RESEARCH

Open Access

Go together: providing nonverbal awareness cues to enhance co-located sensation in remote communication

Minghao Cai* and Jiro Tanaka

*Correspondence:
mhcai@toki.waseda.jp
Waseda University, Tokyo,
Japan

Abstract

Nowadays, people are easy to have a remote conferencing in different scenarios. However, the potential of mobile video communication has yet to be fully exploited. In our research, we explore providing multiple nonverbal awareness cues to create deeper empathy or understanding between users in the telecommunication. We present our design for the use of remote communication between two geographically separated users. This system provides mutual nonverbal communication cues including independent viewpoint and gesture cue to support natural human-to-human interactions. Using our setups, two users are able to perform multiple joint activities including talking, looking and gesturing together. Finally, users can share co-located sensation and feel themselves “go together” side-by-side in the same place. We performed an evaluation to explore our system’s usability and how multiple awareness cues affect the remote communication. It consists of two experiments: an asymmetric work in which an indoor helper instructed a walking worker and a symmetric collaboration in a more realistic scenario. The positive results show that our design could significantly improve the human-to-human interactions and enhance co-located sensation.

Keywords: Remote communication, Gesture, Awareness, Collaboration, Togetherness, 360° view sharing, Mixed reality

Introduction

Today commercial video conferencing systems are more immersive than traditional voice-only phone calls. These techniques are extensively used at work or in daily life to increase the productivity or to reduce the perception of spatial separation and strengthen the connection of the participants to some extent [1]. However, existing video communication systems are still not satisfactory enough to provide users with a togetherness sensation. When it comes to performing a physical task collaboratively, existing commercial video conferencing techniques offer limited ways to employ body language to interact in the remote user’s view, as they tend to focus on providing mutual talking with a mere capture of both user’s face.

In addition to speech communication, body language including gazes, hand postures and continuous gestures also is an irreplaceable type of mutual human-to-human interaction. Diverse body languages are important in communication and collaboration,

because themselves can act as shared sources of information and treated as mutual visible awareness cues. For example, in “operation teaching” scenario when a expert tries to teach a learner to operate tools, gesture cues (e.g. pointing or fingering) and gaze cues (e.g. eye contact and joint attention) provide rich context information that mere speech cannot clearly reveal.

In our research, we are exploring how to provide multiple nonverbal awareness cues to create deeper empathy or understanding between users in remote communication scenario. A rapid growth in communication infrastructure with increasing bandwidth and pervasiveness, and advances in hardware that can capture people and their surrounding environment afford the possibility for representing body language and provide opportunities for enhancing co-located feelings in joint activities. In this paper, we present our remote communication system design—“Go Together” which refers to a joint activity when two users go together in the same world with natural communication. We provide visual representations of mutual nonverbal awareness cues into the design. This system allows natural communication which is enabled by a combination of three types of user activities: “talk together”, “look together” and “gesture together”, as illustrated in Fig. 1. In “talk together” activity, two users are allowed to have real-time voice communication. In “look together” activity, two users share their viewpoint as awareness cue for joint attention while they still can see the shared environment independently. In “gesture together”, two users perform mutual hand gesture communication as a visible awareness cue. Being allowed to perform these activities during the conversations, users can experience a co-located sensation, feeling they accompany each other going together side-by-side and communicating smoothly in a same physical environment, which we called “go together” feeling.

In this paper, we first present our system design and the implementation. We then present our detail design for supporting viewpoint sharing and gesture interaction as mutual awareness cues to enhance the human-to-human interactions. After that, we present our evaluation to show how mutual these nonverbal communication cues (viewpoint and gesture) can enhance a co-located sensation. The evaluation consists of two experiments. In the first experiment, we compare the user behaviors in an asymmetric collaborative work with our system and two control conditions. The positive results demonstrate that our system can support more efficient remote instructions. In

the second experiment, we test our system in a symmetric collaboration scenario. We aim to understand the manner in which users perform mutual interaction and whether they share a the co-located sensation—"go together" feeling. The results demonstrate that our remote communication design can enhance human-to-human interactions by improving user connection and common feeling. Before we give our conclusion, we discuss the advantages of our "go together" communication, and its implications for remote collaborative interface design, and the potential applications.

Related work

Mixed reality for remote communication

Sharing media space

Since the emergence of technology supporting collaborative works [2, 3], how to construct an effective remote instruction has been a substantial topic in the design of remote communication interfaces. There have been an increasing number of attempts focusing on realizing human-to-human interactions in media spaces. Previous studies have explored methods of helping users use a controller to create and control some virtual annotations overlapping the remote environment to achieve a certain degree of user instruction [4]. However, the communication was relatively restricted and it was difficult to reveal spatial instructions in 3D physical space with simple 2D annotations. More recently, rapid growth of Mixed Reality (MR) techniques make it possible to engage multi-users in a shared media space (virtual or augmented). Introduced as a mix of both augmented reality and virtual reality [5], MR can engage users through a seamless combination of the virtual world and the real physical world. Several attempts have been made to diversely engage separated users in multi-user activities such as joint games [1, 6–8]. These projects show us the possibility of reducing the perception of spatial separation and strengthening the connection of the participants to some extent in remote collaboration.

However, previous space-sharing designs still have some limitations. Although some studies bring multi-users in a shared virtual reality (VR) space [9], much fewer research can provide users a co-located feeling of being together in a physical space. Another problem is, in the most of space-sharing designs, users are required to remain within a specific workspace (usually a desk-scale space) because multiple complicated sensors used for body motion tracking are usually immobile. This problem limits the mobility and makes it difficult to establish a communication for the remote collaboration which requires users move around freely in a larger workspace. In this paper, we are aiming to realize an effective way to share the physical world of a local user with a remote-access user, finally to enhance co-located sensation of users. What's more, our system allows the local user to move freely in the shared world with a tele-presence of the remote-access user, which improves the flexibility and makes our system suitable for mobile collaborative work such as remote tour guidance and remote rescue assistance.

