Fisheye Stereo Camera-based Distance Measurement Tool for Desktop and Virtual Reality

Shovan Shakya, University of South Florida

Mentors: Eric Martin, Drew Bewley, Steve Haddock

Summer 2023

Keywords: Computer Vision, Depth Map, Augmented Reality

ABSTRACT

Stereoscopic cameras have been used by MBARI for various usages such as distance estimation and 3D reconstruction. Fisheye Stereo Cameras can provide a wider and hemispherical field of view that allows for more immersive Augmented Reality experience. This paper discusses the development of an application that combines the abilities of stereo camera systems like distance estimation using fisheye lenses that can be used in a desktop environment as well as with a VR headset both for real time as well as post dive analysis.

INTRODUCTION

Using Virtual Reality or Augmented Reality for piloting UROVs has been a debated and discussed topic. There is a lot of promise in using Virtual Reality headsets for piloting UROVs due to their immersive display capabilities. VR headsets and related software are already being used to train pilots and control drones in both civil and military applications. A team from Olin College of Engineering was able to integrate MBARI's fisheye stereoscopic cameras with VR and include live telemetry information as UI overlay. Stereoscopic cameras can also be used to determine distance. The project's goal was to see whether a distance estimation tool developed using this stereoscopic camera can be integrated into the VR System or be made as a separate desktop application. Performance metrics also needed to be tested, such as whether the tool can be used in Realtime Feed or on pre-recorded videos and what the frame rates are. Usability based on feedback from the ROV pilots as well as scientists was also a necessary part of the process.

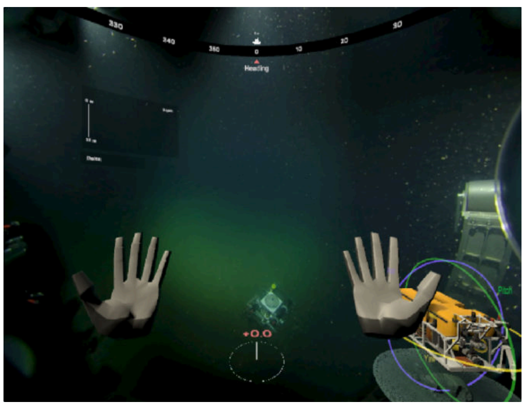


Figure 1 (left): Olin Team's VR UI. **Figure 2 (right):** Testing VR in test tank.

MATERIALS AND METHODS

CAMERA SETUP

The custom camera system utilizes two Z-Cam E2G cameras, outputting an HDMI source with a resolution of 3840 x 2160 @ 29.97 frames per second and is fitted with a Fujinon C-mount Fisheye lens with a super wide lens of 185 degrees.

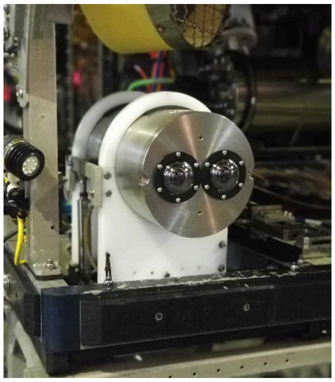

Figure 3: Camera

STEREO CALIBRATION

Stereo calibration is the first step, and its accuracy determines how good the disparity map as well as the distance estimation is. The raw feed from the cameras is then rectified using values from the calibration.

