

# Towards Automated Sample Collection and Return in Extreme Underwater Environments

---

**Gideon Billings**

Naval Architecture and Marine Engineering  
University of Michigan  
Ann Arbor, MI 48109, USA  
gidobot@umich.edu

**Matthew R. Walter**

Toyota Technological Institute at Chicago  
Chicago, IL, 60637, USA

**Oscar Pizarro**

Australian Centre for Field Robotics  
University of Sydney  
NSW 2006, Australia

**Matthew Johnson-Roberson**

Naval Architecture and Marine Engineering  
University of Michigan  
Ann Arbor, MI 48109, USA

**Richard Camilli**

Applied Ocean Physics and Engineering  
Woods Hole Oceanographic Institution  
Deep Submergence Laboratory  
Woods Hole, MA, 02543, USA

## Abstract

In this report, we present the system design, operational strategy, and results of coordinated multi-vehicle field demonstrations of autonomous marine robotic technologies in search-for-life missions within the Pacific shelf margin of Costa Rica and the Santorini-Kolumbo caldera complex, which serve as analogs to environments that may exist in oceans beyond Earth. This report focuses on the automation of ROV manipulator operations for targeted biological sample-collection-and-return from the seafloor. In the context of future extraterrestrial exploration missions to ocean worlds, an ROV is an analog to a planetary lander, which must be capable of high-level autonomy. Our field trials involve two underwater vehicles, the SuBastian ROV and the Nereid Under Ice (NUI) hybrid ROV for mixed initiative (i.e., teleoperated or autonomous) missions, both equipped 7-DoF hydraulic manipulators. We describe an adaptable, hardware-independent computer vision architecture that enables high-level automated manipulation. The vision system provides a 3D understanding of the workspace to inform manipulator motion planning in complex unstructured environments. We demonstrate the effectiveness of the vision system and control framework through field trials in increasingly challenging environments, including the automated collection and return of biological samples from within the active undersea volcano, Kolumbo. Based on our experiences in the field, we discuss the performance of our system and identify promising directions for future research.

# 1 Introduction

A growing body of evidence suggests that the Earth is not unique in containing liquid water (Lewis, 1971; Gaeman et al., 2012; Malin and Edgett, 2000; Khurana et al., 1998), an essential ingredient for carbon-based life. Recent indications of water geysers emanating from moons of Saturn and Jupiter, including Enceladus (Nimmo et al., 2007) and Europa (Arnold et al., 2019), suggest that they may contain subsurface oceans with active hydrothermal venting (Lowell and DuBose, 2005; Hsu et al., 2015). Here on Earth, ocean floor hydrothermal systems and cold seep sites have long been known to host diverse chemosynthetic ecosystems that rely on the redox potentials of deep Earth fluids emitted from these sites to derive biochemical energy (Jannasch and Wirsén, 1979; Brooks et al., 1987), and may serve as analogs for oases of life elsewhere in our solar system and beyond. However, exploration for life within the distant oceans of Europa and Enceladus remains a daunting technological challenge. Robotic submersible vehicles equipped with manipulators provide a practical means for sample analysis and collection, enabling flexibility and dexterity without requiring precise and energetically costly positioning of the vehicle. Planetary landers such as the Mars Rovers have historically relied on human teleoperated manipulation using manually generated scripts (Lehner et al., 2018; Fong et al., 2001; Leger et al., 2005) to collect samples. However, teleoperation of robotic subsea vehicles within these putative ocean worlds is impractical because of high communication latencies (e.g., on the order of an hour for Europa). Thus, robotic missions must be capable of fully automated manipulation.

Marine robotic platforms such as remotely operated vehicles (ROVs) equipped with manipulators provide a useful testbed to develop automated manipulation and sampling technologies as analogs for space missions. Although Earth’s gravitational constant is higher than Europa and Enceladus, these moons’ estimated ice thicknesses of up to 30 km (Iess et al., 2014; Billings and Kattenhorn, 2005) are expected to present operational challenges, such as extreme pressure, near-freezing temperatures, and corrosion that are similar to Earth’s deep ocean environments. While autonomous underwater vehicles (AUVs) have been used for under-ice surveys for nearly 50 years (Francois and Nodland, 1972), deep ocean missions that require sample collection and return using manipulator arms are generally conducted using ROVs under direct human piloted control with cable-tethered communication. Only limited attempts at autonomous manipulation have been made in natural ocean environments (Marani et al., 2009; Shim et al., 2010; De Novi et al., 2010; Sanz et al., 2013; Peñalver et al., 2015; Sivčev et al., 2018b). The comparative lag in subsea manipulator autonomy behind terrestrial systems can be at least partly attributed to commercial systems being historically designed for direct teleoperation, with limited command modes and feedback, low control loop frequency, and poor repeatability Sivčev et al. (2018a). Despite these challenges, we demonstrate an automation framework that is compatible with existing commercial manipulator systems and that automates many high level tasks, while reducing risk through visual based scene understanding and pilot supervision.

In this paper, we consider the challenge of automated subsea manipulation and sample collection using existing ROV platforms as a technology analog for an under-ice exploration missions to Europa or Enceladus. We discuss the challenges that deep seafloor environments pose to automated robotic intervention and propose an architecture that overcomes many of these challenges. The system that we describe can be integrated on existing ROVs with minimal hardware requirements, namely, a vehicle-mounted stereo camera and a manipulator-mounted fisheye camera. We investigate the practical use of our perception methods to estimate the vehicle configuration, dynamically localize tools, and ground the transform between the natural scene reconstruction and the structured vehicle workspace. The manipulator control and vision processes that serve as the basis of this automation framework can be readily adapted to a variety of hardware configurations, making them suitable for a wide range of robotic platforms, including space flight systems. We demonstrate the flexibility of this framework through separate field trials performed with two different classes of ROVs equipped with substantively different manipulators. Figure 1 shows a conceptual diagram of how our system integrates with an ROV. In the current system implementation, a topside machine performs all processing using camera and manipulator data streamed from the vehicle over a high-bandwidth tether.

We conducted testing and field trials in progressively challenging environments, initially in laboratory settings and tank testing, followed by 11 dive missions at the Central American Pacific shelf margin of Costa Rica

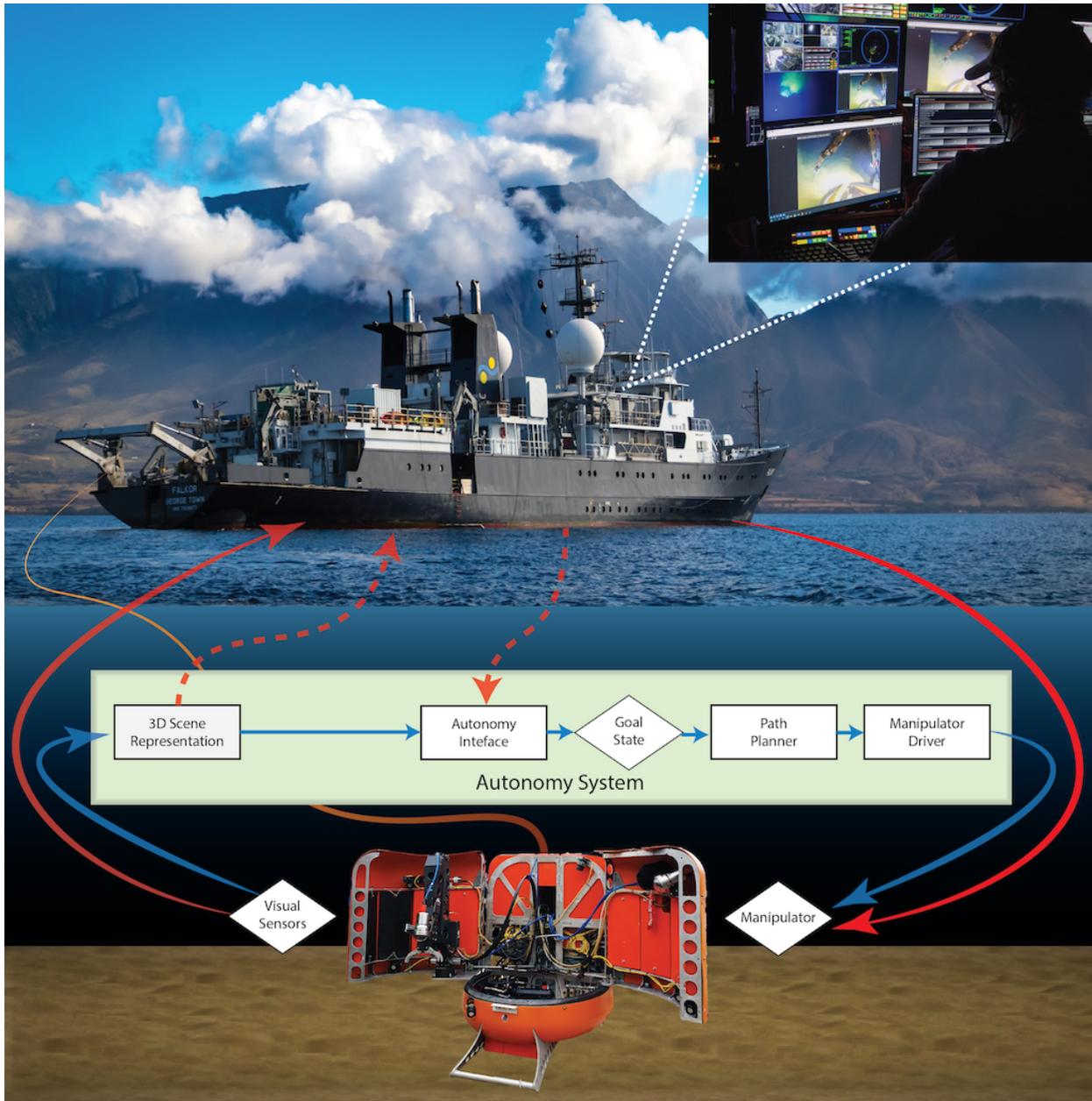


Figure 1: Conceptual graphic of the our control system for an underwater intervention vehicle. The autonomy system runs on a topside desktop computer with visual sensor data and manipulator coms streamed over a high bandwidth tether from the vehicle. Solid red flow lines represent standard teleoperated control from a surface ship. Blue flow lines represent our automated system. Red dashed lines represent interfacing between the pilot and the autonomous system, where, in this work, the pilot acts as the high level task planner and interfaces with the automated system through a graphical scene representation and task level controller. Eventually, the pilot would be replaced with an automated mission planner that could issue high level tasks.

to operational depths of approximately 1800 m. This area of the Costa Rican accretionary prism is a well studied region with localized ocean floor fluid expulsion sites that host diverse assemblages of extremophile organisms (Hensen et al., 2004; Krause et al., 2014; Levin et al., 2015; Sahling et al., 2008; Silver et al., 2000). Following the completion of the Costa Rica expedition, our team conducted a series of five dive missions to depths of 500 m within the potentially hazardous craters of the Kolumbo and Santorini calderas. These sites

of active volcanism contain localized areas of high-temperature hydrothermal venting causing environmental hypercapnia (Camilli et al., 2015; Carey et al., 2013), which host non-calcifying chemosynthetic organisms that may resemble those that arose early on Earth, prior to the advent of its oxidizing atmosphere.

The paper is organized as follows. We begin in Section 2 with an overview of previous work on automated underwater manipulation. Section 3 briefly outlines our strategy for mission planning and operations, and describes in detail the architecture of our automated manipulation system. Section 4 examines the results of experimental missions during field demonstrations using the *SuBastian* ROV (Campbell et al., 2019) and *NUI* Hybrid Remotely Operated Vehicle (HROV) (Bowen et al., 2014). Section 5 draws on these field results and experiences to discuss advances as well as the limitations and potential failure modes of our perception and control methods and examines how this research may help to advance both automated ROV operations here on Earth and future space flight missions to explore for life within ocean worlds elsewhere. Section 6 identifies promising directions for future research.

## 2 Background

There is a rich body of literature on underwater vehicle manipulator system (UVMS) control. This section provides a brief review of the work most related to our approach which have demonstrated their methods in experimental trials. For a thorough discussion of the prior work on UVMS systems, we refer the reader to Sivčev et al. (2018a).

Ishimi et al. (1991) and Broome et al. (1995) describe some of the pioneering work on automating UVMSs, where demonstrations included 3D graphical renderings of an ROV’s configuration and workspace, real-time visualization of manipulator motion plans, and Cartesian space end-effector control. More recent works under the large-scale research projects RAUVI (De Novi et al., 2010), TRIDENT (Sanz et al., 2013), TRITON (Peñalver et al., 2015), PANDORA (Cieslak et al., 2015), and MARIS (Simetti et al., 2017) focus on tightly coupled control of the 140 kg displacement *Girona 500* AUV outfitted with a customized electric manipulator to perform free-floating intervention tasks. The PANDORA project explores the ability to learn the vehicle and manipulator trajectories by demonstration. The other projects combine vehicle and manipulator motion generation under a task priority framework, where the manipulator control law is a function of the vehicle velocity. Building on these works, the MERBOTS project (Youakim et al., 2017) offers a significant advancement towards automated UVMS control by integrating the ROS-based MoveIt! motion planning framework with the intervention AUV to generate combined vehicle and manipulator motion trajectories in Cartesian space for free-floating intervention tasks. While this body of work provides key advancements towards automated free-floating intervention, limitations make it difficult for many actively operated UVMSs to adopt these methods. Among them, integrating such a tightly coupled control system with existing UVMS platforms would require significant modification to the software architecture, which is particularly problematic for commercial systems. Additionally, the dynamic coupling effect between the vehicle and manipulator during free-floating intervention can strongly affect the trajectory tracking performance, necessitating very slow actuation of the vehicle and manipulator. Lastly, this control approach is designed for high-precision electric manipulators that support velocity-based control, whereas most manipulators on operational UVMS platforms are hydraulic and support only position set point commands with limited precision and repeatability.

Hydraulic manipulators have orders of magnitude higher power-to-weight ratios compared to their electric counterparts and are generally more reliable, making them the manipulator of choice for commercial ROV systems. Though recent commercial electric manipulators have entered the market, their significantly higher power requirements make them practical only for ROVs that have power supplied over a tether. For vehicles like the *NUI* HROV, which carries all power onboard, low-power hydraulic manipulators remain the most practical choice. However, the limited precision and feedback of hydraulic manipulators present challenges for automation, and little work exists that addresses these challenges. Hildebrandt et al. (2009) demonstrated precision control of a hydraulic manipulator to plug a deep-sea connector. Shim et al. (2010) perform pre-

programmed motion following and operator control of a hydraulic work class manipulator. Using Cartesian space end-effector control, they demonstrate operator-guided push-core sampling in the deep ocean. Zhang et al. (2019) perform visual servoing and target grasping with a custom 7-DoF hydraulic manipulator. Sivčev et al. (2018b) demonstrate impressive visual servoing of a working class hydraulic manipulator using position-based control, with feedback provided by fiducials detected from a wrist-mounted camera. Their results include grasping and turning T-bar valves and tracking targets in motion with the end-effector.

Our system builds on these prior approaches to UVMS control, where we demonstrate the effective integration of the MoveIt! motion planning framework (Coleman et al., 2014) with a work class ROV manipulator system for automated planning and control in obstructed scenes. We take a decoupled approach to manipulator control that assumes the vehicle holds station (i.e., rests on the bottom) during the manipulation task. This assumption is motivated by the goal of having the system widely transferable among existing ROV systems. This decoupled approach enables our manipulator control system to be integrated externally from the existing UVMS control systems, providing high-level autonomy with flexibility to be integrated onto a wide array of ROV classes and manipulator arms, including both electric and hydraulic systems.

Important to automating UVMSs are the problems of visual scene understanding and target localization, whether the target be a tool to grasp, a valve to turn, or a sample location in an unstructured environment. Subsea perception is a particularly challenging problem for a number of reasons: turbidity degrades image quality; evenly lighting the scene is very difficult; variable wavelength-dependent absorption and scattering properties of the water column attenuate light and reduce color and photometric contrast; and gathering underwater datasets for developing computer vision methods is expensive. Despite these challenges, computer vision remains the primary means of performing target localization for automated UVMS platforms. Most prior works on UVMS automation rely on fiducials or known geometric shapes that retain high contrast underwater. Marani et al. (2009) use large spherical markers attached to a target and a circle shape edge detector algorithm to localize the marker from a video feed. Sanz et al. (2013) localize a black box object on a harbor seabed by first constructing a visual mosaic from a pre-intervention survey dive with a downward-facing stereo camera, and then matching an image template of the black box to the mosaic. Cieslak et al. (2015) localize a known panel during intervention operations by registering interest points against a template image. They then estimate the orientation of valves on the panel based on edge detection. Peñalver et al. (2015) and Youakim et al. (2017) use fiducial markers to localize a panel with *a priori* known relative positions of the turn valves and connector plugs. Youakim et al. (2017) also use fiducials on the end-effector of the manipulator to update the manipulator calibration in real-time. Simetti et al. (2017) use color and geometric shape segmentation of RGB images to detect the pose of a cylindrical pipe of known size. Under the DexROV project, Birk et al. (2018) process stereo point clouds into a 3D occupancy map, while also using fiducial markers to detect and localize a panel with known structure that was projected into the planning scene.

Building on the long history of fiducials as a robust visual cue for underwater computer vision methods, our work extends the use of fiducials to detect the pose of graspable tools carried on-board the ROV, estimate dynamic vehicle configurations in real-time, and ground the relative reference frames in the planning environment. We demonstrate the use of fiducials in a way that is practical for field deployments with an underlying vision system that can effectively localize tools and target objects within the workspace, as well as reconstruct the workspace for obstructed motion planning. Fiducials also enable the collection of annotated image datasets in natural deep seabed environments that support the development of advanced perception methods for scene reconstruction, and target detection and localization.

### 3 System Overview

The following sections provide an overview of the different components of our system, including the mission architecture for field operations and the methods for perception and control.

