Visual-Acoustic SLAM for Underwater Caves

Sharmin Rahman[1], Alberto Quattrini Li[2], and Ioannis Rekleitis[1]

***Abstract*— Underwater caves are extremely challenging environment for *perception*, due to the absence of natural light and the highly unstructured nature of such environment, making it also dangerous and cognitively heavy even for highly skilled divers. This paper presents an overview of our previous works for underwater cave mapping which combines data from multiple sensors to assist the divers by reducing the cognitive loads. The challenges of the underwater environment augmented by the complete absence of natural light and the effects of sharp shadows are discussed together with the contributions of the different sensing modalities. A tightly-coupled keyframe-based SLAM framework with loop-closing and relocalization capabilities combining visual, inertial, depth, and acoustic sensors has been described together with the design of a sensor suite for collecting data in the challenging environment of underwater caves. Experimental results illustrate the accuracy and robustness of the proposed methodology from a cavern at Ballroom, Ginnie Springs, FL, USA.**

I. Introduction

In this paper, we show our efforts towards underwater cave reconstruction – starting from offline stereo vision only 3D reconstruction to the current state where we combine Sonar, visual, inertial, and water pressure information for real-time robust state estimation. We discuss the challenges of underwater environments; the feasibility of using and combining different sensors for robotic operations; present results from those methods and provide future work directions.

Exploration of underwater environments with autonomous robots could assist us in a variety of scenarios, ranging from historical studies to health monitoring of coral reef and underwater infrastructure inspection – e.g., bridges, hydroelectric dams, water supply systems and oil rigs. More specifically, mapping underwater structures – caves, shipwrecks, etc. – is crucial for several fields, such as, marine archaeology, Search and Rescue (SaR), resource management, hydrogeology, and speleology, and has many broader impacts in terms of economy, conservation, and scientific discoveries. However, due to the highly unstructured nature of such environments, navigation by human divers could be extremely dangerous and labor intensive. Hence, employing an underwater robot is an excellent fit to build the map of the environment while simultaneously the robot localizes itself in the map.

Currently, most of the efforts for mapping caves are performed by divers that need to take measurements manually using a grid and measuring tape, or using hand-held sensors [1], and data is post-processed afterwards. Autonomous Underwater Vehicles (AUVs) present unique opportunities to automate this process; however, there are several open problems that still need to be addressed for reliable deployments, including real-time robust Simultaneous Localization and Mapping (SLAM).

Fig. 1. Cave in Mexico with the dive light illuminating part of the walls.

Underwater environment suffers from light and color attenuation, haze, scattering, and dynamic obstacles, such as marine life and particulates. In addition to the above challenges, in underwater cave environments there is a complete absence of natural light. The only lighting available comes from artificial lights brought in by the robot or by divers; see Fig. 1 where a cave is illuminated by a strong video light. It is worth noticing in Fig. 1 a second narrow beam of blueish light that extends from the top of the image; this beam is produced by a dive light held by a second diver. The complete lack of natural (ambient) light results in harsh shadows. In contrast to most vision applications where the light remains constant, at least for brief periods of time, in the cave environment, the light source (or sources) move as much as the camera. In some configurations the light source is carried by a different person – see Fig. 1 – while in other the light is attached (albeit not rigidly) to the sensing apparatus – see Fig. 2(a) and Fig. 2(c).

In order to produce a robust and accurate estimate of the pose of the sensors and a map of the environment we augment a state-of-the-art Visual-Inertial state estimation package, OKVIS [2], with acoustic and depth sensor data, and with loop closing capabilities. In particular, a mechanical scanning sonar, which returns range measurements based on acoustic information, and a depth sensor, which provides depth measurement from the water pressure, are introduced to aid the visual-inertial system. Furthermore, a pre-pro-

[1]S. Rahman and I. Rekleitis are with the Computer Science and Engineering Department, University of South Carolina, Columbia, SC, USA srahman@email.sc.edu, yiannisr@cse.sc.edu

[2]A. Quattrini Li is with the Department of Computer Science, Dartmouth College, Hanover, NH, USA alberto.quattrini.li@dartmouth.edu

cessing step is introduced to alleviate the water effects on the visual data. The different components are modular so depending on the application, they can be activated on demand. In the experimental data collected in an underwater cave, we qualitatively show that the integration of multiple sensors improves the quality of the state estimation and provides a real-time dense 3D reconstruction.

