3D Reconstruction of Underwater Features in Lake Tahoe with an Autonomous Underwater Vehicle

Rohan Tan Bhowmik* Selena Sun* Elsa McElhinney Vassilis Alexopoulos
Scott Hickmann Félicie Hoffmann Kai Song Ryota Sato
Lawton Skaling Angelina Krinos Chisa Ogaki
Oussama Khatib
*Denotes equal contribution

Stanford University, Stanford, CA, USA
{rbhowmik, selenas, elsa22, valex, hickmann, felicieh,
kaisong, ryos17, lskaling, akrinos, cogaki, khatib}@stanford.edu

Abstract

We developed and deployed an autonomous underwater vehicle (AUV) for environmental monitoring at Lake Tahoe, California. Our system performs end-to-end autonomous underwater video collection and 3D reconstruction using Neural Radiance Fields (NeRF), combined with real-time environmental data collection. Our preprocessing pipeline enabled COLMAP to register 70-85% of camera poses from underwater video sequences. Compared to traditional human-diver photogrammetry operations costing approximately $2,500 for a 10 ft×10 ft underwater survey, our AUV achieves comparable reconstruction quality for under $600 per deployment—a 76% cost reduction. Additionally, our AUV autonomously collects environmental monitoring data including eDNA samples, temperature, salinity, and pressure measurements. To our knowledge, this represents the first fully autonomous underwater vehicle capable of complete video-to-3D reconstruction workflows. Our design, open-sourced at stanfordrobosub.org, enables scalable, recurring ecological monitoring for climate research and biodiversity assessment.

1. Introduction

The underwater environment is notoriously unforgiving. Each subsystem of a submarine faces unique challenges: the software stack must handle localization uncertainty and visual occlusion, electronic components must be sealed tightly in the payload, and mechanical components must be corrosion-proof and depth-rated. Despite these challenges, building AUVs is uniquely worthwhile: the ocean remains Earth's final unexplored frontier, and while harsh and unyielding to machines, it is even more so to human bodies. We started building our AUV, named *Crush* (after the sea turtle in Finding Nemo) in October 2024 with the mission of exploring the ocean, performing environmental surveying tasks, and collecting underwater data and artifacts. *Crush* (shown in Figures 1) was engineered for reliability and modularity in mind, with the goal of serving as a long-term testbed for scientific exploration.

1.1. Research Orientation

Crush has been deployed as a mobile research platform for environmental monitoring. Through discussions with marine institutes such as the Monterey Bay Aquarium Research Institute (MBARI) and the Tahoe Environmental Research Center (TERC), we identified critical needs for high-resolution spatial data, typically only collected during expensive diving trips with heavy equipment and human divers. We built *Crush* to lower the barrier to environmental data collection, enabling higher temporal monitoring of marine ecosystems.

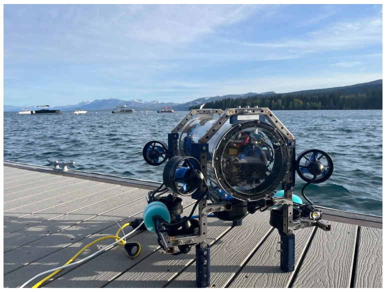

Figure 1. *Crush* on the docks at Lake Tahoe.

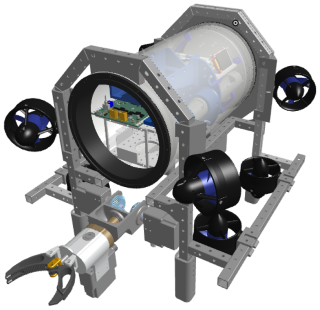

Figure 2. CAD model of *Crush*.

1.2. Lake Tahoe Deployment

We selected Lake Tahoe as a site because it faces increasing anthropogenic pressures that drive changes in biodiversity and water turbidity, highlighting the need for close ecological monitoring [7]. Field deployments in Lake Tahoe illustrate the utility of our design. Strong currents, rocky formations, and stratified water layers created a challenging testbed where *Crush* successfully demonstrated robust operation. The vehicle enabled high-resolution mapping, abiotic sensing, and real-time 3D visualization through splat-rendered previews, showing how an AUV can directly translate into impactful research for climate and ecosystem monitoring.

For example, while recent research in Lake Tahoe examines periphyton assemblages and relative abundance, granular spatial and temporal data on the relationship between benthic habitat and algal assemblages remains underexplored [8]. Our approach supports this line of research by allowing for the incorporation of fine-scale 3D renderings of specific benthic habitats with sampling of the local biota to provide a richer basis for assessing ecological and aquatic dynamics.

