Towards Multi-Camera Active Perception of Underwater Structures

Marios Xanthidis, James Johnson, Jason M. O'Kane, and Ioannis Rekleitis

***Abstract*— Mapping and monitoring underwater environments are topics of progressively increasing importance, but they also introduce several new challenges in robotics, due to the unique underwater conditions. Underwater robots operating close to underwater structures should be equipped with robust localization modules, robust navigation pipelines capable of safely navigating the underwater robot by sensing and avoiding obstacles, and collecting the necessary observations of the surroundings. Especially, tasks that require visual inspection executed by autonomous underwater robots are significantly challenging due to the visibility limitations of the underwater domain. We propose a new active perception framework for underwater robots utilizing arbitrary multi-camera configurations, which safely navigates the robot in close proximity to target structures. It also produces motions that encourage the robot to actively track multiple visual objectives, while dealing effectively with limited FOVs and sensing ranges. Our novel formulation of the active perception problem provides necessary building blocks and components for allowing the robot to track areas of interest that could interchangeably assist mapping, monitoring, or localization tasks. Although the approach is initially targeted to multi-camera systems, we show that it could be readily adapted for heterogeneous configurations with other sensors such as sonars and LIDARs. Preliminary results in simulation showing the strong potential of the proposed technique, together with applications and future extensions of the method are discussed.**

I. INTRODUCTION

Underwater operations using Autonomous Underwater Vehicles (AUVs) is a research topic that attracts attention of an increasing number of researchers and organizations in both academic and industrial settings. Pushed by recent advancements in hardware, real-time state estimation, and motion planning techniques, this strong current towards realizing autonomy in the underwater domain is supported by many potential applications, such as marine archaeology, underwater infrastructure inspection and maintenance, energy and resource utilization, public security, and environmental monitoring. Moreover, due to climate change threatening maritime infrastructure and requiring constant monitoring in isolated and hard to reach marine environments, underwater autonomy is becoming more essential than ever.

Currently, most —if not all— essential underwater tasks are performed by human operators directly who risk their health or even lives, or remotely by controlling ROVs which require significant human resources and logistics. In the first case, even excluding the risks, operations are constrained by water conditions, and especially the hard limitations on maximum depth and operation duration set by human biology. In the second case, even excluding the deployment costs, tethered operations are limited to unconfined spaces and generally uncluttered environments. Also, the human operators are controlling the ROV based on limited information from its sensors, which potentially leads in delays and underutilization of the platform's capabilities due to very conservative and overcautious operation.

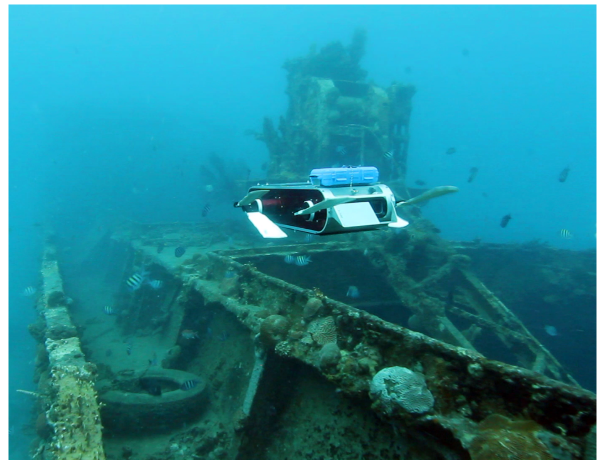


Fig. 1. Aqua2 AUV navigating over the Pamir shipwreck, Barbados.

Autonomous Underwater Vehicles (AUV) could perform tasks underwater without additional motion, depth, and duration limitations and especially without any risk to human lives. But AUVs deal with other important issues that are raised in the underwater environment, both with hardware and software. A major bottleneck of underwater autonomy is robust vision-based SLAM [1], [2], leading many state-of-the-art platforms to rely completely on dead-reckoning in order to avoid the visibility challenges of the underwater domain such as color attenuation, turbidity, and lack of color saturation, illumination, and feature-rich areas. A second important bottleneck is real-time motion planning, which should deal with motion uncertainty and safe operations in proximity to underwater structures and cluttered environments for monitoring, mapping, and exploration purposes.

Previous work has addressed these issues providing robust solutions for visual-inertial-based underwater state-estimation, by introducing SVIn [3] and SVIn2 [4], and a complete robust underwater navigation framework, called AquaNav [5], which was tested on the Aqua2 AUV — see Figure 1— in simulation, pool, and open-water trials. Though AquaNav provided safe and efficient paths that avoided obstacles in real-time, it had no consideration on

M. Xanthidis, J. Johnson, J. M. O'Kane, and I. Rekleitis are with the Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA. [mariosx, jvjjl]@email.sc.edu, [jokane, yiannisr]@cse.sc.edu

This work was made possible through the generous support of National Science Foundation grants (NSF 1659514, 1849291, 1943205, 2024741).

the future visibility of the few feature-rich areas of the underwater domain, which SVIn2 and many other vision-based SLAM techniques rely upon. Additionally, for the same reasons, AquaNav lacks the awareness needed for inspection, mapping, and monitoring purposes. Thus, introducing a new methodology for active perception will not only extend autonomy and minimize state-estimation uncertainty by driving the robot towards feature-rich areas, but also it will enable it to observe areas or objects of interest.

This paper proposes a novel formulation for active perception and a novel framework for a real-time perception-aware underwater navigation, called AquaVis. The proposed pipeline builds on the existing AquaNav pipeline and enables an underwater robot with an arbitrary multi-camera configuration to perceive multiple visual objectives, extracted automatically, along the path for mapping, monitoring, or localization purposes, by introducing two novel cost-functions in the optimization process. These visual objectives can be perceived online and from a desired distance, while at the same time safely reaching the desired goal. Though our primary focus lies on multi-camera configurations, as shown in our experiments, other sensors such as sonars and LIDARs, or combinations of sensors of different capabilities and attributes, could be used within our approach to offer more robust performance.

II. Related Work

The problem of active perception was first introduced by Aloimonos *et al.* [6] and Bajcsy [7] and then by Feder *et al.* [8] in the context of exploration. Additional work followed, developing active perception techniques for pose and map uncertainty reduction [9]–[18], but they only considered the 2D case, with no clear indication on being able to scale for applications in 3D – our target domain.

