TIUI: Touching Live Video for Telepresence Operation

Article *in* IEEE Transactions on Mobile Computing · September 2021

DOI: 10.1109/TMC.2021.3112559

4 authors:

Yunde Jia
Beijing Institute of Technology
284 PUBLICATIONS **4,452** CITATIONS
SEE PROFILE

Yanmei Dong
Beijing Institute of Technology
9 PUBLICATIONS **15** CITATIONS
SEE PROFILE

Bin Xu
Beijing Institute of Technology
7 PUBLICATIONS **28** CITATIONS
SEE PROFILE

Che Sun
Beijing Institute of Technology
6 PUBLICATIONS **27** CITATIONS
SEE PROFILE

TIUI: Touching Live Video for Telepresence Operation

Yunde Jia, *Member, IEEE, Yanmei Dong, Bin Xu, and Che Sun*

Abstract This paper presents a framework for telepresence operation by touching live video on a touchscreen. Our goal is to enable users to use a smartpad to teleoperate everyday objects by touching the objects' live video they are watching. To this end, we coined the term "teleinteractive device" to describe such an object with an identity, an actuator, and a communication network. We developed a touchable live video image-based user interface (TIUI) that empowers users to teleoperate any teleinteractive device by touching its live video with touchscreen gestures. The TIUI contains four modules — touch, control, recognition, and knowledge — to perform live video understanding, communication, and control for telepresence operation. We implemented a telepresence operation system that consists of a telepresence robot and teleinteractive devices at a local site, a smartpad with the TIUI at a remote site, and communication networks connecting the two sites. We demonstrated potential applications of the system in remotely controlling telepresence robots, opening doors with access control panels, and pushing power wheelchairs. We conducted user studies to show the effectiveness of the proposed framework.

Index Terms—Telepresence operation, user interface, touching live video, touchscreen gesture, mobile robotic telepresence

1 Introduction

As mobile smart computing and communication networking are becoming pervasive, we can easily use smart mobile devices (e.g., smartpads and smartphones) to video chat with family members, video monitor houses, and watch live shows [1] [2]. It would be natural for people to expect to touch and operate objects in the live video they are watching. For example, when children are watching the live video of a toy car, they may curiously touch and drag the toy on the screen. What is disappointing is that they will not get the expected response, such as moving, stopping, or turning. Another example is a class of mobile robotic telepresence systems, such as Beam [3] and Ohmni [4], which are basically mobile video-conferencing systems. A mobile robotic telepresence system allows remote users to teleoperate a telepresence robot to video chat with local persons with some degree of mobility. But most existing systems do not allow remote users to teleoperate local objects around them, such as opening doors or turning on/off power, so remote users have had to request local help, and they have described this experience as "feeling disabled" [5]. Therefore, there is a great need to enable remote users to teleoperate local objects they are watching on live video, while also teleoperating telepresence robots [6].

The purpose of this paper is to explore telepresence operation in which users can teleoperate everyday objects by touching the live video they are watching. Figure 1 illustrates an example of teleoperating an object by touching its live video. A user at a remote site (Figure 1a) uses a smartpad to watch the live video of an access control panel, and taps the image of the number keys on the panel to enter the password to remotely open the door at a local site (Figure 1b). This process is very similar to directly tapping the physical number keys at the site. The object in this example is an access control system, and it can be any other everyday object, as long as the object has three elements: an identity, an actuator, and a communication network. We coined the term "teleinteractive device" to refer to such an object with these three elements. For convenience, we often use teleinteractive device and object interchangeably from now on. Using touchscreen and computer-vision techniques, we developed a touchable live video image-based user interface, called TIUI, which empowers users to teleoperate any teleinteractive device by touching its live video with touchscreen gestures. A touchable live video refers to a live video containing teleinteractive devices, which can be touched to perform teleoperation. In contrast to traditional graphical user interfaces (GUIs), the TIUI only contains a touchable live video and looks like a plain video window, and there are no explicit graphical buttons and menus.

We implemented a telepresence operation system that is capable of not only teleoperating a telepresence robot for video-mediated communication with local persons, but also teleoperating teleinteractive devices. We specifically use a telepresence robot because it gives remote users mobile telepresence experience and also enables the remote users to actively acquire live video by using on-board cameras to avoid occlusion and unclear images. The TIUI facilitates remote users to teleoperate a telepresence robot for exploring any place and control teleinteractive devices without requesting local help. An earlier version of this study was published in [7].

The remainder of the paper is organized as follows. Section 2 surveys related work. Section 3 presents the methodology of telepresence operation. Section 4 describes the system architecture, as well as the hardware and software components. Section 5 introduces touchscreen gestures for the TIUI. Section 6 demonstrates examples of potential applications. Section 7 reports the evaluation of the proposed

- *Yunde Jia, Yanmei Dong, Bin Xu and Che Sun are with the School of Computer Science at Beijing Institute of Technology and Beijing Laboratory of Intelligence and Technology, Beijing 100081, PR China. Email: {jiayunde, dongyanmei, xubin47, sunche}@bit.edu.cn*

(a) Remote site (b) Local site

Fig. 1. An example of performing telepresence operation tasks with our system. (a) A user at a remote site is using a smartpad with the TIUI to press the keys on the panel in a live video to open a door, which is very similar to pressing the physical keys in person. (b) A telepresence robot at a local site is capturing the live video of a password access control panel that the user is watching, and the top right corner is an enlarged view of the password access control panel.

framework. Section 8 discusses some issues, and we conclude this work in Section 9.

2. Related Work

There are many video-mediated teleoperation scenarios such as vehicle teleoperation [8], seeing like a planetary rover [9], emotive gloves over distance [10], and teleoperated surgical robot teamwork [11]. As our work focuses on telepresence operation by touching live video via a telepresence robot, we only review live video based teleoperation and mobile robotic telepresence.

2.1 Teleoperation through Live Video

An early attempt to teleoperate physical objects through live video is the work of Tani et al. in 1992 [12]. They used a common monitor-mouse-keyboard interface to manipulate the live video of real-world objects for remotely controlling an electric power plant, including clicking button images for controlling and dragging the 2D or 3D model of a physical object for positioning. Our study extends this work and aims to present a framework for telepresence operation of everyday objects by touching live video on a touchscreen with touchscreen gestures. Hashimoto et al. [13] presented a touchscreen-based interaction system that can remotely control a multi-degree-of-freedom robot. They touched the 3D computer graphic (CG) model of the robot overlaid on a live video from a ceiling camera to perform 3D motion, and then the robot moved to match this model to achieve precise positioning. Kasahara et al. [14] used a smartpad as a first-person view to control an actuated object through an augmented reality-mediated mobile interface with the 3D CG model of the object. These techniques focus on remotely controlling specific objects, especially matching the model of an object based on its live video image to achieve precise positioning. Our work focuses on a framework of telepresence operation, which allows us to teleoperate everyday objects by touching live video. We believe that matching the model of an object based on its image for positioning mentioned above can be integrated into our framework.

Some work used a ceiling camera as a third-person view to explore live video-based interactions by touching live video. Seifried et al. [15] developed a video-based user interface on an interactive table surface for controlling media devices in a living room. Guo et al. [16] designed two user interfaces, touch (touchscreen) and toy (tangible), for interaction between a single user and a group of robots. There are also interaction interfaces on a smartpad using the live video from a ceiling camera to support sketching the expected path for an indoor robot [17] [18] [19]. Kato et al. [20] used a smartpad to control multiple mobile robots simultaneously by manipulating a vector field on the live video. These methods were designed to remotely control objects located at the same site as users, and the users could obtain visual feedback by directly seeing the objects. Unlike these methods, our work aims to teleoperate objects at a distance, and users can only see the objects through live video and teleoperate these objects by touching their live video.

Touchable live video is one of the most important concepts in our work. Kim and Park [21] introduced a touchable video stream that is generated by rendering haptic information onto an RGB-D video stream in mixed/augmented reality. Sung et al. [22] reported a touchable video/audio device in which haptic feedback data is combined with a traditional audiovisual stream to improve user visual/haptic immersiveness in the 3D virtual environment. The touchable videos above are video streams generated in a virtual environment through the fusion of a real video and haptic information. Different from these touchable videos, the touchable live video we define is a purely live video of a real environment, but simply contains teleinteractive devices that can be teleoperated by touching their live video images.

2.2 Mobile Robotic Telepresence

Paulos and Canny (1998) developed the first telepresence robot [23] that includes a mobile robot base attached with

a human-height pole and audio/video communication devices mounted on the pole, which is the basic configuration of telepresence robots and is still popular. Since then, many telepresence robots with the basic architecture have emerged. Readers can refer to a review of the literature [24]. So far, there has been a wide range of applications related to mobile robotic telepresence, such as health care [25], education [26], workplaces [27], and public places [28] [29]. These systems allow remote users to operate telepresence robots through computers/laptops by using traditional user interfaces, including keyboards, mice and joysticks. There are some mobile robotic telepresence systems [3] [4] [30] that allow users to use a smartpad as a navigation controller to teleoperate a telepresence robot. In these systems, the user interfaces were designed by simply converting physical keyboards and/or joysticks to graphical buttons [31] [32]. Different from the existing systems, the user interface in our system allows users to teleoperate a telepresence robot by touching live video instead of pressing graphical buttons or icons. In addition, most existing systems do not allow remote users to teleoperate local objects around them, so the remote users have to request local help. Our system allows remote users to teleoperate everyday objects at a local site by touching their live video image, which can alleviate the need for local help. It should be noted that although some mobile robotic telepresence systems allow users to touch the live video of a docking station for automatic docking or recharging of a telepresence robot (e.g., Beam [3]), the touch action there is only an instruction for users to guide the robot to the docking station, not for the users to teleoperate the docking station. In fact, those docking stations are robot accessories, not teleinteractive devices. It is easy to design a docking station as a teleinteractive device, so that remote users can directly touch its live video to teleoperate it for docking.