II. Related Work

Acoustic sensors have been the first choice for underwater SLAM and navigation for a long time. Using such sensors, – e.g., Doppler Velocity Log (DVL), ultra-short baseline (USBL), and multibeam imaging sonar – many underwater navigation algorithms [3], [4], [5], [6], [7] have been developed. A recent example includes Sunfish [8] – a human-portable autonomous underwater vehicle capable of planning, SLAM, exploration, and control. The AUV performed a successful underwater cave exploration using a real-time SLAM system combining expensive sensors, including a multibeam sonar which provides a fan of sonar beams, an underwater dead-reckoning system based on a fiber-optic gyroscope (FOG) IMU, acoustic DVL, and pressure-depth sensors.

In the last few years, both pure visual and visual-inertial odometry (VO and VIO, respectively) systems have gained maturity and are capable of performing real-time accurate navigation for indoor and outdoor environments covering a large area. Though designed for small scale indoor environment, PTAM [9] is the first one of such algorithms which provides method for mapping based on *keyframes*, efficient tracking and mapping running in two parallel threads, camera pose estimation for every frame, and relocalization after tracking failure. Later on, other VO systems, based on both direct method and indirect (feature-based) methods, have been developed – e.g., ORB-SLAM [10], SVO [11], LSD-SLAM [12], DSO [13]. For improved accuracy and robustness, filtering based – e.g., MSCKF [14], ROVIO [15], REBiVO [16] – and non-linear optimization based – e.g., OKVIS [2], visual inertial ORB-SLAM [17], VINS-Mono [18] – VIO systems have been developed showing excellent performance.

However, due to low visibility, low contrast, haze, and scattering, vision-based navigation and exploration is very challenging and often results into failure. In our recent work [19], we present a comprehensive study and performance analysis of state-of-the-art open-source visual odometry algorithms in different underwater environments. Most of the vision based underwater navigation algorithms are developed for experiments in open areas with natural lighting or artificial lighting that completely illuminates the field-of-view. However, in the highly unstructured nature of underwater environment, data collection and exploration based on DVL and sonar while diving is expensive and sometimes also not suitable. Corke et al. [20] compared acoustic and visual methods for underwater localization showing the viability of using visual methods underwater in some scenarios. Hence, combination of visual and acoustic sensor opens the scope for the design and development of underwater navigation and mapping algorithms using both sensors. Our proposed approach [21], [22] shows the feasibility of such a technique in underwater domain.

III. Technical Approach

In this section, first we present a brief overview of a custom-made sensor suite used in the data collection process. Second, we describe our tightly-coupled non-linear optimization based SLAM system with loop-closing and relocalization capabilities fusing Sonar, visual, inertial, and pressure sensor.

A. Sensor Suite Overview

The sensor suite is equipped with two IDS UI-3251LE cameras in a stereo configuration, Microstrain 3DM-GX4-15 IMU, Bluerobotics Bar30 pressure sensor, Intel NUC which has Linux operating system and runs Robot Operating System (ROS) [23], and IMAGENEX 831L mechanical scanning Sonar. The cameras are synchronized by a trigger which captures 15 frames per second; the Sonar provides *range* and *heading* information by scanning over a plane over $360°$, with angular resolution of $0.9°$; the IMU provides angular velocity and linear acceleration measurements at a rate of $100\,\mathrm{Hz}$; and the depth sensor gives water-pressure measurements at $1\,\mathrm{Hz}$. A 5-inch LED display has been added to provide visual feedback to the diver as well as to interact via AR tags [24] to start/stop recording, change camera and Sonar parameter, and to run the SLAM algorithm onboard. The PVC tube enclosure for the electronics has been designed and tested to ensure the device is waterproof up to 100 meters. One of the design criteria of such a sensor suite was to ensure the ease of deployment in different modes – e.g., hand-held, single or dual Diver Propulsion Vehicle (DPV) – see Fig. 2. Please refer to our work [25], [26] for the detailed specifications of the hardware and software components.

B. Notations and States

In the state estimation framework, the reference frames are denoted as C for Camera, I for IMU, D for Depth, S for Sonar, and W for World. Let us denote ${}_X\mathbf{T}_Y = [{}_X\mathbf{R}_Y | {}_X\mathbf{p}_Y]$ the homogeneous transformation matrix between two arbitrary coordinate frames X and Y, where ${}_X\mathbf{R}_Y$ denotes the rotation matrix with corresponding quaternion ${}_X\mathbf{q}_Y$ and ${}_X\mathbf{p}_Y$ represents the position vector.