Recently, 3D active perception approaches have been developed for quadrotors but none of them could be applied directly to the problem considered in this study, because they either use direct photometric methods that perform poorly underwater [19], [20], do not provide sufficient obstacle avoidance guarantees for observations in proximity [19]–[29], consider only a single visual objective (or a single cluster of visual objectives) that should be always visible [24]–[33], are trained to operate only in known environments, or similarly to the previous works they assume a mobile robot that can perform lateral motions [21], [34]. Another common theme on these techniques is an attempt to tightly couple the objective with a low-level controller which might be ideal for the applications considered, but does not allow for developing high-level strategies for exploration, mapping and monitoring applications. Moreover, any active perception method applied to Aqua2 [35], a robot with past attempts to uncover its dynamics [36]–[38] but have yet to provide a complete model, should deal with motion planning and control in a loosely coupled or decoupled way, in order to utilize effectively its separate control module [39]. Additionally, due to the infeasibility of lateral motion by the platform, multiple visual objectives have to be considered along the path, assisting localization using a forward looking camera.

There are only few works dealing with active perception underwater, mostly attempting loop-closing in the context of coverage [40], [41], or curiosity-based exploration [42]. Other related applications to our objective have been applied in simulation for 3D coverage of shipwrecks [43], and real open water trials which avoided rocks and maximized visibility of corals [44], were based on deep learning techniques, trained upon motion commands obtained by human operators. Such techniques might not require odometry and mapping, a very challenging problem underwater, but in generally are unable to fully exploit the kinematic abilities of the agile Aqua2 platform, are limited to navigating only to similar environments to the training set, and very reactive with a short decision window that was compensated by following predefined local goals.

On the other hand, AquaVis, the framework proposed in this paper, attempts to provide a general systematic framework for active perception originally targeted to the underwater domain, but without explicit limitations on applications to other domains, such as aerial or space robotics, or even mobile manipulators. AquaVis as a purely model-based geometric method does not require a training set, thus it does not overfit to any suboptimal decisions of a human operator, and can be easily adjusted for different tasks by incorporating third-party packages for object recognition, or by automatically detecting feature rich areas to assist SLAM.

Unlike the previous methods discussed in this section, the generality and the applicability of the proposed technique extends to arbitrary exteroceptive multi-sensor configurations, regardless of the range sensor placement, FOV, capabilities, or type. Additionally, AquaVis decides actions that allow the robot to observe multiple objectives along the path, from a desired distance dealing with turbidity, which surpasses limitations of previous techniques and provides a more deliberative planning with a longer decision horizon. A very recent work [45] has also provided a technique with similar attributes within an underwater exploration framework, utilizing additional actuators for controlling the sensors independently, computationally expensive sampling-based motion planning approaches, and dead reckoning. AquaVis is applied on an agile robot with static sensor units, and inherits the robust navigation performance of AquaNav in cluttered environments, the typical type of environment for the target underwater structures considered. More importantly, it offers limited computational costs and fast replanning times, similarly to AquaNav, which is crucial for online application with SLAM.

III. The Proposed System

The goal of AquaVis is to produce safe and efficient paths that encourage observations of points of interest by prioritizing visibility over path length. An instance showing the desired behavior is provided in Fig. 2. The next paragraphs will explain the basic components of our proposed pipeline: providing a short overview of AquaNav, presenting the online extraction of visual objectives, our novel formulation for the

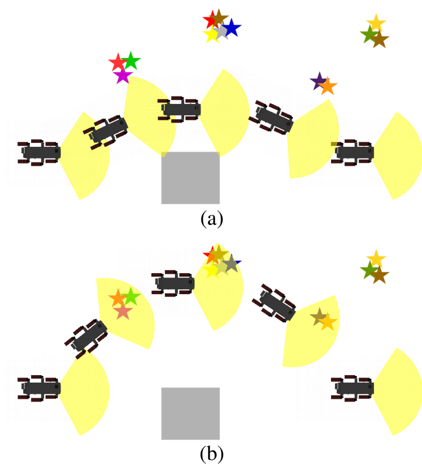


Fig. 2. An environment with obstacles (grey) and feature-rich areas indicated with stars. (a) AquaNav, considers only avoiding obstacles and minimizing the path length. (b) AquaVis, the new method introduced here, navigates the robot safely by avoiding obstacles, while at the same time observing nearby visual objectives.

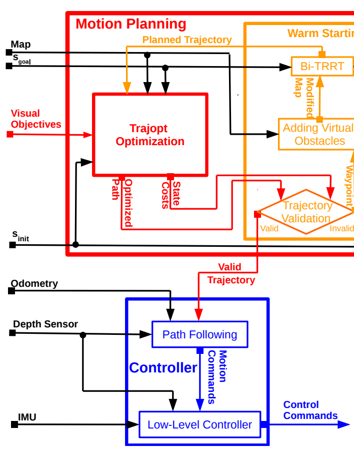


Fig. 3. System architecture of AquaVis, which is based on AquaNav. AquaVis alters the core planning component by incorporating visual objectives, shown with red, while modules for warm-starting, shown with orange, and path following, shown with blue, are kept the same.

active perception problem, and finally the motion planning modifications.

A. AquaNav

AquaNav [5] is a robust autonomous underwater navigation package, capable of real-time replanning, that produces efficient paths to operate safely in very challenging cluttered underwater environments. It operates on pointclouds produced by SVIn2 [4], a robust visual-inertial-based SLAM package for underwater operations. AquaNav deals with the expected motion uncertainty of Aqua2, by quickly producing minimum-length paths that satisfy clearance guarantees with a powerful path-optimization-based framework called Trajopt [46]. Potential local minimum issues are resolved rapidly with a novel warm-starting technique which alters the map to reinitialize the optimization with initial solutions that avoid previously detected local-mimimum areas, by addition of virtual obstacles and replanning with a very fast sampling-based motion planner called BiTRRT [47]. Finally, a waypoint navigation path follower is employed to generate the motion commands that are fed to the low level controller.

