Intuitive and Interactive Robotic Avatar System for Tele-Existence: TEAM SNU in the ANA Avatar XPRIZE Finals

Beomyeong Park[1] · Donghyeon Kim[2] · Daegyu Lim[2] · Suhan Park[2] · Junewhee Ahn[2] · Seungyeon Kim[3] · Jaeyong Shin[2] · Eunho Sung[2] · Jaehoon Sim[2] · Junhyung Kim[2] · Myeong-Ju Kim[2] · Junhyeok Cha[2] · Gyeongjae Park[2] · Hokyun Lee[2] · Seungbin You[2] · Keunwoo Jang[4] · Seung-Hun Kim[2,5] · Mathew Schwartz[6] · Jaeheung Park[7,8]

Abstract
Avatar robots enable the teleoperation and telepresence of an operator with a rich and meaningful sense of existence in another location. Robotic avatar systems rely on intuitive interactions to afford operators comfortable and accurate robot control to perform various tasks. The ability of operators to feel immersed within a robot has drawn interest in multiple research fields to explore the future capabilities of such systems. This paper presents a robotic avatar system based on a custom humanoid robot, TOCABI, with a mobile base. Its teleoperation system was developed in response to the ANA Avatar XPRIZE. Combining the life-size humanoid robot and the mobile base allows for improved mobility and dexterous manipulation. The robotic avatar system comprises the robot/base and an operator station that incorporates haptic feedback devices, trackers, a head-mounted display, gloves, and pedals. These devices connect the robot-environment interaction and operator-avatar robot experience through visual, auditory, tactile, haptic, and kinesthetic feedback. Combining the untethered battery-operated and Wi-Fi-enabled robot with these sensory experiences enables intuitive control through the operator's body movement. The performance of the robotic avatar system was evaluated through user studies and demonstrated in the ANA Avatar XPRIZE Finals, represented by Team SNU, where it completed 8 of the 10 missions, placing the team eighth among the 17 finalists.

Keywords Avatar system · Teleoperation · Telepresence · Humanoid robot · Human robot interaction · Human machine interface

✉ Jaeheung Park
park73@snu.ac.kr

Beomyeong Park
bpark@ihmc.org

Donghyeon Kim
kdh0429@snu.ac.kr

Daegyu Lim
dgyo3784@snu.ac.kr

Suhan Park
psh117@snu.ac.kr

Junewhee Ahn
june992@snu.ac.kr

Seungyeon Kim
sy07.kim@samsung.com

Jaeyong Shin
jasonshin0537@snu.ac.kr

Eunho Sung
eunho526@snu.ac.kr

Jaehoon Sim
simjeh@snu.ac.kr

Junhyung Kim
john3.16@snu.ac.kr

Myeong-Ju Kim
myeongju@snu.ac.kr

Junhyeok Cha
threeman1@snu.ac.kr

Gyeongjae Park
rudwo1301@snu.ac.kr

Hokyun Lee
hkleetony@snu.ac.kr

Seungbin You
ysb127@snu.ac.kr

Keunwoo Jang
jang90@kist.re.kr

Seung-Hun Kim
ksh1018@snu.ac.kr

1 Introduction

Robotic avatar systems facilitate secure and seamless operator control of a robot from a considerable distance while obtaining a lifelike perception of the robot's surrounding environment and status. Consequently, robots have the advantage of being able to replace humans in dangerous places: in space, underwater, etc. Furthermore, they are expected to provide everyday assistance and support to individuals with disabilities [1–3]. Within the field of avatar robot research, the effective creation of tele-existence − the experience of being in another location − is investigated through a combination of telepresence and the corresponding teleoperation [4]. In addition, the DARPA Robotics Challenge demonstrated the importance and feasibility of robots replacing humans in hazardous environments through remote teleoperation [5–7].

Teleoperation, which has been extensively studied in the field of robotics, finds application in various domains such as industrial, surgical, space, and underwater contexts [8–15]. The objective of teleoperation research is to develop intuitive interfaces for operators. The joystick, a device that operators can easily familiarize themselves with, has been used in teleoperation research for a long time [16, 17]. However, the joystick interface faces challenges in directly conveying factors such as contact between the robot and the surrounding environment or the reactive forces occurring when the robot manipulates objects during tasks. To address this limitation, haptic controllers with feedback capabilities have been adopted as an alternative interface [18, 19]. Haptic controllers enable force feedback, allowing the operator to perceive a

counterforce when the robot manipulates an object, thereby improving teleoperation performance.

Although joystick controllers or haptic controllers are intuitive interfaces, they have fewer degrees of freedom (DoF) than required to control humanoid robots. Therefore, to teleoperate humanoid robots with multiple DoFs, such as TOCABI [20], more intuitively, a different form of interface is desired. Current research explores how operators control robots by moving their entire bodies, a departure from earlier techniques that relied solely on hand-operated joysticks and haptic controllers. Motion capture methods involve attaching devices to the operator's body to track their movement and then employing the measured movements to teleoperate the robot. One method of measuring human motion involves attaching markers to the body and recording their positions through cameras [21–23]. Another approach is to attach multiple Inertial Measurement Unit (IMU) sensors to the operator's body to estimate the position and orientation of the operator's limbs [24–27]. These methods utilize the positions of markers attached to a person's body to calculate retargeted robot joint angles through inverse kinematics. As a result, drawbacks, such as tracking errors or pose discrepancies, may arise due to the different DoFs and differences in body lengths between the human and the robot. Furthermore, the system complexity increases as numerous sensors need to be placed on each limb of the human body [28].

Methods using an exoskeleton-type interface have been investigated to directly measure the angles of the operator's joints [29–31] and to teleoperate the robot without relying on inverse kinematics (IK). An advantage of the exoskeleton interface is the allowance for force feedback at the contact points between the operator's body and the exoskeleton [30, 32]. However, exoskeletons come with the drawback of increased inertia in the interface device due to the addition of actuators for force feedback.

In the recent gaming market, there have been significant advancements in virtual reality (VR) controller interfaces, allowing users to experience feedback through the vibration of the VR controller. In research, VR controllers have been used as teleoperation interfaces [33–36]. VR controllers offer significant convenience compared to motion capture or exoskeleton devices, as they only require the operator to grasp them, eliminating the time needed to equip motion capture interfaces. However, the information obtained through VR controllers is limited to the position and rotation of the operator's hands, leading to pose discrepancies between robots and human users. Commercial motion trackers,[1] in comparison to traditional motion capture interfaces, offer convenience in terms of wearability [37–42]. They allow for a relatively precise matching of the operator's posture and the robot's.

Mathew Schwartz
cadop@njit.deu

1   Institute for Human and Machine Cognition (IHMC), Pensacola, FL, USA

2   Department of Intelligence and Information, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea

3   Samsung Advanced Institute of Technology, Samsung Electronics, Suwon, Republic of Korea

4   Korea Institute of Science and Technology, Seoul, Republic of Korea

5   Intelligent Robotics Research Center, Korea Electronics Technology Institute, Gyeonggi, Republic of Korea

6   New Jersey Institute of Technology, New Jersey, USA

7   Department of Intelligence and Information, Graduate School of Convergence Science and Technology, ASRI, RICS, Seoul National University, Seoul, Republic of Korea

8   Advanced Institutes of Convergence Technology (AICT), Suwon, Republic of Korea

---

[1] VIVE tracker (3.0). 2021. https://www.vive.com/us/accessory/tracker3/.

However, the drawback is that these motion trackers cannot provide force feedback to the operator, and problems can occur when the motion tracking camera is accidentally blocked, preventing it from capturing the state of the operator [43].

Another commercial product, Head-Mounted Displays (HMD) for virtual reality (VR), have been researched to provide users with a realistic display for a more immersive sensation, allowing them to experience telepresence while remotely controlling the robot [44–48].

To accelerate the development of robotic avatars, the ANA Avatar XPRIZE competition was established [49]. Unlike other robot competitions that typically involve developers operating their robots to complete predefined missions [50], the ANA Avatar XPRIZE Finals and Semifinals took a different approach. In this competition, randomly assigned judges used and evaluated each team's systems. Within an hour, each team had to complete the setup of the avatar system in the operator room, where robots cannot be directly seen and where the judge controls the robot remotely. In addition, each team had to provide instructions to the judge on how to use the system. The evaluation of teleoperation technology in this competition was based on the successful completion of missions and the subjective scores given by two judges. Therefore, the system must provide an intuitive and user-friendly interface, even for non-experts. In the semifinals, most of the robots were powered through a tethered connection and supported by a crane while performing tasks. However, the finals required the robots to move around freely [51], meaning they needed to be untethered and equipped with mobile robotic systems capable of providing various sensory feedback.

Most finalists prepared robots that mimicked the operator's actions using a teleoperation interface with haptic feedback capability [52–55]. Furthermore, most robots incorporated wheels for efficient long-distance movement (which will be discussed in Sect. 6.4.) Our team, Team SNU, also used a mobile base to enhance mobility, meeting the Finals requirements while preventing the robot from falling over. In the semifinals, our system featured a tracking marker-based interface that mimicked the operator's actions, facilitating the robot's movement. However, it had limitations, such as the absence of haptic feedback for the operator [38]. We developed and integrated a haptic device into our Semifinals interface that used markers to address these limitations. The updated system provides haptic feedback directly to the operator's wrist. Additionally, we developed gloves to provide tactile and kinesthetic feedback, along with robot hands capable of manipulating the tools required for the finals.

This paper outlines the Robotic Avatar System developed by Team SNU for the ANA Avatar XPRIZE Finals. The remainder of this paper is organized as follows. Section 2 provides an overview of the robotic avatar System consist-

ing of an operator station and an Avatar robot. Sections 3 and 4 provide detailed explanations of the proposed teleoperation and telepresence system, respectively. In Sect. 5, we present our user studies for evaluating our avatar system's performance. Section 6 describes the missions and results of the ANA Avatar XPRIZE, offering insights gained from the competition, and Sect. 7 concludes the article.

2 Proposed Robotic Avatar System

In this section, the robotic avatar system of Team SNU is introduced. Figure 1 shows the structure of our robotic avatar system, and a concise summary of its specifications is presented in Table 1. Our robotic avatar system is described in terms of both hardware and software.

2.1 Hardware Structure of Robotic Avatar System

We made several modifications to our system between the competition's semifinals [38] and finals. These changes included the integration of robot hands and a mobile base for the Avatar robot, as well as the incorporation of haptic feedback devices and haptic gloves for the operator station, as shown in Fig. 1. *Operator station* refers to equipment setup for remotely controlling robots. The following sections will provide detailed explanations of the operator station and the Avatar robot.

2.1.1 Avatar Robot

Our Avatar robot is comprised of a humanoid robot featuring two camera sensors, one speaker and microphone, one wrist camera, two robot hands, and a mobile base (Fig. 1).

- **TOCABI**: The humanoid robot TOCABI (TOrque Controlled compliAnt BIped) was used for the Avatar robot [20]. Its height is 1.8 m, and its weight is 100 kg. TOCABI has 8 DoF in each arm, 6 DoF in each leg, 3 DoF in the waist, and 2 DoF in the neck. The payload of each arm is approximately 5 kg without the robot hands and approximately 3 kg with the robot hands. Various sensors were integrated to capture and transmit environmental information. Two head cameras, a speaker, and a microphone were equipped on the head to convey the environment information and facilitate communication between the operator and individuals near the robot [56]. The head cameras are See3CAM_24CUG_CHL of *e-con Systems*,[2] and for integrated speaker and micro-
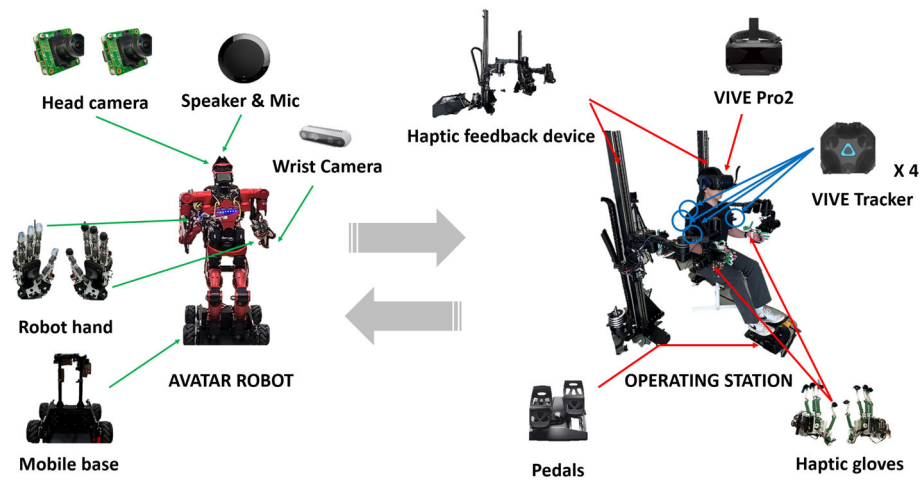

Fig. 1 Illustration of the robotic avatar system structure, which consists of an Avatar robot and an operator station. The Avatar robot consists of TOCABI, two head cameras, one speaker and microphone, one wrist camera, two robot hands, and one mobile base. The operator station includes a head-mounted display (HMD) (VIVE Pro2), a haptic feedback device, four trackers (VIVE Tracker3), a pair of haptic gloves, and foot pedals. The haptic feedback devices are connected to the haptic gloves worn by the operator. Two VIVE trackers are attached to each upper arm, one to the back of the operator and one to the chair. The operator wears the HMD with feet placed on the pedal. (Color figure online)

Table 1 Specifications of the robotic avatar system

Specifications	Values
DoF of TOCABI	33 DoF
Payload of TOCABI (each robot arm)	3 kg
DoF of the robot hand	8 DoF
Maximum force of the robot finger	4.92 N
Mobile base type	Omnidirectional
Velocity of the mobile base	0.75 m/s (moving forward) 0.5 m/s (moving laterally) 0.5 rad/s (rotating)
Maximum force of haptic feedback device	13.5 N (Software limit)
Maximum force of the finger of haptic gloves	1.4 N
FOV of the head cameras	104.6° (H) & 61.6° (V) 1920 × 1200 resolution
Video image transmission	Up to 114 Hz and 100 Mbps
Latency between operator and controller	10–20 ms
Feedback types	Audio, visual force, and tactile

phone functionality, a Jabra Speak 410[3] was employed. Each wrist of the robot featured a 6-axis force/torque sensor (F/T) from ATI,[4] used to measure the weight of objects held by the robot. An Intel RealSense Depth Camera D435[5] was attached to the left wrist. The wrist camera played a crucial role in identifying the surface characteristics of objects, particularly in conjunction with the force sensors on the robot hand fingers (Sect. 4.4). We used the wrist camera to capture images directly to train the algorithm for surface detection and classification.

- **Robot Hands**: Figure 2 shows our finger module with two DoFs and robot hands with eight DoFs [57]. The robot hand consisted of four modular fingers, as shown in Fig. 2c. Considering the velocity and torque of the joints, two types of DYNAMIXEL actuators were used for the fingers: XC330-M288-T[6] for the adduction/abduction joints (A/A) and XC330-M181-T[7] for the flexion/extension joints (F/E). For robust movement of the A/A joint, the XC330-M288-T actuator was cho-

[3] https://www.jabra.com/business/speakerphones/jabra-speak-series/jabra-speak-410.

[4] https://www.ati-ia.com/products/ft/sensors.aspx.

[5] https://www.intelrealsense.com/depth-camera-d435/.

[6] https://emanual.robotis.com/docs/en/dxl/x/xc330-m288/.

[7] https://emanual.robotis.com/docs/en/dxl/x/xc330-m181/.

sen. The F/E joint, responsible for gripping and pushing objects, needed to move faster than the A/A joint and, therefore, used the XC330-M188-T actuator. The three F/E joints are coupled by an internal 4-bar linkage. Compared to the three-fingered hand used in semi-finals [38, 58], the robot for the finals featured four fingers. The final competition required a robot hand capable of grasping various objects, such as a switch bar, a canister, a drill, and a stone (Sect. 6.1). When handling a drill, one finger must be dedicated to pressing its trigger button for activation, making it difficult for a robot with only two remaining fingers to grasp the drill securely. Therefore, the robot hand for the finals was developed with four-finger modules. The grasp taxonomy evaluation confirmed the robot hand's capability to perform 15 out of 16 motions [57]. To improve grasp stability, the fingertips of the right hand were covered with silicone, and 3D force sensors (Optoforce)[8] were attached to the fingertips of the left hand (Fig. 2c). Additionally, the palm of the robot hand was designed to mimic the shape of a human palm to increase the contact area during object grasping. The maximum force of a finger module was experimentally measured as 4.92 N.