Viewing independence for remote conferencing

One of the important issues in remote conferencing is how to establish a camera to provide an appropriate viewing perspective of the remote user. In traditional remote conferencing systems, cameras used to capture the environment usually exhibit a limited

capture angle and certain blind spots, although some wide-angle cameras have a larger range of view. In most cases, the viewpoint of a remote-access user is fixed to one who carry and control cameras because the cameras are mounted on top of the head or held hands. Such problem not only restricts the remote user's freedom of viewing but might also causes discomfort of disorientation. Previous work by Chang et al. [10] attempted to address such a limitation by using multiple combined cameras with an adjustable shooting direction mechanism. Their work suggest that providing remote-access users with an independent control of the viewpoint can help regulate conversation flow and arouse a feeling of being together, although in this work, users are still restricted by a certain degree of blind spots.

To address the problems we found, in this paper, we present our design to provide two users with independent viewing of the shared world. We explore how the amount of remote view independence affected collaboration. And we found that increase view independence led to faster task completion and more effective assistance in both gestural and verbal communication.

Mutual gesture for remote communication interfaces

Hand gesture for multi-user communication

Hand gesture is known to be an important cue for face-to-face conversations [11, 12]. Mutual gesture awareness can benefit collaborative activities as it is usually related to the awareness of physical surroundings. Mutual gesture is considered as a visible cognitive awareness cue, and it provides rich context information that other body cues cannot reveal, which contributes significantly to a recipient's understanding. For example, in an "operation teaching" scenario, an instructor may guide a learner with gestures (e.g. showing a series of operating method with machines). Or, in a "joint shopping" scenario when two users go shopping together, gestures provide assistance to speech (e.g. pointing out interesting products). Hand gesture has been found useful to detect misunderstanding to overcome the lack of expression of directional information in the conversation.

Gestural instructions for remote collaboration

A number of researches in HCI has investigated how hand motions (without fine finger gestures), mainly used with controllers or pointers, affect collaborations in video conferencing systems [13–16]. Some researchers began to explore conveying gestures over distance. Tang et al. [17] built a system that share live images of captured local user's arms on a remote shared tabletop display. Although the system provides gesture in remote collaboration, one limitation of such systems is that they only provide 2D capture of hands or arms without any depth information. Several systems have captured users' hands in 3D and share hand embodiments in a shared media space [16, 18].

However, these studies mainly support simple or static hand gestures, with the help of controllers, instead of continuous hand gestures with fine finger movements. Another limitation is that most of these systems only provided a unidirectional gesture interaction. Only one user's gestures, mainly the remote-access user's, are captured and used as supporting instructions in the collaboration. Such unequal interaction is unable to be treated as an effective mutual awareness cue in a user-to-user interaction.

To address this problem, our system aims to provide a mutual gesture cue in the remote collaboration and allows users perform mutual free-hand gestures (consisting of hand motions and fine finger gestures) to enhance the understanding of conversation in progress.

Conveying mutual gesture cues in remote collaboration

Based on prior findings in observation studies, we have explored our attempts to enhance a feeling of togetherness in remote communication by maintaining a presence of hand gestures of multi-users. We previously built a prototype called “Trip Together”, a remote pair sightseeing system that supports gestural communication [19]. It is designed to bridge a gestural communication between a user remaining indoor and a remote user going outside. It investigated supporting hand gestures to realize a spatial navigation and direction guidance during mobile sightseeing. We then built an improved prototype for a co-shopping scenario in which the communication involves the environment and objects [20, 21]. The positive feedback of these studies has inspired our research of supporting spatial gestural interaction to assist remote collaboration to enhance user connection.

In this paper, we explore how to convey a mutual gesture cue in remote collaboration and how it affects the amount of communication efficiency. And we found that mutual gesture interactions led to higher accuracy rate, more user confidence, and it could cooperate with speech interactions.

Overview and implementation

System overview

This system is designed for remote communication between two users in geographically separated place. The goal is to maintain mutual gesture interactions and sharing viewpoint as nonverbal awareness cues and finally to enhance a co-located sensation—users are going together in the shared physical world. As shown in Fig. 2, two users include a walking user walks around a physical environment that is shared, and an indoor user remains in a geographically separated indoor environment and tries to assess the shared world through an immersive experience. The indoor user may be a professional engineer who has knowledge to maintain a machine but have difficulties in reaching the remote work scene. He/she may ask a local worker to be a walking user and guide the worker do some operations to finish the work collaboratively.

In this system, we provide both users with independent viewing of the shared world, while they still keep being aware of each other’s viewpoint. A design of sharing viewpoint is used to obtain a joint attention that enhances the accompanying feeling. The design of the 360° view sharing is constructed to help an indoor user view and immerse themselves in the remote physical world. We also provide an effective mutual gesture communication which is developed to reinforce human-to-human interactions, especially for the transmission from the indoor user to the walking user. Captured gestures of the walking user are shared to the indoor user through the panoramic view sharing. In the meantime, the indoor user also gestures freely and his/her gestures are simultaneously captured by the depth sensor and presented back to the walking user’s display superimposing on the physical world.

Implementation

We implement our system using C# in Windows 10. The system software is constructed with Unity 3D engine.

Walking user setup

To realize our system, on the walking user side, we design and implement a set of mobile augmented reality (AR) setup (Fig. 2, upper right). Ricoh Theta S [22] omnidirectional camera is used to capture the shared surrounding, which comprises two fisheye lenses each with a about 195-degree field of view. The two fisheye lenses generate real-time videos with slightly limited overlapping field of views and finally produce a full 360° spherical panorama. The camera is connected it to a laptop (2.5 GHz, 8 GB RAM, Window 10) over USB (1280 * 720, 15 fps) for live video streaming (2500 bit rate) to the remote side using Real Time Messaging Protocol via the Internet. The walking user wears an optical head-mounted display (OHMD)—Epson Moverio BT-300 [23], which displays system information with transparent binocular display (921,600 pixels, 16:9). The Moverio BT-300 also contains an orientation sensor to track the walking user's head movements.

Indoor user setup

We implement a set of wearable virtual reality (VR) setup for the indoor user (Fig. 2, bottom right). We use an Oculus Rift cv1 [24] as the head-mounted display (HMD) for the indoor user, which includes a pair of screens, one for each eye, providing a 110° viewable angle. This HMD is connected to and drove by a desktop PC (2.8 GHz,

8 GB RAM, AMD Radeon RX480, Window 10). A tracking sensor is used to provide a full 6 degree of freedom rotational and positional tracking of the head movements of the indoor user. We choose a Leap Motion sensor [25] for the extraction of the user's gestures which is reconstructed back to the walking user. The sensor is mounted to the front side of the Oculus HMD with a three-dimensional (3D) printing support. The core software is constructed using a Unity 3D engine, which runs on the desktop PC.