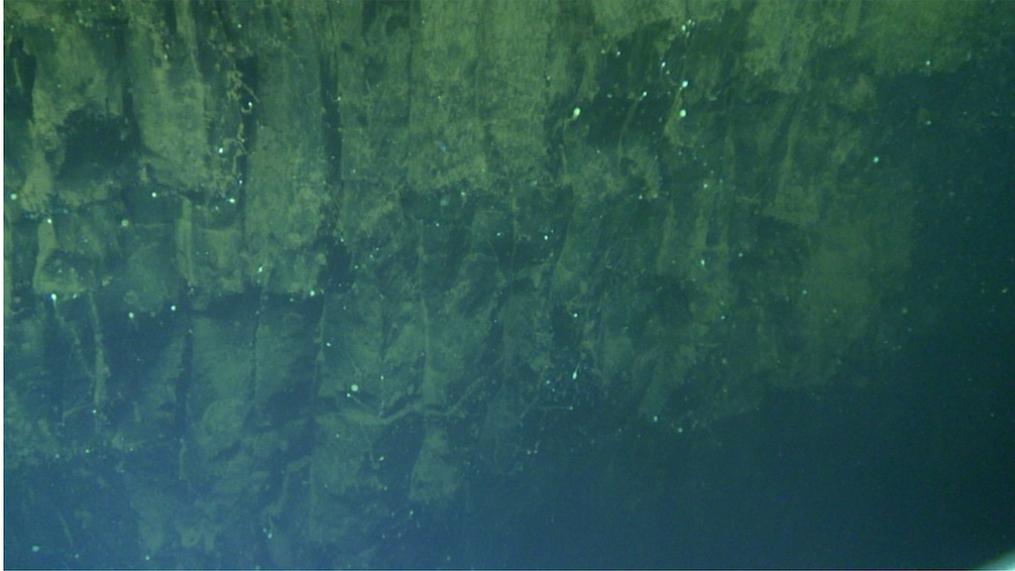


Figure 2: Photograph taken by the *NUI* vehicle within the Kolumbo volcano crater that shows an overhanging vertical wall of columnar lava. Colonization of the lava surfaces by relatively uncommon lollipop sponges (*Stylocordyla pellita*) are visible as white dots within the image.

### 3.1 Mission and Vehicle Platform Architecture

Field demonstration and validation include two research cruises, conducted east of the Cocos and Caribbean tectonic subduction zone along Central America’s Pacific continental margin (9.0 N 84.5 W), and within the Kolumbo and Santorini Calderas of the Hellenic volcanic arc in the southern Aegean Sea (36.52 N 25.48 E and 36.45 N 25.39 E, respectively). The sites, which are known to host oases of chemosynthetic communities associated with hydrothermal and seafloor hydrocarbon seeps, were chosen as NASA TRL-6 demonstration locations for analog astrobiology exploration missions. These campaigns utilized a sequentially nested survey method with a coordinated team of heterogeneous robotic platforms that relied on automated planing tools to rapidly synthesize vehicle missions in response to newly acquired information (Vrolijk et al., 2021; Ayton et al., 2019). To better approximate an analog space flight mission scenario, surface ships operated as orbiters, conducting multibeam sonar bathymetric mapping of the Pacific (Vrolijk et al., 2019) and Aegean (Nomikou et al., 2019) campaign sites, with coverage areas of 2.000 km<sup>2</sup> at 30 m resolution and 48 km<sup>2</sup> at 10 m resolution, respectively. These maps informed the mission planning for autonomous underwater gliders (AUG), which acted as long-range in-situ reconnaissance drones, conceptually similar to NASA’s *Ingenuity* and *Dragonfly* vehicles, conducting reconnaissance missions of between 1 km and 500 km in length at standoff distances to within 15 m of ocean floor obstacles in order to identify potential areas of scientific interest (Vrolijk et al., 2021; Duguid and Camilli, 2020). Automated AUG mission planning considered resource (e.g., time and power) and risk constraints (Timmons et al., 2016, 2019), and adaptively replanned missions based on inferred sites of scientific interest that correlated with the presence of active ocean floor hydrocarbon cold seeps and hydrothermal vents. Using information gained by the surface ship sonar and AUG missions, the automated planning process then generated viable mission sequences that the ROV used to investigate areas of highest estimated information gain (Vrolijk et al., 2021). During these missions, the ROV acted as a lander, outfitted with a manipulator for automated sample collection and return. The hazardous deep ocean environments explored as part of these ROV missions are considered probable analogs for environments (Fig. 2) that may exist on other ocean worlds.

The two ROVs used for these demonstration campaigns, *SuBastian* and *Nereid Under Ice (NUI)* are substantially different in design and purpose (Table 1). *SuBastian* is an exemplar of modern deep ocean work class ROVs, with its power, communications, and navigation net provided via an armored cable by its at-

Table 1: Comparison of *SuBastian* and *NUI* configurations.

	Depth rating meters	Displacement kilograms	Lateral excursion (tethered) meters	Power draw (typical) watts	Endurance hours	Manipulator type	Manipulator reach meters	Payload capacity kilograms
<i>SuBastian</i>	4500	3200	< 500	40000	unlimited	2 x 7-DoF	1.9	200
<i>NUI</i>	2000	2000	20000	2500	6 to 8	7-DoF	1.3	100



(a) Wrist-mounted fisheye camera



(b) Vehicle-mounted stereo camera

Figure 3: The vision system for autonomy is composed of (a) a wrist-mounted fisheye camera and (b) a vehicle-mounted stereo pair (shown here mounted on the *SuBastian* ROV). The vision system can be easily integrated onto existing vehicles.

tendant surface ship, the *R/V Falkor*. *SuBastian* is equipped with twin 7-DoF Schilling Titan-4 hydraulic manipulator arms (Schilling Robotics, Davis, California) and is a fully teleoperated vehicle that can operate at horizontal excursions of up to 500 m laterally from the *R/V Falkor*. In contrast, *NUI* is a HROV that relies on its own battery power and uses an un-armored fiber optic link (roughly the diameter of a human hair) for optional communication with an attendant surface ship, and can operate as both an ROV and an AUV. When in tethered ROV mode, *NUI*'s power and telemetry architecture enables lateral excursions of up to 20 km from the attendant surface ship. To aid hydrodynamic efficiency, *NUI* has articulating bow doors that can be closed and act as a fairing during transits and AUV missions. In contrast to *SuBastian*'s twin Titan-4 architecture, which is configured to maximize ROV work area and dexterity, *NUI*'s starboard door is equipped with a single, custom 7-DoF hydraulic manipulator (Kraft Telerobotics, Overland Park, Kansas) that is optimized for energy efficiency. This emphasis on efficiency comes at the expense of reductions in available payload and work space, the usable range of motion, control precision, lighting field, and available viewing perspectives.

### 3.2 Perception

A vehicle capable of automated intervention must have an effective means to self-localize within the environment and visually reconstruct the workspace to complete the mission tasks. We adopt a vision system consisting of computer vision cameras, that takes into consideration three primary criteria. First, the system must be capable of generating a 3D reconstruction of the manipulator workspace, enabling the motion planner to avoid obstacles and generate safe, collision-free paths. Second, the system must be able to localize a

set of known objects, such as tools, and guide the manipulator to grasp them. Third, the system must easily integrate with existing robotic platforms. Our vision system is composed of a vehicle chassis-mounted stereo camera pair with a fixed-baseline and a manipulator wrist-mounted fisheye camera (Fig. 3). The stereo pair observes the manipulator workspace, including part of the tool tray and the scene working area. The system uses the stereo to generate 3D point clouds of the workspace for scene reconstruction, assist with localizing tools in the tool tray, and visually track dynamic vehicle reference frames that are otherwise not observable (e.g., the position of the *NUI* HROV doors). The wrist-mounted fisheye camera provides a wide-angle view of the scene, and is used to detect objects and acquire dynamic viewpoints of the scene, which may be occluded or outside the field-of-view of the stereo pair. The wide field-of-view of the fisheye compared to a perspective camera enables clear views of objects and scene context at both close and far range (Fig. 4), which is advantageous for manipulation.

All three cameras are Blackfly model BFLY-PGE-50S5C-C (FLIR, Wilsonville, OR). The stereo cameras use the VS Technology SV-0614H 6 mm f/1.4 lens (VS Technology Corporation, Tokyo, JP), and the fisheye lens is the Fujinon FE185C086HA-1 2.7 mm f/1.8 (Fujinon, Tokyo, JP). The camera housings are custom fabricated with titanium shells and dome viewports (Sexton Corporation, Salem, OR), with a depth rating of 6000 m. A hardware trigger synchronizes the cameras. We calibrate the cameras using images of a checkerboard that the ROV manipulator moves throughout each camera’s field-of-view while the vehicle is submerged. We calibrate the stereo cameras using the ROS stereo camera calibration package. We calibrate the fisheye camera using the Kalibr toolbox (Kannala and Brandt, 2006). Because the usable field-of-view for the fisheye camera is less than  $180^\circ$  due to occlusions from the housing, we use the pinhole projection model with equidistant distortion. We verify that both the stereo and fisheye calibrations achieve sub-pixel reprojection errors for the checkerboard corners.

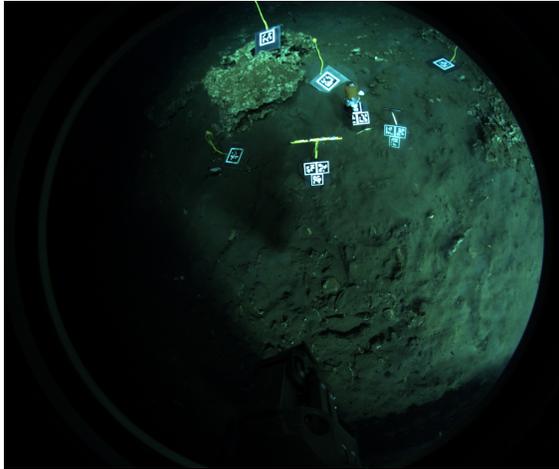
### 3.2.1 Tool Handle Pose Estimation

Tools carried by the ROV must be localized by the vision system before they can be grasped. It is general practice in ROV operations to use a single type of handle on every tool to provide consistency for ROV pilots. Given a known type of tool and its model, the vision system need only localize the handle for a tool to be grasped and manipulated.

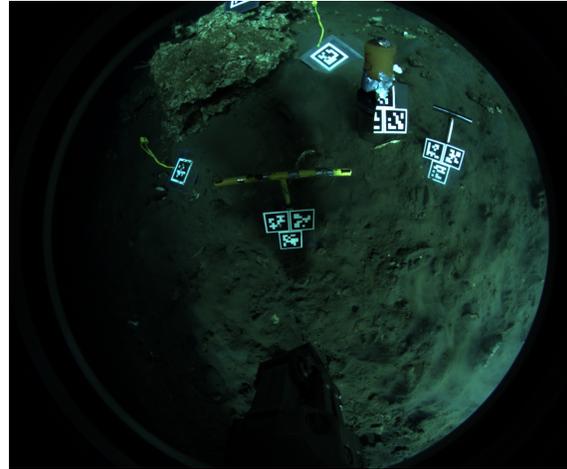
Using data collected with our vision system during the field trials, we developed a novel deep learning-based method, SilhoNet (Billings and Johnson-Roberson, 2019) and SilhoNet-Fisheye (Billings and Johnson-Roberson, 2020), that estimates the pose of tool handles detected from the wrist-mounted fisheye camera, without the need for fiducials. SilhoNet uses an intermediate silhouette representation to regress the detected object poses. This silhouette representation improves pose regression performance and facilitates training the network on synthetic data, which is especially beneficial when real training data is limited, as is the case for underwater environments. This method achieves promising results on the recorded datasets, but was not ready for integration with the system during the field trials.

During our field demonstrations, we relied on AprilTag markers (Olson, 2011) to localize the tool handles. Our choice of the AprilTag marker was motivated by the results of dos Santos Cesar et al. (2015), which show that AprilTags yield the best performance in underwater environments, with the lowest sensitivity to turbidity and variable lighting conditions in comparison to other popular fiducial markers. In this study, the minimum marker size detectable in an image was approximately 20 pixels, which, for the  $50 \times 50$  mm markers used in our system, equates to an expected maximum detection range of approximately 1.0 m for the fisheye camera and 2.4 m for the stereo cameras. These distances are within the typical working ranges of the manipulators used in our demonstrations. We designed 3D printed mounts that screw onto the t-handle bases and hold AprilTag vinyl stickers (Fig. 5, right).

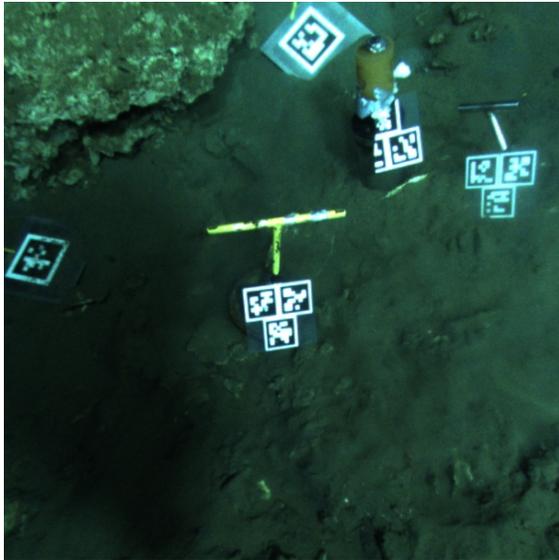
We use the ROS TagSLAM package (Pfrommer and Daniilidis, 2019) to detect the fiducials from the wrist-mounted fisheye camera. TagSLAM is built on the GTSAM (Dellaert, 2012) factor graph library and uses the ISAM2 (Kaess et al., 2012) incremental optimizer for efficient run-time performance. TagSLAM



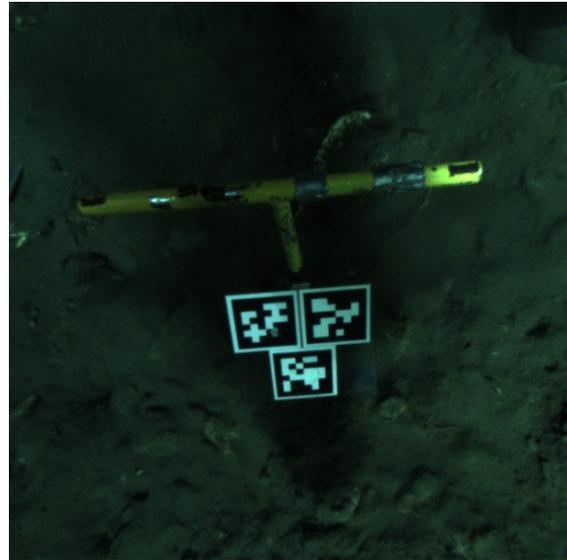
(a) Far fisheye view



(b) Close fisheye view



(c) Far perspective view



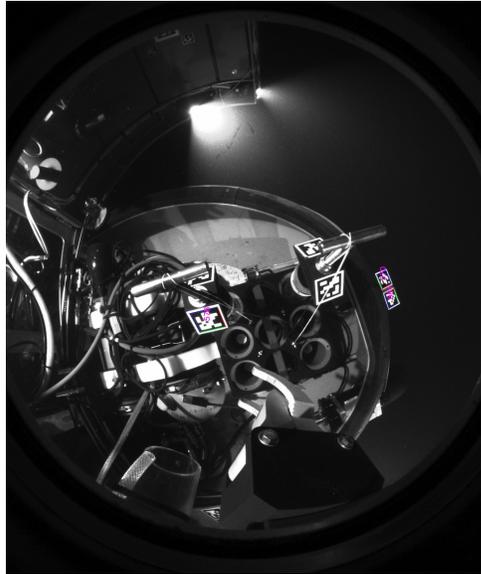
(d) Close perspective view

Figure 4: A comparison of (top) the full view of the wrist-mounted fisheye camera in an underwater scene at close and far range compared to (bottom) a  $60^\circ$  perspective rectification, which illustrates the significant increase in the field-of-view provided by a fisheye lens compared to a conventional perspective lens. This increased field-of-view provides significantly better contextual awareness to the vision and manipulation systems, especially when working at close range to the target, which is typical for manipulation tasks.

operates in a transform tree completely separate from the world planning environment. Within the TagSLAM environment, the fisheye camera is set as the origin, while the tools with the tag mounts are set as dynamic objects. We optimize the pose of each detected tool with respect to the fisheye camera frame using TagSLAM. The optimized tool pose with respect to the fisheye frame is projected into the world frame through the manipulator kinematics. If the fisheye camera loses sight of a tool, the tool pose within the world scene remains static until the tool is tracked again with TagSLAM.



(a) AprilTag mount for tools



(b) Fisheye view of tools in tool tray

Figure 5: A single type of t-handle was used to manipulate the different tools. The vision system localizes the t-handles using (a) AprilTags affixed to 3D-printed mounts located beneath the t-handle. These tags are detected in (b) images of the ROV tool tray from the wrist-mounted fisheye camera.

### 3.3 Control

While many existing methods tightly couple vehicle and manipulator motion planning and control, our approach decouples the manipulator and imaging system from other systems on the ROV. This makes it easier to integrate the system with different ROVs and also minimizes risk to the vehicle, as the automation system runs independently of the vehicle’s software stack. This approach also mimics standard ROV operation procedures, in which one pilot controls the vehicle while another pilot controls the manipulator. Our system seeks to replace the direct pilot control of the manipulator with a high-level automation interface that naturally integrates with standard ROV operational procedures. A current limitation of this control approach is a fixed-base assumption while the manipulator is activated. During manipulator operations, the ROV is assumed to be set down on the seabed and essentially acts as a fixed-base manipulator platform during a sampling tasks. When a manipulator command is executed, our system assumes that the scene state remains static until the activation is completed. This assumption of fixing the vehicle position before activating the manipulator follows the standard practice for operating work-class ROVs.

