Towards a Simulator for Egocentric Robotic Avatar Manipulation in Extreme Environments

Aaron Smiles[1], Changjae Oh[1], Ildar Farkhatdinov[1]

***Abstract*— Robotic manipulators can be difficult to remotely operate in certain situations, such as adverse weather (heavy rain, snow, fog, or glare from the sun), poor lighting, or visual impairment of the operator. Another scenario is underwater. Remotely operated vehicle (ROV) pilots at the National Oceanography Centre (NOC) report the *underwater illusion*: an underwater visual effect that causes them to misjudge distances to objects with the end effector. ROV pilots can benefit from mixed reality (MR) ques to aid manipulation tasks. However, performing experiments in extreme or deep sea environments is difficult, so simulated environments are required. The ROV used by the NOC features an imaging sonar and two main manipulator cameras. This extended abstract describes early work on stereo vision for informed augmented reality distance ques and future directions into 3D reconstruction and a telerobotic synthetic environment. Preliminary stereo vision results are promising for gauaging distance to objects in a scene, which will be integrated into the simulation and validated in user tests.**

I. INTRODUCTION

The Internet of Skills [1] proposes the ability for expert skills to be *teleported* to remote, inaccessible, or hostile places. This means that specialist skills could be accessed anywhere, at any time. This also means that critical tasks can be performed in extreme environments without risk to human life e.g., search and rescue missions during natural disasters, surgeons performing emergency operations on patients in other countries, space engineers performing repairs to space stations without needing to do a spacewalk, or ocean engineers performing benthic station maintenance in turbulent waters without leaving shore [2], [3].

Mixed reality could revolutionise traditional telerobotics by providing increased spatial and situational awareness, reducing cognitive load [4]. However, performing experiments in hostile envirnoments or under adverse conditions is risky, due to the human life involvement and use of expensive equipment. Hence, simulation is the ideal training and test bed, but realistic simulators in hostile environments are under-researched [5].

Realistic simulation will require real world physics and replica sensor data. The remotely operated vehicle (ROV) used by the National Oceanography Centre (NOC) features two main forward cameras for manipulation tasks, as well as an imaging sonar. To begin with this research investigates stereo vision to triangulate distance information from objects to display on screen.

*This work was supported by the ESPRC CDT Data-Centric Engineering

[1]School of Electronic Engineering and Computer Science, Queen Mary University of London, UK
{a.l.smiles,c.oh,i.farkhatdinov}@qmul.ac.uk

Fig. 1. Stereo vision AR system overview: a) stereo images rendered as raw image game objects in Unity from a stereo pair of Logitech C505 cameras, b) OpenCV processed camera data broadcast as a virtual camera using OBS into Unity Vuforia plugin as an AR camera, c) PluraView stereoscopic screen and Geomagic Touch haptic controller, d) Unity 3D model of manipulator with laser pointer distance from object

II. METHOD

Fig. 1 shows key components of the proposed system. The stereo camera images (a) are used for performing stereo vision applications. The stereo images are first processing in OpenCV, then sent to Unity as virtual cameras via OBS (b). The scene will be displayed in 3D on a stereoscopic screen (c) and the haptic controller (right side of the screen) is used to control the simulated robotic arm (d). The 3D model of the robotic manipulaotr will be translated into augmented reality (AR) using the Vuforia Unity plugin. The OpenCV-Unity system will be integrated using sockets and virtual camera middleware.

The stereo vision system is needed to obtain distance from objects using two parallel cameras and consists of the following steps [6], [7]:

- **Camera calibration**: Capture a batch of checkerboard (or fiducial marker) images from different angles to