- **Mobile Base**: The robot TOCABI was equipped with a mobile base with mecanum wheels and was used for navigation during the final missions. In the semifinals, bipedal walking was used. The mobile base included a chair that provided a seat for TOCABI (Fig. 3). The CTM300-7R mecanum wheels allowed the robot to move in any direction, while a mini PC (NUC11TNK) controlled its operations. The mobile base was powered by a lithium-ion battery pack 21700, 7S6P battery.[9] The mobile base's dimensions were 0.66 m in width (0.44 m without wheels), 0.68 m in length, and 0.58 m in height, including the wheels. The mobile base had a weight of approximately 50 kg. To improve the visibility of the robot's surroundings, lights were attached to both sides under the chair.
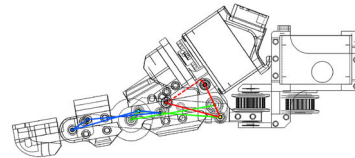
2.1.2 Operator Station

The operator station enables the operator to teleoperate with the robot and provides the sensory information input of the robot to the operator. Figure 1 shows the entire operator station and its components. We used three commercial products
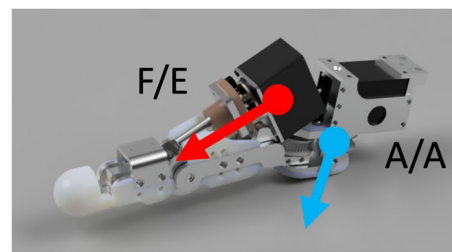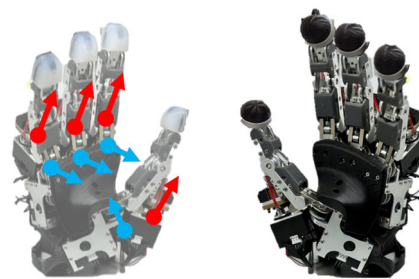
(a) The finger module

(b) Schematic of the finger module [56].

(c) Robot hands

Fig. 2 Illustration of the finger module and robot hand. **a** Each finger module has two actuators. Blue and red arrows represent the axes of the actuators. The red arrow denotes the flexion/extension (F/E) joint axis, while the blue arrow indicates the adduction/abduction (A/A) joint axis. **b** The schematic figure illustrates three links moved by the F/E actuator. The red link corresponds to the metacarpophalangeal link, the green link pertains to the proximal link, and the blue link relates to the distal link. **c** Each robot hand consists of four finger modules. Silicon fingertips are attached to the right hand. Force sensors are attached on the left hand. (Color figure online)

(VIVE Pro2, VIVE Tracker3, and rudder pedals) and two manufactured devices (haptic feedback device and haptic gloves).

- **VIVE Pro2**: VIVE Pro2 of *HTC* is used as the HMD for sending the visual and auditory information to the operator (Sect. 4.1). The HMD provides the operator with visual information for telepresence and shows the operator the interface needed to operate the Avatar robot. Furthermore, TOCABI's head moves in response to the operator's head movements sensed through the HMD.
- **VIVE Tracker3**: Four HTC VIVE trackers V3 are used to measure the upper body motion of the operator (Sect. 3.1). Two VIVE trackers are attached to each upper arm, one to the back of the operator, and the other to the
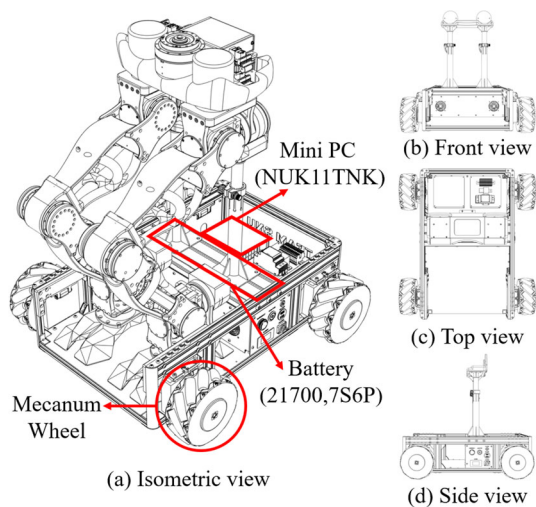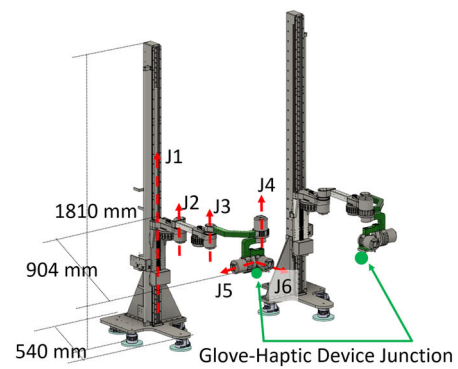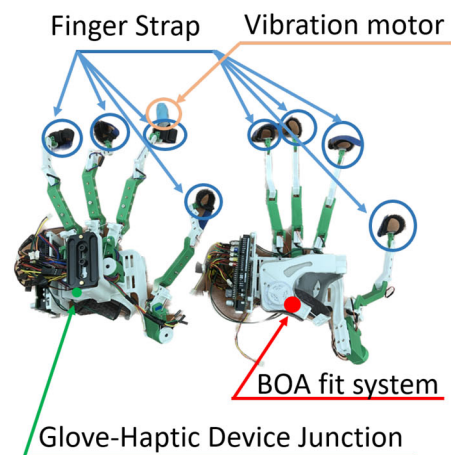
(a) Isometric view

(b) Front view

(c) Top view

(d) Side view

Fig. 3 Snapshots of the mobile base [59]. The mobile base includes the chair where TOCABI is sitting. **a** Isometric view of the mobile base when the TOCABI is seated. The mobile base includes four mecanum wheels, one Battery (21700, 7S6P), and a Mini PC (NUK11TNK). **b** The front view provides a detailed view of the robot's chair, comprising two bars and a section designed for the robot's hip. **c** The top view reveals the bottom box area of the mobile base, which serves as the placement space for TOCABI's feet. **d** The side view shows buttons for turning on the mobile base and its Mini PC and a display that indicates the base's battery status. (Color figure online)

(a) Haptic feedback devices.

(b) Haptic gloves.

Fig. 4 Snapshots of haptic feedback devices and gloves for teleoperation. The *Glove-Haptic Device Junction* connects the haptic feedback devices and the gloves. A finger strap is attached to the end of each finger. A vibration motor is on the index finger of the left hand. The BOA fit system is installed on each palm. (Color figure online)

back of the chair to serve as a reference frame(Fig. 1). The position and orientation of the trackers are measured optically through external base stations.

- **Haptic Feedback Devices**: Our developed haptic feedback devices can accurately measure the position and orientation of the operator's wrist while also providing force feedback to the operator (Sect. 4.2). The devices that provide haptic feedback are displayed in Fig. 4. They are 1.8 m tall and 0.5 m wide. Each device consists of six joints, with the first joint using a prismatic actuator and the remaining joints using revolute actuators. The workspace of the haptic feedback device is a cylinder with a 0.9 m radius centered on the J1 axis. The height of the cylinder is 1.8 m, ensuring that the haptic device's workspace adequately covers most of the human operator's workspace. A wrist connector between the haptic gloves and feedback device is next to the 6th axis.
- **Haptic Gloves**: The gloves are also developed to measure the movement of the operator's fingers (Sect. 3.2) and to deliver tactile and kinesthetic feedback to the operator (Sect. 4.3 and 4.4): tactile for delivering roughness of the object, and kinesthetic for delivering whether the robot grasps the object. The gloves can exert a maximum force of 1.4 N on the operator's fingers (Sect. 4.3). The strap of each finger is connected to the middle phalanx of the

operator's finger. For ease of use, the BOA fit system[10] is installed in the palm of the gloves. The vibration actuator of the left index finger is placed on the operator's fingertip and allows them to perceive the roughness of an object through vibration (Sect. 4.4). Each finger measures the joint position of the F/E and A/A joints of the operator.

- **Pedal**: T.Flight rudder pedals from *Thrustmaster*[11] are used as the controller of the mobile base (explained in Sect. 3.3). The operator can use the pedal with their feet while seated. The pedal commands the mobile base to *Drive*, *Rotate*, *Move Left*, *Move Right*, and *Reverse*. Switches attached to both sides of the pedals are used to

---

[10] BOA fit is a unique design that integrates a dial, lace, and guide configuration throughout the shoe, https://www.boafit.com/en-us.

[11] https://www.thrustmaster.com/en-us/products/t-flight-rudder-pedals/.

change the driving mode of the mobile base or control the VR interface.

2.2 Software Structure of the Avatar Robot System

Our software structure is illustrated in Fig. 5. The operator station is comprised of two computers, while the Avatar robot consists of four. Table 2 provides an overview of each computer's specifications. The operator station and Avatar robot are connected through a Wi-Fi network. The HMD data is transmitted via TCP/IP while the robot data is transmitted via Robot Operating System (ROS) over TCP/IP. The TOCABI PC runs on Ubuntu 20.04 [60] and the Operator PC uses Windows to operate Unity3D. To transfer data between Windows and Ubuntu systems through ROS messages, we used Win-Ros.[12]

The Operator PC is connected to various devices such as HMD, VIVE tracker, Haptic gloves, and pedal. In order to facilitate real-time voice communication with low latency, we use an open-source tool called Mumble.[13] Mumble is installed on both the Operator PC and the Processing PC to transmit sounds between the people around the Avatar robot and the operator. The head cameras on the Avatar robot capture the surrounding environment with a resolution of 1920 x 1200 and a field of view of 104.6° horizontally and 61.6° vertically. The video manager on the Processing PC transmits video to the Operator PC at a rate of up to 114 Hz and 100 Mbps. On the Operator PC, Unity3D in OpenVR receives the video data through TCP/IP and adjusts the 2D video image to fit the VR screen of the HMD. Each image from the head camera is projected to each eye of the operator through the HMD. The camera image from the mobile base is also transmitted to the HMD via TCP/IP.

The VIVE trackers and haptic feedback devices are used to teleoperate TOCABI's arms for manipulation. Real-time position and orientation information for the VIVE trackers is obtained through their open-source API at a rate of 90Hz. The operator's hand position and orientation are calculated using the forward kinematics of the haptic feedback device at 2000 Hz. To determine the relative position between both hands, the distance between two haptic feedback devices is measured. ROS topics for the VIVE trackers and the haptic feedback device are published to the motion retargeting algorithm on TOCABI PC at a rate of 100 Hz. The motion retargeting algorithm calculates the desired joint positions and velocities of the TOCABI's upper body with the transmitted data (Sect. 3.1). The TOCABI PC provides F/T sensor data to the Haptic PC via ROS for force feedback.

The haptic gloves measure the finger movement of the operator. Measured data is transmitted from the Haptic glove API on the Operator PC to the Hand Controller in the Processing PC via ROS. The Hand Controller maps the operator's hand motion to the robot hand using the transmitted glove data. The Hand Controller and Recognition Algorithm send feedback information to the glove via ROS. The current of TOCABI's finger motor, I_{finger} is measured and transmitted to the Haptic glove API. The Haptic glove API provides kinesthetic feedback to the operator, indicating whether TOCABI's hand has grasped an object or not. The opto-force sensors on the robot hand's fingertip measure the contact forces of the fingertips. The Recognition algorithm in Recognition PC distinguishes the surface of the stone under the palm of the robot's hand. With the measured force of the robot fingertip and the recognized object, the information of the object that the robot hand grasps or touches is delivered to the operator through the vibration motor of the gloves(Sects. 4.3 and 4.4).

The pedals are used to drive the mobile base. The output of the pedals is transmitted from the pedal API of the Operator PC to the Mobile Controller of the Mobile PC via ROS. The Mobile Controller calculates the velocity of the wheels using the transmitted pedal command (Sect. 3.3).

2.3 Untethered System

In order for our system to operate untethered, we have implemented batteries for the Avatar robot and wireless communications between the Avatar robot and the operator station. The two batteries, Tattu Plus 22,000 mAh 22.2V Lipo battery,[14] are carried on both sides of TOCABI's waist. Batteries supply the rated voltage to the robot PC and Elmo drivers through voltage conversion, with further details explained in [20]. The mobile base uses a separate PC and power supply from the robot and is powered through the attached battery of the mobile base.

For communication between the Operator station and the Avatar robot, a Netgear R7800 router was used.[15] During the ANA Avatar XPRIZE Finals, the XPRIZE network was provided at the venue and team garage.[16] Our operator station (Operator PC, and Haptic PC) is connected to the XPRIZE network via an Ethernet line. The Avatar robot (Processing PC, TOCABI PC, Recognition PC, and Mobile PC) is connected wirelessly via the router.

---

[12] Package for developing ROS on Windows: https://wiki.ros.org/win_ros.

[13] https://www.mumble.info/.

[14] https://genstattu.com/tattu-plus2-25c-22000mah-6s1p-xt90-smart-lipo-battery.html.

[15] https://www.netgear.com/kr/home/wifi/routers/r7800.

[16] The venue served as the location for the ANA Avatar XPRIZE Competition. Meanwhile, the team garage is where participating teams prepare for the competition.
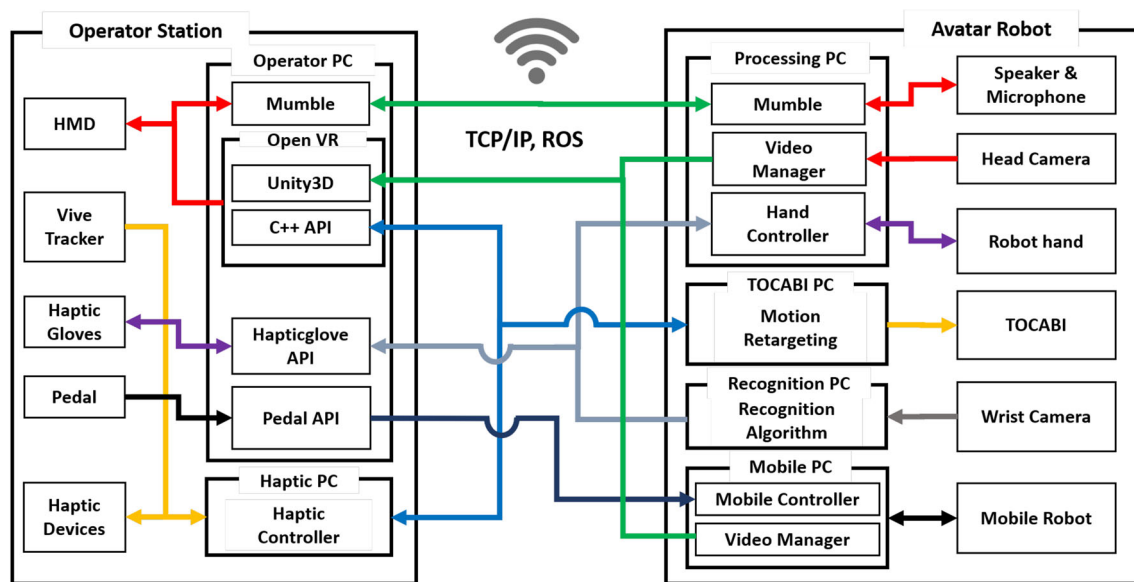
Fig. 5 Software system structure of our proposed system. The wireless connection between the Operator station and the Avatar robot uses TCP/IP and ROS. Each arrow indicates where the data comes from and where it goes to. The Operator station includes the Operator PC and Haptic PC. HMD, VIVE trackers, Haptic gloves, and pedals are connected to the Operator PC. The haptic feedback devices are connected to the Haptic PC. Three PCs are wirelessly connected to the Processing PC: TOCABI PC, Recognition PC, and Mobile PC of the Avatar robot. The head cameras, speaker & microphone, and robot hands are connected to the Processing PC. TOCABI is connected to the TOCABI PC. The wrist camera is connected to the Recognition PC. The mobile base is connected to a Mobile PC. (Color figure online)

Table 2 Specifications of the robotic avatar system computers

PC	CPU	RAM (GB)	Storage	OS
Operator station				
Operating PC	Intel Core i7-11800K	16	512 GB of NVMe SK hynix SSD	Windows 10
Haptic PC	Intel Core i7-10700	16	250 GB of NVMe SAMSUNG SSD	Ubuntu 20.04
Avatar Robot				
Processing PC	Intel Core i5-1135G7	64	512 GB of NVMe SSD	Ubuntu 20.04
TOCABI PC	Intel Core i7-10700	32	250 GB of NVMe SAMSUNG SSD	Ubuntu 20.04
Recognition PC	Intel Core i9-12900H	16	1 TB of SSD	Ubuntu 20.04
Mobile PC	i5-1135G7	64	250 GB of M.2. SSD	Ubuntu 20.04

3 Teleoperation

This section describes the three types of teleoperation: upper body operation, hand operation, and mobile base operation.

3.1 Upper Body Operation

The operator's upper body movements are tracked by haptic feedback devices and VIVE trackers to control TOCABI. We have combined haptic feedback devices and VIVE trackers to accurately measure the position and orientation of the operator's hand while enabling the Avatar robot to mimic the operator's upper body movements simultaneously. Additionally, force feedback can be delivered to the operator by the haptic feedback devices.

