

# Underwater Dense Mapping with the First Compact 3D Sonar

Chinmay Burgul<sup>1\*</sup>, Yewei Huang<sup>2\*</sup>, Michalis Chatzisprou<sup>1</sup>,  
Ioannis Rekleitis<sup>1</sup>, Alberto Quattrini Li<sup>2</sup>, and Marios Xanthidis<sup>3</sup>

**Abstract**—In the past decade, the adoption of compact 3D range sensors, such as LiDARs, has driven the developments of robust state-estimation pipelines, making them a standard sensor for aerial, ground, and space autonomy. Unfortunately, poor propagation of electromagnetic waves underwater, has limited the visibility-independent sensing options of underwater state-estimation to acoustic range sensors, which provide 2D information including, at-best, spatially ambiguous information. This paper, to the best of our knowledge, is the first study examining the performance, capacity, and opportunities arising from the recent introduction of the first compact 3D sonar. Towards that purpose, we introduce calibration procedures for extracting the extrinsics between the 3D sonar and a camera and we provide a study on acoustic response in different surfaces and materials. Moreover, we provide novel mapping and SLAM pipelines tested in deployments in underwater cave systems and other geometrically and acoustically challenging underwater environments. Our assessment showcases the unique capacity of 3D sonars to capture consistent spatial information allowing for detailed reconstructions and localization in datasets expanding to hundreds of meters. At the same time it highlights remaining challenges related to acoustic propagation, as found also in other acoustic sensors. The datasets used in our evaluation will be publicly released to support reproducibility and foster further research by the community: Dataset Repository.

## I. INTRODUCTION

Mapping underwater environments is a time consuming, labor intensive, and often dangerous operation. Overhead and cluttered environments are particularly challenging due to entanglement hazards. Underwater caves (and the inside of wrecks) present many dangers, even for highly skilled people, with more than 350 deaths since 1969 in USA alone [1], [2]. However, Karst aquifers (underwater caves) account for 25% of potable water worldwide [3]. One of the main challenges in mapping underwater environments is the limited visibility. Even in clear waters, after a few meters, the scene becomes blurred and finding features is extremely challenging. Furthermore, limited illumination and color absorption [4], [5] result in most vision-based odometry/SLAM systems to under-perform [6], which in turn affects autonomous pipelines that rely upon them [7].

<sup>1</sup> Chinmay Burgul, Michalis Chatzisprou, and Ioannis Rekleitis are with the Mechanical Engineering Department, University of Delaware, Newark, DE, USA. [cmburgul,michalis,yiannisr]@udel.edu

<sup>2</sup> Yewei Huang and Alberto Quattrini Li are with Dartmouth College, Hanover, NH, USA {yewei.huang, alberto.quattrini.li}@dartmouth.edu

<sup>3</sup> Marios Xanthidis is with the Aquaculture Robotics and Automation Group, SINTEF Ocean, marios.xanthidis@sintef.no

\* Authors contributed equally.

This work was supported by the Research Council of Norway (EchoNav: NO-359447), National Science Foundation (NSF 2545370, 1943205, 2024541, 2144624)

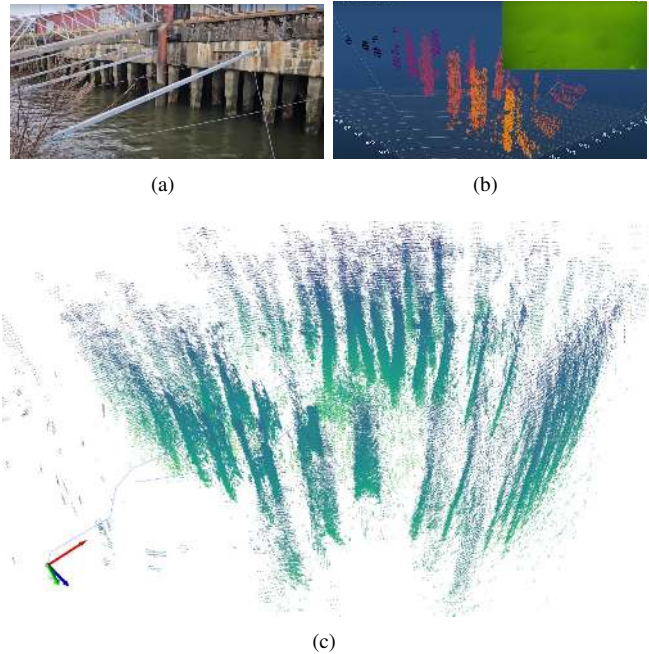


Fig. 1: **3D sonar map reconstruction** in a pier piling environment with limited camera visibility. (a) Pier piling environment, (b) single 3D sonar scan and camera visibility, and (c) reconstructed pointcloud results.

On the other hand, acoustic sensors — despite their popularity — suffer from several inherent limitations. A primary drawback is their high angular uncertainty; for instance, the Ping360 from Blue Robotics [8], a 2D mechanical scanning sonar, has a beam angle of  $25^\circ$ , which introduces substantial directional ambiguity. Reducing this uncertainty requires narrower beams, but this comes at the expense of range. For example, the Imagenex pipe profiling sonar [9], another mechanical scanning sonar, achieves a beam width of only  $1.4^\circ$ , yet its effective range is limited to about 6 meters. 3D sonars may resolve some of these disadvantages, as showcased in preliminary works utilizing probabilistic scan matching [10]. Though, logistic challenges in these devices in terms of power demands, size, and weight hindered broad introduction to underwater platforms, limiting data collection to fly-overs in open-water environments.

In contrast, newly introduced 3D sonars, such as the Sonar 3D-15 from Water Linked [11], are compact, provide a LiDAR-like pointcloud (as shown in Fig. 1), and do not suffer from the aforementioned disadvantages, allowing for data collections in complex cluttered overhead environments. Notably, this sensor can be readily integrated into most popular underwater platforms, providing acoustic measurements offering unique 3D spatial information.

This paper exploits the advantages of the Sonar 3D-15 with deployments spanning several field trials over different environments including a low-visibility dock, a submerged container, a swimming pool for testing the reflectance properties of different materials, a cavern, and a 300 m penetration of a cave system. With the exception of the dock and the container, visual, inertial, and water depth measurements were collected at the same time. This enabled the development of new calibration procedures, the introduction of SLAM frameworks that integrate such sensors, the collection of dedicated datasets, and the characterization of this novel 3D range sensor.

