Effects of Augmented Reality on the Performance of Teleoperated Industrial Assembly Tasks in a Robotic Embodiment

Filippo Brizzi, *Student Member, IEEE*, Lorenzo Peppoloni, *Member, IEEE*, Alessandro Graziano, Erika Di Stefano, Carlo Alberto Avizzano, *Senior Member, IEEE*, and Emanuele Ruffaldi, *Member, IEEE*

***Abstract*—Teleoperation in robotic embodiments allows operators to perform and program manipulation tasks with better accuracy, dexterity, and visualization than what is possible with traditional human–robot interaction paradigms. However, the perception of cues (e.g., egocentric distances) relevant to task execution, is known to be distorted in virtual environments due to many factors, which can be grouped into technical, human, and methodological categories. This phenomenon becomes more pronounced in a low-cost/encumbrance setup, where the dynamic environment is captured with color and depth (RGB-D) cameras and presented in a virtual environment. In this paper, the effects of augmented reality (AR) are evaluated as a tool to deliver additional information, which helps in overcoming the differences in perception between telepresence and actual presence. The AR feedback is used to improve the embodiment illusion and to guide the operator during task execution. The AR setup, comprising an RGB-D camera and a head-mounted display, is integrated with the Baxter robot and evaluated by involving 22 participants in an experiment while they execute a pick-and-place task, taking into account their expertise in AR/virtual reality (VR) and gaming. The use of AR results in enhancing the accuracy and efficiency of the task performance, besides significantly reducing the effect of the differences in skillfulness between the participants. Furthermore, it is found that the sense of presence and embodiment for the participant is positively affected by different types of AR.**

***Index Terms*—Augmented reality (AR), humanoid robots, intelligent manufacturing systems, telerobotics, wearable sensors.**

I. Introduction

RECENT advances in robotics are characterized by an impressive evolution of humanoid robots that culminated in the DARPA Robotics Challenge in June 2015. Humanoid robots are designed to operate in human environments with minimal modifications. This type of robot can be used, particularly, in hazardous scenarios, such as rescue missions and industrial manufacturing. Despite humanoid robots' capability to autonomously perform a gamut of tasks, these robots still need human guidance when executing complex tasks, typically by means of teleoperation. In the particular case of teleoperation through body-based interfaces, humanoid robots have proved to be ideal, thanks to the capability of straightforward mapping between the human operator's motion and the robot's motion [1]–[3].

Nevertheless, there are several aspects that make teleoperation a nontrivial task, such as delay and high latency in end-to-end communication, visualization issues of remote environment, difficulties in identifying the right objects to interact with, and in judging the objects' distances from the robot end effector [4]. Producing a teleoperation system that can provide the operator with the same quality and quantity of visual information as what would be possible when the operator is physically present in the remote place is an extremely complex task. The less complex methods utilize a static camera and a monitor [5], [6] to provide a monovision feedback, where the perception of distances is constrained by the lack of parallax. Such constraints are usually overcome through the learning process of the operator. When the teleoperation scenario requires an advanced form of visual feedback, either multimodal feedback is integrated [7] or the operator is "cheated" by utilizing some advanced visualization technique to increase his/her illusion of being in the remote place [8], [9]. Complex visual feedback usually requires implementation of a virtual version of the remote environment. But, the use of virtualized environments leads to perceptual issues, such as the perception of egocentric distances [10], and loss of visual acuity and contrast [11], which have been empirically proven to influence the action capabilities of the body [12].

Augmented Reality (AR) has proven [13] to be a viable solution to overcome visual feedback limitations by providing additional information to the operator. AR solutions for teleoperation can be divided into two categories, which are mutually complementary: embodiment enhancement and virtual fixture.

Embodiment enhancement refers to the capability of making the operator feel that he/she is in the remote environment, and he/she is the robot [14]. Such capability is obtained by placing the camera on the head of the robot and showing the received

Manuscript received May 25, 2016; revised December 22, 2016, July 27, 2017, September 28, 2017, and October 30, 2017; accepted December 1, 2017. Date of publication December 29, 2017; date of current version March 13, 2018. This work was supported in part by the TAUM Project cofinanced by the European Structural Fund for Regional Development, in part by the Programme POR CReO FESR 2007–2013 of Regione Toscana, and in part by the H2020 European Project RAMCIP under Grant 643433. This paper was recommended by Associate Editor Julie A. Adams. *(Corresponding author: Emanuele Ruffaldi)*

The authors are with the PERCRO Laboratory, TeCiP Institute, Scuola Superiore Sant'Anna, Pisa 56100, Italy (e-mail: fi.brizzi@sssup.it; l.peppoloni@sssup.it; a.graziano@sssup.it; e.distefano02@gmail.com; c.avizzano@sssup.it; e.ruffaldi@santannapisa.it).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2017.2782490

image to the operator by means of a head-mounted display (HMD). The vision problem associated with this camera-based approach is generically considered trivial, although it is often one of the major causes of poor performance in teleoperation tasks. The motivation behind using this method is enabling the operator to see the current position of the end effector by eliminating the problems caused by the limited field of view and low resolution of the cameras, and the bulky size of the robots, which obstructs the scenes view.

The second category of AR solutions employs virtual fixtures [15] to help the operator in overcoming the difficulties of perceiving the remote environment, and to guide him or her in accomplishing the tasks. Virtual fixtures refer to virtual images or objects, overlain on the remote scene, to provide the operator with visual cues that highlight the points of interest and useful information for accomplishing the given task [16]. Virtual fixtures have been proved to be capable of speeding up the execution of a teleoperation task, especially in high-latency scenarios [17]. Virtual fixtures also serve as an effective tool in teleoperation by providing sensory substitution, particularly in perceiving force feedback in the absence of haptic device [18].

Although AR has been successfully applied to reduce the rate of task error in many robotic teleoperation scenarios, no studies were so far carried out to assess the impact of AR multiple features.

This paper studies the effects of AR in a generic industrial assembly scenario. In particular, this paper addresses the use of both task-related and nontask related features and their combination, by quantifying the features' effects not only on the task performance, but also on the operator's sense of telepresence and embodiment. The study shows that specific findings are associated with different types of operators' expertise in AR/virtual reality (VR) and gaming.

The remainder of this paper is organized as follows. Section II describes the experiment's rationale, setup and trials, followed by data analysis and participants' demographic particulars. Section III reports the results of experimental data. Section IV discusses the results and its implications for future development.

II. Materials and Methods

The fundamental question addressed in this paper is how different types of AR features impact the operator's experience and performance. The other issues addressed by this paper include: 1) understanding the extent to which excessive visual information can be detrimental; and 2) evaluating the nontask specific AR features. Therefore, the focus was on exploring the effects of three-dimensional (3-D) AR feedback on the presence, embodiment, and ease of task execution, while accounting for the possible effects of participants' expertise. Measures were taken to minimize the learning effect that could arise during the execution of the experimental trials.