Due to its modularity, AquaNav is a very general framework that could become the basis of more specialized navigation planners in the future, and can be extended not only to different domains but also to different and more challenging navigation problems. We regard AquaVis, the proposed navigation planner of this work, the first instance showcasing the strong potential of AquaNav, by inheriting its unique safety and efficiency features, along with its upgrading prospects to tackle very challenging problems, such as underwater active perception, in real-time. The pipeline of AquaVis is shown in Fig. 3, and it is worth noting that the main difference with AquaNav is the new input of a set of visual objectives to be tracked along with novel constraints added to the optimization problem.

B. Extracting Visual Objectives Online

In the AquaVis pipeline as depicted in Fig. 3, a set of visual objectives to be tracked, if possible, represented by 3D points, is considered an input to enhance further the modularity and the potential applications of the proposed framework. Such input could be determined, for example, directly by the user, or by other independent third-party object recognition modules that detect corals [48]–[50], or other Aqua2 underwater robots [51].

In the absence of any specific monitoring task, in hopes of reducing position uncertainty in the notoriously challenging underwater domain [2], [52], we propose tracking the expected few but necessary feature-rich areas. Thus, a fast online method that identifies such areas during operations was developed to inform planning. Our approach is to treat the extraction of the feature-rich areas from the pointcloud obtained by the sensors as an unsupervised density-based clustering problem, in order to identify clusters with dense features. The process starts by applying the popular DBSCAN [53] algorithm on the 3D pointcloud. DBSCAN is very fast and highly parameter free, requiring as parameters only the maximum allowed distance for neighboring 3D points and the minimum number of points per cluster. Then, the centroids of each sufficient cluster are computed and used as the visual objectives.

A large number of accumulated visual objectives may negatively affect the replanning time of AquaVis, but this problem could be resolved easily by limiting the amount of visual objectives considered with policies as simple as using priority queue, or more sophisticated ones, such as updates based in proximity and relevance.

C. Active Perception Formulation

Most active-perception techniques utilizing optimization-based methods are focusing on producing a tightly coupled system between the sensor unit and the controls. Such approaches might offer faster computation time of motions that are dynamically feasible, something crucial for application on quadrotors, but the necessity for strictly convex objective functions restricts applications to a single sensor and a single visual objective. In our approach we treat the problem more

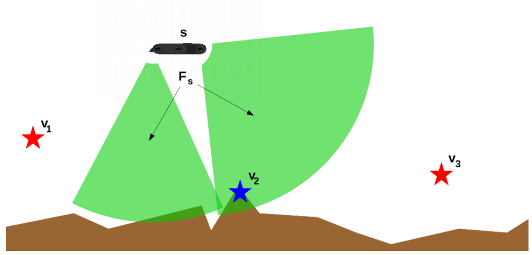
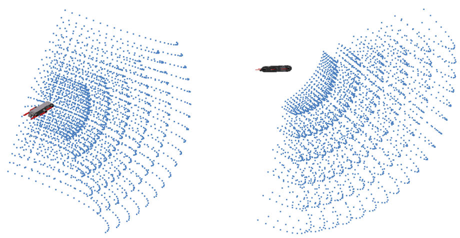


Fig. 4. The visibility manifold F_s for 2 cameras mounted on the robot is shown in light green. Visual objectives v_1, v_2, and v_3 are indicated with stars. Only v_2 is visible because it is inside F_s, while v_1 and v_3 are not observable from the robot's current state s.



(a) (b)

Fig. 5. Different perspectives of the projected points of the $F_s^{\sim}$ visibility set approximating the F_s visibility manifold corresponding to the front camera.

deliberatively by replacing the necessity for dynamically feasible paths with solutions that guarantee clearance, due to the expected motion uncertainty of Aqua2. The core path-optimization-based planning module utilized by AquaVis, similarly to AquaNav, is Trajopt. Following this concept, novel cost functions were developed for tracking multiple objectives, using multiple sensors, for a forward moving robot with 3D Dubins' kinematics.

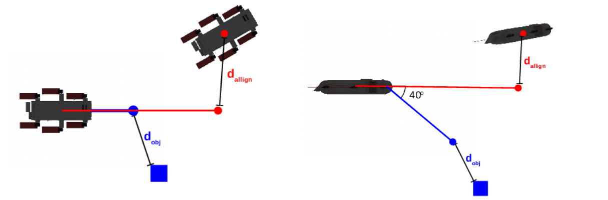
1) Visibility Constraints: Let $s \in SE(3)$ describe the robot's state (position and orientation) and a set V of 3D points representing the visual objectives in some common fixed coordinate frame. Also, let F_s represent the visibility manifold of the sensor units of the robot at state s, where for each point $o \in F_s$ the point o is visible by the robot at state s from at least one sensor. The visibility manifold is the union of the FOVs of all the sensors. Then each visual objective $v \in V$ is either visible from at least one sensor of the robot at state s ($v \in F_s$), or not visible at all ($v \notin F_s$). An example of the formulation as applied to the sensor configuration of an Aqua2 robot is shown in Figure 4.

Processing the true visibility manifold directly as a geometric polytope could be very challenging, so an approximation of the F_s manifold denoted as $F_s^{\sim}$ is used, formed by a representative collection of points, Fig. 5. Ideally, these projected points will sufficiently cover the area of the visibility manifold up to a desired resolution.

The key idea for the novel visibility cost function is to project points in front of the sensors and then attempt to minimize the distance d_{obj} between the closest nearby visual objective $v \in V$ with its closest point $f \in F_s^{\sim}$. More formally, the resulting cost function is the following:

$$\mathrm{Vis(s)} = \min_{v \in V} \min_{f \in F_s^{\sim}} ||f - v|| \tag{1}$$

Upon successful convergence of the Trajopt optimization



(a) (b)

Fig. 6. Top (a) and side (b) views of a state using the novel constraints during optimization. The blue square indicates a visual objective, and the red circle marks the next waypoint. Minimizing d_{obj} will result on observing the objective, while minimizing d_{align} will force the robot to be consistent with the kinematics assumed during path execution and planning.

process, by definition, at least one visual objective will be visible by at least one sensor in each state.