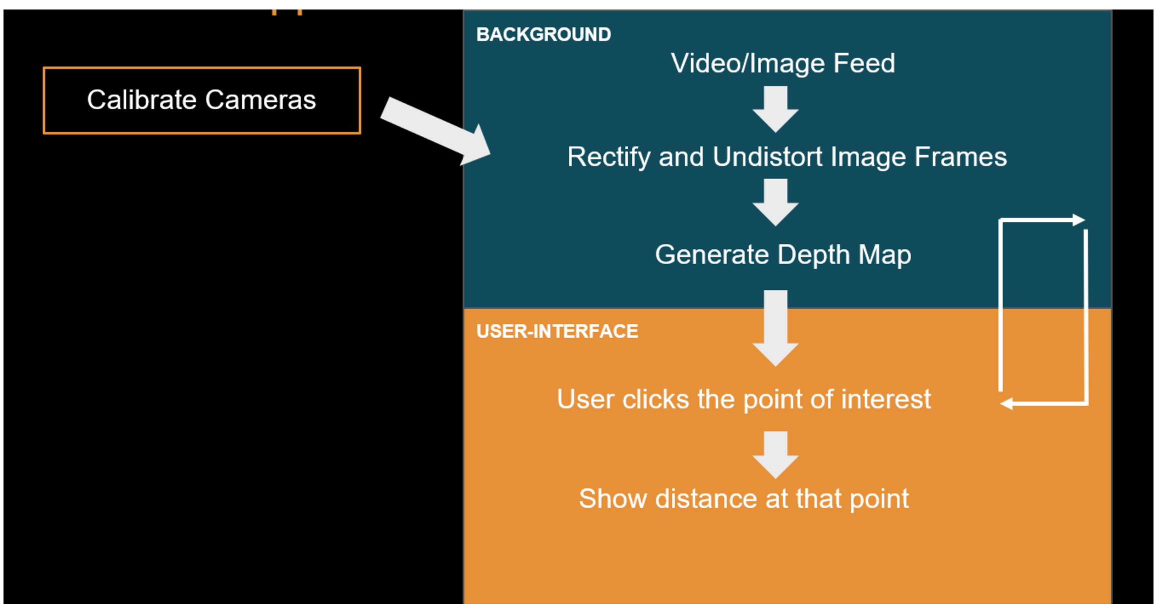

Figure 3: Overview of workflow

OpenCV stereo calibration begins by capturing images of a known pattern (usually a chessboard) from two cameras simultaneously. Each camera's intrinsic parameters (focal length, principal point, and distortion) are estimated, and then both cameras are jointly calibrated to find their extrinsic parameters (rotation and translation). This produces the essential and fundamental matrices, after which images can be rectified so that corresponding points lie on the same horizontal line for stereo matching and 3D depth estimation.

