Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

Can You Easily Perceive the Local Environment? A User Interface with One Stitched Live Video for Mobile Robotic Telepresence Systems

Yanmei Dong, Yunde Jia, Weichao Shen & Yuwei Wu

To cite this article: Yanmei Dong, Yunde Jia, Weichao Shen & Yuwei Wu (2020) Can You Easily Perceive the Local Environment? A User Interface with One Stitched Live Video for Mobile Robotic Telepresence Systems, International Journal of Human–Computer Interaction, 36:8, 736-747, DOI: 10.1080/10447318.2019.1685194

To link to this article: https://doi.org/10.1080/10447318.2019.1685194

Published online: 05 Nov 2019.

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
2020, VOL. 36, NO. 8, 736–747
https://doi.org/10.1080/10447318.2019.1685194

Taylor & Francis
Taylor & Francis Group

Can You Easily Perceive the Local Environment? A User Interface with One Stitched Live Video for Mobile Robotic Telepresence Systems

Yanmei Dong, Yunde Jia, Weichao Shen, and Yuwei Wu

Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing, China

ABSTRACT
Many existing mobile robotic telepresence systems have equipped with two cameras, one is a forward-facing camera for video communication, and the other is a downward-facing camera for robot navigation. However, the two live videos from these two cameras would cause some confusion which makes it difficult for a remote operator to perceive the local environment. In this paper, we propose to use a user interface with one stitched live video instead of two live videos for mobile robotic telepresence systems. We used a video stitching algorithm to stitch the two live videos into one live video through which a remote operator can well perceive the local environment. We conducted a user study to investigate the difference between one stitched live video and two separate live videos in the user interface. The results show that the user interface with one stitched live video improves task efficiency, the number of errors, and remote operators' feelings of presence, and enables remote operators to concentrate on the work they are doing.

1. Introduction

A mobile robotic telepresence system incorporates a video conferencing device into a mobile robot, allowing a remote operator to teleoperate the robot as his/her embodiment to actively telecommunicate with other people in a way similar to face-to-face interaction (Kristoffersson, Coradeschi, & Loutfi, 2013). In recent years, mobile robotic telepresence systems are increasingly common in various everyday contexts, such as elderly people support (Cesta, Cortellessa, Orlandini, & Tiberio, 2016; Koceski & Koceska, 2016), remote education (Cha, Chen, & Mataric, 2017; Rudolph et al., 2017), and remote meeting attendance (Neustaedter, Venolia, Procyk, & Hawkins, 2016; Rae & Neustaedter, 2017).

Existing mobile robotic telepresence systems (BeamPro, 2018; Double2, 2018; QB, 2018) are usually equipped with a forward-facing camera (FF-camera) for video communication and a downward-facing camera (DF-camera) for robot navigation. Some of the systems, e.g., QB (2018), use a single video window to alternately display the two live videos from the FF-camera and DF-camera by manual switching. Some of the systems, e.g., Double2 (2018) and BeamPro (2018), use two video windows to simultaneously display the two live videos, allowing a remote operator to teleoperate the robot while telecommunicating with local persons. The two videos show different parts of a local environment, as shown in Figure 1(b). Boll (2017) found that one needs to frequently switch his/her attention between the two videos to see the different parts. For example, he/she looks at targets in the front scene through the video in the upper window and looks at the floor through the video in the lower window for robot navigation. The remote operator also has to pay much time to familiarize himself/herself with the two videos (Boll, 2017), and to find the relationship among the different parts of the local environment in the videos. Therefore, using two videos to display the local environment would cause some confusion and decrease task performance (Nielsen, Goodrich, & Ricks, 2007).

To alleviate the remote operator's burden and confusion when interacting with the local environment, we propose to use a user interface with one stitched live video for mobile robotic telepresence systems. We conducted a user study through a between-subject controlled laboratory experiment to investigate the difference between one stitched live video and two separate live videos in the user interface. The user study with 18 participants consists of a tele-exploration task and a telecommunication task. We analyzed the experimental data to find the difference. The results show that the user interface with one stitched live video improves remote operators' feelings of presence in the local environment, task efficiency, the number of errors, and enables remote operators to concentrate on the work they are doing.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 lists the hypotheses we make. The user study design is described in Section 4, including the introduction of our telepresence system, participants, environment setup, tasks, measures, experimental procedure, and data analysis methods. Section 5 shows the results of hypothesis testing. Section 6 discusses the study

CONTACT Yuwei Wu ✉ wuyuwei@bit.edu.cn Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, China.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hihc.

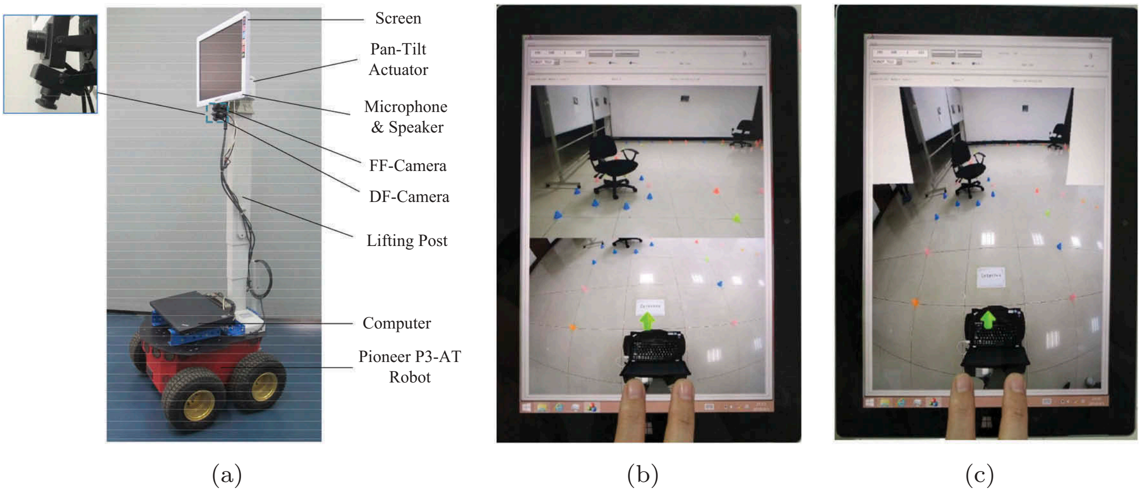

Figure 1. The Mcisbot robot and the two user interfaces. (a) The camera configuration of the Mcisbot. (b) The user interface with two separate live videos, i.e., the FF-video and the DF-video. (c) The user interface with one stitched live video that is stitched from the FF-video and the DF-video. The green arrows indicate the current moving direction of the robot.

findings, and Section 7 describes some implications for the design of mobile robotic telepresence systems. Section 8 describes the limitations of our work and the corresponding future work that can be done for further study. We conclude this work in Section 9.

2. Related work

2.1. Video feedback

Mobile robotic telepresence systems are typically designed to provide social interaction based on a video conferencing system that can improve the remote operator's feelings of presence in a local environment. Through a video conferencing system, a remote operator can acquire visual information of a local environment, which is critical in improving remote operator's user experience, keeping robot safe, and helping a remote operator to complete remote tasks (Berisha, Kölle, & Griesbaum, 2015; Boll, 2017; Neustaedter et al., 2016; Rae & Neustaedter, 2017; Rae, Venolia, Tang, & Molnar, 2015; Rebola & Eden, 2017; Yang, Jones, Neustaedter, & Singhal, 2018).

