

## Recommended Paper

# A Remote Communication System to Provide “Out Together Feeling”

CHING-TZUN CHANG<sup>1,a)</sup> SHIN TAKAHASHI<sup>2</sup> JIRO TANAKA<sup>2</sup>

Received: April 11, 2013, Accepted: September 13, 2013

**Abstract:** In this paper, we set out to define the *out together feeling* as the experience when two people at different locations feel as though they are together. In other words, it makes a pair of users, one outdoors and the other indoors, feel as if they are both outdoors together. To determine a set of interaction methods to enable indoor and outdoor users to interact and share the *out together feeling*, we carried out preliminary experiments to observe the basic elements of communication between people who are really together. We then carried out an experiment in which indoor and outdoor users communicated via a videophone and observed the interaction patterns of each user as they attempted to achieve a given goal. From the analysis of these data, we defined three basic elements that are required to achieve the *out together feeling*: (1) both users can freely peruse the outdoor user’s surroundings, (2) know where each other is looking, (3) and can communicate non-verbally using gestures. Using these basic elements, we designed and implemented a system called WithYou. This consists of two subsystems: a wearable system for the outdoor user and an immersive space for the indoor user. The indoor user wears a head-mounted display (HMD) and watches video from a pan-and-tilt camera mounted on the outdoor user’s chest. Thus, the indoor user can look around by simply turning their head. The orientation of the outdoor user’s face is also displayed on the HMD screen to indicate where they are looking. We experimentally evaluated the system and, based on an analysis of the subjects’ response to questionnaires and video recordings, we were able to assess the level to which the *out together feeling* was achieved.

**Keywords:** tele-presence, communication support, wearable mobile, human robot interaction

## 1. Introduction

With the use of high-speed mobile networks, it is now possible to provide high-bandwidth and stable mobile communications in outdoor environments. In addition, mobile video communication systems such as the videophone have become feasible. However, the potential of mobile video communication has yet to be fully exploited. One reason for this is that most video communications systems developed to date primarily assume face-to-face communication, which is not always helpful for users who may want to focus on other information, such as body language and gestures, or to both look at something, such as a distant object.

There are, however, other possibilities for mobile video communication. For example, using a videophone, users can shoot and send a video stream of their surroundings to an indoor user. They can then share images of a place and talk about it. This type of communication may allow users to feel more as if they are together in the same place. However, simply sending images is not sufficient to realize such a feeling. For example, sharing the focal direction naturally is important to initiate conversation about the shared video images. Body language and gestures are

also an important aspect of communication, and so should also be shared.

Our final goal is to make full use of remote video communication technology and to design interaction methods to realize the *out together feeling*, a sensation shared by two people at different locations (one indoors and one outdoors) whereby it feels as if they are both in the same outdoor environment. It is a form of telepresence for outdoor environments. Although both users may be going out, in this research, we assume that one user is going out (i.e., outdoor users) and the other stays inside (i.e., indoor users).

The purpose of this work was first to determine the basic elements of the *out together feeling*. To that end, we designed two experiments to examine which types of communication methods people use when they are actually together in an outdoor environment, and which types they employ when making a videophone call. In these experiments, subjects determined their own mission target, such as to purchase something or survey a point of interest. We observed the interactions using video recordings, and asked the participants to complete a questionnaire about their experiences. From these results, this paper figures out the basic elements of interaction between the subjects.

The second aim of this work was to design interaction methods

<sup>1</sup> Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba, Ibaraki 305–8577, Japan

<sup>2</sup> Division of Information Engineering, Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Ibaraki 305–8577, Japan

<sup>a)</sup> tsubaki@iplab.cs.tsukuba.ac.jp

The initial version of this paper was presented at the Sixth International Conference on Collaboration Technologies: CollabTech2012 held in Sapporo, Japan, on August 28–30, 2012, under the sponsorship of SIGGN. This paper was recommended to be submitted to IPSJ Journal by the chairman of SIGGN.

to realize the *out together feeling* by implementing a video-based communication system. We designed and implemented a system called WithYou to provide the *out together feeling* between an indoor user and an outdoor user. The outdoor user shares the remote environment via a head-mounted display (HMD) worn by the indoor user, while the outdoor user wears a pan-and-tilt camera mounted on the outdoor user's chest. WithYou enables the indoor user to freely look around the surroundings of the outdoor user, and also makes each user aware of the direction that other is looking.

Finally, we evaluated the system by performing an experiment on a real street. In this experiment, subjects were asked to use our system to achieve their own mission target. The experiment was videoed and analyzed to compare with (1) the first experiment that two subjects actually go out, and with (2) the second videophone experiment.

The contribution of this paper is the design and implementation of the interaction methods that realize the *out together feeling* on the video-based communication system. In this paper, three basic functions (i.e., view surroundings freely, notice the focus of each other and gesture communication) were implemented to achieve the concept of *out together feeling*.

The remainder of this paper is organized as follows. First, Section 2 describes related work. Next, the *out together feeling* is defined in Section 3. Section 4 describes the results of the two preliminary experiments (communication between people who are together outdoors, and communicating via a videophone). Sections 5 and 6 describe the design and implementation of the WithYou system. Section 7 describes an experimental analysis of the system. Section 8 concludes the paper.

## 2. Related Work

### 2.1 Remote Instruction and Support

In remote instruction and support systems, indicators are supplied with a clear remote live image, and can instruct and support operators who are at the remote location.

