Intelligent Human Systems Integration (IHSI 2023), Vol. 69, 2023, 145–152
https://doi.org/10.54941/ahfe1002831

An Augmented-Reality-Assisted Immersive System for Robotic Arms Teleoperation

Claudio Loconsole[1,2], Leonardis Daniele[2], and Antonio Frisoli[2]

[1]Universitas Mercatorum, Roma, 00186, Italy
[2]PERCRO Laboratory, Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna, San Giuliano Terme (PI), 56010, Italy

ABSTRACT

Robotic teleoperation allows humans to remotely operate to perform several tasks in different fields, from social to high-risk situations. In this paper, we propose a novel immersive system which exploits augmented reality (AR) to provide effective feedback to the operator, simplifying the task and optimizing the performance during teleoperation. The AR cues are conveyed to the operator through a head-mounted display (HMD) equipped with hand tracking and gesture recognition system to control the remote robotic arms for teleoperation, and a head orientation system to control the orientation of a remote RGB-D camera. This setup allows to reduce the bandwidth required for video streaming (both 2D and point cloud) and to improve the quality of the remote interaction in real-world tasks.

Keywords: Robotic teleoperation, Augmented reality, Head-mounted display, RGB-D camera, Hand tracking, Gesture recognition

INTRODUCTION

Advancements in the robotic field have allowed humans to use robots for demanding tasks in an autonomous way to assist other humans in several tasks including remote activities. As example, robots are used in high-risk scenarios (e.g., Klamt *et al.*, 2019) which are too dangerous for human operators or in social applications (e.g., Schwarz *et al.*, 2021). However, the development of assistance robots for remote tasks has been accelerated by involving a human operator in the loop (Leeper *et al.*, 2012) through teleoperation (Niemeyer *et al.*, 2008) which combines robot skills with the operator's capabilities. In particular, for manipulation tasks, body-based control methods including gestures have been proved to be more effective for operator dexterity (Almeida *et al.*, 2014).

On the other hand, the strategy and the technology to be used to provide visual information to the operator makes teleoperation a complex task. Delay and high latency in end-to-end communication, visualization issues of the remote environment, difficulties in identifying the right objects to interact with, and in judging the objects' distances from the robot end effector are only some of the issues to be overcome for implementing teleoperation

(Nitsch *et al.*, 2012). Augmented Reality (AR) has proven (Mosiello *et al.*, 2013) to be a viable solution to overcome visual feedback limitations by providing additional information to the operator for both embodiment enhancement and virtual fixture. In detail, embodiment enhancement can be obtained by using a camera mounted on the head of the remote robot and showing the captured images (or point clouds) through a Head-Mounted Display (HMD) to the operator, thus allowing the operator to feel as he/she is in the remote environment (Zahorik *et al.*, 1998). Virtual fixtures, instead, are used to help the perception of the operator of the remote environment (Sayers *et al.*, 1998) and speeding up the execution of tasks (Xia *et al.*, 2012) by providing him/her visual cues, such as lines, virtual objects or numerical information, in order to highlight points of interest and to improve the spatial perception of the scenario. Furthermore, especially in absence of haptic feedback on the operator side, AR allows the operator to receive indirect sensory information (Kitagawa *et al.*, 2005).

However, in augmented reality-assisted teleoperation an important issue is represented by the high latency for video streaming. In fact, high latency can deeply affect the operator performance during tasks (Martins *et al.*, 2015).

In this paper, we propose an augmented reality-assisted immersive system for robotic teleoperation aiming at reducing latency of the video streaming to the operator, and an accurate control for robotic telemanipulation of real objects.

The proposed system can be used for several tasks in several scenarios including social applications and telemanipulation in high-risk situations. In the former scenario, an operator can remotely interact with a person in demonstrative social tasks (i.e., hand-shaking or playing board games). In the latter scenario, our system, suitably integrated with a robotic base allowing its mobility, aims at performing teleoperated activities such as door opening, valve turning and switch operating without local human intervention at the remote side.

SYSTEM DESIGN

As shown in Figure 1, the operator wears a Head Mounted Display (HMD), whereas on the remote side the system includes both two robotic arms each equipped with a robotic hand. An RGB-Depth (RGB-D) camera shooting a real-world workspace is fixed at the robotic base through a pivoting support. In Figure 2, the system architecture is reported.

The head orientation angle values are sent to the remote controller to pilot the robotic base. The rotation of the robotic base and of the RGB-D camera allows the operator to remotely have the same point of view of the remote workspace scene and to freely rotate his/her head resulting in a corresponding movement of the camera on the remote side.

