Ego-to-Exo: Interfacing Third Person Visuals from Egocentric Views in Real-time for Improved ROV Teleoperation

Adnan Abdullah[1], Ruo Chen[1], Ioannis Rekleitis[2], and Md Jahidul Islam[1]

[1] RoboPI Laboratory, Dept. of ECE, University of Florida, FL 32611, USA.
{adnanabdullah@,chenruo@,jahid@ece.}ufl.edu,
[2] AFRL, Dept. of CSE, University of South Carolina, SC 29208, USA.
yiannisr@cse.sc.edu.

Abstract. Underwater ROVs (Remotely Operated Vehicles) are unmanned submersibles designed for exploring and operating in the depths of the ocean. Despite using high-end cameras, typical teleoperation engines based on first-person (egocentric) views limit a surface operator's ability to maneuver the ROV in complex deep-water missions. In this paper, we present an interactive teleoperation interface that enhances the operational capabilities via increased situational awareness. This is accomplished by (**i**) offering on-demand third-person (exocentric) visuals from past egocentric views, and (**ii**) facilitating enhanced peripheral information with augmented ROV pose in real-time. We achieve this by integrating a 3D geometry-based Ego-to-Exo view synthesis algorithm into a monocular SLAM system for accurate trajectory estimation. The proposed closed-form solution only uses past egocentric views from the ROV and a SLAM backbone for pose estimation, which makes it portable to existing ROV platforms. Unlike data-driven solutions, it is invariant to applications and waterbody-specific scenes. We validate the geometric accuracy of the proposed framework through extensive experiments of 2-DOF indoor navigation and 6-DOF underwater cave exploration in challenging low-light conditions. A subjective evaluation on 15 human teleoperators further confirms the effectiveness of the integrated features for improved teleoperation. We demonstrate the benefits of dynamic Ego-to-Exo view generation and real-time pose rendering for remote ROV teleoperation by following navigation guides such as cavelines inside underwater caves. This new way of interactive ROV teleoperation opens up promising opportunities for future research in subsea telerobotics.

1 Introduction

Unmanned submersible vehicles such as ROVs (Remotely Operated Vehicles) play a crucial role in subsea inspection, remote surveillance, and underwater cave exploration [1–3]. They are particularly useful in inspecting deep-water structures and surveying confined spaces that are beyond the reach of human scuba divers [4, 5]. In a typical mission, ROVs are controlled by human operators from a surface vessel, who are responsible for the safe and efficient maneuvering of the vehicle [6, 7]. The control consoles for teleoperation typically offer real-time

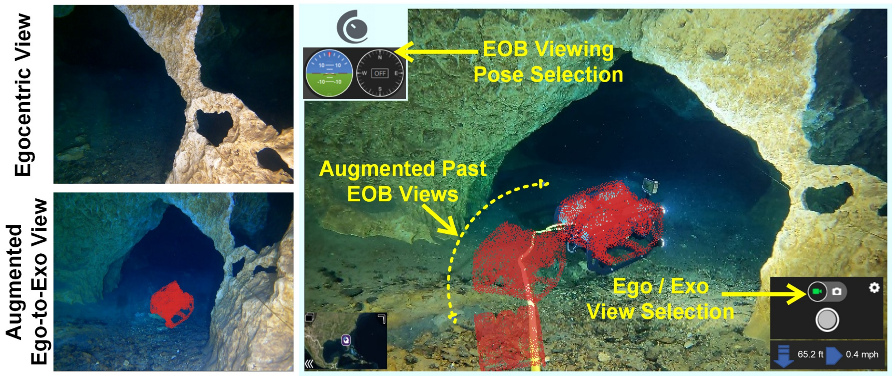


Fig. 1: The proposed Ego-to-Exo teleoperation interface is demonstrated for an underwater cave exploration scenario with an ROV. The traditional console interfaces are based on egocentric views (top left), which is limiting and disorienting to a surface operator in noisy low-light conditions. Our method generates on-demand exocentric views from a given EOB viewpoint, *i.e.*, third-person views from behind the ROV. This interactive viewpoint selection is integrated into a standard BlueROV2 console for an improved teleoperation experience.

data such as the egocentric video feed, pose, velocity, depth, etc. State-of-the-art ROVs can also include autonomous features for atomic tasks such as hovering [8], following navigation guidelines (*e.g.*, cavelines inside underwater caves [9, 10]), gripping objects [11, 12], trajectory estimation, etc.

While the subsea industries and agencies such as NOAA and naval defense teams deploy underwater ROVs with high-end cameras, sonars, and IMUs [2, 13] – safe and efficient teleoperation remains a challenge in adverse visibility conditions and around complex or sensitive structures. The typical first-person feeds from an ROV camera provide very limited information in landmark-deprived underwater scenes. The operators on the surface can only see the egocentric view, often without global or peripheral semantic information around the ROV. Although ROVs can use artificial lights to enhance visibility in low-light scenes, their bright light get reflected and back-scattered by suspended particles directly at the front camera [9], creating glare and large blind spots for the operator. Additionally, the autonomous and semi-autonomous features of ROVs also become erroneous without peripheral positioning in such noisy sensing conditions.

In this paper, we address these issues by introducing an AR (augmented reality) inspired ROV teleoperation interface that generates third-person (exocentric) perspectives as well as provides interactive control choices for viewpoint selection. As shown in Fig. 1, the proposed console can generate multiple exocentric views from past egocentric images, with a virtual ROV projected on the images as if it were taken by a third person following the robot. Our earlier work introduced the idea of EOB (Eye On the Back) visuals [14], which produces a single third-person view from immediately behind the ROV to facilitate better teleoperation. This work develops the AR-based end-to-end pipeline for generating on-demand exocentric perspectives given any EOB viewpoints. Additionally, we integrate the feature for geometrically accurate ROV positioning into those views for an interactive ROV teleoperation experience in challenging underwater applications. With these features, an operator can *slide across* multiple EOB viewpoints for the best view, see where the ROV is, and get comprehensive peripheral information for safe and efficient teleoperation.

Specifically, we design an efficient Ego-to-Exo (egocentric to exocentric) view generation framework integrated into a monocular visual SLAM system for underwater ROV teleoperation. The proposed Ego-to-Exo algorithm keeps track of the ROV camera poses and exploits a buffer of egocentric views for exocentric view synthesis. We then transform and project a pre-sampled 3D model of the ROV, in the form of a point cloud, into those views to generate realistic augmented visuals with more peripheral information. As seen in Fig. 1, such views offer comprehensive information of the surrounding scene with global semantics. In addition to the views, the integrated SLAM system provides real-time pose and map updates for atomic tasks [15, 16] such as obstacle avoidance, object following, next-best-view planning, etc.