Shared-View by Ōta et al. [3] is a method of directing cardiopulmonary resuscitation by remote environment operation. The system users are the operator and the director. The operator wears an HMD and a head-mounted camera, and follows the instructions of the director, who guides the emergency resuscitation using the HMD screen and voice instruction, and sees an image of what the operator sees at the remote location.

GestureMan by Kuzuoka et al. [4] employs a robot to create a remote working direction system. The director's head movements cause the head of the robot to rotate, and three cameras are mounted on the robot's head to transmit real-time images to the director. GestureMan also provides a pointing function, using a controllable arm with a laser pointer. The director uses a joystick to control the arm of the robot and provide remote instruction. Moreover, the laser pointer can be used to indicate a given position by touching the screen on the director's side.

Koizumi et al. [7] employed a teleoperated communications robot with the aim of developing a system to interact with human activities at train stations or shopping centers. The operator could communicate with visitors using voice and video. In addition,

the operator could also monitor live images taken by remote cameras.

Michaud et al. [6] employed a telepresence robot for home care assistance. Their system, Telerobot, employed a mobile video-phone robotic platform with a waypoint navigation feature. The operator may give orders for the robot to move to a specified position by clicking the waypoint displayed on a 2D map.

Both of our research and these works support for local users to grasp remote situation via live images. The distinctive aspect of our research is that our research aims at more equal treatment of a local user (i.e., an indoor user) and a remote user (i.e., an outdoor user), instead of regarding them as an instructor and a worker. For example, joint attention can be possible from the focusing behavior of either end of users. Furthermore, our system not only provides a method for communicating video images but also allows multiple methods of interaction between users.

### 2.2 Virtual Activities & Communication Support via Robot

In researches which provides the feeling of communicate to others via robot has been widely known. In many cases, indicator operate the remote robot and perform a human to join activities.

Tsumaki et al. [1] proposed and developed a wearable robotic system called Telecommunicator, which allowed the local site user to communicate with others at a remote site. Telecommunicator is a wearable robotic device mounted on the user's shoulder, which consists of a rotatable video camera and a simple arm. The users are divided into the local site user and remote site user; the former wears an HMD and controls the remote camera by turning the head. Live images are displayed on the HMD at the local site.

Kashiwabara et al. [2] developed a system called Teroos, which involves a wearable avatar to enhance the feeling of participation in joint activities between local and remote users. The avatar is controlled remotely by the local user, and a pan-and-tilt camera and a rotatable eye for virtual expression of the eye movement were mounted on the avatar, to provide a sense of presence to the remote user.

Mebot by Adalgeirsson et al. [5] employed a telepresence robot to provide social expression. Mebot had two arms and a head with pan-and-tilt ability. A smartphone or a tablet personal computer (PC) could be mounted on the head of the robot to show the face of the indicator in real-time. In addition, the indicator can see the remote image through the smartphone camera. The indicator controlled the arm of the robot using a joystick, and the head of robot could pan and tilt automatically in response to head movements of the indicator.

Compared to these works, our work focuses on the direction in which users are facing, and our system makes use of these directions both for the indoor and the outdoor user. Furthermore, a "joint attention" mode helps users focus on the same object together. In our system, the indoor user can view the remote surroundings easily by turning his or her head, and the process does not involve keyboard or mouse control. Such an intuitive and immersive space involving an HMD provides the indoor user with greater telepresence.

### 3. The Out Together Feeling

The *out together feeling* is a sensation shared by two people at different locations that feels like they are going out together. In *out together feeling*, two people equally feel a sense of doing something together with the remote partner. It is a kind of telepresence technology for use in outdoor environments. Although both users may be going out, in this research, we assume that one user is going out (i.e., outdoor users) and the other stays inside (i.e., indoor users). The following cases are example applications:

**Communication among separated family members** Separated family members, such as parents and a child who is studying abroad, may communicate using WithYou. The parents have an interest in the circumstances of the school where the child is studying, but visiting is expensive. Using WithYou, the child can take his or her parents on a virtual tour of the school and the town where the child is living.

**Virtual Travel** Virtual travel is another potential use of WithYou. A travel guide walking at a popular sight-seeing destination guides virtual tourists, who can look around the place and ask the guide about the attractions.

**For people to go out or go home** People may have difficulty leaving their home for a number of reasons. (e.g., health problems or disabilities). Using WithYou, they can virtually go outdoors or virtually go home.

### 4. Preliminary Experiment

To investigate how people communicate when they are actually together in an outdoor environment, as well as when communicating remotely using a videophone, we conducted experiments to examine their communication methods, i.e., the actions people take when they are together in an outdoor environment, and when they are using a videophone.

#### 4.1 Experiment of Going Out Together (Experiment A)

##### 4.1.1 Purpose

The aim of this experiment was investigate how people communicate with each other when they are outdoors and together at the same location. In this experiment, a pair of subjects went shopping together in the Akihabara electronics district of Tokyo, Japan. Subjects were able to choose their own mission target (such as to buy something or survey products), which they attempted to achieve during the experiment. The main purpose of this experiment was to identify the basic communication skills people use when they are outdoors together.

##### 4.1.2 Method

Four subjects (i.e., two pairs) participated in the experiment, during which they were encouraged not to consider that they were part of an experiment and, simply, to go out shopping together. **Figure 1** shows a still from the video recording. The experiment was performed at Akihabara, which was decided upon after taking the interests of the subjects into consideration.