This formulation offers many options for developing different policies in the future, such as enforcing visibility of visual objectives from multiple sensors by using an $F_s^{\sim}$ representing the intersection of the FOVs of the desired sensors, maximizing visibility of many objectives by attempting to minimize the sum of the distances between each visual objective and its closest point of $F_s^{\sim}$, and alternate sensing of the target object between different sensors (such as cameras, LIDARs, or sonars) with respect to proximity by choosing the relevant parts of $F_s^{\sim}$ for each state during optimization.

Though the above formulation is powerful and could lead to superior performance in terms of path quality and visibility, at the same time it can add a severe computational overhead during planning, degrading the desired real-time performance of the AquaVis pipeline. However, both desired behavior and real-time replanning can be achieved at the same time, by significantly reducing the approximation quality of $F_s^{\sim}$, to only a single point projected at the center and at a desired distance d_{vis} of each sensor. This might seem reductive but it could be argued that such a trade-off is necessary due to the very limited computational resources of Aqua2, shared by the notoriously computationally heavy SLAM modules. Our experimental results showcase the robustness of the method despite this reduction.

More formally, let C represent the set of sensors, d_{vis} be a desired distance for projecting the points, T_w^s be the homogeneous transformation from the common fixed world frame to the robot's local frame at state s, and T_r^c be the homogeneous transformation from robot's local frame to the local frame of sensor $c \in C$. Then the redacted approximated visibility manifold can be constructed as:

$$F_s^{\sim} = \bigcup_{c \in C} \left\{ T_w^s T_r^c \begin{bmatrix} d_{\text{vis}} & 0 & 0 & 1 \end{bmatrix}^T \right\} \tag{2}$$

The above simple formulation in conjunction with Equation 1 guarantees that multiple visual objectives could be observed by multiple sensors along the path, and that at least one objective will be observed at each time if possible. Fig. 6 shows an instance of the proposed cost function and the distance d_{obj}, shown with blue.

2) Kinematic Constraints: The formulation described in the previous section solves the active perception problem sufficiently for holonomic robots, capable of moving freely in $SE(3)$ with no explicit kinematic limitations. Such a robot will always orient itself and move towards the most convenient visual objectives. Aqua2, though, and the majority of the mobile robots, are non-holonomic robots. Especially, in the case of Aqua2, the controller allows only for 3D Dubins' locomotion, and the path follower inherited by AquaNav performs a simple waypoint navigation. Due to the waypoint navigation, the kinematic constraints assumed during planning is that the robot will always face and move towards the next waypoint. This was not an issue that had to be addressed for AquaNav, since the only constraints affecting motion planning were path minimization, that minimizes rotations, and obstacle avoidance, that had no significant effect violating the kinematic assumption. By adding the visibility constraint, the robot will be encouraged to face directly the visual objectives, diverging significantly from the desired orientation towards the next waypoint.

The above important issue was resolved by adding another constraint, similar in nature, applied to each state that allows the robot to both observe the visual objectives while at the same time forces it to facing towards the next waypoint. Such alignment was achieved by projecting a single point in front of the robot at a specific distance and then minimizing the distance d_{align} of this point with the next waypoint. Fig. 6 shows an instance of such distance with red.

More formally, let p_s denote the 3D position coordinates of state s at a common fixed coordinate system, $S = [s_1, s_2, \dots, s_{n-1}, s_n]$ be the trajectory to be optimized, where $s_i,\ s_{i+1} \in S$ are two consecutive states, that correspond to the two consecutive waypoints p_{s_i} and $p_{s_{i+1}}$. Finally, let $len(S)$ return the total length of the S path, $av(S) = \frac{len(S)}{n-1}$ return the average distance between two consecutive states, let ϵ be a positive value. Then for each state s_i the kinematic constraint aligning properly the robot to produce valid trajectories is:

$$A(s_i) = \left\| T_w^{s_i} \begin{bmatrix} av(S)-\epsilon \\ 0 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} p_{s_{i+1}}^T \\ 1 \end{bmatrix} \right\|, \tag{3}$$

Similar to Equation 2 the first element of the first vector is the distance that the point will be projected. The distance needs to be automatically adjusted during optimization according to the continuously changing path length, while at the same time both maintains waypoints of equal distance and encourages minimal paths by being reduced by a small positive value ϵ.

By adding Equation 1 and Equation 3 as cost functions in the optimization formulation of AquaNav, a robust active perception behavior emerges as shown in the next section.

IV. Experimental Results

Simulation experiments were conducted to validate the robustness of the proposed AquaVis pipeline within the Gazebo simulation environment [54]. To simulate the output that is expected to be produced by vision-based SLAM techniques, we placed two simulated LIDARs with the same FOV as the cameras' configuration on the real robot, as shown in Fig. 4. The front stereo cameras have a horizontal FOV of $120°$, a vertical of $90°$, and they are tilted downwards by $40°$, while the back camera has the exact same FOV but it is tilted downwards by $90°$. The LIDARs output depth images of resolution 100×75 which are potentially thousands more than the expected during real deployment, to show the capabilities of the pipeline to deal with large inputs online.

We developed two different versions of AquaVis: (i) AquaVis-Mono that utilizes only the front camera both for extracting visual objectives and for registering obstacles, and (ii) AquaVis-Dual that additionally utilizes the back camera for informing the objectives extraction method and for sensing. Moreover, we set the maximum range of the sensors to different distances – $3\,\text{m}$ for the front cameras and $6\,\text{m}$ for the back one – to validate the behavior of AquaVis for an AUV employed with a heterogeneous multi-sensor system with different range capabilities. The desired distance from the visual objectives was set to $1.5\,\text{m}$, half of the maximum range of the front cameras, to encourage observations from the front camera system, and the desired clearance to $0.6\,\text{m}$ similarly to the original AquaNav pipeline which produces a safe behavior for the expected operating speed of the robot at $0.4\,\text{m/s}$.

Our primary motivation for this work is to enhance underwater operations both for state estimation and for mapping, inspection, and monitoring missions. So we compared AquaVis-Mono, AquaVis-Dual and the original AquaNav framework as baseline in two different environments.