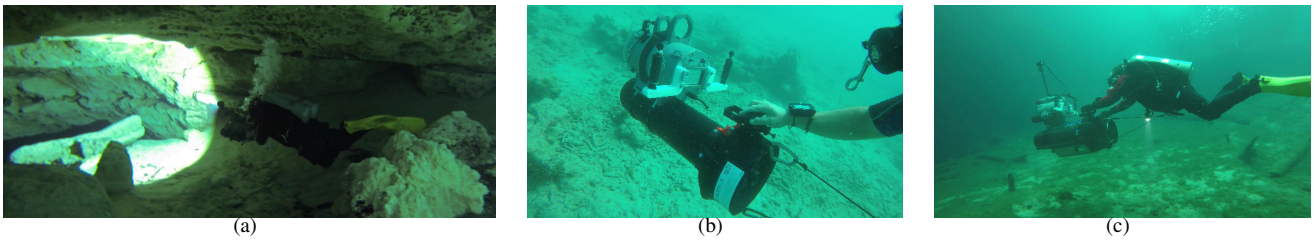
The state of the robot R is defined as $\mathbf{x}_R$:

$$\mathbf{x}_R = [{}_W\mathbf{p}_I^T, {}_W\mathbf{q}_I^T, {}_W\mathbf{v}_I^T, {\mathbf{b}_g}^T, {\mathbf{b}_a}^T]^T \quad (1)$$

Containing the position ${}_W\mathbf{p}_I$, the quaternion ${}_W\mathbf{q}_I$, the linear velocity ${}_W\mathbf{v}_I$. All of them are defined in the IMU reference frame I with respect to the world reference frame W. Furthermore, the gyroscope and accelerometer bias $\mathbf{b}_g$ and $\mathbf{b}_a$ are also estimated and placed in the state vector.

The corresponding error-state vector is defined in minimal coordinates, while the perturbation takes place in the tangent space:

$$\delta\boldsymbol{\chi}_R = [\delta\mathbf{p}^T, \delta\mathbf{q}^T, \delta\mathbf{v}^T, \delta{\mathbf{b}_g}^T, \delta{\mathbf{b}_a}^T]^T \quad (2)$$



(a) (b) (c)

Fig. 2. Deployment methods of the Stereo Rig (a) hand-held (b) on a single DPV, (c) on dual DPV.

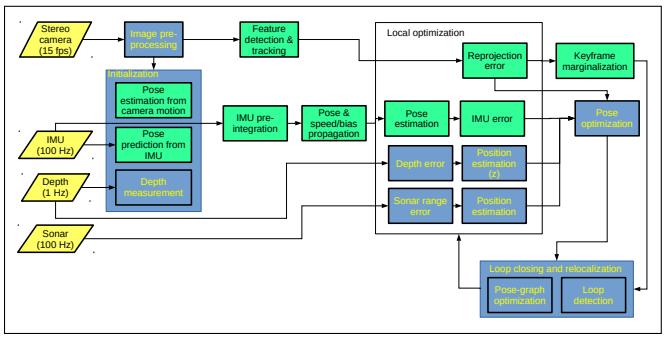


Fig. 3. Overview of the proposed approach. In yellow are the sensor feeds and their frequency; in green the OKVIS [2] components; in blue the components introduced to handle acoustic and depth data, underwater visual effects, and loop closure.

C. Tightly-coupled SLAM system using Sonar, visual, inertial, and pressure sensor

As shown in Fig. 3, we augmented OKVIS [2], a state-of-the-art open-source visual-inertial package to combine acoustic and depth information. In [22], we introduce a two-step refinement of the *scale* at the initialization: we refine the initial scale factor from the stereo camera using depth measurements, which is further refined by aligning the IMU measurements with stereo vision.