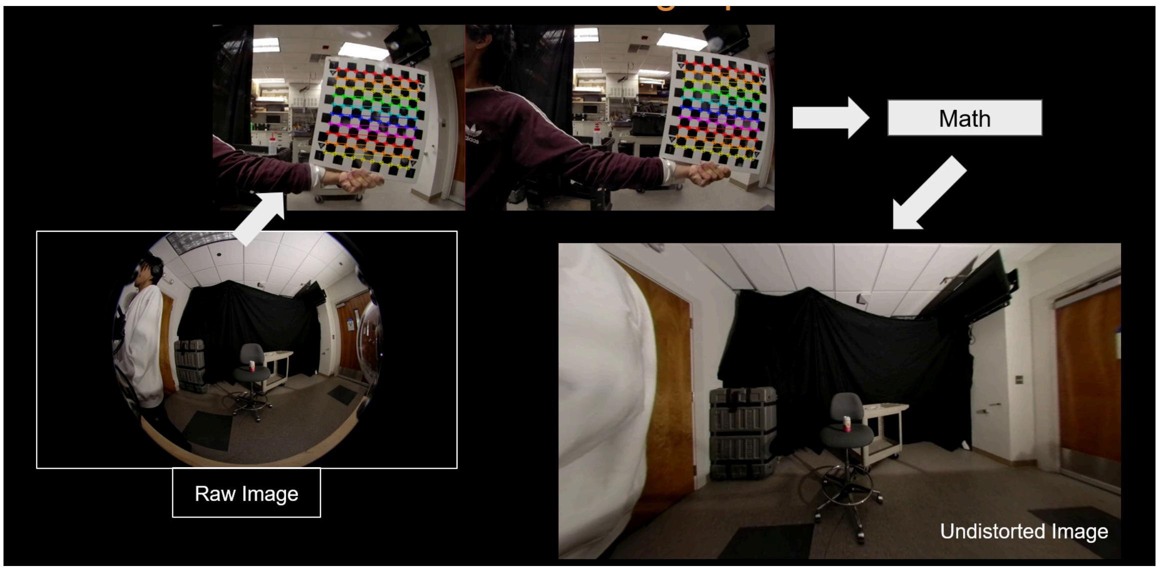

Figure 4: Calibration process

DISPARITY
In stereo vision, disparity is the difference in x-coordinates of corresponding features in the left and right images. OpenCV provides stereo matching methods (e.g., block matching or semi-global matching) to compute a disparity map from these images. Once disparity is known, distance (depth) for each pixel can be calculated using camera parameters such as the focal length and the distance (baseline) between the two camera centers. This theory is applied in our approach to determine the distance.

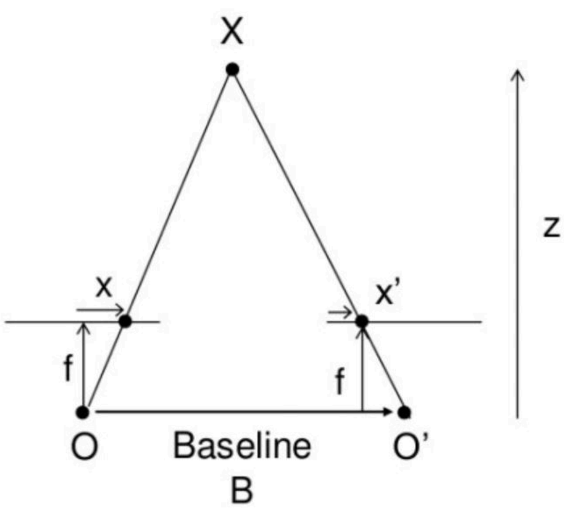

Figure 5: Calibration process

parameters such as the focal length and the distance (baseline) between the two camera centers. This theory is applied in our approach to determine the distance.

Using OpenCV's window method, we are able to add UI elements such as being able to point and click on the screen at the desired location to retrieve the distance.