There are already many teleoperation robots that require little or no local help because they are equipped with robotic manipulators. Typical examples are anthropomorphic or humanoid robots, like TELESAR [33], Robonaut [34], HRP [35], and Rollin' Justin [36]. Readers can refer to the latest review [37] for detailed teleoperation systems with robotic manipulators. These robots usually have exquisite anthropomorphic structures with a higher level of dexterity and can mimic human actions to physically contact objects to complete teleoperation tasks, such as pushing, pulling, screwing, etc. They require dedicated hardware and well-trained operators, and are usually used in hazardous exploration and other situations where it is difficult for humans to be present. Unlike these systems, our system enables ordinary users to use a smartpad to teleoperate everyday objects by touching the objects' live video they are watching, without the need for robots with manipulators to physically contacting the objects.

3. METHODS

The basic concepts of our work are teleinteractive device, touchable live video, touchscreen gesture, TIUI, and telepresence operation. In this section, we describe the methods of telepresence operation by touching live video based on the concept-driven design principle proposed by Stolterman & Wiberg [38].

3.1 Teleinteractive Devices

A teleinteractive device has three elements: an identity, an actuator, and a communication network, as mentioned above. The identity refers to the name or designation of a device with an IP address, which ensures that a remote user can accurately teleoperate the device. The actuator refers to a component that can be powered to perform a specific state change of the device, such as switching, moving, vibrating, displaying, etc. The communication network of a teleinteractive device is used to connect the device to the Internet in a wired or wireless way.

Since expounded by Mark Weiser in his landmark article [39], more everyday objects are becoming smart by embedding processors, sensors, and actuators, and connecting to the Internet. A smart object is an autonomous object with sensing, processing, and networking capabilities [40]. Smart objects (or devices) are a class of teleinteractive devices, but many teleinteractive devices may not be smart devices because they can neither sense nor have autonomy. Generally speaking, teleinteractive devices are much simpler and cheaper than smart devices. There are a number of teleinteractive devices, such as WiFi toys and WiFi appliances. A simple teleinteractive device is a WiFi power switch or button [41].

3.2 Touchable Live Video

We often watch live video, but if we touch the live video, we won't get any response. If the live video of a teleinteractive device is displayed on a touchscreen, we can establish the relationship between a teleinteractive device and its live video by using object recognition algorithms, and then we can remotely control this device by touching its live video. We call the live video of teleinteractive devices touchable live video.

Object recognition is a prerequisite for achieving teleoperation by touching live video. Object recognition is one of the most important tasks in computer vision, which has rapidly developed in recent decades. One can easily access computer vision resources (e.g., OpenCV [42]). However, most existing computer vision algorithms are limited in terms of reliability and robustness in applications. In most cases, people prefer to use a 2D barcode to identify an object, and use computer vision algorithms to lock and track the appearance of this object.

3.3 Touchscreen Gestures

Touchscreen gestures have become a primary way to interact with smartpads and smartphones. A touchscreen gesture used to touch live video for teleoperation not only maps the user's intention to the corresponding live video of an object, but also extends the mapping to teleoperate the object.

The new gestures normally require minimal learning, as Blackler et al. [43] argued that intuitive interaction should be contingent upon the user's prior experience and familiarity with technology. Operating an object is usually related to the perceived affordance of the object [44], which has been commonly used in interaction design. When we

see the image of a door on a touchscreen, for example, we touch the knob image on the door or press the button image on the side of the door, and the door should open accordingly. Another example is to drag a volume slider in a live video to change the speaker's volume. Therefore, human actions of operating objects can be converted into touchscreen gestures. We can define more touchscreen gestures inspired by the human actions of manipulating physical objects. In this work, we define a set of touchscreen gestures with which users can touch live video of an object to teleoperate it. Touchscreen gestures will be described in detail in Section 5.

3.4 TIUI

In order to enable users to teleoperate objects by touching live video, we developed a touchable live video image-based user interface (TIUI). The TIUI addresses the following four issues:

(1) *What is the intention of a user's touchscreen gesture?*
(2) *How does the system perform remote control?*
(3) *Which object in a live video is being touched?*
(4) *How does a user embed knowledge into the system by touching live video to facilitate teleoperation?*

To this end, we introduce four corresponding modules: touch, control, recognition, and knowledge. The touch module is used to detect and interpret gestures. The control module is used to transform the gestures into control instructions for teleoperation. The recognition module is used to identify an object being touched, and to lock and track the object. Thus, the teleoperation relationship between the user and the object being touched is established and maintained until another object is touched. The task of the knowledge module is to embed knowledge into the system to improve the performance of teleoperation. The TIUI first executes the touch module to detect and interpret the user's gestures, and then executes the control, recognition, or knowledge modules to perform corresponding teleoperations based on the gestures. The explanation of the modules and their relationships is given in Section 4.3.

3.5 Telepresence Operation

Telepresence operation is a combination of video-mediated communication and video-mediated teleoperation. Video-mediated teleoperation refers to not only using live video as visual feedback but also touching live video to remotely control a telepresence robot or teleinteractive devices. Note that telepresence operation allows both remote users and local users to see each other through live video, which is different from teleoperation. Using mobile smart computing and networking technologies, we can easily realize telepresence operation by touching live video.

Telepresence operation is characterized by the following three dimensions:

Intuitive. *Users can intuitively and naturally teleoperate an object by touching the object's live video they are watching, which is similar to the action of operating the physical object.*
Flexible. *Users can flexibly teleoperate any object in the touchable live video, as long as the object is identified by the recognition module. The system also has good scalability.*
Mobile. *Users can use a mobile smart device (smartpad) with the TIUI to connect to the Internet anywhere for both video-mediated communication and video-mediated teleoperation.*

The key to realizing telepresence operation is to integrate the interpretation of touchscreen gestures and the recognition of touchable live video images to design the TIUI. We developed a telepresence operation system to evaluate the proposed methods and demonstrate the potential applications.

4. System Overview

In this section, we describe the telepresence operation system, including system architecture, hardware and software components.

4.1 System Architecture

Figure 2 shows the system architecture. The system consists of a telepresence robot and teleinteractive devices at a local site, a mobile smart device (e.g., a smartpad) with the

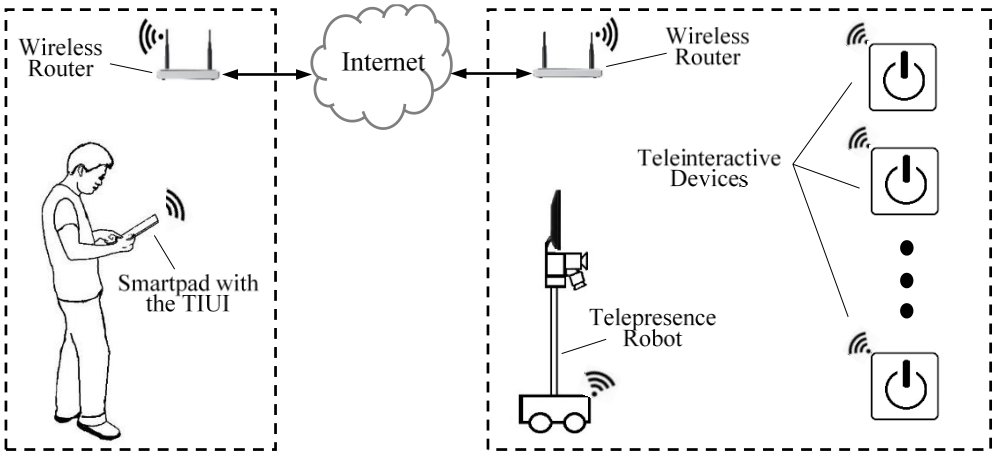


Fig. 2. System architecture. The system consists of a telepresence robot and teleinteractive devices at a local site (dashed rectangle on the right), a smartpad with the TIUI used by a user at a remote site (dashed rectangle on the left), and communication networks connecting the two sites. We use a power button symbol plus WiFi symbol to represent a teleinteractive device (far right), and use thin lines to indicate the names of the devices, and double arrow lines to indicate wired network connections.

TIUI at a remote site, and communication networks connecting the two sites. Users can use a smartpad with the TIUI at a remote site to teleoperate a telepresence robot at a local site by touching live video with touchscreen gestures and chat with local persons. Users can also touch the live video of a teleinteractive device they are watching to make the TIUI connect to the device via networks, so that the users can teleoperate this device.

A telepresence robot is a typical teleinteractive device because it has the three elements of a teleinteractive device. But in view of the special role of a telepresence robot in our system, it is not included in the category of teleinteractive devices for simplicity of presentation: (1) a telepresence robot is regarded as the embodiment of remote users to interact with local persons, and (2) remote users can teleoperate the telepresence robot to actively capture the live video of the local site so that they can be better aware of the local site. We use a power switch symbol plus a WiFi symbol to indicate a teleinteractive device (on the far right in Figure 2), considering that most teleoperations involve simply pressing buttons.

In this work, live videos are captured by two cameras installed on the pan-tilt on the vertical post of the telepresence robot. The configuration of fixed overhead cameras with or without pan-tilt-zoom (PTZ) commonly used in surveillance and monitoring systems can be regarded as a special case of our system.

4.2 Hardware Components

Teleinteractive devices and a telepresence robot are the main members of our system. We set up a smart room in our lab as a testbed, as shown in Figure 3, containing a telepresence robot (labeled A) and some teleinteractive devices, such as a door with password access control (labeled B), a cheerful lit tree (labeled C), a table lamp (labeled D), a power wheelchair (labeled E), and a power curtain (labeled F).

4.2.1 Teleinteractive Box for Teleoperation

In our system, all teleinteractive devices are off-the-shelf products, and each is equipped with a teleinteractive box (similar to set-top box) developed in our lab, making it a teleinteractive device. A teleinteractive box shown in the lower right corner of Figure 3 is added to a power wheelchair to replace the joystick for remotely controlling the wheelchair. A teleinteractive box is basically composed of a WiFi unit for networking, a 2D barcode for identifying a teleinteractive device or an object, and an actuator for changing the state of the object.