The cost function $J(\mathbf{x})$ for the tightly-coupled non-linear optimization includes the IMU error $\mathbf{e}_s$, the reprojection error $\mathbf{e}_r$, the depth error e_u and the sonar error $\mathbf{e}_t$, :

$$J(\mathbf{x}) = \sum_{i=1}^{2}\sum_{k=1}^{K}\sum_{j\in\mathcal{J}(i,k)} \mathbf{e}_r^{i,j,k^T}\mathbf{P}_r^k\mathbf{e}_r^{i,j,k} + \sum_{k=1}^{K-1}\mathbf{e}_s^{k^T}\mathbf{P}_s^k\mathbf{e}_s^k + \sum_{k=1}^{K-1}\mathbf{e}_t^{k^T}\mathbf{P}_t^k\mathbf{e}_t^k + \sum_{k=1}^{K-1}e_u^{k^T}P_u^k e_u^k \quad (3)$$

with i denoting the camera index – $i = 1$ for left, $i = 2$ for right camera used in a stereo camera – and the landmark index j observed in the k^{th} camera frame. $\mathbf{P}_r^k$, $\mathbf{P}_s^k$, P_u^k, and $\mathbf{P}_t^k$ denote the information matrix of visual landmarks, IMU, depth, and sonar range measurement for the k^{th} frame respectively.

The reprojection error is calculated based on the difference between a keypoint measurement in the camera coordinate frame C and the corresponding landmark back-projection based on the stereo projection model. The IMU error term combines all the accelerometer and gyroscope measurements utilizing the *IMU pre-integration* approach described by Forster *et al.* [27] between successive camera frames and represents the *pose*, *speed*, and *bias* error between the prediction based on the previous and the current states. Both the reprojection error and the IMU error term follow the formulation described by Leutenegger *et al.* [2].

The sonar range error, introduced in our previous work [21], represents the difference between the 3D point that can be derived from the range measurement and a corresponding visual feature in 3D. In poor visibility and low contrast environment where vision fails to detect features, Sonar provides additional features and helps in mapping the surroundings. The depth error term can be calculated as the difference between the rig position along the z direction and the water depth measurement provided by a pressure sensor. Depth values are extracted along the *gravity* direction which is aligned with the z of the world W – observable due to the tightly coupled IMU integration. This can correct the position of the robot along the z axis. For the detailed formulation of the above error terms, please refer to our previous work [21], [22].

Ceres Solver nonlinear optimization framework [28] optimizes $J(\mathbf{x})$ then to estimate the state of the system.

Loop-closing and *relocalization* is achieved using the binary bag-of-words place recognition module DBoW2 [29]. A pose-graph maintains the connections between keyframes where a node represents a keyframe and an edge between two keyframes exists if there is sufficient overlap between them. With every new frame in the local window, the loop-closing module searches for loop candidates in the BoW database. When a candidate is found with enough match, feature correspondences are obtained to establish connection between the current frame and the loop candidate frame. Then, a PnP RANSAC is performed to obtain the geometric validation. The relocalization module is responsible for aligning the current keyframe pose in the local window with the loop candidate keyframe by sending the drift in pose to the windowed sonar-visual-inertial-depth optimization thread.

In our contour based reconstruction [30] for underwater environment, we showed that the cone-of-light created in the boundary of the light and dark area due to the lighting variations and the prominent edges in the scene help in denser 3D reconstruction. Therefore, we employ all three types of features – i.e., tracked features in the local window,

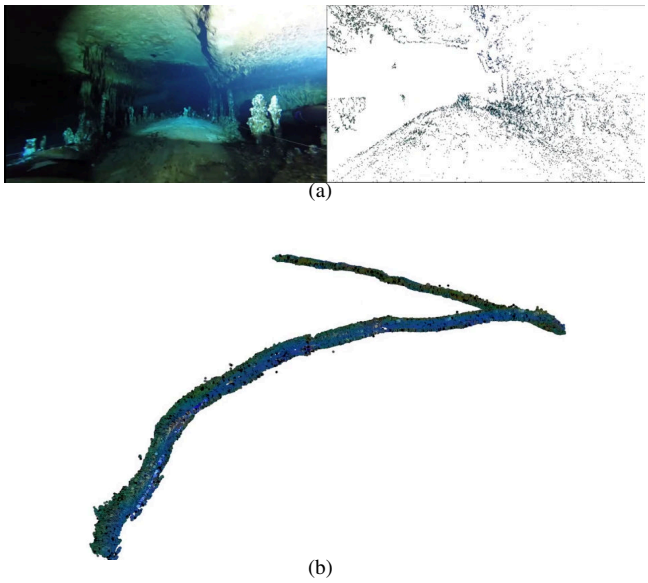


Fig. 4. Early experiments with vision only mapping of an underwater cave. (a) Sample image and the reconstruction from the shadow contours generated by the video light over consecutive frames. (b) Reconstruction of an eight minute sequence (approximately $240\,\mathrm{m}$) through a cave in Sistema Camillo, Mexico.