Figure 6 shows a diagram of our system architecture. We use the MoveIt! Motion Planning Framework (Coleman et al., 2014) to integrate the outputs of the perception system into the planning scene and to generate collision-free motion plans. MoveIt! directly supports a diverse set of state-of-the-art motion planners and inverse kinematic (IK) solvers. For this work, we used the RRT\* planner (Karaman and Frazzoli, 2011) with the KDL IK solver. We visualize the planning scene using RVIZ, with out-of-the-box integration with MoveIt!. We generated a kinematic description of the manipulator and vehicle from CAD models, and configured a motion planning environment with MoveIt!. A low-level driver for the manipulator exposes a position trajectory control interface to MoveIt! and interprets motion plans as command packets that it sends to the manipulator. Most work class hydraulic manipulators support only position setpoint commands. For this work, the system encodes the target joint positions and sends them directly to the manipulator valve controller.

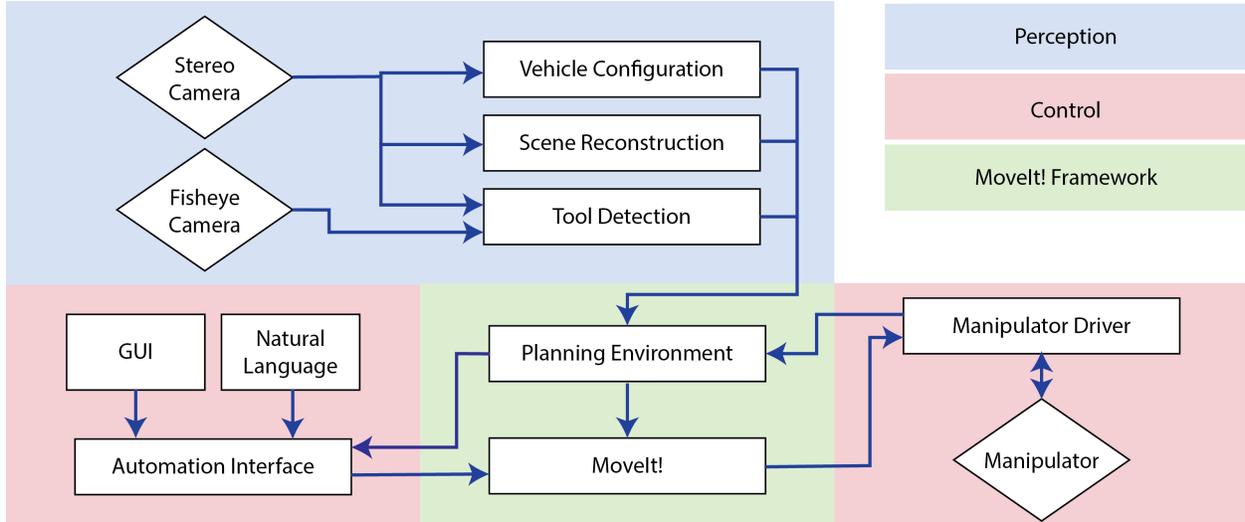


Figure 6: A diagram of the overall system, where rectangular blocks represent processes and diamond-shaped blocks represent hardware. Blocks in blue relate to perception. Blocks in red relate to (left) high- and (right) low-level control. Blocks in green are part of the MoveIt! framework around which our system is built. Our system uses the stereo camera to estimate the vehicle configuration (e.g., the pose of the doors on the *NUI* HROV), generate point clouds of the scene that can be fused to produce a 3D reconstruction of the scene, and assist with tool localization. The fisheye camera is used to localize tools and obtain dynamic viewpoints of the workspace. For low-level control, a driver implements a position-based trajectory controller, which integrates between MoveIt! and the manipulator valve controller. For high-level control, we implemented an automation interface to MoveIt! that supports high-level commands. In this work, we implement this interface using a graphical front-end as well as a preliminary demonstration using natural language.

### 3.3.1 Calibration Procedure

Figure 7 illustrates the coordinate frame transforms that must be calibrated for motion planning and kinematic-based control of the manipulator. The end-effector pose follows from the kinematic chain of transforms from the manipulator base frame through each consecutive link, where each transform is parameterized by the joint angle. Hydraulic manipulators generally provide limited joint feedback from position sensors like potentiometers or resolvers, which must be calibrated to the kinematic model. For this work, we assume a linear interpolation between the feedback values at the joint limits. We calibrate the hand-eye transform between the fisheye camera and the wrist link by detecting the fisheye-relative pose of an AprilTag positioned at a fixed location relative to the manipulator base. We perform these detections for a set of different kinematic configurations and then optimize the hand-eye transform using the ROS `easy_handeye` package (Tsai et al., 1989). When the manipulator base is rigidly fixed relative to the stereo pair, as would be the case for most ROV configurations, the transformation between the stereo pair and the manipulator base frame is calibrated by detecting the pose of a vehicle-affixed AprilTag in the scene in both the left stereo and the wrist-mounted fisheye cameras. We transform the pose of the fiducial from the fisheye camera frame to the manipulator base frame through the kinematic chain, giving the stereo-to-base frame transform as the difference in the tag pose between the two frames. For the *NUI* vehicle, where the stereo is not fixed relative to the manipulator, we used a different approach to estimate the stereo-to-base transform in real-time (see Section 4.3).

### 3.3.2 Pick-and-Place Interface

Our autonomy framework targets manipulation operations that involve pick-and-place tasks, such as taking a push-core sample. We implement a simple front-end interface (Fig. 8) that allows a user to step through

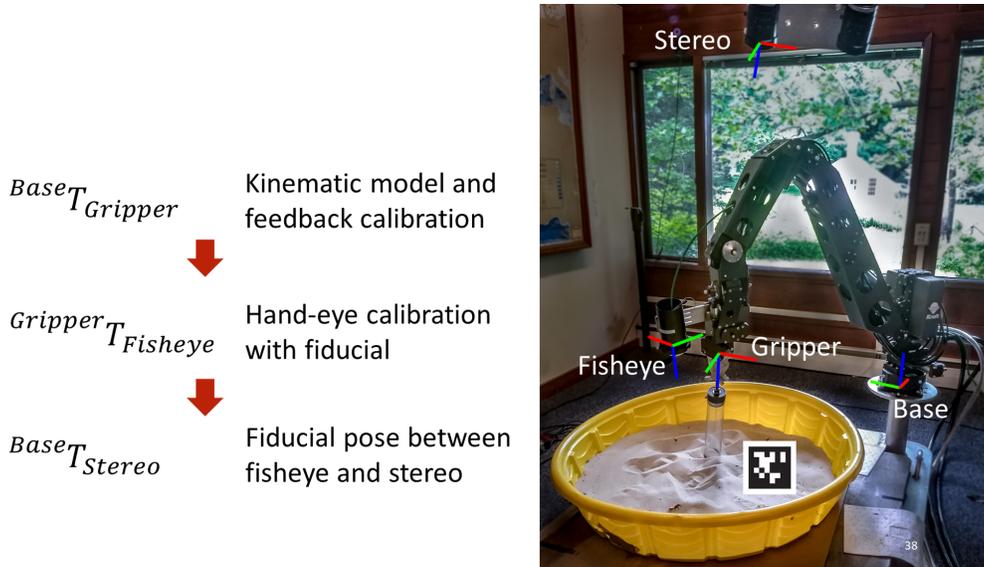


Figure 7: (right) An image of our testbed consisting of a Kraft TeleRobotics manipulator, a fisheye camera mounted to the end-effector, and an overhead stereo camera. Together with the manipulator base frame, there are four reference frames (left) which must be calibrated in order to fuse sensor data into a common reference frame and to plan the motion of the arm. Calibration is performed in the order shown on the left, where each transform enables calibrating the next in a bootstrapping manner. The fiducial in the image is included to indicate that AprilTags were placed statically in the workspace to obtain the Gripper-to-Fisheye and Base-to-Stereo calibrations.

a pick-and-place state machine that automates each step of the process, while maintaining a high level of safety through human oversight. The interface visualizes the manipulator motion plan at each step and only proceeds to execute the plan after the operator provides confirmation. The interface allows the user to select a target among a set of tools detected in the scene and then activate a sequence of automated steps to grasp and manipulate the target using pre-defined grasp points. An interactive marker enables the user to indicate the desired sample location in the 3D planning scene. Besides the pick-and-place state machine controller, the interface enables one-click planning of the manipulator to a set of pre-defined poses, immediate stopping of any manipulator motion, and opening and closing of the gripper. The MoveIt! planning environment also allows the operator to command the manipulator to an arbitrary configuration within the workspace through an interactive 3D visualization.

### 3.4 System Precision

The maximum precision of our system is limited by both kinematic and visual factors. The KRAFT manipulator uses 11 bit encoders, for an approximate per-joint angular resolution of  $0.176^\circ$ . When the arm is fully extended to 1.3m, the angular resolution for the shoulder joints equates to a metric arc length resolution of 4 mm at the end-effector. However, this estimate does not account for non-linear effects in the hydraulic actuators, bias in the joint actuation, inaccurate feedback from the joint sensors, or flexing of the arm’s mounting base/vehicle door, any of which may significantly reduce the kinematic accuracy of the system. The visual factors that limit precision include the accuracy of localizing the AprilTags from the fisheye camera and the resolution of the stereo reconstruction. Visual precision is dependent on the metric resolution of a pixel projected into the world. For a tag that is 1 m from the fisheye camera, the pixel metric resolution is 1.3mm, which is the expected best precision for localizing the tags. High distortion of tags near the edges of the fisheye image is expected and has been observed to reduce the localization accuracy. When processing the stereo images to produce depth maps, the maximum working distance can be tuned based on the maximum disparity over which a feature match is searched across a rectified image pair. In

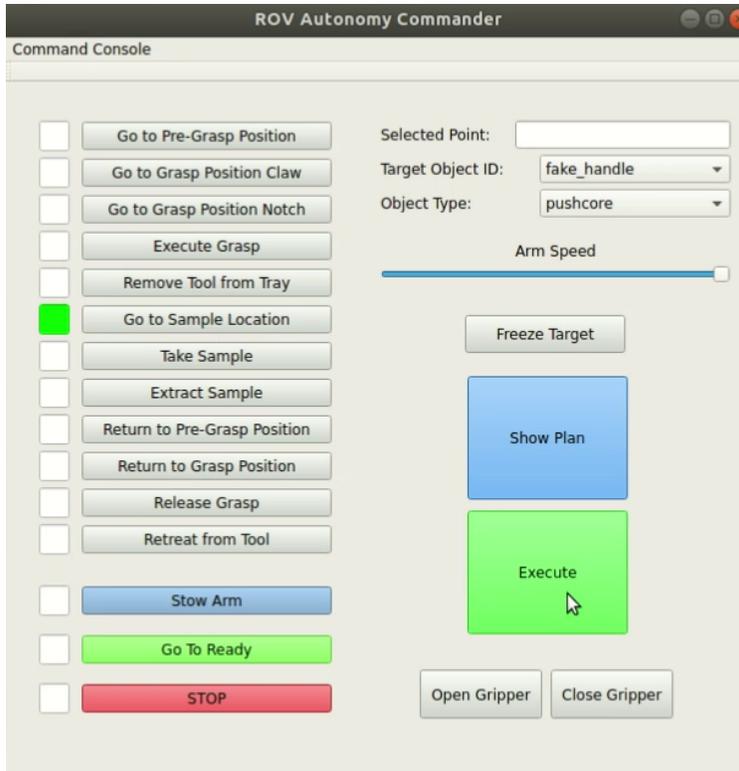


Figure 8: A simple interface to the automated system allows the user to configure and step through the automated pick-and-place pipeline. The motion plan for each step is visualized in the planning scene and is only executed upon confirmation by the user, which provides a high-level of safety for the system to be deployed on ocean-going systems.

our system, the maximum practical distance we target for stereo reconstruction is 3 m, which is well beyond the manipulator reach and beyond which lighting and haze effects severely degraded the image quality. For a viewing range of 3 m, the metric pixel resolution in the stereo view is 1.7 mm. Due to feature smoothing by the SGM correlation window, the actual reconstructed spatial resolution is coarser. In practice, we have observed that the kinematic accuracy is the limiting factor on the precision of our system, due to the many sources of kinematic error in hydraulic manipulator systems.

## 4 Experiments and Field Results

### 4.1 Automated Pick-and-Place Demonstration on Testbed

To prove the viability of our system before deploying it in the field, we demonstrated the full pick-and-place pipeline on a hardware testbed (Fig. 9) that mimics the configuration of the vision system and manipulator as they would be mounted on an ROV. The testbed includes a Kraft TeleRobotics manipulator identical to the one that we use for the field deployments with the *NUI* vehicle. The planning environment simulates the manipulator being mounted on the *NUI* HROV. The stereo point cloud is projected into the planning scene to inform placement of the sample marker. As described previously, we estimate the t-handle pose from the wrist-mounted fisheye camera by detecting the AprilTags mounted below the handle. We executed each step of the automated pick-and-place interface successfully, with no manual control input. The system grasped the t-handle based on the detected pose, and the planner found and executed a manipulation path to the sample location marker, which was placed at a non-trivial angle, touching a rock in the scene. The rock was

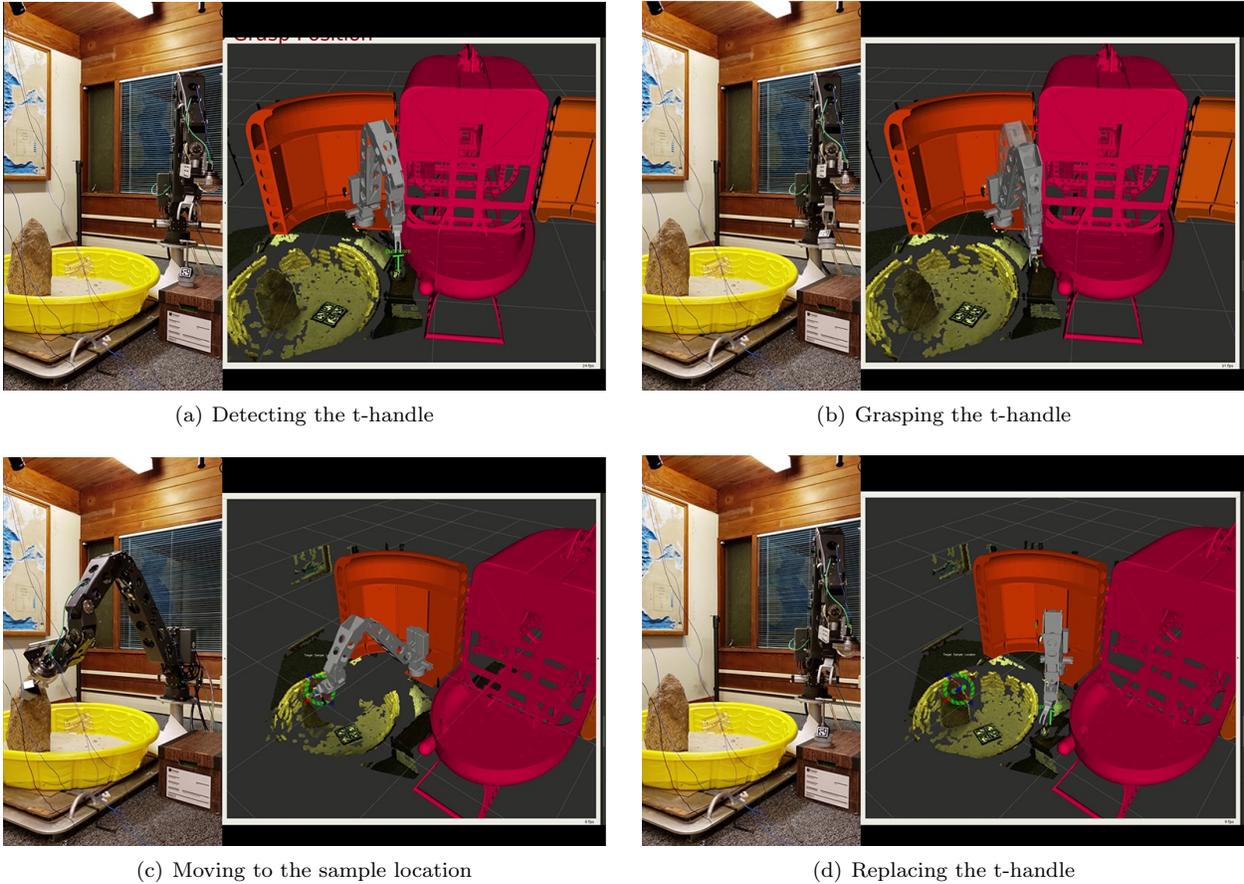


Figure 9: We demonstrated fully autonomous pick-and-place with a t-handle on a testbed with the same camera and manipulator hardware used on the *NUI HROV*. First, (a) the t-handle was detected from the fisheye camera using the AprilTags, and the handle pose was projected into the planning scene. Next, (b) the manipulator was commanded to grasp the t-handle via the autonomy interface. Subsequently, (c) a sample location was set in the planning scene with an interactive marker based on the projected stereo point cloud, the manipulator planned a motion to reach the sample location, and executed the plan after the user verified it. The manipulator was then (d) commanded to return the t-handle to the location where it was first grasped. The rock in the environment was placed in a delicate balance on its end, yet the manipulator was controlled with enough precision to bring the tool into direct contact without knocking it over.

placed on its end in a delicately balanced position, and the manipulator was able to bring the tool into contact with the rock with enough precision that the rock remained standing. The tool was then returned to the position from where it was grasped. This full experiment was repeated multiple times, though not without some grasp failures, due to noise in the visually estimated pose of the t-handle. However, the interface made it easy to recover from any failed step of the state-machine without ever requiring the operator take manual control of the manipulator.

#### 4.2 Real-Time Scene Reconstruction and Data Collection at the Costa Rican Pacific Shelf Margin

Demonstrations at the Pacific continental margin were conducted during a two-week research cruise aboard the R/V *Falkor* using the *SuBastian* ROV. The automated manipulation component of this expedition focused on a demonstration of the vision system and data collection to aid the development of visual methods.