The hand tracking is used to control the position of the end-effector of the robotic arms (corresponding to the centre of mass of the robotic hands). The hand gesture, instead, is used to control the robotic hands for both grasping/releasing objects and for replicating some specific gestures.

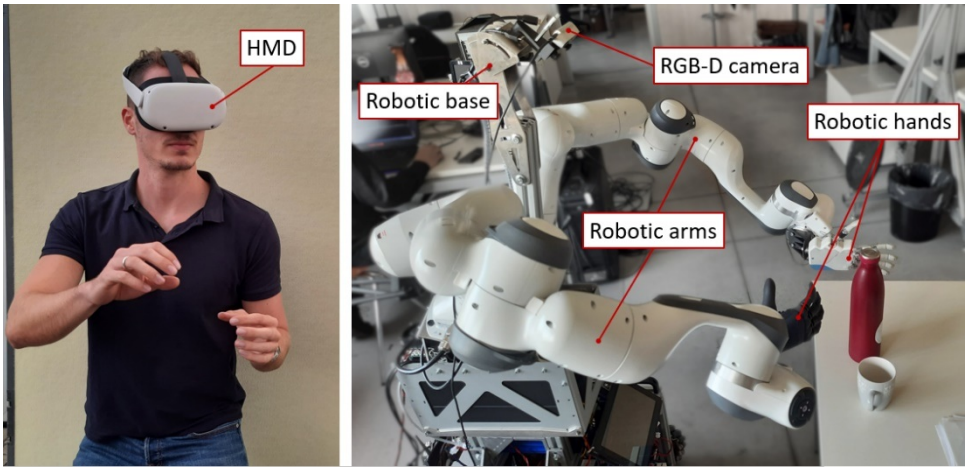

Figure 1: The AR-assisted immersive system: the operator side (left) and the remote side (right).

Application of the presented methods is tailored to the robotic platform developed at the IIM institute of the Scuola Superiore Sant'Anna (Pisa, Italy). However, the presented methods can be generalized to similar robotic platforms with bilateral arms and a pivoting sensorized head.

Robotic Arms

The core of the remote robot is composed of two dexterous robotic arms, arranged in a bilateral configuration. The devices are Franka Emika Panda robotic arms, with maximum payload of 3kg each at the end effector and 7 degrees of freedom. These robots are particularly suitable for dexterous manipulation and physical interaction with humans: all the seven revolute joints are equipped with torque sensors, making the robot sensitive to physical interaction at all the segments of the kinematic chain, and enabling the robot for safe collaborative tasks.

An impedance control modality with closed-loop force feedback is provided, already calibrated for gravity and friction compensation. This provides a smooth and stable operation with safety limits in terms of maximum force exchanged between the robot and the environment.

The 7 dofs allow to better adapt the workspace of the robot in tracking the hand of the human operator, providing also an additional dof used to maximize distance from each joint end-stops. It has to be noted the robotic arm is not anthropomorphic (the kinematic chain does not respect the conventional spherical wrist- anthropomorphic base). Hence, in the human-operator-tracking modality this can lead to very different workspace configuration between the human and the robot (poses comfortable for the human can result in pose close to the limit of certain joints of the robot). To diminish such occurrences, the control modality of the robot was programmed to use the redundant degree of freedom (7 available dofs with respect to a 6dofs position-rotation tracking task) to maximize distance from the end-stops of each joint.