All these questions were sought to be answered by undertaking an experiment that simulates an industrial assembly scenario, which forms one primary application field for AR. The tasks involve a pick-and-place operation by teleoperating a robot, seen from an egocentric point of view. The task execution is parametrized in terms of completion time and placement accuracy. Furthermore, the operators' hand trajectories were analyzed, together with their subjective sense of telepresence and embodiment, in different conditions. Also, the combinations of the chosen features were evaluated.

A. Augmented Reality

To create accurate and effective AR feedback, it is necessary to extract as much information as possible from the remote environment, exploiting all the available sensors. In the presented scenario, the available sensors were a color and depth (RGB-D) camera and the robot's joint encoders. Exploiting the state-of-the-art computer vision algorithms, it is possible to track the pose of the target objects to be manipulated. The calibration between the robot and the camera allows for colocating the robot and the remote environment by obtaining their exact relative poses. This information can be used as an AR feature, because it is not directly inferable by humans owing to the difficulty in perceiving distances accurately in a virtual scenario [19].

Different types of augmented information were implemented. The features can be categorized into two classes: *embodiment* and *visual virtual fixtures*. These two categories of AR features were chosen because they are considered to be possibly the most informative ones to aid task execution.

The embodiment class comprises information that can help the operator in improving the overall sensation of embodiment and illusion of presence. In the setup used for this paper, the embodiment class comprises two features. The first feature enables the operator to explore the remote environment by changing the virtual viewpoint with head and body movements. The second, and novel feature, is a virtual 1:1 scale model of the remote robot that is animated by the real robot's movements [see Fig. 1(a)]. This feature allows the operator to see the position of the robot's end effector when it is not visible in the camera's field of view. The display helps the operator understand the robot's position in the remote environment.

The chosen visual virtual fixtures, listed in Table I, provide different types of information, which can be classified into two main subgroups, according to the information they deliver: *manipulation information* and *task information*. The fixtures that belong to the manipulation group [see Fig. 1(b)] deliver additional information, relating to object manipulation, such as distances from the object's grasp points (green bar) and robot grip closure (blue bar). The fixtures also include a red 3-D beam representing, in real time, the optimal trajectory between the robot end effector and the closest grasp point of the object. The visual virtual fixtures that belong to the task group are used to highlight information relating to the task execution. The fixtures are characterized by a green 1:1 3-D mesh of the real object placed in the task's target pose [see Fig. 1(c)]. AR features that deliver information, relating to the task execution, have been demonstrated to significantly reduce error rates in assembly tasks [20].

B. Implementation

The AR feedback component was developed using an robot operating system (ROS)-integrated framework for high-

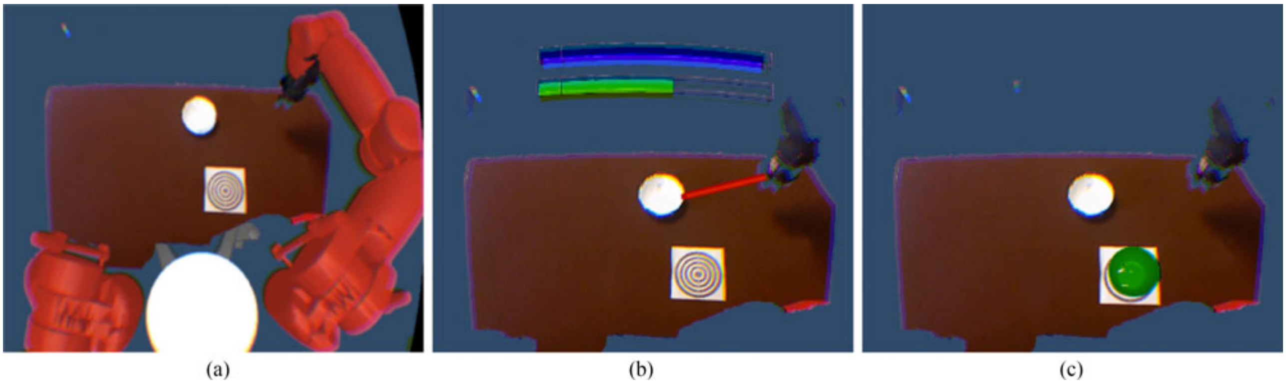


Fig. 1. Three different groups of AR features (a)–(c) used in the experiment. (a) *Embodiment*. Image of the 1:1 scale model of the remote robot animated by the real robot's movements. The overlaid model allows us to see the position of the robot limbs even if not in the real camera view space. The picture shows the precise overlay of the 3-D model of the robot with the pointcloud from the 3-D camera. (b) *Manipulation information*. Blue bar: gripper closure. Green bar: distance from target. Red beam: shortest trajectory from the robot's end effector to the bowl's grasp point. (c) *Task information*. In the target pose, a green 3-D mesh of the task object (bowl) is placed. The placement of the object succeeds if the real bowl and the green one perfectly match.

TABLE I
IMPLEMENTED VISUAL CUES AND THE CORRESPONDING CLASSES AND GROUPS

Class	Group	Feature
Embodiment	–	3-D robot model
Virtual fixture	Manipulation information	Trajectory to the object grasping point (3-D beam)—Distance from the target object (color bar)—Gripper closure (color bar)—Mesh of the object to be grasped
Virtual fixture	Task information	Objects target poses

performance AR and mixed reality (MR), called compact components (CoCo). CoCo is composed of a core library and several modules, each specialized in managing particular elements of an AR/MR application [21]. CoCo provides for high-performance execution on modern multicore machine, thanks to the support available for parallel programming constructs. The use of CoCo for the MR display, instead of the ROS's RViz viewer, is motivated by the high visual feedback requirements and VR devices support. The CoCo library is used to create a 3-D mesh from camera streams and to augment it with additional information. In particular, the first CoCo module receives and decompresses the video and depth streams from the camera. The color channel is streamed using H.264 compression (435 kbps on average) and the depth channel with zn16 compression from OpenNI2 (21 Mbps on average). The second module reconstructs a pointcloud with interpolated points from the decompressed buffers. This operation is performed to enhance the quality of the mesh of the virtual scene. The third module does the graphic rendering for the augmented scene. The end-to-end latency of the visualization is 89 ms, as computed after synchronizing the robot and graphics computers with the precision time protocol [22].