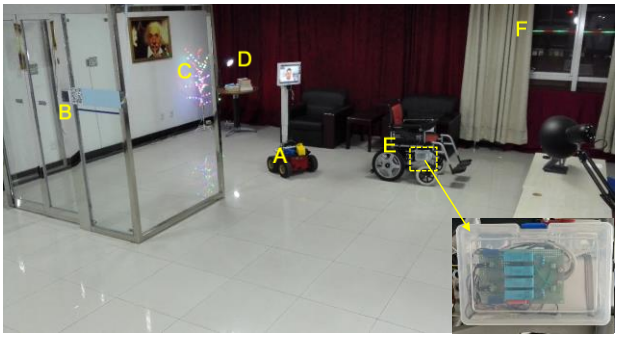


Fig. 3. A picture of a smart room set up in our lab as a testbed, containing a telepresence robot (labeled A), a door with password access control (labeled B), a cheerful lit tree (labeled C), a table lamp (labeled D), a power wheelchair (labeled E), and a curtain (labeled F). The lower right corner is an enlarged view of a teleinteractive box we developed for the wheelchair.

Actuating modes can be roughly divided into three categories: point, line and plane. Point actuation is the output of a specific amount of power to start a device, for example, to drive a relay to perform button functions. Point actuation is common and can achieve all 0-1state transitions, such as turning on/off power and switching channels. Line actuation is the output of varying power to change the state of a device. Typical examples of line actuation include dimming lights, adjusting volume, and opening curtains. Plane actuation is the output of two varying powers (as a 2D vector) to control mobile robots or other vehicles to move in 2D space or on the surface. The conventional plane actuating is performed through a keyboard and joystick-based interface, which can control an object (e.g., a mobile robot and a power wheelchair) to move. We developed a teleinteractive box that can replace a joystick, which enables users to use a smartpad with the TIUI or the GUIs to teleoperate a power wheelchair. The box on the power wheelchair is the most complicated of all, but its cost is only tens of US dollars.

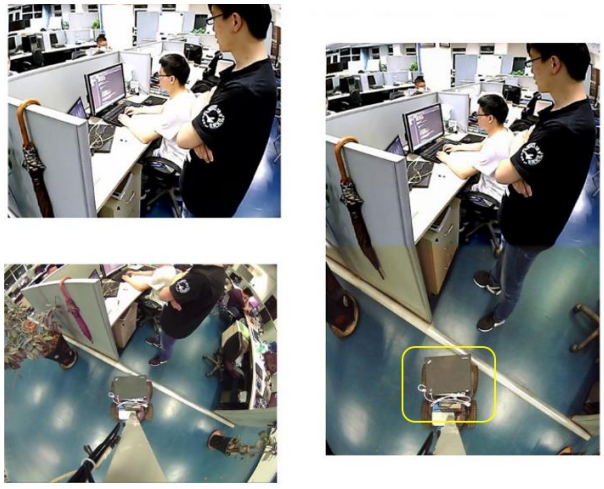
4.2.2 Telepresence Robot

We assembled a telepresence robot, named Mcisbot [7] (labeled A in Figure 3), in our lab as a mobile testing platform for telepresence operation through the TIUI. The Mcisbot is composed of a robot base and a robot head based on the configuration of the telepresence robot [23]. We used a Pioneer 3-AT robot at hand as a robot base and developed a robot head to test certain functions of the system. The head includes an LCD, a speaker, a microphone, a forward-facing camera (FF camera), and a downward-facing camera (DF camera), which are mounted on the pan/tilt together. The FF camera is used for video communication and object recognition, and the DF camera is used for robot navigation. The head is fixed on a vertical post on the robot and can be moved to adjust the robot height from 1200 mm to 1750 mm, covering the height of school-aged children to adults. We also installed the robot on a cheap mobile base to assemble a low-cost telepresence robot [45].

As the embodiment of remote users, a telepresence robot should help remote users better perceive the situation of a local site through the two cameras, the FF camera and the DF camera. We follow the two principles when selecting these two cameras: seeing forward clearly and seeing downward panoramically:

Seeing forward clearly. *The FF camera captures video (FF video) with high resolution for remote users to see objects in front clearly. High-quality FF video facilitates object recognition.*

Seeing downward panoramically. *The DF camera captures video (DF video) with a very large field of view for remote users to look down at the ground from a panoramic view for navigation.*



(a) (b)

Fig. 4. Illustration of a stitched video image. Left top: an image from the FF camera. Left bottom: an image from the DF camera. Right: the stitched image. We can see the entire mobile robot base (yellow rectangle).

According to the two principles, we used a high-resolution camera with a field of view (about 70 degrees diagonal) as the FF camera and a fisheye lens camera (about 170 degrees diagonal) as the DF camera. We adopted our video stitching algorithm [46] to stitch the FF video and DF video into one video, named FDF video, as shown in Figure 4. Through the FDF video, remote users can clearly see objects in front, and can also see the robot base (enclosed by a yellow rectangle in Figure 4b) and a panoramic view of its surroundings. The FDF video makes it easier for remote users to perceive the robot's surroundings, without having to use their brains to integrate the two video together, reducing the burden of teleoperation [47].

4.3 Software Components

The TIUI is the most important software part of our system. As mentioned earlier, the TIUI contains four modules: touch, control, recognition, and knowledge. The following is a description of the four modules and their relationships.

4.3.1 Touch Module

The touch module is used to perform the detection and interpretation of touchscreen gestures. Detection aims to obtain the position data of fingers touching the touchscreen, and interpretation is a mapping from the data to touchscreen gestures. The touch module allows remote users to perform the touchscreen gesture on a live video while watching the live video, which forms a human-in-the-loop control of telepresence operation. Thus, remote users play a dominant role in the loop according to the strategy of human-centered teleoperation [48].

4.3.2 Control module

The control module is used to transform touchscreen gestures into instructions to complete teleoperation tasks. One category of teleinteractive devices shares the same instruction set. Different categories of teleinteractive devices have different instruction sets, for example, all lighting devices share the same lighting instruction set. Once the object being touched in a live video is automatically recognized and locked, the touchscreen gestures can be mapped to the actions of the object with the corresponding instruction. In other words, the control module allows remote users to remotely control the object using touchscreen gestures.

4.3.3 Recognition Module

Most existing remote-control systems, including mobile robotic telepresence systems, adopt a one-to-one direct teleoperation strategy, in which remote users use a controller that directly connects to a specific object to teleoperate it without identification. In smart environments, remote users can use a smartpad with the TIUI that automatically connects to multiple local teleinteractive devices. If users want to remotely control many local teleinteractive devices, i.e., one-to-many teleoperation, the first task they would do is to recognize which object is being touched in a live video. The recognition task includes object detection, localization, identification, and tracking. In our system, the FF camera mounted on a telepresence robot captures the live video of objects in front of the robot, and then the recognition module is used to recognize the object being touched. As a result, the teleoperation relationship between the user (via the TIUI) and the object is established.

In this work, we also use a 2D barcode to identify an object, and then use computer vision algorithms to detect, localize, and track the object. If a tracked object is lost, the system immediately returns to the touch module to wait for the user's gesture to re-identify the object. We can also use the recognition module to perform path planning and obstacle detection for navigation [45].

4.3.4 Knowledge Module

Since knowledge acquisition involves complex cognitive processes (perception, communication, and reasoning), it is a major bottleneck in AI systems [49]. The task of the knowledge module is to embed knowledge into the system to improve the performance of teleoperation. Users can place markers on the live video of a local environment through the TIUI, such as marking obstacles, doors, paths, and other objects of interest. A marked object can be automatically locked using the recognition module. The knowledge module can also help users embed rich information into a live video, such as spatial relationship and semantic description.

4.3.5 Relationship between Modules

Motivated by the autonomous robot architecture proposed by Rodney Brooks [50], we adopt a layered structure to describe the relationship among the touch, control, recognition, and knowledge modules, as shown in Figure 5. Touch and control modules form the first layer, touch and recognition modules form the second layer, and recognition and knowledge modules form the third layer. The first layer only contains the touch and control modules, which is a conventional direct teleoperation strategy (i.e., one-to-one control strategy) designed for teleoperation of a specific object. In this work, since there is only one telepresence robot at a local site, we use a smartpad to

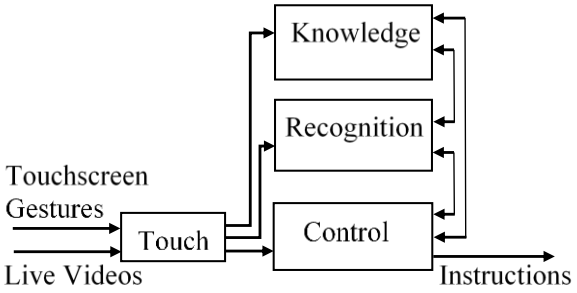


Fig. 5. Software scheme of the TIUI. The TIUI contains the four modules: touch, control, recognition, and knowledge. The touch module interprets touchscreen gestures based on live videos to decide which module to switch to. The control module calls instructions under guidance of the recognition and knowledge modules to realize the network connection between the TIUI and local devices and teleoperation of the local devices.

connect directly to the robot through the communication network, and then use the touch and control modules to perform robotic teleoperation without the recognition and knowledge modules. However, when the telepresence robot is in an autonomous or semi-autonomous state (e.g., the telepresence robot is in the state of following as mentioned in Section 6.3), the recognition and knowledge modules are still needed. On the basis of the first layer, a second layer is added to expand the one-to-one control in the first layer to one-to-many control. Knowledge as the third layer of addition can improve the performance of the control and recognition modules. It is obvious that the recognition and knowledge modules can improve the performance of each other. The control, recognition, and knowledge modules can be combined for various intelligent systems, such as autonomous robot systems, intelligent interaction systems, and intelligent surveillance systems. Our system falls into the category of intelligent interaction systems.