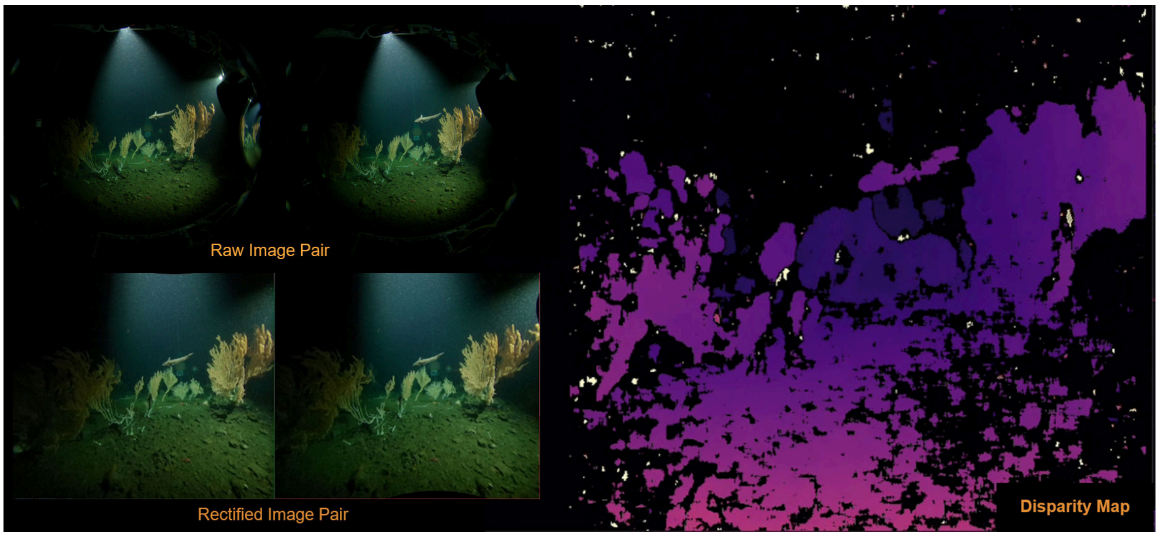

Figure 6: Disparity Map


Figure 7: OpenCV distance window

FATHOMNET TRACKING

FathomNet is an open, expert-labeled image database of marine life from MBARI, designed to advance computer vision and machine learning research for understanding and conserving ocean ecosystems. A YOLO tracking model pretrained on FathomNet was also integrated as part of automated distance labelling on detected animals, which allowed for retrieval of pixel coordinates from the bounding box which was then fed into the input of the disparity distance calculation.

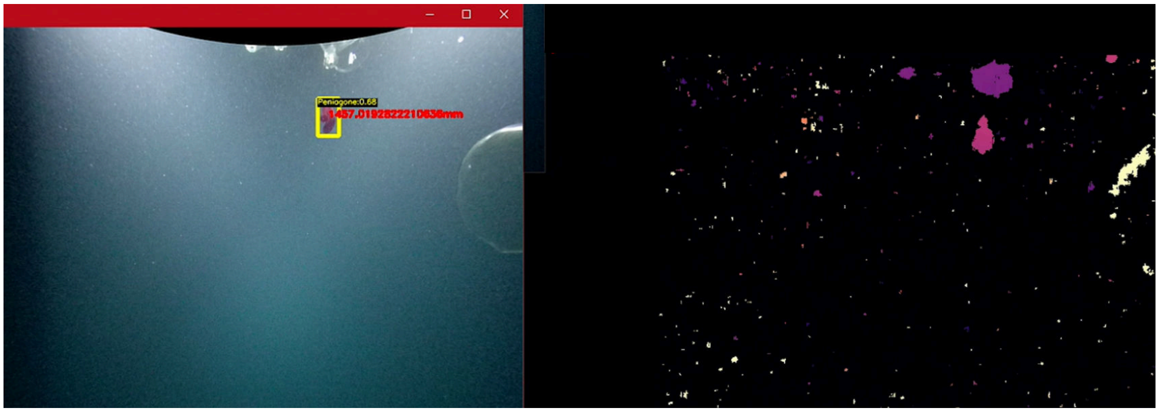

Figure 7: FathomNet Tracking Integration

VR INTEGRATION

Since the Python OpenCV was running separately from the Unity System, sockets were used to transfer disparity frames from the Python program to the Unity System. The disparity frames were then overlaid on the camera feed to provide an intuitive method to understand the disparity map in VR. Addition of UI elements such as being able toggle the depth map, and adjusting transparency were also added.