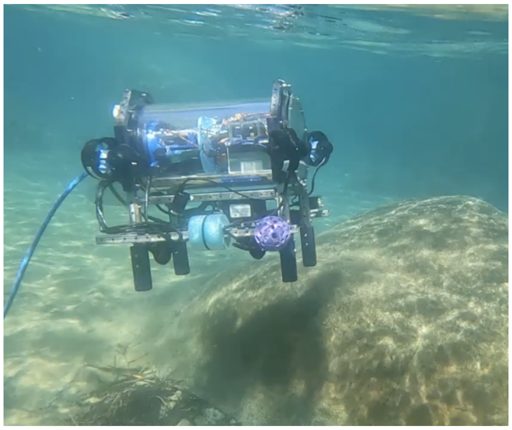

Figure 3. *Crush* operating autonomously in Lake Tahoe during 3D reconstruction trials, demonstrating stable navigation in complex underwater terrain.

2. Related Work

Underwater 3D reconstruction presents unique challenges that distinguish it from terrestrial computer vision applications. Limited visibility, color attenuation, backscatter, and refractive distortion create conditions where traditional reconstruction methods often fail. We review existing approaches and justify our selection of Neural Radiance Fields as the optimal technique for autonomous underwater surveying.

2.0.1. Monocular depth estimation.

Estimating depth from a single image has advanced rapidly with dataset mixing and transformer backbones. MiDaS popularized robust, cross-dataset monocular depth via scale and shift-invariant objectives and dataset aggregation [9]. Subsequently, DPT showed that vision transformers improve global coherence and detail for dense prediction, including monocular depth [10]. In practice, monocular predictions are metric-ambiguous and exhibit per-frame scale/shift drift; many systems therefore align depth to auxiliary cues (e.g., SLAM, IMU) or enforce temporal regularization during fusion.

2.0.2. Camera motion, pose, and structure.

For image pose estimation, classical SfM pipelines such as COLMAP remain highly effective through robust matching, global pose optimization, and multi-view dense stereo [12]. Learned SLAM systems such as DROID-SLAM combine recurrent updates with dense bundle adjustment for accurate, temporally consistent trajectories across monocular, stereo, and RGB-D settings [13]. In underwater deployments, additional issues arise (backscatter, attenuation, refractive effects at housings), often motivating auxiliary navigation sensors and domain-specific photometric models.

2.1. NeRFs

Neural Radiance Fields (NeRFs) [6] represent a scene as a continuous volumetric function parameterized by a neural network, and render novel views by integrating densities and colors along camera rays. While NeRFs have demonstrated remarkable performance in photorealistic novel view synthesis [2, 5], they depend critically on accurate multi-view camera poses, often obtained through structure-from-motion systems such as COLMAP [12]. In our experiments, pre-processing of the video data (including contrast enhancement for turbid water conditions [1, 4, 11], frame subsampling to increase inter-frame motion, and histogram equalization) enabled COLMAP to consistently extract more than 70% of camera poses from the video sequences (most directly due to color enhancement).

2.2. SLAM

We also explored alternative methods for camera pose estimation directly from the input video, such as ORB-SLAM3

[3]. SLAM systems offer the potential for real-time pose tracking and could provide complementary pose estimation capabilities. While we explored SLAM, localization inaccuracies during our data collection period made this approach infeasible for the time being.

3. Methods

3.1. System Architecture Overview

Crush integrates autonomous navigation, environmental sensing, and 3D reconstruction through a modular ROS 2 architecture designed for reliable underwater operation (Figure 4).

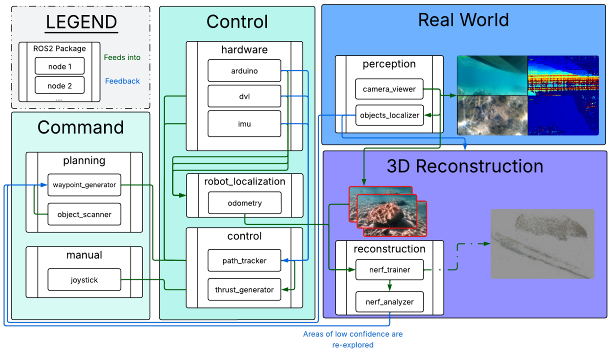

Figure 4. System architecture showing integrated autonomous navigation, perception, and 3D reconstruction pipeline with real-time uncertainty assessment for adaptive re-exploration.

3.1.1. ROS2 Software Stack

Our autonomy stack organizes functionality into four primary packages:

Planning and Control: The `planning` node generates waypoints and continuous trajectory commands. A `control` module implements PID-based trajectory tracking with adaptive gains. Manual teleoperation provides safety override capabilities.