An ROS-integrated teleoperation setup, analogous to one previously presented by Peppoloni *et al.* [23], was used to evaluate the effects of AR in a remote teleoperation framework. The systems architecture is shown in Fig. 2. On the operator's side, a wearable device captures the movements of the operator's upper limb and his pinch grip position through inertial sensors and a custom haptic device [24]. The raw sensor data were sent, via wireless, to the main computational unit, where several ROS nodes reconstruct the operator's motion and combine them with the grip position to generate a control signal for the teleoperated robot (Baxter robot, Rethink Robotics, Boston, MA, USA). On the remote robot side, the environment in which the robot was acting was captured through a Kinect 360 camera (Microsoft, Redmond, WA, USA) placed on the top of the Baxter's head, as shown in Fig. 4. The camera was not actuated and its field of view was fixed with respect to the robot's pose. From the captured pointcloud, a virtual scene was created, which was augmented by the main computational unit with AR information coming from the ROS control node. The visual appearance of the robot model used in the embodiment feedback is based on the universal robot description format of the Baxter robot. The 3-D AR scene was sent as a visual feedback to the operator's side and visualized with an HMD (Oculus Rift DK2, Facebook, Menlo Park, CA, USA). Further details about the motion reconstruction algorithms, the haptic interface, and the control architecture can be found in the prior work of the authors [23]–[25].

C. Participants

In total, 22 operators (16 male, 6 females), aged between 23 and 40 years, all right handed, participated in the study after giving their informed consent for participation. Their familiarity with AR and VR systems and video games was assessed using a Likert scale related to expertise (1–7), and in terms of number of hours the operators spend on these technologies per week (1 for less than 1 h, 2 for 2–5 h, and 3 for more than 5 h). For this, they were given the option of choosing between three answers: less than 1 h, between 2 and 5 h, and more than 5 h. For what concerns expertise score AR/VR resulted in an average score of 5.05 ($\pm$ 2.01), whereas video games in 5.38 ($\pm$1.80). The time spent on AR/VR resulted in an average score of 1.48 ($\pm$0.75), whereas video games in 1.62 ($\pm$0.92).

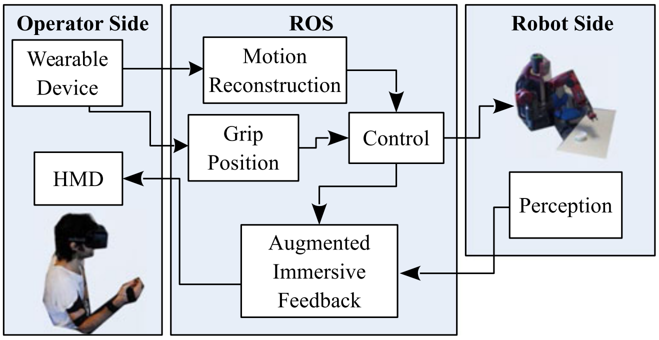


Fig. 2. System's setup. On the operator's side, the upper limb motion and the grip position were captured through a wearable interface. Operator's motion was reconstructed and used together with the grip position, to generate a control signal for the teleoperated robot. The remote scene was captured and virtualized; AR information was added to the virtual scenario, which can be visualized on the operator's side as a 3-D visual feedback.

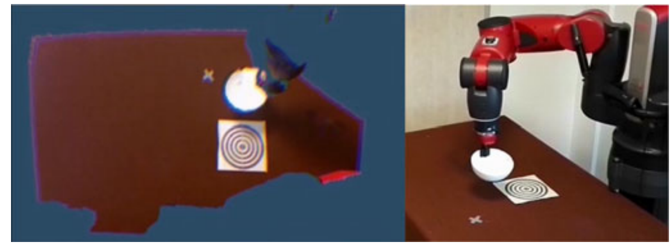


Fig. 3. Virtual and the real remote environments captured simultaneously. The image on the left shows the 3-D environment captured from the camera and rendered in the HMD, whereas the right image shows the real environment from an external point of view. In both images, it is possible to see the robot grasping the target object (bowl), the starting position of the object (cross marker), and the target position (circular marker).

D. Experimental Protocol

The participants completed a single experimental session. Prior to the commencement of the session, the participants were asked to read a written description of the experiment and completed a questionnaire to assess their expertise. After donning the wearable interface, the participants were asked to perform a series of familiarization trials, followed by experimental trials that involved performing remotely a pick-and-place task of an object by teleoperating the Baxter robot, simulating the classical step of an industrial assembly task. The object was a plastic bowl, which measured 16 cm in diameter and 7 cm in height. The bowl was chosen because it had radial symmetry that helped in grasping without introducing any rotational error while placing it. To normalize the task execution among the participants, the starting and target positions of the bowl on the table were fixed. A post-experiment questionnaire assessed participants' experience with AR during the trails (see Appendix for the specific questions).

1) Familiarization: Approximately 5-min familiarization trials allowed the participants to practice the task so that they can perform the task correctly in its entirety and minimize the effects of their learning during the experimental trials. In particular, the participants had to perform the task of each trial within a specified time frame. The allocated time frame ensured the time for the robot for estimating the end-effector motion base on participant's upper limb motion, control of the robot over the full manipulation workspace, and proper grasp and release of the remote object with the remote robot gripper. During the first trial, the participants were allowed to share the working space with the robot (i.e., without wearing the HMD) and teleoperate it during several pick-and-place tasks. During the second familiarization trial, the participants had to teleoperate the robot, using the visual feedback (i.e., wearing the HMD) without AR information, thus visualizing only the virtual remote scene. During familiarization, the target object was the same as used in the experiment, but the pick-and-place locations were randomized. The participants performed the trials until they and the investigators were confident that the task could be performed in its entirety. Overall, the participants required 3–4 repetitions of the pick-and-place trial to become proficient in executing the aforementioned aspects of the task.

2) Experimental Trials: The participants were required to perform the pick-and-place task of grasping and moving the object as accurately as possible to its target pose. The participants executed the task under the following five different modalities of visual feedback:

1) no AR information;
2) only AR information relating to the manipulation class is visualized in the feedback;
3) only AR information, related to the embodiment group is visualized;
4) only AR information, relating to the task group is visualized; and
5) full AR feedback (all the features relating to embodiment, manipulation, and task are activated simultaneously).

Fig. 3 shows the remote scenario, wherein the starting point (the cross on the table) and the target point (the circular marker) are visible, and the robot is grasping the target object. The left image shows the virtual scene, as displayed to the participants, whereas the right image shows the real environment.

To minimize the learning effects, the trial order was randomized. To simulate the effect of noncolocation between the robot and the operator, the operator was acoustically isolated from the environment and was asked to teleoperate the robot over the Internet. For full experimental setup see Fig. 4.

To assess the effects of various types of AR feedback, several variables, relating to task execution performance and task execution modality, were monitored. The task completion time and the accuracy of each placement of the object were used as metrics for defining the execution performance. The accuracy was expressed in terms of the Euclidean distance between the actual position and the target position; no orientation error was possible in measuring this metric because the object had radial symmetry.

Before the commencement of each task, the robot arm was moved to a predefined rest position. For computing the completion time of the task, the timer was configured so that it started as soon as the operator started moving the arm to grasp the object, and stopped once the operator released the object and the system tracked its pose.