The significance of this work resides in the simplicity of a 3D-geometric formulation of the Ego-to-Exo problem, and its integration into a real-time SLAM system. A monocular vision-only pipeline is chosen to ensure generalized utility and computational efficiency. The proposed Ego-to-Exo solution, EOB viewpoint-based parameterization, and ROV point cloud projection – are carried out by closed-form solutions to ensure real-time performance. As opposed to data-driven approaches for exocentric view synthesis, the proposed framework is invariant to the changes in waterbody style, scene geometry, and application scenario – making it transferable to any underwater ROV platforms.

We demonstrate the performance and utility of the proposed Ego-to-Exo framework through a series of experiments. First, we prove the geometric validity of our estimation with 2D indoor navigation scenarios utilizing a TurtleBot4. We quantify the geometric validity based on ground plane estimation errors and reprojection errors of known reference points in the scene. Next, we conduct experiments for underwater cave exploration scenarios where a BlueROV2 is teleoperated inside underwater caves and grotto systems at various geographical locations. Remote teleoperation in such cave systems includes unique challenges such as low visibility, turbid water conditions, moving shadow effects, and hazy blind spots due to artificial lights. We evaluate the usefulness of the proposed TeleOp console with 15 human subjects in such challenging scenarios. They use the interactive features such as viewpoint selection to visualize augmented ROV poses on the map in real-time. The subjects rank the complexity, efficiency, and other aspects of the system in a questionnaire as well as provide open-ended feedback regarding their cognitive workload.

We further discuss the observations for a variety of challenging cases of caveline detection and following in noisy low-light conditions inside multiple underwater caves. We demonstrate that the generated exocentric views, the augmented robot pose, and the rendered map embed significantly more information and global context about the scene to facilitate safe and efficient robot control.

2 Related Work & Applications

2.1 ROV Deployment & Teleoperation

Safe and efficient robot teleoperation is crucial for marine biology, oceanography, offshore energy, and infrastructure maintenance. Underwater ROVs enable researchers to study deep-sea ecosystems and collect samples from depths beyond the limits of human scuba divers. A few prominent applications are as follows.

Subsea structure inspection. Underwater structures operate in challenging remote environments and are susceptible to various hazards with significant economic and environmental consequences. Corrosion/erosion combined with tensile stresses caused by waves and seismic activities lead to cracking, leakage, and malfunction which are regularly inspected by ROVs [1, 17]. ROVs are also deployed for 3D mapping and acoustic profiling of marine structures [18, 19].

Underwater infrastructure maintenance & security. The underwater energy sector [20] and submerged data centers [21, 22] are experiencing rapid growth with increasing demand for regular maintenance and surveillance operations. Once ROVs are teleoperated to the site, they need to perform various autonomous and semi-autonomous tasks to inspect, monitor, and ensure safety and integrity of those infrastructures.

Marine archaeology and cave exploration. Exploring underwater heritage sites [23] and caves [4] are important to unravel archaeological history as well as for water resource management. Intelligent ROV capabilities enable effective mapping, sampling, and exploration of remote sites and challenging overhead environments. Moreover, seasonal monitoring of oceanographic features is performed for photometric studies as well [24–26].

Beyond these sectors, improved ROV teleoperation interfaces based on our proposed Ego-to-Exo augmented visuals will be useful in coastal habitat restoration and surveying, aquaculture, and search-and-rescue missions as well.

2.2 Third-Person Views for ROV Teleoperation

A common issue reported by ROV operators is that using a remote vision platform for teleoperation is like looking through a "*soda straw*" [27]. This is because the typical ROV controller interfaces are based on egocentric *first person camera* views – which provide no peripheral vision, resulting in significantly reduced situational awareness [28, 29]. As such, there have been multiple methods proposed to mitigate the issue and provide the operator with more visual data.

Two types of commonly used approaches exist for generating an exocentric view for unmanned ground and aerial vehicles. The first method uses an external camera to provide the exocentric view such as using a fixed camera to record the vehicle's trajectory from a distance [30], mounting the camera on a UAV (unmanned aerial vehicle) that flies above the target [31–34], attaching the camera to an elevated position on the robot [35], utilizing a follower ROV with a camera to provide the external view [36], or using a fish-eye camera to create a top-down perspective [37]. The second method utilizes integrated sensors on the robot, such as LiDAR (Light Detection and Ranging) to generate a point

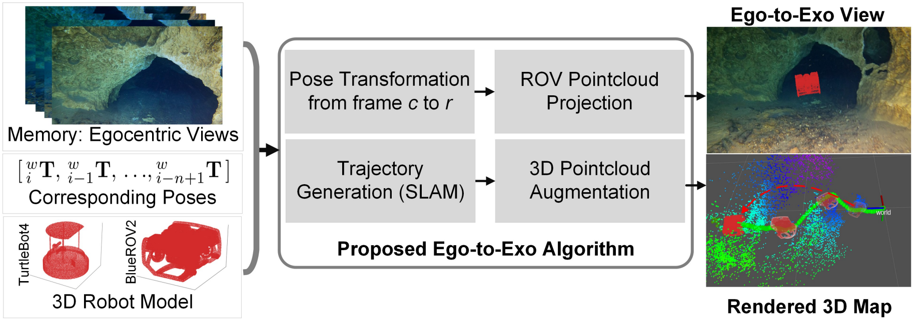


Fig. 2: The proposed Ego-to-Exo problem formulation is shown. The algorithm maintains a buffer of past egocentric views and corresponding ROV poses from the monocular SLAM system. Given an EOB viewpoint r, an exocentric view is synthesized with 3D robot's pose for more informative third-person perspectives.

cloud of the surrounding environment [38, 39] and use that toward creating an augmented/virtual environment for interfacing and teleoperation [40–42].