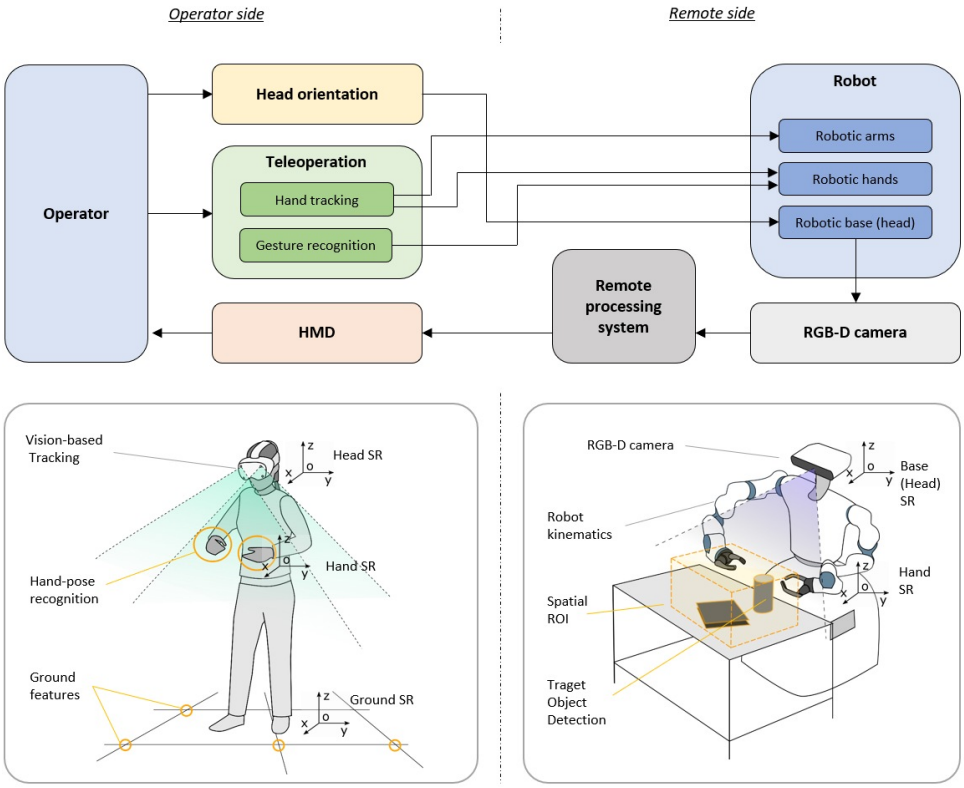

Figure 2: Implemented system architecture.

Robotic Head

Regarding the sensorized head, it is equipped with a Realsense D435 RGB-D camera (FOV 87°x58°, depth range 0.28-2.0 meters with <2% error).

The depth information is obtained by the camera through an embedded IR laser projector, projecting on the environment a grid of visual features, and an IR-stereo camera, acquiring the projected features.

The depth information is obtained by the camera through an embedded IR laser projector, projecting on the environment a grid of visual features, and an IR-stereo camera, acquiring the projected features. The depth information is matched by the sensor to the pixels of a conventional RGB image. The robotic head has been designed with a pivoting neck in order to extend the FOV of the RGB-D sensor. Conventional DC geared motors with optical encoder have been used to actuate the head on 2 dofs (pan and tilt). A closed loop position control has been implemented.

Operator's VR Harness

The Operator's hardware consists of just an Oculus Quest 2 3D vision system. The simplicity and wearability of the operator's 3D harness is one of the key-point of the proposed method. The Oculus HMD provides an embedded and robust tracking system based on artificial-vision, referenced to grounded visual features (i.e. floor, wall and ground references in the physical room).

Moreover, the vision-based tracking system is capable of smooth hands recognition with continuous finger pose estimation. These features allow to reference all the required systems of reference at the operator's side (ground, hands and head) without introducing additional tracking systems and calibration procedures between them. Also, system operation is performed with bare-hands, with the operator wearing only the HMD: no additional tracking devices (i.e., gloves or markers) have to be worn, eventually limiting comfort and dexterity of the user.

Teleoperation Methods

Tracking and Systems of Reference

A relevant aspect in teleoperation is the setting of the system of references and position estimation between different local components and between the local and remote side.

Local Operator Systems of Reference

Regarding the operator side, position estimation and the relative systems of reference are based on the HMD embedded tracking system. The tracking system is based on artificial vision processing data acquired by four cameras embedded onto the device and pointed at the outside environment. Cameras extract and track a set of visual features from static elements of the environment (i.e., ceil, walls and floor of the room), estimating in turn a relative transform between the device, fixed at the user's head, and the environment. It corresponds to the relative pose of the head with respect to the ground system of reference.

An additional feature of the Oculus system is the tracking and pose estimation of the users' hands. Position and rotation of the hand palm is estimated with respect to the device system of reference, which in turn is in a known position and rotation with respect to the ground. Estimation of fingertips position is also made available by the system with respect to the hand palm, hence providing full information of the hand pose to be used in manipulation and interaction tasks.

Remote Robot Reference System

Regarding the remote robotic side, the system uses the kinematic model of the robot and data of the angular sensors placed at the revolute joints to reconstruct the relative position of the hands with respect to the robotic base. The same approach is used for the head, thus estimating the relative position and orientation of the RGB-D camera with respect to the base.

Physical objects of interest in the remote location are, identified by processing the RGB-D camera data. Their pose can be estimated with respect to the camera head (using depth information and location in the frame of the centroid) and, in turn, with respect to the base of the robot and to the hands of the robot.