5. Touchscreen Gestures

The touchscreen gestures are designed to perform the following three types of teleoperations:

(1) *Users touch the live video of an object with the touchscreen gestures to teleoperate the object.*
(2) *Users touch the live video of a telepresence robot with the touchscreen gestures to teleoperate the telepresence robot.*
(3) *Users touch the live video of a local environment with the touchscreen gestures to place markers on the live video.*

The first two are the basic types of telepresence operation. The third is an enhancement by embedding user knowledge (such as passages and obstacles) into the system. A typical example of marking is marking obstacles or road edges for navigation. In this paper, we mainly define touchscreen gestures for the first two types. Based on the basic touchscreen gestures [51], we define two types of gestures: one-finger gestures and two-finger gestures.

5.1 One-Finger Gestures

In daily lives, we often use one finger to perform most operations on various panels and interfaces of everyday objects or devices, because they usually contain switches, buttons and/or sliders. For a joystick, it can be seen as a combination of multiple buttons and arrow keys. In fact, most touchscreen gestures for interacting with smartpads and smartphones are performed with one finger. Therefore, we can choose one-finger gestures to touch live video to perform teleoperation of teleinteractive devices.

Figure 6 shows the four one-finger gestures: tap, press, drag, and lasso. When a user taps an object or object's 2D barcode in a live video, the TIUI will recognize the object to obtain its ID, so that the TIUI connects to the object through networks. A marker (e.g., a colored spot or rectangle) is automatically marked on the live video image of the object connected to the TIUI. A press gesture refers to a long press on a touchscreen, which can be seen as an extension of the tap gesture. One can use a press gesture to remotely control the stop or pause of an object. For example, when a user presses a live video of a moving object, the TIUI will recognize the object and connect to it through networks, and as a result, the object will stop moving until this press gesture ends. A lasso gesture can lasso a certain area in a live video, which can alleviate the challenges in image segmentation and improve the performance of object recognition. A drag gesture can make a marked object move with the movement of the finger. When a user drags a marked object in a live video, the corresponding physical object will move with the user's gestures. For example, when a chess player drags a live video image of huge physical chess piece on a touchscreen, the chess piece (with powered wheels) on the field for the audience to watch will move accordingly.

Two or more one-finger gestures can be combined to perform combined teleoperation actions. For example, one finger performs a pressing gesture on a live video of a moving toy car to stop the car, and the other finger makes a tapping gesture to open the door of the car.

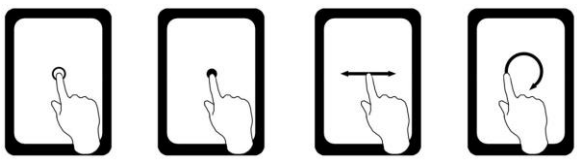


Fig. 6. One-finger gestures for teleoperating teleinteractive devices.

5.2 Two-Finger Gestures

Two-finger gestures are specifically designed for teleoperation of a telepresence robot because two fingers can perform more complex gestures than one finger. Figures 7 and 8 depict two-finger gestures and corresponding action diagrams of the robot base and robot head. Note that the directions of the two-finger gestures and the robot actions are opposite. This is because the live video is captured by a telepresence robot, and remote users watch the live video of the robot in a first-person camera view. In contrast, remote users are in a third-person camera view when watching the live video of teleinteractive devices, and the remote users' gestures and the actions of teleinteractive devices must be in the same direction. Using two-finger gestures to touch a live video to teleoperate a telepresence robot is similar to the process of browsing a panoramic image with

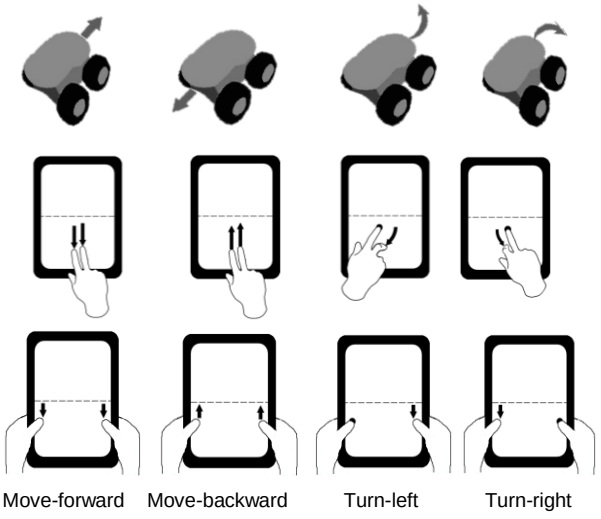


Fig. 7. Two-finger gestures for teleoperation of a robot base. Bottom: gestures made with two fingers of two hands. The dashed line indicates a dividing line between the FF image and the DF image, which may not be displayed on the TIUI. Middle: gestures made with two fingers of one hand. Top: action diagrams of the robot base corresponding to the gestures.

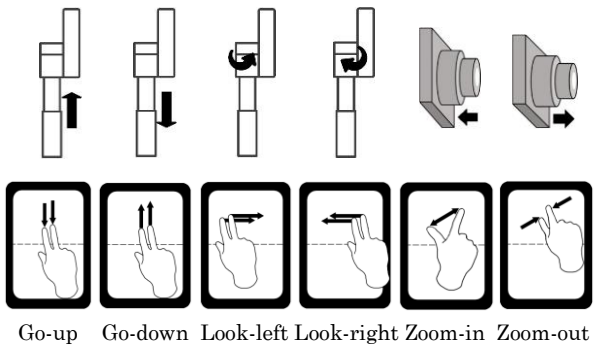


Fig. 8. Two-finger gestures for teleoperation of a robot head with a PTZ camera. The top is the action diagrams of the robot head corresponding to the gestures in the bottom.

gestures on a smartpad, which is natural and intuitive. In our system, these two-finger gestures are translated into the instructions to remotely control the robot base and robot head to perform corresponding actions.

In order to distinguish whether a two-finger gesture controls the robot base or the robot head, we make full use of the feature that an FDF video in the TIUI consists of an upper zone and a lower zone. The dashed line shown in Figures 7 and 8 is a dividing line between the upper and lower zones, and this line may not be displayed on the TIUI. The upper zone corresponds to the FF video of objects to be teleoperated, and the lower zone corresponds to the DF video of the robot base, the ground, and the robot's surroundings for navigation (Figure 4b). Thus, when users touch the upper zone, they can remotely control the robot head to see the objects clearly; when touching the lower zone, they can remotely control the robot base to move around.

The design of two-finger gestures for controlling a robot base is inspired by the observation of human stroke actions in boating and skating. The gestures are similar to human skating strokes. The speed of movement depends on how far the finger slides on the touchscreen or how often the finger touches the touchscreen. Some users prefer to use two fingers of one hand, while others prefer to use two fingers of two hands, as shown in Figure 7.

Two or more two-finger gestures and/or one-finger gestures can be combined to perform more complex head actions. For example, the robot head shaking or nodding gesture can be seen as a combination of several head gestures.

6. Application Examples

We demonstrate three example applications: teleoperation of a telepresence robot, telepresence operation of everyday objects, and telepresence operation of a power wheelchair.

6.1 Teleoperation of a Telepresence Robot

We demonstrate how users use the TIUI on a smartpad to teleoperate the telepresence robot (Mcisbot) with two-finger gestures. Figure 9 shows that a remote user touches the

Fig. 9. Two-finger gestures performed at the lower zone of the TIUI to teleoperate the robot base. Bottom: the TIUI used by a user at a remote site. Top: corresponding positions of the robot at a local site. (a) Reference position. (b) Moving forward. (c) Turning left. (d) Turning right.

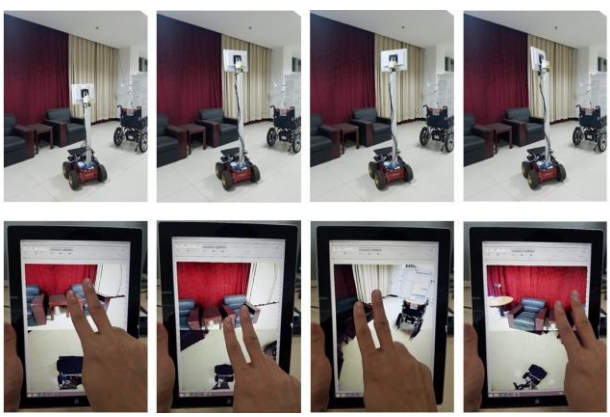


Fig. 10. Two-finger gestures performed on the upper zone of the TIUI to teleoperate the robot head. Bottom: a user is performing two-finger gestures on the TIUI. Top: corresponding actions of the robot. (a) Two fingers touching the TIUI while the post is at a height of 1200 mm. (b) Sliding two fingers down from the position in (a) to raise the post to a height of 1750 mm. (c) Sliding two fingers to the left to look to the right. (d) Sliding two fingers to the right to look to the left.

lower zone of the live video with two-finger gestures to teleoperate the robot base to move forward into the room (Figure 9b), turn left to see the portrait of Albert Einstein (Figure 9c), and then turn right to see the depth of the room (Figure 9d). The top row in Figure 9 shows the corresponding actions of the robot.

Figure 10 shows that a user touches the upper zone of the live video with two-finger gestures to teleoperate the robot head. When a user slides the two fingers down from the position in Figure 10a, the vertical post will rise to a height of 1750 mm (Figure 10b). A user slides the two fingers to the left to view the image on the right (look to the right, Figure 10c) and slides the two fingers to the right to view the image on the left (look to the left, Figure 10d).

6.2 Telepresence Operation of Everyday Objects

Our system allows remote users to teleoperate teleinteractive devices in the smart room (Figure 3) by touching live video with one-figure gestures. Figure 11 shows that a user uses the TIUI to teleoperate the telepresence robot to the door and recognize the password access control panel, and then taps the password to open the door. This process is similar to how the user taps the keys on the physical panel to open the door on site. The top views of Figures 11a and 11d display the closed door before teleoperation and the opened door after teleoperation, respectively. This means that our system allows a remote user to teleoperate everyday button-controlled objects by touching live video.

