Teleoperating Humanoids Robots using Standard VR Headsets: A Systematic Review

Lucio Davide Spano[a]
Department of Mathematics and Computer Science, University of Cagliari, Via Ospedale 72, 09124, Cagliari, Italy

Keywords: Teleoperation, Virtual Reality, Humanoid Robots, Systematic Review, Meta Analysis.

Abstract: The recent development of both Virtual Reality (VR) and the availability of multipurpose Humanoid Robots and their development platforms fostered the combination of such technologies for supporting teleoperation tasks. Many technical solutions are documented in the literature and studies discussing the benefits and limitations of different solutions. In this paper, we survey a total of 23 papers written between 2017 and 2021 that employ a consumer VR headset for teleoperating a humanoid robot, applying the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. We identify the characteristics of the hardware setup, the software communication architecture, the mapping technique between the operator's input and the robot movements, the provided feedback (e.g., visual, haptic, etc.), and we report on the identified strengths and weaknesses on the usability level (if any). Finally, we discuss possible further research directions in this field.

1 INTRODUCTION

In the last years, we witnessed the convergence of two important trends. The first is the wide availability of Virtual Reality (VR) head-mounted displays (HMD) at a low price, pushing the development of many interfaces based on such devices, both at the consumer and the research level. The second is the development of commercial versions of humanoid robots that fostered their application in different scenarios. In addition, modular development frameworks like the Robot Operating System (ROS) (Whitney et al., 2018) ease the communication between software components and the robot hardware.

In such a scenario, there have been many attempts to close the circle between the virtual and the real world, supporting the teleoperation of humanoid robots through a Virtual Reality interface. If implemented effectively, such an interaction can support physical interactions in remote places that keep a high sense of presence for the user. Use cases include remote learning and teaching (Botev and Lera, 2020), remote manufacturing (Lipton et al., 2018) or other tasks where a remote embodied presence is convenient. Such a fascinating topic fostered a high amount of research, especially in the last five years, after the hardware and software platforms we mentioned became available at reasonable prices. Such a rapid evolution requires a step back for understanding the lesson we learnt from the different attempts and possible improvements and open research questions.

In this paper, we survey a total of 23 papers written between 2017 and 2021 that employ a consumer VR headset for teleoperating a humanoid robot, applying the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. We categorise the proposed solution describing the employed hardware and software solutions, highlighting the identified strong and weak points for each of them. In addition, we analyse the validation sections of the surveyed papers for identifying open questions related to the teleoperator experience.

2 METHOD

This review aims at analysing the last five years of development in the VR interfaces for teleoperating humanoid robots. Through the literature analysis, we want to highlight the current practices and identify the possible improvements to their development and evaluation methodologies. We applied the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Moher et al., 2009) guidelines for identifying the papers of interest for our review,

---

[a] https://orcid.org/0000-0001-7106-0463

Spano, L.
Teleoperating Humanoids Robots using Standard VR Headsets: A Systematic Review.
DOI: 10.5220/0010723200003060
In *Proceedings of the 5th International Conference on Computer-Human Interaction Research and Applications (CHIRA 2021)*, pages 308-316
ISBN: 978-989-758-538-8; ISSN: 2184-3244

applying the four phases for getting the final set.

We included in the review the papers that meet the following criteria:

1. **Published between 2017 and 2021.** We explored the literature published in the last five years.
2. **Support the Teleoperation Task.** The proposed interface must support the control of a robot by an operator when they are not in the same location.
3. **Consumer VR Headset.** The proposed interface must use a consumer VR headset for implementing the immersive interface (e.g., an Oculus Rift, HTC Vive, etc.).
4. **Control of a Humanoid Robot.** The controlled robot must resemble entirely or partially the a human body.
5. **Full Paper.** The paper must contain the description of a mature work (e.g., should not contain an idea or a work in progress implementation).

Figure 1 summarises the four phases that led to the selection of the 23 papers. The initial set of papers was build using the following digital libraries and search engines:

- ACM digital library
- IEEE Xplore
- Springer Link
- Google Scholar

In order to replicate the database query, we looked for all the items in the library published between 2017 and 2021, including the words *teleoperation*, *humanoid*, *robot*, *Virtual Reality* (or VR) in the paper text. After collecting the starting set, we removed from the list the papers that seemed not relevant according to their title and/or abstract. Finally, we examined the full text of 83 papers, and we excluded those not meeting the described criteria.

In the next sections, we will discuss the identified techniques, describing the hardware configuration, the supported tasks and the validation of the proposed approaches. Table 1 summarises the different aspects identified in the surveyed papers.

