Implementation of a Stereo Vision System for a Mixed Reality Robot Teleoperation Simulator

Aaron Lee Smiles(✉), Kitti Dimitri Chavanakunakorn, Bukeikhan Omarali, Changjae Oh, and Ildar Farkhatdinov

Queen Mary University of London, London, UK
{a.l.smiles,c.oh,i.farkhatdinov}@qmul.ac.uk
https://www.robotics.qmul.ac.uk/

Abstract. This paper presents the preliminary work on a stereo vision system designed for a mixed reality-based simulator dedicated to robotic telemanipulation. The simulator encompasses a 3D visual display, stereo cameras, a desktop haptic interface, and a virtual model of a remote robotic manipulator. The integration of the stereo vision system enables accurate distance measurement in the remote environment and precise visual alignment between the cameras' captured scene and the graphical representation of the virtual robot model. This paper delves into the technical aspects of the developed stereo system and shares the outcomes of its preliminary evaluation.

Keywords: Teleoperation · Stereo Vision · Mixed Reality

1 Introduction

Augmented, virtual, and mixed reality have the potential to enhance the quality of human-machine interfaces in robot teleoperation. For example, virtual reality-based control interfaces combined with depth cameras can improve scene understanding [10,12], allow efficient workspace mapping for telemanipulation tasks [11], and provide easier dexterous tactile telemanipulation [5]. It was also demonstrated that using multiple visual displays and interfaces increases the cognitive load demand during teleoperating remote underwater vehicles [8]. A taxonomy of mixed reality (MR) based robot teleoperation modes for hazardous environments was proposed in [16] demonstrating MR advantages for specific use-case scenarios.

Training human operators to efficiently and safely complete remote manipulation tasks is an important aspect to which MR can contribute [1,4,7,14]. Usually such teleoperation training simulators use virtual reality (VR) to model the remote scene and the robot. However, a more enhanced training experience can be achieved if a real visual stream (live or recorded) of the remote environment can be combined with the virtual model of the robot using augmented reality

F. Iida et al. (Eds.): TAROS 2023, LNAI 14136, pp. 494–502, 2023.
https://doi.org/10.1007/978-3-031-43360-3_40

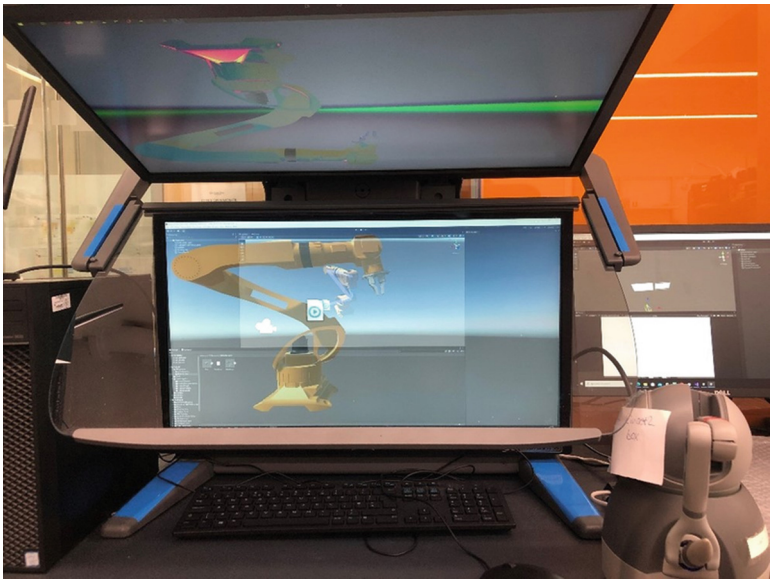

Fig. 1. 3D PluraView stereo display as part of the experimental setup with Touch haptic controller for teleoperating the virtual model of a remote robot-manipulator. The visual scene on the screen shows a combined view of the virtual robot and the stereo view from the remote environment or its emulator.

(AR), MR and stereo vision. Stereo vision allows improved scene understanding through the addition of depth perception and enables correct virtual robot geometry and motion alignment with the visual representation of the remote environment as demonstrated in surgical vision applications [9].