In summary, the contributions of this paper are as follows:

- 1) Characterization of the Sonar 3D-15 sensor's performance across different material types and surface properties.
- 2) A novel calibration strategy for estimating the extrinsic parameters between the camera and sonar.
- 3) A loosely coupled integration of sonar and VIO into a common estimation process, showing their complementary sensing capabilities.
- 4) Introduction of several datasets, including sonar, visual, inertial, and water depth data, which would be made publicly available upon acceptance.

## II. RELATED WORK

Simultaneous Localization and mapping (SLAM) are key to enabling autonomy in underwater environments, providing the foundation for reliable control and planning. A wide range of sensor configurations have been explored, each tailored to the specific challenges of different environments and tasks. Johannsson et al. [12] utilized imaging sonar and Doppler Velocity Logs (DVL) for harbor surveillance. The Sunfish team [13], [14] integrated a multibeam sonar, fiber-optic gyroscope (FOG), Inertial Measurement Unit (IMU), DVL and pressure sensors for cave exploration and mapping. SVIn2 [15] combined visual-inertial sensing with a pipe profiling sonar, and pressure sensor data, extending the framework of OKVIS [16], to map underwater caverns and shipwrecks; this approach was later extended to incorporate magnetometer data in [17]. Turtlmap [18] relied on DVL, inertial, and pressure sensors for localization in low-texture environments and a depth camera for mapping, while [19] utilized side-scan sonar (SSS) with dead-reckoning to address localization in open-water settings.

While proprioceptive sensors such as IMUs, DVLs, and magnetometers provide valuable localization cues, they lack environmental awareness for mapping. Vision-based approaches are constrained to short operating ranges due to light attenuation and turbidity in underwater environments, whereas traditional sonar-based methods, although resilient to visibility conditions, introduce their own challenges [20]. Low-cost scanning sonars, such as Ping360 and pipe-profiling sonars, exhibit angular uncertainty and low-dimensional sensing. Imaging sonar as well as side-scan sonars generate raw acoustic imagery that is subject to geometric distortion and uneven ensonification, causing the

same region to appear significantly different when viewed from varying perspectives. The Sonar 3D-15 extends these capabilities with a range of up to 15 meters. Unlike acoustic imaging sensors, its pointcloud is free from geometric distortion or vertical uncertainty, although it often displays uneven ensonification on smooth surfaces. In this work, we evaluate a sonar-only approach in low visibility conditions and additionally explore fusion with Visual-Inertial sensors in environments with satisfactory visibility.

Calibrating the Sonar 3D-15 extrinsic parameters with a visual sensor appears conceptually similar to LiDAR-camera calibration. Existing approaches are broadly categorized into *target-based* and *targetless* methods. In camera-pointcloud target-based calibration, past work [21] reviews the use of variety of targets, including 2D ArUco markers, checkerboards, and 3D objects (e.g., boxes, spheres, or other structured geometries). Targetless methods exploit scene geometry such as structural regularities [22] or line-plane correspondences [23]. However, conventional flat or smooth calibration target surfaces are not reliably perceived by the tested 3D sonar, and target-less methods, relying on line-plane or geometric structure, are generally absent in natural underwater environments. Moreover, sonar sensors suffer from acoustic reverberation and acoustic multipath inference in confined environments (e.g., swimming pools), where hard, flat pool walls reflect acoustic pulses creating delayed echoes that return to the sensor from different angles at different times. This interference overwhelms the sensor's ability to distinguish a desired target due to the presence of such noise. Similar practical challenges arise in using IMU-LiDAR calibration [24], which additionally requires specific excitation patterns that are difficult to perform underwater. This motivated us to develop our own 3D sonar-camera calibration approach, which is presented in Section III-B.

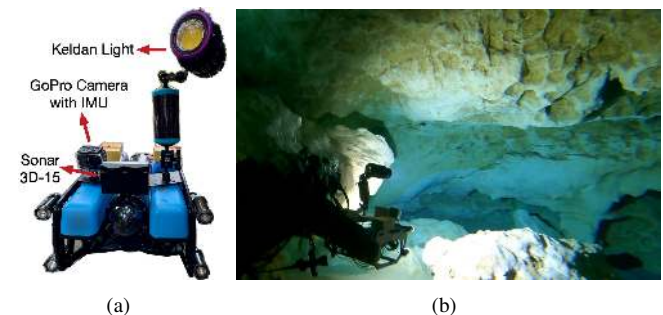


Fig. 2: Our customized BlueROV2 setup (a) and the experiment environment (b).

## III. SONAR CHARACTERIZATION AND CALIBRATION

In this paper, the following experimental setup was used. The primary sensor is a reproduction unit of Sonar 3D-15 from Waterlinked [11]. The Sonar 3D-15 is an advanced 3D sonar with approximately a  $90^\circ$  by  $40^\circ$  field of view in navigation mode, the only mode tested and available at the time of this reporting. The navigation mode provides a range of up to 15 meters with an update frequency of 5 Hz. The

sensor was rigidly mounted on a BlueRov2 platform [25]. The BlueRov2 is equipped with a Raspberry PI camera, an IMU and a pressure sensor measuring the current depth, thus providing globally consistent measurements along the Z-axis. Furthermore, a GoPro Hero Black (10 and 13) with an internal IMU is mounted next to the sonar, facing forward. Illumination is provided by four 1500-lumen Subsea Light [26] and a 20000-lumen Keldan light [27]. Fig. 2 shows the experimental setup deployed inside a cavern in Florida.

#### A. Sonar Characterization and Material Response

During the experiments the sonar operated in low (1.2 MHz) frequency, producing a beam of  $0.6^\circ$  horizontal and  $2.4^\circ$  vertical angle. The returned range measurements have 4 mm resolution. Physically, the Sonar 3D-15 measures 0.08 m height by 0.12 m width by 0.04 m depth and weights 390 g submerged.

To select a suitable material for sonar-camera calibration and assess the performance of the 3D sonar with different surfaces, experiments were conducted to evaluate 3D sonar responses across different materials. Selected materials included those frequently encountered in underwater environments, such as bricks, concrete, PVC pipes, and wood. Additionally, easily available items typically used for calibration were included, such as a glass sphere, popular for 2D sonar calibration, and a metal checkerboard with 3D-printed tag for camera calibration. The camera image and corresponding sonar point-cloud of the test materials are shown in Fig. 3. The experiment was conducted in an indoor swimming pool with smooth walls, which introduced multipath effects in the collected data. The test materials were suspended in the middle of the water to minimize the influence of the pool walls. To further mitigate their impact on the material experiment results, the sensor range was limited to  $r = 2\text{m}$  for visualization in Fig. 3.