Adapting the aforementioned methods from terrestrial or aerial domains to underwater environments presents inherent challenges. Firstly, sending diver-robot teams [43] is not always an option in complex deep-water missions – which are the majority of use cases for ROVs. Secondly, UGVs that utilize past egocentric views [44, 45] primarily rely on GPS-based localization that does not apply to GPS-denied underwater environment. Unlike underwater ROVs, ground vehicles generally operate on a 2D plane with limited pitch and roll variations over rough terrain. Thirdly, installing an external visual system requires significant hardware modifications, *e.g.*, they need to be rugged and pressure-sealed, recalibrated for buoyancy and motion dynamics, and additional tether integration for high-speed exocentric data transfer. Even with all the structural modifications, an external camera will provide a single additional third-person perspective. We attempt to address these issues by synthesizing geometrically accurate exocentric views into a monocular SLAM pipeline given a dynamic viewpoint.

3 Problem Formulation

We formulate the Ego-to-Exo problem as a 3D geometric algorithm that involves generating an on-demand EOB (Eye On the Back) view, and then projecting the robot on it for an augmented rendering of the scene; see Fig. 2. The proposed method has the following computational components.

3.1 Curating Pose and Image Queue from Monocular SLAM

A monocular SLAM algorithm such as ORB-SLAM3 [46] provides a continuous solution for estimating and tracking camera poses from a sequence of monocular images. We use an ORB-SLAM3-based framework to obtain camera poses of each keyframe location to eventually construct the trajectory map of the teleoperated robot. Our adapter SLAM framework initiates the trajectory estimation process by building a pose buffer of length n: ${}^w\mathbf{T} \triangleq [{}^w_i\mathbf{T}, {}^w_{i-1}\mathbf{T}, \ldots, {}^w_{i-n+1}\mathbf{T}]$, where, ${}^w_i\mathbf{T} = [{}^w_i\mathbf{R}_{3\times3} \,|\, {}^w_i\mathbf{t}_{3\times1}]$ denotes camera pose at instance i in global ($world$) frame of reference. The corresponding raw egocentric views $\mathbf{I}$ for each instance is also stored in a queue $\mathbf{I} \triangleq [\mathbf{I}_i, \mathbf{I}_{i-1}, \cdots \mathbf{I}_{i-n+1}]$. These memory buffers are updated instantaneously as the robot pose changes during teleoperation. We use an empirically tuned threshold to trigger an update only when the pose change is significant to avoid unnecessary updates (when the robot is static).

3.2 Generating Exocentric Perspective

Given the pose memory ${}^{w}\mathbf{T}$ and egocentric views $\mathbf{I}$, we formulate the Ego-to-Exo problem of estimating an exocentric view from a reference location r, looking toward the robots current location c, where $r, c \in [i{-}n{+}1,\ i]$ and $r < c$. Typically, c is set to i (most recent available frame), and r remains a free variable with n known samples in memory – to mimic the EOB viewpoint generation.

We use the ROV point cloud model $\mathbf{P}$ of size $3 \times m$ as prior. These m points are transformed from current camera pose ${}^{w}_{c}\mathbf{T}$ to reference pose ${}^{w}_{r}\mathbf{T}$ using:

$$\tilde{\mathbf{P}} = ({}^{w}_{r}\mathbf{R}^{-1}\,{}^{w}_{c}\mathbf{R}) \cdot \mathbf{P} + ({}^{w}_{c}\mathbf{t} - {}^{w}_{r}\mathbf{t}), \tag{1}$$

where $[{}^{w}_{c}\mathbf{R} \,|\, {}^{w}_{c}\mathbf{t}]$ and $[{}^{w}_{r}\mathbf{R} \,|\, {}^{w}_{r}\mathbf{t}]$ represent the pose for current and reference (target) location in world coordinate, respectively. The transformed point cloud $\tilde{\mathbf{P}}$ is then projected onto the target image plane by using camera intrinsics $\mathbf{K}$ as:

$$\left[\mathbf{u}\ \mathbf{v}\ \mathbf{1_{m\times 1}}\right]^{T} = \lambda_1\, \mathbf{K} \cdot \tilde{\mathbf{P}}. \tag{2}$$

Here, $\mathbf{u}$ and $\mathbf{v}$ vectors denote the pixel locations (u,v) on image $\mathbf{I}_r$ for projection; λ_1 is the scale.

3.3 3D ROV Rendering and Map Update

While the SLAM system constructs a map of the surrounding, the proposed algorithm simultaneously renders the 3D ROV model on the same spatial context. As illustrated in Fig. 2, the ROV points $\mathbf{P}$ are transformed to the current camera location and projected based on the relative pose information ${}^{w}_{i}\mathbf{T}$ as follows:

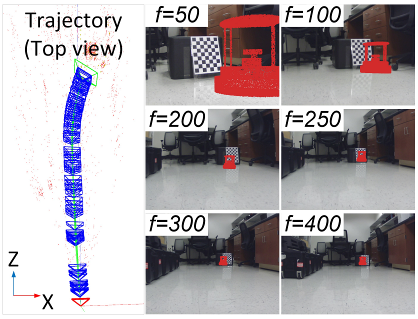
$$\tilde{\mathbf{P}}_{map} = \lambda_2\, {}^{w}_{i}\mathbf{R} \cdot \mathbf{P} + {}^{w}_{i}\mathbf{t}. \tag{3}$$

Here, λ_2 is the scaling factor for the ROV model. Note that our mapping and projection method is up to scale, like all monocular SLAM-based systems [47, 48]. While the scale can be resolved with additional sensor fusion, the augmented visuals of Eq. 3 are sufficient for teleoperation.

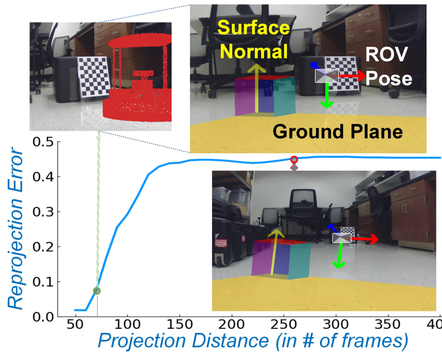
4 Implementation & Evaluation

4.1 Implementation Details

The framework is implemented using ROS Noetic in an Ubuntu 20.04 environment, running on an Intel Core i9 processor with 16 GB of RAM. A ROS node for ORB-SLAM3 is integrated as the monocular SLAM backbone. The buffer queue size n is set to 100 and the frame separation threshold is set to 0.001 unit (up to scale). The ROV point clouds are generated by sampling 3D mesh models of BlueROV2 and TurtleBot4; $10,000$ points are sampled for each model. The scaling parameters λ_1 and λ_2 are empirically tuned once for each test sequence according to the scale of the SLAM-generated map, ensuring that the rendered ROV model appears realistic in size.