The aim of this work is to present a stereo vision-based robot teleoperation simulator and validate the stereo vision component of the system. The stereo vision system is based on earlier works [2,13] and is extended to include egocentric stereo vision and an MR-based robot simulator. The materials presented in the following sections are preliminary work demonstrating the integration of the stereo vision system with a virtual robot teleoperation simulator.

2 Implementation

2.1 System Overview

The stereo vision-based teleoperation interface built for this study is shown in Fig. 1. It comprises a stereo display to represent video flow from a remote robot's cameras (PluraView 3D[1]), a desktop haptic interface (Touch haptic interface[2]) and a custom-designed MR-based robotic manipulator simulation. The aim of

[1] www.3d-pluraview.com.
[2] www.3dsystems.com/haptics-devices/touch.

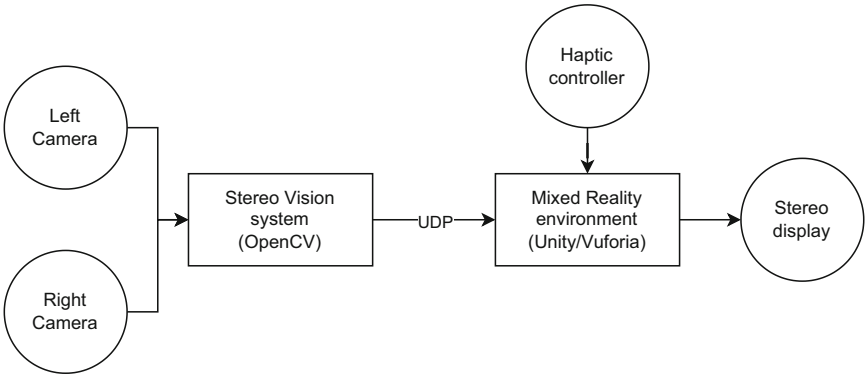

Fig. 2. High-level system diagram describing the flow of the system. The two mono camera inputs are used to create the stereo vision system in OpenCV, which sends the processed images and distance estimate data via UDP sockets into Unity. The Unity environment renders a simulated robotic end effector, which can be controlled via the haptic controller, in AR using the Vuforia plugin in the images from the webcams. These augmented camera images are then displayed on the stereo display, including the distance estimate, which can be viewed in 3D.

the proposed simulator is to provide a robot teleoperation training environment that combines a virtual remote robot model with the stereo visual flow (live or pre-recorded) from a remote physical environment. Such a simulation system provides a realistic training environment as a stereo representation of the remote physical environment is used, as well as the overall system is low-cost and safe as a virtual model of the robot is used.

A high-level flow diagram depicting the main components and inputs/outputs is shown in Fig. 2. The proposed operation workflow for the stereo vision subsystem consists of first obtaining camera images from a pair of stereo webcams. The stereo images rectified with OpenCV are sent to the Unity 3D platform used for rendering the stereo view of a remote scene is displayed using the 3D PluraView stereoscopic screen. Desktop Touch haptic controller is used to operate the simulated robotic end-effector. The 3D model of the robot is rendered in augmented reality (AR) using the Vuforia Unity plugin[3]. A Model Target is generated of a real object using the Vuforia Model Target Generator[4], so the Vuforia engine can recognise the object in the camera and trigger an AR event (a Unity GameObject) that is the same, or near-identical, distance as the real object from the virtual and real cameras. The stereo distance data of the real object from OpenCV is sent to Unity via UDP, which is then deducted from the virtual end-effector distance from the virtual object.

[3] https://library.vuforia.com/getting-started/vuforia-engine-package-unity.
[4] https://library.vuforia.com/objects/model-targets.