System design

The system design consists of the following main parts:

- A Look together: independent viewpoint.
- B Gesture together: mutual mid-air gesture interaction.
- C Virtual pointing assistance.

Look together

In “look together” activity, two users are provided with independent viewpoint control to improve their freedom. In the meantime, by sharing users' viewpoint as an awareness cue for joint attention, we enhance the connection between users and extend the sense of co-presence.

Independent viewpoint

In this system, we use a compact omnidirectional camera to capture a high qualified full 360° spherical panorama of the shared environment. With the help of a mobile PC, a live transmission of the panorama to indoor user is achieved via the Internet. The indoor user wears a head-mounted display (HMD), which comprises of high-resolution optic displays for each eye as the video display, to access the real-time venue and manipulate the view directions through simple head movements. While being provided the omnidirectional capture, the indoor user navigates freely and changes to a different view of the shared world, just like the one personally on the scene (Fig. 3).

The advantage of sharing a full omnidirectional panorama view of the environment is that it provides the indoor user an independent viewpoint of the shared world. While the camera is placed over the walking user's left shoulder (mounting on a metal support carried on the walking user's back), the indoor user's viewpoint is prevented from being influenced by the walking user's head motions.

Sharing viewpoint

Sharing viewpoint design is used to make users be aware of their partner's viewpoint. During remote communication, it is important for the users to know the direction in which the partner is looking, especially when users try to talk about something related to the environment such as spatial distribution or objects in the field of vision. By knowing their partner's viewpoint, users can easily join in the same view, which provides the possibility of obtaining a "joint attention moment". Achieving joint attention, users can share interesting points and grasp a potential conversation topic for discussion. Because both users are provided with independent viewpoints in this system and can look around the shared world freely, it becomes difficult for the users to determine the partner's current direction of focus to obtain a joint attention moment, especially for the walking user. This might result in disorientation and restricts the users' feeling of togetherness.

We design a 3D head avatar to track and show the indoor user's head movements. The avatar presenting in the field of vision (see Fig. 4b). We use a motion tracking sensor placed on the desk to track the indoor user's head movements. When the indoor user changes viewpoint during the conversation, the avatar simultaneously rotates to follow the user's head movements. Seeing the avatar, the walking user can easily keep being aware of the indoor user's viewpoint during the communication.

Auto-reorientation of Head Avatar We intend to make the walking user capable of naturally telling where the indoor user is currently looking at by checking the head avatar displayed in the OHMD. However, the walking user also looks independently and can change the viewpoint during the conversation. A problem must be addressed when the walking user turns his/her head and looks around to a viewpoint that is different from that of the indoor user, the head avatar will not be correctly oriented to

the indoor user's actual viewing direction. The reason is that the two users use different world coordinates. This might cause misunderstandings in communications.

We design an auto-reorientation function to solve this problem.

- Step 1 *Head movement detection* When the walking user turns the head, the head movements will be extracted by a compact orientation sensor in the OHMD. It is used as rectification data sent back to the system rectification component.
- Step 2 *Auto-reorientation* The rectification component works before the system renders the head avatar and the virtual hands in the walking user's OHMD. After receiving the rectification data, the rectification component will be active and rotate back the head avatar and virtual hands.
- Step 3 *GUI* Render corrected head avatar in the walking user's GUI. Consequently, the head avatar of the indoor user is always auto-reoriented with the world-fixed coordinate and forced toward the indoor user's exact viewing direction in the shared world.

When the indoor user simply looks to right side, he/she directly sees the walking user's profile (over the shoulder) in the shared world. So the indoor user can be easily aware of the walking user's viewpoint (see Fig. 4a). The users can then easily acquire a joint attention and share interesting points.

Gesture together: mutual mid-air gesture interaction

In "gesture together", mutual hand gesture is used as a visible awareness cue. Two users perform free-hand gestures naturally for communication and collaboration.

Gesture from the walking user

Our system provides the mid-air hand gestures of the walking user in the shared world (for example, as shown in Fig. 5a, the walking user makes hand gestures that show in the indoor user's view). These gestures may be related to physical objects (such as catch and point) or are simply used as an awareness cue to assist the understanding of the conversation in progress.

Gesture from the indoor user

Some previous research has shown that, to an extent, depth-based hand gestures recognition exhibits the advantage of accuracy and robustness [15, 26]. This allows no wearable or attached sensors on the hands while the in-house user makes some gestural input freely, which extends freedom and comfort. In our design, we choose a depth recognition approach for extracting the hand gestures of the indoor user. Our system represents the gestures with a pair of 3D human-skin hand models superimposing in the walking user's view (see Fig. 5b). These hands are presented in the left part of the field of view.

A compact depth camera is used to extract the real-time depth data of hand gestures. When the depth camera receives gesture data, the system processes the data with the following stages:

- Stage 1 *Leap Motion SDK* The raw data is sent from the depth camera via a USB cable to the host Windows computer, where it is processed by controller software. We use the Leap Motion SDK [27] to extract raw gesture data with almost 200 frames per second. In each frame, the SDK provides data concerning each finger (including bones metacarpal bone, proximal phalanx bone, intermediate phalanx bone and distal phalanx bone), palm of each hand and each upper arm, which includes information concerning 3D position (x, y, z coordinates with reference to the camera itself), orientation, moving speed, and moving direction. The system runs a dedicated process to receive gesture data. The data is then sent to the reconstruction module to reconstruct the user's hands.
- Stage 2 *Hand Model Reconstruction* Each hand model consists of interconnected components representing fingers, palms, and arms. When the reconstruction module receives a frame of the extracted gesture data by matching these data with the hand models, the system precisely reconstructs a pair of 3D mid-air hands in the virtual environment. The current reconstructed hands are then sent to the controller module as an event to update the previous hands.
- Stage 3 *Controller* When the controller component receives reconstructed hands, it updates the 3D position and posture of the corresponding hand objects. Therefore, consecutive frames of data generate a series of consecutively changing hand form mid-air gestures. In other words, once the indoor user makes gestures in front of the depth camera, the virtual models are matched to present the same gestures almost instantaneously.
- Stage 4 *GUI* Finally, the system renders the reconstructed hands in the walking user's OHMD. They are presented on the left side of the field of vision superimposing on the physical world, providing a side-looking perspective. The side-looking perspective enhances the feeling of staying together while the walking user still has a good view of the physical world without being disturbed by the overlapping hand models. These reconstructed hands are also rendered in the indoor user's GUI with a first-person perspective (FPP).