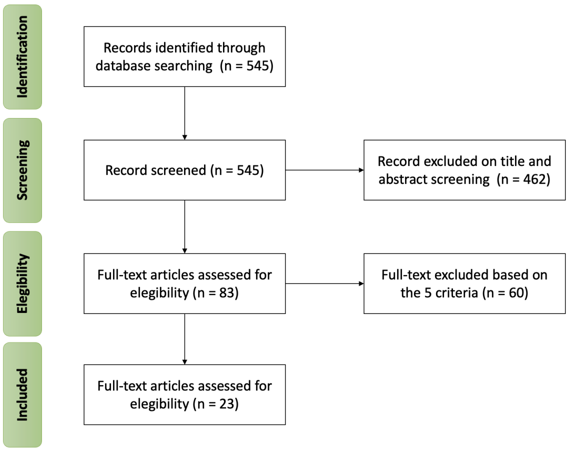

Figure 1: The four phases of the literature search performed using the PRISMA guidelines.

3 TELEOPERATION TECHNIQUES

According to (Lipton et al., 2018) we can distinguish the teleoperation systems into three categories. The first and most simple is employs a *direct mapping* (see Figure 2): there is no intermediate space between the user and the robot. In such a case, the sight of the human is bound to the robot vision, and a high refresh rate (minimum 60Hz) is required for preventing motion sickness. For controlling the robot movements (or the movement of its parts) in direct mapping scenarios, we can employ a piloting controller (e.g., through a joypad, joystick, keyboard and mouse etc.) while wearing the headset. This usually breaks the sense of presence but requires less effort in mapping the input towards the robot actuators. Other solutions employ motion capturing devices such as the Kinect, Leap Motion etc., which capture the user's body movement. In this case, the problem is how to map the user's movements towards the robot, since simple mappings require a good correspondence between the structure of the two bodies, for avoiding fatigue in the user or impossible movements.

In the *cyber-physical approach* the user controls a virtual twin of the robot in a Virtual Reality environment (see Figure 3. There is a mapping between the relevant part of the real world in the remote setting, including the virtual counterpart of the robot. On the one hand, they usually have the advantage of supporting simulations for training the operator. On the other hand, they require tracking a high amount of data for

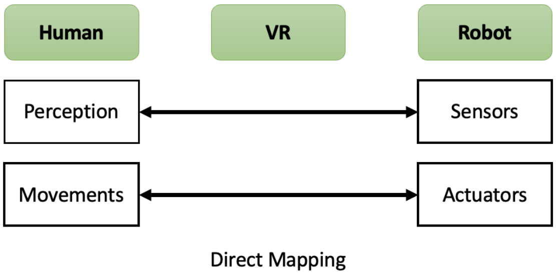

Figure 2: The representation of a direct mapping teleoperation techniques. The user's perception and movements are directly mapped to the robot's sensors and actuators.

Table 1: Summary of the surveyed papers.

Paper	Robot	HMD	Controllers	Movements	Framework	Technique
(Martinez-Hernandez et al., 2017)	iCub, Pioneer LX mobile	Oculus Rift DK2	Custom gloves (Force)	Head Orientation,	YARP	Direct
(Kilby and Whitehead, 2017)	InMoov	Oculus Rift DK2	IMU sensors	Arm control, grasp	Custom	Direct
(Chen et al., 2017)	Baxter	Oculus Rift DK2	Kinect2, Geomagic Touch	Head Orientation, grasp	Custom	Direct
(Mizuchi and Inamura, 2017)	ROS-based robots	Oculus Rift, HTC Vive, FOVE	Kinect2, Leap Motion, Perception Neuron	Head Orientation, robot control	ROS	Cyber-physical
(Lipton et al., 2018)	Baxter	Oculus Rift	VR Controllers	VR controllers	ROS	Homunculus
(Zhang et al., 2018)	PR2	HTC Vive	VR Controllers	VR Controllers	Custom	Cyber-physical
(Bian et al., 2018)	Baxter	Oculus Rift	Kinect	Head Orientation, arm control, grasp	Custom	Direct
(Whitney et al., 2018) and (Whitney et al., 2020)	ROS-based robots	Unity-compatible HMD	VR Controllers, Kinect	Head Orientation, arm control, grasp	ROS	Direct
(Spada et al., 2019)	NAO	Oculus Rift	Cyberith Virtualizer	Head Orientation, Navigation	Custom	Direct
(Xi et al., 2019)	UR2	HTC Vive	VR Controllers	Arm control, grasp	Custom	Direct
(Elobaid et al., 2019)	iCub	Oculus Rift	VR Controllers, Cyberith Virtualizer	Head Orientation, arm control, navigation	YARP	Direct
(Cardenas et al., 2019)	Telebot-2	HTC Vive	Wearable, VR Controller	Head Orientation, arm control, grasp	Custom	Direct
(Gaurav et al., 2019)	Baxter	HTC VIve	VR Controllers	Head Orientation, grasp	ROS	Direct
(Hirschmanne et al., 2019)	Pepper	Oculus Rift	Leap Motion	Head Orientation, arm control, grasp	Custom	Direct
(Lentini et al., 2019)	ALTER-EGO	Oculus Rift	Myo armband, IMU, Wii Balance Board	Head Orientation, arm control, grasp		Direct
(Girard et al., 2020)	InMoov	Oculus Rift DK2	IMU sensors	Head Orientation, robot control	Custom	Direct
(Orlosky et al., 2020)	iCub	Oculus Rift DK2	N/A	Head Orientation,	Custom	Cyber-physical
(Botev and Lera, 2020)	Qtrobot	Oculus Rift	N/A	Head Orientation, robot control	ROS	Direct
(Omarali et al., 2020)	Panda	Oculus Rift	VR Controllers	arm control, grasp	Custom	Cyber-physical
(Nakanishi et al., 2020)	Toyota Human's Support	Oculus Rift	VR Controllers	Head Orientation, arm control, grasp, navigation	ROS	Direct
(Zhou et al., 2020)	Baxter	HTC Vive	VR Controllers	Head Orientation, arm control, grasp, navigation	ROS	Cyber-physical
(Wonsick and Padır, 2021)	NASA Valkyrie	HTC Vive	VR Controllers	Head Orientation, arm control, grasp, navigation	ROS	Cyber-physical