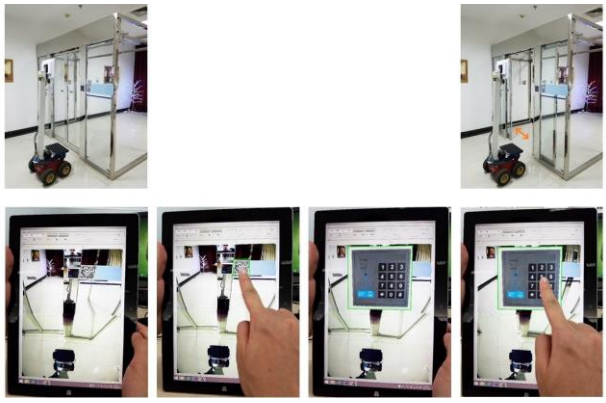
Another example of the telepresence operation of objects is to remotely open a curtain, as shown in Figure 12. A user teleoperates the telepresence robot to look for the curtain (Figure 12a) in the live video, and then lassos the curtain image via the lasso gesture for image segmentation (red circle in Figure 12b). The segmentation result guides the TIUI to recognize and locate the curtain (green rectangle in Figure 12c). Then the user remotely controls the opening of the curtain by the drag gesture (Figure 12d), similar to the user opening the curtain manually on site. This example demonstrates the effect of using a drag gesture to teleoperate objects with line actuating.

6.3 Teleoperation of a Power Wheelchair

Telepresence robots can be used in elderly care and healthcare to help prevent problems related to loneliness and social isolation [52]. Elderly people usually regard telepresence robots as representatives of family members or caregivers [53]. Commercially available telepresence robots, such as Giraff [54] and Ohmni [4], are designed for elderly and disabled people. However, these robots barely address the teleoperation of local objects. If a user is remotely controlling a telepresence robot in the elderly person's home to chat with an elderly person, and the elderly person wants the user to push the wheelchair for a walk, the user has to go to the site in person to meet the need of the elderly person, as illustrated in Figure 13a.

Our system allows a remote user to teleoperate a telepresence robot, not only for video chatting with local elderly persons, but also for remotely "pushing" a power wheelchair, as shown in Figure 13b. We put quotation marks on "pushing" because the telepresence robot has no physical contact with the wheelchair. It is not really pushing the wheelchair. In this application, we converted the physical joystick into a soft joystick overlaid on the live video of the wheelchair (wheelchair backrest). Using the TIUI, a user touches the wheelchair image with the soft joystick gesture (one-figure gesture) to teleoperate it, similar to using the physical joystick to teleoperate the wheelchair. Once a user teleoperate the wheelchair, the telepresence robot is in an autonomous state of following, providing the TIUI with the live video of the wheelchair and its surrounding environment.

Figure 14 shows how a user pushes a wheelchair remotely using a telepresence robot. We added a teleinteractive box to the power wheelchair, which cost about 400 US dollars in total. First, a remote user uses the TIUI to teleoperate the robot to the back of the wheelchair (Figure 14a). Then, the user touches the live video image of a 2D barcode on the backrest of the wheelchair with a one-finger gesture



(a) (b) (c) (d)

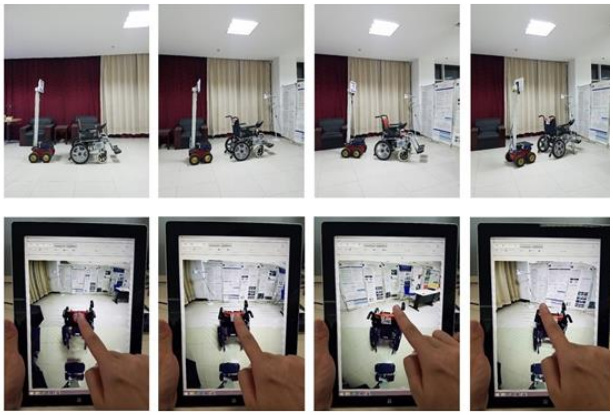
Fig. 11. Using the TIUI to remotely open a door with password access control. (a) A user is looking for the password panel in the live video (bottom), and the door is closed (top). (b) The user lassos the 2D barcode for identification (bounding box on the bar). (c) The panel is identified (bounding box on the panel). (d) The user taps the key images to enter the password to open the door (bottom), and accordingly, the door opens (orange double arrow on the top).

(a) (b) (c) (d)

Fig. 12. Using the TIUI to remotely open the curtain. (a) A user drives the telepresence robot to the curtain (bottom), and the curtain is closed (top). (b) The user lassos the curtain image (red circle) for image segmentation. (c) The TIUI is guided by the segmentation result to recognize and locate the curtain (green rectangle). (d) The user teleoperates the curtain using a drag gesture (bottom), and the curtain is opened (top).

(a) (b)

Fig. 13. Pushing a wheelchair. (a) A local user is pushing a wheelchair. (b) A remote user is remotely "pushing" the same wheelchair via a telepresence robot.



(a) (b) (c) (d)

Fig. 14. A user uses the TIUI to remotely "push" a wheelchair with the soft joystick gestures. Bottom: (a) the wheelchair image is marked with a color square, and the wheelchair is ready to be pushed, (b) the user touches the image of the wheelchair with the soft joystick gestures to remotely push the wheelchair forward, (c) to the right, and (d) to the left. Top: the corresponding views of the telepresence robot and the wheelchair.

to recognize and locate it. The identified wheelchair image is marked, which means that it is ready to be pushed. Third, the user touches the wheelchair backrest with the soft joystick gestures to remotely push the wheelchair forward (Figure 14b), to the right (Figure 14c), and to the left (Figure 14d). The speed of the wheelchair can reach 1.5meters per second.

7. Evaluations

We conducted laboratory experiments to evaluate the proposed framework. The experiments include two tasks: teleoperating a telepresence robot along the ∞-shaped course and teleoperating everyday objects.

7.1 Teleoperating a Telepresence Robot

In order to compare the TIUI and the GUIs, we designed a conventional touchscreen GUI overlaid on a live video, as shown on the far right of Figure 15a. Touchscreen GUIs usually contain graphical buttons and arrow keys, which are converted simply from mouse/keyboard and joystick.

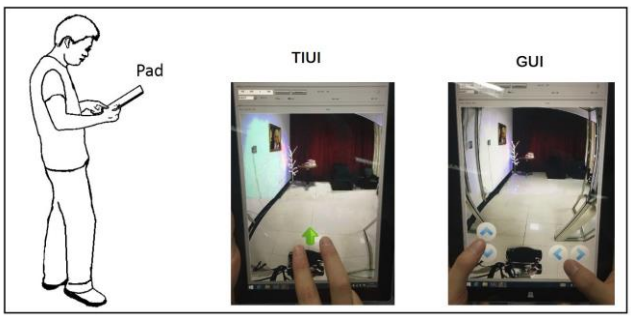
We recruited 20 people from a local university for the study, 6 females and 14 males, aged between 20 and 27 (M = 23, SD = 1.662). All of them use computers in their daily lives, and some have a little experience in operating mobile robots (M = 1.22, SD = 0.752). We use a five-point scale ranging from 1 (not familiar at all) to 5 (very familiar). Participants also reported how familiar they were with smartpads or tablets (M = 4.91, SD = 0.223).

We set up the ∞-shaped course out lined with green tape for the user study of teleoperating the telepresence robot, as shown in Figure 15b. The inner and outer ring radii are 800 mm and 1600 mm, respectively. The ∞-shaped course is a loop route along which the participant teleoperates the robot for multiple laps, and it contains the same left and right turning routes. In this study, we asked the participants to use the TIUI and the GUI respectively to teleoperate the robot along the course for two laps as quickly as possible. The order of using the two user interfaces was counterbalanced across participants.

7.1.1 Measures and Analysis

We adopted the evaluation methods used in [55], [56], and [57] to make objective and subjective measures. The objective measures include the number of practice laps, task completion time, and the number of times the robot touches the route borders. The task completion time is from when the experimenter issued the start command ("3, 2, 1, Go") to when the robot returned to the start/end position for the second time (two laps). The subjective measures seek to obtain maneuverability, flexibility, convenience, participant skill level, and interface preference. We administered a post-task questionnaire to obtain the subjective measures. The questions are as follows:

Q1 How skilled were you using the GUI to teleoperate the robot?
Q2 How skilled were you using the TIUI to teleoperate the robot?
Q3 I was satisfied with the maneuverability of the TIUI.
Q4 I was satisfied with the maneuverability of the GUI.



(a)

(b)

Fig. 15. Illustration of the experimental setup. (a) Two user interfaces (TIUI and GUI) used by a participant to teleoperate the robot. (b) The ∞-shaped course out lined with green tape, where the area enclosed by the two yellow lines is the start/end position of the robot.

Q5 I felt that using the GUI to teleoperate the robot was flexible.
Q6 I felt that using the TIUI to teleoperate the robot was flexible.
Q7 I felt it was convenient to use the TIUI to teleoperate the robot.
Q8 I felt it was convenient to use the GUI to teleoperate the robot.
Q9 Which user interface (GUI or TIUI) do you prefer?

The questionnaire adopts a five-point scale, ranging from 1 (unskilled) to 5 (very skilled) for questions Q1 and Q2, and ranging from 1 (strongly disagree) to 5 (strongly agree) for questions Q3 to Q8.

We ran a paired t-test analysis to determine the effect of the TIUI and the GUI on the objective measures. To test for statistical significance, we used a cut-off value of $p < .05$.

7.1.2 Procedure

The participants first filled in the pretest questionnaire to obtain demographic information, and then were instructed how to use the two interfaces (TIUI and GUI) to teleoperate the telepresence robot in an open area. They learned to teleoperate the telepresence robot by using the two user interfaces without time limit (so that they were indeed familiar with teleoperating without worrying). Most participants were familiar with teleoperating within less than 10 minutes. After instruction, we asked the participants to report their skill level in using the two interfaces.

