a cross-layer optimization for deciding when and how to share.

Once we integrate these different components, our plan is to deploy the system in an archaeological expedition to map large shipwrecks.

In this work, a novel multi-robot approach was presented for mapping of large and challenging underwater structures, such as shipwrecks, or energy and aquaculture infrastructure. The main contribution of this work is twofold: (a) to present the main components necessary for enabling such mapping – i.e., 1) robust state estimation, where we presented a robust system that is able to switch between state estimators according to their reliability, 2) dense mapping, where we presented approaches that can run in real-time and and with low-cost sensor configuration, 3) team coverage with distal and proximal observers – and (b) discussing the insights and interesting research questions, towards a fully-integrated system for underwater structure mapping.

Acknowledgments

The authors would like to thank the National Science Foundation for its generous support (NSF 1943205, 1919647, 2024741, 2024541, 2024653, 2144624) and the Research Council of Norway for support through the project Resilient Robotic Autonomy for Underwater Operations in Fish Farms (ResiFarm, NO-327292). We also thank Halcyon Dive Systems for their support with equipment.

References

1. Mai, C., Pedersen, S., Hansen, L., Jepsen, K.L., Yang, Z.: Subsea infrastructure inspection: A review study. In: 2016 IEEE International Conference on Underwater System Technology: Theory and Applications (USYS). pp. 71–76. IEEE (2016)
2. Maurelli, F., Carreras, M., Salvi, J., Lane, D., Kyriakopoulos, K., Karras, G., Fox, M., Long, D., Kormushev, P., Caldwell, D.: The PANDORA project: A success story in AUV autonomy. In: OCEANS 2016-Shanghai. pp. 1–8. IEEE (2016)
3. Palomer, A., Ridao, P., Ribas, D.: Inspection of an underwater structure using point-cloud SLAM with an AUV and a laser scanner. Journal of field robotics 36(8), 1333–1344 (2019)
4. Rekleitis, I., Dudek, G., Milios, E.E.: On multiagent exploration. In: Vision Interface. pp. 455–461. Vancouver, BC, Canada (Jun 1998)
5. Kurazume, R., Hirose, S.: An experimental study of a cooperative positioning system. Autonomous Robots 8, 43–52 (2000)

6. Koenig, N., Howard, A.: Design and use paradigms for Gazebo, an open-source multi-robot simulator. In: Proc. IROS. pp. 2149–2154 (2004)
7. Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., Furgale, P.: Keyframe-based visual-inertial odometry using nonlinear optimization. Int. Journal of Robotics Research 34(3), 314–334 (2015)
8. Mourikis, A.I., Roumeliotis, S.I.: A multi-state constraint Kalman filter for vision-aided inertial navigation. In: Proc. ICRA. pp. 3565–3572. IEEE (2007)
9. Campos, C., Elvira, R., Rodríguez, J.J.G., Montiel, J.M., Tardós, J.D.: ORB-SLAM3: An accurate open-source library for visual, visual-inertial and multi-map SLAM. IEEE Trans. on Robotics (2021)
10. Rosinol, A., Abate, M., Chang, Y., Carlone, L.: Kimera: an open-source library for real-time metric-semantic localization and mapping. In: ICRA. IEEE (2020)
11. Joshi, B., Rahman, S., Kalaitzakis, M., Cain, B., Johnson, J., Xanthidis, M., Karapetyan, N., Hernandez, A., Quattrini Li, A., Vitzilaios, N., Rekleitis, I.: Experimental Comparison of Open Source Visual-Inertial-Based State Estimation Algorithms in the Underwater Domain. In: Proc. IROS. pp. 7221–7227 (2019)
12. Quattrini Li, A., Coskun, A., Doherty, S.M., Ghasemlou, S., Jagtap, A.S., Modasshir, M., Rahman, S., Singh, A., Xanthidis, M., O'Kane, J.M., Rekleitis, I.: Experimental comparison of open source vision based state estimation algorithms. In: Proc. ISER (2016)
13. Rahman, S., Quattrini Li, A., Rekleitis, I.: SVIn2: An Underwater SLAM System using Sonar, Visual, Inertial, and Depth Sensor. In: Proc. IROS (2019)
14. Roumeliotis, S.I., Rekleitis, I.: Propagation of uncertainty in cooperative multi-robot localization: Analysis and experimental results. Autonomous Robots 17(1), 41–54 (Jul 2004)
15. Aloimonos, J., Weiss, I., Bandyopadhyay, A.: Active vision. Int. Journal of Computer Vision 1(4), 333–356 (1988)
16. Bajcsy, R.: Active perception. Tech. Rep. MSCIS-88-24, Un. of Pennsylvania Department of Computer and Information Science (1988)
17. Feder, H.J.S., Leonard, J.J., Smith, C.M.: Adaptive mobile robot navigation and mapping. The Int. Journal of Robotics Research 18(7), 650–668 (1999)
18. Frolov, S., Garau, B., Bellingham, J.: Can we do better than the grid survey: Optimal synoptic surveys in presence of variable uncertainty and decorrelation scales. Journal of Geophysical Research: Oceans 119(8), 5071–5090 (2014)
19. Chaves, S.M., Kim, A., Galceran, E., Eustice, R.M.: Opportunistic sampling-based active visual slam for underwater inspection. Autonomous Robots (2016)
20. Girdhar, Y., Giguere, P., Dudek, G.: Autonomous adaptive exploration using realtime online spatiotemporal topic modeling. Int. Journal of Robotics Research 33(4), 645–657 (2014)
21. Karapetyan, N., Johnson, J., Rekleitis, I.: Coverage path planning for mapping of underwater structures with an autonomous underwater vehicle. In: MTS/IEEE OCEANS - Singapore. Singapore (Virtual) (2020)
22. Karapetyan, N., Johnson, J., Rekleitis, I.: Human diver-inspired visual navigation: Towards coverage path planning of shipwrecks. Marine Technology Society Journal, "Best of OCEANS 2020" 55(4), 24–32 (2021)
23. Manderson, T., Higuera, J.C.G., Wapnick, S., Tremblay, J.F., Shkurti, F., Meger, D., Dudek, G.: Vision-based goal-conditioned policies for underwater navigation in the presence of obstacles. Robotics: Science and Systems (2020)
24. Vidal, E., Palomeras, N., Istenič, K., Gracias, N., Carreras, M.: Multisensor online 3d view planning for autonomous underwater exploration. Journal of Field Robotics 37(6), 1123–1147 (2020)