Virtual pointing assistance

As this system aims to enhance the feeling of togetherness, it is important that users transmit instructions smoothly, especially when communication involves spatial and directional context. Supporting a spatial pointing cue is an effective means to enhance

Fig. 6 Pointing cue for instructions: **a** a pointing instruction from the indoor user to the walking user; in **b**, a pointing gesture to start the pointing cue; in **c**, a zoomed in view shows how our system simultaneously supports the virtual 3D arrow: the starting position (finger-tip) and orientation (direction from the proximal interphalangeal of the finger to the finger-tip)

remote cooperation works/applications [15]. In this system, we design a simple freehand pointing gesture that allows the indoor user to use a virtual 3D arrow that presents in the walking user's view to assist a pointing instruction. It is used to draw users' attention as a spatial cue during navigation tasks or selecting instructions (see Fig. 6a).

Pointing Gesture The indoor user makes a pointing gesture to trigger and keep performing the pointing cue. Our system uses a heuristic approach for the gesture recognition based on the depth-based tracking. One important aspect of our technique is that, using the depth sensor, our system can keep tracking the 3D structure of the user's hand including different finger joints and extract both the 3D position and orientation of the indoor user's fingers. Such recognition implementation requires no calibration or precedent training. To activate the pointing technique, the user extends only the thumb and index finger and ensures that the angle between them is larger than the set threshold (see Algorithm 1 and Fig. 6b, c).

Algorithm 1 Detection of pointing gesture

```

procedure Pointing Activated
   $q \leftarrow$  angle between the thumb and index finger
  if index finger extended and thumb extended
    then if other fingers not extended then
      if  $q > \theta$  then
        Pointing function activated
      end if
    end if
  end if
end procedure
  
```

This virtual 3D arrow, which is composed of a long thin cylinder and a red-color tapered tip highlighting the direction, actually assumes a three-dimensional vector as a prototype. The virtual 3D arrow ejects from the tip of the indoor user's index finger and illustrates the spatial rectilinear direction of the spot that the user intends to point toward, presenting in the walking user's view of the shared world. When the user changes the position and orientation of the index finger, this virtual arrow changes simultaneously to follow slight changes and to match the exact pointing direction.

Evaluation

In this section, we introduce our evaluation design, analysis of the results and our observations. We performed two experiments in different use case scenarios to evaluate our system with both asymmetric and symmetric tasks.

Comparative experiment: comparison with two condition

First experiment is a comparative experiment. This work aims to evaluate the usability of the system and whether it can provide effective human-to-human instructions. We adopt a within-subject design for deriving accurate evaluations and a better user's appreciation of our system.

Comparison

We evaluated our system (*Full-functional condition*) against two different conditions that correspond to different levels of immersion: *No-gesture condition* and *Baseline condition*. Under all these conditions, verbal communication was supported via Internet IP phone call.

In the **Full-functional condition—go together**, a full-functional system was supported, including verbal communication, our unique design which we call Independent Viewpoint, and Side-by-side Communication design.

Fig. 7 a Workspace for scenario 1 (b, c) samples of experiment results

In the **No-gesture condition—look and talk together**, we disabled the transmission of hand gestures, so the indoor user only can look and talk to the walking user without gestural interactions. Two users were still provided with independent viewpoints and the head avatar was used to indicate the indoor user's head movements. Real-time verbal communication is supported.

In the **Baseline condition—talk together**, we disabled all mutual nonverbal awareness cues and provided verbal communication. Under this condition, only a baseline function of telecommunication is supported. Only a real-time video capture of the local surrounding is shared along with verbal communication. The local surrounding capture is provided from a fixed built-in camera of the smart glasses worn by the walking user, which always makes the indoor user's viewpoint synchronously follow the walking user. The indoor user browses the shared video in the HMD without viewpoint control.

Experiment design

This experiment exhibited an asymmetric collaboration in which the two users played different roles. The task involved a collaborative arrangement of disordered boxes. In this task, the walking user worked as a **worker** picking up selected boxes and placing them in designated locations under the instruction of the indoor user. The indoor user worked as a **helper** instructing the walking user to complete the task. We call this scenario "arranging boxes", which simulates a daily scenario of arranging disordered objects in a room. It was considered representative as it involved several instructions and guidance which could be found in practical real-world tasks. Several past studies have demonstrated similar tasks in the evaluation of remote collaboration [19, 28, 29].

Participants

We recruited 12 participants who were students or researchers at our department. They were between the ages of 20 and 28 years and possessed regular computing skills. The participants were randomly grouped into six pairs. In each pair, one participant assumed the role of the walking user while the other assumed the role of the indoor user.

Task and procedures

This study was performed in a room-scale workspace (see Fig. 7). The workspace consisted of two office desks and a shelf placed in the corner. The shelf included multiple lattices that were arranged with the desks in an L-type distribution. The entire workspace could not be seen at once by the users in the field of vision without turning the head to change the viewpoint while the worker (walking) walked in the workspace.

Before the start of the experiment, participants were given 10 min to practice in advance, and an explanation of the system was also included.

At the beginning of each trail, there were ten boxes placed on the desks. The helper (indoor) instructed the worker (walking) to pick up the selected boxes (one by one, five in total) and directed the worker to place them into the lattices (one lattice, one box, Fig. 7b, c present two samples randomly chosen from the experiment results). The worker was provided with sufficient space and allowed to walk freely in the workspace.

Each pair of participants were given three trails for the three conditions, and the order of conditions was randomized to maintain the counterbalance.

Data collection

We collected both quantitative and qualitative data.

During the experiment, we measured the total task completion time of each trail. We also counted the number of adjustments, which refer to the number of times the walking user failed to follow instructions correctly and additional adjustment was then needed to complete the box placement task under the instructions of the indoor user.