In Fig. 6, the coordinates of the operator delivered to the Avatar robot using a haptic device and VIVE trackers are shown. Figure 6a shows how the operator uses the haptic feedback devices, VIVE trackers, haptic gloves, HMD, and pedal. The haptic feedback devices measure the position and orientation of the hand of the operator, $\mathbf{p}_{ha}^{o}$ and $\mathbf{R}_{ha}^{o}$. The VIVE trackers and the HMD measure the position and orientation of the chest, upper arm, shoulder, and head. The measured coordinates of the operator are mapped into the coordinates of the Avatar robot according to the diagram in Fig. 7. The pose required for *Pose Calibration* in Fig. 7 involves attaching both arms to the body and forming an 'L'

(a)

(b)

Fig. 6 Description of coordinates of the operator and the robot. Each arrow indicates which part of the operator is mapped into the Avatar robot. The yellow circles indicate the origin of each coordinate. The red, green, and blue arrows are the orientation axis in x, y, and z, respectively. The wrist and upper arm use three orientations in x, y, and z axes. The upper body uses the orientation in x and y axis, while the head uses the orientation in y and z axis. **a** Illustration of the coordinates of the operator. The operator positions their arms in an 'L' shape for pose calibration. **b** Illustration of the coordinates mapped to the Avatar robot. (Color figure online)

shape with the arms. The method used to map the orientation of the shoulder, upper arms, chest, and head was introduced in our previous research [37]. From this *Pose Calibration*, the initial position of the operator's hands, $\overline{\mathbf{p}}^o_{ha,i}$, are obtained. The next step is to calculate the desired velocity of the robot hand. This is done using the following formula:

$$\mathbf{p}^r_{ha,d} = \overline{\mathbf{p}}^r_{ha,i} + a(\mathbf{p}^o_{ha} - \overline{\mathbf{p}}^o_{ha}), \qquad (1)$$

$$\mathbf{R}^r_{ha,d} = \mathbf{R}^o_{ha}, \qquad (2)$$

$$\dot{\mathbf{p}}^r_{ha,d} = \mathbf{K}(\mathbf{p}^r_{ha,d} - \mathbf{p}^r_{ha}), \qquad (3)$$

where $\mathbf{p}^r_{ha,d}$, and $\dot{\mathbf{p}}^r_{ha,d}$ are the desired position and velocity of the robot hand. $\overline{\mathbf{p}}^r_{ha,i}$ is the initial position of the robot corresponding to the 'L' pose. $a \in [1.0, 1.3]$ is the scaling
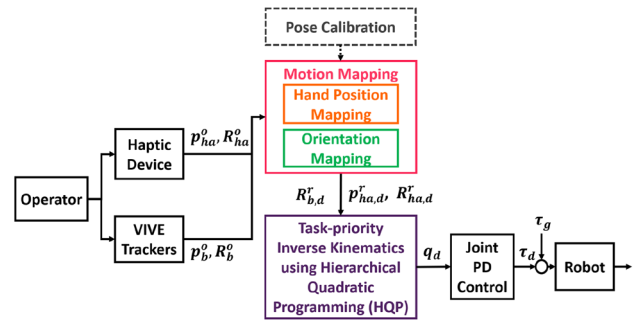

Fig. 7 Diagram of motion mapping. $\mathbf{p}^o_{ha}$ and $\mathbf{R}^o_{ha}$ are the position and orientation of the hand of the operator. $\mathbf{p}^o_b$ and $\mathbf{R}^o_b$ are the position and orientation of the body of the operator. The body includes upper body, upper arm, shoulder, and head, $b \in \{upperbody, upperarm, shoulder, head\}$. $\mathbf{p}^r_{ha,d}$, $\mathbf{R}^r_{ha,d}$ and $\mathbf{R}^r_{b,d}$ are the desired position and orientation of the robot. $\mathbf{q}_d$ is the desired joint angle. τ_d and τ_g are desired torque and gravity torque, respectively. (Color figure online)

factor that represents how much the robot's hand moves in proportion to the distance moved by the operator's hand [37]. When $a = 1.0$, the robot hand moves the same distance as the movement of an operator's hand. Additionally, when $a = 1.3$, the robot hand moves 1.3 times the distance of the movement of an operator's hand. This is based on the ratio of the operator's arm length to the robot's arm length. K is the feedback gain for tracking the desired position of the robot's hand.

The angular velocities of TOCABI's joints are calculated using hierarchical quadratic programming (HQP)-based inverse kinematics [37]. Table 3 displays the task priorities related to Fig. 7. The top priority is to ensure safety by restricting joint angle, joint velocity, and hand velocity. The second priority is the orientation of the chest, whereas the third priorities are the position and orientation of the hand, and the orientation of the head. The head orientation is used to control the visual feedback in the HMD. The orientation control of the upper arm and shoulders has the lowest priority, which aims to make the robot's pose similar to that of the operator. The optimal joint velocities of the upper body are computed by solving the HQP problem in (4) while adhering to the task priorities in Table 3,

$$
\begin{aligned}
\min_{\dot{q}} \quad & \rho_p \left\| \mathbf{J}_p \dot{\mathbf{q}} - \dot{\mathbf{x}}_{p,d} \right\|^2 + \|\dot{\mathbf{q}}\|^2_A \\
\text{s.t} \quad & \underline{\dot{\mathbf{x}}}^r_{ha} \leq \mathbf{J}_{ha} \dot{\mathbf{q}} \leq \overline{\dot{\mathbf{x}}}^r_{ha} \\
& \mathbf{K}_q(\underline{\mathbf{q}} - \mathbf{q}_{k-1}) \leq \dot{\mathbf{q}} \leq \mathbf{K}_q(\overline{\mathbf{q}} - \mathbf{q}_{k-1}) \\
& \underline{\dot{\mathbf{q}}} \leq \dot{\mathbf{q}} \leq \overline{\dot{\mathbf{q}}} \\
& \mathbf{J}_n \dot{\mathbf{q}} = \mathbf{J}_n \dot{\mathbf{q}}^*_{p-1}, \forall n \in 2, \ldots, p-1, (p \geq 3),
\end{aligned}
\qquad (4)
$$

where p is the pth priority task in Table 3. $\dot{\mathbf{x}}_{p,d}$ is the desired velocity of the pth priority, and $\mathbf{J}_p$ is the Jacobian matrix of the pth priority. ρ_p is the weighting value for the control

Table 3 Priority of the tasks

Priority	Task	DoF
1 (Highest)	Joint limit, Joint velocity limit, Hand velocity limit	$23 + 12$
2	Chest orientation	3
3	Hand position, Hand orientation, Head orientation	$12 + 2$
4 (Lowest)	Upper arm and shoulder orientation	4

error and should be much larger than 1 to control the desired motion accurately. The first term of the cost function in Eq. (4) is to minimize the velocity error of the pth task, $\|\mathbf{J}_p\dot{\mathbf{q}} - \dot{\mathbf{x}}_{p,d}\|^2$. $\|\dot{\mathbf{q}}\|_A^2$ is the regularization term that is being weighted by the inertia matrix of the robot, $\mathbf{A}$, which minimizes the kinematic energy of the robot. $\mathbf{J}_{ha}$ is the Jacobian matrix of the hand, $\underline{\dot{\mathbf{x}}}_{ha}^r$ and $\overline{\dot{\mathbf{x}}}_{ha}^r$ are the lower and upper limits of the velocity of the robot hand, respectively. $\underline{\mathbf{q}}$ and $\overline{\mathbf{q}}$ are the lower and upper limits of the upper body joint angle, while $\underline{\dot{\mathbf{q}}}$ and $\overline{\dot{\mathbf{q}}}$ are the lower and upper limits of the upper body joint velocity, respectively. $\mathbf{J}_n$ is the Jacobian matrix of higher priority than p. $\dot{\mathbf{q}}_{p-1}^*$ is the optimal value obtained from the previous hierarchy. The optimized $\dot{\mathbf{q}}^*$ is then integrated into the desired joint position, $\mathbf{q}_d$. To avoid self-collision, the method introduced in [61] was applied to TOCABI. If a self-collision is detected, TOCABI will halt its motion and notify the operator, allowing the operator to move away from the self-collision situation before resuming the robot's operation. The desired torque, $\boldsymbol{\tau}_d$, is calculated using the proportional-derivative (PD) control method in the joint space with gravity torque compensation. The latency between the operator and the controller is around 10 to 20 ms, and the operator is barely aware of the delay.

3.2 Hand Operation

Exoskeleton-type gloves are equipped on the operator's hands to control the robot hand. The mapping of the operator's pose to the robot hand is explained in Fig. 8. The glove shown in Fig. 4 has four linkage-type fingers and can measure the operator's finger joint angles of F/E and A/A. The maximum and minimum F/E angles of the operator's finger are measured using five mapping gestures [62]. Theses values are then linearly mapped to the maximum and minimum F/E angles of the robot hands to enable the operator's finger motions to control the robot hand. This method is an extension of a previous study that mapped human actions onto a robot hand with three fingers and has now been expanded to four fingers [62]. The A/A motion of the operator's finger is similarly mapped to the robot hand. The A/A movement of the robotic hand enables it to grasp various shapes of objects stably.

Posture		Measured finger joint value
Plate		Maximum Extension of four fingers Minimum A / A of Thumb and Index
Pinch		A / A Pinch of Thumb FE Pinch of four fingers
Finger Flexion		Maximum Flexion of three fingers (except Thumb)
Thumb Flexion		Maximum Flexion of Thumb
Sphere		Maximum A / A of Thumb and Index

Fig. 8 Five gestures for mapping operator's hand pose to the robot hand. Each figure illustrates the measured finger joints for each posture. (Color figure online)

3.3 Mobile Operation

The mobile base of the robot has four mecanum wheels, which enable it to move in any direction. Out of the available choices of joysticks, 3D Rudder pedals, and flight pedals, we selected the flight pedal as the interface for controlling the mobile base. We excluded the joystick because the operator's hands needed to control the robot's arm remotely. Additionally, we ruled out the 3D Rudder pedal because the operator would have to continuously pay attention to keep it in a neutral position when not actively moving the mobile base. The procedure for mobile base teleoperation is illustrated in Fig. 9.

The buttons labeled as ① and ② can only be pressed in one direction, requiring a separate reverse button to switch
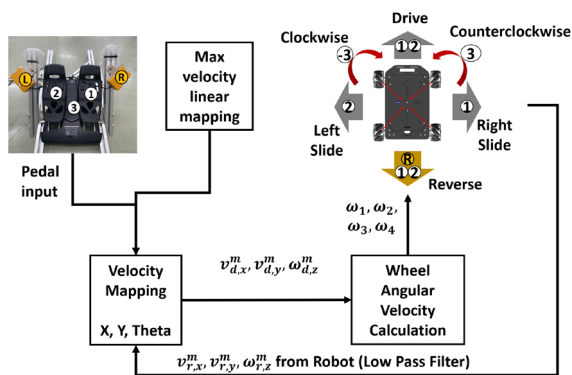
Fig. 9 Illustration of the mobile base operation procedure: The button Ⓛ changes the HMD menu. The button Ⓡ changes the three modes of the mobile base: Drive, Reverse, and Parking. The buttons ① and ② operate the mobile base. When ① or ② button is pushed, the mobile base moves in the respective direction: left for the ① button and right for the ② button. Pressing both the buttons of the ① and ② activates the mobile base in drive or reverse mode. If the pedal (③) is rotated, the mobile base also rotates accordingly: clockwise ③ or counterclockwise ③. (Color figure online)

to the reverse mode. When in Parking mode, the mobile base will not respond to pedals commands. Upon generating a pedal command, the desired velocities ($v_{d,x}^m$, $v_{d,y}^m$, $\omega_{d,z}^m$) are mapped using the pre-defined maximum velocity. Here, the subscript m stands for the mobile base. For the ANA Avatar XPRIZE Finals, we set a constant maximum velocity of 0.75 m/s for Drive mode, 0.5 m/s for side movement, and 0.5 rad/s for rotation. The pedals input value ranges from 0 to 1, depending on the degree to which the pedal is pressed, and is then used to determine the desired velocity by considering the maximum velocity value. The desired angular velocities of each wheel (ω_1, ω_2, ω_3, and ω_4) are calculated using the desired velocity of the mobile base and the kinematics according to the equation below, as described in [63].

$$\begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \end{bmatrix} = \frac{1}{R} \begin{bmatrix} 1 & 1 & -(l_1 + l_2) \\ 1 & -1 & l_1 + l_2 \\ 1 & -1 & -(l_1 + l_2) \\ 1 & 1 & l_1 + l_2 \end{bmatrix} \cdot \begin{bmatrix} v_{d,x}^m \\ v_{d,y}^m \\ w_{d,z}^m \end{bmatrix}, \qquad (5)$$

where R is the radius of the wheel, l_1 is the width of the mobile base (the distance from the center to the center of the wheels), and l_2 is the length of the mobile base (the distance from the center to the center of the wheels).

4 Telepresence

Telepresence technology enhances the operator's perception, creating the feeling as if they are physically present at the location of the Avatar robot. Our robotic avatar system caters to three of the five senses: vision, hearing, and touch. Addi-
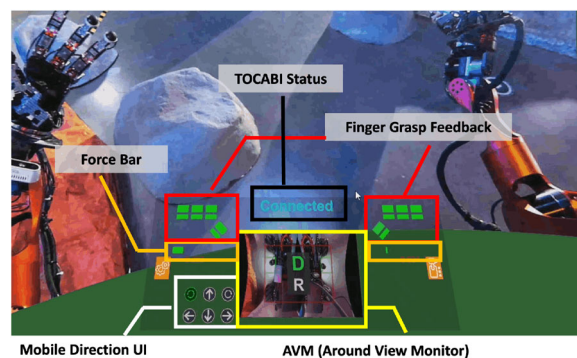

Fig. 10 Snapshot of the HMD view: the surroundings and UI are presented to the operator. Each rectangle signifies the type of information being transmitted to the operator. (Color figure online)

tionally, the operator can sense the robot's interactions with the environment through force feedback. In this section, we will elaborate on the HMD that provides vision and hearing in TOCABI, as well as the Around View Monitor (AVM) that enables vision around the mobile base. We will also detail the force, tactile, and kinesthetic feedback mechanisms that convey to the operator the robot's interactions with objects or the environment.

4.1 Head Mounted Display

4.1.1 Visual Feedback

Our robotic avatar system provides the operator with visual and auditory information through an HMD for telepresence [47, 64]. To achieve this, we utilize the HTC VIVE Pro 2 HMD, which offers a resolution of 2448 × 2448 pixels per eye. We transmit the video using the *TurboJPEG* Codec after capturing the image through a USB camera on the robot with OpenCV. The image is then encoded encoding using Python's *TurboJPEG* and then sent via TCP. In Unity, the image is received through TCP using the *TurboJpegWrapper*[17], decoded, applied to a Unity texture, and then displayed on the operator's HMD with a latency of 100 ms. The HMD device has a microphone and speaker which allows the operator to hear and communicate with individuals nearby TOCABI. The visual image captured by TOCABI's head camera is transmitted to the HMD.

4.1.2 User Interface

The HMD not only displays the robot's surroundings but also presents a user interface (UI) to assist the operator in teleoperating the robot. The operator can perceive the

---

[17] Using GitHub - Sklinay/AS.TurboJpegWrapperForUnity: Libjpeg-Turbo wrapper for.Net - Working fine with Unity and cross-platform, https://github.com/Sklinay/AS.TurboJpegWrapperForUnity.
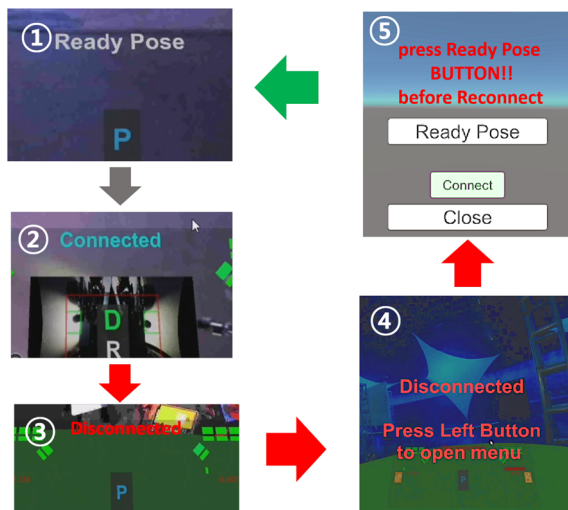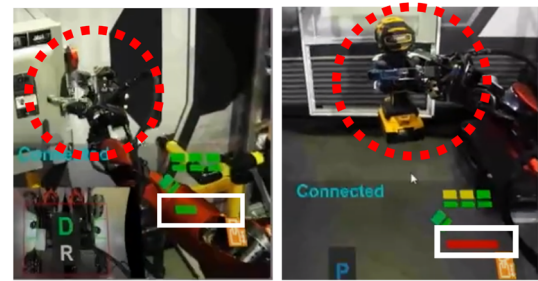
(a) Activating the switch. (b) Lifting the drill.