The early mobile robotic telepresence system, uses a conventional camera with a limited view angle to capture a live video of a local environment which makes it difficult for a remote operator to well perceive a local environment due to the keyhole effect in viewing (Woods, Tittle, Feil, & Roesler, 2004). Giraff (2018) uses a webcam with a wide-angle lens for both robot navigation and video communication in hospital and home care. A remote operator could adjust the robot head to look forward or look at the ground using the same camera. VGO (2018) is equipped with an automatic or manual tilting camera to achieve the same goals. However, manually adjusting the robot head or camera viewpoint would increase the burden on a remote operator to perceive a local environment, and also distract the operator's attention. In addition, more degrees of control freedom would result in higher user learning requirements and effort (Tsui & Yanco, 2013).

Keyes, Casey, Yanco, Maxwell, and Georgiev (2006) proposed to use two web cameras in a mobile robotic telepresence system, a forward-facing camera and a downward-facing camera, for improving the remote operator's situation awareness. Recent commercially available mobile robotic telepresence systems, such as QB (2018), BeamPro (2018), and Double2 (2018), are all equipped with two cameras (i.e., an FF-camera and a DF-camera) for providing two separate live videos of a local environment. However, displaying a local environment in the two videos would cause some confusion, and a remote operator has to integrate scene information in the two videos to perceive a large field of view of the local environment (Chen, Haas, & Barnes, 2007), and he/she has to switch his/her attention between the two videos to see them (Boll, 2017). When the FF-camera and the DF-camera are with different lenses or the two videos are shown at different scales, a remote operator has to pay time to familiarize himself/herself with the FF-video and the DF-video (Boll, 2017). In contrast, we propose to stitch the FF-video and the DF-video into one video with a large view for mobile robotic telepresence systems.

Voshell, Woods, and Phillips (2005) showed that a user interface is more effective when it can help the remote operator to recognize the relationship between multi-displayed images. Lazewatsky and Smart (2011) presented a panorama-based user interface to help the remote operator to perceive a large view of the local environment. They used a motorized pan-tilt camera to create a static panorama by rotating the camera to scan the local environment. Specifically, the panorama consists of multiple image patches, and each patch was captured at a fixed position. Except the image corresponding to the current camera position, other regions of the panorama remain unchanged until next scanning. Therefore, the panorama is only suitable for teleoperation and telecommunication in a static environment. Tsui et al. (2015) used an augmented reality user interface to provide a larger field of view of the local environment for people to teleoperate the telepresence robot to visit a remote art gallery. The larger field of view was created from multiple conventional cameras by putting these videos together. However, the larger field of view looks like a simple cropping of several videos or several scenes on the user interface. Differently, we used a video stitching algorithm to create a stitched live video from an FF-video and a DF-video, and investigate the difference between one stitched live video and two separate live videos in tele-interaction.

2.2. *Effects of field of view*

For the effects of field of view, Shiroma, Sato, Chiu, and Matsuno (2004) investigated the effects of different camera images on teleoperation when using a conventional camera, an omnidirectional camera, and a fisheye lens camera. They found that if the telepresence robot is in the center of the live video image, the image could help to enhance teleoperation efficiency since the remote operator could obtain a full view of the robot's surroundings. Compared with the wide-angle landscape view in social interaction, the wide-angle portrait view was proved to be useful to improve the quality of interaction, and a bigger vertical view angle could improve the driving experience (Kiselev, Kristoffersson, & Loutfi, 2014). Comparisons on three widths of field of view (narrow, wide-angle, and panoramic) on collaboration were conducted to show that a wider view is beneficial to improve task efficiency and decrease collisions, but user interfaces with the wider view is more difficult for use (Johnson, Rae, Mutlu, & Takayama, 2015). Heshmat et al. (2018) developed a user interface with a 360-degree video which was viewed through head-mounted displays. They investigated the benefits and challenges of the use of mobile robotic telepresence systems in an outdoor activity, geocaching. Their results showed that the remote operator had feelings of presence with the telepresence robots in geocaching. They also compared the difference between the 360^o view and a wide-angle field of view, and showed that driving with the 360^o view was harder than the wide-angle field of view. We propose to use one stitched live video from a wide-angle camera (i.e., the FF-camera) and a fisheye camera (i.e., the DF-camera), instead of two separate live videos. The DF-camera locates the robot near the center of DF view to provide a full view of the robot's surroundings. We investigated how the stitched live video and two separate live videos affect the remote operator's feelings of presence, concentration on the task, task performance, perceived task success, and perceived ease of using the user interface.

3. Hypotheses

By using the user interface with one stitched live video compared to two separate live videos, we made five hypotheses to investigate the difference between one stitched live video and two separate live videos in tele-interaction task.

- **Hypothesis 1**: The task performance will be improved. The task performance will be improved. Previous work has shown that the use of a larger field of view can reduce the effects of cognitive tunneling (Thomas & Wickens, 2000). Compared to the user interface with two separate videos, the stitched video has a larger in field of view. Therefore, we predict that the task performance, including task completion time and the number of task errors, will be decreased by using the user interface with one stitched live video.
- **Hypothesis 2**: The perceived task success will be improved. Perception of task success is important since it reports remote operators' confidence on task success, which helps them to gain a good experience of using the user interface. Showing the local environment in one stitched video would be more intuitive for accessing the environment information than showing in two separate videos. Thus, we predict that the remote operator would feel better in perceived task success since he/she can have a better perception on the robot's location and surroundings (Hestand & Yanco, 2004; Yanco & Drury, 2004).
- **Hypothesis 3**: The remote operator will be more concentrated on the task. With one stitched live video, a remote operator can directly perceive the local environment from one video with a larger field of view, instead of mentally integrating information in two videos to obtain a larger field of view. Compared to using the user interface with two separate videos, we predict that remote operators using the user interface with one stitched video could be more concentrated on the task since there is no need for information integration and attention switching among different videos (Chen et al., 2007).
- **Hypothesis 4**: The perceived ease of using the user interface will be improved. Video information shown on the user interfaces with one stitched video and two separate videos is the same since the video information is captured from the same cameras. The difference between the two user interfaces is that the environment information perception from the user interface with one stitched video is more intuitive and easier. We predict that the remote operator would feel easier in using the user interface with one stitched video, as two separate videos provide remote operators with more complex video information and require greater cognitive processing (Chen et al., 2007).
- **Hypothesis 5**: The remote operator's feelings of presence will be improved. The scene information shown on the user interface with one stitched live video is more similar to the real world since the scene is shown as a whole, while in the user interface with two separate videos, the scene is shown in two different parts. Therefore, we predict that the feeling of presence by using the user interface with one stitched live video would be more similar to the way of physically being in the local environment, i.e, the remote operator's feelings of presence would be increased (Schubert, Friedmann, & Regenbrecht, 2001).

4. User study design

We conducted a between-subject controlled laboratory experiment to test the hypotheses, in which all participants were divided into two groups, one group used the user interface with one stitched video to participate the experiment, and the other group used the user interface with two separate videos to take part in the experiment. Note that with the between-subject setting, each participant was only tested on one condition, such that we can avoid interference between the experiments of using two user interfaces (MacKenzie, 2012). Participants remotely drove a telepresence robot using the user interface with one stitched live video or two separate live videos to complete two tasks, i.e., tele-exploration and telecommunication. In both tasks, a participant played a role as the remote operator. In the telecommunication task, an experimenter was located in the local environment to communicate with the participant. We denote the experimenter as "the local person".