The evaluation includes practice, test, and questionnaires and interviews. In the practice phase, the participants used one of the two interfaces to teleoperate the robot along the ∞-shaped course until they thought they could perform the task, and the number of practice laps was recorded. In the test phase, the participants used one interface to teleoperate the robot to move along the course for two laps, and the task completion time and the number of times the robot touched the route borders were recorded. Then, they continued to use another interface to do the same task. To verify the various events that occurred during the experiment, we used three cameras to record the experimental process. A reward was added to encourage the participants to teleoperate the robot as quickly as possible, and a penalty of 5 seconds was imposed each time the robot touched the route borders.

In the questionnaire, the participants were asked to evaluate maneuverability, flexibility, convenience, participant skill level, and interface preference. Finally, the experimenters interviewed the participants to collect their comments and suggestions.

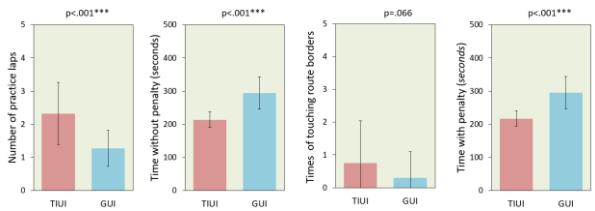
7.1.3 Results

During instruction, most participants took less than 3 minutes to become familiar with teleoperating the telepresence robot using the GUI and took up to 9 minutes using the TIUI. The average time taken to learn to use the GUI and the TIUI was approximately 1 minute and 6 minutes, respectively.

Figures 16 and 17 show the results of the objective and subjective measures of teleoperating the robot. Figure 16a shows that the participants made more practice laps using the TIUI ($M_{TIUI} = 2.33$, SD = 0.936) than using the GUI ($M_{GUI} = 1.28$, SD = 0.550), and the difference is significant, t (19) = -5.21, $p < .001$. Participants felt that their skills in using the GUI ($M_{GUI} = 4.54$, SD = 0.419) were higher than their skills in using the TIUI ($M_{TIUI} = 4.36$, SD = 0.563) (Figure 17a), t (19) = 1.75, p = .048. Some participants said they used the TIUI for the first time, so they wanted to spend more time becoming more familiar with the touchscreen gestures. Some participants said that they spent more time practicing the TIUI because the touchscreen gestures are a bit complicated.

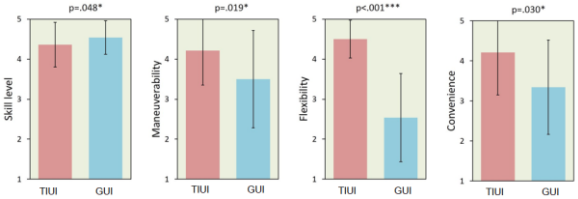
Figure 16b shows the average task completion time. We found a major effect on this measure ($M_{TIUI} – M_{GUI} = -81$), t (19) = 7.60, $p < .001$. The average time for completing the task using the TIUI and the GUI was 213 seconds (SD = 21.045) and 294 seconds (SD = 48.853), respectively. Figure 16c shows that there was no significant difference in the number of times the robot touched the route borders ($M_{TIUI} – M_{GUI} = 0.45$), t (19) = -1.58, p = .066. Figure 16d shows the average task completion time with penalty (i.e., task completion time + 5 seconds × the number of times the robot touched the route borders) for each participant. We found that the major effect on this measure (i.e., the average task completion time with penalty) is still maintained ($M_{TIUI} – M_{GUI} = -78.75$), t (19) = 7.49, p<.001. Thus it can be seen that participants achieved better performance of task completion time using the TIUI than using the GUI.

Participants reported that they were more satisfied with the maneuverability of the TIUI than that of the GUI ($M_{TIUI} - M_{GUI} = 0.71$), t (19) = -2.24, p = .019 (Figure 17b). Comparing the TIUI and the GUI on flexibility, there was a significant difference ($M_{TIUI} – M_{GUI} = 1.96$), t (19) = -8.50, p< .001 (Figure 17c). All participants felt that the flexibility of using the TIUI to teleoperate the robot was higher than that of



(a) (b) (c) (d)

Fig. 16. Objective measures of teleoperating the telepresence robot using the TIUI and GUI, respectively. (a) The average number of practice laps. (b) The average task completion time. (c) The average number of times the robot touched the route borders. (d) The average task completion time with penalty). (***) denotes p<0.001.



(a) (b) (c) (d)

Fig. 17. Subjective measures of teleoperating the telepresence robot using the TIUI and GUI, respectively. (a) The average level of skill. (b) The average satisfaction with the maneuverability. (c) The average flexibility. (d) The average convenience. (***) denotes p<0.001 and (*) p<0.05.

the GUI. Figure 17d shows a major effect on the convenience of the user interfaces ($M_{TIUI} - M_{GUI} = 0.85$), t (19) = -2.00, p= .030. Participants said that they can use the TIUI to control the robot smoothly, but when using the GUI, they felt unsmooth because they always aligned their fingers with the graphical buttons and often switched between the buttons to turn left or right.

Seventeen of the twenty participants said that they would choose the TIUI to teleoperate the robot if they had one more chance, while the remaining three participants said that they preferred the GUI as it is very simple. Participants who chose the TIUI said that they could complete the task faster and touch the route borders less when using the TIUI because they become more skilled with more practice. Furthermore, most participants reported that they enjoyed the TIUI very much because of the intuitive and flexible interactions.

7.2 Teleoperating Everyday Objects

To evaluate the telepresence operation of everyday objects, we created a scenario for visiting and assisting elderly person in the smart home (see Figure 3). The purpose is to enhance the usefulness verification of the proposed framework, as proposed by Greenberg and Buxton [58]. In the scenario, an elderly person has a limited ability to drive a wheelchair, and additional assistance is needed. There are some everyday objects with teleinteractive boxes in the room, such as lighting devices, media devices, doors with password access controls, wheelchairs, curtains, and other furniture, such as sofas, tables, and chairs. The furnishings and appliances provide reference points for participants (who are in another place specified as the remote site) to orient themselves in the live video of the room (specified as the local site).

For the study of teleoperating everyday objects, we recruited another 24 participants from the same local university as the previous study, 12 females and 12 males, whose ages ranged from 18 to 25 years (M = 20.13, SD = 2.242). All of them use computers in their daily lives, but they had no experience in operating robots. Participants reported how familiar they were with smartpads or tablets (M = 3.53 and SD = 1.523). We still used a five-point scale ranging from 1 (not at all familiar) to 5 (very familiar). We invited a member of our laboratory to play an elderly person, and all participants acted as visitors.

The experiment consisted of two sessions: visiting the elderly person in person and visiting the elderly person using the telepresence robot. In the first session, the participants personally visited the elderly person's home and operated everyday objects. In the second session, the participants used the TIUI to teleoperate the telepresence robot to the elderly person's home, and then teleoperated the same objects.

7.2.1 Procedures

Participants were instructed in the visit procedure before performing the task. We asked all the participants to strictly follow the procedure to make the visits of different participants as consistent as possible. Referring to the elderly person's room shown in Figure 3, we presented the following visiting procedure for the first session:

(1) *The participant walks to the door (B) of the elderly person's room, enters the access control password, and opens the door.*
(2) *The participant enters the room, and taps the lighting button to turn on the lights (C and D).*
(3) *The participant can see the elderly person sitting on the sofa, and says hello to the elderly person.*
(4) *The participant asks the elderly person to sit in the wheelchair (E), and then pushes the wheelchair to the curtain (F).*
(5) *The participant opens the curtain, and the elderly person looks out of the window.*

After the participants completed the first session, they moved to another room (remote site) to proceed with the second session.

In the second session, the participants performed the same visit as the first session by teleoperating the telepresence robot using the TIUI. The participants were instructed how to use the TIUI to teleoperate the robot in an open area. This instruction process is the same as the process mentioned in Section 7.1, and the participants were able to become familiar with teleoperating the telepresence robot using the TIUI without time limit until they thought they were proficient. Then, we showed them the procedure for the second session as follows:

(1) *The participant uses the TIUI to teleoperate the robot (A) to the door (B), and taps the password on the live video image of the access control panel to open the door.*
(2) *The participant teleoperates the robot into the room, and then taps the lighting button to turn on the lights (C and D).*
(3) *The participant can see the elderly person sitting on the sofa in the live video, and says hello to the elderly person.*
(4) *The participant asks the elderly person to sit in the wheelchair (E), and then teleoperates the wheelchair to the curtain (F).*
(5) *The participant uses the lasso gesture to lasso the curtain image and then to drag the curtain; the curtain opens with the drag gesture, and the elderly person looks out of the window.*

After completing the second session, each participant was asked to complete a questionnaire.

7.2.2 Measures

Measures include the user-friendliness of the interaction, the convenience of teleoperation, the utility of the robot, and the feeling of "being there". The questionnaire consists of seven questions, each based on a five-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). The questions are as follows:

Q1 Teleoperating objects using the TIUI was user-friendly.
Q2 Teleoperating objects using the TIUI was convenient.
Q3 The telepresence robot could act as my embodiment to complete teleoperation tasks.
Q4 When tapping the buttons of the access control panel in the live video on the TIUI to open the door, I felt as if I were tapping the physical buttons on site.
Q5 When dragging the curtain in the live video on the TIUI, I felt as if I were dragging the physical curtain on site.
Q6 When pushing the wheelchair in the live video on the TIUI, I felt as if I were pushing the physical wheelchair on site.
Q7 When tapping the light button image on the TIUI to turn on the light, I felt as if I were tapping the physical light button on site.

7.2.3 Results

The participants responded very positively to the telepresence operation system, as depicted in Figure 18. More than half of the participants reported that they quickly learned to use the system. Most participants said that it is user-friendly to teleoperate teleinteractive devices by touching live video (M = 4.04, S = 0.550). They deemed the TIUI intuitive and natural to use (M = 4.50, SD = 0.590). They experienced the high utility of telepresence operation using the TIUI. They truly had the feeling of "being there" (M = 4.25, SD = 0.532) and reported that they teleoperated everyday objects as if they were doing so on site, including opening the door (M = 4.58, SD = 0.504), drawing the curtain (M = 4.63, SD = 0.495), pushing the wheelchair (M = 4.38, SD = 0.711), and turning on the lights (M = 4.29, SD = 0.690).