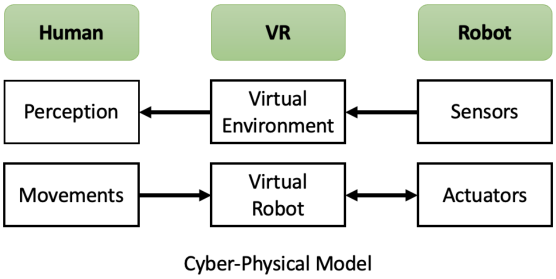

Figure 3: The representation of a cyber-physical teleoperation techniques. The robot sensors influence the representation of a virtual environment, which is perceived by the user, who controls a virtual representation of the robot for interacting with it.

supporting the simulation mapping the relevant part of the remote setting.

The *homunculus* model (see Figure 4) exploits the concept of a control room, where multiple displays and objects are spread in the virtual world. The user interacts with such control elements, which are in turn mapped towards the robot. Different displays and objects may employ different mapping techniques, thus increasing the adaptation to specific tasks. In addition, removing the need to have a virtual replica of the remote setting requires less data for simulations.

In the next sections, we will summarise the work in the surveyed papers grouping them by teleoperation technique. Most of them belong to the direct mapping (68.2%), followed by cyber-phisical approaches (27.2%) and homunculus (4.6%). We will present them mainly in cronological order of publication, grouping only follow-up work.

3.1 Direct Mapping

The direct mapping technique supports the teleoperation by rendering the video or the raw output of the robot's visual sensors to provide context information and action feedback to the teleoperator. Such a technique is simple to implement, but it suffers the delays between the user's actions and the sensor feedback rendering, which may be perceivable for the users. However, the simplified implementation and the intuitiveness of the mapping makes it the default choice in many teleoperation interfaces.

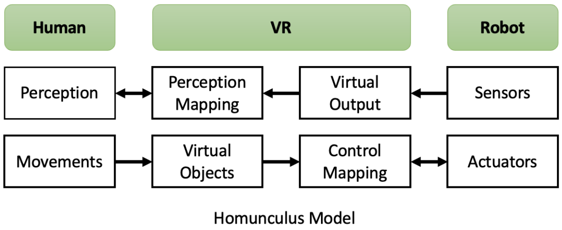

Figure 4: The representation of the homunculus teleoperation techniques. The user selects among different displays and objects in a virtual control room, supporting different mappings between perception and sensors and between movements and actuators.

(Martinez-Hernandez et al., 2017) experimented with the iCub and the Pioneer XL mobile for reaching telepresence in a remote setting. The hardware configuration includes an Oculus Rift DK2 playing the camera video and audio stream on the HMD. The HMD orientation is mapped to the robot's head movements. As for the hand control, they created a pair of custom gloves, providing vibration according to the pressure data coming from the iCub hands.