25. Xanthidis, M., Kalaitzakis, M., Karapetyan, N., Johnson, J., Vitzilaios, N., O'Kane, J., Rekleitis, I.: Aquavis: A perception-aware autonomous navigation framework for underwater vehicles. In: Proc. IROS. pp. 5387–5394 (2021)
26. Wenhardt, S., Deutsch, B., Angelopoulou, E., Niemann, H.: Active Visual Object Reconstruction using D-, E-, and T-Optimal Next Best Views. In: Proc. CVPR (2007)
27. Potthast, C., Sukhatme, G.S.: A probabilistic framework for next best view estimation in a cluttered environment. Journal of Visual Communication and Image Representation 25(1), 148–164 (2014)
28. Kim, A., Eustice, R.M.: Active visual SLAM for robotic area coverage: Theory and experiment. The Int. Journal of Robotics Research 34(4-5), 457–475 (2015)
29. Daudelin, J., Campbell, M.: An adaptable, probabilistic, next-best view algorithm for reconstruction of unknown 3-d objects. IEEE Robotics and Automation Letters 2(3), 1540–1547 (2017)
30. Fraundorfer, F., Heng, L., Honegger, D., Lee, G.H., Meier, L., Tanskanen, P., Pollefeys, M.: Vision-based autonomous mapping and exploration using a quadrotor mav. In: Proc. IROS. pp. 4557–4564. IEEE (2012)
31. Vidal, E., Hernández, J.D., Istenič, K., Carreras, M.: Online view planning for inspecting unexplored underwater structures. IEEE Robotics and Automation Letters 2(3), 1436–1443 (2017)
32. Hepp, B., Nießner, M., Hilliges, O.: Plan3d: Viewpoint and trajectory optimization for aerial multi-view stereo reconstruction. ACM Transactions on Graphics (TOG) 38(1), 4 (2018)
33. Smith, N., Moehrle, N., Goesele, M., Heidrich, W.: Aerial path planning for urban scene reconstruction: a continuous optimization method and benchmark. In: SIGGRAPH Asia. p. 183 (2018)
34. Shade, R., Newman, P.: Choosing where to go: Complete 3d exploration with stereo. In: Proc. ICRA. pp. 2806–2811 (2011)
35. Hollinger, G.A., Englot, B., Hover, F.S., Mitra, U., Sukhatme, G.S.: Active planning for underwater inspection and the benefit of adaptivity. The Int. Journal of Robotics Research 32(1), 3–18 (2013)
36. Golodetz, S., Cavallari, T., Lord, N.A., Prisacariu, V.A., Murray, D.W., Torr, P.H.: Collaborative large-scale dense 3d reconstruction with online inter-agent pose optimisation. IEEE transactions on Visualization and Computer Graphics 24(11), 2895–2905 (2018)
37. Kapoutsis, A.C., Chatzichristofis, S.A., Doitsidis, L., de Sousa, J.B., Pinto, J., Braga, J., Kosmatopoulos, E.B.: Real-time adaptive multi-robot exploration with application to underwater map construction. Autonomous robots 40(6) (2016)
38. Paull, L., Seto, M., Leonard, J.J., Li, H.: Probabilistic cooperative mobile robot area coverage and its application to autonomous seabed mapping. The International Journal of Robotics Research 37(1), 21–45 (2018)
39. Dudek, G., Jenkin, M., Prahacs, C., Hogue, A., Sattar, J., Giguere, P., German, A., Liu, H., Saunderson, S., Ripsman, A., Simhon, S., Torres-Mendez, L.A., Milios, E., Zhang, P., Rekleitis, I.: A visually guided swimming robot. In: Proc. IROS. pp. 1749–1754 (2005)
40. Huang, G.: Visual-inertial navigation: A concise review. In: Proc. ICRA (2019)
41. Shkurti, F., Rekleitis, I., Scaccia, M., Dudek, G.: State estimation of an underwater robot using visual and inertial information. In: Proc. IROS. pp. 5054–5060 (2011)
42. Rahman, S., Quattrini Li, A., Rekleitis, I.: Sonar Visual Inertial SLAM of Underwater Structures. In: IEEE Int. Conf. on Robotics and Automation. pp. 5190–5196. Brisbane, Australia (May 2018)