Fig. 12 Snapshots of force feedback through the HMD. Each circle indicates the object grasped by the robot hand, while the color and length of each rectangle indicate the object's weight. **a** The snapshot depicts the moment when TOCABI activates the switch. As shown in the white rectangle, the force bar is short and green, signaling that the force feedback from the switch is light. **b** The snapshot captures TOCABI lifting the drill. As shown in the white rectangle, the force bar is long and red, indicating the substantial weight of the drill. (Color figure online)

4.2 Force Feedback

The proposed system has two ways to provide force feedback: visual feedback and haptic feedback. When lifting objects, the F/T sensor on the wrist detects changes in force and torque induced by the object. The weight of the object is then calculated using the wrist orientation and the F/T sensor values. The weight information is then displayed on the force bar on the HMD. The changes in the force bar on the HMD corresponding to different objects are illustrated in Fig. 12. For example, when TOCABI lifts a light switch, the force bar appears as a short green bar, as shown in Fig. 12a. On the other hand, when lifting a heavy-weight drill, the force bar turns red and increases in length as depicted in Fig. 12b.

The haptic feedback devices allow the operator to feel the force feedback physically. The acquisition and conveyance of force feedback to the operator are depicted in Fig. 13. The forces exerted on the robot's hand, denoted as F_{sensor}, are measured with the F/T sensor placed on the wrist. F_{sensor} reflects only the weight of the object by removing the contribution of the robot hand, and it is represented in the robot base frame. When TOCABI is not holding anything, gravity, and friction compensation are added to the haptic feedback device as follows:

$$\tau_d^{haptic} = \tau_{gravity}^{haptic} + \tau_{friction}^{haptic}, \tag{6}$$

where τ_d^{haptic} is the input torque for the haptic feedback device, $\tau_{gravity}^{haptic}$ and $\tau_{friction}^{haptic}$ are the gravity and friction torques for the haptic feedback device, respectively. $\tau_{gravity}^{haptic}$ is determined through computations based on the CAD model of the haptic feedback device. The $\tau_{friction}^{haptic}$ was calculated
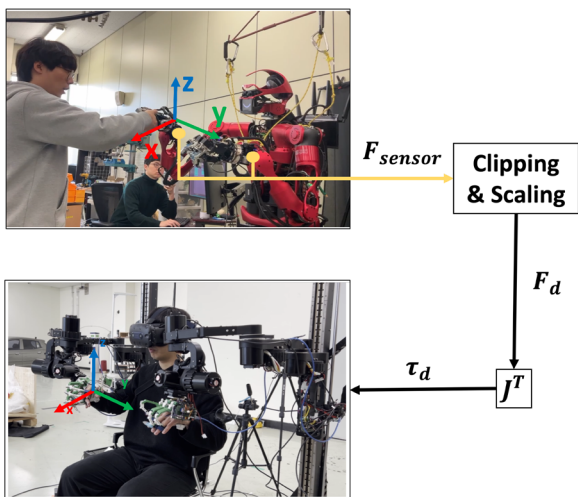

Fig. 11 Display of the UI according to the robot's status. The stages of the robot connection process are marked with ①, ②, and ③. The corresponding notices for each state are as follows: *Ready Pose* in ①, *Connected* in ②, and *Disconnected* in ③. ④ and ⑤ represent the steps when a disconnection occurs. In ④, the notice is *Disconnected Press Left button to open menu*, and in ⑤, it is *Press Ready Pose BUTTON!! before Reconnect*. (Color figure online)

scene viewed by the robot through the HMD, as depicted in Fig. 10. The lower section of Fig. 10 explains the information conveyed to the operator, which includes the Mobile Direction UI and AVM for mobility, Finger Grasp Feedback for kinesthetic feedback assistance, and the Force Bar for force feedback. The UI provides information on the connection status between the operator and the Avatar robot, finger grasp feedback, and force feedback. The AVM illustrates the surroundings of the mobile base along with the current driving mode, which could be D (Drive), P (Parking), and R (Reverse). The Mobile Direction UI indicates the direction in which the mobile base is moving.

The UI provides the operator with real-time information about the robot's status. As shown in Fig. 11, the robot can be in one of three states: *Ready Pose*, *Connected*, or *Disconnected*. The *Ready Pose* represents the initial pose of TOCABI, which the operator needs to replicate before connecting to the robot. The robot's movement is only enabled in the *Connected* state, and transitioning from *Ready Pose* to *Connected* facilitates this motion.

In case of an emergency or singularity occurrence, the robot system automatically switches to the *"Disconnected"* mode while updating the information on the HMD screen. After that, the operator can reset the robot's state to the *Ready Pose*. Once this reset has been performed, the robot becomes operational again and transitions back to the *Connected* state. This mechanism ensures safe control of the robot and prevents potential damage during its operation.

Fig. 13 An illustration of force feedback from the robot to the haptic feedback device of the operator. Yellow circles represent F/T sensors attached to the wrists. The force measured by the F/T sensor is calculated as the desired torque through clipping and mapping using the transpose of the Jacobian matrix. Red, green, and blue arrows indicate the force feedback's *x, y,* and *z* directions, respectively. (Color figure online)
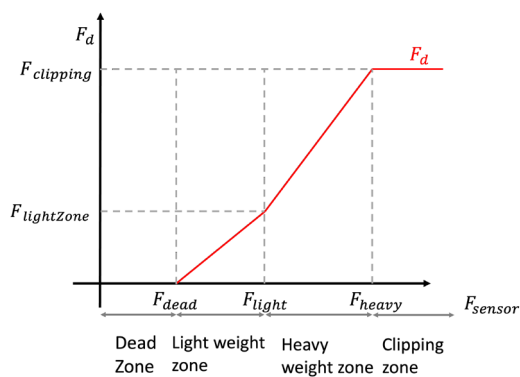

Fig. 14 Desired force for clipping and scaling. F_{sensor} represents the force measured by the F/T sensor attached to the wrist, oriented in the gravity direction. F_d represents the desired force transmitted to the operator as force feedback. F_{dead} serves as a threshold for the dead zone, allowing the neglect of values attributed to uncertainties. In the light-weight zone, $F_d = F_{sensor} - F_{dead}$. In the heavy-weight zone, $F_d = F_{light} + K_{scaling} \times (F_{sensor} - F_{light})$, where $K_{scaling}$ is the coefficient force scaling. If F_d surpasses $F_{clipping}$, then $F_d = F_{clipping}$. (Color figure online)

by adjusting the coefficients of static friction and viscous friction.

TOCABI measures the force, F_{sensor}, exerted when lifting an object, and this force is scaled by a $K_{scaling}$. The scaling factor, $K_{scaling}$, is multiplied to the weight of the object, and the resulting value, F_{clip}, is clipped, ensuring that it does not exceed a specific value as shown in Fig. 14. According to competition regulations, we needed to differentiate between objects weighing a maximum of 32 Oz (around 900 g), and those weighing less (around 300 g). Our system uses a set
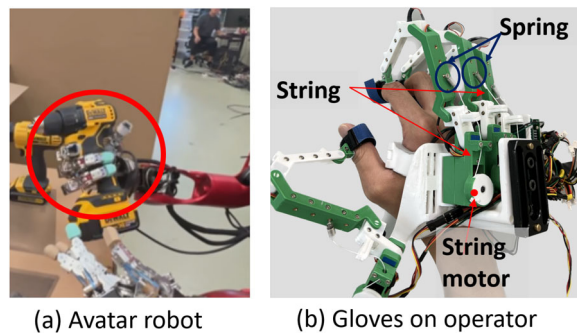

(a) Avatar robot   (b) Gloves on operator

Fig. 15 Illustration of the kinesthetic feedback system. **a** Snapshot depicting the moment when the Avatar robot grasps the object. **b** Realization of the kinesthetic feedback component of the haptic gloves. Each finger of the gloves is equipped with a spring and a string. The string is pulled and released by the string motor, while the spring ensures that the finger returns to its original position when the string motor is not in operation. When the robot hand grasps the object tightly, the current of the finger motor surpasses the threshold. The string motor pulls the string, resulting in the movement of the operator's finger. (Color figure online)

of values called F_{dead}, F_{light}, and F_{heavy} as 2.0 N, 2.5 N, and 5.25 N, respectively. When the measured force (F_{sensor}) falls between 2.5 N and 5.25 N, the difference between the values of F_{sensor} and F_{light} is multiplied by the scaling factor ($K_{scaling}$) to calculate F_d. During the ANA Avatar XPRIZE Finals, we set the value of $K_{scaling}$ as 4. The resulting scaled force, F_d, is then used in the (6).

$$\tau_d^{haptic} = \tau_{gravity}^{haptic} + \tau_{friction}^{haptic} + J_{haptic}^T F_d, \qquad (7)$$

where J_{haptic}^T is the transpose of the Jacobian matrix of the haptic feedback device, and F_d is the scaled reaction force by the object weight.

4.3 Kinesthetic Feedback

The gloves provide kinesthetic feedback to the operator, indicating whether or not the robot's hands have successfully grasped the objects. Figure 15 demonstrates how the kinesthetic feedback is transmitted to the glove. When the robot hand fully grasps the object, the fingers cannot bend, causing their current values to increase and surpass the threshold. This results in the string being pulled by the servo motor (HITech HS-5070MH servo motor[18]) and the spring (MIS-UMI AUA5-15 spring[19]) generating a force that is applied to the operator's finger, trying to extend it. The maximum force of kinesthetic feedback is 3.24 N. The servo motor has

[18] https://hitecrcd.com/products/servos/digital/micro-mini-wing/hs-5070mh/product.

[19] https://us.misumi-ec.com/vona2/detail/110300266030/?HissuCode=AUA5-15.

Mini Vibration Motor   Intel RealSense D435   3-axis Force sensor
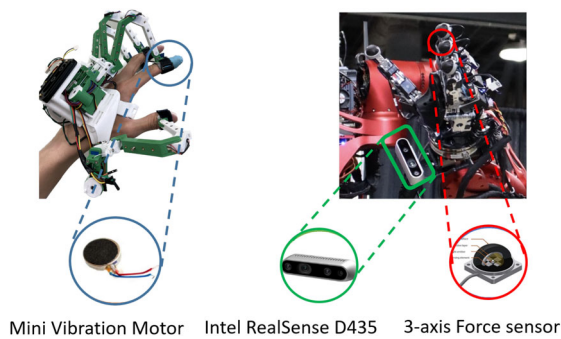
Fig. 16 Illustration of the tactile feedback system. The blue circle in the left figure shows the vibration motor attached to the left index finger of the glove. The green circle and box highlight the Intel RealSense D435 camera affixed to the wrist. The red circle in the right figure indicates the 3-axis force sensor manufactured by Optoforce. These sensors measure the force when the robot hand's fingers make contact with an object. The RealSense camera is used to recognize the surfaces of stones. The vibration motor imparts varying sensations to the user, contingent upon the perceived roughness of the contacted or perceived stones. (Color figure online)

Table 4 Classification performance of model (%)

| Metric	Value
Precision	100.0
Recall	98.7
mAP_{50}	99.5
$mAP_{50:95}$	82.1

a maximum torque of 3.8 kg · cm and a pulley with a 1 cm diameter is used, which generates a maximum force of 38 N. However, when used in the competition, the spring's stiffness parameter is reduced to generate a maximum force of 1.4 N. The spring is limited to prevent it from stretching beyond this maximum force. Even though the maximum load of the spring is 3.24 N, the kinesthetic force is limited to 3.24 N. The operator can perceive the sensation of his finger being pulled, providing a tangible indication that the robot hand has successfully grasped the object.

4.4 Tactile Feedback

During the Avatar XPRIZE Finals, participating teams were challenged to create methods that would provide physical haptic feedback. Our team developed a robotic avatar system that conveys the roughness of surfaces to the operator through the vibration motor on the index finger of the left glove. This approach was crucial in fulfilling one of the competition missions, which required detecting a rough-surfaced stone hidden behind a curtain and out of view (explained in Sect. 6.1). Measuring the roughness of an object using a force sensor attached to a robot finger requires delicate manipulation of the robot finger while maintaining contact between the object and the sensor [65, 66]. This becomes even more
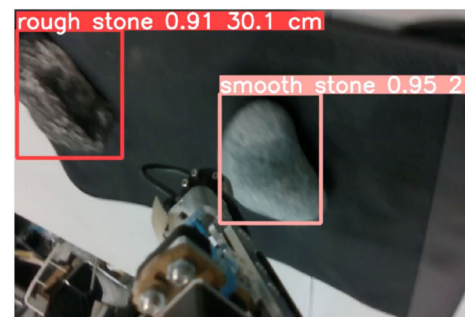
(a)
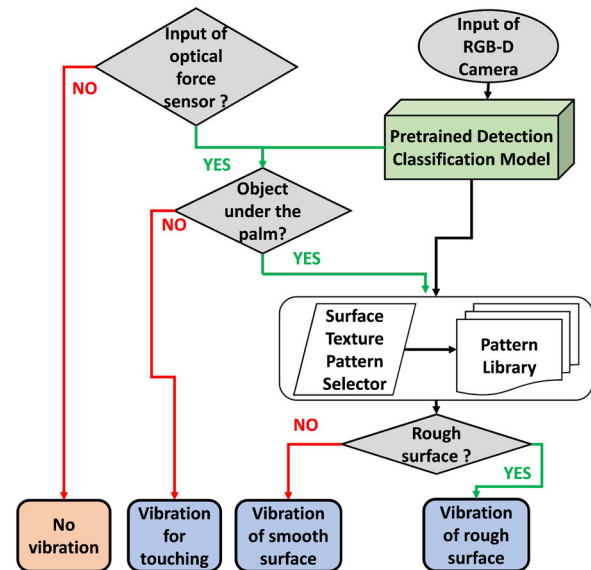
(b)

Fig. 17 Explanation of the recognition result and the strategy of identifying the stone surface. **a** Snapshot of the recognition of the stone surface. Each stone label denotes its surface characteristics, recognition confidence, and the distance from the robot's wrist. It's important to note that this result snapshot is not delivered to the operator. **b** Stone Surface Identification Strategy: Our approach utilizes two sensor inputs, the optical force sensor on the fingertip and the RGB-D camera on the wrist, to identify the stone surfaces. The strategy yields four types of vibration responses. No vibration occurs when there is no input from the optical force sensor. If there is input from the optical force sensor, the YOLO v5 algorithm is employed to detect whether the object is a stone and to distinguish its roughness. In cases where the object is not identified as a stone, vibration is triggered to simulate touching. On the contrary, if a stone is detected, the vibration is determined based on the perceived roughness within the defined parameters (white box). (Color figure online)

challenging when the object is out of sight. To address this, our system uses a more intuitive approach by employing an Intel RealSense camera mounted on the left wrist (as shown in Fig. 16). The YOLO v5 algorithm and roboflow are used to recognize the surface of the stone [67, 68]. Moreover, since the items to be used in the final competition were disclosed by the XPRIZE competition organizers, we were able to pre-

pare datasets in advance by acquiring stones of the same type. The performance of the trained model [69] is presented in Table 4. The mean average precision (mAP) [70] at an intersection over union (IoU) threshold of 0.5 and the mean AP at IoU thresholds ranging from 0.5 to 0.95 are expressed as mAP_{50} and $mAP_{50:95}$, respectively. The recognition algorithm successfully distinguishes between smooth and rough stones, as shown in Fig. 17a.

The process of delivering tactile feedback about the roughness of a stone to the operator is illustrated in Fig. 17b and described in [69]. The robot hand's contact with the object is initially determined based on input values from the Opto-force sensor. If there is no input value from the sensor, the operator does not receive any feedback. The wrist-mounted camera, which faces in the same direction as the palm, helps detect the presence of an object beneath the robot's hand. If the robot hand is in contact but there is no object beneath the palm, the operator receives feedback about the contact. When a stone is beneath the palm, the roughness recognition algorithm comes into play to determine the surface roughness of the object. The dataset used for the algorithm training was obtained from various environments, including dark or bright settings, with certain parts of the rocks obscured. Once the trained recognition system distinguishes the roughness of the stone, corresponding vibrations are transmitted to the operator. A low-frequency vibration is triggered when a smooth surface is detected, while a high-frequency vibration is transmitted upon detecting a rough surface. However, the system can only detect two kinds of stones. Nevertheless, it provides the operator an intuitive perception of the stone's roughness. While we successfully validated our approach in a test bed, we regret that we were unable to test it during the competition.

5 System Evaluation

We evaluated our system through a user study when users encountered tasks for the first time. Participants were given instructions solely through verbal explanations before the evaluation test. The experiment was carried out by ten participants in a single trial. All participants were members of the research team and had a basic understanding of our system. Participants were divided into three groups based on their experience with the system: Beginner (up to 90 min), Intermediate (90–180 min), and Expert (over 180 min). There were 3 Beginners, 4 Intermediates, and 3 Experts.

5.1 Benchmark Tasks

To evaluate the intuitiveness of our avatar system's telepresence and teleoperation, five benchmark tasks were designed based on the ANA Avatar XPRIZE Competition.