(Kilby and Whitehead, 2017) created a system consisting of a 3D printed open-source humanoid robot (InMoov), an Arduino-based mapping between IMU sensors strapped on the user's dominant arm and the controlled robot's arm. The user also wears an Oculus Rift DK2 for visualizing the remote environment. The robot was equipped with cameras whose stream was displayed in the HMD. No other source of input is available.

(Chen et al., 2017) created a teleoperating system based on a Baxter robot equipped with a Kinect V2 on top of his head. The pan and tilt servo motors of the Kinect are controlled through the Oculus Rift DK2 IMU sensors, mapping the natural head movements into changes in the camera viewport. The user controls the robot's arm through a Geomagic Touch joystick, which provides the tactile force of the robot's pinchers, the restoration feedback that puts back the joystick in the rest position (as does the robot arm) and the force feedback during the robot arm control. The communication architecture exploits a wireless network. Three computers and one Arduino implement the communication among the different components of the architecture.

(Bian et al., 2018) detail a mapping technique for transforming the skeleton tracking data acquired with a Kinect V2 towards the arm movements of a Baxter robot. The interface exploits a direct mapping on both the user's sight and movements: the stream of another Kinect V2 sensor is rendered on an Oculus Rift HMD.

(Whitney et al., 2018) introduced the ROS Reality, an interface supporting the teleoperation in VR through the Internet of any robot compatible with the Robot Operating System (ROS). A Unity-compatible VR headset is required for running the interface. Following the ROS modular architecture, ROS Reality is a collection of components supporting the creation of a VR teleoperation interface built on top of ROS components communicating whit the Unity environment. It basically supports a direct mapping between the user's movements and the remote robot, even if a virtual model of the robot is included in the Unity scene for immediate feedback.

(Spada et al., 2019) propose a different configuration for controlling the robot locomotion. Instead of joypad/joystick or keyboard controls for changing the robot's position in the environment, they map the user's movement on a passive omnidirectional treadmill, the Cyberith Virtualizer. Such a configuration increases the direct mapping level in the teleoperation task. They employ an Oculus Rift HMD, and the proposed architecture supports both the teleoperation of a physical robot or the simulation in a virtual environment.

The work in (Xi et al., 2019) focuses on learning manipulation tasks from several human demonstrations. Using Hidden Markov Models, the mapping technique is able to "correct" the user's movements for improving the manipulation success. The testing environment exploits an HTC Vive and its controller for getting the user's input, a Kinect V2 for implementing the robot's view and two Universal Robot3 six-DoF arms.

(Elobaid et al., 2019) elaborated more on the idea of mapping the robot navigation task through a walk-in-place technique. Similarly to (Spada et al., 2019), they employed a Cyberith Virtualizer, but they also included the support to the VR controllers for mapping the arm movements. Overall, such a setup completes the direct mapping effort, but the hardware configuration suffers from a tradeoff between safety and usability. For avoiding possible user's falls, the treadmill exploits a ring and some belts around the user's waist, but this makes it difficult for them to move their arms, and often users hit the ring with the remotes.

(Cardenas et al., 2019) propose an IMU-equipped wearable garment for controlling the movements of the robot's arms and torso. After a calibration step, they use the sensors' output to reconstruct the kinematic of the user's movements and reproduce them through the connected robot, called Telebot-2 and specifically designed for working with such a wearable device. Other tasks such as controlling the view, grasping and the navigation are performed through the standard VR headset controllers and sensors (HTC Vive).

Similarly, (Girard et al., 2020) worked on a control system based on wearable suits, this time employing low-cost components. The controlled robot is 3D printed, based on the InMoov (Langevin, 2014) project. The stream of two high-resolution cameras and binaural microphones positioned on the robot head is sent to an Oculus Rift DK2. The movements are collected through a suit of IMU sensors processed using the Perception Neuron framework.

(Gaurav et al., 2019) propose a Deep Learning-based technique for mapping the human poses and the robot joint angles in a direct mapping. The technical setup includes a Kinect for sensing the environment around the robot, a VR-based visualization of the depth-sensing point cloud and the usage of the HTC-Vive controllers for supporting a Baxter Robot grasping task.

(Hirschmanner et al., 2019) designed a free-hand control interface for the robot teleoperation. Instead of employing standard VR controllers, they exploited a mount for installing a Leap Motion device in front of an Oculus Rift headset. Through such a device, the interface tracks the skeleton joints of the user's hands, which are used to reconstruct the user's shoulder and elbow poses and map them to the robot's joints. They employed a Softbank Pepper in their experiments, but the system may be used on other robots of the same producers.