4.1. Telepresence robot

We have developed in our lab a telepresence robot, called Mcisbot, with a Touchable live video Image-based User Interface (TIUI) (Jia et al., 2015), as depicted in Figure 1(a). The Mcisbot robot contains a mobile robot base and a telepresence robot head. It uses the Pioneer 3-AT as the mobile robot base, and a special robot head is used for displaying the remote operator for local users. The robot head consists of a light LCD screen, a forward-facing camera (FF-camera), a downward-facing camera (DF-camera), and a speaker & microphone, and all together are mounted on a pan-tilt platform held up by a vertical post. The FF-camera with a wide-angle lens can provide a live video for a clear watching of targets or persons in front, allowing the remote operator to telecommunicate with local users or teleoperate the local objects in a way similar to face-to-face interaction. The DF-camera with a fisheye lens locates the robot near the center of the DF-video which provides a complete watching of the ground around the robot base for safe teledriving. The vertical post can be moved to change the robot's height from 1200 *mm* to 1750 *mm*, which is similar to the heights from a pupil to an adult.

In our user study, we limited the height of the robot in a proper value that could provide a clear and full view on both the forward scene and the downward scene. During two tasks, we kept the robot in a consistent height for all participants. Generally, a remote operator could rotate the robot head to watch around since there is a pan-tilt platform in the Mcisbot, but we closed this function for simplicity. As a consequence, all participants needed to turn the robot to change the views.

We used a tablet with the TIUI to remotely drive the robot through wireless communication networks. The TIUI is a new user interface which allows the remote operator to teleoperate the telepresence robot by directly touching the live video images with specific touch gestures. Note that all participants were given enough time to familiarize themselves with the TIUI before the user study. They could touch anywhere of the live video-based user interface to drive the robot. The user interfaces with two separate videos and one stitched video used in our study are shown in Figure 1(b,c), respectively. Videos in the user interface with two videos are original videos captured from the forward-facing camera and the downward-facing camera. "Two separate videos" means two videos without stitching rather than two videos 'separated' by a visible border or spacing. The border between the two videos is hidden, but there still exists a clear boundary between the two videos since illumination in these two videos are different. The stitched live video was stitched from the FF-video and the DF-video through our previous wide-angle and fisheye video stitching algorithm (Dong et al., 2019). The green arrows indicate the current moving direction of the robot.

4.2. Participants

We randomly recruited 18 participants (8 females and 10 males) from the local university for the user study, whose ages range from 20 to 28 (M = 23.39, SD = 1.98). We randomly divided them into two groups, i.e., nine participants for each condition, and each with four females and five males. With a seven-item Likert scale, ranging from "1 = never" to "7 = often", all participants reported their frequency of using tablets or smartphones (M = 6.89, SD = 0.31), and chatting with families or friends through the live video (M = 5.11, SD = 1.63). All participants were also required to tell us their experience of playing online role controlling games (M = 3.83, SD = 2.45), and most of them used a smart mobile device such as tablets or smartphones to play these games. Experience of driving a mobile robot was also reported with the same seven-item Likert scale of frequency (M = 0.83, SD = 0.60), ranging from "1 = never" to "7 = often". Before our user study, some participants had never heard of mobile robotic telepresence systems, and many of the others had no experience in telepresence robot operation. Some participants had experience of controlling a small robot (quite different from telepresence robots) with a mechanical remote controller.

4.3. Environment setup

We chose a real lab as the local environment. A participant could remotely drive the telepresence robot as his/her embodiment to actively attend a group meeting, or find persons and communicate with them in the local environment. There are many office tables, chairs, plants, and clapboards in the local environment. The furniture provides the direction to participants to teleoperate the robot to explore the office, and it also plays a role as obstacles. The road for robot walking is limited in width, and the participant must be careful to avoid a collision to keep the robot safe, even though the Mcisbot robot is equipped with anticollision sonars. During the user study, the user interfaces with one stitched live video and two separate live videos were used interchangeably. Thus, the number of people, location, and movements are basically consistent between the two conditions. The people in the local environment were doing their job as usual, and they did not know what user interface the participant was using.

During the user study, participants were located in another room, specified as the remote environment, to remotely drive the robot through a tablet to explore the office, find someone,

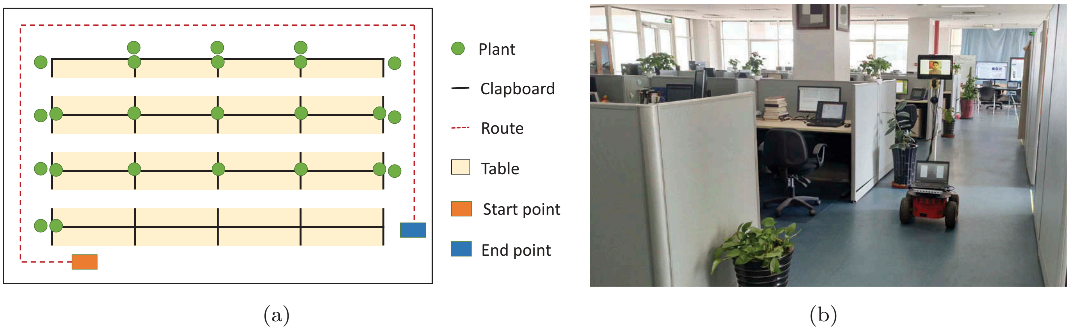

Figure 2. The local environment. (a) The physical arrangement of the local environment. The red-dotted line indicates the route from the start point to the end point for the tele-exploration task (i.e., tele-exploring the office to find all plants). (b) The robot is retemoly driven by a participant in the remote site to explore the environment.

and communicate with him/her. The remote environment is isolated from the local environment. The physical arrangement of the local environment is depicted in Figure 2(a), and a picture of a participant teledriving the robot to explore the office is shown in Figure 2(b).

4.4. Tasks

In order to compare the difference of the user interface with one stitched live video and two separate live videos on tele-interaction, we assigned all participants with two tasks, i.e., tele-exploration and telecommunication.

The tele-exploration task is to remotely drive the Mcisbot robot to explore the office to find all the plants in it, draw the plants' correct locations on a layout map of the local environment, and then go to the end point. The route from the start point to the end point for the exploration is given in Figure 2(a), i.e., the red-dotted line in the figure. The layout map of the local environment used in the user study is shown in Figure 3, and the map only shows the tables, clapboard, the start point, and the end point of the task. Note that the plants are distributed everywhere of the local environment, for example, on the ground and on the clapboards, as shown in Figure 2(b). Most of the plants are occluded by the clapboards and are not visible unless the participant bypassed the clapboards. Since the layout map does not show any plants in the local environment, and the tele-exploration task is a timing task, participants had to explore the local environment to find all plants as quickly as possible. In the task, all participants were required to keep the robot safe without bumping into anything in the local environment, e.g., plants, chairs, tables, and clapboards. Meanwhile, they had to keep the robot away from people walking or standing in the local environment since the robot was "walking" in a real office (many people were working there). Participants were encouraged to finish the tele-exploration as quickly and accurately as possible by a reward, and the time for this task was not limited for every participant.

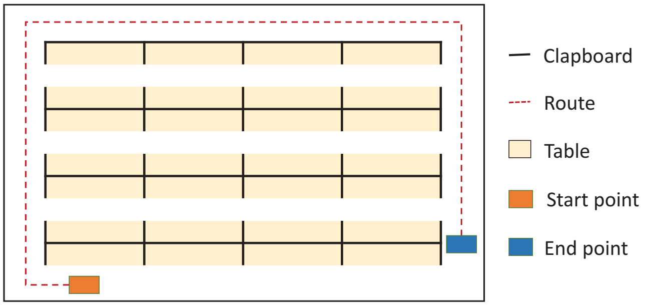

Figure 3. The layout map of the local environment used for the tele-exploration task. Only tables, clapboards, the start point, and the end point are shown on the map.