The results demonstrated that the responses varied with both material type and surface properties at different viewing angles. While the glass sphere and steel calibration board provide strong reflections for the 2D sonar, they are not well-suited for 3D sonar calibration due to multipath effects. Regarding surface properties, smooth objects such as PVC pipe produced noisier detections due to reflections, while rougher materials such as concrete yielded less noisy measurements. Both spray painted and unpainted concrete produced relatively strong signals.

#### B. Calibrating the extrinsic parameters

Based on the material experiments, cinder (concrete) blocks (Fig. 4(a)) were selected due to their rough surface texture, well-defined geometric structure, and convenient availability. Four cinder blocks were placed at the bottom of the water body in a particular arrangement, such that the 16 corners of top surfaces form a distinctive pattern. Non-top surfaces were spray painted to suppress spurious edges and aid corner detection in the camera images. A rectangular mask is used to identify the corners of the four cinder blocks, which are then labeled correspondingly in the visual images.

The top surface of the cinder blocks are extracted from sonar point-clouds by identifying the elevated planar regions. As shown in Fig. 4(b), points on these elevated surfaces were subsequently denoised and clustered into four separate 2D surfaces.

After detecting the corner points of the cinder blocks in both the visual images and sonar pointcloud, correspondences between them are established to form a 3D-2D association. Define a sonar 3D point as  $\mathbf{X} = [x, y, z]^\top$ , and a visual point as  $\mathbf{p} = [u, v]^\top$ . The projection model, with  $s$  a scalar scale factor, is given by:

$$s \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix} = \mathbf{K} \cdot [\mathbf{R}|\mathbf{t}] \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} \quad (1)$$

The intrinsic matrix  $\mathbf{K}$  is obtained through standard camera calibration. The undistorted visual point  $\mathbf{p}'$  is defined as  $[\mathbf{p}', 1]^\top = \mathbf{K}^{-1}[\mathbf{p}, 1]^\top$ . The camera-to-sonar extrinsic  $(\mathbf{R}, \mathbf{t})$  calibration problem can then be formulated as a Perspective-n-Point (PnP) problem using 16 brick corner correspondences. We solve this problem using EPnP [28] algorithm. Fig. 5 illustrates the alignment of camera rays passing through 16 brick corners in the 2D image plane and their corresponding 3D points, along with the estimated camera-sonar extrinsic.

## IV. METHOD DESCRIPTION

#### A. 3D Sonar Scan Registration

Since the pointcloud collected by a compact 3D sonar shares many characteristics with solid-state 3D LiDAR, 3D sonar scan registration is performed using the state-of-the-art LiDAR scan registration algorithm, KISS-ICP [29]. KISS-ICP is a point-to-point ICP algorithm designed to provide a general solution across different LiDAR platforms and application environments. In this paper, we extend the use of KISS-ICP from various LiDAR configurations to 3D sonar (termed Sonar-odom) and evaluate its performance in a series of underwater scenarios. Since the sonar pointcloud is much sparser than the LiDAR pointcloud, we skipped the down-sampling of the source pointcloud and only perform down-sampling on the target pointcloud.

KISS-ICP adopts a constant velocity model for generality. At a current timestamp  $n$ , the initial guess  $\mathbf{T}_{\text{init},n}$  transformation matrix for ICP scan registration is calculated based on the pose estimates  $\mathbf{T}_{n-1}$ ,  $\mathbf{T}_{n-2}$  from previous timestamps:

$$\mathbf{T}_{\text{init},n} = \mathbf{T}_{n-1} \cdot (\mathbf{T}_{n-2}^{-1} \cdot \mathbf{T}_{n-1}) \quad (2)$$

However, ICP is sensitive to the selection of initial guess [30]. Compared to regular 3D LiDARs, which have a wider field of view and higher frequency, the Sonar 3D-15 has a maximum range of only 15 meters with 5 Hz update rate. Furthermore, the complex hydrodynamic conditions in real underwater environments introduce additional uncertainties in the scan registration of the 3D sonar data. As the robot platform utilized is equipped with both a 3D sonar and a GoPro camera (Fig. 2, left), we provided the 3D sonar a more robust initial prior using visual-inertial odometry (VIO) from SVIn [15], which we call Sonar-odom-SVIn-Initial. The VIO