Sonar, and contour features in our tightly couple formulation. The steps involve: 1) computing adaptive thresholding based on the histogram analysis for the light and dark areas; 2) edge detection and filtering to retain the larger continuous contours, and 3) lastly finding the stereo matches of the contour features. Because of the lighting variation and shadow, using direct or indirect method for contour tracking could lead to incorrect estimation and even tracking failure. As such, we used only stereo projection for the matched contour features which is followed by a local bundle adjustment (BA). Please refer to our paper [31] for the details.

IV. Results

Our method shows robust state estimation in challenging underwater environments. We report few examples in caves and caverns, characterized by complete absence of natural light.

Fig. 4 shows the early attempts for underwater cave reconstruction where we exploited the cone-of-light to reconstruct the cave wall using only stereo vision [30]. The point clouds from the individual frames are aligned with the odometry from ORB-SLAM2 [32] *offline* to generate the $240\,\mathrm{m}$ long trajectory in Mexico underwater cave.

The reconstruction result from the full integrated system, with loop closure and semi-dense reconstruction in an underwater cavern in Ginnie Springs (FL) covering a $59\,\mathrm{m}$ trajectory is shown in Fig. 5. The data has been collected by a diver with the sensor suite described in the previous section. As there is no available ground truth, we considered *loop-closure* as a metric of performance evaluation. The system is capable of effectively detecting the loop even if the environment presents self-similarities. In our previous work, we also compared the results with OKVIS [2], VINS-Mono [18], and MSCKF [33] for the validation of our method in the standard datasets as well as in the underwater datasets; please see [22].

V. Discussions and Future Works

An interesting line of future work is increasing the field-of-view (fov) of cameras for 3D reconstruction. This could be done by separating the cameras from stereo setup and arranging them in such a way so that each points away from the other to the opposite directions. This setup could help to perceive a lot of parallax for 3D reconstruction without affecting the *scale* as it can be disambiguated by depth and IMU. In addition, the placement of multiple video lights would also help to provide enough light for each camera. Increasing the number of cameras also could assist for this purpose. As such, in the next iteration of our sensor suite, we will investigate the optimal setup of cameras and video lights with the available on-board computational resources to perform real-time operations.

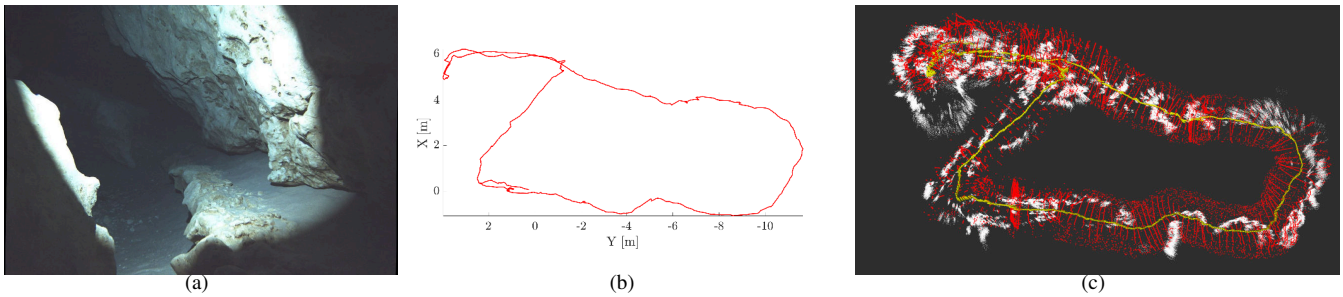
In this work, we used a mechanical scanning profiling sonar for *range* information by scanning over a plane. Combining *imaging* sonar with visual-inertial odometry is another interesting area worth exploring which could lead to several interesting applications based on acoustic information.