- Manipulation and grasping task (Fig. 18a): The purpose of this task was to assess the telemanipulation capabilities. The task involved using TOCABI's arm to reach a specific location and grasp different objects, such as a joystick, water bottle, and stone, placed on a table. The aim was to evaluate the arm and hand operations involved in the task. During the task, the Avatar robot picked up each object sequentially and handed them to a person standing in front of the table. The total time taken for the task was measured from the moment the participant began moving until the task was completed.

- Identifying weight (Fig. 18b): This task was created to test the participant's ability to distinguish weight using force feedback. Three water bottles, weighing 12.7 N, 4.9 N, and 0.24 N respectively, were placed on a table. In the Avatar XPRIZE Competition, the objective was to differentiate between canisters that weighed approximately 12 N and 2.5 N [49]. For this evaluation experiment, we aimed to assess whether the system could differentiate between even smaller weight differences. The Avatar robot lifted the bottles one by one, and the participants had to determine the sequence of the weights based on the force feedback. In the force feedback experiment, participants were required to identify the order of heaviness of the three objects solely using the haptic device, without relying on the information from the HMD. The test was considered successful only if the order was entirely correct.

- Identifying stone surfaces (Fig. 18c): This task tests the ability to identify the roughness of a surface without visual feedback. A curtain obstructed the view between the robot and the table, preventing the participants from seeing the table. This task assesses the ability to determine the roughness of a surface without visual feedback. On the table, two stones were present- one rough and the other smooth. The participants used the robot's left hand to differentiate the roughness of the stones. The success or failure of identifying the rough stone was then measured.

- Mobility (Fig. 18d): During this task, we evaluated how adept participants were at controlling a mobile base using a pedal-shaped interface connected to a visual interface. The start and finish points are marked with rectangles. The robot had to navigate around a large table to reach the finish point, which was approximately 6.5 m away from the start point. The participants had to move the robot to the finish point without colliding with the obstacle. If the robot hit the table, it was considered a failure. We measured the total time it took from the moment the robot departed the start point until the participants believed that the robot's mobile base had completely entered the finish point.
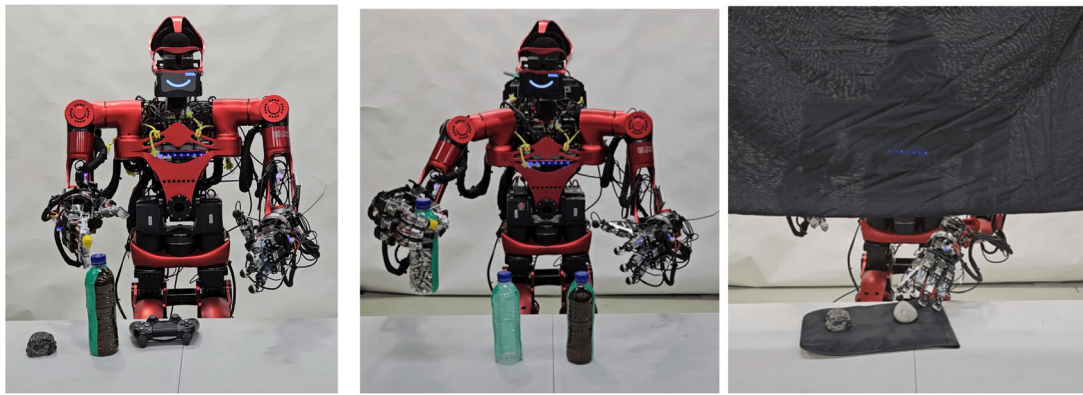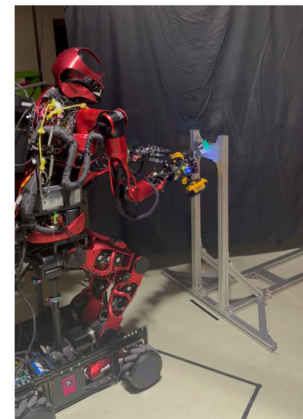
(a) Evaluation test for manipulation and grasping objects.

(b) Evaluation test for force feedback.

(c) Evaluation test for identifying stone surface.
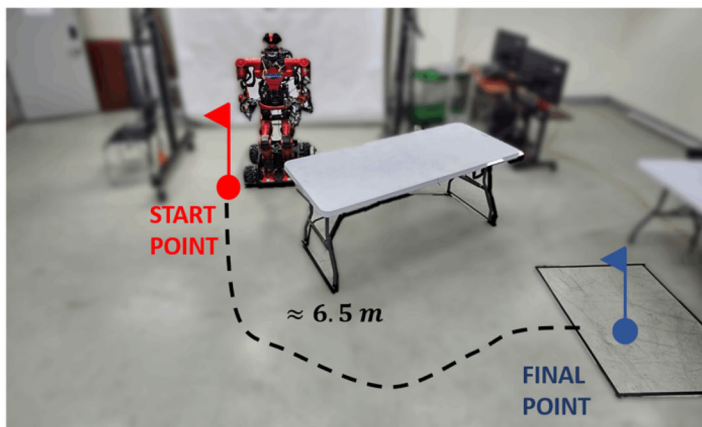
(d) Evaluation test for mobility.

(e) Evaluation test for drill maneuverability.

Fig. 18 Snapshots of benchmark tasks. **a** Test of grasping a stone, an empty bottle, and a joystick controller. **b** Test of force feedback to identify the weight of three bottles (12.7 N, 4.9 N, and 0.24 N). **c** Test of identifying the roughness of stone surfaces (rough and smooth). This curtain effectively blocks the operator from viewing the stones. **d** Test of mobility in an environment with a trajectory length of 6.5 m and one obstacle). **e** Test of drill maneuverability (drill weight: 1.7 kg, bolt height from the ground: approximately 1 m). (Color figure online)

- Drill maneuverability (Fig. 18e): This task was designed to evaluate the participant's ability to perform precise manipulation tasks with heavy tools using a haptic device. The aim was to test whether they could teleoperate both the robot arm and hand to move towards a bolt, which was positioned approximately 1 m above the ground, and then loosen it by activating the drill. The force feedback from the drill was removed during the drill evaluation test, as it was deemed too heavy and could negatively affect the participant's teleoperation. It's important to note that grasping the drill was not evaluated in this task. The total time taken to remove the bolt by the drill was measured from the moment the robot, holding the drill, was placed in front of the workspace.

5.2 Results

The results of the experiment are presented in Table 5, which shows the average completion times and success rates, along with their standard deviation (SD). You can find the experiment in the following link [71]. Each evaluation in Fig. 19 highlights the observed tendencies within each group, despite the small number of participants ranging from 3 to 4 in each experimental group.

- Manipulation and grasping task: The average time taken by each participant to complete the task is 79.5 ± 37.9 s, with completion times ranging from 28 s to 174 s. The standard deviation of 37.9 s indicates that the participant's ability affects their performance in the manipulation and grasping process. In the manipulation and grasping evaluation test, the box plot for each participant group

Table 5 Result of system evaluation experiment

| Participants	Manipulation and grasping (s)	Identifying weight	Identifying stone surface	Mobility (s)	Drill maneuverability (s)	System experience (min/level)
1	79	Success	Success	52	124	80 min/Beginner
2	174	Success	Success	44	28	40 min/Beginner
3	74	Fail	Success	38	27	60 min/Beginner
4	52	Success	Success	40	31	160 min/Intermediate
5	73	Success	Success	58	40	100 min/Intermediate
6	90	Fail	Success	41	28	100 min/Intermediate
7	98	Success	Success	51	35	100 min/Intermediate
8	28	Fail	Success	35	16	180 min/Expert
9	41	Fail	Success	51	28	200 min/Expert
10	86	Success	Success	56	41	240 min/Expert
Mean	79.5	–	–	46.6	39.8	–
Success rate	–	60%	100%	–	–	–
SD	37.9	1.55	0	7.6	28.9	–

is displayed in Fig. 19a. The mean completion times for the beginner, intermediate, and expert groups were 109 ± 56.3 s, 78.3 ± 20.3 s, and 51.7 ± 30.4 s, respectively. It is observed that the duration of the task is inversely proportional to the participant's experience level.

- Identifying weight: The success rate of the participants in the identifying weight task is 60 %. All participants could identify the lightest object, but some had trouble distinguishing between the objects of medium and heaviest weight. The success rate of each group is displayed in Fig. 20. Interestingly, the success rate of identifying weights did not seem to be related to the level of experience with the system, which could be associated to the clipping issue described in Fig. 14. As explained in Sect. 4.2, the desired force transmitted to the participants F_d of each bottle after scaling are 0 N, 11.6 N, and 42.8 N (we use $K_{scaling}$ as 4, which is the same value used in the ANA Avatar XPRIZE). However, the weight of the heaviest bottle exceeds $F_{clipping} = 13.5$ N, so F_d for the heaviest bottle becomes 13.5 N instead of 42.8 N. Therefore, the weight difference between the second heaviest object and the heaviest object becomes only 1.9 N. Also, due to the significant inertia of the haptic device, a participant with less sensitivity might have difficulty distinguishing the 1.9 N difference and the counterforce arising from inertia. For this reason, we speculate that the expert group found it challenging to distinguish between the force feedback arising from the 0.9 N difference and the counterforce from inertia, likely due to their familiarity with the inertia of the haptic device, as they tended to operate the haptic device more quickly.

- Identifying stone surfaces: All participants were able to identify stone surfaces with a success rate of 100 %, regardless of their level of experience with the system.

- Mobility: The average time for the mobility test for all participants was 46.6 s, with a range from 35 s to 58 s. The mobility experiment was the one in which the influence of system experience appeared to be the least significant. As shown in Fig. 19b, the average completion times for each group are 44.6 s, 47.5 s, and 47.3 s, respectively. Surprisingly, having more experience with the system did not reduce the completion time. Thus, the results suggest that the mobile base system is intuitive and allows users with less system experience to perform similar operations as those with more system experience.

- Drill maneuverability: During the drill maneuverability test, the average time taken by all participants was 39.8 ± 28.9 s. The drill experiment was conducted to evaluate the time taken by each participant to complete the task, and the results showed that the shortest time taken was 16 s and while the longest time taken 124 s. This indicates that there was a considerable variation in completion time. The mean and standard deviation for each group are shown in Fig. 19c: 59.7 ± 55.7 s for the beginner group, 33.5 ± 5.2 s for the intermediate group, and 28.3 ± 12.5 s for the expert group. The drill task was a complex experiment that required participants to align the drill held by the robot with a small bolt and then manipulate the drill button. Interestingly, except for the first participant, the mean time taken by the remaining participants to complete the drill task was 30.4 s, with a standard deviation of 7.62 s. This suggests that the proposed system does not pose significant difficulties in performing fine tasks using the drill. The first partic-
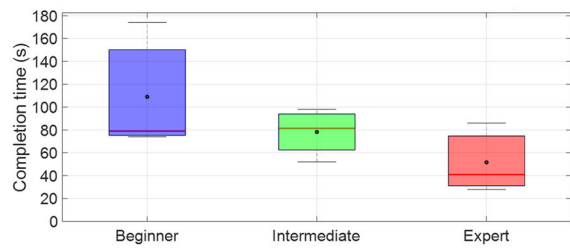
(a) Box and whisker plot of the evaluation of manipulation and grasping.
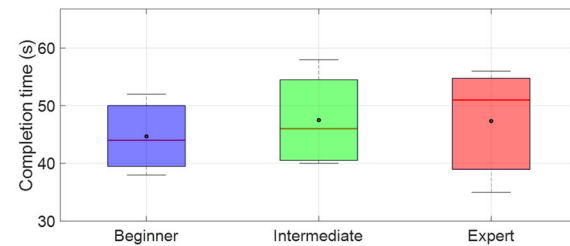
(b) Box and whisker plot of the evaluation of mobility
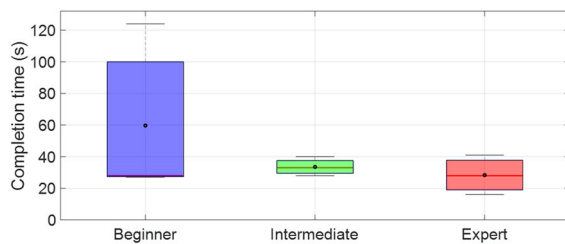
(c) Box and whisker plot of the evaluation of drill maneuverability

Fig. 19 Box and whisker plots of each evaluation test: manipulation and grasping, mobile and drill maneuverability. The X-axis of each figure shows the groups of participants. The Y-axis of each figure shows the completion time of each evaluation test. The blue boxes correspond to the beginner group, the green boxes to the intermediate group, and the red boxes to the expert group. Additionally, the red lines within the boxes represent the median values, while the black circles represent the mean values. (Color figure online)

ipant took over 120 s because he had difficulty keeping the drill perpendicular to the wall while attempting to remove the bolt. The group of beginners showed more variability than any other group in the experiment.

The average time taken to complete the three measured tasks was 213.3 s for the beginner group, 159.3 s for the intermediate group, and 127.3 s for the expert group. This indicates that participants with more experience tend to handle the system more efficiently. However, the small number of participants in each group makes it difficult to consider these results to be highly conclusive. Therefore, it would be beneficial to conduct user studies with a larger number of participants in future research on avatar systems. Additionally, if we train participants who have never used the system
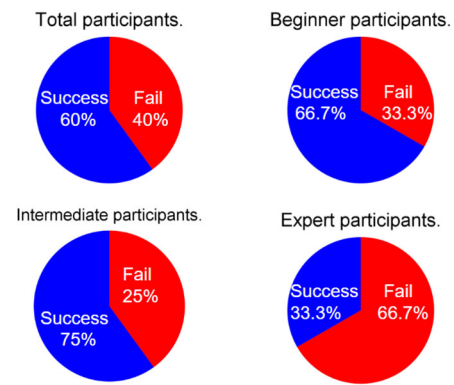
Fig. 20 The success rate of the evaluation of identifying weight. The blue parts show the success rate, and the red parts show the failure rate. The graphs show, clockwise from the left top, the ones for the total participants, the beginner group, the expert group, and the intermediate group, respectively. (Color figure online)

before on how to operate the avatar system and then compare their task performance based on the duration of their training, we can better analyze the relationship between the amount of experience with the system and the ability to use it effectively.

6 ANA Avatar XPRIZE Finals

In this section, the missions and results of the ANA Avatar XPRIZE will be outlined. We will also provide a brief introduction and analysis of the interface used by other teams, and discuss the valuable lessons and insights gained from our participation in the competition.

6.1 Missions of ANA Avatar XPRIZE Finals

During the Avatar XPRIZE finals, teams were ranked based on their scores, with a maximum of 15 points available. The Avatar ability was worth 10 points, the Operator experience was worth 3 points, and the Recipient experience was worth 2 points. The competition tested the avatar system's ability through 10 missions conducted by the operator. These missions allowed operators to evaluate the system's performance and effectiveness. The locations where these ten missions were carried out are described in Fig. 21. Detailed descriptions of each mission can be found in Table 6. In the Avatar Ability category, teams scored one point for a pass and zero points for a fail. To proceed to the following missions, teams had to succeed in the current mission, and if the allocated time passed before they succeeded, their trial would end.