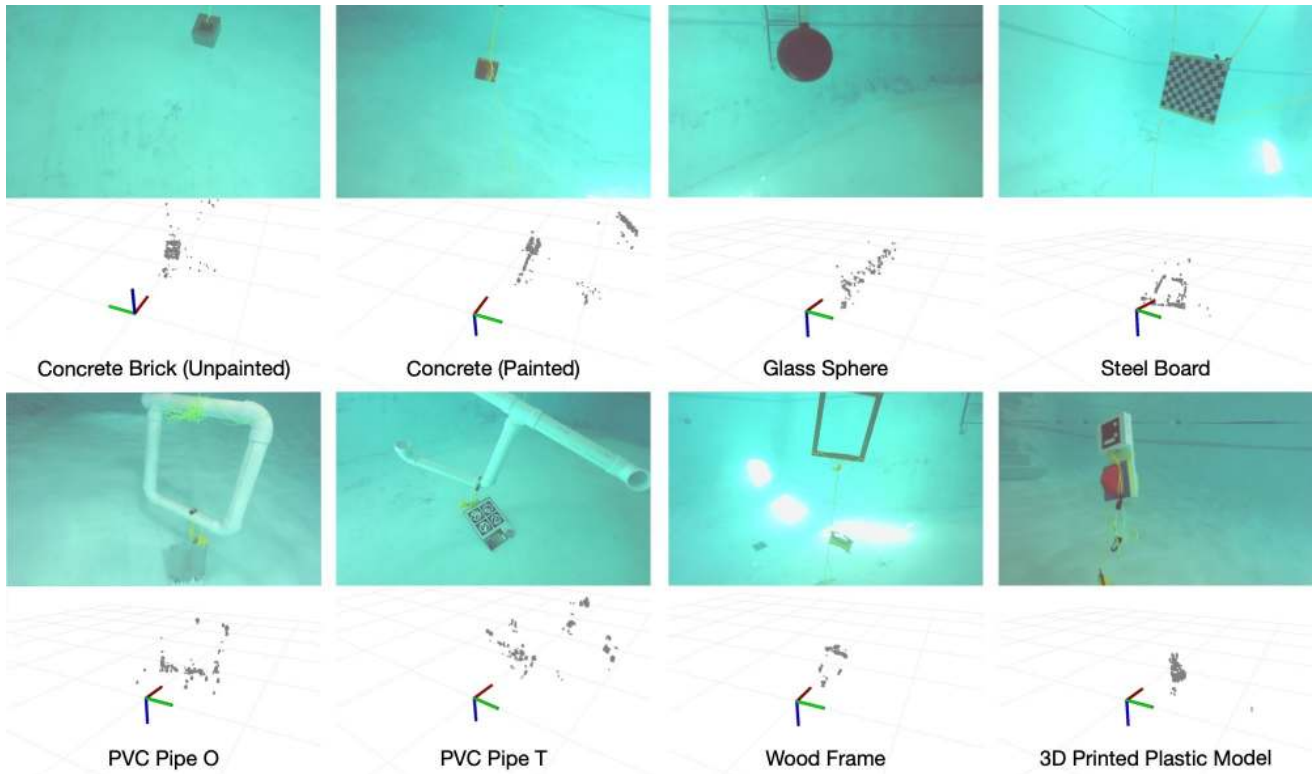


Fig. 3: The sonar response on different materials commonly found in underwater environments and used for calibration.

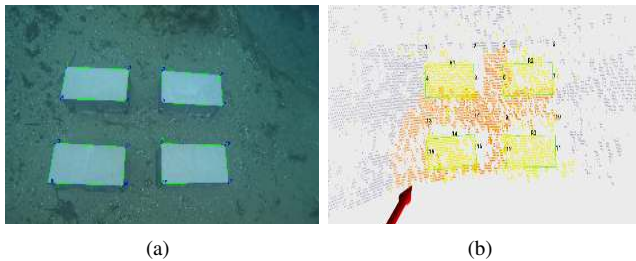


Fig. 4: Calibration setup. (a) Camera image of the four cinder blocks showing the top-surface borders marked in green and detected corners in blue. (b) Pointcloud from the Sonar 3D-15 displaying detected cinder blocks in yellow with borders highlighted in green.

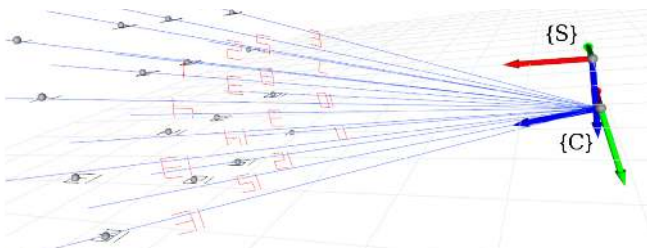


Fig. 5: Rays from the camera projected through 16 cinder block corners in the 2D image plane to their corresponding 3D points. Sonar 3D corner points are annotated in black and visual points in red. The camera frame  $\{C\}$  and sonar frame  $\{S\}$  are also illustrated in the figure using the calibrated extrinsic transformation.

estimation in SVIn is performed at a higher frame rate (30 Hz camera rate). Therefore, we adopt the SVIn odometry from the closest timestamp,  $\mathbf{T}_{\text{VIO},n'}$  (with  $n' \approx n$ ), as the initial guess for 3D sonar ICP scan registration. The odometry pose is transformed into the sonar frame,  $s$ , using the extrinsic

parameters  $\Delta\mathbf{T}$  obtained in Section III:

$${}_s\mathbf{T}_{\text{VIO},n'} = \Delta\mathbf{T} \cdot \mathbf{T}_{\text{VIO},n'} \cdot \Delta\mathbf{T}^{-1} \quad (3)$$

The initial guess for the 3D sonar scan registration is then defined as:

$$\mathbf{T}_{\text{init},n} = \mathbf{T}_{n-1} \cdot ({}_s\mathbf{T}_{\text{VIO},n'}^{-1} \cdot {}_s\mathbf{T}_{\text{VIO},n'}) \quad (4)$$

where  $n' \approx n-1$  denotes the VIO timestamp closest to  $n-1$ .

### B. 3D Sonar Loop-Closure and Pose Graph Optimization

Due to the relative limited field of view and low update frequency of the sonar, the odometry-only trajectory shows significant drift, especially over longer durations (Table I). An additional strategy to mitigate drift is loop-closure detection and pose graph optimization, which we call *Sonar PGO*. Similar to LiDAR SLAM [31], we detect loop-closures using a Euclidean distance-based method. Keyframes, denoted as  $\mathcal{F} = \{\mathbf{F}_i\}$ , are selected to ensure a stable and efficient performance. If the current source keyframe  $\mathbf{F}_s$  is geometrically close enough to, and temporally sufficiently separated from a target keyframe,  $\mathbf{F}_t$ , a potential loop-closure is detected. We then perform scan-to-map matching using the target sub-map,  $\mathcal{M}_t = \{\mathbf{F}_{t-w}, \dots, \mathbf{F}_{t+w}\}$ , with a window size of  $w = 10$ .

In pose graph optimization, the sonar pose at each keyframe  $\mathbf{F}_i$  is treated as a state  $\mathbf{X}_i$ . Sonar odometry from scan registration is modeled as a sequential odometry measurement  $\mathbf{Z}_{i-1,i}^o$ , while the scan-to-map loop-closure is modeled as a loop-closure measurement  $\mathbf{Z}_{i,j}^l$ . These measurements are added to the graph as factors  $f(\mathbf{X}_{i-1}, \mathbf{X}_i; \mathbf{Z}_{i-1,i}^o)$

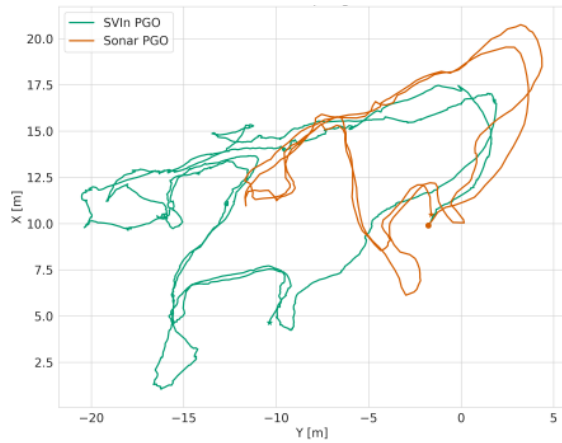


Fig. 6: Trajectories comparison of sonar-only and SVIn on the Ginnie ballroom I dataset.

and  $f(\mathbf{X}_i, \mathbf{X}_j; \mathbf{Z}_{i,j}^l)$ . The overall factor graph at a given time  $n$ , with  $m$  detected loop-closures, can be expressed as:

$$f(\mathcal{X}_n; \mathcal{Z}_n) = f(\mathbf{X}_0) \prod_{i=1}^n f(\mathbf{X}_{i-1}, \mathbf{X}_i; \mathbf{Z}_{i-1,i}^o) \prod_{j=1}^m f(\mathbf{X}_i, \mathbf{X}_j; \mathbf{Z}_{i,j}^l) \quad (5)$$

with  $\mathcal{X}_n = \{\mathbf{X}_i\}$  denoting the set of all states from time 0 to  $n$ , and  $\mathcal{Z}_n = \mathcal{Z}_n^o \cup \mathcal{Z}_n^l$  denoting the set containing both odometry measurements  $\mathcal{Z}_n^o = \{\mathbf{Z}_{i-1,i}^o\}$  and loop-closure measurements  $\mathcal{Z}_n^l = \{\mathbf{Z}_{i,j}^l\}$ .