In the future, we plan to achieve ground truth of the collected data by placing AprilTags along the trajectory [34] for the assessment of our approach. Currently, to validate the *loop-closure* module we compare our method with original OKVIS along with other state-of-the-art SLAM packages in the benchmark datasets where ground truth is available [22]. For the underwater datasets, as there is no available ground truth, we rely on the information from the divers or measuring tape for a rough estimation of the trajectory. We also compared the performances of state-of-the-art SLAM packages in the underwater datasets where most of them fail to track or introduce large drifts in the trajectory over time [22], [19] but our approach successfully tracks.

VI. Conclusions

In this paper, we presented an underwater SLAM system fusing acoustic, visual, inertial, and depth information capable of running real-time on the available computational resource of a custom-made sensor suite. The visual inertial state estimation package, OKVIS has been extended to handle acoustic and depth data, and it was augmented with loop-closing capabilities.

During different field deployments using the sensor suite, it is clear that the turbidity of the water and the coloration of the cave walls have a major effect in the suitability of a vision-based state estimation approach. Fig. 6 illustrates the difference in two deployments occurred in Florida, USA. In Fig. 6(a), the waters are clear as the cave is the outlet of a spring. In addition, the walls are light-colored, almost white, as such the video light illuminates the walls in some distance; even with an over-exposed spot in the middle. In contrast, the same video light used in Turner Sink (Fig. 6(b))



(a) (b) (c)

Fig. 5. (a) Cave environment, Ballroom, Ginnie Springs, FL, USA, with a loop (b) trajectory from tightly coupled Sonar-visual-inertial-depth framework with loop-closing – where the diver started and stopped data collection at the same place (c) contour based reconstruction, red denotes Sonar features and white denotes stereo contour matched features.

barely illuminates the wall next to the sensor, and the light diffuses in the water. The main difference is that the walls in this cave are covered by some micro-organisms and have a dark coloration; in addition the water turbidity is rather high. Stronger and/or additional lights are required for mapping such caves.

The proposed method has been also deployed on an autonomous underwater vehicle (AUV), Aqua2 for real-time navigation which validates the robustness and improved accuracy of the estimated odometry; see Fig. 7 for an initial deployment of the Aqua2 vehicle at the Blue Grotto cavern in FL, USA. The limited field of view of the AUV poses a major challenge for the detection of obstacles especially in an enclosed environment, such as a cave. The current work is expected to open the door for a variety of planning and control related research of AUVs and remotely operated vehicles (ROVs) underwater.

ACKNOWLEDGMENT

The authors would like to thank the National Science Foundation for its support (NSF 1513203, 1637876).

REFERENCES

[1] J. Henderson, O. Pizarro, M. Johnson-Roberson, and I. Mahon, "Mapping submerged archaeological sites using stereo-vision photogrammetry," *International Journal of Nautical Archaeology*, vol. 42, no. 2, pp. 243–256, 2013.

[2] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *Int. J. Robot. Res.*, vol. 34, no. 3, pp. 314–334, 2015.

[3] J. J. Leonard and H. F. Durrant-Whyte, *Directed sonar sensing for mobile robot navigation*. Springer Science & Business Media, 2012, vol. 175.

[4] C.-M. Lee *et al.*, "Underwater navigation system based on inertial sensor and doppler velocity log using indirect feedback Kalman filter," *International Journal of Offshore and Polar Engineering*, vol. 15, no. 02, 2005.

[5] J. Snyder, "Doppler Velocity Log (DVL) navigation for observation-class ROVs," in *MTS/IEEE OCEANS, SEATTLE*, 2010, pp. 1–9.

[6] H. Johannsson, M. Kaess, B. Englot, F. Hover, and J. Leonard, "Imaging sonar-aided navigation for autonomous underwater harbor surveillance," in *Proc. IROS*. IEEE, 2010, pp. 4396–4403.

[7] P. Rigby, O. Pizarro, and S. B. Williams, "Towards geo-referenced AUV navigation through fusion of USBL and DVL measurements," in *OCEANS*, 2006, pp. 1–6.

[8] K. Richmond, C. Flesher, L. Lindzey, N. Tanner, and W. C. Stone, "SUNFISH®: A human-portable exploration AUV for complex 3D environments," in *MTS/IEEE OCEANS Charleston*, 2018, pp. 1–9.

[9] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *IEEE and ACM Int. Symp. on Mixed and Augmented Reality*, 2007, pp. 225–234.

[10] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, 2015.

[11] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "SVO: Semidirect Visual Odometry for Monocular and Multicamera Systems," *IEEE Trans. Robot.*, vol. 33, no. 2, 2017.