##### 4.1.3 Conditions

The participants were briefly informed (for 10 minutes) about the task. They were asked to choose their own mission target, which in this experiment was to buy something or survey prod-



**Fig. 1** Scene from experiment A.

**Table 1** Interaction patterns observed in Experiment A.

Interaction pattern		Times
P1	Notice the direction in which their partner is facing then focus to the same direction.	12
P2	Pointing with a finger.	7
P3	Notice their partner standing still somewhere, and look where they are focusing their attention.	5
P4	Pick something up and look at it together.	5
P5	Describe using a gesture.	3
P6	Turn around to check that the partner is following.	2

ucts with a view to buying. Then the subjects had 20 minutes with which to achieve their task, and were able to move freely in the street and into stores.

During the experiment, two staff members followed the subjects; one videoed the subjects, while the other observed and noted the subjects' methods of communication and interaction. At the end of the task, the subjects were asked to fill out a questionnaire based on the experience.

We videotaped only part of this experiment (e.g., the length of video recorded was 13 minutes and 5 seconds, the time of one task in experiment A was 20 minutes), this was due to the fact that some stores did not permit filming.

#### 4.1.4 Results

We analyzed the result of the experiments (i.e., video, notes, and questionnaire), and identified six typical interaction patterns that appeared during the experiment; they are listed in **Table 1**, together with the frequency with which they occurred. The most important interaction pattern between the subjects was to pick something up and look at it together. Their targets were often electronic products, and they usually took various products in their hands and talked about the specifications and appearance.

In addition, we also found that "pointing with a finger" was an important interaction pattern between the subjects. One of the subjects indicates his interest by pointing with a finger at a sign or a product, which may not be picked up easily. For example, at one point, the two subjects were standing before a price list at a computer shop, with one subject pointing with a finger at the price of a product and talking to the other subject. Following this, a conversation about the price of that product was initiated.

Further patterns are outlined in Table 1. Note that the frequency of communication methods is based on the video data. The result of the experiment indicates that focus and gesture (pointing) are two important elements for activities when two subjects are out together. Based on these results, our system gives high priority to focus sharing and detection between the indoor

and outdoor users.

## 4.2 Experiment of Going Out Together by Videophone Call (Experiment B)

### 4.2.1 Purpose

The aim of this experiment was to observe how people communicate using a videophone when one is outdoors and the other is indoors. In this experiment, two subjects used a videophone to go out together virtually. The outdoor user went to a shopping center in Akihabara, and the indoor user remained seated in a different part of the same shopping center. During the experiment, the mission target was the same as in experiment A; however, the users had to communicate via videophone. The major purpose of this experiment was to observe people's communication skills during the experiment, and determine the differences from experiment A.

### 4.2.2 Method

Four subjects (2 pairs) participated in this experiment, which was performed at a shopping center at Akihabara, Tokyo, Japan. One subject remained at a rest place as an indoor user (Fig. 2, left), while the other subject walked around freely on all floors of the shopping center as an outdoor user (Fig. 2, right).

### 4.2.3 Conditions

The conditions in this experiment were mostly the same as those in experiment A. The difference was that the subjects communicated through a videophone call. To achieve the mission target (e.g., buying something), the indoor user requested the outdoor user to move to a specific floor or location, and to aim the camera at the target. The subjects had 15 minutes to achieve their mission target. During the experiment, both users were videoed, and the outdoor user's actions, gestures, and interaction patterns were observed and noted. However, different to experiment A, we took a video record during the whole experiment. The length of video recorded was 31 minutes and 23 seconds. At the end of the experiment, the subjects were asked to fill out a questionnaire.

### 4.2.4 Results

We identified a number of problems when using the videophone to communicate remotely, which limit the *out together feeling*. It was difficult for the indoor user to see what he wanted to look at. When using the videophone, the shooting direction is controlled entirely by the outdoor user. If the indoor user wants to view a place of the indoor user's interest, the indoor user must ask the outdoor user to move the videophone camera. In addition, the indoor user had difficulty knowing which direction the outdoor user was facing, which also made it difficult for the indoor user



Fig. 2 The outdoor user (left) and the indoor user (right) in experiment B.

to see what they wanted to look at.

Conversations were dominant during the experiment. This situation was partly attributable to the low quality and frame rate of the videophone image.

In experiment A, we found that the subjects often pointed using a finger. However, in experiment B, the most frequent interaction patterns were asking to change the camera direction i.e., the indoor user asked the outdoor user to turn the camera toward a specific direction.

Analysis of the video recording indicates that their purpose is essentially same; the subjects want to share the place of interest. In experiment B, the outdoor user often changed the direction of the camera instead of pointing with a finger.

Table 2 shows the frequency with which each method of interaction was used from an analysis of the video recordings. Tables 3 and 4 show the results of the questionnaire and user com-

Table 2 Frequency of interaction patterns in experiment B.

Interaction pattern		Times
P1	Outdoor user shows product information (such as price or design) to the camera.	7.5
P2	Outdoor user reads the information for indoor user.	4
P3	Indoor user requests the outdoor user to move somewhere.	4
P4	Outdoor user finds a product for the indoor user.	3.5
P5	Indoor user requests the outdoor user to turn the camera to a specific direction.	3
P6	Outdoor user shows something to the camera	3

Table 3 Questionnaire results of experiment B.

Question	Indoor user	Outdoor user
Q1 I felt a sense of doing something together with my remote partner via the videophone.	3	3
Q2 I felt that the videophone is suitable for realizing the <i>Out Together Feeling</i> .	2	2
Q3 Please rank the videophone image quality (1, low – 5, high).	2	2

Scores were given on a scale from 1 to 5, where 5 indicates strong agreement and 1 indicates strong disagreement.