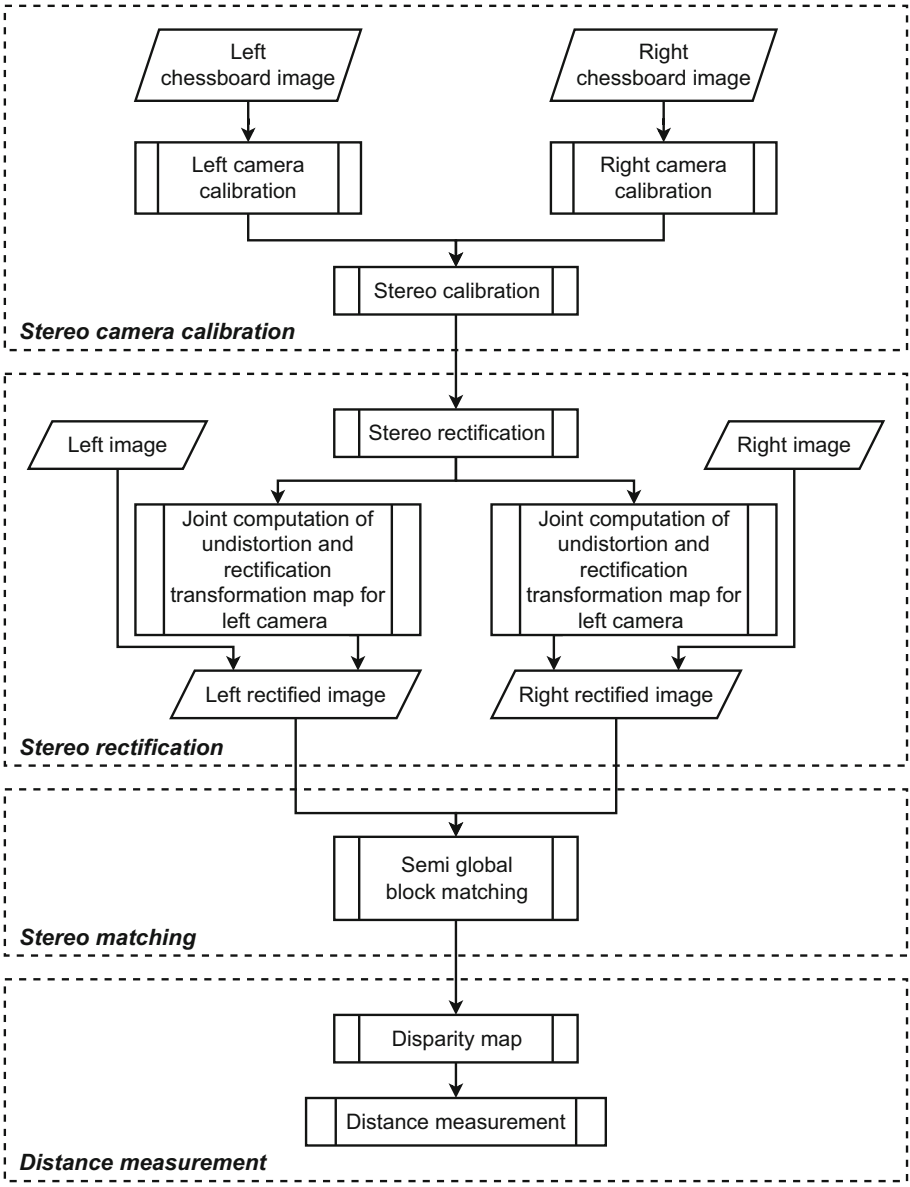

Fig. 3. Flow diagram of the proposed stereo vision distance measuring system.

2.2 Stereo Vision

The stereo vision system is used to estimate the distance to the remote scene's objects using two cameras (Logitech C505) with the optical axis aligned to be parallel. The stereo cameras were set horizontally as close as possible to the

human interpupillary distance (IPD) (average 63 mm), as this also conforms to the zone of comfort [15]. The video flow processing used in the proposed system is similar to earlier works [2,6,13] and represented as a flowchart in Fig. 3 and in the following items:

- **Cameras calibration** is performed by capturing a batch of checkerboard images from different angles to infer the focal length and optical centres of the 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.
- **Rectification and undistortion of the images** is achieved by removing barrel or pincushion lens distortion. This step rectifies the stereo images based on epipolar geometry [2], re-projecting the stereo image planes onto a common plane that is parallel to the line between the cameras.
- **Block/feature matching** is performed by searching for similar graphical features between left and right images (horizontally). Block matching is used to generate a disparity map between the left and right stereo images. Here we calculate the similarity for *each* block in the image, searching for correspondence along horizontal epipolar lines using a window, or "block", commonly a sum of absolute difference (SAD) window. The best match is chosen by the pair with the lowest SAD score.
- **Distance estimation from disparity map** produces a depth map that we can use to calculate the distance to the objects in a remote scene, z, based on the focal length, f, baseline stereo camera distance, b, and disparity, d, using $z = \frac{fb}{d}$ [6].

2.3 Stereo Display

The stereo display (3D PluraView) is used in the robot teleoperation simulator to provide realistic visual feedback with a sense of depth. Parallax images are images that are passed through your left and right eyes to create a sense of depth. All 3D stereo media involves a pair of parallax images that pass to the left and right eyes of a viewer separately and concurrently [3]. By combining two images into a cross-polarised image, the 3D PluraView monitor (Fig. 1) is able to produce stereoscopic visual feedback. The images are displayed from different fixed angles and combined by a beam-splitter mirror in the centre, which is viewed with passive polarised filter glasses. The stereo camera images are used here to create 3D egocentric views of a remote scene.

The 3D model of the robotic end effector was imported into Unity and augmented to the raw camera images using the Vuforia Unity plugin. The model of the manipulator was controlled using the haptic controller (the haptic controller's stylus position was used as a reference position for the robot's end-effector).

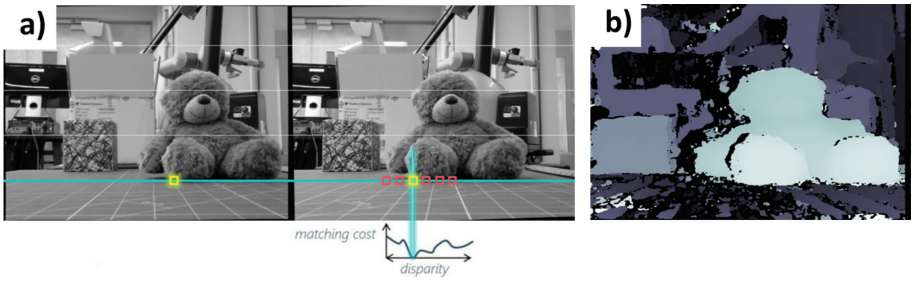

Fig. 4. a) Dense stereo block matching. The scanline (blue) is shown across the left and right images. The left image shows the target/reference block (yellow). The right image shows the scanline (blue), best matching block (yellow), and second-best matching candidates (red). The bottom image shows a plot for the Sum of Absolute Difference (SAD) between the reference block and the sliding window along the scanline in the right image. **b)** An example of stereo block matching and the disparity image obtained with OpenCV.

3 Validation of the Stereo Cameras System

An experiment was conducted to evaluate the accuracy of distance estimation to remote environments objects with the help of the implemented stereo vision system. The cameras were calibrated and the root mean squared reprojection errors were 0.0131 px for the left camera and 0.0131 px for the right, at a baseline distance of 6.3 cm. An example of stereo matching using image disparity is shown in Fig. 4b.

To validate disparity-to-depth estimation for the second study, we conducted a series of distance measurements with the system on objects of different materials as shown in Fig. 5a: a patterned cardboard box, metal bottle, clear bottle, white bottle, fabric teddy toy, and a grey 3D printed plastic part with a rough texture, almost like rough stone or concrete. A tape measure (used for ground truth) was placed in front of the cameras, and object distance measurements were taken at 18 different distances from the cameras within the range at which the system could obtain a reading, which was approximately 45–130 cm. Ten readings were taken at each distance.

The results are shown in Fig. 5. Root mean squared error (RMSE) of distance estimation for each object was: box, 14.57 cm; clear bottle, 16.34 cm; white bottle, 1.01 cm; metal bottle, 2.03 cm; teddy toy, 1.95 cm; 3D printed part 2.5 cm. The average RMSE for all objects was 9.13 cm.