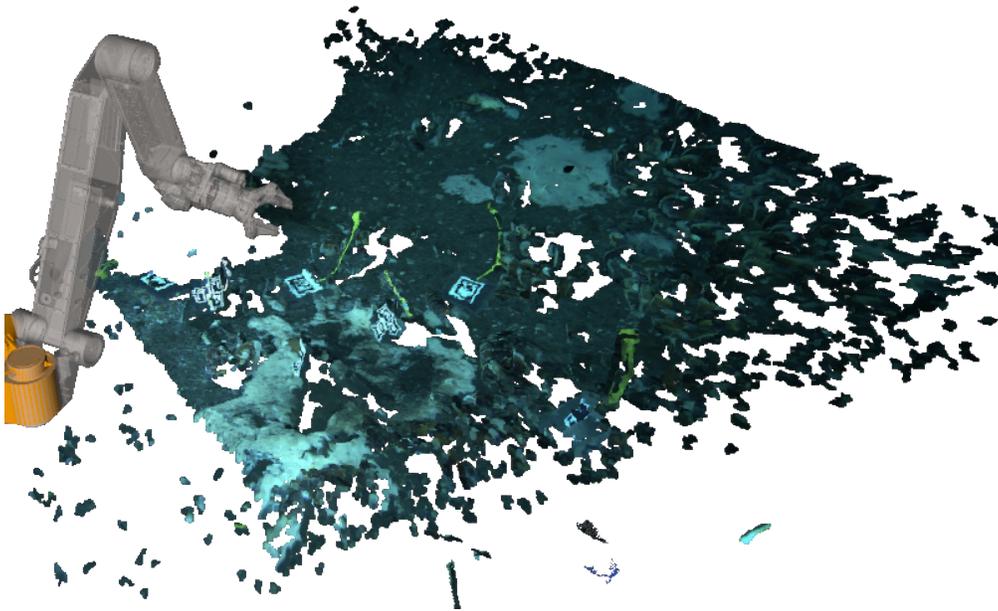


Figure 10: The vision system was integrated on the *SuBastian* ROV operated by Schmidt Ocean Institute, where we demonstrated real-time visualization of the planning scene with a Schilling Titan-4 manipulator and the projected stereo point clouds. This also demonstrates the flexibility of the system to be integrated with different vehicles and manipulators.

The integration time of our system took approximately two days during cruise mobilization, highlighting the relative ease and flexibility with which the system can be implemented on a variety of vehicles and manipulators. Camera image data was streamed over a GigE interface at 3 Hz to a topside workstation, which handled all processing and visualization. Joint encoder feedback from the manipulator was obtained by passively monitoring the serial communication between the manipulator and the ship’s control computer. We visualized the real-time configuration of the manipulator in the 3D planning environment with the stereo point clouds projected into the scene. The point clouds were generated from the stereo imagery using the standard semi-global matching (SGM) method built into the ROS image processing pipeline, and the parameters were hand-tuned to achieve the best results. Figure 10 shows a frame from the real-time visualization captured on the seafloor during one of the dives. A good camera calibration combined with high water clarity, rich seafloor texture, and evenly distributed scene lighting resulted in high quality point clouds. These early results demonstrated the effectiveness of the vision system to capture the 3D structure of the workspace and the ability to fuse the information into a real-time scene representation that is useful for both manipulation planning and 3D visualization of the ROV configuration and planning environment.

During this expedition we collected an extensive dataset (Billings and Johnson-Roberson, 2020) of synchronized stereo and wrist mounted fisheye images along with the manipulator joint feedback from a diverse set of seafloor environments (Fig. 11). AprilTags mounted on plates were dispersed into the scenes to provide ground truth for the camera poses, and three different types of graspable handle objects were also randomly placed into the scenes. This dataset supported the development of our visual methods and is also intended to serve the underwater research community for the development of scene reconstruction, object detection, and pose estimation methods that work robustly in real seafloor environments.

The fisheye images were processed into a standalone dataset with annotated 2D bounding boxes and 6D poses for the handle objects visible in each frame. This dataset was released with the SilhoNet-Fisheye publication (Billings and Johnson-Roberson, 2020). Figure 12 shows sample images from this dataset. The

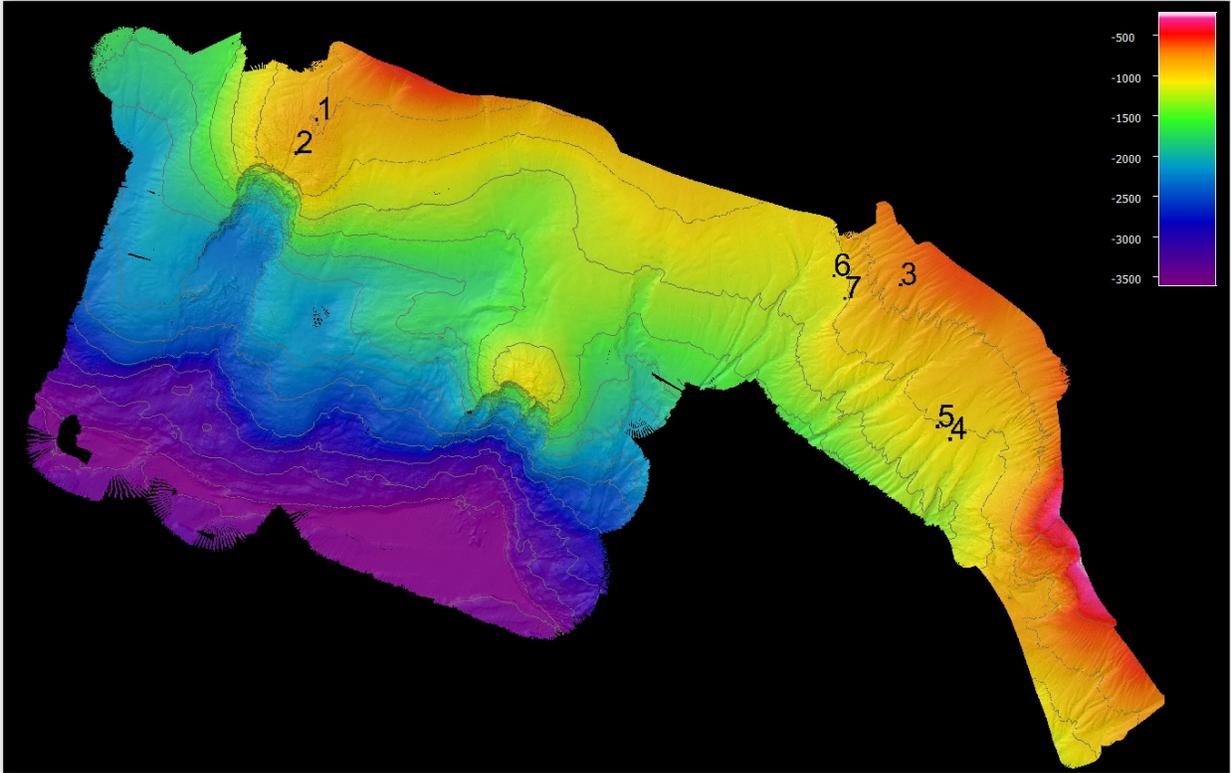


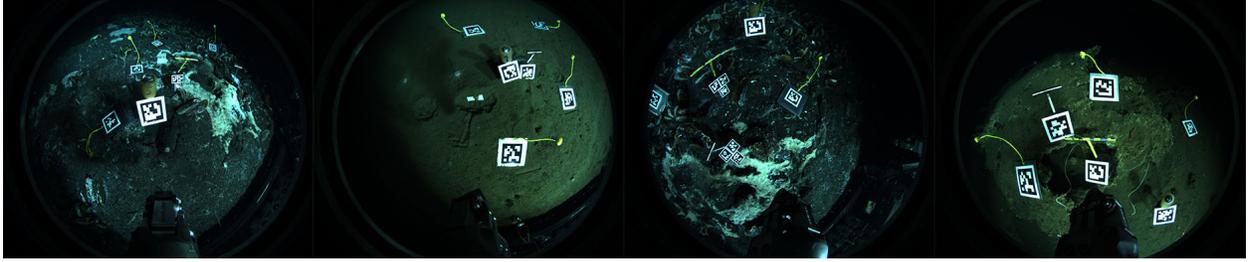
Figure 11: Bathymetric map of the survey area from the 2018 cruise on the Pacific continental margin showing data collection locations at seven different science goal sites, spanning over 62 km (linear distance between Locations 1 and 4) and ranging in depth from 600 m to 1100 m. Depth contours are spaced at 250 m intervals and the map is oriented with North up.

combined dataset of stereo and fisheye imagery with synchronized manipulator joint feedback supported our development of visual methods for scene reconstruction.

### 4.3 Automated Sample Collection and Return within Active Submarine Volcanoes

For exploration of the Kolumbo and Santorini calderas, *NUI*'s manipulator was mounted to the starboard door and the stereo cameras were mounted to the port door (Fig. 14). Having the manipulator and stereo camera on opposite articulating doors allowed for flexibility in configuring the position of the arm according to the specific manipulation task and enabled on-the-fly adjustment of the manipulator and stereo positions separately. Unfortunately, the doors are actuated using hydraulic rams that lack position feedback. For safe motion planning, it was necessary to estimate the door positions in real-time. The estimated door positions were used to update the kinematic configuration of the vehicle in the planning scene. However, we observed that the doors could flex, introducing some error in the kinematic estimates that negatively impacted the accuracy of the stereo point cloud projection into the planning scene. To minimize accumulated error in the transform between the stereo camera frame and the manipulator base frame, the stereo frame was referenced directly to the base frame in the ROS transform tree. The base frame was accurately localised directly from the stereo camera through detection of tags fixed to the manipulator base.

To estimate the door positions in real-time, we affixed AprilTags to the starboard door and to the bow of the vehicle's payload bay (Fig. 15 (left)). The tags on the vehicle bow were mounted at a measured location relative to the vehicle reference frame, with the reference tag's  $Z$ -axis aligned with the  $Z$ -axis of the vehicle reference frame. The door joint axes of rotation were also aligned to the  $Z$ -axis of the vehicle frame,



(a) Sample raw fisheye images from different sequences of the dataset



(b) Sample annotations from a single sequence of the dataset

Figure 12: The fisheye imagery collected during the Costa Rica cruise was processed into a stand-alone dataset (Billings and Johnson-Roberson, 2020). The images are annotated with the bounding box and six-DoF pose of the tool handles placed in the workspace. The top row (a) shows sample raw fisheye images from different sequences of the dataset, and the bottom row (b) shows sample annotations from a single sequence in the dataset. The images are center rectified here only for purposes of visualization.

enabling a simple trigonometric calculation of the door angles based on the relative tag locations in the  $X$ - $Y$  plane. We used the left stereo camera to track the relative pose of the AprilTags and used these estimates as observations in AprilTag-based visual SLAM (Pfrommer and Daniilidis, 2019) (Fig. 15). Figure 16 shows a schematic of the vehicle and visual SLAM system with the relevant transforms in the  $X$ - $Y$  plane used to calculate the door angles. The visual SLAM provided the relative translations between the vehicle tag frame and the starboard tag frame,  $T_{vs}$ , and between the vehicle tag frame and the stereo camera frame,  $T_{vp}$ . Given that the translations between the vehicle tag and the door joint frames,  $T_{os}$  and  $T_{op}$ , were measured and known, the angle of the starboard and port doors,  $\theta_s$  and  $\theta_p$  respectively, were recovered as

$$\theta_s = \arctan \frac{T_{s,y}}{T_{s,x}} - \theta_{s_0} \quad (1a)$$

$$\theta_p = \arctan \frac{T_{p,y}}{T_{p,x}} - \theta_{p_0}, \quad (1b)$$

where

$$T_s = T_{vs} - T_{os} \quad (2a)$$

$$T_p = T_{vp} - T_{op}, \quad (2b)$$

where the  $x$  and  $y$  subscripts indicate the corresponding component of the translation vector, and  $\theta_{s_0}$  and  $\theta_{p_0}$  are the measured angle offsets.

#### 4.3.1 Planner Controlled Biological Sample Collection

The Kolumbo-Santorini expedition resulted in several scientific achievements, including verification of the persistence of Kalliste Limnes (Camilli et al., 2015), 3D reconstruction of extremophile habitats within the

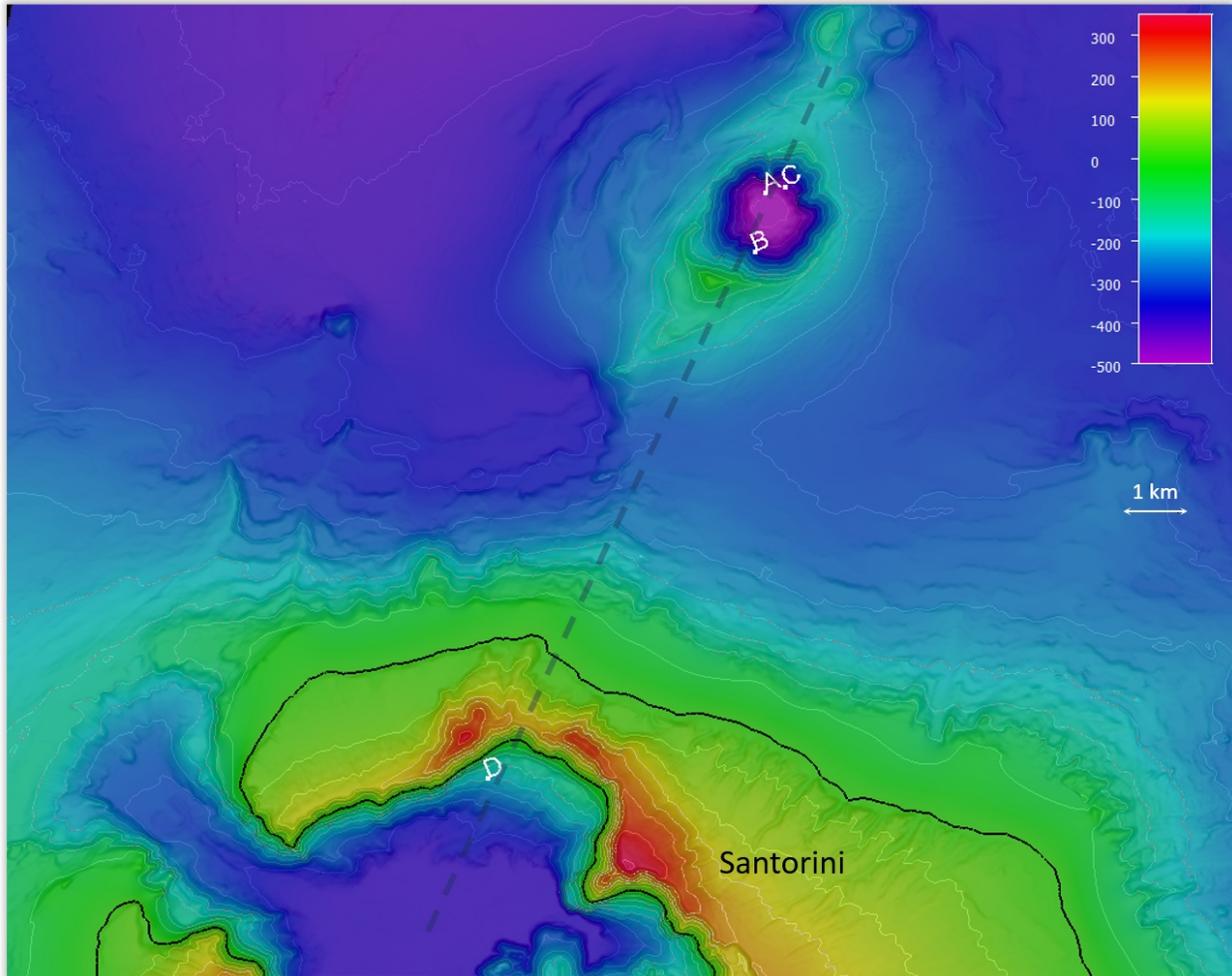


Figure 13: Map of automated sample collection locations, with regional bathymetry adapted from Nomikou et al. (2012, 2013). The sea level contour is indicated in black. The dashed line indicates the Christiania-Santorini-Kolumbo tectonic line (Nomikou et al., 2012). Locations marked A, B, and D indicate automated sample collection and return sites, and location C indicates the site where a natural language proof-of-concept demonstration was conducted. Sampling depths ranged from 240 m to 501 m

calderas’ craters, and sampling of benthic fluids, seafloor sediments, and biological materials. One of the most useful subsea tools for sample collection and return is a “slurp gun” vacuum sampler. For these operations, the slurp nozzle is in close proximity to the sample of interest and a vacuum pump sucks the sample through the hose into a collection chamber. To demonstrate automated slurp collection, we attached the slurp hose to the side of the manipulator, so that the end-effector could be commanded to the desired location to collect the slurp sample. We completed multiple successful sample collections, including that shown in Figure 17, where a slurp sample of a sediment microbial mat was collected using the planner interface to command the manipulator to the desired sample location, after which the manipulator was returned to the home position.

#### 4.3.2 Natural Language Control

Subsea ROV missions require close collaboration between the ROV pilots and scientists. The primary means by which pilots and scientists communicate is through spoken language—scientists use natural language to convey specific mission objectives to ROV pilots (e.g., requesting that a sample be taken from a particular location), while the pilots engage in dialogue to coordinate their efforts. Natural language provides a flexible,



Figure 14: The *NUI* vehicle is outfitted with clam shell doors that can be closed to reduce drag when cruising and opened to perform manipulation tasks. The manipulator is mounted to the starboard door and the stereo cameras are mounted to the port door.