Note we refer to the use of SVIn as an initial guess for sonar registration and loop-closure as the **Fusion approach**. While SVIn originally denotes Sonar Visual-Inertial Navigation, in this work we utilize only its visual-inertial odometry (VIO) component [16], since the sonar sensor used in SVIn2 [15] was a pipe profiling sonar, which differs from our setup.

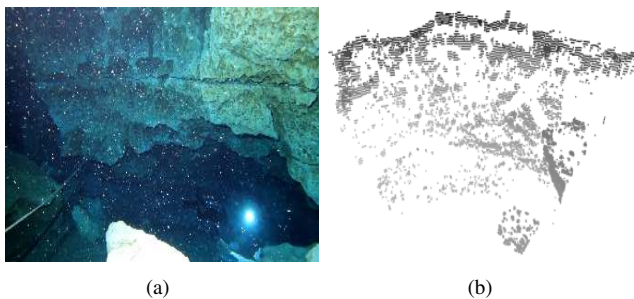


Fig. 7: GoPro camera image (a) and sonar pointcloud (b) captured at nearly the same timestamp. The camera image is filled with floating particulates, whereas the sonar pointcloud is unaffected by them.

## V. EXPERIMENTAL VALIDATION

We evaluate the proposed methods on four datasets with distinct environmental characteristics:

- *Pier Pilling*: A shallow-water dataset containing wood piles, collected using a teleoperated BlueROV2.
- *Container*: A dataset of submerged metal tank with an opening.
- *Peacock Springs Cave*: A long cave trajectory collected using a BlueROV2 equipped with a mounted GoPro13

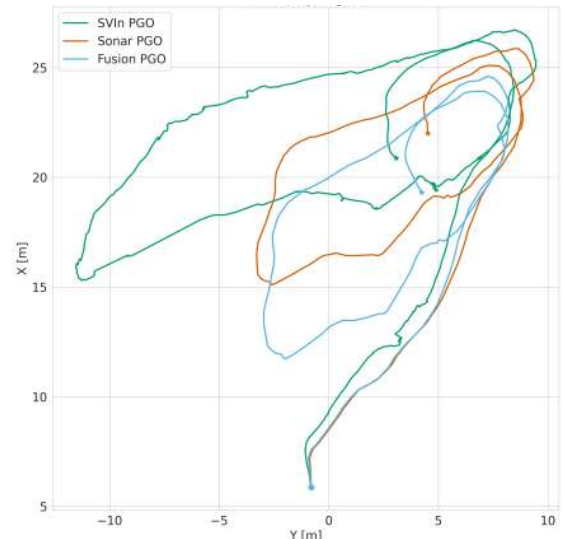


Fig. 8: Trajectories comparison of our Fusion approach, SVIn and Sonar-only on the Ginnie ballroom II dataset.

and carried by cave divers. The divers entered and exited through the same area, resulting in a trajectory with a start-end loop.

- *Ginnie Springs cavern*: A cavern dataset with a larger opening, varying depth, and several narrower passages. The BlueROV2, equipped with a 3D sonar and a Go-Pro10, was carried by divers along looping trajectories. The dataset was collected over two days: Day I with a powerful Keldan light, and Day II with less powerful lights mounted on the BlueROV2.

Since the datasets were recorded using multiple devices with unsynchronized clocks, temporal alignments was performed in post-processing by synchronizing the individual data-streams using raw gyroscope measurements.

In this section, we evaluate the performance of Sonar SLAM (referred to as Sonar PGO) and our visual-sonar fusion approach (referred to as Fusion PGO), comparing them to the state-of-art underwater VIO, SVIn2 (referred to as SVIn PGO), on datasets collected in the Ginnie Springs cave system.

We select three representative datasets for comparison: the Ginnie Springs cave system, the Peacock Springs cave system, and the container dataset. The Ginnie Springs dataset contains floating particulates in the images, bringing challenges for the state-of-art underwater VIO framework. The Peacock Springs dataset is a long-duration dataset that demonstrates the ability of our system to handle extended localization and mapping missions. The container dataset represents a challenging environment that includes a steel container. As discussed in Sec. V, the steel would cause multi-path effect for the 3D sonar, making localization and mapping using sonar pointcloud even more challenging.

Fig. 6 presents the comparison between Sonar PGO and SVIn PGO on a start-end loop trajectory, referred to here as Ginnie ballroom I dataset. All trajectories are aligned using the keyframe poses from first five timestamps of the

Sonar PGO with EVO<sup>1</sup>, and the starting and ending points are marked by a circle and a star, respectively. Due to the presence of floating particulates (Fig. 7), which are detected and misinterpreted as static features, SVIn exhibits a larger start–end re-visitation drift (10.34 m) compared to Sonar PGO (3.31 m). Though raw sonar scan matching suffers from the drift caused by the low sensor update frequency and the constant-velocity motion model used by Kiss-ICP, the longer sonar range (15 meters) ensures that enough loop closures are detected, thereby constraining the overall drift and achieving a better start–end re-visitation drift error.

We evaluate the performance of our fusion approach on the datasets collected in Ginnie Ballroom II, and compared it against two configurations, Sonar-only SLAM and SVIn. Fig. 8 shows the trajectories using various configurations. Similarly, all trajectories are aligned using the first five timestamps of Sonar PGO. Incorporating the SVIn prior into Sonar-PGO reduces overall drift compared to both methods on the Ginnie Ballroom II dataset. The vision-based method drifts when there are floating particulates, whereas the sonar remains unaffected by such occlusions and provides reliable longer range measurements. However, sonar-only approaches suffer from registration errors due to low measurement frequency. This drawback is mitigated by the higher-frequency odometry from SVIn, which supplies robust motion priors for scan alignment. By combining both camera and sonar data, the system exhibits lower drift compared to trajectories using only the camera or the sonar.