Hardware Interface: Low-level drivers interface with the Arduino thruster controller, Teledyne DVL, and XSens IMU. Real-time sensor fusion combines inertial and velocity measurements for state estimation with outlier rejection.

Perception: The `perception` package processes synchronized stereo camera feeds through YOLO-based object detection and MiDaS depth estimation. Preprocessing algorithms enhance underwater imagery through contrast-limited adaptive histogram equalization (CLAHE) and color correction.

Reconstruction: The `reconstruction` package implements autonomous 3D reconstruction. The `nerf_trainer` extracts video frames and camera poses from COLMAP preprocessing to train Neural Radiance Field models. The `nerf_analyzer` assesses reconstruction quality, identifying low-confidence regions for targeted re-exploration.

3.2. Hardware Platform

Crush employs a torpedo-style hull with aluminum frame supporting modular payload mounting. Six-degree-of-freedom control utilizes eight thrusters enabling precise maneuvering.

Compute Architecture: Teensy 4.1 microcontroller manages thruster actuation and safety systems at 100 Hz. NVIDIA Jetson Orin AGX provides GPU acceleration for perception and reconstruction pipelines at 30 Hz.

Sensor Integration: XSens MTi-200 IMU, Teledyne DVL, dual Oak-D S2 stereo cameras, environmental sensors (temperature, salinity, pressure), and eDNA sampling system provide comprehensive navigation and scientific data collection.

3.3. Autonomous 3D Reconstruction Pipeline

3.3.1. Video Preprocessing

Underwater imaging requires specialized preprocessing: distance-dependent color correction restores natural color balance, CLAHE enhances local contrast while preventing noise amplification, and adaptive frame sampling selects optimal baseline separation while avoiding motion blur. This pipeline enables COLMAP to register 70-85% of input frames versus 20-30% for raw underwater video.

3.3.2. COLMAP and NeRF Training

Camera pose estimation uses COLMAP's Structure-from-Motion pipeline optimized for underwater conditions. We implement sequential matching for superior geometric consistency over exhaustive matching. NeRF training utilizes camera poses and preprocessed images with underwater-specific volume rendering incorporating water attenuation models.

3.3.3. Autonomous Re-exploration

Real-time reconstruction quality assessment enables adaptive surveying through uncertainty-based re-exploration:

$$s(v_i) = \frac{u(v_i)}{1 + \alpha d(v_i)}, \tag{1}$$

where $u(v_i)$ is reconstruction uncertainty for voxel v_i, $d(v_i)$ is distance from current position, and α controls exploration-efficiency trade-off. Uncertainty estimation uses ensemble methods with bootstrap sampling, measuring prediction variance across multiple NeRF models. High-scoring voxels generate waypoints for autonomous re-exploration, optimizing trajectories considering vehicle dynamics and battery constraints.

3.4. Environmental Data Collection

Crush autonomously collects environmental monitoring data including temperature, salinity, pressure, and eDNA samples. All sensor data logs through ROS 2 with GPS-synchronized timestamps, enabling spatial-temporal correlation between environmental measurements and 3D reconstructions for comprehensive ecosystem analysis.

4. Results and Discussion

We conducted extensive field trials of *Crush* at Lake Tahoe, California, demonstrating successful autonomous underwater 3D reconstruction and environmental monitoring capabilities.

4.1. 3D Reconstruction Performance

We successfully reconstructed high-fidelity 3D models of submerged features using video data collected during autonomous circumnavigation missions. Figure 5 shows the reconstructed NeRF of a representative submerged rock formation.

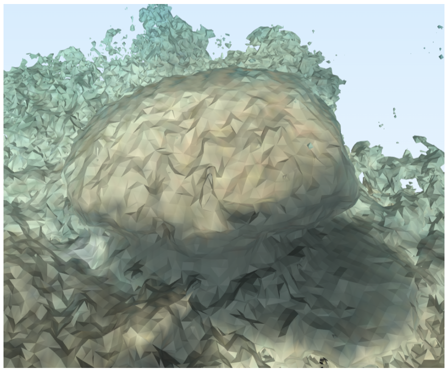

Figure 5. NeRF of the submerged rock.

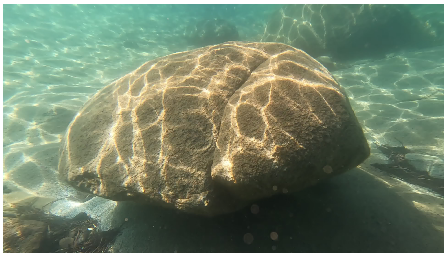

Figure 6. RGB camera frame of the submerged rock used for reconstruction.