(Lentini et al., 2019) introduced ALTER-EGO, a mobile two-wheel and self-balancing robot equipped with variable stiffness actuators. The robot components include sensor and computational power, enabling it to work autonomously. The robot supports a teleoperation mode through a pilot station composed of wearable devices, a Wii Balance Board and an Oculus Rift HMD. The headset controls the robot's head orientation, while two Myo armbands for each arm and an additional IMU device per hand control the robot's arm movements and grasp. Finally, the Wii Balance board allows controlling the robot's velocity by leaning.

(Botev and Lera, 2020) propose the application of a direct mapping telepresence interface in the educational field. The proposed support leverages ROS Bridge and allows connecting Unity to ROS for controlling the robot. In the paper, the operator used an Oculus Rift and its controllers for teleoperating a QTrobot.

(Nakanishi et al., 2020) created a VR teleoperation interface for the Toyota Human Support Robot, based on the Oculus Rift and its controllers, and exploiting the ROS framework for communicating with the robot. The standard robot's hardware was augmented with a Ovrvision Pro 360^o camera streaming the output to the HMD. The work describes the mapping between the human and the robot movements and reports on an extensive test against a set of selected tasks in a home environment and on the standard setting of the WRS robot competition.

3.2 Cyber-physical Approaches

The Cyber-Physical Approaches create a teleoperation interface starting from the robot RGB and/or depth camera and perform a scene reconstruction and/or understanding step before rendering it in VR. The scene is *updated* reading such streams, but not completely replaced. This allows reducing the delays between the user's actions and the visualization of their effects.

(Mizuchi and Inamura, 2017) introduce a framework for managing different types of robots and HMD through the same underlying support. The solution leverages on Unity 3D and the ROS framework, allowing to teleoperate real robots and simulate the interaction in VR without the need to connect a real physical robot. There is not much information on how the supported control gestures, the paper reports tests with depth-camera based tracking devices. The evaluation included discusses only the communication latency.

(Zhang et al., 2018) proposes an interesting solution for implementing a cyber-physical system for teleoperating a robot. Instead of directly mapping the robot camera on the HMD (an HTC Vive), they exploit a depth camera to create a coloured point-cloud view of the surrounding environment and render it in VR to avoid sickness problems. The user controls the robot's arms through the Vive controllers, mapping their position and orientation towards the robot. The trigger button controls the robot's gripper. The authors exploit such a setting for training a deep network that learns how to manipulate objects from the teleoperation data.

(Orlosky et al., 2020) focused on a method for decreasing the discomfort caused by the delay in the transmission of the robot camera stream while wearing an HMD. They propose to use a pre-computed panoramic reconstruction of the environment, decoupling the user's and the robot's view spaces. They tested the configuration using an Oculus Rift DK2 and an iCub robot in teleoperation tasks to assess the effects of latency on head movement and accuracy. The panoramic reconstruction improved the comfort during teleoperation that performance only improved for tasks requiring slow head movements.

(Omarali et al., 2020) introduce a modification in the overall hardware setting for the teleoperation, placing a depth camera for the reconstruction of the remote environment on the arm effector rather than on the robot's head. The authors designed three different grasp techniques (direct, pose-to-pose and point-and-click) to be combined with gestures based on Tilt-Brush for controlling the camera viewpoint and scene navigation. The hardware setup included two Kinect V2 cameras for reconstructing the scene, the Oculus Rift with its touch controllers for the scene visualization and manipulation and a Franka Emika's Panda robot.

(Zhou et al., 2020) proposed TOARS, an advanced teleoperating system designed for controlling a Baxter robot. The system exploits Unity and the ROS framework for building different components, employing deep learning techniques to overcome the delay between the camera capture and the HMD rendering (HTC Vive). The system can reconstruct the environment from the point cloud provided by a Kinect, simplifying the geometry and speeding up the rendering in VR. In addition, it also contains a scene understanding algorithm that can replace the point-cloud representation of know objects with 3D prefabs. Such a feature is reasonable in controlled settings (e.g., assembly tasks) where most of the objects in the scene are known in advance. No evaluation of the teleoperation experience was reported.

(Wonsick and Padır, 2021) propose a VR teleoperation system for the NASA Valkyrie robot, a 32 degree of freedom (DOF) humanoid robot designed to compete in the DRC Trials in December 2013. The VR interface uses an HTC Vive HMD and its controllers, and a waist tracker. The scene rendering is powered by Unity and the robot is controlled through the ROS framework. Differently from other work in this field, the proposed interface splits the interaction into a planning and an executing step. The planning is achieved through different menus and gesture tracking using the Vive controllers. The interface visually shows the plan's results (e.g. the steps in a navigation task) using a ghosting technique (replicas of the relevant robot parts are displayed in the environment).