#### A. Dense Pointcloud Reconstruction

1) *Reconstruction of the Cave System:* We also demonstrate the dense reconstruction results of our fusion approach in the challenging environments of the Ginnie Springs cavern (Fig. 9) and the Peacock Springs (Fig. 10 and 11).

The Sonar 3D-15 enables consistent dense reconstruction, as shown in a zoomed-in view of the interior from the Ginnie Ballroom. As mentioned previously, this dataset contains floating particulates that interfere with the performance of SVIn. However, in the sonar reconstruction results, some basic topographic features of the cave environment are perceptible, demonstrating the effectiveness of our fusion and loop closure detection strategy. In Fig. 9, the limestone structure and a nearby narrow passage can be roughly identified in the dense point cloud, corresponding well to the collected image.

The Peacock Spring dataset presents confined, tunnel-like settings, posing significant difficulties for conventional 3D sensing modalities. Although SVIn achieves a relatively low drift rate, as shown in Tab. I, failures in loop closure due to sparse visual features and viewpoint differences, as illustrated in Fig. 11(a), prevent successful loop closure. In contrast, the fusion approach is largely unaffected by these limitations. As shown in Fig. 10, it effectively reconstructs even the narrowest part of the dense structural layout of the confined cave environment. We generate the

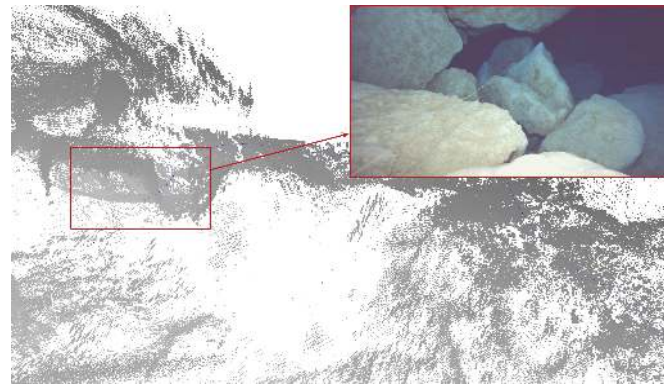


Fig. 9: Detailed pointcloud reconstruction of the cave interior from the Ginnie Ballroom I dataset, using a limestone structure as a reference.

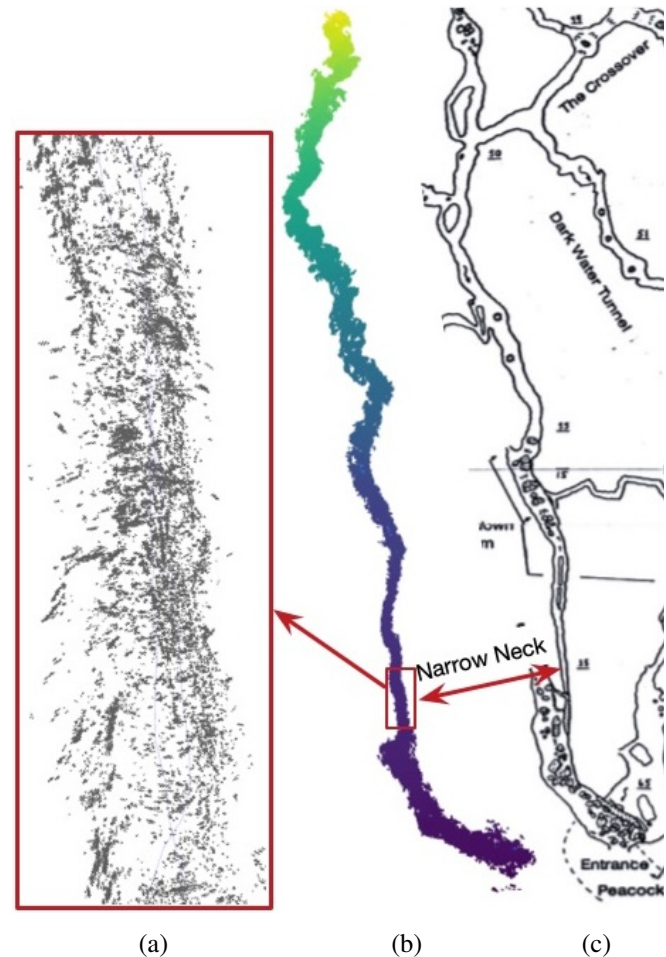


Fig. 10: Results of Fusion approach on a 700 m trajectory inside Peacock Springs cave system: (a) detailed pointcloud of the narrow neck, (b) pointcloud map of the explored cave system, and (c) the corresponding Cave Atlas map.

dense pointcloud from sonar pointclouds associated with the keyframes in the pose graph optimization. Compared to the existing Cave Atlas maps, which are manually surveyed and contain only the topographic relationships of the caves, the reconstructed pointcloud map provides finer structural detail, improved scale accuracy, and a full 3D representation.

The re-visitation error for the start–end loop of around a 700-meter trajectory collected in one of the Peacock Springs cave systems is summarized in Tab. I. The error is calculated

<sup>1</sup>EVO: <https://github.com/MichaelGrupp/evo>

TABLE I: **Re-visitation Error [m]** between the start and end points of the Peacock Springs cave system trajectory under different methods and configurations.

SVIn PGO	Sonar Odom	Sonar PGO	Fusion Odom	Fusion PGO
17.02	250.34	21.51	238.71	6.76

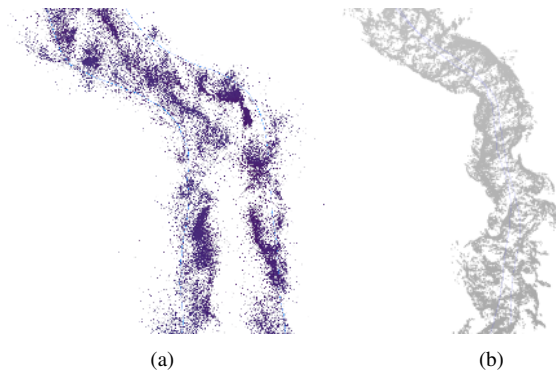


Fig. 11: Comparison of detailed pointcloud reconstructions of the same area in Peacock Caves: (a) SVIn PGO and (b) Fusion PGO approach.

as the Euclidean distance between the starting and ending points of the trajectory. The cave entrance has a diameter of approximately 10 meters; therefore, even if the starting and ending points do not coincide, the re-visitation error should be no greater than 10 meters. Our fusion approach achieves a re-visitation error of 6.76 meters, which is the closest to this theoretical bound.