Table 4 Questionnaire results (user comments).

Question	Answer
During the experiment, what type of interaction did you have most frequently with the indoor user?	(Outdoor user): Reading the price tag to the indoor user, directing the camera to the product.
Please give your general impression of this experiment.	(Outdoor user): Outdoor user turns camera → indoor user checks the live image and requests the outdoor user to change the facing direction.  (Outdoor user): No time to look at the indoor user. It felt no different from a voice call.  (Indoor user): Due to the low image quality and poor focus, small text on products was difficult to read.



ments from experiment B.

The questionnaire results listed in Table 3 show that neither the indoor nor outdoor user thought that the videophone was suitable for realizing the *out together feeling*. In addition, Table 3 also shows that the subjects thought that they neither agreed nor disagreed with the statement, “I felt a sense of doing something together with my remote partner via the videophone.” The user comments in the questionnaire also show that the indoor user frequently asked the outdoor user to turn the videophone camera when they wanted to see something.

#### 4.3 Basic Elements to Realize the Out Together Feeling

From analyzing the interaction patterns observed in experiment A (see Table 1), we noticed that (P1): “notice the direction in which their partner is facing then focus in the same direction” and (P3): “notice their partner standing still somewhere, and look where they are focusing their attention” both relate to the facing direction between partners. Knowing where one’s partner is looking is important. In addition, we observed that gestures, including (P2) pointing with a finger and (P4) picking something up to look at it together, were also important elements of non-verbal communication.

In experiment B, the interaction pattern: (P3): “indoor user requests the outdoor user to turn the camera toward a specific direction” (see Table 2) and the user’s comment: “outdoor user turns camera, then indoor user checks the live image and requests the outdoor user to change the facing direction” indicates that the indoor user frequently asked the outdoor user to turn the videophone camera. Feeling able to freely peruse the surroundings is an important element for achieving the *out together feeling*.

Although a number of aspects are clearly necessary to realize the *out together feeling*, we first define three basic requirements that are necessary for being aware of the existence and actions of one’s partner:

1. The indoor user must be able to freely and naturally peruse the surroundings of the outdoor user.
2. Each user must be able to perceive where the other user is looking without conversation. The focus of a user shows their interest, and it is important that this is conveyed without explicit verbal instruction.
3. Some kind of body actions and gesture are also important when outing together. People are not communicating only with verbal information. We should realize some non-verbal communication bi-directionally.

## 5. The WithYou System

To realize the basic elements of the *out together feeling*, we designed and implemented a system called WithYou. It assumes there are two users: one is outside (outdoor user) and the other is in a room (indoor user) (Fig. 3). The indoor user is defined as the person who uses the system to get the *out together feeling* to go outside virtually.

A wearable device with a pan-and-tilt camera and various sensors is mounted on the outdoor user’s chest. Live images from the outdoor user together with the direction in which they are facing are displayed on the indoor user’s HMD screen. They can also

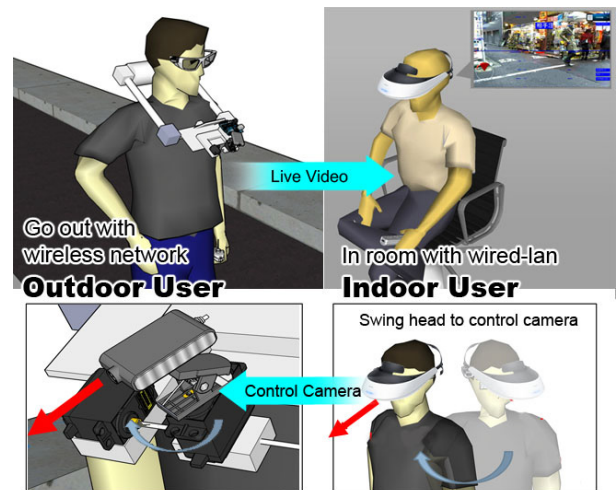


Fig. 3 System overview.

communicate by voice in WithYou. In addition, both of the users use a wireless hand controller to perform hand gestures, which are sent to each other.

The following describes three basic functions of WithYou which correspond to the three basic elements of the *out together feeling* (see Section 4.3):

1. Free viewing for the indoor user and its interaction methods of camera control (i.e., Indoor user is able to look around freely by turning his/her head)
2. Sharing the focus and its interaction (i.e., Both indoor/outdoor users know where the other user is facing and can detect the focusing status of both indoor/outdoor user and notify each other)
3. Gestural communication (i.e., Both users can communicate by performing gestures using the wireless controller)

### 5.1 Free Viewing for the Indoor User

In WithYou, the indoor user can view live images from the camera placed on the outdoor user’s chest. In addition, the direction of the camera is linked to the direction of the indoor user’s head. Thus, the indoor user can look around the surroundings of the outdoor user freely by turning his/her head (Fig. 3 Down). Zooming is also possible for the indoor user, and can be achieved by pressing on one of the buttons on the wireless controller.

More precisely, there are four viewing modes of the camera control:

**Relative view mode** In this mode, the absolute shooting direction is calculated relative to the direction that the outdoor user’s body is facing. In this mode, even if the indoor user does not change the direction of the camera, when the outdoor user turns their body, the absolute shooting direction of the camera is changed. This mode is useful when the indoor user just wants to look in the same direction as the outdoor user.