(a) The TurtleBot4 trajectory during teleoperation is shown; the f numbers indicate the *EOB distance* from current to reference frame.



(b) Reprojection errors for reference checkerboard corners are evaluated for different EOB distances (f). The estimated ground plane is shown with a cube for visualization.

Fig. 3: We conduct 2D indoor navigation tests with a TurtleBot4 to validate the geometric accuracy of our method; here, results are visualized for ground plane estimation and reprojection errors of known reference points in the scene.

4.2 Proof of Concept: 2D Indoor Navigation

Experimental setup. The proof-of-concept experiments are conducted with TurtleBot4, a 2D ground robot that can be teleoperated with egocentric views from its front-facing monocular camera. It has only two degrees of freedom (DOF) for linear and angular velocity - which simplifies the motion kinematics for tracking its instantaneous position and orientation. We teleoperate it to collect visual data with a USB camera at 640×480 p resolution in office, laboratory, and hallway scenarios. The experiments are designed to validate the proposed algorithm by evaluating ground plane estimation and reprojection errors.

Geometric validation: reprojection error analysis. We first evaluate the reprojection errors of known reference points from the generated Ego-to-Exo views and the estimated ROV pose. As shown Fig. 3, we use standard checkerboard corners as reference points from egocentric views and then evaluate the reprojection errors for those points from exocentric views. This test is iterated over different sets of past egocentric images, each corresponding to a different *EOB distance*. As shown in Fig. 3a, a checkerboard is viewed from different EOB distances (further back into the past) indicated by the parameter f. More specifically, f is the number of frames between the current egocentric view and the selected EOB view. The corresponding reprojection error is plotted in Fig. 3b, which shows how the estimation is accurate for lower values of f, and gradually degenerates for $f > 100$. This is consistent with our visual observation of the projected ROV point cloud, *i.e.*, it is on the ground plane with accurate orientation based on the SLAM trajectory estimates.

Geometric validation: ground plane estimation. We also visualize the accuracy of the estimated ground plane in these experiments for qualitative assessments. Fig. 3b illustrates two cases: one for $f = 70$ with a low reprojection error, and another for $f = 260$ with a high error. As seen, the estimated ground plane (and the drawn sample cube) validates the geometric accuracy for the $f = 70$ case. On the other hand, a misaligned ground plane for the $f = 260$ case demonstrates the underlying error in pose estimation as well as in the reprojection process. Essentially, the geometric accuracy of our proposed Ego-to-Exo algorithm depends on the trajectory estimation performance of the SLAM system.

Computational efficiency. We analyze the computational complexity of the Ego-to-Exo algorithm for different configurations to ensure real-time execution in resource-constrained edge devices onboard standard ROV platforms. Table 1 shows the memory requirement of our algorithm for different choices of buffer size. The memory footprint is less than 300 MB for a buffer size of up to 180 frames, making it highly efficient. Table 2 demonstrates that the end-to-end framework maintains a consistent output rate of over 25 FPS (frames per second), making it suitable for integration in existing teleoperation engines. In fact, our algorithm only adds 6.8% overhead on the SLAM backbone for the map update.

Table 1: The memory requirement of Ego-to-Exo algorithm is compared for different buffer length.

Buffer size (# of frames)	50	100	200	300	400
Memory usage (MB)	65	142	301	455	609

Table 2: The computational complexity of the proposed Ego-to-Exo framework integrated into a SLAM backbone is reported. The metrics used are: (**i**) ROS node publish rate of the exocentric image; and (**ii**) the global map update rate.

Method	**SLAM Only**	**SLAM + Ego-to-Exo**
O/P image rate	26 FPS	25.1 FPS
Map update rate	26 FPS	25.3 FPS

4.3 Field Deployment: 3D Underwater Cave Exploration

Experimental setup. We extend our experiments to underwater cave exploration scenarios, where the ROV performs full 6-DOF motions. While the *roll* motion is limited in the standard BlueROV2s, we consider all 6-DOF for teleoperation with the buoyancy change and pressure imbalance caused by water flow at the cave openings. For remote teleoperation, we consider the scenarios where human operators maneuver an underwater ROV from the surface by following the caveline and other navigation markers as a guide [49]. The mission objective is to navigate the ROV 75-300 feet deep inside the cave through its complex structures, and then safely return it to the surface. The videos are recorded at 1920×1080 p resolution with a GoPro11 camera mounted on BlueROV2 and then compressed to 640×480 p the within the Ego-to-Exo framework. In addition to evaluating the geometric accuracy and robustness, we consider how informative the generated Ego-to-Exo views are compared to traditional consoles for ROV teleoperation.

Real-time map update and teleoperation. In addition to the Ego-to-Exo view generation and ROV pose rendering, our framework simultaneously updates a 3D map with extracted feature points from the SLAM system. Fig. 4 shows an

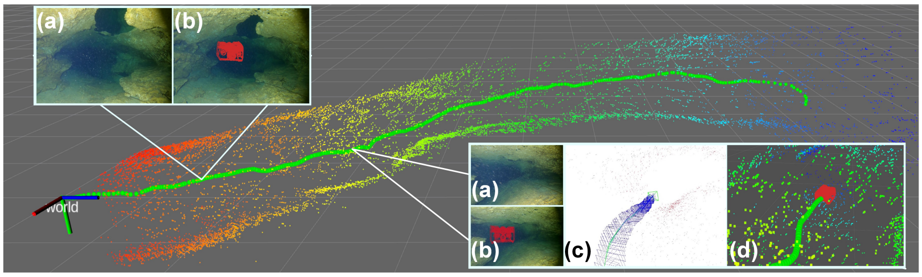


Fig. 4: The reconstructed 3D map of a cave segment in Devil's Springs, FL is shown. The colored points represent the tracked ORB features, whereas the ROV trajectory is shown in green circles. The popup frames show samples of: (a) egocentric images; (b) exocentric images with rendered ROV pose; (c) the underlying camera pose updates; and (d) an exocentric view of the 3D map.

ROV's trajectory mapped during an underwater cave mission in Devil's Springs, Florida. As seen, the generated Ego-to-Exo views embed significantly more peripheral information about the scene. The exocentric view of the ROV pose and its relative distance from cave walls or overhead obstacles are useful to surface operators for obstacle avoidance and efficient decision-making. Additionally, the 3D map shows the ROV's past trajectory and its current pose which are useful to analyze the mission progress, which is not possible in traditional teleoperation consoles. Such a global view of the trajectory is also useful during emergency evacuation and recovery. Beyond cave exploration, these features will be crucial in ROV-based subsea surveillance and search-and-rescue operations as well.