Teleoperation Control Signals

In teleoperation, hands are directly controlled generating a position reference signal form the operator side to the remote robot. It uses the position and

rotation of the hand tracking system with respect to ground. An offset position is used to better match the different operator's head to hands distance with respect to the camera and robotic hands distance. In the implemented robotic system, scaling is one to one, although different scaling of the position reference signals can be adopted for robot kinematic workspace differing from the human. Finger control reference is implemented from pose estimation provided by the hand tracking, linearized between a fully open and fully closed position for each finger.

Similarly to hands teleoperation, the pose of the operator's head with respect to ground is sent as reference to the pivoting actuators of the robotic neck.

Remote Vision and AR Assistance

Video streaming is a critical aspect in teleoperation. Visual channel provides fundamental information to the operator regarding the environment and interaction between the remote robot and manipulated objects. However, due to the high amount of visual data, it is usually the most impacted by latency and throughput of the communication channel. In this paper, we propose a method to reduce the amount of significant data in the scene that needs to be communicated at high rate and low latency.

By using both AI methods and RGB-D information, the aim is to isolate the manipulation region of interest from the surrounding environment. Only data related to the manipulation ROI is sent at a fast communication rate, while the whole image of the environment is updated at a lower frame-rate.

The method uses the following steps:

- at the remote side, the image is acquired and processed by a convolutional neural network providing classification of common objects in the scene and location in the RGB image;
- centroid of each recognized object is associated to the depth information, hence to the cartesian position;
- the object with the cartesian position closest to the mid-point between the two robotic hands is selected as the target object;
- a ROI is centered on the target position and used to select in the matched RGB-D data only points inside that volume. A spherical ROI is used in the example of Figure 3;
- the RGB image is compressed with only pixels of the ROI and sent to the operator side.

While on one hand the proposed method increases the computational effort, it significantly reduces the amount of data that needs to be sent on the communication channel. In the shown example, the processed RGB image, compressed in jpg format, has size of 45 kbytes, whereas the original image has size of 72 kbytes, thus resulting in a reduction of the bandwidth of 36%. Image d) in Figure 3 is obtained by recomposing at the operator side the fast updated manipulation ROI with the environment, updated at lower frame rate.

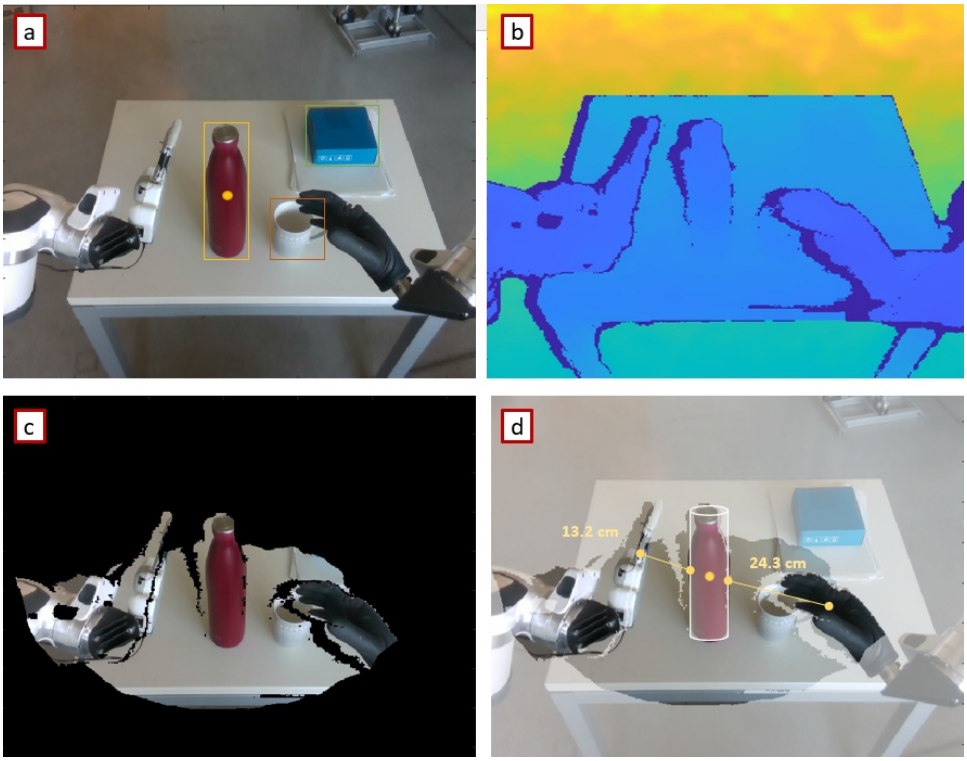

Figure 3: a) The RGB image with object detection. b) Depth information from the RGB-D camera. c) The processed RGB image selecting a ROI based on object detection and depth information. d) The reconstructed image at the operator side, evidencing the fast update ROI, the surrounding environment and AR cues (e.g., the distance between each robotic hands and the object boundaries).