Since the tele-exploration task is a timing task, participants possibly could not experience carefully of the user interface with one stitched live video and two separate live videos, but concentrated on finding all plants. Therefore, we assigned every participant with the other task, i.e., the telecommunication task, which is to drive the Mcisbot robot to find and communicate with a local person. In this task, the participants needed to find the specific local person, approach him/her, communicate with him/her, and take a walk with him/her through the specific user interface (with one stitched live video or two separate live videos). The local person was working on his/her seat for all participants, and he/she did not know which user interface the participant was using. We kept the local person unchanged for all participants. During telecommunication task, the local person did not talk about anything about the mobile robotic telepresence system, thus, participants' answer to the questionnaire would not be affected by the local person. After communicating and taking a walk with the local person, the participant could actively wander around to further experience the mobile robotic telepresence system with the specific user interface. The time for this task was limited up to 10 min, including about 5 min for the participant to be together with the local person.

4.5. Measures

To evaluate the performance and remote operators' experience of using the user interfaces with one stitched live video and two separate live videos, we utilized both objective and subjective measures. The objective measures include task completion time and the number of errors. The subjective measures include task perceived task success, perceived ease of using the user interface, concentration on the task, and feelings of presence. Subjective measures were acquired from two post-task questionnaires, in which the questions were designed following the related works (Johnson et al., 2015; Kiselev et al., 2014; Rae, Mutlu, & Takayama, 2014) on effects of field of view.

The objective measures of the tele-exploration task include task completion time and the number of errors, by using one stitched live video and two separate live videos, respectively. Task completion time is timed from when the robot left the start point to when the robot finished the task and stopped at the end point. The number of errors is the sum of the number of plants that are drawn on the wrong places and the number of plants that are not found. The number of collisions that the telepresence robot bumping into something is also recorded to measure situation awareness during the tele-exploration task. The subjective measures include perceived task success, perceived ease of using the user interface, concentration on the task, and feelings of presence. We administered a post-task questionnaire to acquire the subjective measures. The questions in the questionnaire include “I felt that the task performance was the best, i.e., both the task completion time and the number of errors were the least,” “I felt that it was easy to navigate through the user interface,” “I could concentrate on the task without frequently switching attention between navigating and searching plants,” “I felt that exploring the office to find all plants was as if I was physically being in the office.” Each question is with a seven-point scale, ranging from “1 = Strongly Agree” to “7 = Strongly Disagree”.

In the telecommunication task, we only investigated subjective measures through another post-task questionnaire since participants could wander around as they like. The subjective measures contain perceived ease of using the user interface, concentration on the task, and feelings of presence. The questions include “I felt that it was easy to find the local person, approach him/her, and take a walk with him/her through the user interface,” “I could concentrate on the local person, and did not need to frequently switch attention from looking at the ground to looking at the local person,” and “I felt that finding the local person, approaching him/her, telecommunicating and taking a walk with him/her were as if I was physically being with the local person.” Each question is with a 7-point scale, ranging from “1 = Strongly Agree” to “7 = Strongly Disagree”.

We listed the corresponding relationship among hypothesis, measures, and tasks in Table 1 for a clear view. The questions for subjective measures on the tele-exploration task and the telecommunication task are similar. For example, in the tele-exploration task, the question corresponds to the measure of concentration on the task is ‘I could concentrate on the task without frequently switching attention between navigating and searching plants’, and the corresponding question in the telecommunication task is ‘I could concentrate on the local person, and did not need to frequently switch attention from looking at the ground to looking at the local person.’ The tele-exploration task is to explore the office to find all the plants in it. During the exploration, a remote operator has to switch attention among navigation and searching for plants. The telecommunication task is to drive the Mcisbot robot in the office to find and communicate with a local person. Here, a remote operator has to switch attention between driving the robot and looking at the local person when the remote operator was taking a walk with the local person. Thus, the questions of the two tasks are similar and only the targets are different, i.e., searching for plants in the tele-exploration task while searching for the local person or looking at the local person in the telecommunication task.

4.6. Procedure

The procedure contains three stages, i.e., preparation, experiment, and interview. We randomly divided the participants into two groups for using the user interfaces with one stitched live video and two separate live videos, respectively. Then, the two groups of participants alternately participated in the experiment using the corresponding user interface.

In the preparation stage, every participant was asked to fill in a pre-task questionnaire for obtaining demographic information and then given an overview of the experiment. An experimenter instructed each participant how to use the TIUI with one stitched live video or two separate live videos to drive the Mcisbot robot. Every participant had up to 25 min to get familiar with the TIUI and driving skills, under the help of the experimenter. The familiarization session was also used to minimize the bias of participants’ experience. When 25 min had elapsed or participants indicated that they were ready for the experiment, the participant stopped the familarization, and the experimenter disconnected the robot. We recorded the familiarization time of each participant for further analysis after the experiment. The time is about 5 to 25 min for each participant.

After the instruction, it is the evaluation of the effects on the tele-exploration task. Driving the robot to the start point, the experimenter reconnected the system and told the participant to find all plants in the office and draw their correct locations on the layout map (Figure 3), and then started the timer. Every participant was encouraged to complete the task as quickly and accurately as possible by receiving an extra dollar as a reward. They were required to keep the robot safe without bumping into anything in the office (e.g., plants, chairs, tables, and clapboards), and keep the robot away from people that walking or standing in the office, because the robot was “walking” in a real office, where many people were working. When the participant finished the task and drove the robot to the end point, the experimenter stopped the timer, disconnected the robot, and then asked the participant to fill in the corresponding post-task questionnaire to evaluate the perceived task success, perceived ease of using the user interface, concentration on the task, and feelings of presence. During the experiment, the experimenter recorded the task completion time, the number of errors, and the number of collisions that the robot bumping into anything.

After filling in the post-questionnaire of the tele-exploration task, we carried on to evaluate the effects on the telecommunication task. At the begining, we made the local person and the participant meet each other in the remote environment, and then the local person came back to his/her seat for working. After the robot was reconnected, the participant started to drive the robot away from the start point to find the local person. When the robot was close to the local person, the participant made a video call through the

TIUI, waited for the response, and then communicated with the local person through the stitched live video or separate live videos. After video communication, the local person asked the participant to take a walk with him/her, and then the participant could actively wander around the office to further experience the mobile telepresence system with the specific user interface. When the participant indicated to finish or 10 min had elapsed, the experimenter disconnected the robot, and then provided the participant with the corresponding post-task questionnaire to evaluate the perceived ease of using the user interface, concentration on the local person, and feelings of presence. Note that the post-task questionnaire was administered after each of two tasks, and every participant first participated in the experiment with the tele-exploration task and then the telecommunication task. In our study, there are no order effects on this procedure since the tasks were not compared to each other.

Finally, an experimenter interviewed every participant for gathering their comments and suggestions on the video(s) in the user interface. The interview took about 15 min. For each participant, the time cost of the whole experiment is about 70–100 min, from the preparation stage to the end of the interview. To preserve the experiment process for further analysis, the experimenter filmed the process of the two tasks, including the Mcisbot robot's behavior in the local environment and the participant behavior on the tablet. After the user study, each participant of the user study was paid about $ 22 USD, including the extra dollar.

4.7. Analysis methods

We ran a simple t-test analysis on the practice time, task performance (including task completion time and the number of errors), perceived task success, and checked the effects of the user interfaces with one stitched live video and two separate live videos. We compared the average scores of the two user interfaces (the average score of using the user interface with one stitched live video minus the average score of using the user interface with two separate live videos, denoted as $M_{one-two}$), and tested whether the difference of the two user interfaces was zero or not. The number of collisions was not analyzed since the robot's bumping into something was almost not happen during the whole user study. We conducted the one-way fixed-effects analysis of variance (ANOVA) to test the difference between the user interfaces with one stitched live video and two separate live videos on participants' concentration on the task, perceived ease of using the user interface, and feelings of presence. The input variable is the user interfaces with one stitched live video and two separate live videos. For the test of statistical significance, we used a cutoff value of $p < 0.05$.