[12] J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *European conference on computer vision*. Springer, 2014, pp. 834–849.

[13] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 611–625, 2018.

[14] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proc. ICRA*. IEEE, 2007, pp. 3565–3572.

[15] M. Bloesch, M. Burri, S. Omari, M. Hutter, and R. Siegwart, "Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback," *Int. J. Robot. Res.*, vol. 36, pp. 1053–1072, 2017.

[16] J. J. Tarrio and S. Pedre, "Realtime edge based visual inertial odometry for MAV teleoperation in indoor environments," *J. Intell. Robot. Syst.*, pp. 235–252, 2017.

[17] R. Mur-Artal and J. D. Tardós, "Visual-inertial monocular SLAM with map reuse," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 796–803, 2017.

[18] T. Qin, P. Li, and S. Shen, "VINS-Mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, 2018.

[19] B. Joshi, B. Cain, J. Johnson, M. Kalitazkis, S. Rahman, M. Xanthidis, A. Hernandez, A. Quattrini Li, N. Vitzilaios, and I. Rekleitis, "Experimental comparison of open source vision-inertial-based state estimation algorithms," *Proc. IROS*, 2019, (under review).

[20] P. Corke, C. Detweiler, M. Dunbabin, M. Hamilton, D. Rus, and I. Vasilescu, "Experiments with underwater robot localization and tracking," in *Proc. ICRA*. IEEE, 2007, pp. 4556–4561.

[21] S. Rahman, A. Quattrini Li, and I. Rekleitis, "Sonar Visual Inertial SLAM of Underwater Structures," in *Proc. ICRA*, 2018.

[22] ——, "Svin2: Sonar visual-inertial SLAM with loop closure for underwater navigation," *CoRR*, vol. abs/1810.03200, 2018. [Online]. Available: http://arxiv.org/abs/1810.03200

[23] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, no. 3.2, 2009, p. 5.

[24] X. Wu, R. E. Stuck, I. Rekleitis, and J. M. Beer, "Towards a Framework for Human Factors in Underwater Robotics," in *Human Factors and Ergonomics Society International Annual Meeting*, 2015, pp. 1115–1119.

[25] S. Rahman, A. Quattrini Li, and I. Rekleitis, "A modular sensor suite for underwater reconstruction," in *MTS/IEEE Oceans Charleston*, 2018, pp. 1–6.

[26] Autonomous Field Robotics Lab, "Stereo Rig Sensor documentation," https://afrl.cse.sc.edu/afrl/resources/StereoRigWiki/.

(a) (b)

Fig. 6. Difference in the lighting conditions in caves due to the turbidity of the water and the color of the walls: (a) Good lighting conditions, Ballroom, Ginnie Springs, FL, USA, white walls, high flow, clear water. (b) Bad lighting conditions, Turner Sink, FL, USA, dark walls, slightly tannic water.

Fig. 7. An Aqua2 AUV operating at the Blue Grotto cavern in Florida.

[27] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual–inertial odometry," *IEEE Trans. Robot.*, vol. 33, no. 1, pp. 1–21, 2017.

[28] S. Agarwal, K. Mierle, and Others, "Ceres Solver," http://ceres-solver.org, 2015.

[29] D. Gálvez-López and J. D. Tardos, "Bags of binary words for fast place recognition in image sequences," *IEEE Trans. Robot.*, vol. 28, no. 5, pp. 1188–1197, 2012.

[30] N. Weidner, S. Rahman, A. Quattrini Li, and I. Rekleitis, "Underwater Cave Mapping using Stereo Vision," in *Proc. ICRA*, Singapore, May 2017, pp. 5709–5715.

[31] S. Rahman, A. Quattrini Li, and I. Rekleitis, "Contour based Reconstruction of Underwater Structures Using Sonar, Visual, Inertial, and Depth Sensor," *Proc. IROS*, 2019, (under review).

[32] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, 2015.

[33] Research group of Prof. Kostas Daniilidis, "Monocular MSCKF ROS node," https://github.com/daniilidis-group/msckf_mono, 2018.

[34] E. Westman and M. Kaess, "Underwater AprilTag SLAM and calibration for high precision robot localization," Robotics Institute, Carnegie Mellon University, Tech. Rep. CMU-RI-TR-18-43, Oct. 2018.