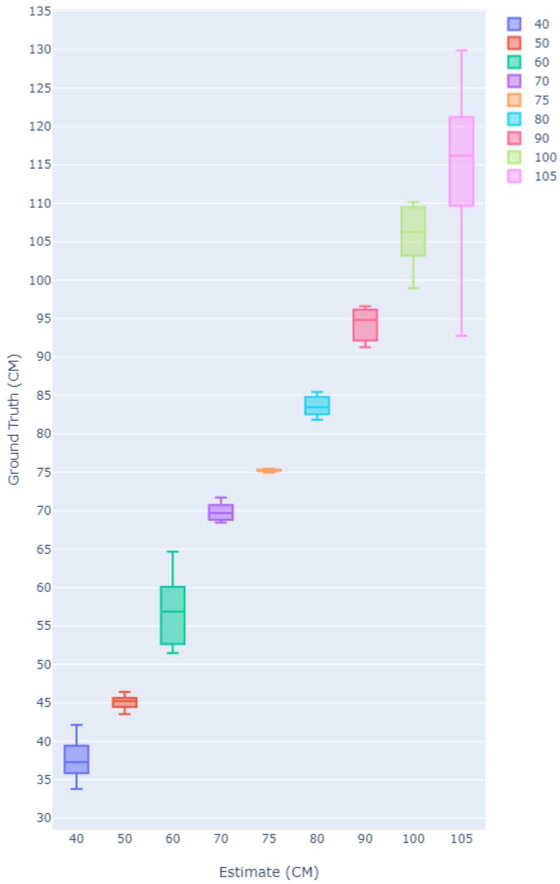

Fig. 2. Results from testing the stereo vision system to estimate the distance of an object at different distances from the disparity map

infer the focal length and optical centres of your camera (intrinsics). The calibration should not be *biased* towards any corner or region, so we calibrate with images of the chessboard everywhere in the field of view.

- **Rectify and undistort images**: Remove lens distortion.
- **Feature matching**: Look for similar features (horizontally) between left and right images.
- **Disparity map (to distance)**: Create a depth map that we can use to calculate distance Z of objects in the scene, obtained using the focal length f, baseline stereo camera distance B, and disparity d in equation (1) [7].

$$Z = \frac{fB}{d} \tag{1}$$

The PluraView monitor creates a 3D image by combining two images into a cross-polarised image. The images are displayed from different fixed angles and combined by a beam-splitter mirror in the center, which is viewed with passive polarised filter glasses [8]. The stereo camera images are used here to create a 3D egocentric view of a scene.

Touch is a device that utilizes motorized force feedback on the user's hand, while also allowing 6-degree positional sensing for the manipulation of 3D objects on a screen [9]. This will be used to control the position of the end effector in 3D space and to open/close the gripper.

Fig. 3. Unity AR model of robotic manipulator from an egocentric perspective

III. RESULTS

The final disparity to depth calculations (Fig. 2) obtained promising results with an RMSE of 6.6cm between 40-105cm distance range. However, at the *sweet spot* of 75cm, the accuracy showed an error of only 31mm.

ROV manipuators typically have a maximum reach of 2m and are usually positioned to work within the middle of this range, so the results shown are encouraging for ROV manipulation.

The AR model (Fig. 3) is not accurate when the desired position is outside the scope of the *image target* used by Vuforia to map Unity game objects in AR using webcams.

IV. DISCUSSION AND FUTURE WORK

Stereo vision was conducted using unsyncronised cameras, since ROV cameras are separated. The stereo camera baseline used here was 71mm which is good for obtaining short distance depth. The cameras on the ROV are over a meter apart. This would need to be considered when deploying on mission and the camera arrangement adjusted for the purpose of stereo vision. However, further investigations will be made using similar arrangements, as well as using RGBD (Microsoft Kinect) data for point clouds: for i) stablising the AR model, ii) 3D reconstruction (as replica sonar data).

REFERENCES

[1] M. Dohler, T. Mahmoodi, M. A. Lema, M. Condoluci, F. Sardis, K. Antonakoglou, and H. Aghvami, "Internet of skills, where robotics meets AI, 5g and the tactile internet," in *2017 EuCNC*.

[2] M. Wonsick and T. Padir, "A systematic review of virtual reality interfaces for controlling and interacting with robots," vol. 10, no. 24, p. 9051, publisher: Multidisciplinary Digital Publishing Institute.

[3] M. de la Cruz, G. Casañ, P. Sanz, and R. Marín, "Preliminary work on a virtual reality interface for the guidance of underwater robots," vol. 9, no. 4, p. 81.

[4] S. Livatino, D. C. Guastella, G. Muscato, V. Rinaldi, L. Cantelli, C. D. Melita, A. Caniglia, R. Mazza, and G. Padula, "Intuitive robot teleoperation through multi-sensor informed mixed reality visual aids," vol. 9, pp. 25 795–25 808.

[5] E. Bonetto, C. Xu, and A. Ahmad, "GRADE: Generating realistic animated dynamic environments for robotics research," 2023. [Online]. Available: 2303.04466 [cs]

[6] K. Sadekar, "Depth estimation using stereo camera and opencv (python/c++)," p. 2021. [Online]. Available: https://learnopencv.com/depth-perception-using-stereo-camera-python-c/

[7] A. Kaehler and G. Bradski, *Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library*. Reilly Media, Inc, 2016.

[8] 3d PluraView - 4k monitor, 3d-stereo display - specifications 3d PluraView. [Online]. Available: https://www.3d-pluraview.com/en/specifications

[9] Touch. [Online]. Available: https://www.3dsystems.com/haptics-devices/touch