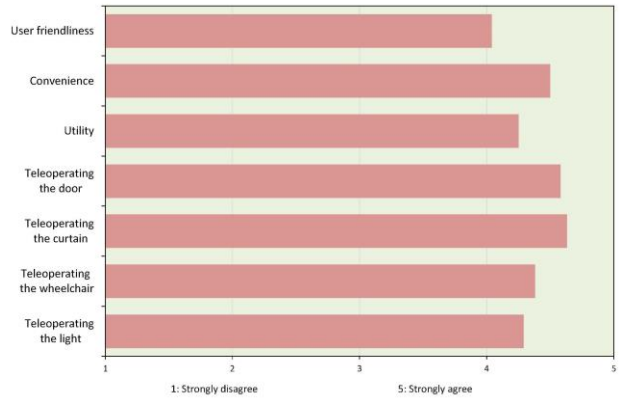


Fig. 18. Interaction efficiency, convenience, utility, and user friendliness of using the telepresence operation system through the TIUI.

8. Discussion

Our study confirmed that the TIUI enables remote users not only to teleoperate a telepresence robot for video chat with local persons, but also to teleoperate local everyday objects by touching live video with touchscreen gestures. In this section, we discuss the implications, limitations, and future work.

8.1 Implications

Our TIUI is superior to the GUIs in three aspects: (1) As long as we see a touchable live video on a touchscreen, we can teleoperate an interesting object in the video with touchscreen gestures; (2) Touching the touchable live video of a physical object for teleoperation has a better feeling of "being there" than touching its corresponding graphical button (or icon); (3) There is almost no visual burden because the TIUI allows users to teleoperate objects by touching objects' live video they are watching, without distracting them by looking for the corresponding buttons of the objects for teleoperation.

Existing mobile robotic telepresence systems lack the capability of teleoperating local objects and need local help. Our TIUI empowers users not only to teleoperate a telepresence robot but also to teleoperate local objects by touching live video without local help. To enhance users' awareness of a local site, we argue that stitching the FF and DF videos together into one video (FDF video) can enable remote users to look forward clearly for object recognition and look downward panoramically for robot navigation.

We have introduced the recognition module and knowledge module into the telepresence operation system. These modules enable the system to recognize local objects and then to connect them to the TIUI for telepresence operation. In other words, our framework for telepresence operation is based on computer vision (CV) and artificial intelligence (AI). We envision that CV and AI will greatly facilitate the development of telepresence operation systems with help of live video understanding and reasoning.

8.2 Limitations

Our work has the following major limitations. The first is that the system cannot perform 3D teleoperation tasks, such as 3D motion of robot arms [13] or 3D actuated objects [14]. There have been many excellent 3D representation algorithms [59] [60] [59]and 3D modeling algorithms [61] [62]. Our work aimed to teleoperate everyday objects by touching live video in a pervasive and compact way, and we did not focus on the use of 3D models and 3D representations that require high computational costs.

The second limitation is that teleinteractive devices located in the lower zone of the TIUI cannot be teleoperated, because we assumed that teleinteractive devices are located in front of the FF camera and should appear in the upper zone of the TIUI. For example, when a vacuum cleaner robot is close to the robot base, it enters the field of view of the DF camera corresponding to the lower zone of the TIUI, and we cannot use the touchscreen gestures defined for the upper zone to teleoperate it.

Our evaluation was only conducted with users who are 18–27 years old with proficient touchscreen skills. We have not performed user studies on other age groups, nor have we invited elderly and/or disabled people to participate in the study, resulting in limited evaluation results.

In order to minimize the impact of network latency, the remote site and the local site used the same local wireless network. Network lag is an important concern that will need to be addressed in application systems. We did not consider many other important issues, such as security and privacy, which may seriously affect the usefulness of the system.

8.3 Future Work

Currently, our system runs at approximately 10 frames per second on a laptop (ThinkPad x200) on the robot, including image stitching, object recognition, localization and tracking, etc. All of these are basic and simple algorithms to ensure that the system runs in real time. We will adopt new computing strategies (hardware and software) to improve system performance.

We will extend the proposed framework to many areas of daily lives and workplaces, as long as there are teleinteractive devices or objects equipped with teleinteractive boxes. If a local person wearing cameras replaces the

telepresence robot to capture live video, it constitutes a remote collaborative work environment based on touchable live video. If a local person wears a teleinteractive device, the remote user can perform virtual touch interaction by touching live video. If you want animals to wear teleinteractive devices to perform telepresence operation by touching live video, you first need to investigate the animals' feelings to ensure the protection of animals.

9. Conclusions

This paper has presented a framework for telepresence operation of everyday objects by touching live video with touchscreen gestures. We developed a touchscreen user interface, TIUI, which can empower remote users not only to teleoperate a telepresence robot but also to teleoperate any teleinteractive device at a local site by touching the live video they are watching. The TIUI can support pervasive telepresence operations, and it is easy to learn and easy to use. We implemented a low-cost telepresence operation system that can perform the telepresence operation of everyday objects towards potential applications.

Acknowledgments

This paper was supported partly by the National Science Foundation of China under Grants No. 61673062 and No. 61773062. The authors would like to thank Dr. Pei Mintao, Dr. Wu Yuwei and other laboratory members: Shen Jiajun, Dong Zhen, Hou Jingyi, Yang Min, Shen Weichao, and Chen Xianda, for their helpful work on the earlier version of this paper. We thank Professor Saul Greenberg for his insightful and valuable suggestions to our manuscript. The authors also thank the reviewers for their insightful and valuable suggestions.

The main contributions of Yanmei Dong, Bin Xu, and Che Sun are the system implementation and evaluation. Corresponding author is Che Sun.

References

[1] C. Neustaedter, J. Zrocyk, A. Chua, A. Forghani, and C. Pang, "Mobile video conferencing for sharing outdoor leisure activities over distance," *Hum.-comput. Interact.,* vol. 35, no. 1, pp. 103-142, 2020.

[2] E. R. McClure, Y. E. Chentsova-Dutton, S. J. Holochwost, W. G. Parrott, and R. Barr, "Look at that! Video chat and joint visual attention development among babies and toddlers," *Child development*, vol. 89, no. 1, pp. 27-36, 2018.

[3] "Beam website," *https://suitabletech.com/*, Oct., 2020.

[4] "Ohmnilabs website," *https://ohmnilabs.com/*, Oct., 2020.

[5] M. K. Lee and L. Takayama, "Now, I have a body: Uses and social norms for mobile remote presence in the workplace," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 33-42, May, 2011.

[6] V. Kaptelinin, P. Björnfot, K. Danielsson, and M. U. Wiberg, "Mobile Remote Presence Enhanced with Contactless Object Manipulation: An Exploratory Study," in *Proc. Extended Abstracts Hum. Factors Comput. Syst.*, pp. 2690-2697, May, 2017.

[7] Y. Jia, B. Xu, J. Shen, M. Pei, Z. Dong, J. Hou, M. Yang, "Telepresence interaction and operation by touching live video images," *arXiv: 1512.04334*, 2015.

[8] T. Fong and C. Thorpe, "Vehicle teleoperation interface," *Auton. Robots*, vol. 11, no. 1, pp. 9-18, 2001.

[9] J. Vertesi. "Seeing like a rover: How robots, teams, and images craft knowledge of mars". *University of Chicago Press*, 2015.

[10] S. Singhal, C. Neustaedter, Y. L. Ooi, A. N. Antle, and B. Matkin, "Flex-N-Feel: The design and evaluation of emotive gloves for couples to support touch over distance," in *Proc. ACM Conf. Computer Support. Coop. Work Soc. Comput.,* pp. 98-110, Mar. 2017.

[11] H. R. Pelikan, A. Cheatle, M. F. Jung, and S. J. Jackson, "Operating at a Distance - How a Teleoperated Surgical Robot Reconfigures Teamwork in the Operating Room," in *Proc. ACM on Hum.-comput. Interact.*, Vol. 2, No. CSCW, pp. 1-28, Nov. 2018.

[12] M. Tani, K. Yamashi, K. Tanikoshi, M. Futakawa, and S. Tanifuji, "Object-oriented video: interaction with real-world objects through live video," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, pp. 593-598, May, 1992.

[13] S. Hashimoto, A. Ishida, M. Inami, and T. Igarashi, "Touchme: An augmented reality based remote robot manipulation," in *Proc. 21st Int. Conf. Artif. Reality Telexistence*, vol. 2, pp. 61 -66, 2011.

[14] S. Kasahara, R. Niyama, V. Heun, and H. Ishii, "exTouch: Spatially-aware embodied manipulation of actuated objects mediated by augmented reality," in *Proc. ACM Int. Conf. Tangible, Embed. and Embodied Interact.*, pp. 223-228, Feb., 2013.

[15] T. Seifried, M. Haller, S. Scott, F. Perteneder, C. Rendl, D. Sakamoto, and M. Inami, "CRISTAL: Design and implementation of a re-mote control system based on a multi-touch display," in *Proc. ACM Int. Conf. Interact. Tabletops Surfaces*, pp. 33-40, 2009.

[16] C. Guo, J. E. Young, and E. Sharlin, "Touch and toys: New techniques for interaction with a remote group of robots," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 491-500, Apr., 2009.

[17] D. Sakamoto, K. Honda, M. Inami, and T. Igarashi, "Sketch and Run: A stroke-based interface for home robots," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 197-200, Apr., 2009.

[18] K. Liu, D. Sakamoto, M. Inami, and T. Igarashi, "Roboshop: Multilayered sketching interface for robot housework assignment and management," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 647-656, May, 2011.