The judges scored the Operator and Recipient Experience tasks: 0 points for Never/Poor, 0.5 points for Sometimes/Fair, and 1 point for Always/Good. Table 6 provides detailed explanations of the judge's evaluations.
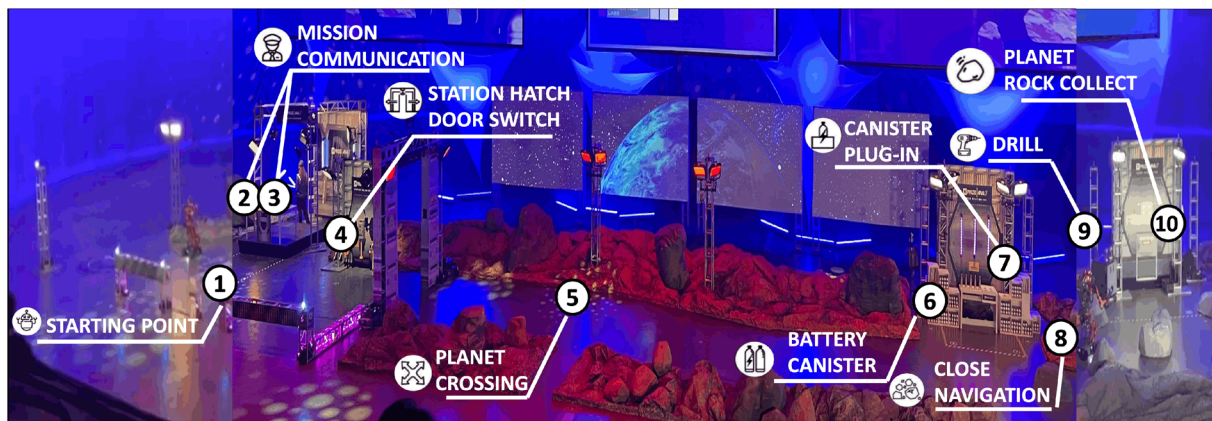
Fig. 21 Description of ANA Avatar XPRIZE Finals missions and test course. The operator is in another room. (Color figure online)

Table 6 Missions of the ANA Avatar XPRIZE finals

|	Mission	Scored point	Testing contents
Avatar ability	1	Was the Operator able to move the Avatar system to the designated area?	Basic mobility
	2	Was the Operator able to utilize the Avatar system to introduce themselves to the mission commander?	Audio and video
	3	Was the Operator able to utilize the Avatar system to confirm (or reiterate) the mission goals?	Audio and video
	4	Was the Operator able to utilize the Avatar system to activate the switch?	Grasping
	5	Was the Operator able to move the Avatar system to the next designated area?	Advanced mobility over distance
	6	Was the Operator able to distinguish the heavy canister through the Avatar system?	Ability to identify weight
	7	Was the Operator able to use the Avatar system to lift and place the heavy canister into the designated slot?	Manipulation
	8	Was the Operator able to navigate the Avatar system through a narrow pathway to get to the designated area?	Navigation and mobility
	9	Was the Operator able to use the Avatar system to utilize a drill within the domain area?	Advanced manipulation
	10	Was the Operator able to feel the texture of the object without seeing it and retrieve the requested item using the Avatar system?	Haptics
Operator experience	1	Did the Avatar System enable the Operator to feel present in the remote space and convey appropriate sensory information?	–
	2	Did the Avatar System enable the Operator to clearly understand (both see and hear) the Recipient?	–
	3	Was the Avatar System easy and comfortable to use?	–
Recipient experience	1	Did the Avatar Robot enable the Recipient to feel as though the remote Operator was present in the space?	–
	2	Did the Avatar Robot enable the Recipient to clearly understand (both see and hear) the Operator?	–
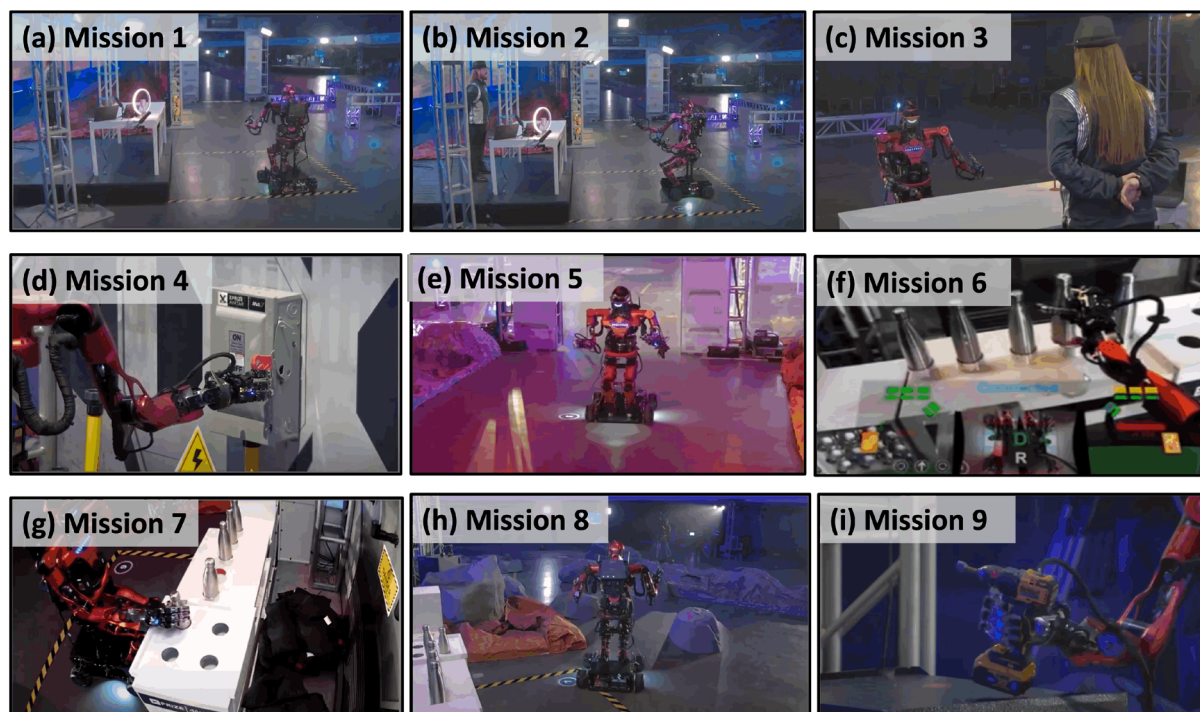
Fig. 22 Snapshots of Team SNU performing the final missions in DAY 1. **a** Mission 1. **b** Mission 2. **c** Mission 3. **d** Mission 4. **e** Mission 5. **f** Mission 6. **g** Mission 7. **h** Mission 8. **i** Mission 9. Team SNU did not try Mission 10. (Color figure online)

6.2 Result of the ANA Avatar XPRIZE Finals

In the Avatar Finals, Team SNU received a score of 12.5 points for DAY 1[20] and DAY 2[21] (8 points for Avatar Ability, 4.5 points for Judge Experience). Figure 22 shows the missions our team carried out during the two days of the Final competition. Team SNU attempted 9 out of the 10 missions over two days, completing 8 of them. However, mission 9, which involved grasping a drill and unscrewing the bolt with the drill, was unsuccessful.

In this section, we provide a detailed explanation of the reason behind the failure of mission 9. The task required grabbing a drill from a table, turning it on, and then moving to the next wall to remove a bolt by unscrewing it. On DAY 1, there were several factors that led to the failure of mission 9, as shown in Fig. 23. In Fig. 23a, b, it is evident that the operator grabs the drill while the orientation of the drill and the robot hand is not aligned. Consequently, the index finger could not fully push the drill button. Furthermore, the operator could not position the drill perpendicular to the wall due to the unaligned orientation of the drill. As a result, mission 9 on DAY 1 failed.

On DAY 2, mission 9 failed again. The reasons behind this failure are shown in Fig. 24. The operator on DAY 2 successfully grasped the drill and placed the index finger on the drill button, as shown in Fig. 24a, b. However, after moving the robot into the wall (as shown in Fig. 24c, d), the operator attempted to move the mobile base instead of moving the arm. As shown in Fig. 24, even though the robot hand holding the drill hit the wall, the operator continued to move the mobile base, causing the drill to rotate in the hand. Consequently, mission 9 failed.

As described in Sect. 5.2, the drill task was a difficult task that an operator with limited experience with our avatar system could not proficiently perform. Furthermore, the system has limitations, such as the difficulty in showing the robot hand holding the drill from various angles and the inability to detect obstacles and collisions between the robot and its surroundings, akin to depth perception. The system's failure to provide appropriate guidance to the operator on whether it is better to move the mobile base or manipulate the robot arm for task execution might have led to the failure of the drill task.

6.3 Analysis of the ANA Avatar XPRIZE Finals

Table 7 shows the scores of the selected finalists; 12 teams were selected to participate in the DAY 2 test. The time shown in Table 7 is the time when each team succeeded

---

[20] Day 1 of Team SNU https://youtu.be/lOnV1Go6Op0?si=Wogj15Z9pR_OG2sp&t=11086.

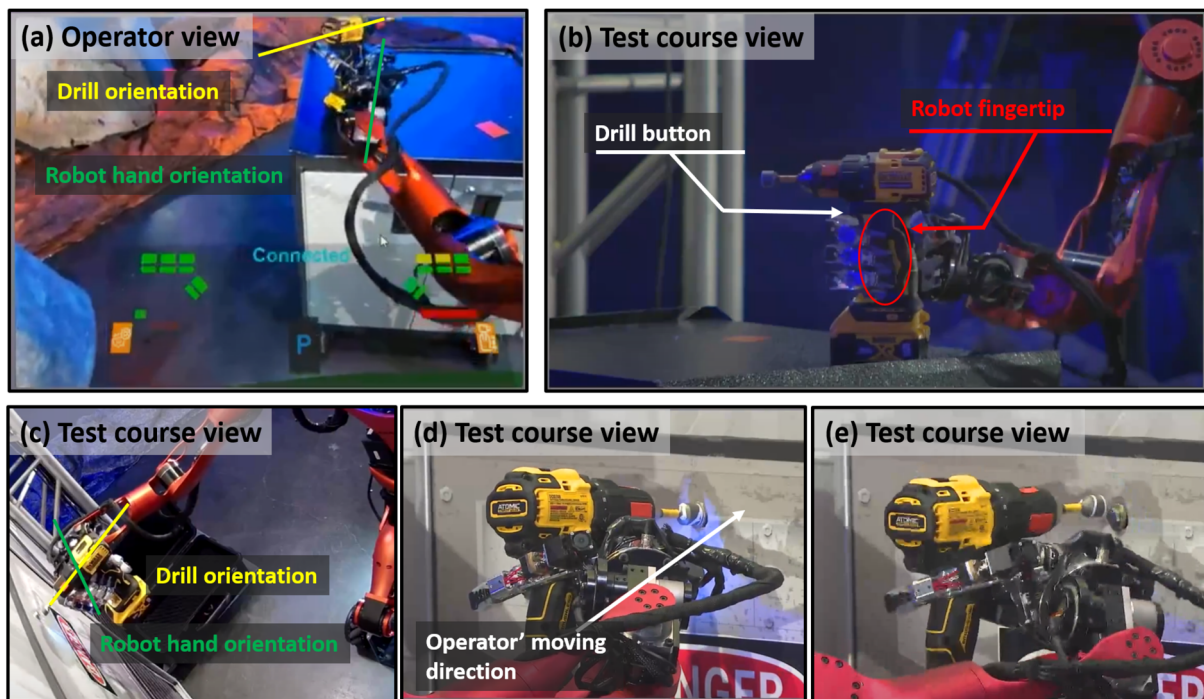[21] Day 2 of Team SNU https://youtu.be/lOnV1Go6Op0?si=d-_nO87YH0aUSzLZ&t=41633.

Fig. 23 Snapshots of the drill mission failure on DAY 1: **a** the operator view when the orientation of the drill and the orientation of the robot hand are not aligned while holding the drill. **b** Test course view when the robot fingertip is not placed on the drill button. **c** The test course view when the robot hand is approaching a wall with the drill misaligned in the robot hand. **d** The test course view reveals a discrepancy between the operator's moving direction toward the wall and the orientation of the robotic hand. **e** The test course view when the misalignment of orientation between the drill and the robot hand causes the drill button to be released. (Color figure online)

in its last mission. Only four teams, namely NimbRo [72], Pollen Robotics [73], Team Northeastern [53], and AVA-TRINA [54], completed all the missions. Also, only four teams (NimbRo, Pollen Robotics, i-Botics [74], and Inbiodroid [75]) were able to achieve the perfect judge score of 5. Furthermore, only two teams, NimbRo and Pollen Robotics, were able to achieve a perfect score in both the task and judge evaluations. These scores reveal that the avatar system, which receives the maximum scores from the judges, does not guarantee perfectly executed missions.

In Fig. 25, the execution times of Team SNU and the top 5 team's missions are compared. The scoreboard on the released video[22] measures each mission execution time, but there may be some emerging errors. As shown in Fig. 25, missions 2 and 3 required the Recipient and Operator to converse about the overall mission, and most teams completed the mission within a similar execution time.

Our team's performance on missions 6 and 7 was on par with the top 3 teams, suggesting that our robotic avatar system is adept at providing force feedback to the operator and accurately relocating objects to specified locations, much like the top-ranking teams.

Nine teams made an attempt at mission 9, with a success rate of six out of nine. This particular mission necessitated the lifting and manipulation of weightier objects in comparison to mission 6. Additionally, the task involved the precise pressing of the drill button, and the haptic feedback system posed a challenge for the operator, given the object's weight.

Only four teams succeeded in completing mission 10 and had varying completion times. The method of preparation for each team to measure and convey the texture of the stone to the operator varied, resulting in varying execution times required for mission 10. Nimbro [72] used an audio sensor on the robot finger for detecting and a vibrotactile actuator for the feedback [76]. Team Northeastern [53] also used an audio sensor (microphone on the wrist) for tactile feedback [53]. Team AVATRINA used the LiDAR camera on the gripper to detect the surface [54]. Pollen Robotics also used an audio system for stone surface detection. (Pollen Robotics has not published research findings, but the final competition video[23] shows microphones and thin white plates attached to both grippers. Additionally, in the last mission, the white plate is used to scrape on the stone surface.) Interestingly, not a
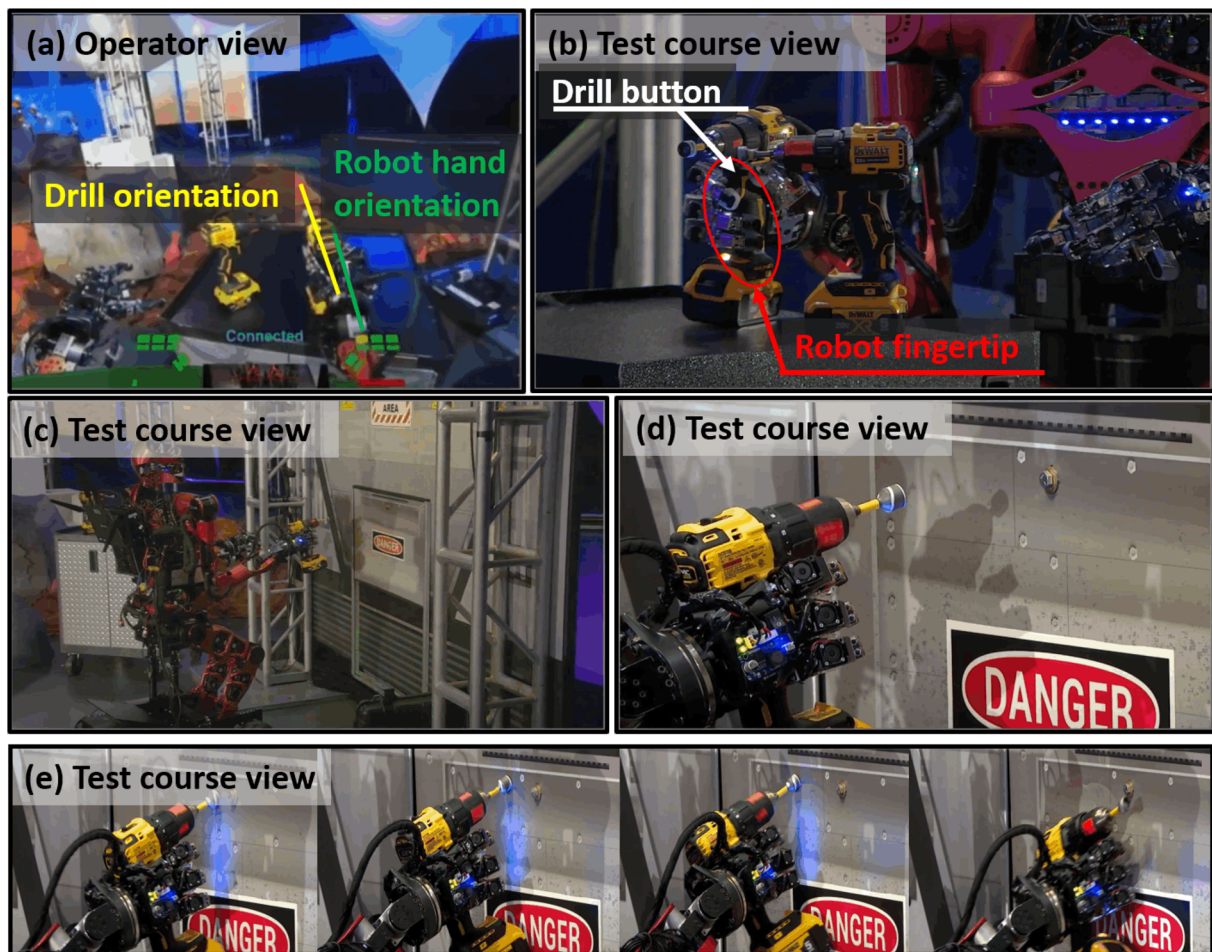
Fig. 24 Snapshots of the drill mission failure on DAY 2: **a** The operator view when the orientation of the drill and the orientation of the robot hand are aligned while holding the drill. **b** Test course view showing the robot's index fingertip placed on the drill button. **c** Test course view capturing the robot approaching the wall while holding the drill. **d** Test course view before the robot holding the drill collides with a wall. **e** Snapshots illustrating the drill rotating within TOCABI's hand after it hits the wall. The operator in DAY 2 did not stop the mobile base in front of the wall, causing the robot to continue moving towards the wall. As a result, snapshots depict the robot's hand unintentionally colliding with the wall. (Color figure online)

single team utilized direct feedback through contact with the robot hand to differentiate the roughness of the stone.

6.4 Comparison of Avatar Systems

The robotic avatar systems of ANA Avatar XPRIZE Finals are briefly compared in Table 8. While it is possible to rank each team based on their competition performance, it is difficult to say which team's system was the best approach. Therefore, in this section, we examine the methods that the participating teams used most frequently.

