ATSEA: Augmented Telepresence in robot SEAfloor plastic cleanup

1st Aaron Lee Smiles
School of Electronic Engineering and Computer Science
Queen Mary University Of London
London, UK
a.l.smiles@qmul.ac.uk

2nd Changjae Oh
School of Electronic Engineering and Computer Science
Queen Mary University Of London
London, UK
c.oh@qmul.ac.uk

3rd Ildar Farkhatdinov
School of Engineering and Materials Science
Queen Mary University Of London
London, UK
i.farkhatdinov@qmul.ac.uk

***Abstract*—Not enough robots are used in the fight against ocean plastics. ROVs controlled by human operators can be used to collect ocean waste, but human depth perception underwater is obscured, especially on 2D screens. This can be made worse when objects are transparent and backgrounds are near blank, common in underwater scenes. We present work on an underwater augmented, or mixed reality, telepresence system that utilises stereo vision for 3D plastic bottle object detection and audiovisual aids to assist with ROV telemanipulation during plastic waste collection. Initial tests with an unsynchronised pair of stereo cameras showed that block matching algorithms had difficulty finding matches with transparent materials. Using a synchronised stereo camera and a YOLO-based model, qualitative results are promising. We hypothesise that with our mixed reality telepresence system we can increase seafloor plastic bottle collection against traditional ROV end effector collection operation whilst reducing cognitive load. Supplementary material, including demo videos, can be found here: https://github.com/aaronlsmiles/ATSEA.**

***Index Terms*—ocean plastics, stereo vision, mixed reality, telepresence, object detection, ROVs**

I. Introduction

Most of the current efforts to clean up ocean plastics have focused on surface litter, which accounts for only 1% of all ocean plastic. An estimated 99% of ocean plastics are below the surface or sink to the seafloor, according to the TOPIOS (Tracking Of Plastic In Our Seas) project[1]. Plastic bottles are a big contributor, with around 8 million tons of them entering the ocean each year.

To properly clean our oceans, robots will need to be involved. No one solution will solve the issue, and both AUVs and ROVs will be required, since each has its advantages and disadvantages. The SeaClear project[2] deploys collaborative AUVs for cleanup, while the Rozalia Project[3] deploys a small ROV that is controlled by a human operator with a system that includes a small 2D monitor. While [1] make the case for AUVs, we argue that ROVs are overlooked and human intervention will be required in a number of situations.

[1]TOPIOS: Tracking Of Plastic In Our Seas, European Research Council Starting Grant, 2017-2022, http://topios.org/

[2]SEACLEAR: Search, Identification and Collection of Marine Litter with Autonomous Robots, Horizon 2020 grant agreement No 871295, https://seaclear-project.eu/

[3]Rozalia Project, https://www.rozaliaproject.org/innovation-technology

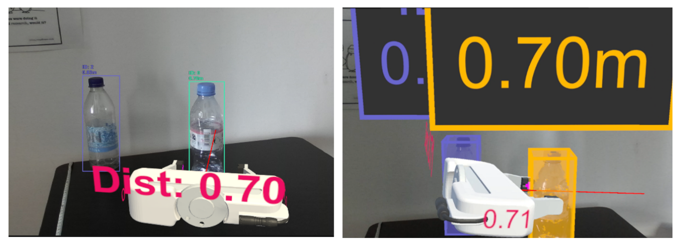

Fig. 1. Demonstration of system in air. Left: 2D custom detector in Unity showing 2D bounding boxes of detected bottles and AR end effector (recorded from 2D screen). Right: 3D custom detector in Unity showing 3D bounding boxes and AR end effector (recorded from VR HMD screen).

The biggest contributor to plastic waste is ghost nets, in which other plastic waste, as well as sea creatures, can get caught. These pose a significant entanglement risk to underwater vehicles too, which would require human intervention. The manipulators on ROVs would be suited to cutting the robot and potentially marine life free of any entanglement.

We interviewed ROV operators and found that they report the "underwater illusion" that affects their sense of depth underwater, especially on 2D screens. Our hypothesis is that this would be further affected when dealing with transparent objects, such as plastic bottles, and aim to answer the question of how we can improve the efficiency of transparent object collection with ROV manipulators using a mixed reality (MR) telepresence system that utilises stereo vision and object detection for providing MR audiovisual (AV) aids.

We focus on techniques to support manual collection using ROVs, but our clear bottle perception research will be useful to future autonomous research, such as reinforcement learning techniques for mapping seafloor litter [2].

Models have been deployed for detecting sub-surface marine litter. The authors [1] focus on litter floating in the water column, whereas, [3] focus on seafloor plastics.

We have developed a MR system for simulated plastic waste collection with a virtual end effector (EE) using synchronised stereo vision and a convolutional neural network (CNN) for real-time bottle detection (YOLOv7-tiny) in Unity. Demo videos can be found at: github.com/aaronlsmiles/ATSEA.