3.3 Homunculus

The only sample of a Homunculus (Lipton et al., 2018) The authors instantiate the homunculus technique exploiting a Baxter robot, an Oculus Rift and the associated Touch Controllers. The VR environment is based on Unity, while the communication relies on the ROS platform. Through this setting, they compare the direct mapping with an implementation of the homunculus technique on assembly manipulation tasks, but the result cannot show differences in the user's performance.

4 EVALUATING A TELEOPERATION INTERFACE

In different papers we registered a simple qualitative assessment of the interface, which does not provide many information for further development of the teleoperating techniques. For instance, in (Martinez-Hernandez et al., 2017) the authors reports only a qualitative comments on the interface without any formal evaluation. Bian et al. (Bian et al., 2018) show only that an operator was able to conclude a grabbing and pushing task trough the proposed interface.

The most commonly investigated aspect in the teleoperation interface evaluation is the effectiveness, i.e., the ability of a human operator to conclude a manipulation or a navigation task correctly. While this is surely the key aspect in this subject since it demonstrates the feasibility of the approach, little attention is devoted to the operator's experience. Completing a task successfully is only the beginning of a good experience. For a real-world adoption of such a technique, the interface should require a reasonable effort (e.g., comparable with the effort required in a co-located setting) and provide a good level of satisfaction for the human user.

For instance, in (Kilby and Whitehead, 2017) the evaluation was performed on a gross and a fine-grained motor task. The gross motor tasks was dropping a cube of 15cm off a platform. The fine grained motor task was grasping a mug from the handle without dropping it. People liked the direct mapping of the arm movements to the robot and the first person view through the HMD. In (Chen et al., 2017), the evaluation includes a grabbing task of Lego blocks and stacking them together. They tested whether including or excluding the VR and haptic feedback affected the user's rating and effectiveness and found a preference for the system, including such components. Whitney et al. (Whitney et al., 2020) describe a sample setting for teleoperating a Baxter robot in 24 manipulation tasks, showing good performance of human teleoperators, but also some limitations in the task success related to a low perception of the force exerted by the robot in grabbing or pushing objects.

Another common aspect included in the evaluation of a teleoperation system is the level of accuracy obtained in the movement mapping. In particular, such an assessment is reported when the authors provide a custom kinematic mapping technique between the teleoperator and the robot movements. Examples of this approach are available in (Spada et al., 2019) and (Nakanishi et al., 2020). In the latter work, there is an attempt to provide reusable guidelines: they register a low difference between the planned and the executed movements and also a high success rate in the different tasks. In addition, they remark that tasks requiring precise alignment of the orientation of the gripper with force interaction are challenging. Unfortunately, there is no evaluation of the teleoperator experience.

We also registered a few evaluations that include data on the teleoperator experience with the proposed interface. They are not common in interfaces employing a direct mapping, which usually focus on accuracy. An exception is the work by (Hirschmanner et al., 2019). The authors evaluated the usability of the proposed approach on two tasks: grabbing and pouring. They compared their teleoperation approach against the kinesthetic guidance, during which the robot's joints are moved by hand into the position desired. The results show that the users preferred the teleoperation mode. They completed the operation faster, putting less effort into the task requiring the control of two arms.

Instead, the evaluation of cyber-physical approaches usually includes some comparison with other techniques (e.g., a direct mapping). This may be explained considering that the increased complexity in the interface development should be justified with some advantages on the user's side. For instance, Omarali et al. (Omarali et al., 2020) show that the in-hand camera and the proposed visualization technique based on point-cloud reconstruction (which tries to overcome occlusion problems) improves the user's scene understanding if compared against a standard point-cloud visualization. In addition, they show that users prefer gestures over physical movements for scene navigation. Wonsik and Padir (Wonsick and Padır, 2021) compare their VR teleoperation interface of the NASA Valkyrie robot to the currently available 2D counterpart, highlighting a lower workload, a higher awareness of the remote environment, but a more complex hardware setup.

The most complete user-centric evaluation we found in our review is the one available in a follow-up work by Whitney et al. (Whitney et al., 2020). The group compared four interface types for teleoperating a robot using the ROS Reality framework: i) Direct Manipulation (basically the kinesthetic guidance, where the user moves the robot's arm and wrist for completing the task), ii) Keyboard and Monitor, iii) Positional Hand Tracking with Monitor and iv) Positional Hand Tracking with Camera Control, which is basically the same configuration discussed in the previous work (Whitney et al., 2018). The evaluation procedure required the participant to teleoperate the robot for stacking three cups, starting with the direct manipulation first for familiarising with the robot motion. Then, they proceeded with the following conditions in randomised order. The results are encouraging for the VR version: it was faster than the other virtual conditions (but slower than the direct manipulation). It registered a lower workload and higher perceived usability.