Accuracy and robustness. Due to the complex scene geometry and absence of a smooth ground plane inside underwater caves, we adopt a homography estimation approach for the performance validation. As shown in Fig. 5, April-Tag [50] corners are used as reference points for reprojection. Specifically, we compute the homography transformation between the egocentric and synthesized exocentric views to visualize the reprojection errors. We use a sample 2D logo and project it onto the reference April-Tag surface using the homographic transform. The unskewed planar projection validates the accuracy of the Ego-to-Exo pose estimation and point cloud rendering processes.

Observations: strengths and limitations. Our experiments reveal some key strengths of the proposed teleoperation framework. First, the generated exocentric views closely resemble the actual EOB views during a smooth trajectory, which is usually the case for subsea exploration and surveying tasks. Second, the buffer memory works as a backup during a temporary failure of the SLAM system, typically observed at turning corners or due to abrupt motion. In such cases, our algorithm retains historical poses from its buffer memory; teleoperators can utilize this for situational awareness to safely anchor or pause the mission until communication is restored. On the other hand, its heavy dependency on the SLAM backbone leads to some inherent limitations. Feature-based monocular SLAM systems often fail in feature-deprived noisy underwater scenes, which leads to inaccurate pose tracking and thus inaccurate Ego-to-Exo view synthesis.

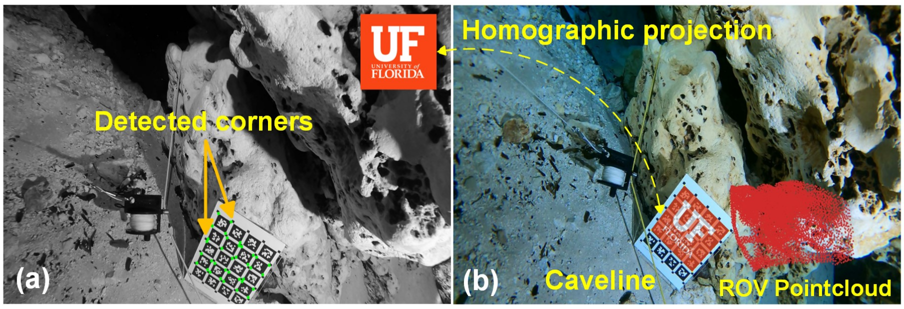


Fig. 5: A snapshot from our cave exploration scenario is shown: (a) Egocentric view with detected reference points; and (b) The synthesized Ego-to-Exo view. We use a sample logo and perform the homographic projection on the reference surface to demonstrate the accuracy in pose estimation.

Tracking 6-DOF ROV motion from monocular vision is particularly challenging with no additional sensor to recover the scale information [51, 52]. We observe some instances where the estimated ROV pose is incorrectly scaled in the rendering. To address this, multi-sensor fusion-based underwater SLAM backbones [53, 54] can be utilized in more critical applications.

4.4 Subjective User Study

The user study is conducted with multiple underwater cave exploration data. A BlueROV2 recorded egocentric video feeds inside the caves at up to 100-meter penetrations. Later on, 15 human participants, between the ages of 21-32 with no/little teleoperation experience, evaluate the ease of operation with our developed console and compare it to traditional consoles. Their feedback is recorded using the System Usability Scale (SUS) [55], with our interface achieving an average SUS score of 77.5. We also formulate an independent set of questions that reflect the teleoperator's preference for the novel features of our method. The individual questions and corresponding scores are presented in Table 3. Some key observations from this study are listed below.

(i) The obtained SUS score is fairly above median (68) and is considered `Good` for user experience; it is slightly below the `Excellent` (score: 80.3) category.
(ii) Post-operation feedback from our ROV operators suggests that the exocentric views are more useful for safe ROV maneuvers.
(iii) The synthesized 3D map provides a better sense of the ROV's global location and improves spatial awareness of the operators.
(iv) The operators report a significantly less workload (perceived cognitive load) in conducting complex tasks such as object following and structure mapping.

5 Improved Underwater ROV Teleoperation

Multiple augmented viewpoints. We validate the utility of our proposed Ego-to-Exo teleoperation interface through further experiments on underwater cave exploration data. Our expedition in cave segments at Devil's Springs,

Table 3: A total of 15 human subjects provide feedback to the following two sets of questions: Question $1-10$ from SUS [55], and the rest are custom designed. The responses are scaled from 1 (strongly disagree) to 5 (strongly agree).

#	Questions	mean, std
	System Usability Scale (SUS)[55]	
1	I think that I would like to use this system frequently.	4.3, 0.6
2	I found the system unnecessarily complex.	2.0, 0.7
3	I thought the system was easy to use.	4.3, 0.4
4	I would need the support of a technical person to use this system.	2.0, 0.6
5	I found the various functions in this system were well integrated.	4.0, 0.8
6	I thought there was too much inconsistency in this system.	2.3, 0.7
7	I would imagine that most people would learn to use it quickly.	4.4, 0.5
8	I found the system very cumbersome to use.	2.0, 0.6
9	I felt very confident using the system.	3.7, 0.7
10	I needed to learn a lot before I could get going with this system.	1.4, 0.5
	Custom Questions	
11	The proposed exocentric view is beneficial for ROV teleoperation.	4.5, 0.5
12	I found the EOB distance tuning feature useful to get the best view.	4.5, 0.5
13	The generated 3D map provides a better understanding of the ROV's global location and its surroundings.	4.6, 0.9

Florida reveals that when ROVs move slowly against strong currents, extending the exocentric viewpoint distance can significantly improve teleoperation. This is achieved by tuning the queue parameters r, c, and n in the proposed TeleOp interface. We consistently find that exocentric views are more informative, especially for about 5-10 seconds preceding the ROV position during navigation. The multiple preceding views offered by our interface are particularly useful for mapping large structures such as newly discovered cave segments or shipwrecks. As Fig. 6 shows, the synthesized viewpoints provide more spatial context, enabling operators to control the ROV efficiently around complex underwater structures.