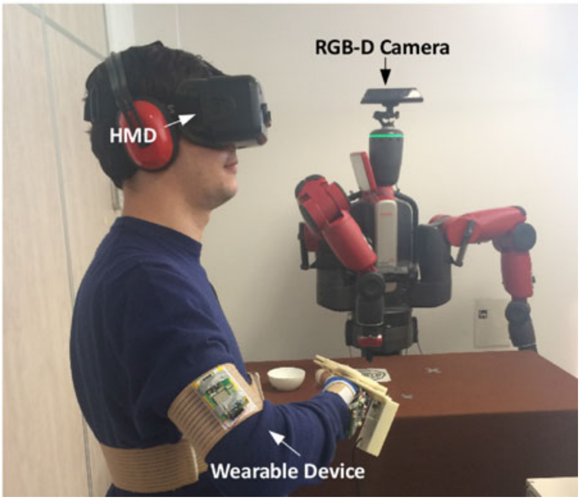


Fig. 4. Full experimental setup. The participant teleoperate the robot over the local network and he/she is acoustically isolated from the environment. The robot-side environment is captured with an RGB-D camera and sent to the operator as a 3-D visual feedback through an HMD.

The trajectories of eight operators' right hands were also recorded at 100 Hz, which denotes the sampling frequency of the inertial sensors (MPU9150, Invensense, San Jose, CA, USA) mounted on the wearable device.

E. Data Analysis

Analysis of the order effect was carried out to ensure that no learning effect occurred between successive trials. The data relating to completion time and placement accuracy were grouped according to their trial order numbers and compared using one-way ANOVA. To quantify the differences in the placement errors and execution times between the five AR feedback modalities, all the distributions were first checked for normality, using the Lilliefors test. Student's *t*-test was used to test for the differences in mean when the resulting data was normally distributed, and MannWhitney *U*-test when the resulting data was not normally distributed.

The smoothness of the participant's right-hand trajectory in the main motion plane (*xy*) was used as an indicator of skillfulness in the task execution [26]. The effects of different AR feedback conditions on the trajectory smoothness were of particular interest.

The smoothness of the trajectory was defined as the normalized jerk ($\hat{J}$), a metric commonly used to determine smoothness

$$\hat{J} = \sqrt{1/2 \int_t j^2(t) \frac{D^5}{\mathrm{L}^2} dt} \tag{1}$$

where D is the duration and L is the length of the trajectory. Low normalized jerk is indicative of smooth trajectory, and a high jerk of a less-smooth trajectory.

Owing to the complexity of the setup, execution anomalies (e.g., unexpected occlusion and grasp problems) may arise, which can drastically increase the values of the variables of interest. Taking these considerations in account and considering

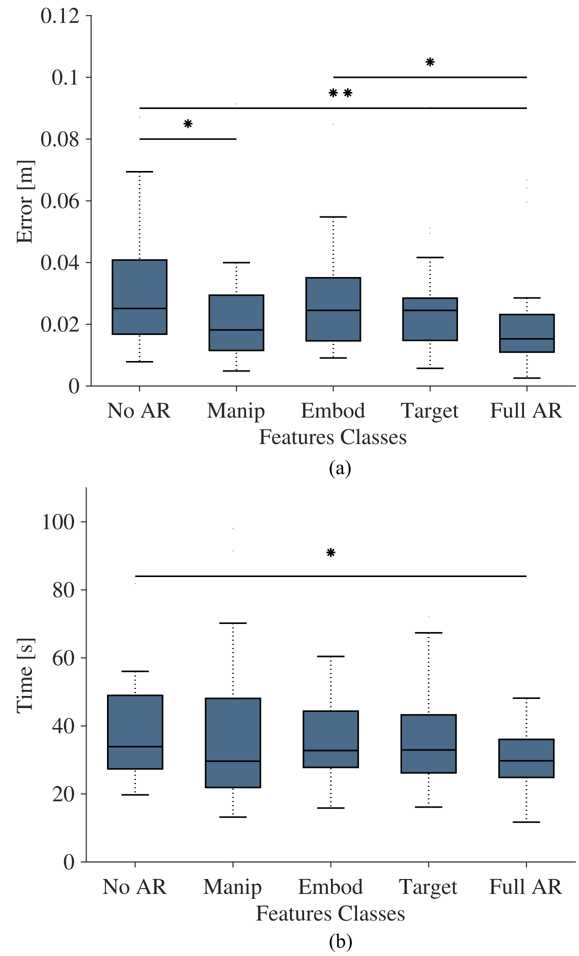


Fig. 5. Results for the variables relating to task execution without considering expertise (20 participants per boxplot). (a) Accuracy for different AR feedback modalities together with significance results ($^{*}p \leq 0.05$, $^{**}p < 0.01$). (b) Execution time for different AR feedback modalities together with the significance results. Each boxplot depicts the statistics of the 8 participants for which this information is reliably available: each box represents the 25th–75th percentile, the black line in the middle is the median, and bars above and below connected with dashed line represent maximum and minimum values.

the possible impact on the statistical analysis, it was decided to eliminate, from each visualization modality, the participants with the maximum and minimum performance scores, resulting in 20 participants per visualization modality.

An additional analysis was performed to account for the effects of training and personal experience, which are known to strongly affect the performance of teleoperation tasks. This analysis divided the participants into two groups according to their self-reported experience in AR/VR and video games. The reported average hours spent weekly interacting with AR/VR environment and playing video games were used to decide whether to insert them into the expert group (at least 2 h weekly) or nonexpert group (less than 2 h weekly). For each condition, the participants with the maximum and minimum scores were removed from the two groups (*expert* and *nonexpert*), resulting in 18 participants for each AR modality.

Each test was considered significant at 95% confidence level ($p \leq 0.05$).