Most disparity-to-depth results in the experiment showed to be normally distributed, besides some outliers due to noise in the disparity map. The clear bottle showed inconsistent readings, often returning 0 cm, or a null reading, as the disparity map had difficulty finding matches due to transparency, but performed accurately on those that it matched.

Overall, the system performed well. The aggregate RMSE was 9.13 cm, but this is skewed by the noisy pixel outliers and transparent bottle. In comparison,

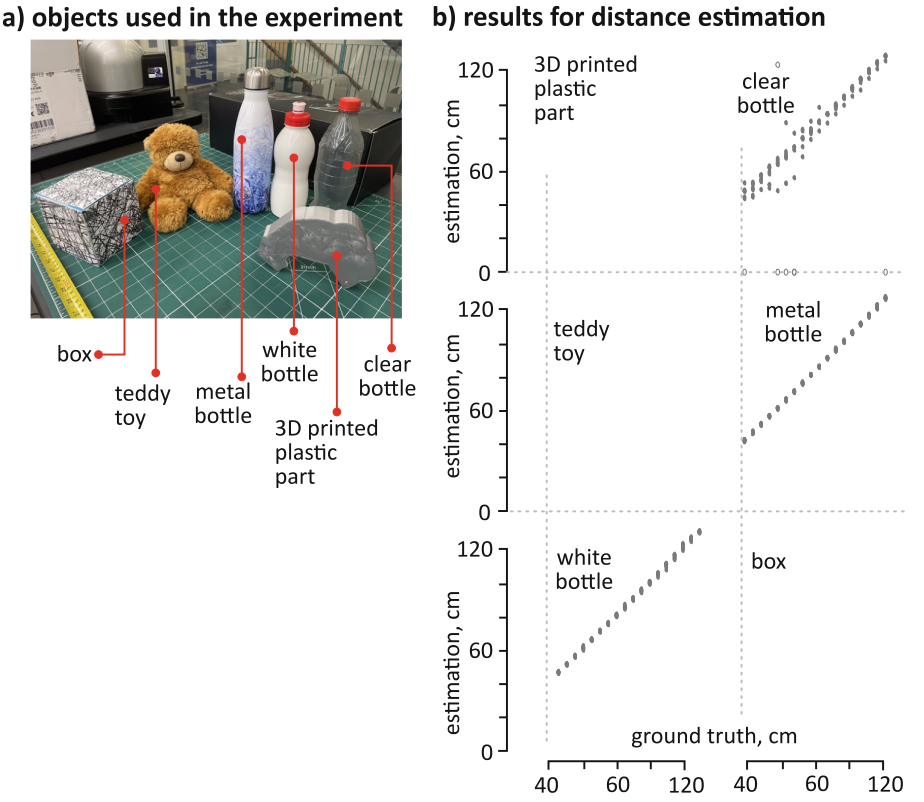

Fig. 5. a) The objects used for estimating distance using the stereo vision sub-system. From left to right: a patterned cardboard box, teddy toy, metal bottle, white bottle, clear bottle, and a 3D printed plastic part. **b)** Results from testing the stereo vision system to estimate the distance of various objects/materials at different distances.

the white bottle had an RMSE of 1.01 cm. In practice, when we track an object and estimate its distance, a mask is used, and the distance of the object is estimated from an average of the points in the disparity mask, which negates the error due to noise or gaps in the disparity map.

4 Conclusion

This paper presented an experimental prototype of a stereo vision-based simulator for robotic teleoperation. The study evaluated the performance of the stereo vision system that is integrated into the simulator.

Stereo vision was conducted using unsynchronised cameras since the ROV cameras are separated. The stereo camera baseline used here was 63 mm, which demonstrated sufficiently accurate distance estimations within 1.2 m range. Having a larger baseline distance between the cameras will increase the operational

range. For example, the cameras on the remotely operated robot at the UK National Oceanography Centre are attached over a meter apart.