The most commonly used forms of the Avatar robot combined a wheeled base with dual arms: nine teams utilized a humanoid-type upper body, six teams used two manipula-
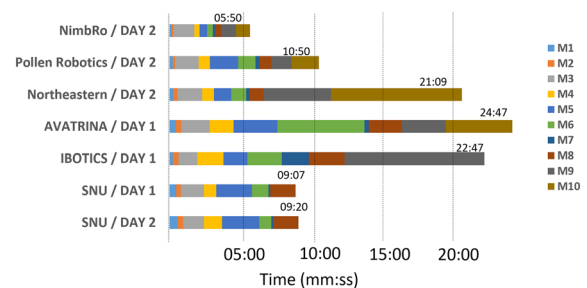

Fig. 25 Comparison of mission execution time for top 5 teams and Team SNU. M1: Moving to commander desk. M2: Reporting to the commander. M3: Receiving and confirming missions. M4: Activating switch. M5: Traveling to the next task. M6: Identifying the heavy canister. M7: Placing canister into the slot. M8: Navigating a narrow path. M9: Using the drill to remove the door. M10: Identifying the rough textured rock and retrieving it. (Color figure online)

Table 7 Result scores of the 12 teams in the ANA Avatar XPRIZE Finals last day

Rank	Team	Total	Score task	Judge	Time (MM:SS)
1	NimbRo	15.0	10	5.0	05:50
2	Pollen Robotics	15.0	10	5.0	10:50
3	Team Northeastern	14.5	10	4.5	21:09
4	AVATRINA	14.5	10	4.5	24:47
5	i-Botics	14.0	9	5.0	22:47
6	Team UNIST	13.5	9	4.5	17:39
7	Inbiodroid	13.0	8	5.0	16:25
8	Team SNU	12.5	8	4.5	09:07
9	AlterEgo	12.5	8	4.5	10:27
10	Dragon Tree Labs	11.0	7	4.0	23:43
11	Avatar-Hubo[a]	9.5	6	3.5	17:00
12	Last Mile	9.0	5	4.0	09:30

The scores and total time of each team are obtained through the scoreboard

[a]The score for task 6 of Avatar-Hubo was updated after their turn was completed. Therefore, the total time for Avatar-Hubo was recorded based on the competition video, and it might be incorrect

tors, and two teams used one manipulator for manipulation.[24] Additionally, 14 teams employed wheels, two used legs, and one used a combined wheel-leg robot for mobility. Only iCub [81] and Janus [82] used bipedal locomotion. During the Avatar XPRIZE Finals, we (Team SNU) used a legged humanoid robot, primarily relying on the mobile base with the robot seated. Thus, Team SNU falls into the category of teams that use wheels. Avatar-Hubo utilized a robot capable of transforming between bipedal walking mode and wheel mode [79]. Avatar-Hubo utilized bipedal walking mode for manipulation missions and wheel mode for mobility missions. 10 teams used robotic hands, and 5 teams used grippers. Two teams, AVATRINA [54] and Cyberselves|Touchlabs, used both robotic hand and gripper.

All teams, except Dragon Tree Labs and Last Mile, have employed a method of remotely controlling robots based on the gestures of operators. Dragon Tree Labs and Last Mile [80] used the joystick controller or mouse. The other teams used different interfaces to mimic the operator's behavior: eight teams used trackers, five teams used a haptic arm, five teams used VR controllers, and one team used an exoskeleton.[25] Within this context, a differentiation has been established between a haptic arm and an exoskeleton, based on the connection between the operator and the device. Specifically, it is determined whether the connection is limited to a single point, such as the wrist, or if it extends across multiple locations throughout the body. The haptic feedback device developed by Team SNU was classified as a haptic arm

since it involves a connection between the operator and the device at a single point. Additionally, it was clarified during the ANA Avatar XPRIZE Workshop that Pollen Robotics' exo-elbow is not intended as a remote control device, but rather as a device designed for providing haptic feedback.[26] The consensus among numerous teams appears to favor an intuitive teleoperation interface that mimics the operator's gestures. Among the total teams, 4 teams used one device for the teleoperation interface, while 13 teams used two or more devices. 12 teams used gloves to control robotic hands or grippers, while the remaining teams remotely operated both robot arms and hands or grippers through a single teleoperation interface.

The ratio of teams that used hands and teams that used feet as the interface to control robot movement was similar: nine teams used feet (3D pedal, one-foot pedal, and trackers), and eight teams used hands (VR controller, flight joystick, and mouse).

14 teams used VR-capable HMD as the interface to deliver telepresence to the operator. Two teams used a larger monitor, and one team used a regular monitor.

6.5 Lessons Learned

The preparation and testing for the ANA Avatar Finals indicated future research directions for us and the community. What we learned during the preparation and the testing of the ANA Avatar XPRIZE is presented as follows:

---

[24] The distinction between 'two manipulators' and 'humanoid type' is whether the 'two arms' are separately attached to the robot, or whether they are integrated with the torso to form a unified body.

[25] In this sentence, 'teams' refers to separate groups using each technology.

---

[26] Also in the video, they explained the operator received feedback through the exoskeleton. https://youtu.be/ZZR957IssHA?si=cGhUJqFJea9my_nl&t=144.

Table 8 Comparison of robots and operation systems participating at ANA Avatar XPRIZE

Rank	Team	Robot type	Teleoperation interface	Mobile interface	Telepresence interface
1	NimbRo [52, 72]	Two manipulators Mobile robot Hands	Haptic arm Tracker Gloves	3D round pedal	VR-HMD
2	Pollen Robotics[a] [73]	Humanoid upper body Mobile robot Gripper	Exo-elbow VR Controller	VR Controller	VR-HMD
3	Team Northeastern [53]	Two manipulators Mobile robot Grippers	Haptic arm Gloves	One-foot pedal	Big monitor
4	AVATRINA [54]	Two manipulators Mobile robot Hand/Gripper	Trackers Glove/VR Controller	VR Controller	VR-HMD
5	i-Botics [74, 77]	Humanoid robot with wheels Hands	Haptic arm Gloves	3D plate pedal	VR-HMD
6	Team UNIST [55]	Humanoid upper body Mobile robot Hands	Haptic arm Trackers Gloves	3D round pedal	VR-HMD
7	Inbiodroid [75]	Humanoid upper body Mobile robot Hands	Exoskeleton controller Gloves	2D feet pedal	VR-HMD
8	Team SNU	Humanoid robot with legs Mobile robot Hands	Haptic feedback device Trackers Gloves	3D flight pedal and switches	VR-HMD
9	AlterEgo [78]	Humanoid upper body Two-wheeled mobile base Hands	VR Controller	VR Controller	VR-HMD
10	Dragon Tree Labs	Two manipulators Mobile robot Grippers	Stick Controller	Flight joystick	Wide curved monitor
11	Avatar-Hubo [79]	Transformable humanoid robot Legged and wheels Hands	Trackers Gloves	Game Controller	VR-HMD
12	Last Mile [80]	One manipulator Mobile robot Gripper	Stick Controller Mouse	2D joystick	Monitor
13	AvaDynamics	One manipulator Mobile robot Gripper	VR Controller	VR Controller	VR-HMD
14	iCub [81]	Humanoid robot with legs Legged locomotion Hands	Trackers Gloves	Trackers on feet	VR-HMD
15	Tangible	Two manipulators Mobile robot Hands	Trackers Gloves	Joystick	VR-HMD
16	Cyberselves -Touchlab	Two manipulator Mobile robot Hand/Gripper	Gloves	Feed pedal	VR-HMD
17	Janus [82, 83]	Humanoid robot with legs Legged locomotion Hands	Tracker Gloves	Trackers on feet	VR-HMD

[a]Pollen Robotics have not published their research

- **Fast Networking With Low Latency:** The teleoperation in real-time requires fast communication between the robot and the operator. During the competition, there were communication delays at the venue on the qualification day. Some teams reported encountering issues related to network disconnection or drops, such as unexpected network drops in UNIST [55] and networking disconnection in AVATRINA [54]. Instances of disconnection or drops in networking can result in delays between the operator and the robot, making real-time remote control or immediate feedback unfeasible. Therefore, low-latency fast networking is indispensable for a robotic avatar system.

- **Intuitive and Ergonomic Teleoperating System with Force Feedback:** Using a haptic device and VIVE trackers together had the advantage of the robot closely tracking the operator's movements, faithfully mimicking the actions, and providing force feedback to the operator. However, to provide force feedback, the operator device must include actuators, thereby increasing the inertia of the device and diminishing the system's ergonomic qualities. Team Northeastern also mentioned that the difference in inertia in master and slave is unsuitable for teleoperation, as it causes disturbance to the operator [53]. We infer that this contributed to the execution time of the mission demonstrated using the drill in Sect. 5 and our failure of the drill mission twice in the ANA Avatar competition. We believe that compensating for inertial-induced unintended movements in the haptic device could prevent this failure. This can be achieved by installing F/T sensors between the operator's hand and the haptic device, measuring unintended movements through these sensors, and applying necessary corrections. This approach resembles Nimbro's method, which involves detecting the operator's movements using an F/T sensor attached to the operator's arm and controlling the operator arm accordingly [84]. An alternative approach involves separating force feedback from the operator system, employing a method akin to AVATRINA [54], where force information is transmitted through the visual system.

- **Visual Feedback Should Provide a Wide Field of View and Diverse Perspectives:** Visual information emerges as the most effective means for operators to comprehend the robot's surroundings. Nonetheless, the current system's cameras provide the operator with a narrower field of view (FoV) compared to human natural vision, resulting in the operator receiving less visual information through the robot's perspective than they would with their own eyes. Furthermore, the limited visual information conveyed can lead the operator to make mistakes. In our team's case, as emphasized in Sect. 6.2, the operator faced difficulty visually confirming whether he

securely grasped the drill during Mission 9. Additionally, the operator failed to determine if the mobile platform was continuously moving towards the wall, resulting in instances where the hand holding the drill collided with the wall. As another example, in the case where the iCub team collided with the door frame due to the operator's mistake,[27] a wider FoV would likely have reduced the chances of such operator errors. A system like Nimbro, offering six degrees of freedom to move the camera [72], or a system like Northeastern, providing depth information through lasers as visual data [53], could have potentially prevented these issues. Considering this, the avatar system should be developed to offer visual feedback with a wide field of view and allow the operator to easily move the camera to see the object from various angles.

- **Difficulty of Bipedal Walking:** During the competition, four teams used bipedal robots: Team SNU, Team Avatar-Hubo [79], iCub [81] and Janus [82]. However, only two teams, iCub and Janus, attempted bipedal walking in the competition. Despite the competition venue being visibly flat and suitable for bipedal robots to navigate, both teams that attempted bipedal walking faced challenges and did not achieve satisfactory results. While it is inherent that bipedal walking on flat ground may be slower than wheeled movement, the reality is that slight variations in the floor make the surrounding environment uneven. Ultimately, developing robots capable of navigating in 3D environments remains a challenge.

- **Need to Develop a Robot Hand that can Move and Feel Similar to a Human Hand:** The missions of the ANA Avatar XPRIZE underscored the significance of developing robotic hands that can exhibit flexible movements similar to a human hand and receive tactile feedback in contact with objects. On Day 1, during the drill mission, the operator grabbed the drill, with the robot index finger's middle phalanx link pressing the drill's button. While the kinesthetic feedback indicated successful drill grasping, it failed to specify which finger pushed the drill button. Furthermore, even if the operator had detected a change in the drill's orientation, the robot hand could not rotate the drill without placing it back on the table. Although our developed robot hand effectively grasped objects with in-hand motion. Team UNIST also mentioned the challenge of developing a robot hand capable of moving as freely as a human hand [55]. Additionally, the feasibility of tactile feedback for each finger link could significantly enhance the operator's ability to sense object roughness. In our team's case, we used an RGB camera and recognition algorithm to discern the

---

[27] On Day 1 of the iCub trial: https://youtu.be/lOnV1Go6Op0?si=c6DtZzaw9UWrT7le&t=5026.

roughness of stones [69]. To perceive the stone's roughness, the four teams that completed the mission employed sound feedback [53, 72] or LiDAR [54] to perceive the stone roughness. To enable a robot hand to perceive the object's roughness similarly to a human, it necessitates the development of tactile sensing throughout each finger's links in the robot hand. Additionally, structural and control advancements are required to develop a robot hand capable of free movements, resembling the motions of a human hand.

- **Shared Autonomy Control:** Many teams spent significant time attempting to grasp the drill, activate it, and unscrew the screw, with only a few teams achieving success. While the advanced teleoperation system enabled the operator to control the robot step by step as desired, it posed challenges in cases requiring fine control, making precise robot manipulation more demanding. AVATRINA implemented semi-autonomy technology, distinct from shared autonomy, and reported its advantageous impact on the operator [54]. The development of shared autonomy control, integrating both manual control by a human and autonomous control by a robot, could address these challenges in remote operation. The operator's remote manual control moves the robot's position or approaches the target object. Simultaneously, the robot's autonomous control aligns its hand with the object or adjusts the position and direction of the held tool to match that of the target object. This streamlined approach reduces the time required for executing remote operations using a robot.

7 Conclusion

This paper comprehensively describes our robotic avatar system, comprising the humanoid robot TOCABI and an operator station facilitating remote control. Our system is designed to provide operators with an intuitive teleoperation experience, ensuring an immersive telepresence. The effectiveness of the proposed system was validated through self-conducted evaluation tests and participation in the ANA Avatar XPRIZE Finals. It allows operators to remotely control the Avatar robot based on their movements, providing a physical stimulus of haptic feedback that enables the operator to sense the weight of objects and distinguish the roughness of surfaces.

During the ANA Avatar XPRIZE Finals, our robotic avatar system empowered operators to complete 8 out of 10 missions with just one hour of training. However, limitations were identified, notably the significant inertia of the haptic feedback device, which poses challenges for precise remote control. Additionally, although the participants in the evaluation tests demonstrated a high task completion rate in

evaluation tests, the method for discerning the roughness of stones, although not attempted in the Finals, still presents a gap compared to human perception. Our plans involve developing an advanced, intuitive teleoperation interface with minimal inertia, informed by extensive user studies. Furthermore, ongoing research is essential to strike a balance between autonomous controls facilitating fast and precise robot manipulation, potentially surpassing human capabilities, and providing exact haptic and control feedback.

Declarations

References

1. Takeuchi K, Yamazaki Y, Yoshifuji K (2020) Avatar work: telework for disabled people unable to go outside by using avatar robots. In: Companion of the 2020 ACM/IEEE international conference on human–robot interaction, pp 53–60

2. Obo T, Hase R, Kobayashi K, Sueta K, Nakano T, Shin D (2020) Cognitive modeling based on perceiving-acting cycle in robotic avatar system for disabled patients. In: 2020 international joint conference on neural networks (IJCNN). IEEE, pp 1–6

3. Barbareschi G, Kawaguchi M, Kato H, Nagahiro M, Takeuchi K, Shiiba Y, Kasahara S, Kunze K, Minamizawa K (2023) "I am both here and there" parallel control of multiple robotic avatars by disabled workers in a café. In: Proceedings of the 2023 CHI conference on human factors in computing systems, pp 1–17

4. Tachi S, Tanie K, Komoriya K, Kaneko M (1985) Tele-existence (i): design and evaluation of a visual display with sensation of presence. In: Theory and Practice of robots and manipulators: proceedings of RoManSy'84: the fifth CISM-IFToMM symposium. Springer, pp 245–254

5. Kim S, Kim M, Lee J, Hwang S, Chae J, Park B, Cho H, Sim J, Jung J, Lee H et al (2018) Team SNU's control strategies to enhancing robot's capability: lessons from the Darpa robotics challenge finals 2015. The DARPA robotics challenge finals: humanoid robots to the rescue. Springer, Berlin, pp 347–379

6. Krotkov E, Hackett D, Jackel L, Perschbacher M, Pippine J, Strauss J, Pratt G, Orlowski C (2018) The Darpa robotics challenge finals: results and perspectives. The DARPA robotics challenge finals: humanoid robots to the rescue. Springer, Berlin, pp 1–26

7. Spenko M, Buerger S, Iagnemma K (2018) The DARPA robotics challenge finals: humanoid robots to the rescue, vol 121. Springer, Berlin

8. Schreiber G, Stemmer A, Bischoff R (2010) The fast research interface for the Kuka lightweight robot. In: IEEE workshop on innovative robot control architectures for demanding (Research) applications how to modify and enhance commercial controllers (ICRA 2010). Citeseer, pp 15–21

9. Brogårdh T (2007) Present and future robot control development-an industrial perspective. Annu Rev Control 31(1):69–79

10. Zhang P (2010) Advanced industrial control technology. William Andrew, Norwich

11. Sanfilippo F, Hatledal LI, Zhang H, Fago M, Pettersen KY (2015) Controlling Kuka industrial robots: flexible communication interface Jopenshowvar. IEEE Robot Autom Mag 22(4):96–109