**Absolute view mode** This mode provides an absolute and stable view for the indoor user. In this mode, the indoor user’s view does not change if he don’t turns his/her head. This mode helps the indoor user to focus on something without being disturbed. This is achieved by compensating for the

motion of the outdoor user via the pan-and-tilt camera. For example, when the indoor user is watching something in front of the camera and the outdoor user turns his body to the right 30 degrees, the system rotates the camera to the left 30 degrees so that the absolute shooting direction of the camera does not change. Note that the pan-and-tilt camera is limited to 180 degrees of rotation (i.e., 90 degrees to the left and 90 degrees to the right), so the system cannot compensate if the outdoor user turns around. In this situation, the camera rotates to the limit, but then the absolute shooting direction changes. For example, if indoor users facing to the right 30 degrees then enabled “absolute view mode,” the correction range will be  $(90 - 30 = 60)$  degrees to the left, and  $(90 + 30 = 120)$  degrees to the right. The system will correct the camera’s shooting direction if the outdoor user’s turn is less than this range.

**“Follow Me” mode** In this mode, the direction of the camera is fixed to the direction that the outdoor user’s head is facing. Thus, the indoor user’s view follows that of the outdoor user. In this mode, the system displays the message “outdoor user’s view” to both users. This mode is helpful when the outdoor user wants to show something to the indoor user.

**“Pointing with finger” mode** In this mode, both users can control the camera using a wireless controller. Users can control the camera by pointing with the controller to the intended direction. This mode is helpful when the outdoor user wants to show something to the indoor user, and when the indoor user wants to look at something.

**5.2 Indoor-user Graphical User Interface (GUI)**

In order to share the focus between users, it is important to know in which direction the other user is facing. To inform the current status of the remote camera to the indoor user, a graphical user interface (GUI) is overlaid on the indoor user’s view (i.e., the camera image). The GUI shows the following information (see Fig. 4):

- The indoor user’s facing angle (green line)
- The outdoor user’s facing angle (red line) and focus point (red grid). Position of the red grid means the head’s facing of outdoor user (i.e., point of focus)
- Other information such as the tilt angle of the remote camera, which user is controlling the camera, the focus status of each user, the camera zoom, and system messages are also displayed.

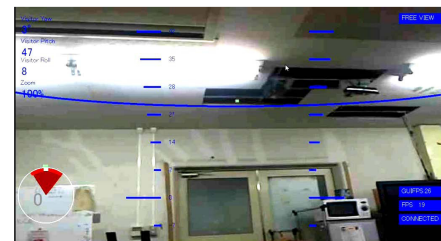
The length and direction of the green and the red line represents the horizontal facing angle of the indoor and the outdoor user, respectively. For example, if an indoor user faces front, the facing angle of the user is zero degrees, and the line is not displayed (just a dot is displayed at the center). If he turns to the right 60 degrees, the line from the center to the right is displayed, where its length is two-thirds (=  $60/90$ ) of the half width of the screen.

The indoor user can know easily if they are facing the same direction by checking whether the green and the red line are the same length and direction. In addition, a round shape displayed on the lower left of the GUI also shows the facing direction of



Indicator	Meaning
Green line	Horizontal facing angle of indoor (local) user.
Blue line	Vertical facing angle of indoor (local) user.
Red line	Horizontal facing angle of outdoor (remote) user.
Red grid	Point of focus of outdoor (remote) user.
Blue box on right	(Up) Focusing status of indoor/outdoor user and notification of joint attention. (Down) connection status and current FPS number.

Fig. 4 GUI on the indoor user’s HMD screen.



when the indoor user is facing up



when the indoor user is facing front



when the indoor user is facing down

Fig. 5 Relation of GUI and indoor user’s facing direction.

both the indoor and outdoor users.

On the other hand, the vertical camera direction is represented by the horizontal blue line. If the indoor user changes his/her vertical facing direction (i.e., facing up or down), the vertical position of the blue line moves up or down. The green and red indicator also follows the blue line. Figure 5 shows the relation of GUI and the indoor user’s facing direction.

**5.3 Sharing the Focus and Joint Attention**

Sharing the focus means that both users know where the other

is looking. As described above, the angle that the outdoor user is facing is displayed on the indoor user's GUI as a red line (see Fig. 4). The outdoor user can monitor the direction in which the indoor user is facing by observing the camera mounted on their chest.

However, these functions are not sufficient to enable both users to focus on the same object or location (i.e., joint attention). It remains difficult to communicate the focus and, in particular, to be confident that the other user is focusing on the same object. In order to achieve joint attention, we designed the interface to notify one user of the focus status of the other.

To help the users to achieve joint attention, it is important to share not only the focus but also the focusing status. We assume a user is in "focusing" mode if the direction in which the user is facing is rotated from the center by more than 15° and the focus remains static for more than three seconds. When a user enters focusing mode, the system sends a notification and plays a sound to the other user. This provides a hint of the partner's actions, and may be a prompt for a topic of conversation.

In addition, the system also notifies the users when joint attention has been achieved. When one user is in focusing mode, and the other user focuses in the same direction, the system recognizes this situation as joint attention, and sends a notification to both users. With this notification, they are aware that they are looking in the same direction, which aids remote communication (see Fig. 6).

5.4 Gestural Communication

Gestural communication, including physical touching, is frequently used in addition to vocal conversation. Examples include tapping on the shoulder or waving the hands. These are also important to realize the *out together feeling*. WithYou uses a wireless motion sensor device to achieve gestural communication, and the user performs a gesture by grasping the device.