However, the proposed method allows us to select to send RGB image and depth map to the operator side to locally calculate the point cloud or to send to the operator side directly the filtered 3D point cloud.

In addition, the proposed approach allows us also to choose the system processing the RGB-data and producing the AR cues to be showed to the operator through the HMD. Proposed AR cues (see Figure 3.d) deal with showing to the operator the distance between the center of each robotic hand and the boundaries of the object of interest. This way, it is possible for the operator to finely pilot the robotic hand grasping since the operator can make a hand gesture at a proper distance from the object to be manipulated.

CONCLUSION

In this paper, we presented an immersive system for robotic teleoperation which allows an operator to finely perform grasping and manipulation tasks in real-world. In detail, the proposed system allows to reduce data size in video streaming and in turn latency, since this can strongly affect the immersive experience of the operator. Also, the system provides to the operator visual cues by means of Augmented Reality to correctly perceive the distance and the dimensions during fine manipulation tasks.

The proposed system can be used for several tasks in several scenarios including social applications and high-risk situations. In the former scenario, an operator can remotely perform social tasks (i.e. hand-shaking or playing board games). In the latter scenario, our system, suitably integrated with a robotic base allowing its mobility, aims at performing physical teleoperated activities such as door opening, valve turning and switch operating without local human intervention.

Future works deal with providing to the operator haptic feedback in telemanipulation tasks to improve the immersive experience and sensory information from the remote environment.

REFERENCES

Almeida, L., Patrao, B., Menezes, P., and Dias, J. (2014) 'Be the robot: Human embodiment in tele-operation driving tasks', *In The 23rd IEEE international symposium on robot and human interactive communication*. IEEE, pp. 477–482.

Kitagawa, M., Dokko, D., Okamura, A. M., and Yuh, D. D. (2005) 'Effect of sensory substitution on suture-manipulation forces for robotic surgical systems', *The Journal of thoracic and cardiovascular surgery*, 129(1), pp. 151–158.

Klamt, T., Rodriguez, D., Baccelliere, L., Chen, X., Chiaradia, D., Cichon, T., ... and Behnke, S. (2019) 'Flexible disaster response of tomorrow: Final presentation and evaluation of the CENTAURO system', *IEEE robotics & automation magazine*, 26(4). IEEE, pp. 59–72.

Leeper, A. E., Hsiao, K., Ciocarlie, M., Takayama, L. and Gossow, D. (2012) 'Strategies for human-in-the-loop robotic grasping', *In 7th ACM/IEEE HRI*. ACM, pp. 1–8.

Martins, H., Oakley, I., and Ventura, R. (2015) 'Design and evaluation of a head-mounted display for immersive 3D teleoperation of field robots', *Robotica*, 33(10), pp. 2166–2185.

Mosiello, G., Kiselev, A., and Loutfi, A. (2013). 'Using augmented reality to improve usability of the user interface for driving a telepresence robot', *Paladyn, Journal of Behavioral Robotics*, 4(3). De Gruyter, pp. 174–181.

Niemeyer, G., Preusche, C. and Hirzinger, G. (2008) 'Telerobotics', *In Springer handbook of robotics*. Springer, pp. 741–757.

Nitsch, V and Färber, B. (2012) 'A meta-analysis of the effects of haptic interfaces on task performance with teleoperation systems', *IEEE Transactions on Haptics,* 6(4). IEEE, pp. 387–398.

Sayers, C. P., Paul, R. P., Whitcomb, L. L., and Yoerger, D. R. (1998). 'Teleprogramming for subsea teleoperation using acoustic communication', *IEEE Journal of Oceanic Engineering*, 23(1). IEEE, pp. 60–71.

Schwarz, M., Lenz, C., Rochow, A., Schreiber, M., & Behnke, S. (2021) 'NimbRo Avatar: Interactive immersive telepresence with force-feedback telemanipulation', *In 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 5312–5319.

Xia, T., Léonard, S., Deguet, A., Whitcomb, L., and Kazanzides, P. (2012). 'Augmented reality environment with virtual fixtures for robotic telemanipulation in space', *In 2012 IEEE/RSJ International Conference on Intelligent Robots and System,*. IEEE, pp. 5059–5064

Zahorik, P., and Jenison, R. L. (1998) 'Presence as being-in-the-world', *Presence*, 7(1). MIT Press Direct, pp. 78–89.