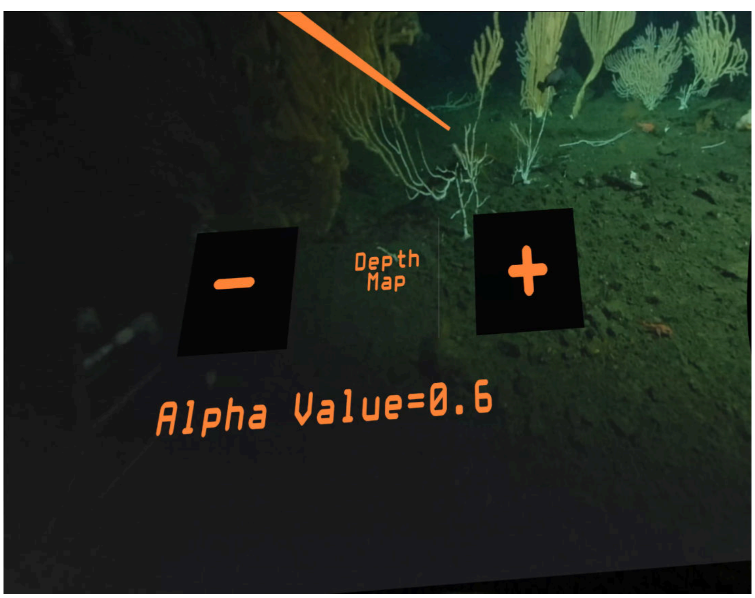

Figure 8: Unity UI addition

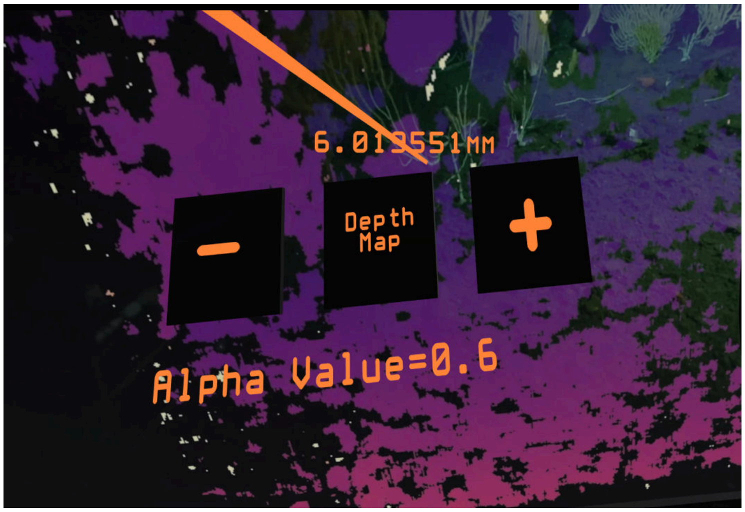

Figure 9: Depth Map VR

RESULTS

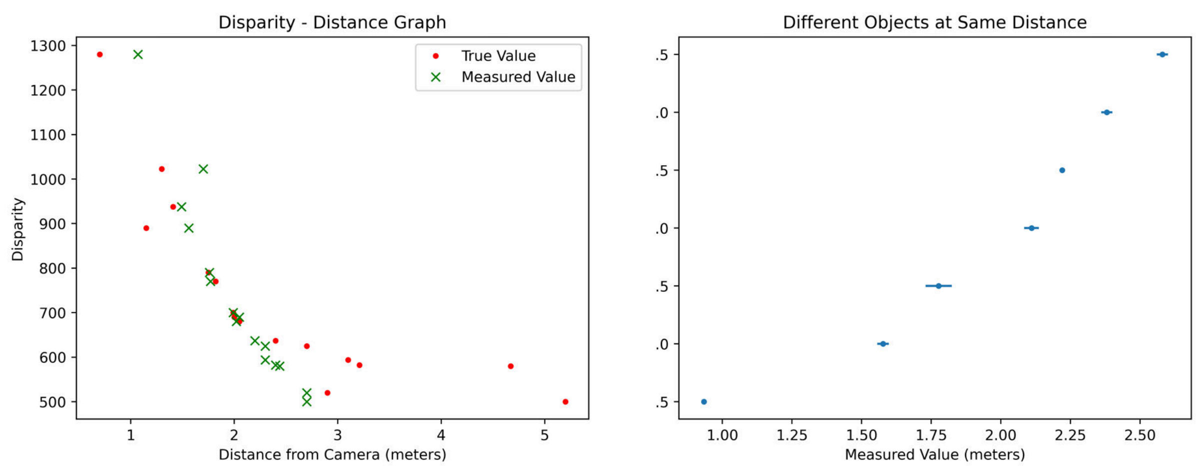

Figure 10: Estimated Distance vs True Distance

The performance of the distance estimation tool was evaluated using multiple datasets collected both on land and underwater. The primary metric for success was the accuracy of the measured distance compared to the true distance from the camera system.