The system analyzes the acceleration data from the motion sensors, recognizes the operation, and sends the result to the other user. For example, the user can virtually tap their partner on the shoulder by shaking the device up and down to imitate the action of tapping on the shoulder. When this gesture is identified, the system plays a sound, vibrates the device, and shows a text

message to inform the remote user. The user can also perform a hand waving gesture by shaking the device left and right. When this motion is identified, the system displays a hand waving animation, and plays a sound. Just sending a notification is also possible by pressing a button on the controller, which displays a message and plays a sound at the remote side.

6. Implementation

6.1 System Overview

Figure 7 shows the system overview. It consists of basically two parts: the outdoor user's device and the indoor user's device. They communicate via a network (which may be the Internet). For example, a live image from the outdoor user device is sent to the indoor user's device. Various sensor data such as facing directions, focusing status, and system messages are sent to each other. Details of those devices are described later in this section.

6.2 Wearable Device of the Outdoor User

Figure 8 shows the wearable device for the outdoor user. It consists of a gyro sensor, two geomagnetic sensors, a pan-tilt camera, a mono LCD monitor, and a wireless hand controller. It is worn around the outdoor user's neck. He also carries a mobile computer on his back. Figure 9 shows the LCD monitor that displays the status of the system and the facing angle of the indoor user.

A camera is mounted on the chest of outdoor user and, therefore, the shooting direction will not change if the outdoor user turns their head. We chose to place the outdoor user's camera on their chest because it is more stable there than on other body

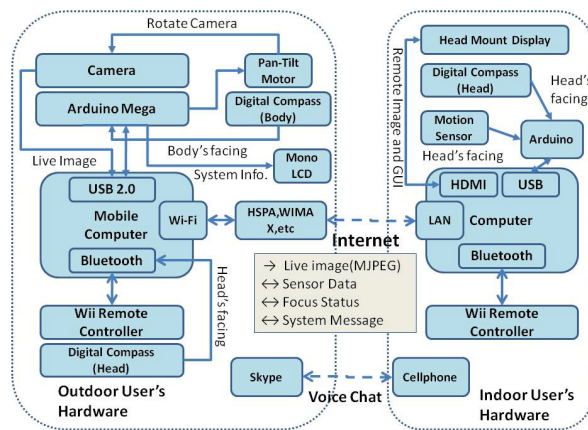


Fig. 7 System hardware overview.

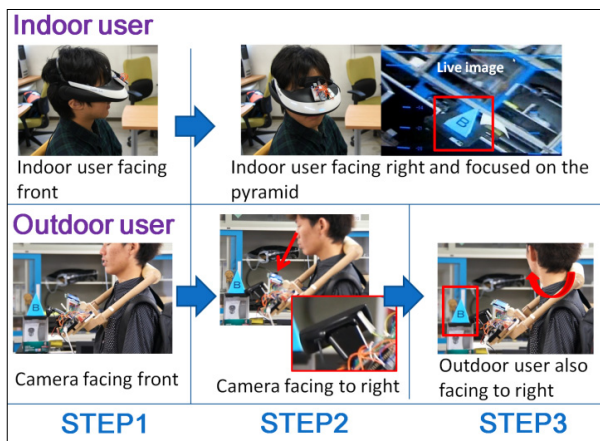


Fig. 6 Flow of joint attention.

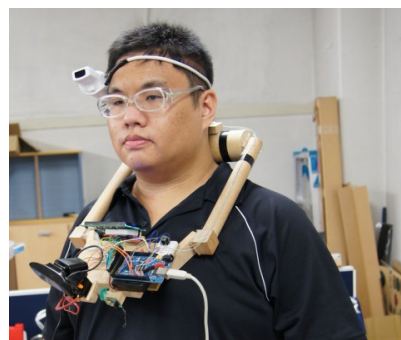


Fig. 8 Wearable device of the outdoor user.



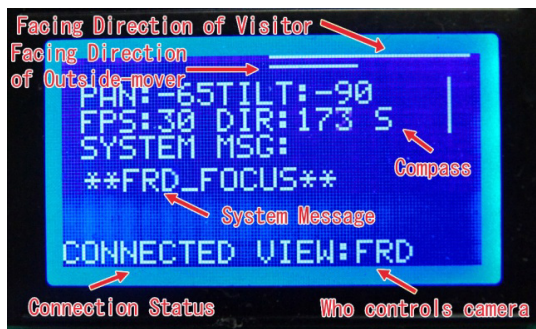


Fig. 9 LCD on the outdoor user side.



Fig. 10 Wearable device of the indoor user.

parts, such as the head or shoulders.

To achieve rapid and wide rotation of the pan-and-tilt camera worn by the outdoor user, we use two high-speed servomotors to control the two axes of rotation. The camera can pan 180 degrees and tilt 130 degrees. A USB camera (Logicool C910) was mounted on the motor system, and an embedded microprocessor (Arduino-mega) controlled these servomotors. In addition, the camera has a built-in digital-zoom function. Therefore, the indoor user can zoom in or out of remote images.

We used two geomagnetic sensors (i.e., digital compasses) to detect the direction that the outdoor user is facing – one for the body and one for the head.

### 6.3 Wearable Device of the Indoor User

The indoor user wears an HMD, as shown in Fig. 10, and holds a wireless hand controller. A geomagnetic sensor — a 3-axis gyro sensor and a 3-axis motion sensor — are mounted on the HMD to measure the horizontal and vertical direction in which they are facing. This direction is linked to the direction of the camera that the outdoor user wears.