A. Assisting Vision-Based State Estimation

The first environment, called the Boxes environment, aims to test the capabilities of the three frameworks to produce motions that provide good features and robustify underwater SLAM in environments where they are concentrated in few sparse feature-rich areas; a very common real scenario during underwater deployments. The results are shown in Fig. 7 for the AquaNav, the AquaVis-Mono, and the AquaVis-Dual pipelines respectively, with the first row showing the resulted trajectories, the second row the features tracked by the front camera, and the last row the features tracked by both cameras. The robot started from an initial position from which only a small segment of the first box was visible only from the front camera, and a goal set $25\,\text{m}$ forward.

Notice that AquaNav focused on minimizing the path length without regard for tracking features, while AquaVis-Mono was able to track few feature-rich areas with the front camera and all of them with the back. On the other hand, AquaVis-Dual tracked all the obstacles with both cameras, since the back camera feed was informing the path planning and assisting visibility from the front cameras. Such result showcases also the capabilities of AquaVis for heterogeneous sensor systems. For example the back camera could be considered equivalent to a sonar that has significantly better range than a camera underwater. Such sensor configurations could drive the robot towards potentially feature-rich areas detected by sonars from distance, and assist a visual SLAM module using the front stereo cameras.

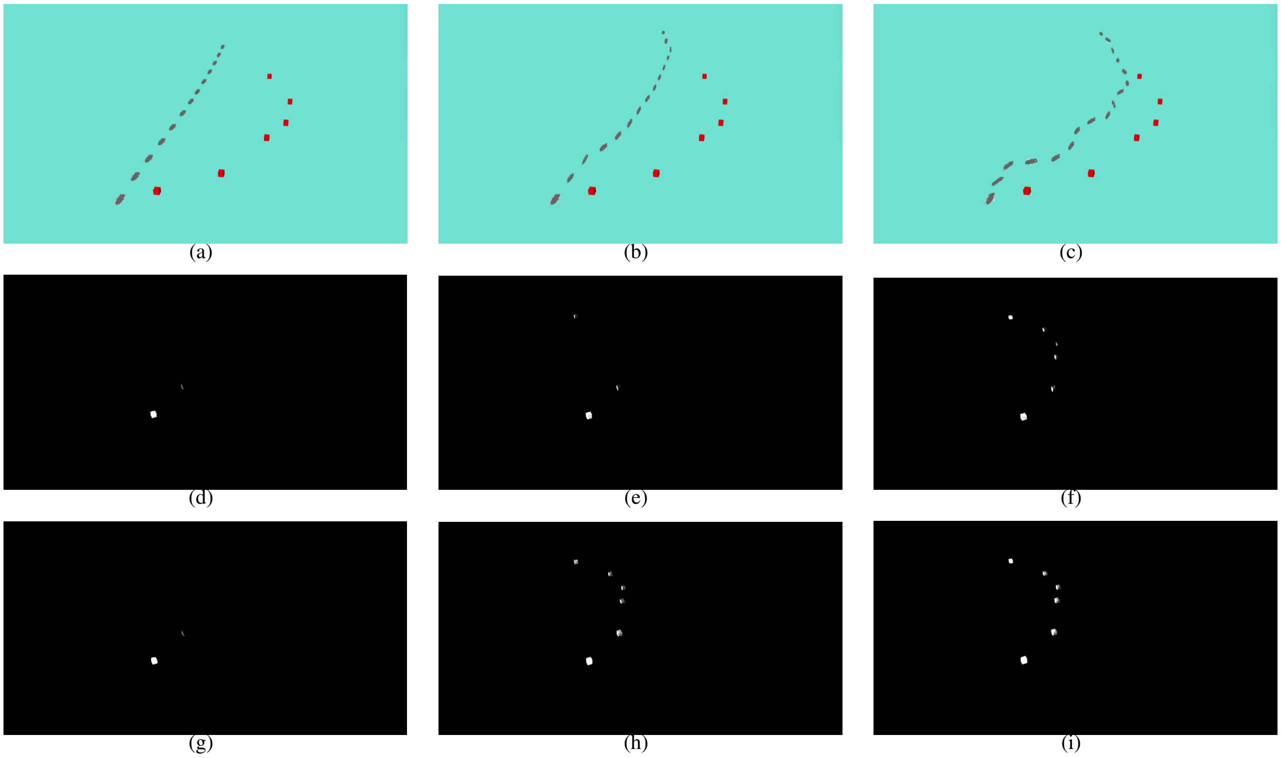


Fig. 7. First row: Trajectories produced by (a) AquaNav, (b) AquaVis-Mono, and (c) AquaVis-Dual for the Boxes environment. Second row: The corresponding point cloud obtained by the front camera. Third row: The corresponding point cloud obtained by both cameras. As expected, adding a second back camera of larger sensing range, informs planning robustifying further the behavior, while the limited range of the front cameras for AquaVis, leads to similar behavior with the uninformed AquaNav pipeline.