43. Qin, T., Cao, S., Pan, J., Shen, S.: A general optimization-based framework for global pose estimation with multiple sensors. arXiv preprint arXiv:1901.03642 (2019)
44. Rahman, S., Quattrini Li, A., Rekleitis, I.: SVIn2: A Multi-sensor Fusion-based Underwater SLAM System. International Journal of Robotics Research (July 2022)
45. Joshi, B., Modasshir, M., Manderson, T., Damron, H., Xanthidis, M., Quattrini Li, A., Rekleitis, I., Dudek, G.: Deepurl: Deep pose estimation framework for underwater relative localization. In: IROS. pp. 1777–1784. Las Vegas, NV (2020)
46. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: NeurIPS (2014)
47. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 4104–4113 (2016)
48. Schönberger, J.L., Zheng, E., Frahm, J.M., Pollefeys, M.: Pixelwise view selection for unstructured multi-view stereo. In: Eur. Conf. on Computer Vision (2016)
49. Hirschmüller, H.: Stereo processing by semiglobal matching and mutual information. PAMI 30(2), 328–341 (2008)
50. Spangenberg, R., Langner, T., Adfeldt, S., Rojas, R.: Large scale semi-global matching on the CPU. In: IEEE Intelligent Vehicles Symposium Proceedings. pp. 195–201 (2014)
51. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. IJCV 47(1-3), 7–42 (2002)
52. Hu, X., Mordohai, P.: A quantitative evaluation of confidence measures for stereo vision. PAMI 34(11), 2121–2133 (2012)
53. Merrell, P., Akbarzadeh, A., Wang, L., Mordohai, P., Frahm, J.M., Yang, R., Nistér, D., Pollefeys, M.: Real-time visibility-based fusion of depth maps. In: ICCV (2007)
54. Hu, X., Mordohai, P.: Least commitment, viewpoint-based, multi-view stereo. In: 3DIMPVT. pp. 531–538 (2012)
55. Woodham, R.J.: Photometric method for determining surface orientation from multiple images. Optical Engineering 19(1), 139 – 144 (1980)
56. Drbohlav, O., Chaniler, M.: Can two specular pixels calibrate photometric stereo? In: ICCV. vol. 2, pp. 1850–1857 (2005)
57. Tsiotsios, C., Angelopoulou, M.E., Kim, T.K., Davison, A.J.: Backscatter compensated photometric stereo with 3 sources. In: CVPR (2014)
58. Tsiotsios, C., Kim, T., Davison, A., Narasimhan, S.: Model effectiveness prediction and system adaptation for photometric stereo in murky water. Computer Vision and Image Understanding 150, 126–138 (2016)
59. Roznere, M., Quattrini Li, A.: Real-time model-based image color correction for underwater robots. In: Proc. IROS (2019)
60. Mur-Artal, R., Montiel, J.M.M., Tardós, J.D.: ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Trans. Robot. 31(5), 1147–1163 (2015)
61. Shkurti, F., Chang, W.D., Henderson, P., Islam, M.J., Higuera, J.C.G., Li, J., Manderson, T., Xu, A., Dudek, G., Sattar, J.: Underwater multi-robot convoying using visual tracking by detection. In: Proc. IROS. pp. 4189–4196 (2017)
62. Xanthidis, M., Karapetyan, N., Damron, H., Rahman, S., Johnson, J., O'Connell, A., O'Kane, J., Rekleitis, I.: Navigation in the presence of obstacles for an agile autonomous underwater vehicle. In: Proc. ICRA. pp. 892–899 (2020)
63. NOAA National Marine Sanctuaries: Shipwrecks. https://monitor.noaa.gov/shipwrecks/, accessed: 2022-04-14