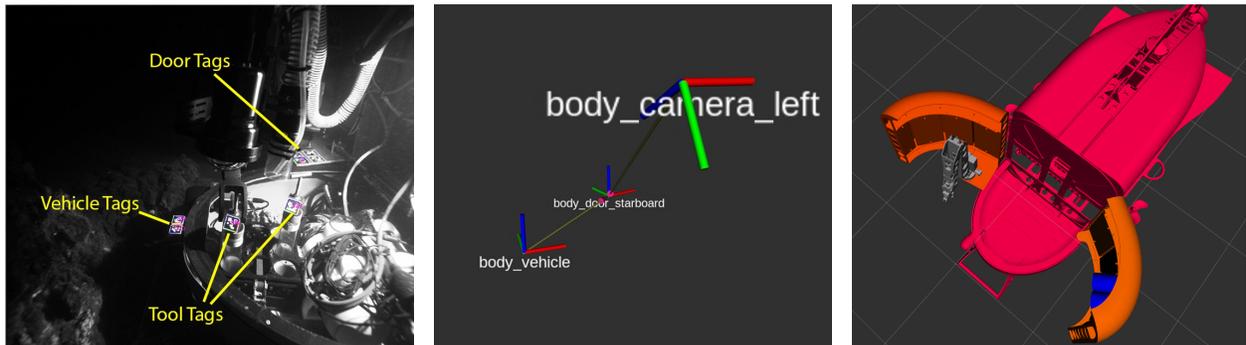


Figure 15: Fiducial-based visual SLAM from the left stereo camera was used to estimate the door angles in real-time using (left) tags mounted to the front of the vehicle frame and at the base of the manipulator on the starboard door. SLAM provided estimates of (middle) the relative transformations between the camera and the tag frames that were used (right) to estimate the door angles and update the vehicle model in the planning scene. The left stereo camera was also used in conjunction with the wrist mounted fisheye for (left) fiducial-based localization of tools.

efficient, and intuitive means for people to interact with our automated manipulation framework. The inclusion of a natural language interface would support our goal to realize a framework that can be integrated seamlessly with standard ROV operating practices and may eventually mitigate the need for a second pilot.

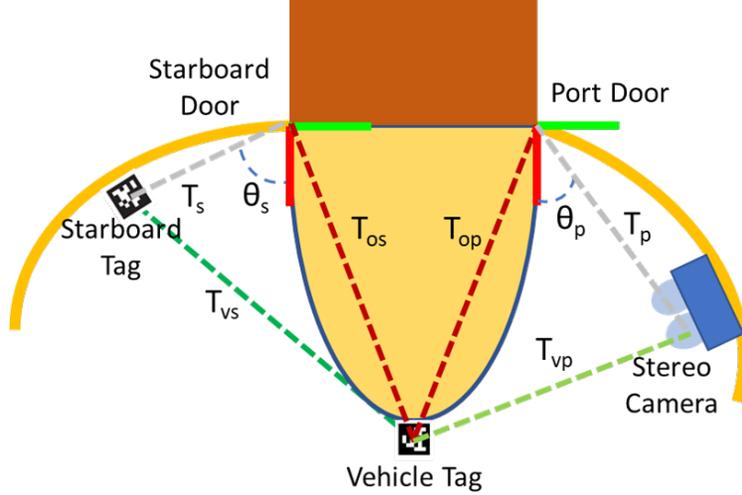


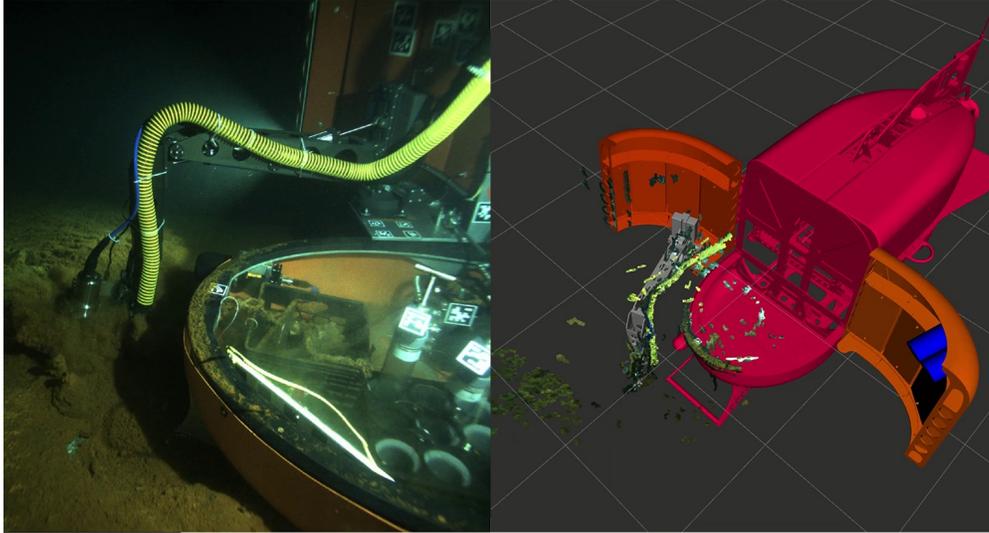
Figure 16: A 2D schematic of the *NUI HROV* that relates the visual SLAM from the left stereo camera to the door positions. The green dashed lines represent transformations estimated from SLAM. The red dashed lines denote known transformations computed from the vehicle kinematic model and the measured position of the tags. The grey dashed lines represent the calculated transforms with respect to each door reference frame, which have a trigonometric relation to the door angles,  $\theta_s$  and  $\theta_p$ .

Using the *NUI HROV*, we performed a proof-of-concept demonstration of an architecture that allows user control of an ROV manipulator using natural language provided as text or speech using a cloud-based speech recognizer. We frame natural language understanding as a symbol grounding problem (Harnad, 1990), whereby the objective is to map words in the utterance to their corresponding referents in a symbolic representation of the robot’s state and action spaces. Consistent with contemporary approaches to language understanding, we formulate grounding as probabilistic inference over a learned distribution that models this mapping. In particular, given the syntactic parse of a natural language command  $\Lambda$ , we employ maximum a posteriori inference over the power set of referent symbols  $\mathcal{P}(\Gamma)$

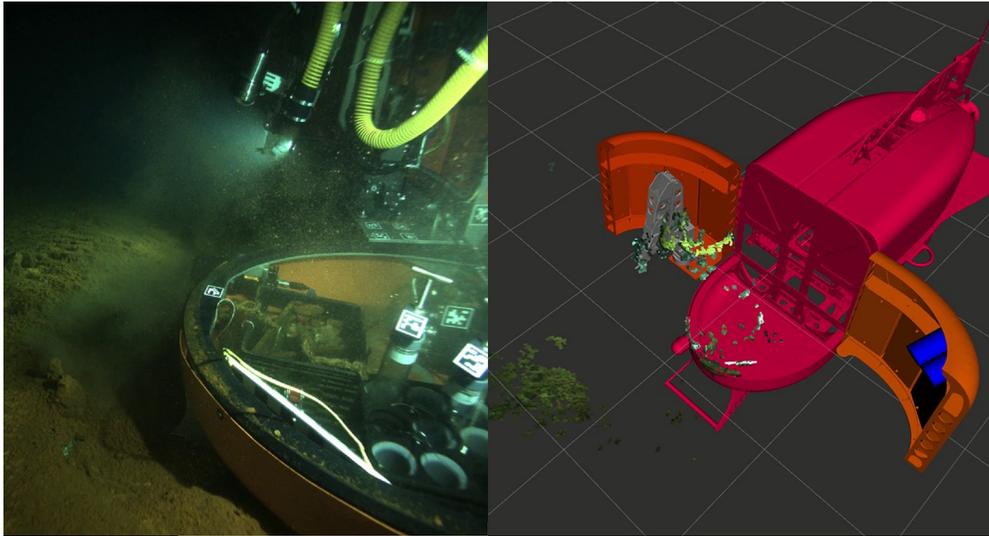
$$\Gamma^* = \arg \max_{\mathcal{P}} p(\Gamma|\Lambda, S). \quad (3)$$

where  $S$  is a variable that denotes the robot’s model of the environment (e.g., the type and location of different tools). We model this distribution using the Distributed Correspondence Graph (DCG) Howard et al. (2014), a factor graph (Fig. 18) that approximates the conditional probabilities of a Boolean correspondence variable  $\phi_{i_j}$  that indicates the association between a specific symbol  $\gamma_{i_j} \in \Gamma$ , which may correspond to an object, action, or location, and each word  $\lambda_i \in \Lambda$ . Critically, the composition of the DCG factor graph follows the hierarchical structure of language. The model is trained on corpora of annotated examples (i.e., words from natural language utterances paired with their corresponding groundings), whereby we independently learn the conditional probabilities for the different language elements, such as nouns (e.g., “the pushcore”, “tool”, and “tool tray”), verbs (e.g., “retrieve”, “release”, and “stow”), and prepositions (e.g., “inside” and “towards”). Together with the fact that the factor graph exploits the compositional nature of language, the DCG model is able to generalize beyond the specific utterances present in the training data.

For our initial implementation, the space of symbols  $\Gamma$  included the tools that the arm was able to grasp and the different steps that comprised the state machine underlying the pick-and-place pipeline. Figure 19 presents an example from a deployment at the Kolumbo caldera in which natural language was used to initiate path planning to the sample location and then to command the manipulator to execute the planned path. Several tests were conducted in which the manipulator was commanded through natural language input to move to a location specified by the sample marker in the planning interface and then return to the home position. These tests demonstrated the flexibility of our system to incorporate different operational modalities through high level abstraction.



(a) Taking slurp sample



(b) Returning to home position

Figure 17: An example of a successful planner controlled slurp collection of a bacterial mat, with the yellow slurp hose attached to the manipulator. The manipulator was (a) commanded to the desired slurp location through the automated planning interface and then (b) directed to return to its home position following the slurp collection.

#### 4.4 Performance Analysis

We evaluated the overall accuracy of the calibrated kinematic and visual system on the testbed. For this evaluation, we placed an AprilTag grid in the scene and activated every joint of the manipulator while keeping the tag grid in view of the fisheye camera. We used TagSLAM Pfrommer and Daniilidis (2019) to generate a visual SLAM estimated trajectory of the fisheye camera, and we used the manipulator joint feedback to also generate a kinematic based trajectory. These trajectories are plotted against each other in figure 20. The overall mean error between the kinematic and visual based trajectories is 1.16cm, the maximum trajectory error is 3.27cm, and the standard deviation is 0.65cm. These results are a conservative estimate of the system calibration accuracy as there are several sources of error: the visual SLAM accuracy degrades when the tags

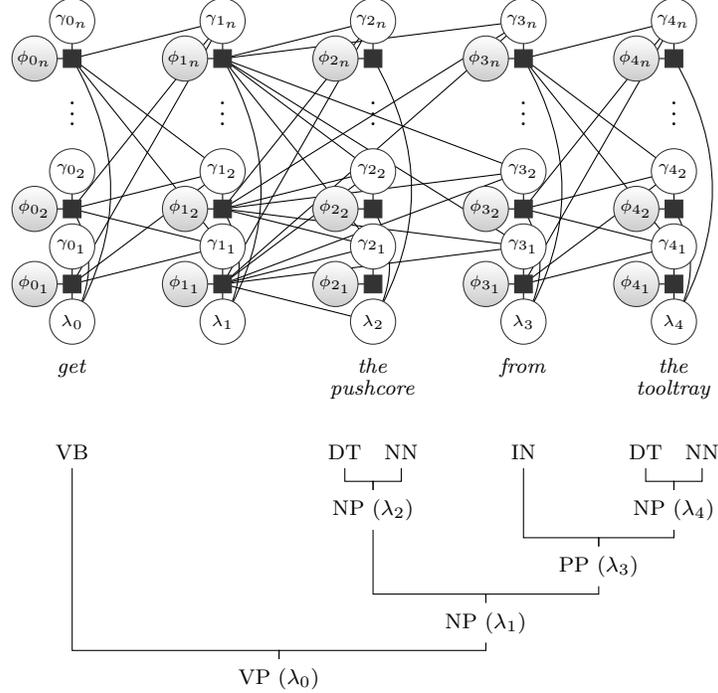
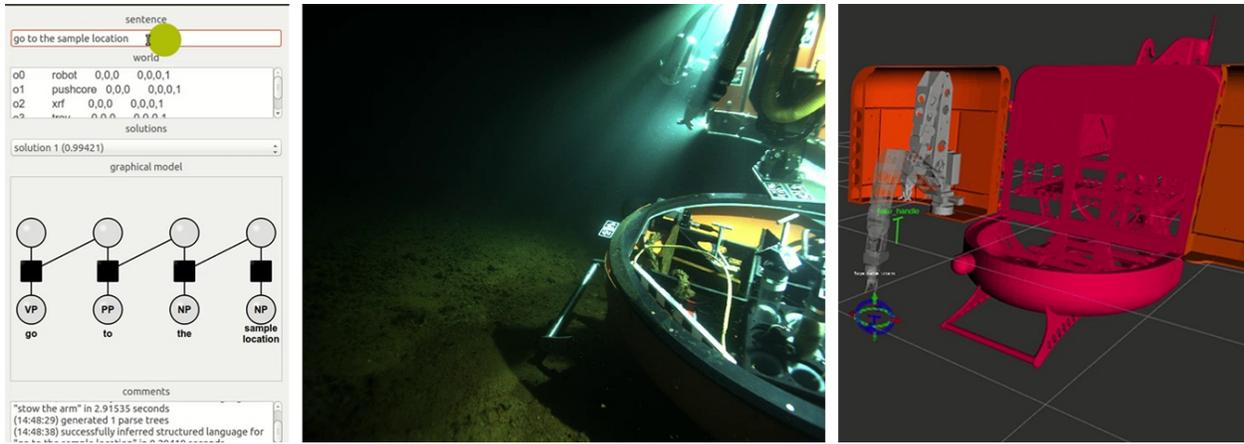


Figure 18: A visualization of (top) the DCG factor graph for the expression “get the pushcore from the tooltray” aligned with (bottom) the associated parse tree. Shaded nodes denote observed random variables, while those rendered in white are latent.

are near the edge of the fisheye image; the agreement between the kinematic and visual based pose of the fisheye camera depends on the accuracy of the hand-eye calibration; the joint feedback and fisheye images are not synchronized; and the SLAM and kinematic reference frames were mapped to each other through a single fisheye frame estimate of the tag grid pose, projected from the fisheye frame through the kinematic chain to the manipulator base frame. However, we have demonstrated in our experimental trials that the system accuracy is good enough to perform high level automation tasks.

The first five joints of the KRAFT manipulator are controlled through joint position set-point commands. We analyzed the manipulator control response for bias or hysteresis, as these are known issues with hydraulic actuation. Figure 21 shows plots of the commanded versus feedback positions during actuation of each joint of the testbed manipulator. The figure also shows a histogram of errors, binned at  $0.5^\circ$ , between the commanded and followed joint trajectories. All of the joints except the wrist pitch exhibit small bias and no major hysteresis is evident. The wrist pitch exhibits a bias of approximately  $1.5^\circ$ , which is significant, but did not prevent completion of high level automation tasks. Figure 22 shows the same plots for the NUI HROV manipulator made from data recorded during the field trials in Greece. The elbow and wrist joints exhibit little bias or hysteresis. However, both of the shoulder joints exhibited high bias, particularly the shoulder yaw joint, which had a bias of approximately  $8^\circ$ . This high bias prevented the completion of a pick-and-place manipulation task during the field trial. Our current control system relies on the manipulator valve controller to move to the desired set-point and does not account for bias in the control response. It will be critical in the future to incorporate an adaptive controller into the system that can account for bias and hysteresis in the hydraulic actuators.



(a) Language commands the system to plan a path to the sample location



(b) Language command to execute planned path

Figure 19: Demonstration of a proof-of-concept framework that enabled operators to interact with our autonomous manipulation architecture using natural language. Given input in the form of free-form text, either entered by the operator or output by a cloud-based speech recognizer, we (left) infer the meaning of the command using a probabilistic language model. (a) In the case of the command to “go to the sample location”, our system (top-right) determines the goal configuration and solves for a collision-free path in configuration space. (b) Given the command to “execute now”, the manipulator then (bottom-right) executes the planned path to the goal.

## 5 Discussion and Future Work

During the course of our field trials, we identified operational challenges and failure modes for both the manipulator and vision systems. Addressing these issues is necessary to improve the robustness of the system and is the objective of ongoing research.