[19] J. A. Frank and V. Kapila, "Path Bending: Interactive human-robot interfaces with collision-free correction of user-drawn paths," in *Proc. ACM Int. Conf. Intell. User Interfaces*, pp. 186-190, Mar., 2015.

[20] J. Kato, D. Sakamoto, M. Inami, and T. Igarashi, "Multi-touch interface for controlling multiple mobile robots," in *Proc. Extended Abstracts Hum. Factors Comput. Syst.*, pp. 3443-3448, Apr., 2009.

[21] S. Kim and J. Park, "Touchable Video Streams: Towards Multisensory and Multi-contact Experiences," *ACM Int. Conf. on Interactive Experiences for Television and Online Video*, pp. 155–160, 2018.

[22] M. Y. Sung, K. Jun, D. Ji, H. Lee, and K. Kim, "Touchable video and tactile audio," in *Proc. IEEE Int. Symp. Multimedia*, pp. 425-431, Dec., 2009.

[23] E. Paulos and J. Canny, "PRoP: Personal roving presence," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 296-303, Apr., 1998.

[24] A. Kristoffersson, S. Coradeschi, and A. Loutfi, "A review of mobile robotic telepresence," *Adv. Hum. Comput. Interaction*, vol. 2013, no. 3, pp. 1-17, 2013.

[25] G. Cortellessa, Francesca Fracasso, A. Sorrentino, et al., "ROBIN, a telepresence robot to support older users monitoring and social inclusion: development and evaluation," *Telemedicine and e-Health*, vol. 24, no. 2, pp, 145-154, 2018.

[26] Naomi T. Fitter and Y. Chowdhury and Elizabeth Cha and L. Takayama and M. Mataric, "Evaluating the effects of personalized appearance on telepresence robots for education," in *Proc.*

ACM Conf. Hum.-Robot Interact, pp. 109-110, Mar., 2018.
[27] M. K. Lee, L. Takayama, "Now, I have a body: Uses and social norms for mobile remote presence in the workplace," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp.33-42, May, 2011.
[28] K. M. Tsui, J. M. Dalphond, D. J. Brooks, M. S. Medvedev, E. McCann, J. Allspaw, D. Kontak, and H. A. Yanco, "Accessible human-robot interaction for telepresence robots: a case study," *Paladyn J. Behav. Robot*, vol. 6, no. 1, pp.1-29, 2015.
[29] C. Neustaedter, G. Venolia, J. Procyk, et al., "To Beam or Not to Beam: A study of remote telepresence attendance at an academic conference," in *Proc. ACM Conf. Computer Support. Coop. Work Soc. Comput.*, pp. 418-431, Mar., 2016.
[30] "Double website". *http://www.doublerobotics.com/*, Oct., 2020.
[31] S. Mayer, A. Tschofen, A.K. Dey, and F. Mattern, "User interfaces for smart things--A generative approach with semantic interaction descriptions," *ACM Trans. Computer-Human Interact.*, vol. 21, no. 2, pp. 12, 2014.
[32] K. M. Tsui, A. Norton, et al., "Iterative design of a semi-autonomous social telepresence robot research platform: a chronology," *Intell. Serv. Robotics*, vol. 7, no. 2, pp. 103-119, 2014.
[33] C. L. Fernando, M. Furukawa, et al., "Design of TELESAR V for transferring bodily consciousness in telexistence," *IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, pp. 5112-5118, Oct., 2012.
[34] J. M. Badger, P. Strawser, et al., "Robonaut 2 and Watson: Cognitive dexterity for future exploration," *IEEE Aero. Conf.*, pp.1-8, 2017.
[35] K. Kaneko, H. Kaminaga, et al., "Humanoid Robot HRP-5P: An electrically actuated humanoid robot with high-power and wide-range joints," *IEEE Robotics Autom. Lett.,* vol. 4, no. 2, pp. 1431-1438, 2019.
[36] P. Birkenkampf, D. Leidner, and N. Y. Lii, "Ubiquitous user interface design for space robotic operation," in *Proc. Symp. Adv. Space Tech. in Robotics and Autom.*, 2017.
[37] S. N. Young and J. M. Peschel, "Review of human–machine interfaces for small unmanned systems with robotic manipulators," *IEEE Trans. Hum. Mach. Syst.*, vol. 50, no. 2, pp. 131-143, 2020.
[38] E. Stolterman, M. Wiberg, "Concept-Driven Interaction Design Research," *Hum. Comput. Interact.*, vol. 25, no. 2, pp.95-118, 2010.
[39] M. Weiser, "The computer for the 21st century," *Scientific American*, vol. 265, no. 3, pp. 94-104, 1991.
[40] G. Kortuem, F. Kawsar, D. Fitton, and V. Sundramoorthy, "Smart Objects as Building Blocks for the Internet of Things," *IEEE Int. Comput.*, vol. 14, no. 1, pp. 30-37, 2010.
[41] L. Wang, D. Peng, and T. Zhang, "Design of smart home system based on WiFi smart plug," *Int. J. Smart Home*, vol. 9, no. 6, pp. 173-182, 2015.
[42] J. Sigut, M. Castro, and R. Arnay. "OpenCV Basics: A Mobile Application to Support the Teaching of Computer Vision Concepts," *IEEE Trans. Educ.*, vol. 63, no. 4, pp. 328-335, 2020.
[43] Blackler, A., Popovic, V. and Mahar, D. "Investigating users'intuitive interaction with complex artefacts," *Applied Ergonomics.*, no. 41, pp. 72–92, 2010.
[44] J. J. Gibson, "The theory of affordances," *Hilldale,* 1977.
[45] J. Shen, B. Xu, M. Pei, and Y. Jia, "A low-cost telepresence wheelchair system," in *Proc. IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, pp. 2452-2457, Oct., 2016.
[46] B. Xu and Y. Jia, "Wide-angle image stitching using multi-homographic warping," in *Proc. IEEE Int. Conf. Image Proc.*, pp. 1467-1471, Sep., 2017.
[47] Y. Dong, Y. Jia, W. Shen, and Y. Wu, "Can You Easily Perceive the Local Environment? A User Interface with One Stitched Live Video for Mobile Robotic Telepresence Systems." *Int. J. Human–Computer Interact*, vol. 36, no. 8, pp. 736–747, 2020.
[48] A. E. Leeper, K. Hsiao, M. Ciocarlie, L. Takayama, and D. Gossow, "Strategies for human-in-the-loop robotic grasping," in *Proc. ACM/IEEE Int. Conf. Hum.-Robot Interact.*, pp. 1-8, Mar., 2012.
[49] A. Kidd (Ed.), "Knowledge acquisition for expert systems: A practical handbook," *Springer Science & Business Media*, 2012.
[50] R. Brooks, "A robust layered control system for a mobile robot," *IEEE J. Robotics and Automation*, vol. 2, no. 1, pp. 14-23, 1986.
[51] C. Villamor, D. Willis, and L. Wroblewski, "Touchscreen gesture reference guide. Touchscreen gesture Reference Guide," *http://www.lukew.com/touch/*, Oct., 2020.
[52] A. Cesta, G. Cortellessa, A. Orlandini, and L. Tiberio, "Long-term evaluation of a telepresence robot for the elderly: Methodology and ecological case study," *Int. J. Social Robotics*, vol. 8, no. 3, pp. 421-441, 2016.
[53] T. C. Tsai, Y. L. Hsu, A. I. Ma, T. Kin, and C. H. Wu, "Developing a telepresence robot for interpersonal communication with the elderly in a home site," *Telemedicine and e-Health*, vol. 13, no. 4, pp. 407-424, 2007.
[54] "Giraff," *http://www.giraff.org/*, Oct., 2020.
[55] S. Johnson, I. Rae, B. Mutlu, and L. Takayama, "Can you see me now? How field of view affects collaboration in robotic telepresence," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 2397-2406, Apr., 2015.
[56] L. Takayama and H. Harris, "Presentation of (Telepresent) self: On the double-edged effects of mirrors," in *Proc. ACM/IEEE Int. Conf. Hum.-Robot Interact.*, pp. 381-388, Mar., 2013.
[57] N. T. Fitter, N. Raghunath, E. Cha, C. A. Sanchez, L. Takayama, and M. J. Matari, "Are we there yet? Comparing remote learning technologies in the university classroom," *IEEE Robotics Autom. Lett.*, vol. 5, no. 2, pp. 2706-2713, 2020.
[58] S. Greenberg and B. Buxton, "Usability evaluation considered harmful (some of the time)," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 111–120, Apr., 2008.
[59] M. Hebert, K. Ikeuchi, and H. Delingette. "A spherical representation for recognition of free-form surfaces," IEEE Trans. Pattern Anal. Mach. Intell., vol. 17, no. 7, pp. 681-690, 1995.
[60] M. Wheeler, Y. Sato, and K. Ikeuchi. "Consensus surfaces for modeling 3D objects from multiple range images," in proc. IEEE int. Conf. Comput. Vis., pp. 917-924, 1998.
[61] H-Y. Shum, K. Ikeuchi, and R. Reddy. "Principal component analysis with missing data and its application to polyhedral object modeling," IEEE Trans. Pattern Anal. Mach. Intell. Vol. 17, no. 9, pp. 854-867., 1995.
[62] I. Sato, Y. Sato, and K. Ikeuchi. "Acquiring a radiance distribution to superimpose virtual objects onto a real scene," IEEE Trans. Vis. Comput. Graphics, vol. 5, no. 1, pp. 1-12, 1999.

Yunde Jia (M'11) received the BS, MS and PhD degrees from the Beijing Institute of Technology (BIT) in 1983, 1986, and 2000, respectively. He is a Professor with the School of Computer Science, BIT, and the team head of BIT innovation on vision and media computing. He serves as the director of Beijing Lab of Intelligent Information Technology. He was a visiting scientist in the School of Computer Science at Carnegie Mellon University (CMU) from 1995 to 1997. In recent years, his interests have extended to vision-based HCI and HRI, intelligent robotics, and cognitive systems.