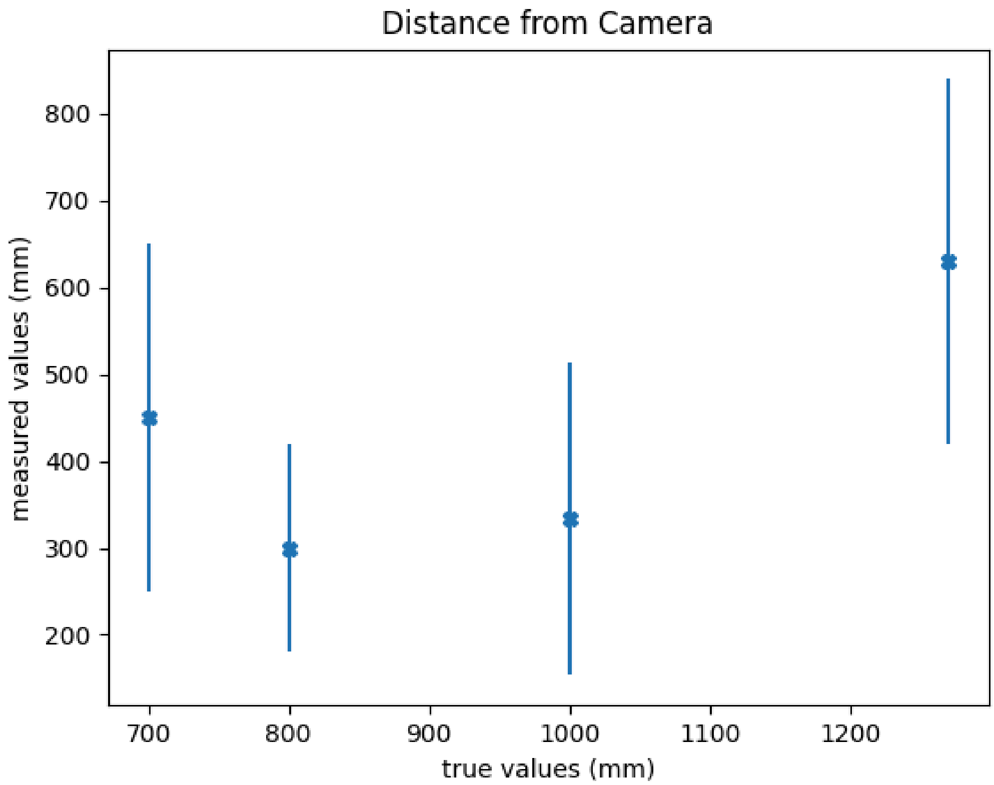

Figure 11: Variance in measured values

Analysis of the collected data, as illustrated in the Disparity Distance Graph (Figure 10), reveals a strong correlation between the system's calculated distance and the true distance for objects up to 3 meters away. Within this range, the system demonstrated a consistent variance of approximately 0.05 meters, which was influenced by the size and texture of the target objects. Figure 10 (right) and Figure 11 further illustrate that measurement precision decreases as the distance to the object increases, with error margins becoming too high for reliable measurements beyond the 3-meter threshold.

The system was successfully implemented for use with both real-time video feeds and pre-recorded video files, confirming its versatility for live piloting and post-dive analysis as outlined in the project goals. Performance was sufficient for interactive use in both desktop and VR applications. The integration of the FathomNet-trained YOLO model was also successful, enabling the system to automatically detect and display distance measurements for marine animals within the video frame, as shown in Figure 7.

Furthermore, the disparity data was successfully transmitted from the Python backend to the Unity VR environment using sockets, allowing for a real-time, interactive depth map overlay that the user could toggle and adjust (Figure 9).