While a stereo vision system was validated, validation of the teleoperation system was out of the scope of this paper. Future work will be dedicated to validating the teleoperation system through subjective user studies. Additional mixed reality human-computer/-robot interaction features will be investigated. In future, we aim to adapt such teleoperation training systems to robot teleoperation in extreme environments [17].

Acknowledgements. Aaron Smiles was funded by the UKRI EPSRC EngD Data-Centric Engineering CDT at Queen Mary University of London (reference 2601988). The work was partially co-funded by the UKRI EPSRC Q-Arena grant EP/V035304/1. We thank Kaustubh Sadekar, Yik Lung Pang, and the National Oceanography Centre for their feedback.

References

1. Adami, P., et al.: Effectiveness of VR-based training on improving construction workers' knowledge, skills, and safety behavior in robotic teleoperation. Adv. Eng. Inform. **50**, 101431 (2021)
2. Adil, E., Mikou, M., Mouhsen, A.: A novel algorithm for distance measurement using stereo camera. CAAI Trans. Intell. Technol. **7**(2), 177–186 (2022)
3. Dodgson, N.: Autostereoscopic 3D displays. Computer **38**(8), 31–36 (2005)
4. Farkhatdinov, I., Ryu, J.H.: Development of educational system for automotive engineering based on augmented reality. In: International Conference on Engineering Education and Research (2009)
5. Giudici, G., Omarali, B., Bonzini, A.A., Althoefer, K., Farkhatdinov, I., Jamone, L.: Feeling good: Validation of bilateral tactile telemanipulation for a dexterous robot. In: Iida, F., et al. (eds.) TAROS 2023, LNAI, vol. 14136, pp. 443–454. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-43360-3_36
6. Kaehler, A., Bradski, G.: Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library. Reilly Media, Inc. (2016)
7. Lin, Q., Kuo, C.: On applying virtual reality to underwater robot tele-operation and pilot training. Int. J. Virtual Reality **5**(1), 71–91 (2001)
8. Livatino, S., et al.: Intuitive robot teleoperation through multi-sensor informed mixed reality visual aids. IEEE Access **9**, 25795–25808 (2021)
9. Murugesan, Y.P., Alsadoon, A., Manoranjan, P., Prasad, P.: A novel rotational matrix and translation vector algorithm: geometric accuracy for augmented reality in oral and maxillofacial surgeries. Int. J. Med. Robot. Comput. Assist. Surg. **14**(3), e1889 (2018)
10. Omarali, B., Denoun, B., Althoefer, K., Jamone, L., Valle, M., Farkhatdinov, I.: Virtual reality based telerobotics framework with depth cameras. In: 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 1217–1222. IEEE (2020)
11. Omarali, B., Javaid, S., Valle, M., Farkhatdinov, I.: Workspace scaling in virtual reality based robot teleoperation. In: Proceedings of the Augmented Humans International Conference 2023, pp. 98–104 (2023)

12. Omarali, B., Palermo, F., Althoefer, K., Valle, M., Farkhatdinov, I.: Tactile classification of object materials for virtual reality based robot teleoperation. In: 2022 International Conference on Robotics and Automation (ICRA), pp. 9288–9294. IEEE (2022)
13. Sadekar, K.: Depth estimation using stereo camera and OpenCV (Python/C++) (2021)
14. Saliba, C., Bugeja, M.K., Fabri, S.G., Di Castro, M., Mosca, A., Ferre, M.: A training simulator for teleoperated robots deployed at CERN. In: ICINCO (2), pp. 293–300 (2018)
15. Shibata, T., Kim, J., Hoffman, D.M., Banks, M.S.: The zone of comfort: predicting visual discomfort with stereo displays. J. Vis. **11**(8), 11 (2011)
16. Szczurek, K.A., Prades, R.M., Matheson, E., Rodriguez-Nogueira, J., Castro, M.D.: Mixed reality human-robot interface with adaptive communications congestion control for the teleoperation of mobile redundant manipulators in hazardous environments. IEEE Access **10**, 87182–87216 (2022)
17. Vitanov, I., et al.: A suite of robotic solutions for nuclear waste decommissioning. Robotics **10**(4), 112 (2021)