In our view, the latter approach should be the standard for further evaluation work in this field. It includes information on both the accuracy and the effectiveness of the evaluated tasks. Still, it also collects standard usability metrics such as the cognitive load through a NASA TLX (Hart, Sandra G. and Staveland, Lowell E., 1988), the overall usability through the SUS questionnaire (Brooke et al., 1996) and other user-experience self-reported metrics through Likert scales. Including such data in the evaluation allows the researchers and the practitioners to better identify possible pitfalls in the teleoperation control, advantages and disadvantages of the input techniques and so on.

Further, defining standard manipulation and navigation tasks to be completed through teleoperation interfaces would be helpful for the replicability of the research and for a systematic comparison of the results. A good starting point may be the settings and the tasks defined for challenges, such as WRS (WRS,). Still, such a standard should emerge from the research community and take into account usability-related aspects.

In summary, while many techniques are available for teleoperating robots in VR, we still need research to evaluate them and create a reliable set of guidelines for adopting such practice in production settings.

5 CONCLUSIONS

In this paper, we surveyed 23 papers written between 2017 and 2021 employing a consumer VR headset for teleoperating a humanoid robot. We categorised the contribution according to the available teleoperation techniques, finding two main competitors: direct mapping and cyber-physical models. The former is easier to implement and understand but suffers the delays between the user's actions and the feedback in the remote location. Cyber-Physical interfaces apply different approaches for mitigating or overcoming such delays, but use virtual replicas of the environment and/or the controlled robot, thus requiring more for both the user and the developers.

We registered a wide variety of solutions for controlling robot movements, especially for the arms, hands, and navigation tasks. We are far away from standard solutions in this regard. Only the head orientation mapped on the HMD internal sensor seems widely accepted.

Finally, there is a lack of consideration of the teleoperation experience in validating the reviewed approaches. There are some remarkable exceptions, but further work is required for moving from the mere assessment of the task completion to a synergistic evaluation of both the robot and the teleoperator performance (and satisfaction). Identifying standard tasks and environments for comparing the different interfaces would foster the adoption of such a perspective.

REFERENCES

World Robot Summit. Partner Robot Challenge. https://worldrobotsummit.org/en/wrs2020/challenge/service/partner.html. Accessed: 2021-08-02.

Bian, F., Li, R., Zhao, L., Liu, Y., and Liang, P. (2018). Interface Design of a Human-Robot Interaction System for Dual-Manipulators Teleoperation Based on Virtual Reality*. In *2018 IEEE International Conference on Information and Automation (ICIA)*, pages 1361–1366.

Botev, J. and Lera, F. J. R. (2020). Immersive Telepresence Framework for Remote Educational Scenarios. In *Learning and Collaboration Technologies. Human and Technology Ecosystems*, pages 373–390. Springer, Cham.

Brooke, J. et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.

Cardenas, I. S., Kim, J.-H., Benitez, M., Vitullo, K. A., Park, M., Chen, C., and Ohrn-McDaniel, L. (2019). Telesuit: design and implementation of an immersive user-centric telepresence control suit. In *Proceedings of the 23rd International Symposium on Wearable Computers*, ISWC '19, pages 261–266, New York, NY, USA. Association for Computing Machinery.

Chen, J., Glover, M., Yang, C., Li, C., Li, Z., and Cangelosi, A. (2017). Development of an Immersive Interface for Robot Teleoperation. In *Towards Autonomous Robotic Systems*, pages 1–15. Springer, Cham.

Elobaid, M., Hu, Y., Romualdi, G., Dafarra, S., Babic, J., and Pucci, D. (2019). Telexistence and Teleoperation for Walking Humanoid Robots. In *Intelligent Systems and Applications*, pages 1106–1121. Springer, Cham.

Gaurav, S., Al-Qurashi, Z., Barapatre, A., Maratos, G., Sarma, T., and Ziebart, B. D. (2019). Deep Correspondence Learning for Effective Robotic Teleoperation using Virtual Reality. In *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*, pages 477–483. ISSN: 2164-0580.

Girard, C., Calderon, D., Lemus, A., Ferman, V., and Fajardo, J. (2020). A Motion Mapping System for Humanoids that Provides Immersive Telepresence Experiences. In *2020 6th International Conference on Mechatronics and Robotics Engineering (ICMRE)*, pages 165–170.