Before analyzing the difference between the stitched live video and separate live videos, we compared the demographic information of the two groups, including average age, gender, experience of using tablets or smartphones, average frequency of chatting with families or friends through video, experience of playing online role controlling games, and experience of driving a mobile robot. We did not find any significant difference on these variables ($p > .05$) between the two groups.

5. Results

Our first hypothesis predicted that the task performance (including task completion time and the number of errors) would be improved by using the user interface with one stitched live video compared to the user interface with two separate live videos. We found complete support for this hypothesis from our data. The average task completion time of the tele-exploration task in seconds are 651.5 (SD =185.26) and 1068.25 (SD =285.97) for one stitched live video and two separate live videos. We analyzed the practicing time to further check the effects on task completion time. As shown in Figure 4(a), the average practicing time of participants using one stitched live video and two separate live videos are 863.33 (SD =286.89) and 827.11 (SD =203.36), respectively. There is no difference between the two groups in practicing time ($M_{one-two} = 36.22$), $t(16) = .29$, $p = .775$, $r = .07$, which means that the effects of practicing time on task completion time is similar. We analyzed the average task completion time by using the two user interfaces ($M_{one-two} = -158.56$), and found that there is a significant effect, $t(16) = -3.69$, $p = .004$, $r = .68$, as

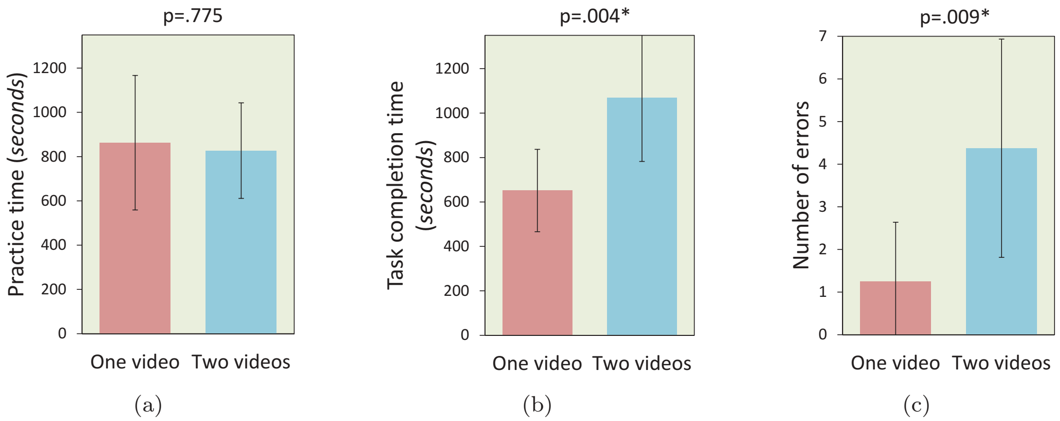

Figure 4. Average of practice time (a), task completion time (b), and the number of errors (c) by using the user interfaces with one stitched live video and two separate live videos.

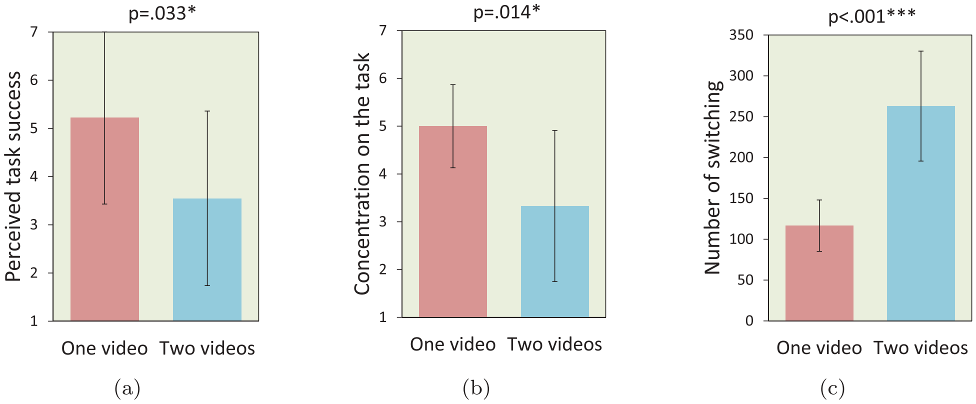

Figure 5. Average of perceived task success (a), concentration on the task (b), and the number of attention switching (c) on the tele-exploration task by using the user interfaces with one stitched live video and two separate live videos.

shown in Figure 4(b). In summary, the task completion time was improved by using the user interface with one stitched live video. For the number of errors, we also found support on the hypothesis ($M_{one-two} = -3.13$), $t(16) = -3.24$, $p = .009$, $r = .63$, as shown in Figure 4(c), and the average error of participants using one stitched live video and two separate live videos are 1.25 (SD =1.39) and 4.38 (SD =2.56), respectively. Note that the error bars show the standard deviation of the data. (***) and (*) denote $p < .001$ and $p < .05$, respectively. “One video” means the user interface with one stitched live video, and “Two videos” indicates the user interface with two separate live videos.

The second hypothesis predicted that the perceived task success would be increased by using the user interface with one stitched live video compared to the user interface with two separate live videos. We found support for this hypothesis ($M_{one-two} = 1.66$), $t(16) = 1.97$, $p = .033$, $r = .44$, as shown in Figure 5(a).

Our third hypothesis predicted that the remote operator would be more concentrated on the task by using the user interface with one stitched live video compared to the user interface with two separate live videos. We analyzed the difference between one stitched live video and separate live videos on each task. As shown in Figure 5(b), participants that used the user interface with one stitched live video were more concentrated on the tele-exploration task than those using the user interface with two separate live videos, $F(1, 16) = 7.69$, $p = .01$, $\eta^2 = .32$. Similarly, we found an extremely significant effect on the telecommunication task, $F(1, 16) = 32$, $p < .001$, $\eta^2 = .52$, as Figure 7(a) shown. Therefore, the hypothesis is completely supported by our data. We annotated the recorded videos of faces of all participants and count the number of attention switching of each participant in the tele-exploration task. The average number of attention switching of participants that used the user interface with two separate videos is $263(SD = 67)$. The corresponding data for participants using the user interface with one stitched video is $117(SD = 31)$. We analyzed these data and found that they are significantly different, $M_{one-two} = -146$, $t(16) = -5.96$, $p < .001$, $r = .83$, as shown in Figure 5(c), which further supports our third hypothesis, i.e., remote operators that used the user interface with one stitched video could be more concentrated on the task.

The fourth hypothesis predicted that the perceived ease of using the user interface on completing the task would be improved through one stitched live video compared to two separate live videos. We found partial support for this hypothesis. We did not find a significant difference in the tele-exploration task, $F(1, 16) = 2.5$, $p = .133$, $\eta^2 = .14$, but significant effects in the telecommunication task was found, $F(1, 16) = 5.70$, $p = .030$, $\eta^2 = .26$, as shown in Figures 6(a) and 7(b), respectively.

The fifth hypothesis predicted that the remote operator would feel more present in the local environment by using the user interface with one stitched live video compared to the user interface with two separate live videos. In the tele-exploration task, the user interface with one stitched live video increased the remote operators’ feelings of presence compared to the user interface with two separate live videos, $F(1, 16) = 5.94$, $p = .027$, $\eta^2 = .27$, as shown in Figure 6(b). From Figure 7(c) we found that similar effects appear in the telecommunication task, $F(1, 16) = 6.38$, $p = .022$, $\eta^2 = .29$. Therefore, the hypothesis is completely supported by our data.