TABLE II
ALL THE MEASURED TIMES (IN SECONDS) AND ERRORS (IN METERS) ARE SHOWED DIVIDED PER PARTICIPANT FOR EACH MODALITY

ID	(1) No AR		(2) Manipulation AR		(3) Embodiment AR		(4) Task AR		(5) Full AR		Order	Expertise
	Time	Error	Time	Error	Time	Error	Time	Error	Time	Error		
1	56.0	0.0168	40.7	0.0132	32.4	0.0292	40.0	0.0251	37.6	0.0129	12345	None
2	54.5	0.0369	27.4	0.0227	118.8	0.0204	14.7	0.0124	48.1	0.0085	23451	AVR and Games
3	23.6	0.0103	24.8	0.0400	29.9	0.0351	22.4	0.0513	11.7	0.0206	34512	AVR
4	32.9	0.0144	54.5	0.0166	33.0	0.0243	35.1	0.0057	44.5	0.0597	45123	None
5	129.4	0.1041	31.7	0.0294	50.6	0.0250	33.3	0.0416	47.0	0.0232	51234	None
6	10.4	0.0408	13.1	0.0369	15.8	0.0847	32.8	0.0277	55.3	0.0667	54321	Games
7	26.4	0.0218	70.2	0.0074	33.3	0.0548	57.0	0.0903	15.4	0.0190	25134	None
8	125.4	0.0559	91.5	0.0163	43.9	0.0389	28.0	0.0195	28.5	0.0641	13524	AVR
9	81.8	0.0144	63.1	0.0914	39.5	0.0329	67.3	0.0495	32.3	0.0026	43215	None
10	38.1	0.0079	98.0	0.0327	44.8	0.0290	72.0	0.0325	30.9	0.0151	12543	None
11	28.2	0.0263	137.9	0.0293	53.0	0.0248	32.7	0.0285	34.4	0.0136	34251	AVR
12	47.4	0.0671	23.3	0.0180	20.8	0.0141	31.1	0.0253	24.7	0.0100	23154	None
13	25.0	0.0193	36.9	0.0389	31.8	0.0392	33.0	0.0175	19.7	0.0116	32514	Games
14	34.8	0.0694	16.5	0.0049	16.9	0.0146	16.3	0.0143	24.9	0.0285	21534	Games
15	38.5	0.0220	10.4	0.0183	25.9	0.0124	30.2	0.0064	32.8	0.0129	45132	None
16	32.8	0.0287	31.9	0.0200	32.0	0.0125	40.2	0.0256	24.9	0.0155	32154	None
17	37.9	0.0259	20.4	0.0248	33.8	0.0191	46.2	0.0068	25.5	0.0193	31245	None
18	30.9	0.0162	25.3	0.0113	60.4	0.0091	57.2	0.0148	41.1	0.0092	45123	AVR and Games
19	50.4	0.0244	41.6	0.0054	46.9	0.0147	96.1	0.0190	32.0	0.0205	43125	None
20	19.7	0.0236	27.5	0.0115	15.8	0.0194	16.1	0.0279	10.6	0.0051	23154	AVR and Games
21	30.6	0.0872	20.3	0.0134	27.9	0.0504	18.9	0.0239	18.0	0.0257	32154	None
22	20.7	0.0357	18.0	0.0114	27.6	0.0119	24.3	0.0236	27.0	0.0110	51432	AVR and Games

Table also reports the competency of each participant, plus the order of the tested conditions.

III. RESULTS

Detailed results are shown in Table II.

A. Order Effect Analysis

The results show no statistical difference either in completion time ($p = 0.3187$) or in accuracy ($p = 0.3688$) among different trials. It is, therefore, plausible to assume statistical significance although not all possible combinations of visualization modalities were tested.

B. Task Execution

Significant group differences were found in the parameters relating to task execution (placement accuracy and execution time) among the five AR feedback modalities, when the analysis is performed on all the 20 participants without accounting for expertise. Lilliefors test shows that both execution time and accuracy results are normally distributed. The accuracy obtained during the task execution, with full AR feedback, is significantly higher ($p = 0.0058$) than that obtained during the task execution with no AR feedback, as also that obtained with the AR features relating to embodiment ($p = 0.0344$). The accuracy obtained during the task execution with AR features relating to manipulation is significantly higher ($p = 0.046$) than accuracy with no AR feedback. The time required to complete the task with full AR feedback is significantly lower ($p = 0.04$) than the time required with no AR feedback. The results are shown in Fig. 5 where part (a) shows the results for accuracy, and part (b) the results for the execution time.

C. Trajectories Analysis

Significant group differences were found in the participant's hand trajectory smoothness values between the AR feedback modalities. Particularly, the normalized jerk on the *xy* plane, obtained during the task execution with AR features relating to embodiment, is significantly lower ($p = 0.002$) than the normalized jerk on the *xy* plane obtained with no AR feedback, with AR feedback relating to the task information ($p = 0.002$), and full AR feedback ($p = 0.029$). The normalized jerk on the *xy* plane, obtained during task execution with AR features relating to manipulation is significantly lower ($p = 0.027$) than the one on the *xy* plane, obtained during the task execution with no AR feedback or with AR feedback, relating to the task information ($p = 0.009$). These statistics, together with the significance results, are shown in Fig. 6.

D. Questionnaires Results

Significant group differences are found in the reported effects of different AR feedback classes on the participant's sense of acting in the remote environment, of embodiment, and on the ease of task execution. Particularly, the AR features relating to manipulation have a significantly stronger effect ($p = 0.024$) on providing the sense of presence in the remote environment, as compared to that of the task-related AR feedback. The virtual robot model has a significantly stronger effect on enhancing the illusion of the embodiment towards the remote robot, as compared to that of the manipulation-related features ($p = 0.044$) or the task-related features ($p = 0.02$). The manipulation-related features are more effective than the task-related features in fa-

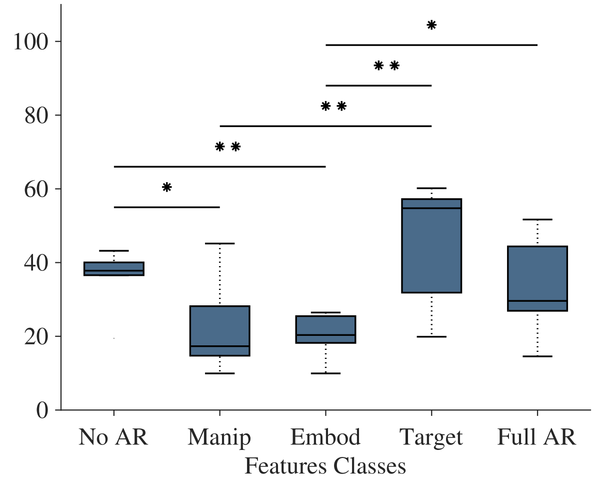


Fig. 6. Results for the smoothness of the participants' hand trajectory. The values of the normalized jerk for the different AR feedback modalities together with the significance results are shown ($^{*}p \leq 0.05$,$^{**}\, p < 0.01$).

cilitating task execution ($p = 0.0138$). These statistics, together with significance results, are shown in Fig. 7.

E. Expertise Analysis

The partitioning of the 22 participants, based on their expertise, resulted in 7 expert participants and 15 nonexpert participants in AR/VR and in 7 expert participants and 15 nonexpert participants in video games. These two expertise types have four participants in common.

The expertise types were analyzed separately and then compared in terms of expert versus nonexpert groups. As discussed in Section II-E, the analysis was performed over 18 participants, after removing the outliers from both expert and nonexpert subgroups for every condition.

Fig. 8 shows the results of the four groups' completion time. Each column is associated with a participant and is grouped, based on visualization modality and expertise. To improve visualization, the bars in the chart are sorted groupwise.

1) AR/VR: The expert group completed the task in significantly less time than the nonexpert group during the task execution with no AR feedback ($p = 0.05$) or with AR features relating to the task information ($p = 0.003$). In terms of the execution time, significant differences are found among the AR/VR games nonexpert groups with different AR feedback modalities. In particular, the time required to complete the task with full AR feedback is significantly less than the time required with no AR feedback ($p = 0.008$) or with the AR features, relating to the task information ($p = 0.015$).