Hart, Sandra G. and Staveland, Lowell E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139 – 183.

Hirschmanner, M., Tsiourti, C., Patten, T., and Vincze, M. (2019). Virtual Reality Teleoperation of a Humanoid Robot Using Markerless Human Upper Body Pose Imitation. In *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*, pages 259–265. ISSN: 2164-0580.

Kilby, C. and Whitehead, A. (2017). A Study of Viewpoint and Feedback in Wearable Systems for Controlling a Robot Arm. In *Advances in Human Factors in Wearable Technologies and Game Design*, pages 136–148. Springer, Cham.

Langevin, G. (2014). Inmoov-open source 3d printed life-size robot. *pp. URL: http://inmoov. fr, License: http://creativecommons. org/licenses/by–nc/3.0/legalcode*.

Lentini, G., Settimi, A., Caporale, D., Garabini, M., Grioli, G., Pallottino, L., Catalano, M. G., and Bicchi, A. (2019). Alter-Ego: A Mobile Robot With a Functionally Anthropomorphic Upper Body Designed for Physical Interaction. *IEEE Robotics Automation Magazine*, 26(4):94–107. Conference Name: IEEE Robotics Automation Magazine.

Lipton, J. I., Fay, A. J., and Rus, D. (2018). Baxter's Homunculus: Virtual Reality Spaces for Teleoperation in Manufacturing. *IEEE Robotics and Automation Letters*, 3(1):179–186. Conference Name: IEEE Robotics and Automation Letters.

Martinez-Hernandez, U., Boorman, L. W., and Prescott, T. J. (2017). Multisensory Wearable Interface for Immersion and Telepresence in Robotics. *IEEE Sensors Journal*, 17(8):2534–2541. Conference Name: IEEE Sensors Journal.

Mizuchi, Y. and Inamura, T. (2017). Cloud-based multimodal human-robot interaction simulator utilizing ROS and unity frameworks. In *2017 IEEE/SICE International Symposium on System Integration (SII)*, pages 948–955. ISSN: 2474-2325.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *PLoS medicine*, 6(7):e1000097.

Nakanishi, J., Itadera, S., Aoyama, T., and Hasegawa, Y. (2020). Towards the development of an intuitive teleoperation system for human support robot using a VR device. *Advanced Robotics*, 34(19):1239–1253. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01691864.2020.1813623.

Omarali, B., Denoun, B., Althoefer, K., Jamone, L., Valle, M., and Farkhatdinov, I. (2020). Virtual Reality based Telerobotics Framework with Depth Cameras. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1217–1222. ISSN: 1944-9437.

Orlosky, J., Theofilis, K., Kiyokawa, K., and Nagai, Y. (2020). Effects of Throughput Delay on Perception of Robot Teleoperation and Head Control Precision in Remote Monitoring Tasks. *PRESENCE: Virtual and Augmented Reality*, 27(2):226–241.

Spada, A., Cognetti, M., and Luca, A. D. (2019). Locomotion and Telepresence in Virtual and Real Worlds. *Human Friendly Robotics*, pages 85–98. Publisher: Springer, Cham.

Whitney, D., Rosen, E., Phillips, E., Konidaris, G., and Tellex, S. (2020). Comparing Robot Grasping Teleoperation Across Desktop and Virtual Reality with ROS Reality. *Robotics Research*, pages 335–350. Publisher: Springer, Cham.

Whitney, D., Rosen, E., Ullman, D., Phillips, E., and Tellex, S. (2018). ROS Reality: A Virtual Reality Framework Using Consumer-Grade Hardware for ROS-Enabled Robots. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1–9. ISSN: 2153-0866.

Wonsick, M. and Padır, T. (2021). Human-Humanoid Robot Interaction through Virtual Reality Interfaces. In *2021 IEEE Aerospace Conference (50100)*, pages 1–7. ISSN: 1095-323X.

Xi, B., Wang, S., Ye, X., Cai, Y., Lu, T., and Wang, R. (2019). A robotic shared control teleoperation method based on learning from demonstrations. *International Journal of Advanced Robotic Systems*, 16(4):1729881419857428. Publisher: SAGE Publications.

Zhang, T., McCarthy, Z., Jow, O., Lee, D., Chen, X., Goldberg, K., and Abbeel, P. (2018). Deep Imitation Learning for Complex Manipulation Tasks from Virtual Reality Teleoperation. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5628–5635. ISSN: 2577-087X.

Zhou, T., Zhu, Q., and Du, J. (2020). Intuitive robot teleoperation for civil engineering operations with virtual reality and deep learning scene reconstruction. *Advanced Engineering Informatics*, 46:101170.