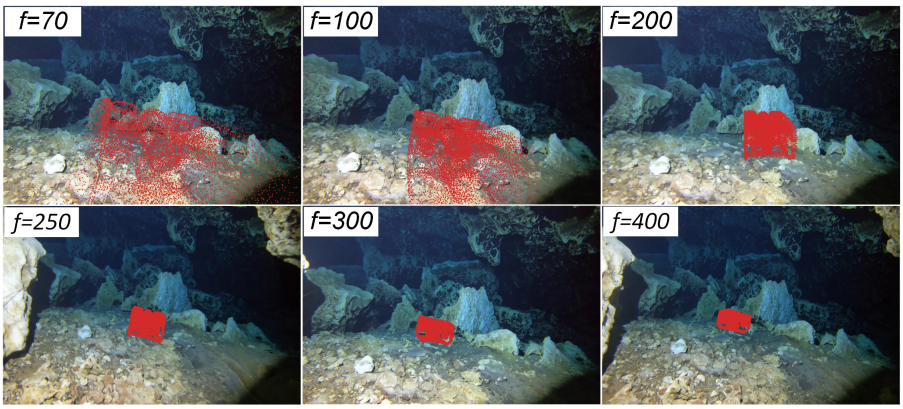


Fig. 6: A demonstration of our *adjustable EOB viewpoint* feature is shown. Teleoperators can *slide* across the EOB distance (f) and find the best exocentric view, which is $f = 200$ in this example.

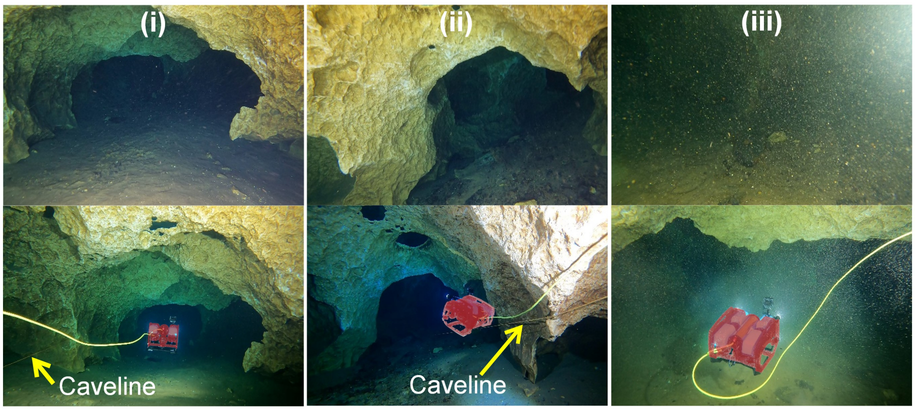


Fig. 7: Three challenging scenarios are shown for ROV teleoperation inside underwater caves: (**i**) caveline is not visible, *i.e.*, blended with the background; (**ii**) caveline is not in the FOV; and (**iii**) front camera-light interactions with suspended particles are causing hazy egocentric views. In all cases, our Ego-to-Exo augmented visuals are clearer and more informative to a surface operator.

Efficient teleoperation in complex missions. We also conducted trials inside an underwater cave system at Orange Grove, Florida, and a grotto system at Hudson, Florida. We observe that maneuvering the robot by following the caveline with egocentric views is challenging because little/no ambient light penetrates inside underwater caves. Despite using powerful lights, problems such as moving shadows and scattered waves create significant blind spots. Consequently, tracking and following the caveline or any other navigation markers [49] without any peripheral view is extremely disorienting to the operator. In some cases, we observe that the cavelines get blended with the texture and features of cave walls in noisy conditions; see Fig. 7. In such scenarios, shifting the viewpoint to exocentric views allows easier identification of cavelines against the surrounding and overhead cave walls. Additionally, the augmented 3D map displays the robot's pose, allowing much safer maneuvering of the vehicle to its desired orientation.

Safer navigation in hazy low-light conditions. Underwater caves present a unique formation of silt and sediment on their floor that results from erosion over extended periods. The silt is susceptible to disturbance from external factors, such as the motion of underwater ROVs or the turbulence generated by their propellers. Although ROV operators pay close attention to avoid contact with floor and cave-walls, it is often unavoidable due to buoyancy imbalance and strong flow of water. Dislodging the sediments results in cloudy or hazy conditions that obscure visibility. Bright lights from the ROV reflect from these suspended particles and make it even more challenging to capture clear imagery of the surroundings. In such cases, third-person EOB views from behind the ROV offer a clearer and more informative perspective for navigation; see Fig. 7. It improves spatial awareness and helps the operator to safely move away from the sediment formations toward open, accessible areas and avoid obstructing other scuba divers in the process.

6 Conclusion & Future Work

This work presents an AR-based framework to synthesize exocentric camera views from egocentric feed in real-time for improved underwater ROV teleoperation. A pose geometry-based closed-form solution is formulated for the proposed Ego-to-Exo problem and then integrated into a SLAM backbone. The end-to-end pipeline only requires a sequence of past egocentric views to generate an exocentric view with the accurate ROV model projected onto it. The proof-of-concept is validated by ground plane estimation and reprojection error analyses in a series of 2D indoor navigation experiments. Subsequent field experiments are conducted for various challenging ROV teleoperation scenarios inside underwater caves. A subjective study proves the advantage of the proposed TeleOp console over traditional systems and demonstrates that the framework: (**i**) offers more informative peripheral views, (**ii**) provides better situational awareness, and (**iii**) facilitates an interactive ROV teleoperation experience. We are currently exploring more comprehensive multi-sensor fusion-based underwater SLAM backbone such as the SVIn2 [53] for more accurate and robust estimation. We further plan to develop and integrate more interactive features that would serve as a simulation platform for ROV teleoperation and navigation research.

Acknowledgements

This work is supported in part by the NSF grants 2330416, 1943205, and 2024741. The authors would like to acknowledge the help from Woodville Karst Plain Project (WKPP), El Centro Investigador del Sistema Acuífero de Quintana Roo A.C. (CINDAQ), Global Underwater Explorers (GUE), Ricardo Constantino, and Project Baseline in providing access to challenging underwater caves.