After each trail, the participants answered a questionnaire consisting of a series of statements with a 7-point Likert Scale (strongly disagree = 1, strongly agree = 7).

Results

We used a Wilcoxon Signed Rank test to analyze the significance of the experiment results across the three conditions. Figure 8a illustrates the mean task completion time for the three conditions. There was a significance difference as the *Full-functional*

condition performed much better than the other two conditions (all $p < 0.01$), and the *No-gesture condition* performed better than the *Baseline condition* ($p < 0.05$).

The numbers of adjustments required to complete the box placement task was measured (see Fig. 8b). There were significant differences with the *Full-functional condition* demonstrating lower adjustments than the other two conditions (all $p < 0.01$).

The results of the questionnaire answered by both users are presented in Fig. 9. The first question based on the Single Ease Question [30]—“*Over all I think this task was easy*” was used to determine the difficulty of the tasks for both users. The *Full-functional condition* was ranked higher than the other conditions by the walking user ($p < 0.05$) and the indoor user ($p < 0.05$). Question “*I could quickly understand instructions.*” answered by the worker and “*I could quickly transmit my instructions.*” answered by the helper were corresponding questions used to test the perceived instruction effectiveness. There were significant differences for both the worker and the helper with the *Full-functional condition* ranking higher than the other two conditions (all $p < 0.05$). The last question “*I enjoyed this trail.*” was used to test the user preference. We observed that both the worker and the helper participants preferred the *Full-functional condition*.

Observation

These statistics results demonstrate a higher quality communication of the system that we proposed. The task completion time was used to examine the efficiency of collaboration. The number of adjustments was used to measure the accuracy of instruction transmission. These demonstrate that our system can provide more efficient instructions during the collaborative task with Independent Viewpoint and Side-by-side Communication designs.

This finding is also supported by the statistical significance obtained in the case of the questionnaire. Based on the responses to the questions, we confirmed that the participants shown higher confidence in transmitting and perceiving the intentions in the communication of our system, and the associated tasks performance was rated significantly easier than the other conditions. Users prefer our system to the other conditions. This is likely because the helper felt more in control of selecting and placing objects, while the worker understood the instructions quickly.

Experiment 2: realistic scenario

In this experiment, we aim to evaluate our system in a more realistic scenario and investigate whether the user can achieve effective mutual interactions. The study exhibited symmetric collaboration in which participants worked more equally than the first experiment. Inspired by our previous research [21], we design a “go and shopping together” task that simulates the situation in which two users walk together side-by-side to a shop to purchase something. This is a physical task collaboratively performed by users which requires users to use gesture cue and share their viewpoint to finish the job in a mobile status. It is suitable for the comprehensive testing of system’s performance.

Participants

We recruited 12 participants between the ages of 21 and 28 years. All participants possess regular computing skills. The participants were randomly grouped into six pairs. In each pair, one participant assumed the role of the walking user while the other participant assumed the role of the Indoor User.

Task and procedures

The go shopping together experiment was performed in a stationery store, a larger space than the workspace in the first experiment. The workspace consists of various type of products with various exteriors. Before the start of the experiment, we explained the use of our system and the task to each participant. Each participant was given 10 min to practice. The task was to look for a product that could interest both users as a little gift (such as a ceramic craft or plastic decoration).

During the experiment, verbal communication was supported via Internet IP phone call. The participants were free to discuss with each other. The walking user walked around and communicated with the Indoor User, and the latter could request the former to move or perform some operation such as picking up and holding some objects with hands. This task was open-ended, and the only requirement was that the participants

must arrive at an agreement to select a product. After the pilot test, we observed that the duration of completion was primarily influenced by the personal preference. Therefore, in this part of the experiment, we did not enforce any time limitation and the task continued until participants find a satisfying object.

At the end of the experiment, participants were asked to fill out a questionnaire. Following this, post-task interviews were conducted. We were primarily interested in user feedback about the operation, language and gestures used in the task, potential application cases and possible opportunities for system modification.

Gesture rate

To investigate user performance, we recorded the following duration data of each pair in the experiment:

- T_t : entire task completion time

T_t consists of T_b and T_c :

$$T_t = T_b + T_c$$

- T_b : the duration in which participants only browsed the environment independently without communicating
- T_c : the duration in which participants took part in user-to-user communication; which includes the moments when users communicated verbally communication, the moments when users made gestural interactions, and the moments both of verbal and gestural communication happened.

T_c can be classified under two categories— T_g and T_v :

$$T_c = T_g + T_v$$

- T_g : the duration when participants performed mutual gestural interaction including those performed with speech
- T_v : the duration when participants communicated without using any gestures (with only speech)

The following formula depicts the relationship between the data:

$$T_t = T_b + T_c = T_b + (T_g + T_v)$$

With the data, we calculate a *Gesture Rate*— R using the following formula:

$$R = T_g / T_c = T_g / (T_g + T_v)$$

This figure is used to measure the statistical proportion of gesturing in user communication. It reveals whether the user can achieve an effective gestural interaction and convey the importance of supporting such mutual gestural communication to a certain extent.

Fig. 10 Gesture rate in Experiment 2

Fig. 11 Mean scores of questionnaire in Experiment 2

As shown in Fig. 10, we observed that the Gesture Rates for all pairs are over 45% with 56% on average. This means participants generally used gestural interactions over the half duration of the communication. This revealed that such mutual gestural interactions could truly assist human-to-human communication.

Questionnaire

We also conducted a questionnaire based on previous evaluation samples [31, 32] to collect qualitative feedback and investigate the co-presence feeling in our system. Figure 11 presents the results of the questionnaire. Question 1 was used to test the usability of independent viewpoint design. Question 2 was used to judge the awareness of user focus. Questions 3 and 4 were used to investigate perceived information understanding and users' confidence of the mutual gestural interaction. Questions 5 and 6 were used to test the co-presence during the communication. In Question 7, we intended to investigate the overall performance and user experience of our system.