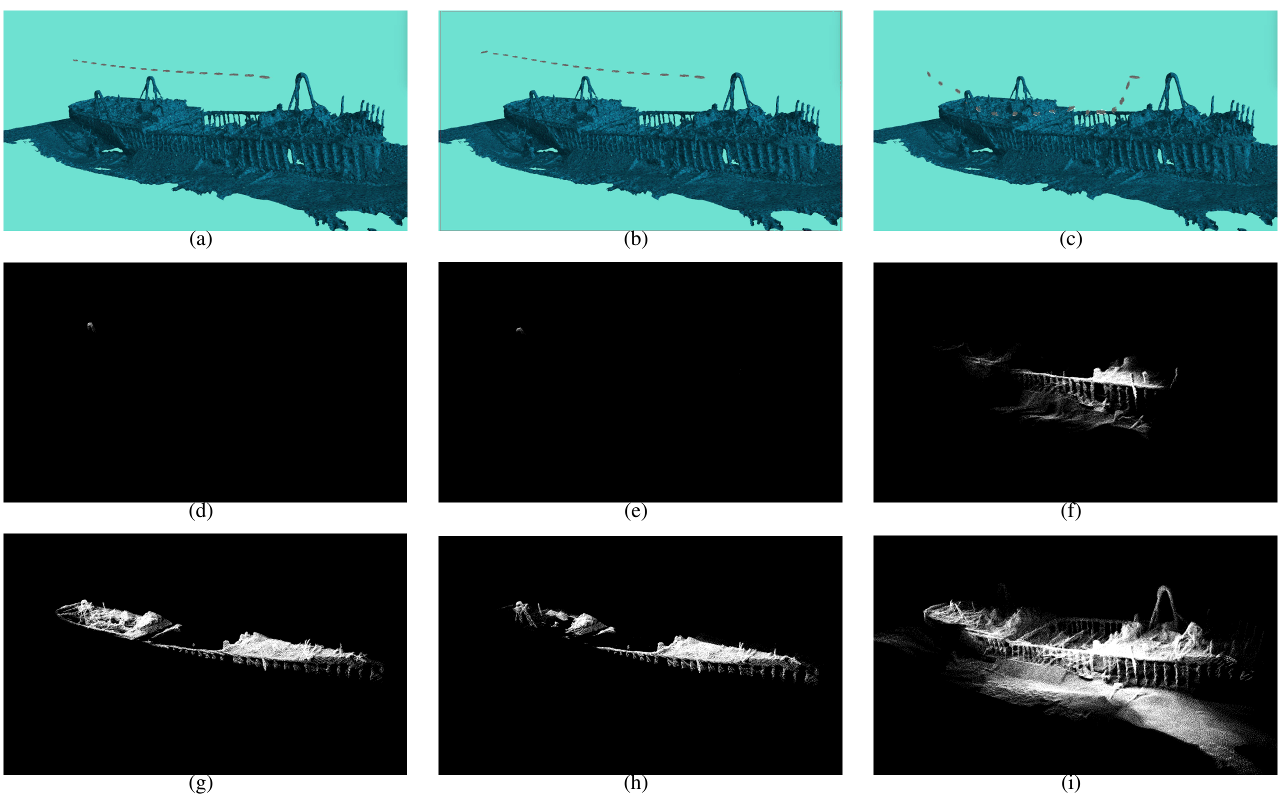


Fig. 8. First row: Trajectories produced by (a) AquaNav, (b) AquaVis-Mono, and (c) AquaVis-Dual for the Shipwreck environment. Second row: The corresponding point cloud obtained by the front camera. Third row: The corresponding point cloud obtained by both cameras. As expected, adding a second back camera of larger sensing range, informs planning robustifying further the behavior, while the limited range of the front cameras for AquaVis, leads to similar behavior with the uninformed AquaNav pipeline.

B. Enabling Mapping and Exploration

The second environment, called the Shipwreck environment, aims to compare the three frameworks on producing motions for mapping, monitoring, and exploring challenging underwater structures, such as shipwrecks; a significant motivation for this study and our previous works. The results are shown in Fig. 8 for AquaNav, AquaVis-Mono, and AquaVis-Dual with the first row presenting the resulting trajectories, the second row the features tracked by the front camera, and the last row the features tracked by both cameras. The initial pose of the robot did not allow it to see any features from the front camera, while a small segment of the shipwreck was visible from the back camera. The goal was set at $45\,\mathrm{m}$ forward, the approximate length of the shipwreck.

AquaNav and AquaVis-Mono had very similar performance, hardly observing any features with the front cameras, while only a small segment of the deck was captured by the back camera, due to its long sensing range. On the other hand, AquaVis-Dual was capable to observe a significant segment of the shipwreck with the short-sighted front cameras, while the majority of the shipwreck with both cameras. Moreover, using AquaVis-Dual the Aqua2 was driven in very close proximity to the shipwreck, which showcases the inherited strong safety and obstacle avoidance guarantees.

The results show great potential for application on robotic platforms with heterogeneous multi-sensor configurations, such as sonars with cameras. The pipeline allows replacing effortlessly the back camera, or even appending the current configuration, with a sonar on the Aqua2, to automatically detect with low resolution potential points of interest. In such configuration, AquaVis could drive the robot towards points of interest detected from distance using sonars to take close visual observations; a very useful attribute for many scientific, commercial, and security underwater operations.

V. Discussion

A. Importance of Contribution

In this study, a novel formulation for active perception was presented for tracking multiple visual objectives using mobile robots carrying arbitrary cameras and range sensors configurations. This novel formulation was presented in the context of underwater navigation, though without any explicit limitations towards applications to other domains. AquaVis, the proposed pipeline based on our novel formulation, to the best of our knowledge, is the first real-time computationally-light active perception underwater navigation method, able to track multiple visual objectives from multiple sensors along the path, while at the same time producing efficient 3D motions with a replanning time of 0.5-1Hz and showing safe performance in proximity to obstacles in cluttered environments for a robot with significant motion uncertainty.

B. Parameter Tuning

The core planning component of AquaVis, inherited by AquaNav is a path-optimization based package called Trajopt. A known challenge with such planners is the excessive parameter tuning often needed for a desired behavior. Indeed, this is potentially a drawback for the proposed method, but, in the authors' experience, tuning until reaching a desired stable performance was not a particularly intensive time consuming process. More specifically, the parameter values could be the same or very similar for similar platforms, and the tuning could be performed in a methodical way due to the intuitive cost functions introduced. No need for extra tuning was identified during addition of extra sensors in our tests.

C. Potential Modifications and Applications

Similarly to AquaNav, AquaVis aims to offer building blocks for underwater operations requiring high-level planning or a locally more sophisticated behavior. Regarding the latter, as noted in previous sections, extracting the visual objectives is an open problem and third-party frameworks could be utilized, while AquaVis could potentially be adjusted for maximizing visibility of multiple objectives simultaneously, obtain multiple observations using multiple sensors of a target visual objective simultaneously, move in a way that alternates between different sensors at different distances for monitoring specific areas of interest, and improving SLAM.

In practice AquaVis is a local planner that takes as inputs state and map estimates, a goal configuration, and visual objectives and produces as output safe motions observing the given visual objectives. By controlling any of these inputs the pipeline could readily be utilized by a high-level planner for mapping, monitoring, exploration of challenging underwater structures, multirobot operations, cooperative localization, human-robot collaboration, and search and rescue missions.

Also, it will be interesting to extend the AquaVis framework to other mobile platforms such UAVs by exchanging the path follower for a robust MPC-based controller. Finally, a maybe unexpected domain that AquaNav and AquaVis could be utilized with only few modification is manipulators and mobile manipulators. Trajopt was originally developed for such systems, and thus we aspire to solve such difficult problems utilizing hyper-redundant complex robotic systems.