Underwater hydraulic manipulators have inherent characteristics which make them especially challenging to automate and can lead to mission failure if they are not accommodated by the planning and control systems. We identify three particular challenges. The first challenge is the sensors that provide joint position feedback (e.g. potentiometers or encoders) can be noisy and prone to drift, resulting in an inaccurate estimate of the manipulator configuration, which can lead to self-collisions or collision with the vehicle or obstacles in the environment. This issue could be mitigated by continuously calibrating the arm using the vision system to detect and compensate for proprioceptive sensor drift. Such a fully automated kinematic calibration

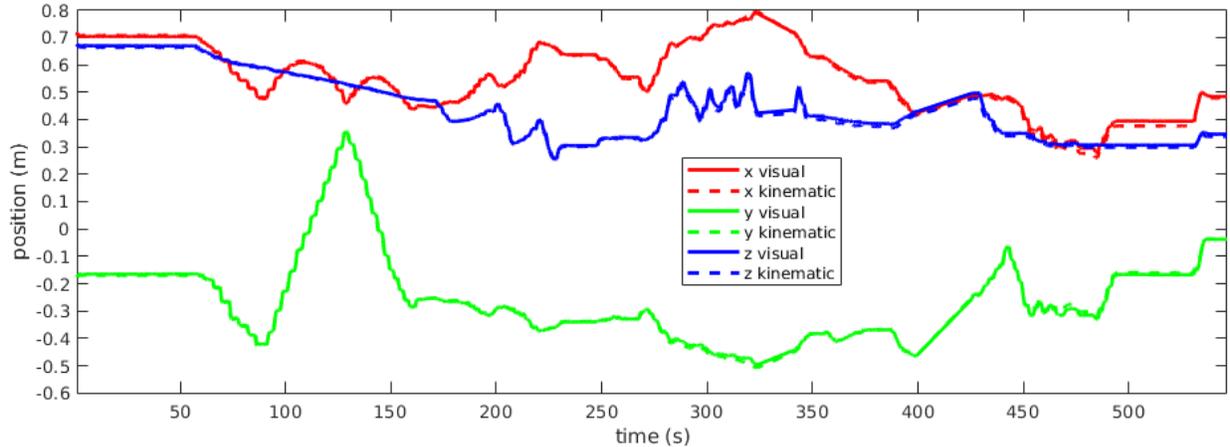
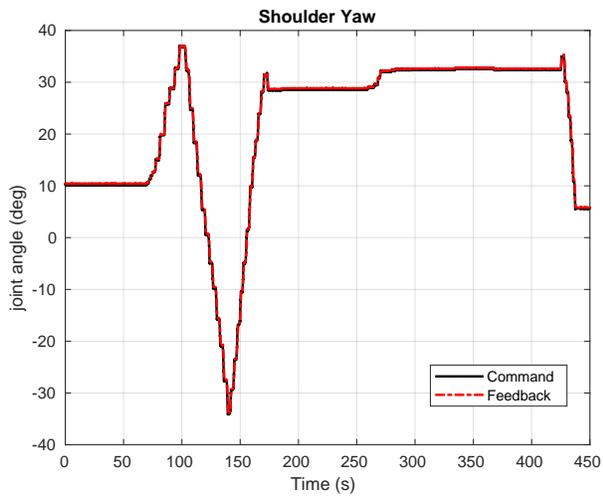


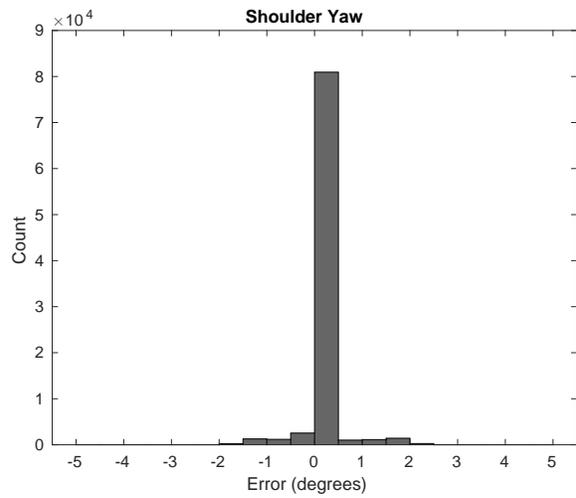
Figure 20: Plot of the TagSLAM estimated trajectory (visual) of the fisheye camera versus the trajectory estimated from the manipulator joint feedback (kinematic). The trajectory is plotted separately for each coordinate axis with respect to the manipulator base frame.

procedure is also a practical necessity for a system to be deployed on a space flight mission and would improve calibration accuracy over the manual procedure used in this report. Our ongoing work seeks to apply a feature-based mapping/structure-from-motion framework that jointly performs scene reconstruction and kinematic calibration of the manipulator using features from the fisheye camera. The second challenge is that hydraulic actuators can be imprecise. Typical hydraulic actuator characteristics include a bias between the commanded and reached joint positions, which we observed in the KRAFT manipulator, and hysteresis, where the offset between commanded and reached positions is variable with the direction of joint actuation and the position of the joint. These actuator effects could be mitigated through an adaptive control strategy that adjusts the joint commands to account for detected anomalies or offsets between the commanded and reached configurations. Sivčev et al. (2018b) reported hysteresis as high as  $1.5^\circ$  in a Schilling Titan 2 manipulator and subsequently learned joint command offsets in a calibration procedure to compensate for it. The third challenge is that complete joint failure is common for underwater manipulators, reducing the degrees-of-freedom by at least one. Mitigation of this failure would require planning level adaptation to determine what manipulation tasks are still feasible. In this under-actuated operational state, the vehicle mobility might be considered within the kinematic planning to compensate for the loss of manipulator dexterity, drawing from the prior work on free-floating intervention.

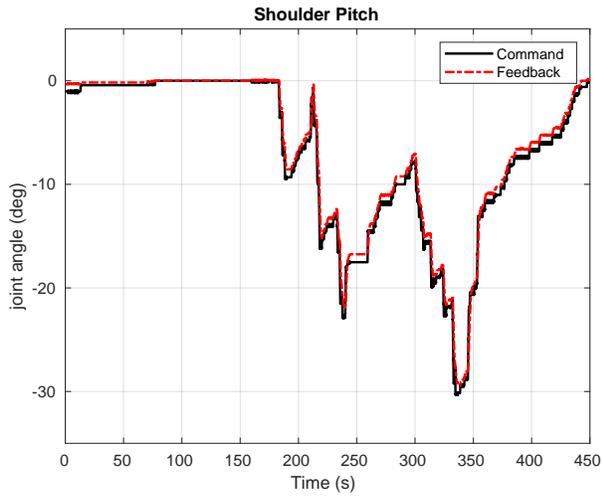
Existing visual reconstruction methods are typically sensitive to lighting, image contrast, and the presence of texture, all of which are highly variable in underwater environments. Figure 23 compares point clouds generated using a standard SGM method from the same stereo camera under two different visual conditions. Under near-ideal conditions that include clear water, uniform illumination, and a richly textured seafloor as was the case during our Costa Rica expedition, the point cloud is highly detailed and exhibits a low amount of noise, resulting in a reconstruction that captures fine details of the scene. During the Kolumbo caldera operations, however, fine-grain unconsolidated sediments and amorphous microbial mats blanketed the seafloor, providing little texture for stereo matching. The illumination was uneven and particulates suspended in the water column caused turbidity and light scattering effects that degraded the quality of the images. Under these conditions, stereo matching is only able to recover the well defined edges of the vehicle, while very little of the seafloor is reconstructed. While these examples represent different extremes in underwater visual conditions, it is critical to develop scene reconstruction algorithms that can operate reliably across this range of conditions to achieve robust autonomy. We are currently investigating information-theoretic ways to exploit our ability to control the pose of the wrist-mounted fisheye camera and an adjacent light source to acquire targeted views and actively illuminate the scene in order to improve and extend reconstruction under both good and degraded visual conditions. Our system implementation currently assumes that the scene is semi-static, i.e., that the ROV position and scene state remain constant



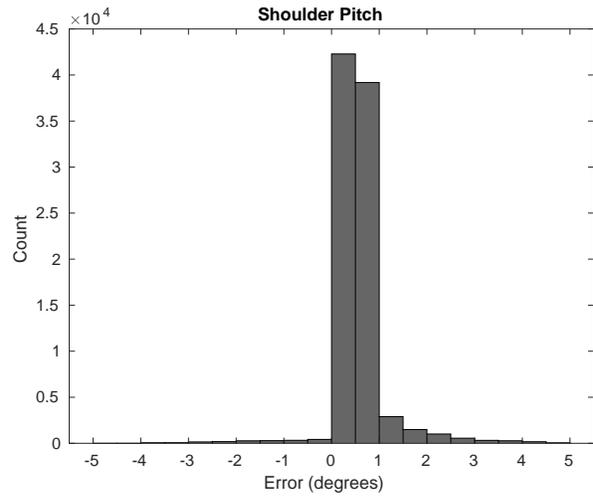
(a) Shoulder Yaw Trajectory



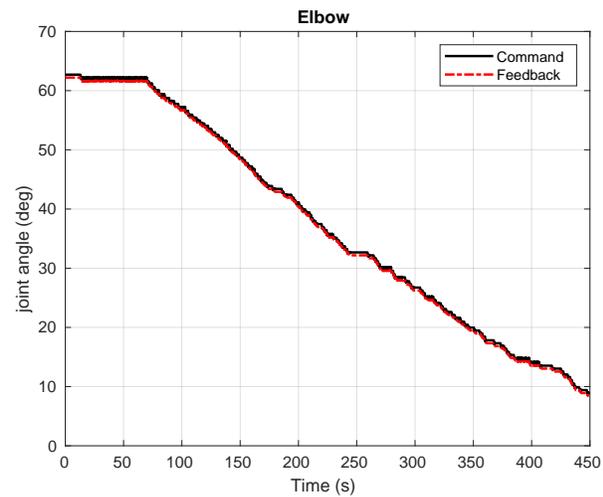
(b) Shoulder Yaw Error Histogram



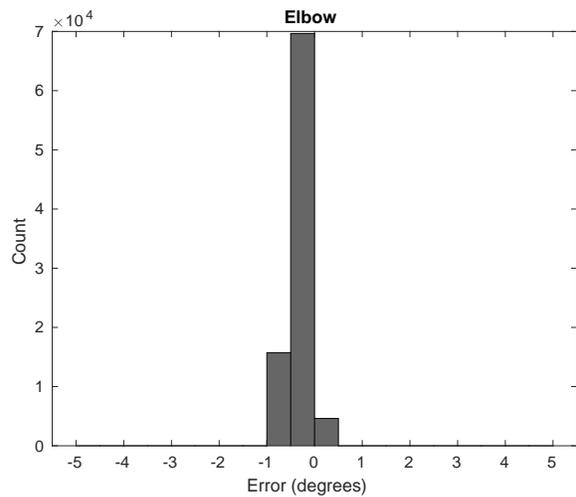
(c) Shoulder Pitch Trajectory



(d) Shoulder Pitch Error Histogram



(e) Elbow Trajectory



(f) Elbow Error Histogram

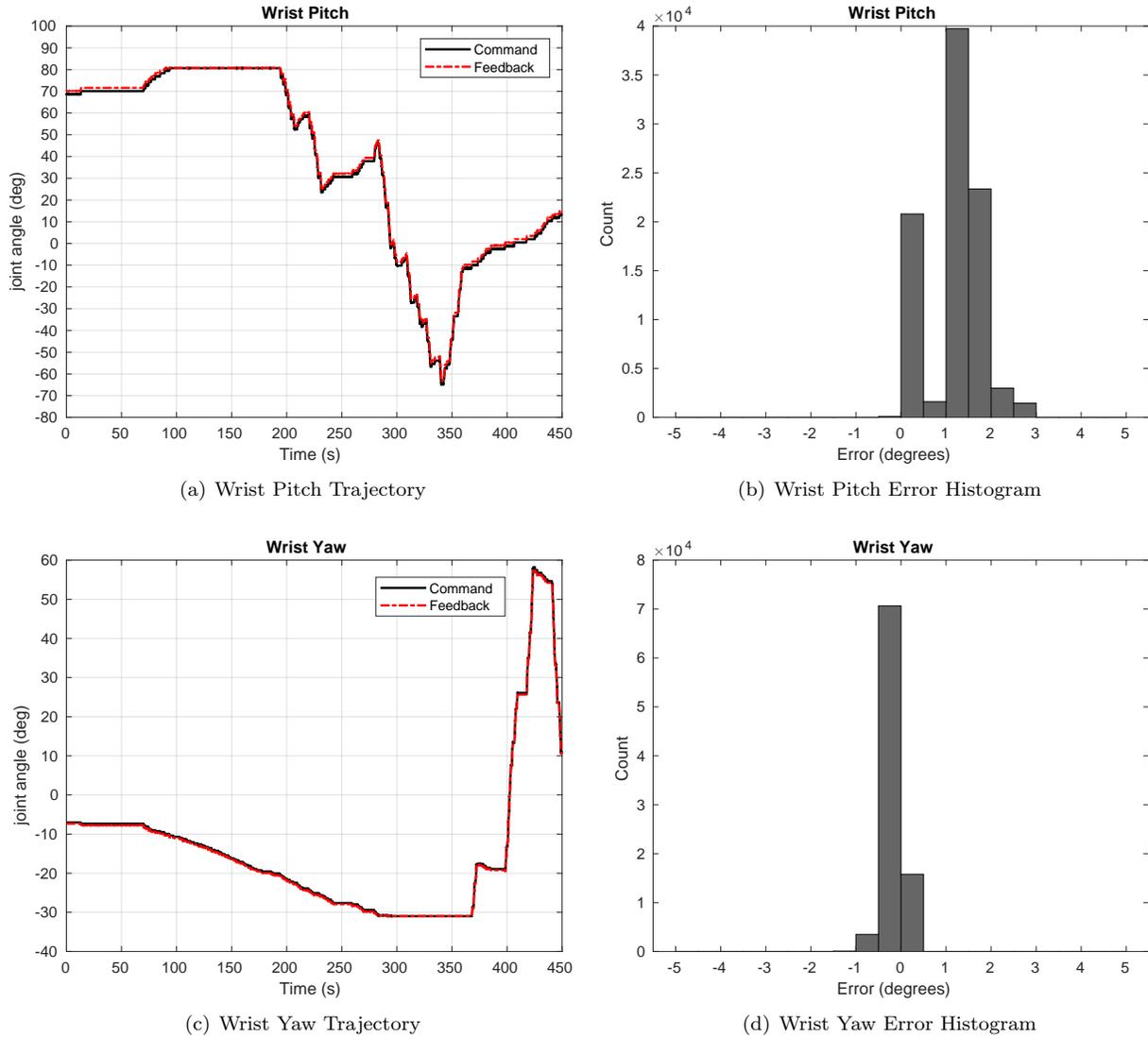
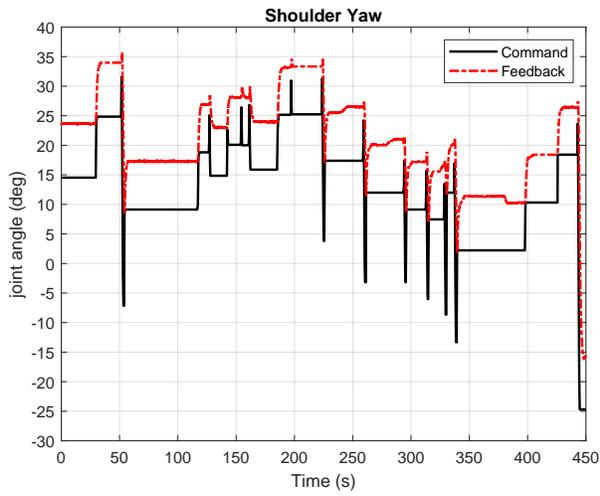


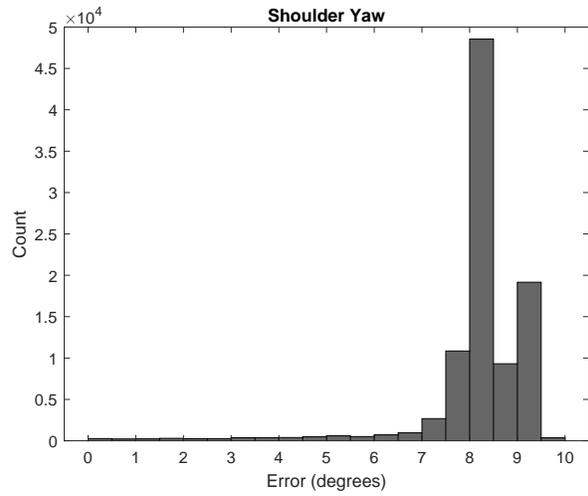
Figure 21: Plot of commanded versus followed joint trajectories for the *testbed* manipulator.

during the execution of a manipulator command. For example, if a command is given to grasp a detected tool, the pose of the tool in the scene is assumed to remain fixed during the execution of the grasp. If the tool were to move due to some disturbance before the grasp was completed, the grasp action would likely fail. Future work may integrate an obstacle-aware visual servoing controller to complete grasps or perform precise tool placement, which would reject disturbances to either the scene or the manipulator during task execution. Because our system relies on visual sensing, any disturbance to the scene that results in degraded water clarity, such as stirring up bottom sediment, can necessitate waiting for the water column to clear before the manipulation task can continue. While the KRAFT manipulator used in our field trials is particularly low power when idle, minimizing the energy cost of waiting for visual conditions to improve, future research may improve robustness of the system to degraded visual conditions by fusing acoustic imaging sonar data into the scene mapping framework. Compared to visual sensors, acoustic signals are not dependent on lighting conditions and are not degraded by haze in the water column or a sparsely textured seafloor.