Observation and feedback

In this section, based on experiment results and our observations, we provide a description of how users achieved a “go together feeling” from different aspects and the general user feedback of our system. It includes following five parts:

Aspect 1 Look together: independent viewpoint The results of Q1 indicated that both users could use independent viewing. Users enjoyed a certain degree of freedom, which relaxed the viewpoint restriction in traditional communication systems. When using our system, the indoor user observed the shared remote scenery and looked around independently using the free viewpoint. Our system enhances the immersive feeling of the indoor user because of the ability to act independently. As participants said: “I could look around at will without asking my partner to change the viewpoint, which was convenient”. Another reason for preference and confidence of use is the ability to view the entire scene in the independent viewpoint. One subject said “I saw the entire environment just like I was really there”. For the walking user, without paying additional attention to contorting the indoor user’s viewpoint (which should always be done with traditional video call), he/she could look around independently and enjoy the experience: “...I could spend time looking around independently. It also increased the efficiency.”

Aspect 2 Look together: viewpoint sharing From the results of Q2, we confirmed that although the walking user and the indoor user viewed independently, they could still share focus awareness easily during the communication. This enhanced the co-presence feeling and improved the efficiency of communication. The indoor user obtained the walking user’s viewing direction from the panoramic sense, while his/her own head motion was transmitted to the latter through the head avatar. Sharing mutual attention was used as one of the nonverbal cues, which assisted users in understanding the messages being relayed and helped users quickly join the same scenery and experience a joint attention as commented by the participants: “When my partner found something interesting, I could quickly find the same thing after a quick confirmation of his viewing direction.” Users felt a more accompanying feeling by being aware of each other’s actions: “I liked seeing the head avatar even though sometimes we did not talk or discuss. Knowing my partner’s situation made me feel accompanied”.

Aspect 3 Gesture together In the results of Q3 and Q4, both users gave positive scores. This indicate that users could perform gestures to transmit their intentions and achieve mutual smooth communication. During the communication, users used mutual gesture interaction as a nonverbal body cue. One reason for perceiving the being together feeling from this interaction is the ability to naturally gesture with a side-by-side perspective in the same world as the real collocated situation. As one indoor participant said, “...seeing her making a gesture was vivid and made me feel like she (walking user) was next to me”, and the partner said “I felt the appearance of hands and gestures were intuitive and convincing.” Another reason is, in some cases,

Table 1 Types of interactions when gesture is used with speech: Gesture-Speech Cooperative Interaction

User	Type of interaction	Example of gesture	Example of speech
Both	Direction guidance	Showing with arm	"This way"
Both	Direct partner's attention	Pointing with finger	"Look at that"
Both	Establishing visual annotation	Counting fingers	"One, two, three..."
Indoor	Instructions related to an object	Pointing with finger (Pointing Cue used)	"Pick up this cup"
Indoor	Instructions related to adjusting the objects	Showing the palm and turning it over	"Turn to the other side"
Indoor	Instructions related to the size	Showing approximate size with two hands	"It's this wide"
Indoor	Confirmation to walking user	Thumb up/ring gesture	"Yes, it's good"
Walking	Query to indoor user	Pick up the object	"Is this correct?"
Walking	Instructions related to an object	Holding a object and turn it to different sides	"Look at this side"

gestures provide more accurate instructions and reduce dependence on language description, which improves efficiency. Users often use hand gestures to indicate their interests and to communicate with the other user. As one said "I felt he knew where I was pointing at, so I just said 'this' or 'that' to identify an object or direction," while another commented "...it (pointing assistance) was very fluid and guided me nicely"

We also observed that information transmission from the walking user to the indoor user was graded slightly higher than that from the indoor user to the walking user. After some post-task interviews with the participants, we determined that the difference in information transmission was likely because the walking user could make gestures and actually touch an object to cause a more visual feedback like depressing an object's surface with fingers.

Aspect 4 Gesture and talk together: gesture-speech cooperative interaction From the result of *gesture rate*, we know that mutual gesture interaction included a noticeable proportion of the entire user communication moments. During the communication, we noticed that users usually intended to gesture with a speech communication, especially when the context was related to the environments or ambiance. We defined this as a *Gesture-Speech Cooperative Interaction*. To measure the pattern that gestures used with speech communication, different gestures made with certain types of phrases were recorded (see Table 1). We found that such interactions exhibit the following features:

- 1 Gesture and speech worked cooperatively to realize the full interaction.
- 2 Hand gestures were used as a visible cue that contained the main directive information.
- 3 Speech was used to draw the recipient's attention and to indicate the start/stop gestures.
- 4 Although users did not always keep talking each time a hand gesture was made, the beginning of the gestures was usually accompanied by speech.

- 5 Speech would be used for the supplementary explanation of gestures. We consider the type in Table 1—line 4 as an example. Users instructed recipients to select an object with a hand gesture (pointing), using speech to explain the following action (pick up). It could be noted that the term ‘that’ in speech explicitly requires the recipient to find something beyond the conversation itself—the information of direction which the speaker instructs; simply listening to it would be inadequate.

General feedback In this experiment, all pairs of participants were able to complete the task successfully and enjoyed the remote communication experience. From the results of Q5 and Q6, we confirmed that both users could receive a co-presence feeling. Users were aware of their partners as the task was being executed and believed that he/she was not alone or secluded, which kept users in close connection. From the results of Q7, we confirmed that users could generally receive common perceptions and experience a “go together” feeling using our system. This was also supported by comments of participants including: “I really enjoyed this collaboration. I was able to feel going together with my partner” and “We could make decisions together just like we were in the same place.” Most participants experienced a co-presence feeling using our system—“I (walking user) found the presence of the head avatar and the hand gestures of the indoor partner to be quite helpful and intuitive, which gave a feeling that my partner was right here with me”, and “I (indoor user) could look around and discuss with my partner, feeling as though I was there going together with my partner”.

Discussion and motivation

This study presents a remote communication system designed to help a walking user and an indoor user, who are in separated spaces to experience a “go together” feeling, which refers to the feeling of going out together and communicating in the same world. We aim to enhance co-located sensation by enhancing human-to-human interactions. Compared to traditional remote conferencing, our system provides nonverbal awareness cues to allow two users perform multi-layered joint activity consists of talking together, looking together and gesturing together. After the evaluation, with our system, people experienced significantly better instruction for asymmetric collaboration. In addition, they could experience a feeling of co-presence in the symmetric communicative work. We also found a noticeable gesture-speech collaborative interaction pattern in the asymmetric collaboration, which might help explain the observation of positive transmission and perception of user interactions.