Significant group differences are found, in terms of the execution time, among the AR/VR expert groups with different AR feedback modalities. In particular, the time required to complete the task with the AR features relating to task information is significantly lower ($p = 0.027$) than the time required to complete the task with the AR features relating to embodiment.

Fig. 8(a) shows the results comparing the execution time between AR/VR experts and nonexperts.

2) Video Games: The expert group completed the task in significantly less time than the time taken by the nonexpert

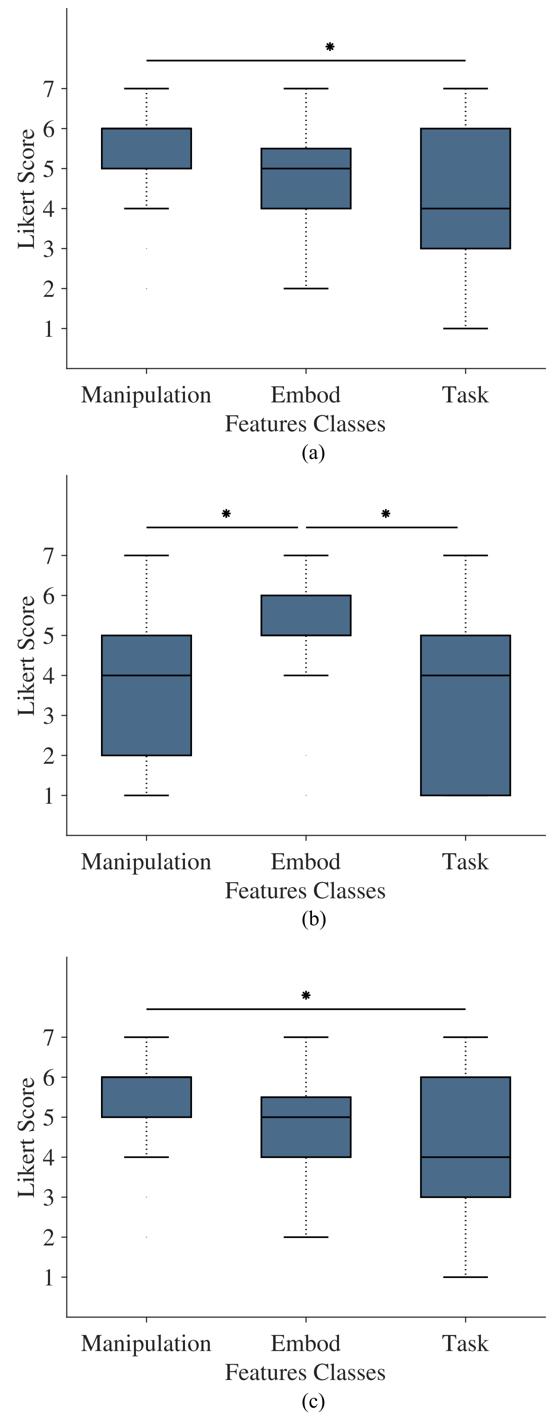


Fig. 7. Reported effects of different AR classes on (a) sense of presence, (b) embodiment, and (c) ease of executing the task. The values from the questionnaires results (from 1 to 7), together with the significance results ($^{*}p \leq 0.05$, $^{**}p < 0.01$), are shown here. Each boxplot depicts the statistics for all the 22 participants.

group with no AR feedback ($p = 0.007$), AR features relating to manipulation ($p = 0.022$), or AR features relating to task information ($p = 0.017$).

Significant group differences are found in terms of the execution time, among the video games nonexpert groups with different AR feedback modalities. In particular, the time required to complete the task with full AR feedback is significantly less

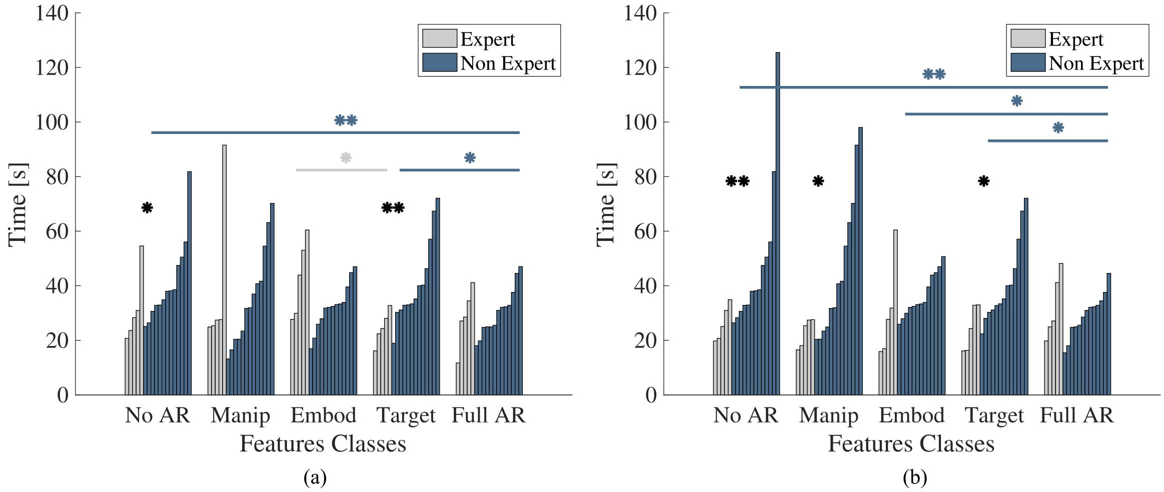


Fig. 8. Task execution time of the expert and nonexpert groups for (a) AR/VR and (b) video games, based on the participants' partitioning, after removing the outliers from each group. Each bar is associated with a participant, with 18 participants per condition. Significance of the comparison between experts and nonexperts for each AR condition is shown with black asterisks. Colored asterisks represent statistical significance of the comparison of different conditions, the expertise group being equal ($^*p \leq 0.05$, $^{**}p < 0.01$). The bars were sorted, based on execution time, to improve visualization.

than the time required with no AR feedback ($p = 0.003$), AR features related to embodiment ($p = 0.02$), or AR features relating to task information ($p = 0.017$). In terms of accuracy, no significant group differences are found between the two groups. Fig. 8(b) shows the results, as also their significance, for comparison of the execution time between video games experts and nonexperts.

IV. Discussions and Conclusion

The experiment revealed some differences between the AR feedback modalities, which possibly answer the research questions (raised in Section II). The AR effects were quantified in terms of placement error of the manipulated object, task completion time, and smoothness of the hand trajectory.

A. AR Effects

Overall, the participants using the full AR feedback had significantly improved execution in terms of both placement error and completion time, as compared to their execution without AR feedback. According to the participants' judgment, the robot model is the most effective AR feature in improving the illusion of embodiment. Information like the target object position, the robot end-effector position or the distance between the robot end effector and the target position can be lost during the execution, due to technological constraints (i.e. , low resolution, nonoptimality of the camera positioning, and view point). According to the experiment's result, all the features that focus in compensating this loss tend to lead to a more skillful and sure execution. The motivation for this result may depend on the responses connected to the improved sense of presence and improved sense of embodiment.