References

1. Rumson, A.G.: The Application of Fully Unmanned Robotic Systems for Inspection of Subsea Pipelines. Ocean Engineering 235, 109214 (2021)
2. Wishnak, S.: New Frontiers in Ocean exploration: the Ocean Exploration Trust. NOAA Ocean Exploration, and Schmidt Ocean Institute 2021 Field Season (2022)
3. Siegel, V., Stone, W., Richmond, K.: Robotic Survey and 3-D Mapping of Underwater Caves using a SUNFISH® Autonomous Underwater Vehicle. LPI Contributions 2697, 1037 (2023)
4. Joshi, B., Xanthidis, M., Roznere, M., Burgdorfer, N.J., Mordohai, P., Li, A.Q., Rekleitis, I.: Underwater Exploration and Mapping. In: IEEE OES AUV Symposium. pp. 1–7. Singapore (Sept 2022)
5. Buzzacott, P.L., Zeigler, E., Denoble, P., Vann, R.: American Cave Diving Fatalities 1969-2007. International Journal of Aquatic Research and Education 3(2), 7 (2009)
6. Konoplin, A.Y., Konoplin, N.Y., Filaretov, V.: Development of Intellectual Support System for ROV Operators. In: IOP Conference Series: Earth and Environmental Science. vol. 272, p. 032101 (2019)
7. Kennedy, B.R., Cantwell, K., Malik, M., Kelley, C., Potter, J., Elliott, K., Lobecker, E., Gray, L.M., Sowers, D., White, M.P., et al.: The Unknown and the Unexplored: Insights into the Pacific Deep-sea Following NOAA CAPSTONE Expeditions. Frontiers in Marine Science 6, 480 (2019)

8. Jin, H.S., Cho, H., Jiafeng, H., Lee, J.H., Kim, M.J., Jeong, S.K., Ji, D.H., Joo, K., Jung, D., Choi, H.S.: Hovering Control of UUV through Underwater Object Detection Based on Deep Learning. Ocean Engineering 253, 111321 (2022)
9. Yu, B., Tibbetts, R., Barna, T., Morales, A., Rekleitis, I., Islam, M.J.: Weakly Supervised Caveline Detection For AUV Navigation Inside Underwater Caves. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 9933–9940. IEEE (2023)
10. Mohammadi, M., Huang, S.E., Barua, T., Rekleitis, I., Islam, M.J., Zand, R.: Caveline Detection at the Edge for Autonomous Underwater Cave Exploration and Mapping. In: IEEE International Conference on Machine Learning and Applications (ICMLA). Jacksonville, FL, USA (Dec 2023)
11. Manjunatha, M., Selvakumar, A.A., Godeswar, V.P., Manimaran, R.: A Low Cost Underwater Robot with Grippers for Visual Inspection of External Pipeline Surface. Procedia computer science 133, 108–115 (2018)
12. Lensgraf, S., Balkcom, D., Quattrini Li, A.: Buoyancy Enabled Autonomous Underwater Construction with Cement Blocks. In: IEEE International Conference on Robotics and Automation (ICRA). pp. 5207–5213 (2023)
13. Elor, A., Thang, T., Hughes, B.P., Crosby, A., Phung, A., Gonzalez, E., Katija, K., Haddock, S.H., Martin, E.J., Erwin, B.E., et al.: Catching Jellies in Immersive Virtual Reality: A Comparative Teleoperation Study of ROVs in Underwater Capture Tasks. In: ACM Symposium on Virtual Reality Software and Technology. pp. 1–10 (2021)
14. Islam, M.J.: Eye On the Back: Augmented Visuals for Improved ROV Teleoperation in Deep Water Surveillance and Inspection. In: SPIE Defense and Commercial Sensing. SPIE, Maryland, USA (2024)
15. Cai, W., Wu, Y., Zhang, M.: Three-dimensional Obstacle Avoidance for Autonomous Underwater Robot. IEEE Sensors Letters 4(11), 1–4 (2020)
16. Palomeras, N., Hurtós, N., Vidal, E., Carreras, M.: Autonomous Exploration of Complex Underwater Environments using a Probabilistic Next-best-view Planner. IEEE Robotics and Automation Letters 4(2), 1619–1625 (2019)
17. Yin, F.: Inspection Robot for Submarine Pipeline based on Machine Vision. In: Journal of Physics: Conference Series. vol. 1952, p. 022034. IOP Publishing (2021)
18. Xanthidis, M., Kalaitzakis, M., Karapetyan, N., Johnson, J., Vitzilaios, N., O'Kane, J.M., Rekleitis, I.: AquaVis: A Perception-aware Autonomous Navigation Framework for Underwater Vehicles. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 5410–5417. IEEE (2021)
19. Xanthidis, M., Joshi, B., Roznere, M., Wang, W., Burgdorfer, N., Quattrini Li, A., Mordohai, P., Nelakuditi, S., Rekleitis, I.: Mapping of Underwater Structures by a Team of Autonomous Underwater Vehicles. In: International Symposium of Robotics Research (2022)
20. Vederhus, L., Pan, Y.: Surface-to-seabed Safety: Advantages of Simulator Practice for Subsea Installation. International Journal of Safety and Security Engineering 6(2), 301–309 (2016)
21. Sheldon, J., Zhu, W., Abdullah, A., Butler, K., Islam, M.J., Rampazzi, S.: Deep note: Can acoustic interference damage the availability of hard disk storage in underwater data centers? In: 15th ACM Workshop on Hot Topics in Storage and File Systems. Best Paper Award. pp. 51–57 (2023), Best Paper Award
22. Curtis, S.: Subsea Fiber: Into the Deep. Optics & Photonics News 34, 32–39 (2023)
23. Eldred, R., Lussier, J., Pollman, A.: Design and Testing of a Spherical Autonomous Underwater Vehicle for Shipwreck Interior Exploration. Journal of Marine Science and Engineering 9(3), 320 (2021)