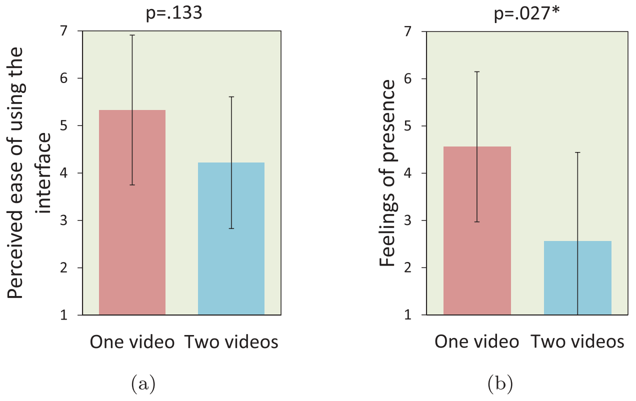

Figure 6. Average perceived ease of using the user interfaces with one stitched live video and two separate live videos (a) and feelings of presence (b) on the tele-exploration task.

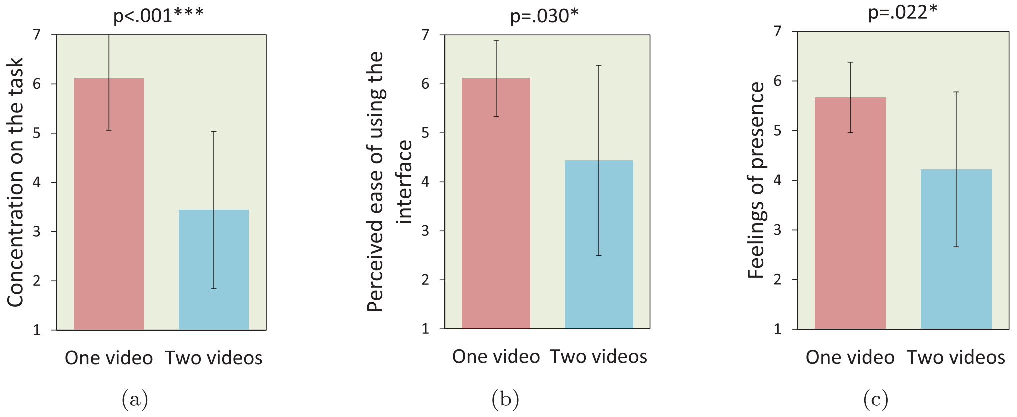

Figure 7. Average subjective measures on the telecommunication task. (a) Concentration on the task. (b) Perceived ease of using the user interfaces with one stitched live video and two separate live videos. (c) Feelings of presence.

For a clear comparison, we list the hypotheses and the corresponding experimental results on Tables 2 and 3 for the tele-exploration task and the telecommunication task, respectively. There are five hypotheses for the tele-exploration task and three hypotheses for the telecommunication task. Note that "P-value" is the p value between the user interfaces on the corresponding measure. As showing in Table 2, the user interfaces with one stitched live video has significant effects on the tele-exploration task in terms of task completion time, perceived task success, concentration on the task, and feelings of presence. In the telecommunication task, significant effects appear on concentration on the task, perceived ease of using the interface, and feelings of presence, as shown in Table 3.

Table 1. Corresponding relationship among hypotheses, measures, and tasks.

Task	Hypothesis	Measure
Tele-exploration	1. Performance	Task completion time Number of errors
	2. Perceived task success	"I felt that the task performance was the best, i.e., both the task completion time and the number of errors were the least."
	3. Concentration on the task	"I could concentrate on the task without frequently switching attention between navigating and searching plants."
	4. Perceived ease of using the interface	"I felt that it was easy to navigate through the user interface."
	5. Feelings of presence	"I felt that exploring the office to find all plants was as if I was physically being in the office."
Telecommunication	3. Concentration on the task	"I could concentrate on the local person, and did not need to frequently switch attention from looking at the ground to looking at the local person."
	4. Perceived ease of using the interface	"I felt that it was easy to find the local person, approach him/her, and take a walk with him/her through the user interface."
	5. Feelings of presence	"I felt that finding the local person, approaching him/her, telecommunicating and taking a walk with him/her were as if I was physically being with the local person."

Table 2. Hypotheses and experimental results of the tele-exploration task.

Hypothesis	Measure	One video	Two videos	*P*-value
1	Task completion time ↓	651.5	1068.25	.004*
	Number of errors ↓	1.25	4.38	.009*
2	Perceived task success ↑	5.22	3.56	.033*
3	Concentration on the task ↑	5	3.33	.014*
4	Perceived ease of using the interface ↑	5.33	4.22	.133
5	Feelings of presence ↑	4.56	2.56	.027*

" ↑ " indicates the smaller value the better, and " ↓ " means the larger value the better.

Table 3. Hypotheses and experimental results of the telecommunication task.

Hypothesis	Measure	One video	Two videos	*P*-value
3	Concentration on the task ↑	6.11	3.44	< .001***
4	Perceived ease of using the interface ↑	6.11	4.44	.030*
5	Feelings of presence ↑	5.67	4.22	.022*

" ↑ " indicates the smaller value the better.

6. Discussions

Our results confirms that task efficiency, perceived task success, concentration on the task, and feelings of presence are improved by using the user interface with one stitched live video compared to the user interface with two separate live videos. From the videos that record each participant' behavior during the task, we observed that participants using the user interface with two separate live videos frequently switched their attention from looking at the DF-video (to look at the robot's surroundings for safe driving) to looking at the FF-video (to look around), and then switched back to looking at the DF-video. In contrast, the number of times to switch attention by using the user interface with one stitched live video is smaller. We owe this improvement to the stitching of the two videos, from which participants can directly perceive a full view of the local environment and are able to look forward during teledriving.

In the tele-exploration task, our results show that task efficiency, the number of errors, perceived task success, concentration on the task, and feelings of presence are significantly improved by using the user interface with one stitched live video. We observed that participants using the user interface with one stitched live video could directly draw the plants, according to what they saw from the user interface. But participants using the user interface with two separate live videos often checked whether the plants was drawn or not, especially when one plant was simultaneously visible in the FF-video and the DF-video, or if there exists similar plants in the scene and they were simultaneously visible in the FF-video and the DF-video. On perceived ease of using the interface, there is no significant difference between one stitched live video and two separate live videos. Participants said that, during the task, they concentrated on driving the robot and searching for the plants since they were encouraged to finish the tele-exploration as quickly and accurately as possible by receiving a reward. They did not pay extra attention to consider whether the user interface was easy to use or not.

In the telecommunication task, remote operators' concentration on task and feelings of presence are improved by using the user interface with one stitched live video, similar to that in the tele-exploration task. The perceived ease of using the interface is also improved by using one stitched live video, different from the tele-exploration task. Participants using the user interface with one stitched live video reported that they were easy to get close to the local person since they could directly see how far the local person was from the robot. Participants using the user interface with two separate live videos said that they could easily approach the local person through the DF-video only when they could see the local person's feet from the DF-video. But when the local person was far away from the robot (the local person only appears in the FF-video), they could not determine the distance very well.