### 2) Operation in Acoustically Challenging Environments:

To challenge the capacity of the sensor on proving valid range responses that could be utilized for reconstruction in confined spaces, a dataset was collected from a submerged steel cylindrical container, as shown in Fig. 12 in poor visibility (Fig. 12(a)). In agreement to previous observations, it is possible that certain sonar scans may be incomplete with some surfaces not returning a response such as the bottom of the door (Fig. 12(b)), though details are covered with relative consistency such as the triangle features at the corners. In the full reconstruction (Fig. 12(c)), the overall shape of the structure is captured well forming very accurately the cylindrical shape of the container, though several outliers are visible.

For instance, due to multipath issues when entering this confined space, inaccurate responses formed a weak, although noticeable outliers on the sides of the wall, at a distance similar to the diameter of the cylinder. Moreover, a large interference pattern emerged in the form of two homocentric rings, on the other open-ended side of the cylindrical structure at a similar distance as its length, indicating further multipath inconsistencies.

## VI. LESSONS LEARNED, FUTURE WORK, AND CONCLUSIONS

One of the key observations was the effect of the field of view on the trajectory estimation. While the wider horizontal field of view ( $90^\circ$ ) allowed tracking of features throughout the experiments, the restricted vertical field of view ( $40^\circ$ ) during operations in the narrow tunnels of the Peacock cave

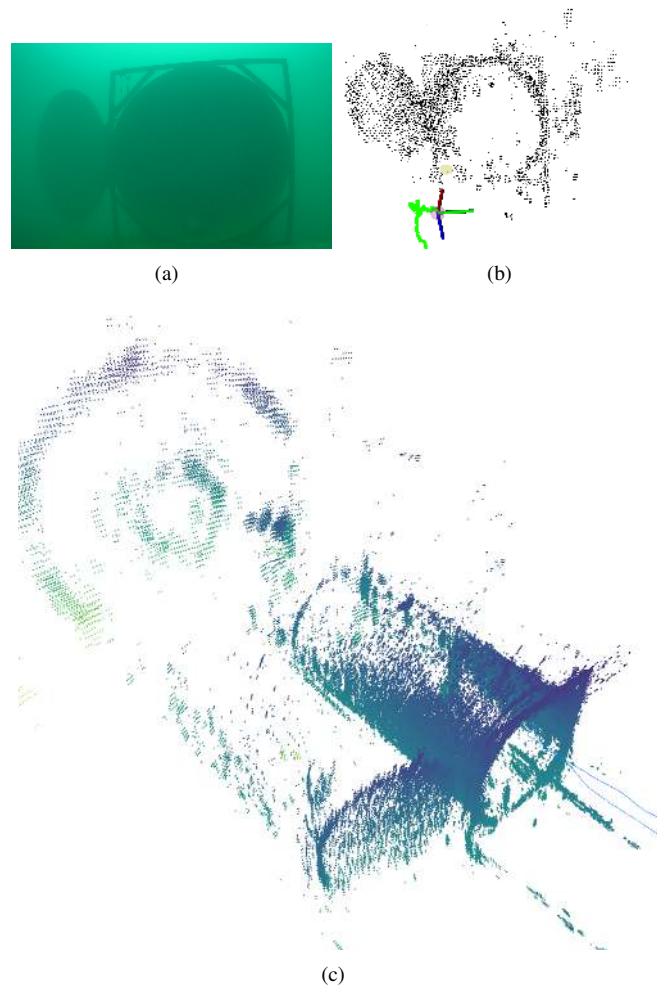


Fig. 12: Reconstruction outliers in acoustically challenging environments. The visibility conditions are shown in (a), a sonar scan registration in (b), with the the resulting reconstruction in (c). Outliers may be present due to multi-path issues.

system, FL, resulted in systematic drift over time. More specifically, as only a small patch of the floor and ceiling was visible for large segments of the trajectory, ICP resulted in incorrect matches and incremental drift on the Z-axis; loop-closures on the way back brought the trajectory back to the starting point. The final goal of this work will be to integrate all the available sources of information, acoustic, visual, inertial, water depth, and possibly magnetic field, into a tightly coupled optimization framework. Contrary to SVIn2 [15] where the pipe profiling sonar and the camera did not share a field of view, the Sonar 3D-15 shares the field of view of the camera, simplifying the correlation of features in the observed environment.

The integration of the range measurements into a coherent volumetric map is also part of the future work. Such a development will assist in multi-session mapping where an environment is explored and mapped over several deployments. Commercial work by Sunfish Inc. [14] has generated an example of visualization. Exploring different representations, such a voxels [32], Surflets [33], or learning based Gaussian Splatting like representations in terms of

accuracy and ease of integration into the state estimation process will open several directions of future research.

Finally, in challenging environments, as shown in our experiments, typical acoustic issues remain, such as multipath effects. Thus future research may focus on online corrections through probabilistic map representations.

Overall, this paper characterized a new 3D sonar in different challenging underwater environments for localization and dense reconstruction, providing the foundation for robust underwater mapping and exploration. By presenting a novel calibration strategy and SLAM pipelines and releasing our datasets, we aim to accelerate research into the unique challenges and opportunities of 3D acoustic sensing for enhanced underwater robot autonomy.

## VII. ACKNOWLEDGMENTS

The authors would like to thank Water Linked AS for their unrestricted support, including access to their hardware and datasets during testing. Furthermore, we acknowledge the help from Woodville Karst Plain Project (WKPP), El Centro Investigador del Sistema Acuifero de Quintana Roo A.C. (CINDAQ), Global Underwater Explorers (GUE), Ricardo Constantino, and Project Baseline over the years in providing access to challenging underwater caves and mentoring us in underwater cave mapping. The authors are also grateful for equipment support by Halcyon Dive Systems, and KELDAN GmbH lights.