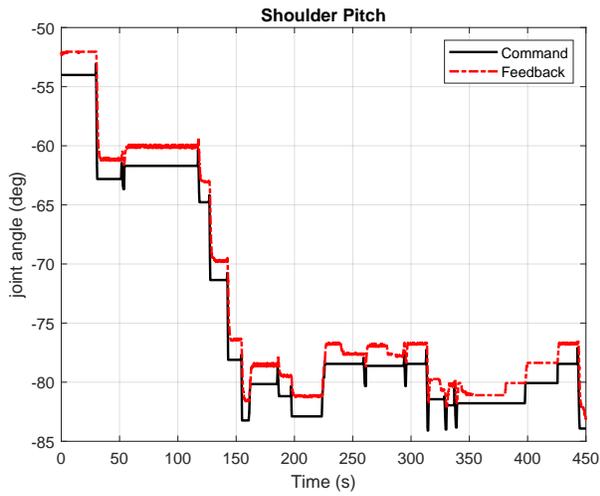
The technology presented in this report can be directly integrated onto terrestrial-based underwater manipulation platforms in order to decrease operational risk, reduce system complexity, and increase overall



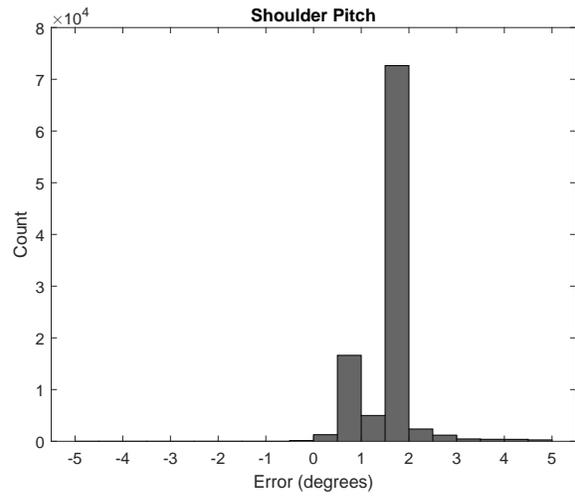
(e) Shoulder Yaw Trajectory



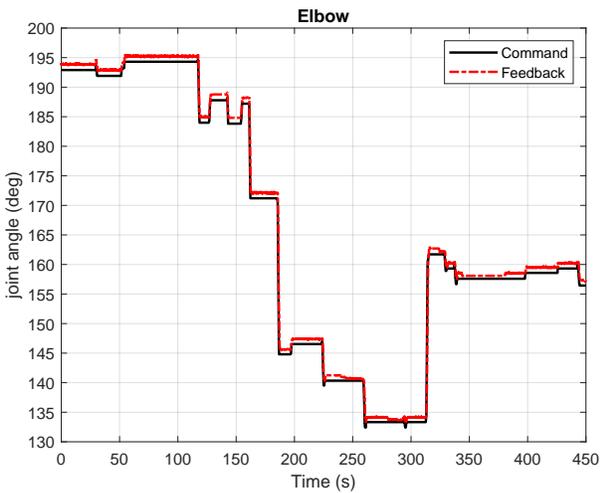
(f) Shoulder Yaw Error Histogram



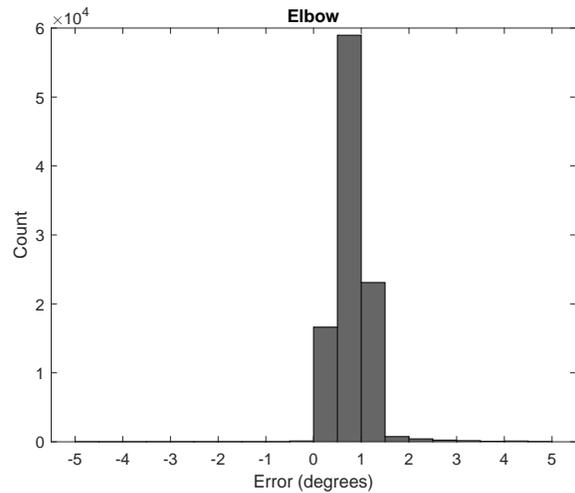
(g) Shoulder Pitch Trajectory



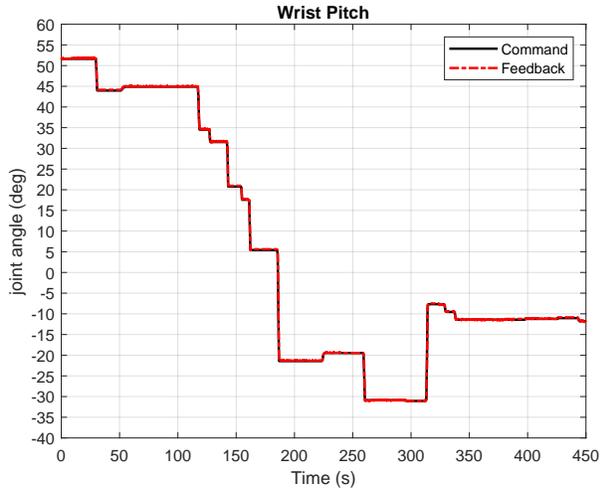
(h) Shoulder Pitch Error Histogram



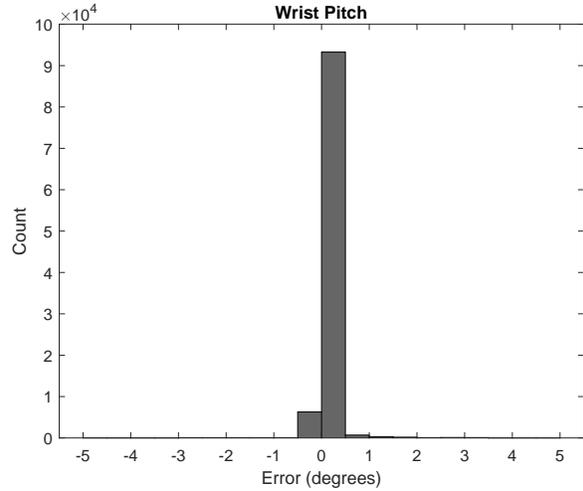
(i) Elbow Trajectory



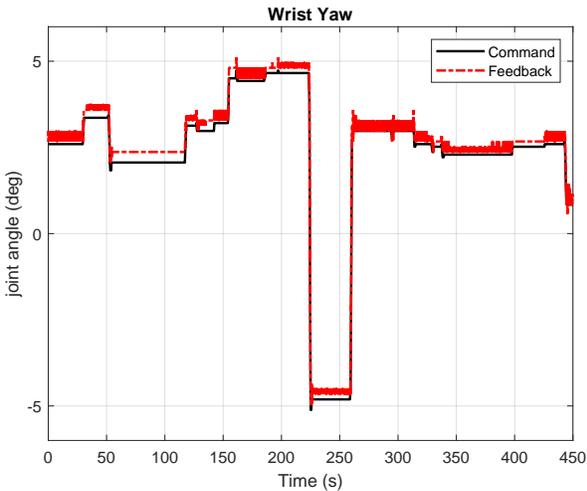
(j) Elbow Error Histogram



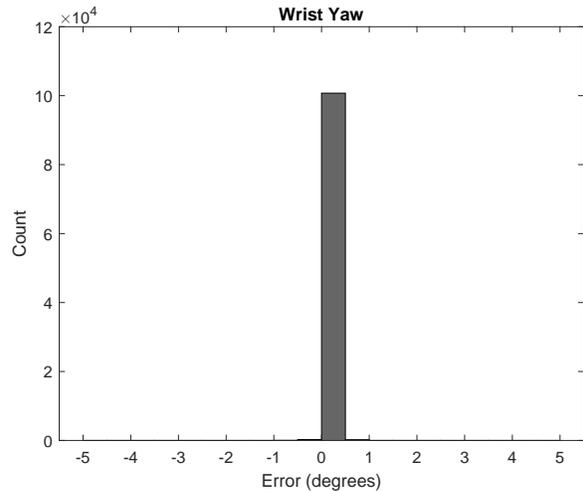
(a) Wrist Pitch Trajectory



(b) Wrist Pitch Error Histogram



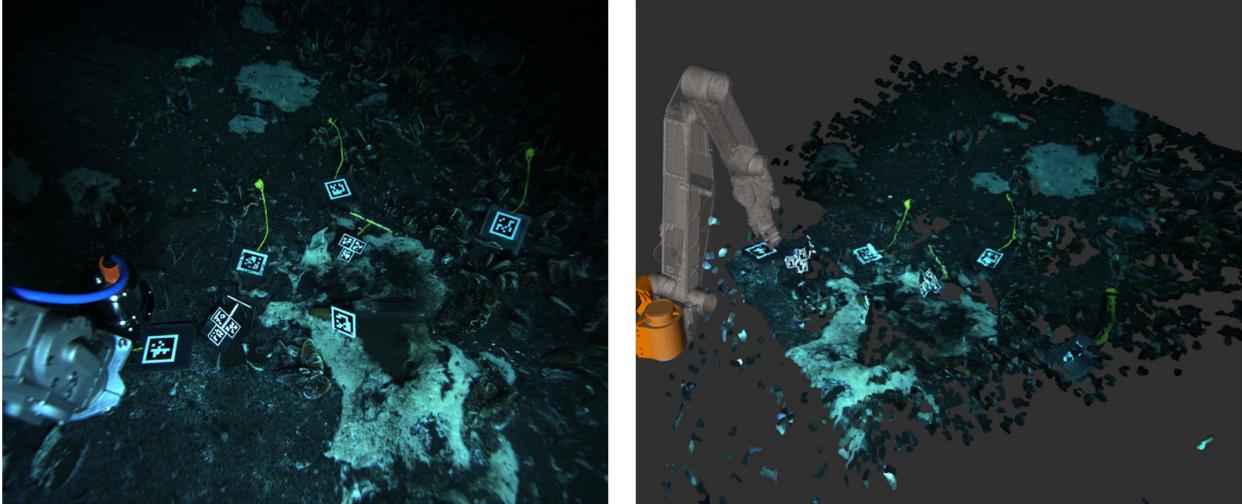
(c) Wrist Yaw Trajectory



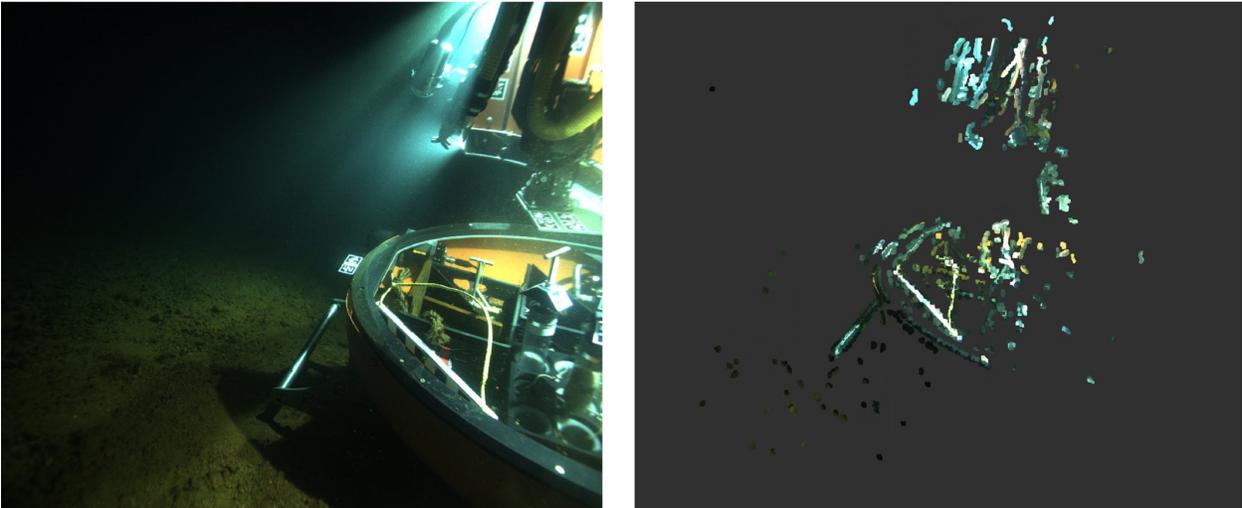
(d) Wrist Yaw Error Histogram

Figure 22: Plot of commanded versus followed joint trajectories for the *NUI* HROV manipulator during the Greece field trials.

efficiency. The current standard for ROV manipulation requires one or more pilots to operate the UVMS based on image feeds from an array of cameras on the vehicle that are displayed on a set of monitors in a ship-side control van. Existing systems do not provide pilots with an estimate of the 3D scene structure, putting the system at risk of collision between the arm and the vehicle or workspace objects. This, together with the cognitive load imposed by having to interpret multiple sensor streams makes it extremely challenging for pilots to establish and maintain situational awareness. The technology presented in this report can be integrated at three different levels with existing ROV systems. At the first and most basic level, the system can act as a decision support tool that provides a detailed real-time 3D visualization of the scene, including the vehicle and manipulator configuration and a reconstruction of the workspace, enabling a pilot to position the manipulator with greater accuracy, speed, and safety. At the second level, the system can be integrated into the manipulator control system for execution monitoring to limit the motion of the manipulator based on scene structure, preventing the pilot from moving the manipulator into collision or a risky configuration (Sivčev et al., 2018c). At the third and highest level, manipulation tasks may be fully automated so that a pilot simply selects a desired function or indicates an intent through some mode of



(e) Good image quality and dense point cloud



(f) Poor image quality and sparse point cloud

Figure 23: The quality of stereo reconstruction is highly dependent on underwater conditions. Here, we compare stereo point clouds generated using the same camera system and stereo matching method, but with images captured within very different seafloor environments. The left images show the view from the left stereo camera, and the right images show the generated point clouds using a SGM-based stereo method. (e) The top row was captured in the clear waters off Costa Rica, with even scene lighting and highly textured seafloor. (f) The bottom row was captured in the Kolumbo caldera, with high backscatter and low texture microbial mats on the seafloor.

communication such as natural language, whereupon the system plans and executes the task while providing visual feedback to the pilot. In this case, it is critical that the pilot be able to override the automated process and take over control of the arm at will.

For teleoperation of ROV manipulators, it is standard practice to stream multiple high-definition (HD) camera feeds at 30 Hz to the operating pilots. In the most bandwidth constrained circumstances, Compressed standard-definition (SD) cameras can be streamed at 10 Hz to the pilots. At lower image resolutions or framerates, it becomes difficult for pilots to teleoperate the manipulator safely. Our system enables high-level command of the manipulator and mitigates the need for continuous image streams back to the controlling

Table 2: Comparison of the bandwidth requirements for direct teleoperation (top two rows) of an ROV manipulator system compared to operating our high-level autonomy system (bottom two rows), running onboard the vehicle with communication through natural language commands and only the necessary scene state feedback to inform the high-level commands.

Mode	Data Type	Bandwidth
Teleoperation Cameras	Compressed SD or HD @ 10–30 Hz	100 KB/s–3 MB/s
Teleoperation Manipulator Coms	2 way $\times$ 15–200 Hz $\times$ 18 B	540 B/s–7.2 KB/s
Natural Language	1 B/letter $\times$ $\sim$ 7 letters/word $\times$ $\sim$ 2.5 words/s	17.5 B/s
Scene State Feedback	State and Compressed Images @ 0.1–1 Hz	3–30 KB/s

pilot. Single image frames need only be sent when a scene change is detected or on request. Future work on the vision system will develop methods for semantic-level scene understanding, which will further reduce the need for direct image streams back to the pilot. For a semantic aware system, natural language is well suited for human-machine interaction and can drastically reduce the data communication load between the vehicle platform and a remote operator by on-boarding data heavy computation (e.g., image processing) onto the vehicle’s local compute system and interfacing with the remote operator through small bandwidth language packets. For our system to operate with pilot oversight, high level commands and sensory feedback need only be streamed at rates which match the dynamics of the scene. In the scenario where the vehicle is set down on the seafloor to collect samples, the relevant scene dynamics can be on the order of seconds, minutes or longer, enabling significant reduction of the communication bandwidth which is vital for remote operations over bandwidth limited connections, such as satellite links. Table 2 shows estimated bandwidth range requirements for the manipulator coms and image streams necessary to support direct teleoperation of an ROV manipulator system compared to the bandwidth requirements for natural language communication with the vehicle and only the necessary scene state feedback to inform the high level commands. In the case of direct teleoperation, the manipulator coms can range from 15 Hz to 200 Hz two-way communication with a typical packet size of 18 B. We estimate the image bandwidth for a single SD or HD camera with compressed data streamed at 10 Hz–30 Hz, though generally multiple camera views are streamed simultaneously back to the pilot for safe manipulator control. In the case of our high-level automation system, the natural language data rates are based on approximate estimates for the average letter count per word and the speech rate. This data rate represents the expected maximum bandwidth load when transmitted in real-time, as language based communication is intermittent and can be compressed. The scene state feedback includes the vehicle state such as the manipulator joint states and semantic information, such as the type and pose of detected tools. However, the visual scene state feedback takes up the bulk of the bandwidth and is assumed to be encoded as a compressed camera frame or view of the 3D scene reconstruction. As demonstrated in the table, communication requirements to support our high-level system reduce the necessary bandwidth load by at least an order of magnitude compared to the requirements of the most limited direct teleoperation modality.

Despite the technological challenges in reaching extraterrestrial worlds, the NASA Science Mission Directorate (SMD) sets its first priority “to discover the secrets of the universe, to search for life, and to protect and improve life on Earth” (NASA Science Mission Directorate, 2020) and “is undertaking a flagship mission to Jupiter’s moon Europa, as its subsurface ocean has great potential to harbor extraterrestrial life.” A Europa mission concept for a surface lander has reached relative maturity, having passed its delta Mission Concept Review (Hand et al., 2021). The sampling system is recognized as being critical to the success of the mission and relies on a robotic arm for “excavation, collection, and presentation (or transfer) of samples to scientific instruments for observation and analysis” (Hand et al., 2017). Due to the anticipated communication limitations, it is likely that the lander will be required to self-select sampling sites, in which “the sampling system would be capable of conducting a sampling cycle in a fully autonomous fashion with no input from ground operators, from target selection to sample delivery. This autonomous capability is to guard against a prolonged telecommunications fault during the short mission lifetime, and will be in place to provide added assurance that the mission threshold science would be met” (Hand et al., 2017). Challenges to

the sampling system will be exacerbated by “poorly-characterized terrain at small scales”, and “the terrain immediately in front of the landing spot must suffice for sampling locations; there is no mobility system that can be used to search for a better site” (Hand et al., 2017). The methods we demonstrated in this report for automated manipulator control and sample collection are directly applicable to operations focused on a Ladder of Life detection mission scenario (Neveu et al., 2018). With the exception of the wrist mounted camera, the manipulator and imaging system used in our demonstrations are very similar to the hardware for the Europa lander concept, consisting of a multi-DoF manipulator and vehicle-mounted stereo pair. A primary limiting factor on the integration of our autonomy methods with the lander would be the available computational power. However, for a stationary lander, the visual processing, which is the primary computational bottleneck, could operate at low-frame rates suitable for extraterrestrial exploration, assuming the environment dynamics are sufficiently slow. The methods we describe are also suitable to run on embedded systems and may be optimized accordingly.

## 6 Conclusions

An exobiology search mission to distant ocean worlds will require a highly automated exploratory vehicle, capable of operating in extreme conditions for an extended period of time. Such a platform will likely be outfitted with a manipulator to maximize the types of samples that could be collected. In this report we describe a vision system and control framework for automating an ROV manipulator. This architecture is readily integrated onto a wide array of vehicle platforms, and we have demonstrated the viability of the system in the field on two ROVs with different manipulators, including the *NUI* HROV which is dynamically reconfigurable. In November of 2019, we demonstrated planner-controlled sample collection and return within active submarine volcanoes that host diverse assemblages of extremophile organisms. These operation locations served as analogs to environments that may exist within other ocean worlds in our solar system and beyond.