References

[1] A. Quattrini Li, I. Rekleitis, S. Manjanna, N. Kakodkar, J. Hansen, G. Dudek, L. Bobadilla, J. Anderson, and R. N. Smith, "Data correlation and comparison from multiple sensors over a coral reef with a team of heterogeneous aquatic robots," in *International Symposium of Experimental Robotics (ISER)*, Tokyo, Japan, Mar. 2016.

[2] 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 Int. Conf. on Intelligent Robots and Systems*, 2019.

[3] S. Rahman, A. Quattrini Li, and I. Rekleitis, "Underwater cave mapping: Stereo visual slam with imu and sonar," in *IROS2017 Abstract*, Vancouver, BC, Canada, 2017.

[4] ——, "An Underwater SLAM System using Sonar, Visual, Inertial, and Depth Sensor," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2019.

[5] 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 *IEEE Int. Conf. on Robotics and Automation*, 2020, pp. 892–899.

[6] J. Aloimonos, I. Weiss, and A. Bandyopadhyay, "Active vision," *Int. journal of computer vision*, vol. 1, no. 4, pp. 333–356, 1988.

[7] R. Bajcsy, "Active perception," Un. of Pennsylvania Department of Computer and Information Science, Tech. Rep. MSCIS-88-24, 1988.

[8] H. J. S. Feder, J. J. Leonard, and C. M. Smith, “Adaptive mobile robot navigation and mapping,” *The Int. Journal of Robotics Research*, vol. 18, no. 7, pp. 650–668, 1999.

[9] R. Bajcsy, Y. Aloimonos, and J. K. Tsotsos, “Revisiting active perception,” *Autonomous Robots*, vol. 42, no. 2, pp. 177–196, 2018.

[10] G. Best, J. Faigl, and R. Fitch, “Online planning for multi-robot active perception with self-organising maps,” *Autonomous Robots*, vol. 42, no. 4, pp. 715–738, 2018.

[11] A. A. Makarenko, S. B. Williams, F. Bourgault, and H. F. Durrant-Whyte, “An experiment in integrated exploration,” in *IEEE/RSJ Int. Conf. on intelligent robots and systems*. IEEE, 2002, pp. 534–539.

[12] C. Stachniss, D. Hahnel, and W. Burgard, “Exploration with active loop-closing for fastslam,” in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, vol. 2. IEEE, 2004, pp. 1505–1510.

[13] I. Rekleitis, “Single robot exploration: Simultaneous Localization and Uncertainty Reduction on Maps (SLURM),” in *Ninth Conf. on Computer and Robot Vision*. IEEE, 2012, pp. 214–220.

[14] ——, “Simultaneous Localization and Uncertainty Reduction on Maps (SLURM): Ear based exploration,” in *IEEE Int. Conf. on Robotics and Biomimetics (ROBIO)*. IEEE, 2012, pp. 501–507.

[15] ——, “Multi-Robot simultaneous Localization and Uncertainty Reduction on Maps (MR-SLURM),” in *IEEE Int. Conf. on Robotics and Biomimetics (ROBIO)*. IEEE, 2013, pp. 1216–1221.

[16] Q. Zhang, D. Whitney, F. Shkurti, and I. Rekleitis, “Ear-based exploration on hybrid metric/topological maps,” in *2014 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*. IEEE, 2014, pp. 3081–3088.

[17] Q. Zhang, I. Rekleitis, and G. Dudek, “Uncertainty reduction via heuristic search planning on hybrid metric/topological map,” in *2015 12th Conf. on Computer and Robot Vision*. IEEE, 2015, pp. 222–229.

[18] R. Martinez-Cantin, N. de Freitas, A. Doucet, and J. A. Castellanos, “Active policy learning for robot planning and exploration under uncertainty.” in *Robotics: Science and Systems*, 2007, pp. 321–328.

[19] C. Forster, M. Pizzoli, and D. Scaramuzza, “Appearance-based active, monocular, dense reconstruction for micro aerial vehicles,” in *Proc. of Robotics: Science and Systems*, Berkeley, USA, July 2014.

[20] G. Costante, J. Delmerico, M. Werlberger, P. Valigi, and D. Scaramuzza, “Exploiting photometric information for planning under uncertainty,” in *Robotics Research*. Springer, 2018, pp. 107–124.

[21] V. Murali, I. Spasojevic, W. Guerra, and S. Karaman, “Perception-aware trajectory generation for aggressive quadrotor flight using differential flatness,” in *Am. Con. Conf. (ACC)*, 2019, pp. 3936–3943.

[22] R. Spica, P. Robuffo Giordano, and F. Chaumette, “Coupling active depth estimation and visual servoing via a large projection operator,” *The Int. Jour. of Rob. Res.*, vol. 36, no. 11, pp. 1177–1194, 2017.

[23] V. Indelman, L. Carlone, and F. Dellaert, “Planning in the continuous domain: A generalized belief space approach for autonomous navigation in unknown environments,” *The Int. Journal of Robotics Research*, vol. 34, no. 7, pp. 849–882, 2015.

[24] B. Penin, R. Spica, P. R. Giordano, and F. Chaumette, “Vision-based minimum-time trajectory generation for a quadrotor uav,” in *IEEE/RSJ Int. Conf. on Int. Robots and Systems*, 2017.

[25] M. Sheckells, G. Garimella, and M. Kobilarov, “Optimal visual servoing for differentially flat underactuated systems,” in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2016.

[26] T. Nägeli, J. Alonso-Mora, A. Domahidi, D. Rus, and O. Hilliges, “Real-time motion planning for aerial videography with dynamic obstacle avoidance and viewpoint optimization,” *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1696–1703, 2017.

[27] T. Nägeli, L. Meier, A. Domahidi, J. Alonso-Mora, and O. Hilliges, “Real-time planning for automated multi-view drone cinematography,” *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, pp. 1–10, 2017.

[28] C. Potena, D. Nardi, and A. Pretto, “Effective target aware visual navigation for uavs,” in *European Conf. on Mobile Robots*, 2017.