12. De Pace F, Manuri F, Sanna A, Fornaro C (2020) A systematic review of augmented reality interfaces for collaborative industrial robots. Comput Ind Eng 149:106806

13. Shu B, Arnarson H, Solvang B, Kaarlela T, Pieskä S (2022) Platform independent interface for programming of industrial robots. In: 2022 IEEE/SICE international symposium on system integration (SII). IEEE, pp 797–802

14. Peters BS, Armijo PR, Krause C, Choudhury SA, Oleynikov D (2018) Review of emerging surgical robotic technology. Surg Endosc 32:1636–1655

15. Darvish K, Penco L, Ramos J, Cisneros R, Pratt J, Yoshida E, Ivaldi S, Pucci D (2023) Teleoperation of humanoid robots: a survey. IEEE Trans Robot 39(3):1706–1727

16. Nawab A, Chintamani K, Ellis D, Auner G, Pandya A (2007) Joystick mapped augmented reality cues for end-effector controlled tele-operated robots. In: 2007 IEEE virtual reality conference. IEEE, pp 263–266

17. Sasaki T, Miyata T, Kawashima K (2004) Development of remote control system of construction machinery using pneumatic robot arm. In: 2004 IEEE/RSJ international conference on intelligent robots and systems (IROS) (IEEE Cat. No. 04CH37566), vol 1. IEEE, pp 748–753

18. Barros, JJO, dos Santos, VMF, da Silva, FMTP (2015) Bimanual haptics for humanoid robot teleoperation using ROS and V-REP. In: 2015 IEEE international conference on autonomous robot systems and competitions. IEEE, pp 174–179

19. Diolaiti N, Melchiorri C (2002) Teleoperation of a mobile robot through haptic feedback. In: IEEE international workshop HAVE haptic virtual environments and their. IEEE, pp 67–72

20. Schwartz M, Sim J, Ahn J, Hwang S, Lee Y, Park J (2022) Design of the humanoid robot Tocabi. In: 2022 IEEE-RAS 21st international conference on humanoid robots (Humanoids). IEEE, pp 322–329

21. Pollard NS, Hodgins JK, Riley MJ, Atkeson CG, (2002) Adapting human motion for the control of a humanoid robot. In: Proceedings 2002 IEEE international conference on robotics and automation (Cat. No. 02CH37292), vol 2. IEEE, pp 1390–1397

22. Montecillo-Puente FJ, Sreenivasa M, Laumond JP (2010) On real-time whole-body human to humanoid motion transfer. In: ICINCO, pp 22–31

23. Koenemann J, Burget F, Bennewitz M (2014) Real-time imitation of human whole-body motions by humanoids. In: 2014 IEEE international conference on robotics and automation (ICRA). IEEE, pp 2806–2812

24. Miller N, Jenkins OC, Kallmann M, Mataric MJ (2004) Motion capture from inertial sensing for untethered humanoid teleoperation. In: 4th IEEE/RAS international conference on humanoid robots, 2004, vol 2. IEEE, pp 547–565

25. Penco L, Clément B, Modugno V, Hoffman EM, Nava G, Pucci D, Tsagarakis NG, Mouret JB, Ivaldi S (2018) Robust real-time whole-body motion retargeting from human to humanoid. In: 2018 IEEE-RAS 18th international conference on humanoid robots (Humanoids). IEEE, pp 425–432

26. Choi S, Kim J (2019) Towards a natural motion generator: A pipeline to control a humanoid based on motion data. In: 2019 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 4373–4380

27. Dallard A, Benallegue M, Kanehiro F, Kheddar A (2023) Synchronized human-humanoid motion imitation. IEEE Robot Autom Lett. https://doi.org/10.1109/LRA.2023.3280807

28. Zhang J, Li P, Zhu T, Zhang WA, Liu S (2020) Human motion capture based on Kinect and IMUs and its application to human-robot collaboration. In: 2020 5th international conference on advanced robotics and mechatronics (ICARM). IEEE, pp 392–397

29. Lee C-H, Choi J, Lee H, Kim J, Lee K-M, Bang Y-B (2017) Exoskeletal master device for dual arm robot teaching. Mechatronics 43:76–85

30. Mallwitz M, Will N, Teiwes J, Kirchner EA (2015) The Capio active upper body exoskeleton and its application for teleoperation. In: Proceedings of the 13th symposium on advanced space technologies in robotics and automation. ESA/Estec symposium on advanced space technologies in robotics and automation (ASTRA-2015). ESA

31. Lee WK, Jung S (2006) FPGA design for controlling humanoid robot arms by exoskeleton motion capture system. In: 2006 IEEE international conference on robotics and biomimetics. IEEE, pp 1378–1383

32. Rebelo J, Sednaoui T, Den Exter EB, Krueger T, Schiele A (2014) Bilateral robot teleoperation: a wearable arm exoskeleton featuring an intuitive user interface. IEEE Robot Autom Mag 21(4):62–69

33. Zhang T, McCarthy Z, Jow O, Lee D, Chen X, Goldberg K, Abbeel P (2018) Deep imitation learning for complex manipulation tasks from virtual reality teleoperation. In: 2018 IEEE international conference on robotics and automation (ICRA). IEEE, pp 5628–5635

34. Lipton JI, Fay AJ, Rus D (2017) Baxter's homunculus: virtual reality spaces for teleoperation in manufacturing. IEEE Robot Autom Lett 3(1):179–186

35. Whitney D, Rosen E, Phillips E, Konidaris G, Tellex S (2019) Comparing robot grasping teleoperation across desktop and virtual reality with ROS reality. In: Robotics research: the 18th international symposium ISRR. Springer, pp 335–350

36. Zhou T, Zhu Q, Du J (2020) Intuitive robot teleoperation for civil engineering operations with virtual reality and deep learning scene reconstruction. Adv Eng Inform 46:101170

37. Lim D, Kim D, Park J (2022) Online telemanipulation framework on humanoid for both manipulation and imitation. In: 2022 19th international conference on ubiquitous robots (UR). IEEE, pp 8–15

38. Park B, Jung J, Sim J, Kim S, Ahn J, Lim D, Kim D, Kim M, Park S, Sung E et al (2022) Team SNU's avatar system for teleoperation using humanoid robot: ANA avatar XPRIZE competition. In: RSS 2022 workshop on towards robot avatars: perspectives on the ANA Avatar XPRIZE competition

39. Caserman P, Garcia-Agundez A, Konrad R, Göbel S, Steinmetz R (2019) Real-time body tracking in virtual reality using a Vive tracker. Virtual Real 23:155–168

40. Almeida L, Lopes E, Yalçinkaya B, Martins R, Lopes A, Menezes P, Pires G (2019) Towards natural interaction in immersive reality with a cyber-glove. In: 2019 IEEE international conference on systems, man and cybernetics (SMC). IEEE, pp 2653–2658

41. Liu H, Zhang Z, Xie X, Zhu Y, Liu Y, Wang Y, Zhu SC (2019) High-fidelity grasping in virtual reality using a glove-based system. In: 2019 international conference on robotics and automation (ICRA). IEEE, pp 5180–5186

42. Yashin GA, Trinitatova D, Agishev RT, Ibrahimov R, Tsetserukou D, (2019) AeroVR: virtual reality-based teleoperation with tactile feedback for aerial manipulation. In: 2019 19th international conference on advanced robotics (ICAR). IEEE, pp 767–772

43. Zhou H, Yang L, Lv H, Yi K, Yang H, Yang G (2019) Development of a synchronized human-robot-virtuality interaction system using cooperative robot and motion capture device. In: 2019 IEEE/ASME international conference on advanced intelligent mechatronics (AIM). IEEE, pp 329–334

44. Kratz S, Ferriera FR (2016) Immersed remotely: evaluating the use of head mounted devices for remote collaboration in robotic telepresence. In: 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN). IEEE, pp 638–645

45. Mimnaugh KJ, Suomalainen M, Becerra I, Lozano E, Murrieta-Cid R, LaValle SM (2021) Defining preferred and natural robot motions in immersive telepresence from a first-person perspective. arXiv preprint arXiv:2102.12719

46. Gaemperle L, Seyid K, Popovic V, Leblebici Y (2014) An immersive telepresence system using a real-time omnidirectional camera and a virtual reality head-mounted display. In: 2014 IEEE international symposium on multimedia. IEEE, pp 175–178

47. Shin J, Ahn J, Park J (2022) Stereoscopic low-latency vision system via ethernet network for humanoid teleoperation. In: 2022 19th international conference on ubiquitous robots (UR). IEEE, pp 313–317

48. Tachi S (2016) Telexistence: enabling humans to be virtually ubiquitous. IEEE Comput Graph Appl 36(1):8–14

49. Behnke S, Adams JA, Locke D (2023) The $ 10 million ANA avatar XPRIZE competition: how it advanced immersive telepresence systems. IEEE Robot Autom Mag 30(4):98–104

50. Kim S, Kim M, Lee J, Hwang S, Chae J, Park B, Cho H, Sim J, Jung J, Lee H et al (2017) Team SNU's control strategies for enhancing a robot's capability: lessons from the 2015 Darpa robotics challenge finals. J Field Robot 34(2):359–380

51. Hauser K, Watson E, Bae J, Bankston J, Behnke S, Borgia B, Catalano MG, Dafarra S, van Erp JB, Ferris T et al (2024) Analysis and perspectives on the ANA avatar XPRIZE competition. Int J Soc Robot. https://doi.org/10.1007/s12369-023-01095-w

52. Lenz C, Schwarz M, Rochow A, Pätzold B, Memmesheimer R, Schreiber M, Behnke S (2023) NimbRo wins ANA avatar XPRIZE immersive telepresence competition: human-centric evaluation and lessons learned. Int J Soc Robot. https://doi.org/10.1007/s12369-023-01050-9

53. Luo R, Wang C, Keil C, Nguyen D, Mayne H, Alt S, Schwarm E, Mendoza E, Padır T, Whitney JP (2023) Team Northeastern's approach to ANA XPRIZE avatar final testing: a holistic approach to telepresence and lessons learned. arXiv preprint arXiv:2303.04932

54. Correia Marques JM, Naughton P, Peng JC, Zhu Y, Nam JS, Kong Q, Zhang X, Penmetcha A, Ji R, Fu N et al (2024) Immersive commodity telepresence with the AVATRINA robot avatar. Int J Soc Robot. https://doi.org/10.1007/s12369-023-01090-1

55. Park S, Kim J, Lee H, Jo M, Gong D, Ju D, Won D, Kim S, Oh J, Jang H et al (2023) A whole-body integrated avatar system: implementation of telepresence with intuitive control and immersive feedback. IEEE Robot Autom Mag 200:300. https://doi.org/10.1109/MRA.2023.3328512

56. Schwartz M, Sim J, Park J (2022) Design and control of a humanoid avatar head with realtime face animation. In: 2022 22nd international conference on control, automation and systems (ICCAS), pp 608–613

57. Sung E, Yu S, Kim S, Park J (2023) SNU-avatar robot hand: dexterous robot hand with prismatic four-bar linkage for versatile daily applications. In: 2023 IEEE/RSJ international conference on humanoid robots. IEEE

58. Kim S, Sung E, Park J (2023) ARC joint: anthropomorphic rolling contact joint with kinematically variable torsional stiffness. IEEE Robot Autom Lett 8(3):1810–1817

59. Lee H, Park G, Shin J, Park B, Park J (2023) Foot-operated telelocomotion interface for avatar robots utilizing Mecanum wheel-based mobile platforms. In: 2023 international conference on control, automation and systems. IEEE

60. Ahn J, Park S, Sim J, Park J (2023) Dual-channel EtherCAT control system for 33-DOF humanoid robot Tocabi. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3272045

61. Koptev M, Figueroa N, Billard A (2021) Real-time self-collision avoidance in joint space for humanoid robots. IEEE Robot Autom Lett 6(2):1240–1247

62. Kim S, Sung E, Park J (2022) 3-finger robotic hand and hand posture mapping algorithm for avatar robot. Korea Robot Soc 17(3):322–333

63. Taheri H, Qiao B, Ghaeminezhad N (2015) Kinematic model of a four Mecanum wheeled mobile robot. Int J Comput Appl 113(3):6–9

64. Shin J, Ahn J, Park S, Park B, Cha J, Park J (2023) Virtual reality based intuitive spatial visual interface for avatar robot system. In: 2023 international conference on control, automation and systems. IEEE

65. Hanif NM, Chappell PH, Cranny A, White NM (2015) Surface texture detection with artificial fingers. In: 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, pp 8018–8021

66. Atapattu S, Senevirathna N, Shan H, Madusanka T, Lalitharatne TD, Chathuranga D (2017) Design and development of a wearable haptic feedback device to recognize textured surfaces: preliminary study. In: 2017 IEEE international conference on advanced intelligent mechatronics (AIM). IEEE, pp 16–21

67. Roboflow I (2022) Roboflow official sites. https://roboflow.com. Accessed 21 Mar 2023

68. Jocher G (2020) YOLOv5 by ultralytics. https://github.com/ultralytics/yolov5. Accessed 21 Mar 2023

69. Park S, Cha J, Park J (2023) Operator-avatar texture feedback approach using hand-eye camera and force sensor. In: 2nd work-

shop toward robot avatars, IEEE international conference on robotics and automation. IEEE

70. Zhu M Recall, precision and average precision. Department of Statistics and Actuarial Science, University of Waterloo, Waterloo (2004)

71. Park B, Park J (2024) Intuitive and interactive robotic avatar system for tele-existence: team SNU—experiment. https://youtu.be/GmxGxU4VXqg. Accessed 18 Jan 2024

72. Schwarz M, Lenz C, Memmesheimer R, Pätzold B, Rochow A, Schreiber, Behnke S (2023) Robust immersive telepresence and mobile telemanipulation: NimbRo wins ANA avatar XPRIZE finals. In: 2023 IEEE-RAS 22nd international conference on humanoid robots (Humanoids). IEEE, pp 1–8

73. Pollen roboticcs. https://www.pollen-robotics.com/reachy/. Accessed 22 Mar 2024

74. van Bruggen J, Brekelmans C, Lieftink R, Dresscher D, van Erp J (2023) I-Botics avatar system: towards robotic embodiment. In: 2nd workshop toward robot avatars, IEEE international conference on robotics and automation. IEEE

75. Inbiodroid. https://inbiodroid.com/. Accessed 22 Mar 2024

76. Pätzold B, Rochow A, Schreiber M, Memmesheimer R, Lenz C, Schwarz M, Behnke S (2023) Audio-based roughness sensing and tactile feedback for haptic perception in telepresence. arXiv preprint arXiv:2303.07186

77. Dafarra S, Pattacini U, Romualdi G, Rapetti L, Grieco R, Darvish K, Milani G, Valli E, Sorrentino P, Viceconte PM et al (2024) iCub3 avatar system: enabling remote fully immersive embodiment of humanoid robots. Sci Robot 9:eadh3834

78. Zambella G, Grioli G, Barbarossa M, Cavaliere A, Lentini G, Petrocelli C, Poggiani M, Rosato G, Sessa E, Tincani V, Bicchi A, Catalano MG (2023) Alter-Ego X: a soft humanoid robot for the ANA Avatar XPRIZE. In: 2nd workshop toward robot avatars, IEEE international conference on robotics and automation. IEEE

79. Vaz JC, Dave A, Kassai N, Kosanovic N, Oh PY (2022) Immersive auditory-visual real-time avatar system of ANA avatar XPRIZE finalist Avatar-Hubo. In: 2022 IEEE international conference on advanced robotics and its social impacts (ARSO). IEEE, pp 1–6

80. Haruna M, Ogino M, Tagashira S, Kashiwa M, Morita S, Koike-Akino T, Imai K, Zuho T, Makita M, Takahashi Y (2023) Avatar technologies of team last mile toward mobile smart device operation service. In: 2nd workshop toward robot avatars, IEEE international conference on robotics and automation. IEEE

81. Dafarra S, Darvish K, Grieco R, Milani G, Pattacini U, Rapetti L, Romualdi G, Salvi M, Scalzo A, Sorrentino I et al (2022) iCub3 avatar system. arXiv preprint arXiv:2203.06972

82. Cisneros R, Benallegue M, Kaneko K, Kaminaga H, Caron G, Tanguy A, Singh R, Sun L, Dallard A, Fournier C et al (2022) Team JANUS humanoid avatar: a cybernetic avatar to embody human telepresence. In: Toward robot avatars: perspectives on the ANA Avatar XPRIZE competition, RSS workshop

83. Cisneros-Limón R, Dallard A, Benallegue M, Kaneko K, Kaminaga H, Gergondet P, Tanguy A, Singh RP, Sun L, Chen Y et al (2024) A cybernetic avatar system to embody human telepresence for connectivity, exploration, and skill transfer. Int J Soc Robot. https://doi.org/10.1007/s12369-023-01096-9

84. Schwarz M, Lenz C, Rochow A, Schreiber M, Behnke S (2021) Nimbro avatar: interactive immersive telepresence with force-feedback telemanipulation. In: 2021 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 5312–5319