### 6.4 Measuring the Facing Direction

WithYou senses the directions that both the indoor and the outdoor user are facing. For the indoor user, a geomagnetic sensor (digital compass) is used to measure the horizontal direction in which the user is facing. Since the geomagnetic sensor measures the absolute angle, we need to convert this into a relative angle, compared with the angle that the outdoor user is facing. To realize this, the indoor user can decide his/her “front” direction anytime simply by pressing the up button on the hand controller. After the “front” direction is determined, the relative angle can be calculated by subtracting the current absolute direction by the absolute front direction. The remote camera rotates relative to its front di-

rection. The direction the outdoor user is facing is also measured using two geomagnetic sensors: one for the body and the other for the head. The relative angle of the head is calculated from the difference between two sensors. The absolute and relative directions of the outdoor user are displayed on the indoor user’s GUI.

### 6.5 Wireless Hand Controller

Both the indoor and outdoor users hold a wireless controller (we used a Nintendo Wii remote controller). The controller has several buttons, a 3-axis gyro sensor, and a 3-axis motion sensor. It was used to control the system settings, for gestural communication and to communicate a pointing direction.

### 6.6 Video/Voice Transmission

The system transmits Motion JPEG data ( $680 \times 480$  pixels) for video communication. The system can transmit up to 25 FPS under wireless local area network (LAN), which corresponds to approximately 100 KB per frame, so 2.5 MB per second for video data. We chose Motion JPEG instead of other advanced video compression protocols, such as H.264/MPEG-4, because Motion JPEG results in less transmission delay.

The system monitors the frame rate and adjusts the video compression and image resolution accordingly to ensure that the video image will be successfully transmitted. We used the JPEG encoder provided by the Microsoft .NET Framework 4.1, where the compression ratio can be adjusted on the fly. In our default settings, the compression ratio was high (i.e., a low data rate) when frame rate is lower than 10 FPS. Also, the image resolution will resize to QVGA ( $320 \times 240$  pixels) when the frame rate is lower than 5 FPS.

For voice communication, we used Skype for the indoor user and a cell phone for the outdoor user. In addition to the built-in audiovisual (AV) transmission features, the user may choose to use an alternative AV conference application, such as Skype, too. In this case, the GUI will overlay to the relevant screen and allow background transparency.

## 7. Evaluation

### 7.1 Purpose

The aim of this experiment (experiment C) was to assess how the communication is changed, compared with experiments A and B, when the WithYou system is used. In other words, this experiment investigates the effectiveness of the prototype system. We evaluated the system in Akihabara. During the experiment, subjects chose their own mission target (e.g., buying something), and used the system freely to achieve their goals. In this experiment, the outdoor user went out and moved around, while the indoor user remained inside a room, and went out virtually using our system.

### 7.2 Method

Four subjects (two pairs) participated in this experiment. One of the pair (the outdoor user) went outside and walked around (see Fig. 11, right), while the indoor user sat in a room (see Fig. 11, left). In this evaluation, all subjects performed both roles.

Before the task began, the subjects practiced using the system



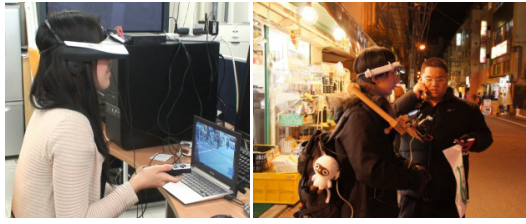


Fig. 11 Indoor and outdoor users' environments in the evaluation.

for 10 minutes. The task in this experiment was to buy something interesting. The indoor user was asked to determine what to buy. The subjects had 20 minutes to achieve the task. The outdoor user was able to move freely in the street and into stores. The indoor user may ask the outdoor user to enter to a store, take something in their hand, and then show it to the camera.

### 7.3 Conditions

This evaluation was executed over two days. On the first day, two subjects stayed in the experiment room, and the other two subjects went out as outdoor users. On the next day, they reversed roles, and conducted the experiment in the same way. Thus, we had two pairs and four experimental sessions. We recorded videos of both subjects. At the end of the task, the subjects were asked to fill out a questionnaire.

The indoor user's system was connected to a wired LAN, and the outdoor user's system was connected to a WIMAX wireless network in the street. Because the WIMAX connection may not be stable and had limited bandwidth, the outdoor user also used a cell phone to communicate verbally with the indoor user. The indoor user used Skype to dial up to the outdoor user's cell phone.

### 7.4 Results

Since the subjects had enough training and explanation, all subjects were able to operate the system comfortably and use all the system functions. During the experiments, the system worked well most of the time, but the system have been broke down at one time during Experiment C.

The frame rate of the video transmission varied during the experiment in the range 1–14 FPS, depending on the time and the location. When the frame rate was low, the indoor users had difficulty knowing which direction they were viewing. In such cases, the outdoor users tended to walk slowly so that the indoor user could recognize the situation. The indoor users frequently asked the outdoor user to stop for a second because they wanted to look around. These interaction patterns are interesting because they are similar to those observed when both partners were actually outdoors. Overall, the subjects communicated well during the experiment.

Tables 5 and 6 show the results of the questionnaire. From the results, three basic elements of *out together feeling* we defined above are mostly achieved by the system WithYou. The first question in Table 5 shows that both users felt a sense of doing something together. The comments of the subjects listed in Table 6 indicate that both users felt they were doing something together, and that they were interacting with each other. The comments of the subjects also show that they all used system features

Table 5 Questionnaire results.