[29] I. Spasojevic, V. Murali, and S. Karaman, “Perception-aware time optimal path parameterization for quadrotors,” *arXiv preprint arXiv:2005.13986*, 2020.

[30] D. Falanga, P. Foehn, P. Lu, and D. Scaramuzza, “Pampc: Perception-aware model predictive control for quadrotors,” in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1–8.

[31] L. Yang, Z. Liu, X. Wang, and Y. Xu, “An optimized image-based visual servo control for fixed-wing unmanned aerial vehicle target tracking with fixed camera,” *IEEE Access*, pp. 68 455–68 468, 2019.

[32] C. Potena, D. Nardi, and A. Pretto, “Joint vision-based navigation, control and obstacle avoidance for uavs in dynamic environments,” in *European Conf. on Mobile Robots (ECMR)*. IEEE, 2019, pp. 1–7.

[33] K. Lee, J. Gibson, and E. A. Theodorou, “Aggressive perception-aware navigation using deep optical flow dynamics and pixelmpc,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1207–1214, 2020.

[34] B. Zhou, J. Pan, F. Gao, and S. Shen, “Raptor: Robust and perception-aware trajectory replanning for quadrotor fast flight,” *arXiv preprint arXiv:2007.03465*, 2020.

[35] G. Dudek, P. Giguere, C. Prahacs, S. Saunderson, J. Sattar, L.-A. Torres-Mendez, M. Jenkin, A. German, A. Hogue, A. Ripsman, *et al.*, “Aqua: An amphibious autonomous robot,” *Computer*, vol. 40, no. 1, pp. 46–53, 2007.

[36] C. Georgiades, M. Nahon, and M. Buehler, “Simulation of an underwater hexapod robot,” *Ocean Eng.*, vol. 36, no. 1, pp. 39–47, 2009.

[37] P. Giguere, C. Prahacs, and G. Dudek, “Characterization and modeling of rotational responses for an oscillating foil underwater robot,” in *IEEE/RSJ Int. Conf. on Int. Robots and Systems*, 2006, pp. 3000–3005.

[38] N. Plamondon and M. Nahon, “Trajectory tracking controller for an underwater hexapod vehicle,” in *IEEE/MTS OCEANS*, 2008, pp. 1–8.

[39] D. Meger, F. Shkurti, D. Cortés Poza, P. Giguère, and G. Dudek, “3d trajectory synthesis and control for a legged swimming robot,” in *IEEE/RSJ Int. Conf. on Int. Robots and Systems*, 2014, pp. 2257–2264.

[40] S. Frolov, B. Garau, and J. Bellingham, “Can we do better than the grid survey: Optimal synoptic surveys in presence of variable uncertainty and decorrelation scales,” *Journal of Geophysical Research: Oceans*, vol. 119, no. 8, pp. 5071–5090, 2014.

[41] S. M. Chaves, A. Kim, E. Galceran, and R. M. Eustice, “Opportunistic sampling-based active visual slam for underwater inspection,” *Autonomous Robots*, vol. 40, no. 7, pp. 1245–1265, 2016.

[42] Y. Girdhar, P. Giguere, and G. Dudek, “Autonomous adaptive exploration using realtime online spatiotemporal topic modeling,” *The International Journal of Robotics Research*, vol. 33, no. 4, pp. 645–657, 2014.

[43] N. Karapetyan, J. Johnson, and I. Rekleitis, “Coverage path planning for mapping of underwater structures with an autonomous underwater vehicle,” in *MTS/IEEE OCEANS - Singapore*, 2020.

[44] T. Manderson, J. C. G. Higuera, S. Wapnick, J.-F. Tremblay, F. Shkurti, D. Meger, and G. Dudek, “Vision-based goal-conditioned policies for underwater navigation in the presence of obstacles,” *Robotics: Science and Systems*, 2020.

[45] E. Vidal, N. Palomeras, K. Istenič, N. Gracias, and M. Carreras, “Multisensor online 3d view planning for autonomous underwater exploration,” *Journal of Field Robotics*, vol. 37, no. 6, pp. 1123–1147, 2020.

[46] J. Schulman, Y. Duan, J. Ho, A. Lee, I. Awwal, H. Bradlow, J. Pan, S. Patil, K. Goldberg, and P. Abbeel, “Motion planning with sequential convex optimization and convex collision checking,” *The Int. Journal of Robotics Research*, vol. 33, no. 9, pp. 1251–1270, 2014.

[47] L. Jaillet, J. Cortes, and T. Simeon, “Transition-based rrt for path planning in continuous cost spaces,” in *2008 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Sep. 2008, pp. 2145–2150.

[48] M. Modasshir, A. Quattrini Li, and I. Rekleitis, “Mdnet: Multi-patch dense network for coral classification,” in *MTS/IEEE OCEANS*, 2018.

[49] M. Modasshir, S. Rahman, O. Youngquist, and I. Rekleitis, “Coral Identification and Counting with an Autonomous Underwater Vehicle,” in *IEEE Int. Conf. on Robotics and Biomimetics*, 2018, pp. 524–529.

[50] M. Modasshir, S. Rahman, and I. Rekleitis, “Autonomous 3D Semantic Mapping of Coral Reefs,” in *Conf. Field and Service Robotics*, 2019.

[51] B. Joshi, M. Modasshir, T. Manderson, H. Damron, M. Xanthidis, A. Quattrini Li, I. Rekleitis, and G. Dudek, “DeepURL: Deep Pose Estimation Framework for Underwater Relative Localization,” in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA, 2020, pp. 1777–1784.

[52] A. Quattrini Li, A. Coskun, S. M. Doherty, S. Ghasemlou, A. S. Jagtap, M. Modasshir, S. Rahman, A. Singh, M. Xanthidis, J. M. O'Kane, and I. Rekleitis, “Experimental Comparison of open source Vision based State Estimation Algorithms,” in *Int. Symposium of Experimental Robotics (ISER)*, Tokyo, Japan, Mar. 2016.

[53] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise.” in *KDD: Proc. of the Second Int. Conf. on Knowledge Discovery and Data Mining*, Portland, Oregon, 1996, pp. 226–231.

[54] N. Koenig and A. Howard, “Design and use paradigms for gazebo, an open-source multi-robot simulator,” in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2004, pp. 2149–2154.