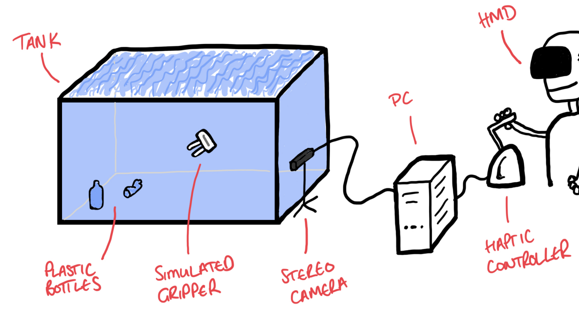

Fig. 2. System overview drawing depicting a tank containing water and plastic bottles (or other ocean waste), the simulated gripper, stereo camera facing into the tank from outside, which is connected to a PC running Unity and the ZED SDK, which is also connected to the haptic controller and VR HMD.

Fig. 3. Sample of synthetic underwater images (middle and right columns) created from one of the air stereo images (top, left) and its depth map (bottom, left) using [5].

II. Implementation

A. System Overview

The stereo vision-based teleoperation interface built for this study comprises a ZED Mini, ZED SDK, Unity, Geomagic Touch, and a virtual reality (VR) head-mounted display (HMD). A sketch (Fig. 2) shows the underwater experimental setup. The stereo camera will give an egocentric view into a tank that can be viewed through a VR HMD, which emulates the view of an ROV operator. The tank contains physical objects, and the haptic-controlled simulated gripper is augmented in the scene.

B. Experimental Protocol

For user studies, the IPD of each participant will be checked to set the HMD lens distance correctly to avoid skewed results from potentially blurry views in VR. The task time limit (1 minute) will remain the same across studies.

The task goal will be to reach each bottle with the centre of the end effector (EE) and click a button on the haptic controller to emulate the gripper being triggered to close/grasp. This must be done without going too far, that is, knocking the object (Figure 37, right), and will be performed in different scenarios:

i) 2D screen with haptic controller but without visual aids,
ii) HMD with haptic controller and no visual aids,
iii) HMD with haptic controller and visual aids.

Upon completion of each task, participants will complete a usability and cognitive load assessment/survey. The system will also record the following measurements:

- Time to complete: total time taken to complete task(s).
- Failed attempts count: number of times the participant either clicked the gripper close button when the object is not within the gripper fingers or the EE is positioned within the object, signifying that the object is knocked.
- Fail attempt type: 0) incorrect grasp area, 1) knocked object.
- Time-out fail: failed to complete task(s) within the time limit.

In-air baseline tests will be performed, followed by underwater tests. Cameras are optimised for viewing underwater scenes through glass, and the same user tests are repeated in the underwater scene with objects submerged in the tank.

III. Conclusion

We [4] first tested an unsynchronised stereo vision system on transparent and opaque plastic and metal materials. The results showed that opaque bottles were much easier to detect than transparent bottles with an RMSE of 1.01 cm for opaque plastic bottles and 2.03 cm for opaque metal bottles, but 16.34 cm for transparent plastic bottles.

The newly developed synchronised stereo system was qualitatively assessed, as can be seen in Fig. 1. The distance of the EE from the stereo cameras matched that of the object detector distances. This is better illustrated in the 3D version, where the occlusions are present and demonstrate that the simulated object's environment scaling matches that of the real world. However, we hypothesise that in-water tests will present less accurate detection results than in-air, hence, we will retrain the detection model with synthetic underwater images of bottles created using [5], as shown in Fig. 3.

Furthermore, initial qualitative tests support the hypothesis that 3D displays provide enhanced perception, as it is much easier to evaluate the position of the EE in the VR HMD, even without the visual aids. However, this will be validated in the user studies that will form part of the final paper.

References

[1] M. Fulton, J. Hong, M. J. Islam, and J. Sattar, "Robotic detection of marine litter using deep visual detection models," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE.

[2] M. Rosynski and L. Buşoniu, "A Simulator and First Reinforcement Learning Results for Underwater Mapping," *Sensors*, vol. 22, no. 14, p. 5384, Jan. 2022.

[3] D. V. Politikos, E. Fakiris, A. Davvetas, I. A. Klampanos, and G. Papatheodorou, "Automatic detection of seafloor marine litter using towed camera images and deep learning," *Marine Pollution Bulletin*, vol. 164, p. 111974, 2021.

[4] A. L. Smiles, K. D. Chavanakunakorn, B. Omarali, C. Oh, and I. Farkhatdinov, "Implementation of a Stereo Vision System for a Mixed Reality Robot Teleoperation Simulator," in *Towards Autonomous Robotic Systems*, ser. Lecture Notes in Computer Science. Springer Nature Switzerland, 2023, pp. 494–502.

[5] J. Wen, J. Cui, Z. Zhao, R. Yan, Z. Gao, L. Dou, and B. M. Chen, "SyreaNet: A Physically Guided Underwater Image Enhancement Framework Integrating Synthetic and Real Images," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023.