## REFERENCES

- [1] P. Buzzacott, C. Nelson, K. Hill, and L. Hires, "Recovery of deceased scuba divers from within flooded subterranean caves," *Forensic science international*, vol. 275, pp. 167–170, 2017.
- [2] J. M. Casadesús, F. Aguirre, A. Carrera, P. Boadas-Vaello, M. T. Serrando, and F. Reina, "Diving-related fatalities: multidisciplinary, experience-based investigation," *Forensic Science, Medicine and Pathology*, vol. 15, pp. 224–232, 2019.
- [3] D. Ford and P. Williams, *Introduction to Karst*. John Wiley & Sons, Ltd, 2007, ch. 1, pp. 1–8.
- [4] M. Roznere and A. Quattrini Li, "Real-time model-based image color correction for underwater robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019. [Online]. Available: <https://arxiv.org/abs/1904.06437>
- [5] S. Skaff, J. Clark, and I. Rekleitis, "Estimating surface reflectance spectra for underwater color vision," in *British Machine Vision Conference (BMVC)*, Leeds, U.K., Sep. 2008, pp. 1015–1024.
- [6] B. Joshi, S. Rahman, M. Kalaitzakis, B. Cain, J. Johnson, M. Xanthidis, N. Karapetyan, A. Hernandez, A. Quattrini Li, N. Vitzilaios, and I. Rekleitis, "Experimental Comparison of Open Source Visual-Inertial-Based State Estimation Algorithms in the Underwater Domain," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, 2019, pp. 7221–7227.
- [7] M. Xanthidis, N. Karapetyan, H. Damron, S. Rahman, J. Johnson, A. O'Connell, J. M. O'Kane, and I. Rekleitis, "Navigation in the presence of obstacles for an agile autonomous underwater vehicle," in *International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 892–899.
- [8] Blue Robotics, "Ping360," <https://bluerobotics.com/store/sonars/imaging-sonars/ping360-sonar-r1-rp/>, 2025, accessed: 2025-09-12.
- [9] Imagenex Technology Corp., "831L Pipe Profiling," <https://imagenex.com/products/831l-pipe-profiling>, 2025, accessed: 2025-09-12.
- [10] A. Ferreira, J. Almeida, A. Martins, A. Matos, and E. Silva, "3duplic: An underwater scan matching method for three-dimensional sonar registration," *Sensors*, vol. 22, no. 10, p. 3631, 2022.
- [11] Water Linked, "Sonar 3D-15," <https://waterlinked.com/shop/wl-21045-2-sonar-3d-15-689/>, 2025, accessed: 2025-09-12.
- [12] H. Johannsson, M. Kaess, B. Englot, F. Hover, and J. Leonard, "Imaging sonar-aided navigation for autonomous underwater harbor surveillance," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 4396–4403.
- [13] K. Richmond, C. Flesher, L. Lindzey, N. Tanner, and W. C. Stone, "Sunfish@: A human-portable exploration auv for complex 3d environments," in *OCEANS 2018 MTS/IEEE Charleston*, 2018, pp. 1–9.
- [14] SUNFISH, Inc., "AUV," <https://sunfishinc.com/>, 2025, accessed: 2025-09-12.
- [15] S. Rahman, A. Quattrini Li, and I. Rekleitis, "SVIn2: A Multi-sensor Fusion-based Underwater SLAM System," *International Journal of Robotics Research*, vol. 41, no. 11–12, pp. 1022–1042, July 2022.
- [16] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.
- [17] B. Joshi and I. Rekleitis, "Enhancing visual inertial slam with magnetic measurements," in *IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan, May 2024, pp. 10012–10019.
- [18] J. Song, O. Bagoren, R. Andigani, A. Sethuraman, and K. A. Skinner, "Turtlmap: Real-time localization and dense mapping of low-texture underwater environments with a low-cost unmanned underwater vehicle," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 1191–1198.
- [19] J. Zhang, Y. Xie, L. Ling, and J. Folkesson, "A fully-automatic side-scan sonar simultaneous localization and mapping framework," *IET Radar, Sonar & Navigation*, vol. 18, no. 5, pp. 674–683, 2024.
- [20] J. McConnell, I. Collado-Gonzalez, and B. Englot, "Perception for underwater robots," *Current Robotics Reports*, vol. 3, no. 4, pp. 177–186, 2022.
- [21] X. Zhang, Y. Xiong, Q. Qu, R. Wang, X. Gao, J. Liu, S. Guo, and J. Li, "Cooperative visual-lidar extrinsic calibration technology for intersection vehicle-infrastructure: A review," *arXiv preprint arXiv:2405.10132*, 2024.
- [22] B. Nagy, L. Kovács, and C. Benedek, "Online targetless end-to-end camera-lidar self-calibration," in *2019 16th International Conference on Machine Vision Applications (MVA)*, 2019, pp. 1–6.
- [23] L. Zhou, Z. Li, and M. Kaess, "Automatic extrinsic calibration of a camera and a 3d lidar using line and plane correspondences," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 5562–5569.
- [24] S. Chen, X. Li, S. Li, Y. Zhou, and X. Yang, "ikalibr: Unified targetless spatiotemporal calibration for resilient integrated inertial systems," *IEEE Transactions on Robotics*, 2025.
- [25] Blue Robotics, Inc., "BlueROV2," <https://bluerobotics.com/store/rov/bluerov2/>, 2022, accessed: 2022-12-23.
- [26] BlueRobotics, "Lumen Subsea Light for ROV/AUV," <https://bluerobotics.com/store/thrusters/lights/lumen-r2-rp/>, 2025, accessed: 2025-09-12.
- [27] KELDAN GmbH, "8XR 20000 lm CRI 86," <https://keldanlights.com/products/compact-lights/1725-video-8xr-20000lm-cri86.html>, 2025, accessed: 2025-09-12.
- [28] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Ep n p: An accurate o (n) solution to the p n p problem," *International journal of computer vision*, vol. 81, no. 2, pp. 155–166, 2009.
- [29] I. Vizzo, T. Guadagnino, B. Mersch, L. Wiesmann, J. Behley, and C. Stachniss, "Kiss-icp: In defense of point-to-point icp—simple, accurate, and robust registration if done the right way," *IEEE Robotics and Automation Letters*, vol. 8, no. 2, pp. 1029–1036, 2023.
- [30] S. Rusinkiewicz and M. Levoy, "Efficient variants of the icp algorithm," in *Proceedings third international conference on 3-D digital imaging and modeling*. IEEE, 2001, pp. 145–152.
- [31] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping," in *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2020, pp. 5135–5142.
- [32] V. Siegel, W. Stone, and K. Richmond, "Robotic survey and 3-d mapping of underwater caves using a sunfish@ autonomous underwater vehicle," in *4th International Planetary Caves Conference*, vol. 2697, 2023, p. 1037.
- [33] V. Chandrasekaran, M. Wakin, D. Baron, and R. Baraniuk, "Surflets: a sparse representation for multidimensional functions containing smooth discontinuities," in *International Symposium on Information Theory, 2004. ISIT 2004. Proceedings.*, 2004, pp. 563–.