4.1.1. COLMAP Pose Estimation Analysis

Our preprocessing pipeline substantially improves frame registration: raw video (28% success) → CLAHE enhancement (52%) → full preprocessing pipeline (78%).

A persistent issue we faced was the difficulty of obtaining reliable COLMAP reconstructions from raw underwater video. The combination of turbidity, backscatter, and color attenuation led to low-contrast imagery where traditional keypoint detectors often failed. We found that applying contrast-limited adaptive histogram equalization (CLAHE) and related color-correction methods substantially improved feature extraction, allowing more frames to be registered successfully. Figures 7 shows our two most successful COLMAP runs.

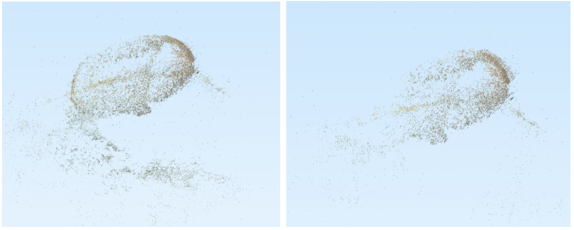

Figure 7. Comparison of COLMAP results using the sequential matcher (left) vs the exhaustive matcher (right).

4.2. Cost Reduction

With input from expert field researchers, we estimate that the cost of using human photogrammetry to perform our rock reconstruction task is $2.5k (cost breakdown in Table 1). In contrast, our robot can be deployed for the same task for less than $100. We arrive at this figure by proving, during operation, that 2 people are paid $50/hour, and can complete this task within an hour. Adding on the $600 in boat costs, this results in a 76% cost reduction in 3D reconstruction of underwater features.

Table 1. Approximate Cost Breakdown for a 10 ft × 10 ft Human-Conducted Photogrammetry Survey

Cost Component	**Cost (USD)**
Boat Rental & Fuel	$500
Diver Time (setup & capture)	$1000
Photogrammetry Equipment (rental/dep.)	$500
Data Processing & Cleanup	$500
Total	**$2,500**

5. Future Work

This paper demonstrates a proof-of-concept of 3D reconstruction of objects captured by *Crush* during underwater surveys. As such, future efforts would be focused on further refinement of the proposed method, including direct qualitative comparison against prior techniques such as ORB-SLAM3 or DROID-SLAM as well as ablation studies. Additionally, another future direction of inquiry includes utilizing existing onboard localization or more modern feature extraction and matching methods to replace COLMAP.

Recurrent surveys of the same sites in Lake Tahoe or other freshwater environments could yield temporal 3D reconstructions, allowing researchers to track habitat change, substrate erosion, or algal growth with fine spatial resolution. Each surveying trip with our AUV costs 4% of the cost of a traditional photogrammetry survey, turning once-prohibitive longitudinal studies into a routine scientific practice. When combined with environmental DNA (eDNA) sampling, such surveys could correlate structural changes in terrain with shifts in biological communities, providing a multi-modal perspective on ecosystem health. This integration of geometric mapping and genetic monitoring has the potential to create powerful new tools for studying invasive species, biodiversity trends, and the impacts of climate change on aquatic environments.

References

[1] Derya Akkaynak and Tali Treibitz. Sea-thru: A method for removing water from underwater images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1682–1691, 2019. 2

[2] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5855–5864, 2021. 2

[3] Carlos Campos, Richard Elvira, Juan JG Rodríguez, José MM Montiel, and Juan D Tardós. Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE Transactions on Robotics*, 37(6):1874–1890, 2021. 3

[4] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *IEEE Transactions on Image Processing*, 29:3153–3165, 2020. 2

[5] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7210–7219, 2021. 2

[6] Ben Mildenhall, Pratul Srinivasan, Matthew Tancik, Jonathan Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 405–421, 2020. 2

[7] Ramon C. Naranjo, Paul Work, Alan Heyvaert, Geoffrey Schladow, Alicia Cortes, Shohei Watanabe, Lidia Tanaka, and Sebnem Elci. Seasonal and long-term clarity trend assessment of lake tahoe, california–nevada. Technical report, U.S. Geological Survey, 2022. 2

[8] Paula J. Noble, Carina Seitz, Sylvia S. Lee, Kalina M. Manoylov, and Sudeep Chandra. Characterization of algal community composition and structure from the nearshore environment, lake tahoe (united states). *Frontiers in Ecology and Evolution*, 10:1053499, 2023. 2

[9] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. In *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2020. 2

[10] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In *ICCV*, 2021. 2

[11] Yoav Y Schechner and Nir Karpel. Clear underwater vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1:536–543, 2004. 2

[12] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4104–4113, 2016. 2

[13] Zachary Teed and Jia Deng. Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras. In *NeurIPS*, 2021. 2