During the interview stage, we asked every participant some questions for gathering their feedback on using the user interfaces with one stitched live video and two separate live videos. Almost all participants using the user interface with two separate live videos suggested stitching the two videos together, they reported "It might be more convenient and easier to use the systems if the two videos could be stitched together." They also said "I needed to frequently switch attention between the two videos during both tasks to look forward and look at the ground," "I needed to take time to adapt to the videos after switching between them," "I needed to find the spatial relationship between the FF-video and the DF-video," "I could only watch one video (the FF-video or the DF-video) at a time, and it was easy to miss plants and ignore obstacles in high places," "The scenes in the FF-video and the DF-video were in a different scales, it was different from the real scene," "I paid a lot of time on looking at the DF-video, and needed to stop driving the robot to search for the plants," and "Remotely perceiving the scene from the two videos was quite different from seeing the scene by physically standing in the local environment." There is only one participant using the user interface with two separate live videos reported "There might be no difference for me to finish the telecommunication task by using one stitched live video and two separate live videos, but the user interface with two separate live videos is inconvenient for the tele-exploration task." Participants using the user interface with one stitched live video are satisfied with the scene shown in the video. They said "It is easy to watch around through the user interface with one stitched live video," "Since the scene of the stitched live video is similar with the real scene, I could adapt to the scene in the video very quickly," "The user interface is user-friendly, and the scene is very intuitive," and "I occasionally needed to switch my attention from the upper part of the stitched live video to the lower part of the video, but it was infrequently." We asked whether these participants would like to use the user interface with two separate live videos to complete the two tasks or not. Almost all participants said that it would be difficult to watch around with two separate live videos compared to one stitched live video. Only one participant said "It might be no difference for me in completing the two tasks by using one stitched live video and two separate live videos."

7. Implications

Most existing mobile robotic telepresence systems are incorporated with an FF-camera and a DF-camera. For a clear comparison, we list the relative features of some systems in Table 4, including the lens type of cameras, amount of videos shown on the user interface (denoted as "Amount"), and whether videos are always shown on the user interface or not (denoted as "Always on"). It seems that commercial products tend to always display the FF-video and the DF-video on the user interface for increasing spatial awareness and no need to pay time to manually switch the two videos (Double2, 2018). However, remote operators still need to switch their attention from one video to the other or switch back for different purposes, since the local environment is shown on two separate live videos, and its spatial structure is also displayed on two parts. According to our results, we believe that, providing a user interface with one stitched live video, commercial mobile robotic telepresence systems will enable remote operators to be more concentrated on the work they are doing, or it would be more similar to physically working in the local environment.

Remote operators' feelings of presence is an extremely important property for remote communication (Song, Kim, & Park, 2019). Better feelings of presence indicates that the local environment displayed on the user interface is more similar to the real scene, such that, remote operators would not need to pay time to familiarize themselves with the scene displayed on the user interface, different from that in (Boll, 2017). Our results verify that remote operators' feelings of presence are significantly improved by using the user interface with one stitched live video compared to two separate live videos. The implication is to display one video with a larger field of view on the user interface. For looking

Table 4. Comparison of the state of the art vs. experiment conditions.

System	FF-camera	DF-camera	Amount	Always on
QB (QB, 2018)	Standard	Fisheye	2	No
Double2 (Double2, 2018)	Wide-angle	Standard	2	Yes
BeamPro (BeamPro, 2018)	Wide-angle	Wide-angle	2	Yes
Ours with two videos	Wide-angle	Fisheye	2	Yes
Ours with one video	Wide-angle	Fisheye	1	Yes

forward with a larger field of view, future work can use two or more conventional cameras (or wide-angle cameras without radial distortion), and then stitching them to achieve telecommunicating or teleoperating in a way more similar to face-to-face interaction. And navigating in a full view of the robot's surroundings can be achieved by using a fisheye lens camera or two more wide-angle cameras. To obtain a larger field of view, including the horizontal and vertical field of view, future work could stitch multiple videos captured from these cameras, by using real-time video stitching algorithms.

With a limited space in a tablet for showing one stitched video of a larger field of view, the scene shown in the user interface becomes smaller. This implies that one stitched live video can be offered as an option in some situations, or to use zoom cameras instead. For example, when remote operators want to walk around to find somebody or something, or to wander around (similar to that in (Boll, 2017; Neustaedter et al., 2016; Rebola & Eden, 2017)), the stitched live video is beneficial for remote operators to search for the target while keeping the robot safe with teleoperating efficiently, especially in a new environment. When remote operators want to focus on something (the speaker or lecture), they can use the specific video (the FF-video or the DF-video) or zoom in the stitched live video. To zoom in the stitched live video or to switch the stitched live video and the specific video is a choice for designers.

8. Limitations and future work

There are several limitations of our user study on investigating the difference between the user interfaces with one stitched live video and two separate live videos. First, we investigated the difference between the user interfaces in the office place, in which the scale of the local environment and the number of people are limited. Future work can explore the effects on larger and crowded scenes, such as museums, classroom, and meeting room. Second, we investigated the effects on a tele-exploration task and a telecommunication task, during which remote operators are going to find plants in the local environment and telecommunicate with local users. The effects of one stitched live video and two separate live videos on richer tasks can be explored in future research, for instance, collaboration tasks that need collaboration between remote operators and local users (Rae et al., 2014), group activities that need remote operates to communicate or interact with different people or multiple people (Berisha et al., 2015), and the task to tele-interact with smart devices in intelligent home (Shen, Xu, Pei, & Jia, 2016).

In future studies, we also can investigate how different features of the mobile robotic telepresence systems affect the visualization requirements. In our study, we fixed the robot head while the robot head mounted on a pan-tilt platform can be rotated to see around. Future studies can enable this function to allow remote operators to "turn" the robot's head and investigate the effect of the user interface with one stitched live video compared to two separate live videos.

9. Conclusion

In this paper, we presented the study of using the live video-based user interface with one stitched live video for robotic telepresence systems. The stitched live video, generated by stitching two live videos captured from a forward-facing camera and a downward-facing camera, can provide a large view of one local environment for remote operators. As a consequence, remote operators can directly perceive the local environment from only one video shown on the user interface. We conducted a user study through a between-subject controlled laboratory experiment, consisting of a tele-exploration task and a telecommunication task, to investigate the difference between the user interfaces with one stitched live video and two separate live videos. The analyses of the experimental data show that task performance and perceived task success are improved by using the user interface with one stitched live video in the tele-exploration task. The effects on the perceived ease of using the user interface are proved to be significant in the telecommunication task. Furthermore, remote operators' feelings of presence and concentration on the task are significantly improved by using the user interface with one stitched live video in both tasks.

Funding

This work was supported in part by the Natural Science Foundation of China (NSFC) under Grants No. 61773062 and No. 61702037, Beijing Municipal Natural Science Foundation under Grant No. L172027.

ORCID

Yuwei Wu http://orcid.org/0000-0002-0263-925X

References

BeamPro. (2018, July 31). Retrieved from https://suitabletech.com

Berisha, A., Kölle, R., & Griesbaum, J. (2015). Acceptance of telepresence robots during group work. In Proceedings of the 14th International Symposium of Information Science (pp. 350–356), Zadar, Croatia.

Boll, S. (2017). Multimedia at CHI: Telepresence at work for remote conference participation. *IEEE MultiMedia*, *24*(3), 5–9. doi:10.1109/MMUL.2017.3051516

Cesta, A., Cortellessa, G., Orlandini, A., & Tiberio, L. (2016). Long-term evaluation of a telepresence robot for the elderly: Methodology and ecological case study. *International Journal of Social Robotics*, *8*(3), 421–441. doi:10.1007/s12369-016-0337-z

Cha, E., Chen, S., & Mataric, M. J. (2017). Designing telepresence robots for K-12 education. In IEEE International Symposium on Robot and Human Interactive Communication (pp. 683–688). doi: 10.1177/17531934177711594

Chen, J. Y., Haas, E. C., & Barnes, M. J. (2007). Human performance issues and user interface design for teleoperated robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and reviews)*, *37*(6), 1231–1245. doi:10.1109/TSMCC.2007.905819

Dong, Y., Pei, M., Zhang, L., Xu, B., Wu, Y., & Jia, Y. (2019). Stitching videos from a fisheye lens camera and a wide-angle lens camera for telepresence robots. *arXiv:1903.06319*.