These findings were probably influenced by the nature of the collaboration and the role of the non-verbal cues, particularly the mutual gesture cue. During the arranging box experiment, the walking user was on the receiving end of the communication for most of the time and relied heavily on spatial instruction from the indoor user. Our system supported the free mid-air gesture interactions between users, particularly the virtual pointing cue for instructions. In the go shopping together experiment, constant mutual communication was needed. Our system remained users in viewing independence and helped them keep aware of their partner’s current viewpoint. Participants also felt that

using gesture communication their partner’s intentions were clearer to them, and they could understand their partner equally well.

Implications for remote communication interface design

In this paper, we proposed our “go together” design to address one of the important aspects of telecommunication design—how to share non-verbal communication and awareness cues. Mutual gesture interaction design addresses the restriction of the explanation approach and reduces the difficulty for users to describe spatial information when they use only speech. Mutual focus awareness design reduces the need for users to confirm a partner’s focus point constantly via speech. The experiment results imply that people developing an asymmetric remote communication interface should consider adding the multiple awareness cues or a similar element to transmit a remote gesture cue and attention interaction. The constant availability of these designs will likely reduce the task difficulty, improving communication efficiency. Similarly, in a symmetric work that involves shared problem solving and negotiation, our go together design would be useful for improving co-presence and accompanying feeling and awareness of the partner during collaboration.

Side-by-side go together vs first-person perspective

In traditional view sharing designs, which usually are found in previous Computer-Supported Cooperative Work (CSCW) [33], the local user mostly perceives the remote venue with the same field of view of the remote user. With such sharing of first-person perspective (FPP) of the content, the remote user acts more like a “stand-in” of the local user rather than a communicating partner (see Fig. 12, upper right). It might lead to misunderstanding and limits the natural communication between users. By contrast, our go together system simulates a side-by-side togetherness with independent viewpoint control, which provides both users with more independence and let them could focus

Table 2 Potential application scenarios

Indoor user	Walking user	Scenario	Collaborative work/communication context
Expert	Local worker	Remote searching	An expert can supervise a local worker in emergency searching and rescue operation
Trainer	Learner	Remote education	A trainer can guide a learner in manual operations on machines to handle problems for the first time
Elderly people	Friends/family	Remote accompany and rich life logging	People who have difficulties to go out around can stay in a comfortable environment and get a virtual sightseeing with their friends to enjoy accompanying moments and rich life logging
Partner	Designer	Discussion and collaboration	Designer and partner work together for indoor decoration

more mutual interaction (see Fig. 12, bottom right). This could enhance a co-located sensation, which is also supported by our user study results.

Application scenarios

Although we tested our system in arranging boxes and shopping scenarios in our experiments, the potential applications are not limited to these scenarios. Our system is also suitable for other remote collaborative works or remote assistance when users are restricted by geographical separation but try to have an instant communication, especially when it comes to performing a mobile physical task collaboratively in which gesture communication is essential. Table 2 illustrates some example of potential application scenarios, but the practical applications of our system are not limited to these.

Conclusion

In this paper, we presented our design for the remote communication between two geographically separated users. We explore providing two kinds of nonverbal awareness cues, viewpoint sharing and mutual hand gestures, to support natural human-to-human interactions and finally to create deeper empathy or understanding between users. With our setups, users are able to perform multi-layered natural interactions: talk together, look together and gesture together. We enhance the co-located sensation of users and provide a feeling that they go together side-by-side in the same environment chatting and gesturing, and being accompanied.

We performed experiments to evaluate our system in both symmetric and asymmetric collaborations. We found a noticeable gesture-speech collaborative interaction pattern, in which gestures were used collaboratively with speech. We also discussed the implications of our “go together” communication for remote collaborative interface design. Overall, we found that our system can convey mutual nonverbal cues to improve performance in an asymmetric object placement task. Our system was also determined to be useful for mutual user interactions and improving co-presence feeling in symmetric work. In both cases, users gave positive feedback. These findings support our belief that

by enhancing human-to-human interactions and expanding users' independence, our design can finally help users experience the feeling of being accompanied and share a certain degree of co-located sensation.

Abbreviations

MR: mixed reality; AR: augmented reality VR: virtual reality; HMD: head-mounted display; OHMD: optical head-mounted display; 3D: three-dimensional; FPP: first-person perspective.

Acknowledgements

Not applicable.

Authors' contributions

MC developed the idea for the study, performed research and collected the data. Both authors designed research, analyzed the data and were involved in writing the manuscript. Both authors read and approved the final manuscript.

Authors' information

Minghao Cai is a student of WASEDA University. His research interests include human computer interaction, tele-presence and multiple interactions. He received a Bachelor's Degree from South China University of Technology and a Master's Degree from WASEDA University in 2015 and 2017.

Jiro Tanaka is a Professor of Graduate School of Information, Production and Systems, WASEDA University from 2016. He worked at Department of Computer Science, University of Tsukuba as an Associate Professor and a Professor from 1993 to 2016. His research interests include ubiquitous computing, interactive programming, and human computer interaction. He received a B.Sc. and a M.Sc. from The University of Tokyo in 1975 and 1977. He received a Ph.D. in computer science from University of Utah in 1984. He is a member of ACM, IEEE and IPSJ.

Funding

Not applicable.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no significant competing financial, professional, or personal interests that might have influenced the performance or presentation of the work described in this manuscript.