A current limitation of our approach is a semi-static vehicle and scene assumption, where the ROV is held stationary and the scene does not change during execution of a manipulator motion, though the vehicle and scene state may change between motions. The vehicle is typically kept stationary by setting it down on the seafloor before manipulation is initiated. This assumption limits the type of sampling tasks that may be performed with the described system. For example collecting samples from a vertical wall, the underside of an ice shelf, or other moving objects would require free-floating control. Free-floating manipulation is an open problem in robotics, and a promising research direction that directly builds on our demonstrated system is obstacle aware disturbance rejection control of the manipulator. This method is similar to obstacle aware visual servoing, using feature based SLAM with the vision system to compensate for vehicle motions and stabilize the end-effector. A disturbance rejection approach would enhance the flexibility of the system to be easily integrated on different vehicles and manipulators without requiring the generation of complex vehicle and manipulator dynamic models.

While the demonstrated system represents a significant step towards autonomous sample collection and return from seafloor environments, more advancements are required before the system can be deployed reliably in a fully automated fashion. In particular, visual methods must be developed that are robust to the optical challenges of the underwater environment in order to enable safe and targeted sample collection and precision tool handling. These methods must be robust to dynamic scenes, insensitive to the intensity inconsistency of underwater lighting and perform well in sparsely featured and low-textured environments. Fusion of sparse feature based methods for SLAM with learning-based methods for dense scene reconstruction and high-level semantic scene understanding, such as segmentation, object detection and tool pose estimation may provide an appropriate path forward to overcome this challenge.

In summary, automated exploration of unstructured seafloor environments is within reach of current underwater robotic technology. More development is needed, particularly in methods for scene reconstruction and understanding, to make this technology sufficiently reliable for fully automated deployment, but results from

our oceanographic expeditions described in this report demonstrate that a wide range of existing ROVs and manipulator systems can be adapted, with moderate effort, for high level automation capabilities.

## **Acknowledgments**

This work was funded under a NASA PSTAR grant, number NNX16AL08G, and by the National Science Foundation under grants IIS-1830660 and IIS-1830500. The authors would like to thank the Costa Rican Ministry of Environment and Energy and National System of Conservation Areas for permitting research operations at the Costa Rican shelf margin and the Schmidt Ocean Institute (including the captain and crew of the *R/V Falkor*, and ROV *SuBastian*) for their generous support and making the FK181210 expedition safe and highly successful. Additionally, the authors would like to thank the Greek Ministry of Foreign Affairs for permitting the 2019 Kolumbo Expedition to the Kolumbo and Santorini calderas, as well as Prof. Evi Nomikou and Dr. Aggelos Mallios for their expert guidance and tireless contributions to the expedition. We would also like to thank Maritech and crew of the *CLV OceanLink*, and the HROV *NUI* crew for their skillful and friendly assistance with integration, testing, and field operations. Finally we would like to thank Prof. Blair Thornton for graciously sharing his underwater camera housing design, which was used in our work described here.

## References

- Arnold, H., Liuzzo, L., and Simon, S. (2019). Magnetic signatures of a plume at Europa during the Galileo E26 flyby. *Geophysical Research Letters*, 46(3):1149–1157.
- Ayton, B., Williams, B., and Camilli, R. (2019). Measurement maximizing adaptive sampling with risk bounding functions. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 7511–7519.
- Billings, G. and Johnson-Roberson, M. (2019). SilhoNet: An RGB method for 6D object pose estimation. *IEEE Robotics and Automation Letters*, 4(4):3727–3734.
- Billings, G. and Johnson-Roberson, M. (2020). SilhoNet-Fisheye: Adaptation of a ROI based object pose estimation network to monocular fisheye images. *IEEE Robotics and Automation Letters*, 5(3):4241–4248.
- Billings, S. E. and Kattenhorn, S. A. (2005). The great thickness debate: Ice shell thickness models for Europa and comparisons with estimates based on flexure at ridges. *Icarus*, 177(2):397–412.
- Birk, A., Doernbach, T., Mueller, C., Luczyński, T., Chavez, A. G., Koehntopp, D., Kupcsik, A., Calinon, S., Tanwani, A., Antonelli, G., Lillo, P. D., Simetti, E., Casalino, G., Indiveri, G., Ostuni, L., Turetta, A., Caffaz, A., Weiss, P., Gobert, T., Chemisky, B., Gancet, J., Siedel, T., Govindaraj, S., Martínez, X., and Letier, P. (2018). Dexterous underwater manipulation from onshore locations: Streamlining efficiencies for remotely operated underwater vehicles. *IEEE Robotics & Automation Magazine*, 25:24–33.
- Bowen, A. D., Yoerger, D. R., German, C. C., Kinsey, J. C., Jakuba, M. V., Gomez-Ibanez, D., Taylor, C. L., Machado, C., Howland, J. C., Kaiser, C. L., et al. (2014). Design of Nereid-UI: A remotely operated underwater vehicle for oceanographic access under ice. In *2014 Oceans—St. John’s*.
- Brooks, J. M., Kennicutt, M., Fisher, C., Macko, S., Cole, K., Childress, J., Bidigare, R., and Vetter, R. (1987). Deep-sea hydrocarbon seep communities: evidence for energy and nutritional carbon sources. *Science*, 238(4830):1138–1142.
- Broome, D., Larkum, T., and Hall, M. (1995). Subsea weld inspection using an advanced robotic manipulator. In *‘Challenges of Our Changing Global Environment’*. *Conference Proceedings. OCEANS’95 MTS/IEEE*, volume 2, pages 1216–1224.
- Camilli, R., Nomikou, P., Escartín, J., Ridao, P., Mallios, A., Kiliyas, S. P., Argyraki, A., Andreani, M., Ballu, V., Campos, R., et al. (2015). The Kallisti Limnes, carbon dioxide-accumulating subsea pools. *Scientific Reports*, 5.
- Campbell, E., Bingham, N., and Williams, J. (2019). 4500 m remotely operated vehicle (ROV SuBastian).
- Carey, S., Nomikou, P., Bell, K. C., Lilley, M., Lupton, J., Roman, C., Stathopoulou, E., Bejelou, K., and Ballard, R. (2013). Co<sub>2</sub> degassing from hydrothermal vents at kolumbo submarine volcano, greece, and the accumulation of acidic crater water. *Geology*, 41(9):1035–1038.
- Cieslak, P., Ridao, P., and Giergiel, M. (2015). Autonomous underwater panel operation by GIRONA500 UVMS: A practical approach to autonomous underwater manipulation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 529–536.
- Coleman, D., Sucan, I., Chitta, S., and Correll, N. (2014). Reducing the barrier to entry of complex robotic software: A MoveIt! case study. *Journal of Software Engineering for Robotics*, 5(1):3–16.
- De Novi, G., Melchiorri, C., Garcia, J., Sanz, P. J., Ridao, P., and Oliver, G. (2010). New approach for a reconfigurable autonomous underwater vehicle for intervention. *Aerospace and Electronic Systems Magazine*, 25(11):32–36.
- Dellaert, F. (2012). Factor graphs and GTSAM: A hands-on introduction. Technical report, Georgia Institute of Technology.
- dos Santos Cesar, D. B., Gaudig, C., Fritsche, M., dos Reis, M. A., and Kirchner, F. (2015). An evaluation of artificial fiducial markers in underwater environments. In *OCEANS 2015 - Genova*, pages 1–6.
- Duguid, Z. and Camilli, R. (2020). Improving resource management for unattended observation of the marginal ice zone using autonomous underwater gliders. *Frontiers in Robotics and AI*, 7:184.

- Fong, T., Thorpe, C., and Baur, C. (2001). *Collaborative control: A robot-centric model for vehicle teleoperation*, volume 1. Carnegie Mellon University, The Robotics Institute Pittsburgh.
- Francois, R. and Nodland, W. (1972). *Unmanned Arctic research submersible (UARS) system development and test report*. Applied Physics Laboratory, University of Washington.
- Gaeman, J., Hier-Majumder, S., and Roberts, J. H. (2012). Sustainability of a subsurface ocean within Triton’s interior. *Icarus*, 220(2):339–347.
- Hand, K., Murray, A., Garvin, J., et al. (2017). Europa lander study 2016 report.
- Hand, K., Phillips, C., Maize, E., Reeves, G., Pitesky, J., Craft, K., Cameron, M., Tan-Wang, G., San Martin, A., Crum, R., et al. (2021). Europa lander mission concept (update 2021). In *Lunar and Planetary Science Conference*, number 2548, page 2120.
- Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42:335–346.
- Hensen, C., Wallmann, K., Schmidt, M., Ranero, C. R., and Suess, E. (2004). Fluid expulsion related to mud extrusion off Costa Rica—A window to the subducting slab. *Geology*, 32(3):201–204.
- Hildebrandt, M., Kerdels, J., Albiez, J., and Kirchner, F. (2009). A multi-layered controller approach for high precision end-effector control of hydraulic underwater manipulator systems. In *OCEANS 2009*.
- Howard, T. M., Tellex, S., and Roy, N. (2014). A natural language planner interface for mobile manipulators. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 6652–6659.
- Hsu, H.-W., Postberg, F., Sekine, Y., Shibuya, T., Kempf, S., Horányi, M., Juhász, A., Altobelli, N., Suzuki, K., Masaki, Y., et al. (2015). Ongoing hydrothermal activities within Enceladus. *Nature*, 519(7542):207–210.
- Iess, L., Stevenson, D. J., Parisi, M., Hemingway, D., Jacobson, R. A., Lunine, J. I., Nimmo, F., Armstrong, J. W., Asmar, S. W., Ducci, M., and Tortora, P. (2014). The gravity field and interior structure of Enceladus. *Science*, 344(6179):78–80.
- Ishimi, K., Ohtsuki, Y., Manabe, T., and Nakashima, K. (1991). Manipulation system for subsea operation. In *Proceedings of the International Conference on Advanced Robotics (ICAR)*, pages 1348–1353.
- Jannasch, H. W. and Wirsen, C. O. (1979). Chemosynthetic primary production at East Pacific sea floor spreading centers. *Bioscience*, 29(10):592–598.
- Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J. J., and Dellaert, F. (2012). iSAM2: Incremental smoothing and mapping using the bayes tree. *The International Journal of Robotics Research*, 31(2):216–235.
- Kannala, J. and Brandt, S. (2006). A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1335–1340.
- Karaman, S. and Frazzoli, E. (2011). Sampling-based algorithms for optimal motion planning. *International Journal of Robotics Research*, 30(7):846–894.
- Khurana, K., Kivelson, M., Stevenson, D., Schubert, G., Russell, C., Walker, R., and Polanskey, C. (1998). Induced magnetic fields as evidence for subsurface oceans in Europa and Callisto. *Nature*, 395(6704):777–780.
- Krause, S., Steeb, P., Hensen, C., Liebetrau, V., Dale, A. W., Nuzzo, M., and Treude, T. (2014). Microbial activity and carbonate isotope signatures as a tool for identification of spatial differences in methane advection: A case study at the pacific Costa Rican margin. *Biogeosciences*, 11(2):507–523.
- Leger, P. C., Deen, R. G., and Bonitz, R. G. (2005). Remote image analysis for Mars Exploration Rover mobility and manipulation operations. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pages 917–922.
- Lehner, P., Brunner, S., Dömel, A., Gmeiner, H., Riedel, S., Vodermayr, B., and Wedler, A. (2018). Mobile manipulation for planetary exploration. In *Proceedings of the IEEE Aerospace Conference*.

- Levin, L. A., Mendoza, G. F., Grupe, B. M., Gonzalez, J. P., Jellison, B., Rouse, G., Thurber, A. R., and Waren, A. (2015). Biodiversity on the rocks: Macrofauna inhabiting authigenic carbonate at costa rica methane seeps. *PLoS One*, 10(7):e0131080.
- Lewis, J. S. (1971). Satellites of the outer planets: Their physical and chemical nature. *Icarus*, 15(2):174–185.
- Lowell, R. P. and DuBose, M. (2005). Hydrothermal systems on Europa. *Geophysical Research Letters*, 32(5).
- Malin, M. C. and Edgett, K. S. (2000). Evidence for recent groundwater seepage and surface runoff on Mars. *Science*, 288(5475):2330–2335.
- Marani, G., Choi, S. K., and Yuh, J. (2009). Underwater autonomous manipulation for intervention missions AUVs. *Ocean Engineering*, 36(1):15–23. Autonomous Underwater Vehicles.
- NASA Science Mission Directorate (2020). Science 2020-2024, A vision for scientific excellence.
- Neveu, M., Hays, L. E., Voytek, M. A., New, M. H., and Schulte, M. D. (2018). The ladder of life detection. *Astrobiology*, 18(11):1375–1402.
- Nimmo, F., Spencer, J., Pappalardo, R., and Mullen, M. (2007). Shear heating as the origin of the plumes and heat flux on Enceladus. *Nature*, 447(7142):289–291.
- Nomikou, P., Carey, S., Bell, K. C., Papanikolaou, D., Bejelou, K., Alexandri, M., Cantner, K., and Martin, J. F. (2013). Morphological analysis and related volcanic features of the Kolumbo submarine volcanic chain (NE of Santorini Island, Aegean Volcanic Arc). *Zeitschrift für Geomorphologie*, 57(3):029–047.
- Nomikou, P., Carey, S., Papanikolaou, D., Bell, K. C., Sakellariou, D., Alexandri, M., and Bejelou, K. (2012). Submarine volcanoes of the Kolumbo volcanic zone NE of Santorini Caldera, Greece. *Global and Planetary Change*, 90:135–151.
- Nomikou, P., Hannington, M., Petersen, S., Wind, S., Heinath, V., Lange, S., Rothenbeck, M., Triebe, L., and Wenzlaff, E. (2019). Advanced mapping of kolumbo submarine volcano (Santorini) using auv abyss.
- Olson, E. (2011). AprilTag: A robust and flexible visual fiducial system. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3400–3407, Shanghai, China.
- Peñalver, A., Pérez, J., Fernández, J. J., Sales, J., Sanz, P. J., García, J. C., Fornas, D., and Marín, R. (2015). Visually-guided manipulation techniques for robotic autonomous underwater panel interventions. *Annual Reviews in Control*, 40:201–211.
- Pfrommer, B. and Daniilidis, K. (2019). TagSLAM: Robust SLAM with fiducial markers. *arXiv preprint arXiv:1910.00679*.
- Sahling, H., Masson, D. G., Ranero, C. R., Hühnerbach, V., Weinrebe, W., Klauke, I., Bürk, D., Brückmann, W., and Suess, E. (2008). Fluid seepage at the continental margin offshore costa rica and southern nicaragua. *Geochemistry, Geophysics, Geosystems*, 9(5).
- Sanz, P. J., Ridao, P., Oliver, G., Casalino, G., Petillot, Y., Silvestre, C., Melchiorri, C., and Turetta, A. (2013). TRIDENT an European project targeted to increase the autonomy levels for underwater intervention missions. In *2013 OCEANS—San Diego*.
- Shim, H., Jun, B.-H., Lee, P.-M., Baek, H., and Lee, J. (2010). Workspace control system of underwater tele-operated manipulators on an ROV. *Ocean Engineering*, 37(11):1036–1047.
- Silver, E., Kastner, M., Fisher, A., Morris, J., McIntosh, K., and Saffer, D. (2000). Fluid flow paths in the middle america trench and costa rica margin. *Geology*, 28(8):679–682.
- Simetti, E., Wanderlingh, F., Torelli, S., Bibuli, M., Odetti, A., Bruzzone, G., Rizzini, D. L., Aleotti, J., Palli, G., Moriello, L., et al. (2017). Autonomous underwater intervention: Experimental results of the MARIS project. *IEEE Journal of Oceanic Engineering*, 43(3):620–639.
- Sivčev, S., Coleman, J., Omerdić, E., Dooly, G., and Toal, D. (2018a). Underwater manipulators: A review. *Ocean Engineering*, 163:431–450.
- Sivčev, S., Rossi, M., Coleman, J., Dooly, G., Omerdić, E., and Toal, D. (2018b). Fully automatic visual servoing control for work-class marine intervention ROVs. *Control Engineering Practice*, 74:153–167.

- Sivčev, S., Rossi, M., Coleman, J., Omerdić, E., Dooly, G., and Toal, D. (2018c). Collision detection for underwater ROV manipulator systems. *Sensors*, 18(4).
- Timmons, E., Ayton, B., Wang, A., Pascucci, N., Zhang, Y., Bhargava, N., Reeves, M., Duguid, Z., Strawser, D., Fang, C., et al. (2019). Risk-bounded, goal-directed mission planning and execution for autonomous ocean exploration. In *Proceedings of the Astrobiology Science Conference*.
- Timmons, E., Vaquero, T., Williams, B., and Camilli, R. (2016). Preliminary deployment of a risk-aware goal-directed executive on autonomous underwater glider. In *PlanRob Workshop, ICAPS (London, UK:)*.
- Tsai, R. Y., Lenz, R. K., et al. (1989). A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Transactions on robotics and automation*, 5(3):345–358.
- Vrolijk, P., Camilli, R., Summa, L., and Nomikou, P. (2019). Cruise FK181210 on RV Falkor.
- Vrolijk, P., Summa, L., Ayton, B., Nomikou, P., Huepers, A., Kinnaman, F., Sylva, S., Valentine, D., and Camilli, R. (2021). Using a ladder of seeps with computer decision processes to explore for and evaluate cold seeps on the Costa Rica Active Margin. *Frontiers in Earth Science*, 9:143.
- Youakim, D., Ridaou, P., Palomeras, N., Spadafora, F., Ribas, D., and Muzzupappa, M. (2017). MoveIt!: Autonomous underwater free-floating manipulation. *IEEE Robotics & Automation Magazine*, 24(3):41–51.
- Zhang, Z., Wang, C., Zhang, Q., Li, Y., Feng, X., and Wang, Y. (2019). Research on autonomous grasping control of underwater manipulator based on visual servo. In *Chinese Automation Congress (CAC)*, pages 2904–2910.