Double2. (2018, July 31). Retrieved from http://www.doublerobotics.com

Giraff. (2018, July 31). Retrieved from http://www.giraff.org/

Heshmat, Y., Jones, B., Xiong, X., Neustaedter, C., Tang, A., Riecke, B. E., & Yang, L. (2018). Geocaching with a Beam: Shared outdoor activities through a telepresence robot with 360 degree viewing. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 359: 1–359:13). Montreal, QC, Canada: ACM.

Hestand, D., & Yanco, H. A. (2004). Layered sensor modalities for improved human-robot interaction. In 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583) (Vol. 3, pp. 2966–2970), The Hague, Netherlands.

Jia, Y., Xu, B., Shen, J., Pei, M., Dong, Z., Hou, J., & Yang, M. (2015). Telepresence interaction by touching live video images. *arXiv preprint arXiv:1512.04334*.

Johnson, S., Rae, I., Mutlu, B., & Takayama, L. (2015). Can you see me now? how field of view affects collaboration in robotic telepresence. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 2397–2406). Seoul, Republic of Korea: ACM.

Keyes, B., Casey, R., Yanco, H. A., Maxwell, B. A., & Georgiev, Y. (2006). Camera placement and multi-camera fusion for remote robot operation. *IEEE International Workshop on Safety Security & Rescue Robotics*. Gaithersburg, MD, USA.

Kiselev, A., Kristoffersson, A., & Loutfi, A. (2014). The effect of field of view on social interaction in mobile robotic telepresence systems. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 214–215). Bielefeld, Germany: ACM.

Koceski, S., & Koceska, N. (2016). Evaluation of an assistive telepresence robot for elderly healthcare. *Journal of Medical Systems*, *40*(5), 121. doi:10.1007/s10916-016-0481-x

Kristoffersson, A., Coradeschi, S., & Loutfi, A. (2013). A review of mobile robotic telepresence. *Advances in Human-Computer Interaction*, *2013*, 3.

Lazewatsky, D. A., & Smart, W. D. (2011). A panorama interface for telepresence robots. In *Proceedings of the 6th International Conference on Human-robot Interaction* (pp. 177–178). ACM. doi: 10.1177/1753193411427832

MacKenzie, I. S. (2012). *Human-computer interaction: An empirical research perspective*. Waltham, MA, USA: Newnes.

Neustaedter, C., Venolia, G., Procyk, J., & Hawkins, D. (2016). To beam or not to beam: A study of remote telepresence attendance at an academic conference. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 418–431). ACM. doi: 10.1017/thg.2016.53

Nielsen, C. W., Goodrich, M. A., & Ricks, R. W. (2007). Ecological interfaces for improving mobile robot teleoperation. *IEEE Transactions on Robotics*, *23*(5), 927–941. doi:10.1109/TRO.2007.907479

QB. (2018, July 31). Retrieved from https://www.anybots.com

Rae, I., Mutlu, B., & Takayama, L. (2014). Bodies in motion: Mobility, presence, and task awareness in telepresence. In Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (pp. 2153–2162), Toronto, Ontario, Canada.

Rae, I., & Neustaedter, C. (2017). Robotic telepresence at scale. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 313–324). Denver, CO, USA: ACM.

Rae, I., Venolia, G., Tang, J. C., & Molnar, D. (2015). A framework for understanding and designing telepresence. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (pp. 1552–1566). ACM. doi: 10.3168/jds.2014-8796

Rebola, C. B., & Eden, G. (2017). Remote robotic disability: Are we ready for robots? *Interactions*, *24*(3), 48–53. doi:10.1145/3086450

Rudolph, A., Vaughn, J., Crego, N., Hueckel, R., Kuszajewski, M., Molloy, M., & Shaw, R. J. (2017). Integrating telepresence robots into nursing simulation. *Nurse Educator*, *42*(2), E1–E4. doi:10.1097/NNE.0000000000000329

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, *10*(3), 266–281. doi:10.1162/105474601300343603

Shen, J., Xu, B., Pei, M., & Jia, Y. (2016). A low-cost tele-presence wheelchair system. In 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 2452–2457), Daejeon, Korea.

Shiroma, N., Sato, N., Chiu, Y., & Matsuno, F. (2004). Study on effective camera images for mobile robot teleoperation. In Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication (pp. 107–112). doi: 10.1016/j.alcohol.2004.06.001

Song, H., Kim, J., & Park, N. (2019). I know my professor: Teacher self-disclosure in online education and a mediating role of social presence. *International Journal of Human–Computer Interaction*, *35*(6), 448–455. doi:10.1080/10447318.2018.1455126

Thomas, L. C., & Wickens, C. D. (2000). Effects of display frames of reference on spatial judgments and change detection (Tech. Rep.). Illinois Univ at Urbana-Champaign Savoyaviation Research Lab.

Tsui, K. M., Dlphond, J. M., Brooks, D. J., Medvedev, M. S., McCann, E., Allspaw, J., … Yanco, H. A. (2015). Accessible human-robot interaction for telepresence robots: A case study. *Journal of Behavioral Robotics*, *6*(1), 1–29.

Tsui, K. M., & Yanco, H. A. (2013). Design challenges and guidelines for social interaction using mobile telepresence robots. *Reviews of Human Factors and Ergonomics*, *9*(1), 227–301. doi:10.1177/1557234X13502462

VGO. (2018, July 31). Retrieved from http://www.vgocom.com

Voshell, M., Woods, D. D., & Phillips, F. (2005). Overcoming the keyhole in human-robot coordination: Simulation and evaluation. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *49*(3), 442–446. doi:10.1177/154193120504900348

Woods, D. D., Tittle, J., Feil, M., & Roesler, A. (2004). Envisioning human-robot coordination in future operations. *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, *34*(2), 210–218. doi:10.1109/TSMCC.2004.826272

Yanco, H. A., & Drury, J. (2004). “Where Am I?” Acquiring situation awareness using a remote robot platform. In 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583) (Vol. 3, pp. 2835–2840), The Hague, Netherlands.

Yang, L., Jones, B., Neustaedter, C., & Singhal, S. (2018). Shopping over distance through a telepresence robot. *Proceedings of the ACM on Human-Computer Interaction*, *2* (CSCW), 191:1–191: 18. doi: 10.1145/3274460

About the Authors

Yanmei Dong received the B.S. and M.S. degrees from Beijing Institute of Technology (BIT), Beijing, China, in 2013 and 2015, respectively. She is currently pursuing a Ph.D. degree in the School of Computer Science, under the supervision of Prof. Mingtao Pei and Yunde Jia.

Yunde Jia (M’11) received the B.S., M.S., and Ph.D. degrees from the Beijing Institute of Technology (BIT) in 1983, 1986, and 2000, respectively. He was a visiting scientist with the Robotics Institute, Carnegie Mellon University (CMU), from 1995 to 1997. He is currently 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. His interests include computer vision, vision-based HCI and HRI, and intelligent robotics.

Weichao Shen received the B.S. degree in control engineering from the School of Automation, Beijing Institute of Technology (BIT), Beijing, China, in 2014. He is currently pursuing a Ph.D. degree in the School of Computer Science, under the supervision of Prof. Yunde Jia. His current research interests include computer vision, unsupervised representation learning, object tracking and 3D reconstruction.

Yuwei Wu received the Ph.D. degree in computer science from Beijing Institute of Technology (BIT), Beijing, China, in 2014. He is now an Assistant Professor at School of Computer Science, BIT. From August 2014 to August 2016, he was a post-doctoral research fellow at Rapid-Rich Object Search (ROSE) Lab, School of Electrical & Electronic Engineering (EEE), Nanyang Technological University (NTU), Singapore. He received outstanding Ph.D. Thesis award from BIT, and Distinguished Dissertation Award Nominee from China Association for Artificial Intelligence (CAAI).