Interestingly, the AR features relating to embodiment gave better execution results with smoother trajectories than the full AR feedback (see Fig. 6). These results possibly indicate that high degree of additional information added to the scene may lead to an execution that could be more accurate, but it is uncertain from motion view point. The motivation probably relies on the fact that the operator could have taken into account too much information at once, during task execution.

The features delivering information about the task execution seem to have negligible impact on task performance. Moreover, compared to the manipulation-related features, these features contributed less to the sense of presence and to task execution, according to the participants' judgment. This result is aligned with the existing literature related to the use of task-related features, which are usually found to improve the task performance [20]. However, the considered scenario did not regard the teleoperation of a robot. These results may also be partially influenced by the chosen task, which does not have a particularly complex or articulated execution. The fact that execution with task-related features is comparable to execution without AR features is further strengthened when the results are evaluated considering the participants' previous expertise. Both groups, who are considered experts in AR/VR, as well as in video games, performed better than their nonexpert counterparts in both no AR and only task-related information modalities. The performance of nonexperts under these two modalities is worse (in terms of completion time) than their execution performance with full AR feedback. The completion time of nonexperts in AR/VR is higher when they used no AR information and when they used the full AR feedback using task-related features compared to the full AR feedback. The completion time of nonexperts in video games' is higher when they used no AR information than when they used the full AR feedback; it is similarly higher when they used task-related features than when they used the full AR feedback.

B. Expertise Effects

Interestingly, the expert gamers completed the task in less time than the nonexpert gamers, even when the AR feedback relating to manipulation information was available. This result may be explained by the fact that video gamers are more proficient at utilizing and relying on synthetic visual cues. This difference is lost with full AR feedback, which can improve the nonexpert groups' performance, and thus level out the difference between experts' and nonexperts' performances.

Overall, the AR feedback affected the performance of the expert groups less than the nonexpert groups. This finding is in line with that of Gwilliam *et al.* [27]. The difference between the effects of AR for the expert gamers is not statistically significant. Instead, the difference of completion time between the conditions with embodiment-related features and task-related features is statistically significant for AR/VR experts. This result could be explained by the fact that experts are more proficient at performing tasks in a virtual scenario. It is plausible to assume that the experts possess skills relating to virtual scenarios, such as judging distances and finding effective strategies of motion. In this regard, AR emerges as a useful tool to assist the first-time operators and less-skilled operators in executing assembly tasks.

C. Conclusions

Although the performance of teleoperation tasks with complex visual feedback is strongly affected by the operator's skill and expertise, additional information delivered with AR seems to help reduce the gap between the performance of expert and nonexpert operators. Therefore, AR could help in shortening the learning curve, so that the operators become proficient in the teleoperation setup and can thus perform better with just a short familiarization with the system. These results may be attributed to the increased sense of presence and embodiment, which benefit from the additional information that can be lost because of technological constraints, but recovered and delivered through AR. The practical implication of these results is that AR feedback limited to task specific information can be useful in supporting expert operators' activity and full AR feedback in supporting nonexpert operators' activity.

Further studies are necessary to fully assess the effects of task-related AR features on task performance. Those studies may require experimental trials with more complex and articulated tasks that possibly involve multiple execution steps. The outcome of this paper would hopefully facilitate future studies about the effect of AR on the learning curve of teleoperation tasks.

Appendix

This is the list of the pre-experimental questions. Where not stated, we used a Likert-type scale at 7 levels.

1) Rate your familiarity with VR/AR.
2) Rate your familiarity with video games.
3) Rate your familiarity with robotic systems .
4) Rate your familiarity with robotics teleoperation.
5) Rate the time you spend weekly using VR/AR applications (mobile, games,...).
6) Rate the time you spend weekly playing video games—Three results as question 5.

And then, the postexperimental questions are as follows.

1) Rate your sense of acting in the remote environment, on the following scale from 1 to 7, where 7 represents your normal experience of acting in the environment.
2) Rate the effect of AR on your sense of acting in the remote environment.
3) Rate the effect of the following visual virtual fixtures on your sense of acting in the remote environment: manipulation (distance represented by the 3-D cylinder, distance, gripper closure, object mesh), embodiment (robot model), and task (mesh of target pose).
4) I had a sense of being the robot.
5) Rate the effect of AR on your sense of being the robot.
6) Rate the effect of the following visual virtual fixtures on your sense of being the robot—As in question 3
7) Rate the effect of AR on facilitating the task execution.

Author Contributions

Study concept: F. Brizzi and E. Ruffaldi. Study design: F. Brizzi, L. Peppoloni, and E. Ruffaldi. Study implementation: F. Brizzi, A. Graziano, and E. Di Stefano. Acquisition of data: F. Brizzi, A. Graziano, and E. Di Stefano. Analysis and interpretation of data: L. Peppoloni, F. Brizzi, and E. Ruffaldi. Drafting of manuscript: F. Brizzi, L. Peppoloni, and E. Ruffaldi. Critical revision: all.

References

[1] W. Song, X. Guo, F. Jiang, S. Yang, G. Jiang, and Y. Shi, "Teleoperation humanoid robot control system based on kinect sensor," in *Proc. 2012 4th Int. Conf. Intell. Hum.-Mach. Syst. Cybern.*, 2012, vol. 2, pp. 264–267.

[2] S. Chang, I. Kim, J. H. Borm, Ch. Lee, and J. O. Park, "KIST teleoperation system for humanoid robot," in *Proc. IEEE RSJ Int. Conf. Intell. Robots Syst.*, 1999, vol. 2, pp. 1198–1203.

[3] N. Miller, O. C. Jenkins, M. Kallmann, and M. J. Mataric, "Motion capture from inertial sensing for untethered humanoid teleoperation," in *Proc. 4th IEEE/RAS Int. Conf. Humanoid Robots*, 2004, vol. 2, pp. 547–565.

[4] V. Nitsch and B. Farber, "A meta-analysis of the effects of haptic interfaces on task performance with teleoperation systems," *IEEE Trans. Haptics*, vol. 6, no. 4, pp. 387–398, Oct.–Dec. 2013.

[5] A. K. Bejczy, W. S. Kim, and S. C. Venema, "The phantom robot: Predictive displays for teleoperation with time delay," in *Proc. IEEE Int. Conf. Robot. Autom.*, 1990, pp. 546–551.

[6] K. Goldberg, M. Mascha, S. Gentner, N. Rothenberg, C. Sutter, and J. Wiegley, "Desktop teleoperation via the world wide web," in *Proc. IEEE Int. Conf. Robot. Autom.*, 1995, vol. 1, pp. 546–551.

[7] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review," *Psychonomic Bull. Rev.*, vol. 20, no. 1, pp. 21–53, 2013.