Received: 26 December 2018 Accepted: 30 April 2019

Published online: 10 May 2019

References

1. Kegel I, Cesar P, Jansen J, Bulterman DC, Stevens T, Kort J, Farber N (2012) Enabling 'togetherness' in high-quality domestic video. *ACM, New York*, pp 159–168
2. Bly SA, Harrison SR, Irwin S (1993) Media spaces: bringing people together in a video, audio, and computing environment. *Commun ACM* 36(1):28–46. <https://doi.org/10.1145/151233.151235>
3. Bolt RA (1980) "Put-that-there": voice and gesture at the graphics interface. In: *Proceedings of the 7th annual conference on Computer graphics and interactive techniques-SIGGRAPH'80*, pp 262–270. <https://doi.org/10.1145/800250.807503>
4. Poupyrev I, Billinghurst M (1996) The go-go interaction technique: non-linear mapping for direct manipulation in VR of the 9th annual. *ACM, New York*, pp 79–80. <https://doi.org/10.1145/237091.237102>
5. Milgram P, Kishino F (1994) A taxonomy of mixed reality visual displays. *IEICE Trans Inf Syst* 77(12):1321–1329
6. Raffle H, Go J, Spasojevic M, Reville G, Mori K, Ballagas R, Buza K, Horii H, Kaye J, Cook K, Freed N (2011) Hello, is grandma there? Let's read! StoryVisit. In: *Proceedings of the 2011 annual conference on Human factors in computing systems-CHI'11*, p 1195. <https://doi.org/10.1145/1978942.1979121>
7. Hunter S, Maes P, Tang A, Inkpen K, Hessey S (2014) WaaZam! Supporting creative play at a distance in customized video environments. In: *Conference on human factors in computing systems (Figure 1)*, p 146. <https://doi.org/10.1145/2556288.2557382>
8. Dey A, Piumsomboon T, Lee Y, Billinghurst M (2017) Effects of sharing physiological states of players in a collaborative virtual reality gameplay. In: *Proceedings of the 2017 CHI conference on human factors in computing systems-CHI'17*, pp 4045–4056. <https://doi.org/10.1145/3025453.3026028>
9. Sra M, Mottelson A, Maes P (2018) Your place and mine: Designing a shared VR experience for remotely located users. In: *Proceedings of the 2018 on designing interactive systems conference 2018*. *ACM, New York*, pp 85–97
10. Chang C-T, Takahashi S, Tanaka J (2014) A remote communication system to provide "out together feeling". *J Inf Process* 22(1):76–87. <https://doi.org/10.2197/ipsjip.22.76>
11. Goodwin C (1986) Gestures as a resource for the organization of mutual orientation. *Semiotica* 62(1–2):29–50
12. Cook SW, Tanenhaus MK (2009) Embodied communication: speakers' gestures affect listeners' actions. *Cognition* 113(1):98–104

13. Duval T, Nguyen TTH, Fleury C, Chauffaut A, Dumont G, Gouranton V (2014) Improving awareness for 3D virtual collaboration by embedding the features of users' physical environments and by augmenting interaction tools with cognitive feedback cues. *J Multimodal User Interf* 8(2):187–197
14. Greenberg S, Gutwin C, Roseman M (1996) Semantic telepointers for groupware. In: Proceedings sixth Australian conference on computer–human interaction, 1996. IEEE, Piscataway, pp 54–61
15. Sodhi RS, Jones BR, Forsyth D, Bailey BP, Maciocci G (2013) BeThere: 3D mobile collaboration with spatial input. In: Proceedings of the SIGCHI conference on human factors in computing systems—CHI'13. pp 179–188. <https://doi.org/10.1145/2470654.2470679>
16. Amores J, Benavides X, Maes P (2015) Showme: a remote collaboration system that supports immersive gestural communication. In: Proceedings of the 33rd annual ACM conference extended abstracts on human factors in computing systems. ACM, New York, pp 1343–1348
17. Tang A, Neustaedter C, Greenberg S (2007) Videoarms: embodiments for mixed presence groupware. In: People and computers XX—engage. Springer, London. pp 85–102
18. Tecchia F, Alem L, Huang W (2012) 3D helping hands: a gesture based mr system for remote collaboration. In: Proceedings of the 11th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry. ACM, New York, pp 323–328
19. Cai M, Tanaka J (2017) Trip together: a remote pair sightseeing system supporting gestural communication. In: Proceedings of the 5th international conference on human agent interaction. ACM, New York, pp 317–324. <https://doi.org/10.1145/3125739.3125762>
20. Cai M, Masuko S, Tanaka J (2018) Gesture-based mobile communication system providing side-by-side shopping feeling. In: Proceedings of the 23rd international conference on intelligent user interfaces companion. ACM, New York, pp 2–122
21. Cai M, Masuko S, Tanaka J (2018) Shopping together: a remote co-shopping system utilizing spatial gesture interaction. In: International conference on human–computer interaction. Springer, Berlin, pp 219–232
22. RICOH: RICOH THETA S (2018). <https://theta360.com/en/about/theta/s.html>. Accessed 20 Dec 2018
23. EPSON: Moverio BT-300 (2018). <https://moverio.epson.com/>. Accessed 20 Dec 2018
24. Oculus: Oculus Rift (2018). <https://www.oculus.com/>. Accessed 20 Dec 2018
25. LEAP MOTION: LEAP MOTION (2018). <https://www.leapmotion.com/>. Accessed 20 Dec 2018
26. Karam H, Tanaka J (2015) Finger click detection using a depth camera. *Procedia Manufacturing* 3:5381–5388
27. LEAP MOTION: Leap Motion's SDK (2018). <https://developer.leapmotion.com/documentation/>. Accessed 20 Dec 2018
28. Piumsomboon T, Lee GA, Hart JD, Ens B, Lindeman RW, Thomas BH, Billinghamurst M (2018) Mini-me: an adaptive avatar for mixed reality remote collaboration. In: Proceedings of the 2018 CHI conference on human factors in computing systems. p 46. ACM, New York. <https://doi.org/10.1145/3173574.3173620>
29. Higuchi K, Yonetani R, Sato Y (2016) Can eye help you?: effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. In: Proceedings of the 2016 CHI conference on human factors in computing systems—CHI'16. pp 5180–5190. <https://doi.org/10.1145/2858036.2858438>
30. Sauro J, Dumas JS (2009) Comparison of three one-question, post-task usability questionnaires. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp 1599–1608. ACM, New York
31. Harms C, Biocca F (2004) Internal consistency and reliability of the networked minds measure of social presence
32. Slater M, Usoh M, Steed A (1994) Depth of presence in virtual environments. *Presence Teleoperators Virtual Environ* 3(2):130–144. <https://doi.org/10.1162/pres.1994.3.2.130>
33. Kasahara S, Rekimoto J (2015) JackIn head: immersive visual telepresence system with omnidirectional wearable camera for remote collaboration. In: Proceedings of the 21st ACM symposium on virtual reality software and technology, vol 23, issue 3. pp 217–225. <https://doi.org/10.1145/2821592.2821608>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)