DISCUSSION

The results confirm that the fisheye stereo camera system is an effective tool for accurate distance measurement in near-field applications, specifically within a 3-meter range of the cameras. The drop off in accuracy beyond this distance is an expected limitation of stereo vision systems. As the distance to an object increases, the disparity value, the difference in the object's position between the left and right images decreases. At distances greater than 3 meters, this value becomes so small that minor pixel level errors in the stereo-matching algorithm are magnified into significant errors in the final distance calculation.

This 3-meter operational range is highly relevant for the intended application of piloting Unmanned Remotely Operated Vehicles (UROVs). Many critical ROV tasks, such as manipulating robotic arms, collecting biological or geological samples, and performing detailed visual inspections of seafloor habitats, occur within this near field envelope. Providing pilots with accurate, intuitive distance information in this range can enhance spatial awareness, improve operational efficiency, and reduce the risk of collision.

The integration of the tool into an immersive VR headset is particularly promising, as it offers a more natural and intuitive understanding of the 3D environment than a traditional 2D monitor. While the current system functions in realmtime, its performance could be enhanced by offloading the computationally intensive disparity calculations and animal tracking to the GPU. Such an improvement would be a step towards integrating a real-time point cloud into the VR environment, which would provide pilots with a true volumetric reconstruction of the underwater scene rather than just a 2D overlay. The successful integration of FathomNet tracking also demonstrates a powerful application for scientific analysis by automating the process of measuring distance to tagged organisms.

CONCLUSIONS/RECOMMENDATIONS

This project developed a fisheye stereo camera tool capable of accurately measuring distances for both desktop and virtual reality applications. The system is reliable for near field tasks, providing accurate measurements up to 3 meters.

For future improvements, we recommend the following:

- **Improve Range:** To increase accuracy beyond 3 meters, a camera system with a wider baseline (distance between cameras) should be explored.
- **Increase Performance:** Integrate GPU-enabled disparity calculation and animal tracking to increase real-time performance and reduce latency.
- **Enhance VR Interface:** Implement hand and eye tracking as user interface methods for a more intuitive and seamless control scheme within the VR environment.
- **Integrate Point Cloud:** Develop the capability to generate and integrate a real-time point cloud into the VR headset, providing users with a complete 3D model of the immediate environment.
- **Expand Functionality:** The FathomNet integration could be expanded to automatically calculate the size of detected organisms, in addition to their distance, providing a more comprehensive analytical tool for researchers.
- **Conduct Usability Testing:** Perform formal usability studies with ROV pilots and scientists to gather feedback and refine the user interfaces in both the VR and desktop environments.

ACKNOWLEDGEMENTS

I would like to extend my sincere gratitude to my mentors, Eric Martin, Drew Bewley, and Steve Haddock, for their invaluable guidance and support throughout this project. A special thank you is also extended to George Matsumoto for organizing the summer internship program, Amy Phung from the WHOI-MIT Joint Program for her assistance with the existing ROVVR software, Kevin Bernard for providing the FathomNet dataset, and to the entire MBARI ROV Piloting Team and ship crew for their operational support and insightful feedback.

REFERENCES

1. Phung, A., Wierzbanowski, C., Gonzalez, E., Shuster, N., & Lu, E. (2021, April). Developing a Control Room in Virtual Reality (VR) to Improve Underwater Remotely Operated Vehicle (ROV) Piloting [Poster presentation]. 25th Annual Conference of the Consortium for Computing Sciences in Colleges, Northeastern Region (CCSCNE).
2. OpenCV. (n.d.). *OpenCV documentation*. Retrieved from https://docs.opencv.org
3. Katija, K., Orenstein, E., Schlining, B. et al. FathomNet: A global image database for enabling artificial intelligence in the ocean. Sci Rep 12, 15914 (2022). https://doi.org/10.1038/s41598-022-19939-2
4. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*
5. Unity Technologies. (n.d.). Unity Real-Time Development Platform. Retrieved from https://unity.com