[8] C. L. Fernando *et al.*, "Design of TELESAR V for transferring bodily consciousness in telexistence," in *Proc. 2012 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2012, pp. 5112–5118.

[9] L. Fritsche, F. Unverzag, J. Peters, and R. Calandra, "First-person teleoperation of a humanoid robot," in *Proc. 2015 IEEE-RAS 15th Int. Conf. Humanoid Robots*, 2015, pp. 997–1002.

[10] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert, "The perception of egocentric distances in virtual environments—A review," *ACM Comput. Surveys*, vol. 46, 2013, Art. no. 23.

[11] M. A. Livingston *et al.*, "Basic perception in head-worn augmented reality displays," in *Human Factors in Augmented Reality Environments*. New York, NY, USA: Springer, 2013.

[12] S. A. Linkenauger, H. H. Bülthoff, and B. J. Mohler, "Virtual arm's reach influences perceived distances but only after experience reaching," *Neuropsychologia*, vol. 70, pp. 393–401, 2015.

[13] G. Mosiello, A. Kiselev, and A. Loutfi, "Using augmented reality to improve usability of the user interface for driving a telepresence robot," *J. Behav. Robot.*, vol. 4, no. 3, pp. 174–181, 2013.

[14] P. Zahorik and R. L. Jenison, "Presence as being-in-the-world," *Presence*, vol. 7, no. 1, pp. 78–89, 1998.

[15] L. B. Rosenberg, "Virtual fixtures: Perceptual tools for telerobotic manipulation," in *Proc. IEEE Virtual Reality Annu. Int. Symp.*, 1993, pp. 76–82.

[16] C. P. Sayers, R. P. Paul, L. L. Whitcomb, and D. R. Yoerger, "Teleprogramming for subsea teleoperation using acoustic communication," *IEEE J. Ocean. Eng.*, vol. 23, no. 1, pp. 60–71, 1998.

[17] T. Xia, S. Léonard, A. Deguet, L. Whitcomb, and P. Kazanzides, "Augmented reality environment with virtual fixtures for robotic telemanipulation in space," in *Proc. 2012 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2012, pp. 5059–5064.

[18] M. Kitagawa, D. Dokko, A. M. Okamura, and D. D. Yuh, "Effect of sensory substitution on suture-manipulation forces for robotic surgical systems," *J. Thoracic Cardiovascular Surg.*, vol. 129, pp. 151–158, 2005.

[19] A. K. Ng, L. K. Chan, and H. Y. Lau, "Depth perception in virtual environment: The effects of immersive system and freedom of movement," in *International Conference on Virtual, Augmented and Mixed Reality*. New York, NY, USA: Springer, 2016, pp. 173–183.

[20] A. Tang, C. Owen, F. Biocca, and W. Mou, "Comparative effectiveness of augmented reality in object assembly," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2003, pp. 73–80.

[21] E. Ruffaldi and F. Brizzi, "CoCo—A framework for multicore visuo-haptic in mixed reality," in *SALENTO AVR, 3rd International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. New York, NY, USA: Springer, 2016, pp. 339–357.

[22] K. Correll, N. Barendt, and M. Branicky, "Design considerations for software only implementations of the IEEE 1588 precision time protocol," in *Proc. Conf. IEEE*, 2005, vol. 1588, pp. 11–15.

[23] L. Peppoloni, F. Brizzi, E. Ruffaldi, and C. A. Avizzano, "Augmented reality-aided tele-presence system for robot manipulation in industrial manufacturing," in *Proc. 21st ACM Symp. Virtual Reality Softw. Technol.*, 2015, pp. 237–240.

[24] A. Graziano, P. Tripicchio, E. Ruffaldi, and C. A. Avizzano, "A haptic datasuit for controlling humanoid robots," in *Proc. 47th Int. Symp. Robot.*, 2016, pp. 1–8.

[25] L. Peppoloni, A. Filippeschi, E. Ruffaldi, and C. A. Avizzano, "A novel 7 degrees of freedom model for upper limb kinematic reconstruction based on wearable sensors," in *Proc. IEEE 2013 IEEE 11th Int. Symp. Intell. Syst. Informat.*, 2013, pp. 105–110.

[26] L. Shmuelof, J. W. Krakauer, and P. Mazzoni, "How is a motor skill learned? Change and invariance at the levels of task success and trajectory control," *J. Neurophysiol.*, vol. 108, no. 2, pp. 578–594, 2012.

[27] J. C. Gwilliam, M. Mahvash, B. Vagvolgyi, A. Vacharat, D. D. Yuh, and A. M. Okamura, "Effects of haptic and graphical force feedback on teleoperated palpation," in *Proc. 2009 IEEE Int. Conf. Robot. Autom.*, 2009, pp. 677–682.

Lorenzo Peppoloni (M'13) received the M.Sc. degree in automation engineering from the University of Pisa, Pisa, Italy, and the Ph.D. degree in perceptual robotics from the Scuola Superiore Sant'Anna, Pisa, Italy, in 2001 and 2016, respectively.

His research interests include biomechanics, sensor fusion, and state estimation.

Alessandro Graziano received the B.S degree from the University of Naples Federico II, Naples, Italy, in 2012 and the M.S. degree in robotic and automation engineering from the University of Pisa, Pisa, Italy, 2012 and 2015, respectively.

His research interests include haptics, telepresence, and teleoperation.

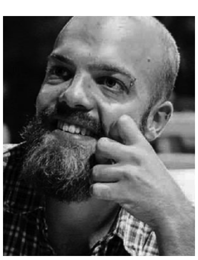
Erika Di Stefano received the M.Sc. degree in robotics and automation engineering from the University of Pisa, Pisa, Italy, in 2015.

Her research interests include computer vision and sensor fusion for robotics applications.

Carlo Alberto Avizzano (SM'17) received the M.Sc. degree in automation from the University of Pisa, Italy, in 1995 and Ph.D. degree in perceptual robotics from the Scuola Superiore Sant'Anna in 2000. He is the currently the Coordinator of PERCRO Laboratory, Pisa, Italy.

His research interests include intelligent robot controllers including autonomous robots as well as human interfaces. He has collaborated/coordinated several EU and National projects.

Mr. Avizzano is an Associated Editor for several international boards and journals.



Filippo Brizzi (S'14) received the M.Sc. degree in computer science jointly from the University of Pisa, Pisa, Italy, and Scuola Superiore Sant'Anna, Pisa, Italy, where he is currently working toward the Ph.D. degree in perceptual robotics at PERCRO Laboratory.

His research interests include augmented reality, haptics, and software architecture for robotics systems.

Emanuele Ruffaldi (M'07) received the M.Sc. degree in computer science from the University of Pisa, Pisa, Italy, and the Ph.D. degree in perceptual robotics from the Scuola Superiore Sant'Anna, Pisa, Italy, in 2002 and 2006, respectively.

His research interests include virtual environments for robotics, machine learning, and human–robot interaction.