24. Islam, M.J., Li, A.Q., Girdhar, Y.A., Rekleitis, I.: Computer vision applications in underwater robotics and oceanography. Computer Vision: Challenges, Trends, and Opportunities pp. 173–196 (2024)
25. Neuner, B., Pascoguin, B.M., Hening, A., Dick, B.: Deployable Wavelength Optimization for Free-space Communication Undersea. In: MTS/IEEE OCEANS - Monterey. pp. 1–5 (2016)
26. Wang, W., Joshi, B., Burgdorfer, N., Batsos, K., Quattrini Li, A., Mordohai, P., Rekleitis, I.: Real-time Dense 3d Mapping of Underwater Environments. In: IEEE International Conference on Robotics and Automation. pp. 5184–5191 (2023)
27. Woods, D., Tittle, J., Feil, M., Roesler, A.: Envisioning Human-robot Coordination in Future Operations. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 34(2), 210–218 (2004)
28. Casper, J., Murphy, R.: Human-robot Interactions during the Robot-assisted Urban Search and Rescue Response at the World Trade Center. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 33(3), 367–385 (2003)
29. Zollmann, S., Hoppe, C., Langlotz, T., Reitmayr, G.: Flyar: Augmented Reality Supported Micro Aerial Vehicle Navigation. IEEE transactions on visualization and computer graphics 20(4), 560–568 (2014)
30. Jangir, R., Hansen, N., Ghosal, S., Jain, M., Wang, X.: Look Closer: Bridging Egocentric and Third-Person Views With Transformers for Robotic Manipulation. IEEE Robotics and Automation Letters 7(2), 3046–3053 (2022)
31. Gawel, A., Lin, Y., Koutros, T., Siegwart, R., Cadena, C.: Aerial-Ground Collaborative Sensing: Third-Person View for Teleoperation. In: IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR). pp. 1–7 (2018)
32. Saakes, D., Choudhary, V., Sakamoto, D., Inami, M., Lgarashi, T.: A Teleoperating Interface for Ground Vehicles using Autonomous Flying Cameras. In: International Conference on Artificial Reality and Telexistence. pp. 13–19 (2013)
33. Inoue, M., Takashima, K., Fujita, K., Kitamura, Y.: BirdViewAR: Surroundings-aware Remote Drone Piloting Using an Augmented Third-person Perspective. In: CHI Conference on Human Factors in Computing Systems. CHI '23, Association for Computing Machinery, New York, NY, USA (2023)
34. Erat, O., Isop, W.A., Kalkofen, D., Schmalstieg, D.: Drone-augmented Human Vision: Exocentric Control for Drones Exploring Hidden Areas. IEEE transactions on visualization and computer graphics 24(4), 1437–1446 (2018)
35. Shiroma, N., Sato, N., Chiu, Y.h., Matsuno, F.: Study on Effective Camera Images for Mobile Robot Teleoperation. In: IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN). pp. 107–112 (2004)
36. Nagatani, K., Kiribayashi, S., Okada, Y., Tadokoro, S., Nishimura, T., Yoshida, T., Koyanagi, E., Hada, Y.: Redesign of Rescue Mobile Robot Quince. In: IEEE International Symposium on Safety, Security, and Rescue Robotics (2011)
37. Sato, T., Moro, A., Sugahara, A., Tasaki, T., Yamashita, A., Asama, H.: Spatio-temporal Bird's-eye View Images using Multiple Fish-eye Cameras. In: IEEE/SICE International Symposium on System Integration (2013)
38. Ferland, F., Pomerleau, F., Le Dinh, C.T., Michaud, F.: Egocentric and Exocentric Teleoperation Interface using Real-time, 3D Video Projection. In: ACM/IEEE International Conference on Human Robot Interaction. p. 37–44. HRI '09, Association for Computing Machinery, New York, NY, USA (2009)
39. Lager, M., Topp, E.A., Malec, J.: Remote Operation of Unmanned Surface Vessel through Virtual Reality -A Low Cognitive Load Approach (03 2018)

40. Livatino, S., Guastella, D.C., Muscato, G., Rinaldi, V., Cantelli, L., Melita, C.D., Caniglia, A., Mazza, R., Padula, G.: Intuitive Robot Teleoperation Through Multi-Sensor Informed Mixed Reality Visual Aids. IEEE Access 9, 25795–25808 (2021)
41. Hing, J., Sevcik, K., Oh, P.: Development and Evaluation of a Chase View for UAV Operations in Cluttered Environments. Journal of Intelligent and Robotic Systems 57, 485–503 (08 2010)
42. Thomason, J., Ratsamee, P., Orlosky, J., Kiyokawa, K., Mashita, T., Uranishi, Y., Takemura, H.: A comparison of adaptive view techniques for exploratory 3d drone teleoperation. ACM Transactions on Interactive Intelligent Systems 9, 1–19 (2019)
43. Islam, M.J., Mo, J., Sattar, J.: Robot-to-robot Relative Pose Estimation using Humans as Markers. Autonomous Robots 45(4), 579–593 (2021)
44. Ito, M., Sato, N., Sugimoto, M., Shiroma, N., Inami, M., Matsuno, F.: A Teleoperation Interface using Past Images for Outdoor Environment. In: SICE Annual Conference. pp. 3372–3375. IEEE (2008)
45. Murata, R., Songtong, S., Mizumoto, H., Kon, K., Matsuno, F.: Teleoperation System using Past Image Records for Mobile Manipulator. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (2014)
46. Campos, C., Elvira, R., Gómez, J.J., Montiel, J.M.M., Tardós, J.D.: ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM. IEEE Transactions on Robotics 37(6), 1874–1890 (2021)
47. Kazerouni, I.A., Fitzgerald, L., Dooly, G., Toal, D.: A Survey of State-of-the-art on Visual SLAM. Expert Systems with Applications 205, 117734 (2022)
48. Macario Barros, A., Michel, M., Moline, Y., Corre, G., Carrel, F.: A Comprehensive Survey of Visual SLAM Algorithms. Robotics 11(1), 24 (2022)
49. Abdullah, A., Barua, T., Tibbetts, R., Chen, Z., Islam, M.J., Rekleitis, I.: CaveSeg: Deep Semantic Segmentation and Scene Parsing for Autonomous Underwater Cave Exploration. In: IEEE International Conference on Robotics and Automation (ICRA). IEEE (2024)
50. Olson, E.: AprilTag: A Robust and Flexible Visual Fiducial System. In: IEEE International Conference on Robotics and Automation. pp. 3400–3407 (2011)
51. 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: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 7221–7227. Macau (Nov 2019)
52. Wu, J., Yu, B., Islam, M.J.: 3D Reconstruction of Underwater Scenes using Nonlinear Domain Projection. In: IEEE Conference on Artificial Intelligence (CAI). pp. 359–361. IEEE (2023), Best Paper Award
53. Rahman, S., Quattrini Li, A., Rekleitis, I.: SVIn2: A Multi-sensor Fusion-based Underwater SLAM System. International Journal of Robotics Research 41(11-12), 1022–1042 (July 2022)
54. Mo, J., Islam, M.J., Sattar, J.: Fast direct stereo visual slam. IEEE Robotics and Automation Letters 7(2), 778–785 (2021)
55. Brooke, J., et al.: SUS- A Quick and Dirty Usability Scale. Usability Evaluation in Industry 189(194), 4–7 (1996)