Question		Indoor user	Outdoor User
Q1	I felt a sense of doing something together with my remote partner.	4.5	4.5
Q2	I was able to look around at the surroundings freely.	3.75	N/A
Q3	I was able to determine the direction my remote partner was facing.	3.75	4.5
Q4	Using the feature "Absolute view for indoor user", I felt that the remote video was stable and that the camera always faced in the required direction.	3.75	N/A
Q5	Did you use the system effectively during experiment? Were you able to operate all the system features?	4	3.5
Q6	Did you think the message of entering focus mode was notified at the correct moment?	4	2.6
Q7	I felt that "Joint attention" was useful and improved activities together with my remote partner.	3	4
Q8	Did you feel the remote image is stable?	3.5	N/A

Scores were given on a scale from 1 to 5, where 5 indicates strong agreement and 1 indicates strong disagreement.

Table 6 Questionnaire results (user comments).

Question	
What was the situation in which you most felt a sense of doing something together with your remote partner?	
character	Answer
Indoor user	When buying something and conversing with my remote partner.
	I felt realistically together when wearing the HMD and seeing a live image.
	At first, I felt I was just watching a video in which the outdoor user is walking. However, after a while I felt like I was I going shopping with them when I began conversation with them.
Outdoor user	When buying something, the situation seems like a voice call; however, the remote partner sees the same thing as I do, so I don't need to explain it.
	When showing something to the indoor user and getting a reaction from them.
	When I was walking, the indoor user focused on a poster at my side and then called to me. After that, I used the "follow me" feature to show the poster to the indoor user.

fully to complete their own mission target (i.e., go shopping and buy something).

#### 7.4.1 How the Out Together Feeling is Achieved

This section describes the details about how the three basic elements of *out together feeling* are achieved in WithYou.

##### Free viewing for indoor users

From the second question (Q2) in Table 5 and the answers in Table 6, the average score of the indoor users is 3.75, the indoor users were able to look around at the surroundings, and both users succeeded in sharing the environment with their partner. We often found that, for example, when the outdoor user approached a store,

the indoor user turned his head to check a price tag or look at some goods placed outside the store, which initiated a conversation. One major difference between this experiment and the videophone communication (experiment B) was that the outdoor user could use both hands, which made it possible to perform gestural interaction such as picking something up and looking at it together more easily than when using the videophone.

**Sharing the focus of each other and joint attention** For the third question (Q3) in Table 5, the average score of the indoor users is 3.75 and the average score of the outdoor users is 4.5. This indicates that both side of users were able to recognize each other's facing direction using our system. However, the score of indoor users is lower than outdoor users' score. This is because outdoor users can grasp the direction of the partner's view by checking the shooting direction of the camera, which is easier to understand than the GUI in the indoor user's view.

From the data in Table 5, the question 6, "Did you think the entering focus mode was announced at the correct moment?," was scored 4 by the indoor users and 2.6 by the outdoor users. This shows that the indoor users understood the meaning of this function well, and felt it worked effectively. One of the outdoor users commented that, "Sound notifications helped me to know that the indoor user was focusing on something." This subject also gave 5 points for questionnaires. On the other hand, other outdoor users experienced difficulty in noticing the focus status. One reason for this is that the outdoor environment was noisy, which made it difficult to hear the notification sound.

**Table 7** shows the frequency of "focusing" during the experiment. This was calculated by inspecting the GUI record in the indoor user's view. The results show that the average number of outdoor user focusing events was larger than those of the indoor user. This is because the indoor users did not turn their head without reason and usually face forward, which is not detected as a focusing state. Indoor users focused at something only when they wanted to see specific products or street scenes. However, the outdoor users turned their head more frequently to see what was part of their surroundings, which led to a greater frequency of detecting a focusing state. The targets of the outdoor users' focus also had more variety. They checked traffic signs, stores or prod-

**Table 7** Frequency of focus interaction.

Times of focus interaction	Indoor User	Outdoor User
Average times of entered "Focusing status" of indoor & outdoor users.	16.25	40.25
Times of joint attention.	5	
Times of entered "Focusing status" of subject A.	15	35
Times of entered "Focusing status" of subject B.	3	19
Times of entered "Focusing status" of subject C.	23	54
Times of entered "Focusing status" of subject D.	24	53

ucts, or just faced a new direction apparently without reason. This difference may be because of the wider field of vision. The use of a wider field of view of the HMD may address this problem.

We also noticed that the "joint attention" mode was not frequently detected during the experiment. This is because the outdoor users' focusing time was typically shorter than expected. From the videos, we observed that the outdoor user constantly monitored the direction of the chest-mounted camera, and noticed the pan-and-tilt rotations immediately. When they noticed the rotation of the camera, they turned their head to the same direction as the camera, but only for a very short time, which was not detected as a focusing state. Therefore, we should recalibrate the threshold of the focusing state so that more "joint attention" events can be detected.

**Gestural interaction** As described above, outdoor users used some (real) gestures many times (See P2 and P4 in **Table 8**). They often point to their interested things and talk to the indoor user. On the other hand, indoor users often used the notification function that sends a notification by pressing a button (see Section 5.4 and P5 in **Table 8**). Other gesture functions such as waving a controller are not well utilized in the experiment. These gestures seems extravagant for the users just for notifying a remote user. More natural and useful gesture functions should be designed to enhance the *out together feeling*.

#### 7.4.2 Comparison with the Experiment A and B

Table 8 shows the frequency of each interaction pattern during the evaluation from analysis of the video recording. In addition to the focus interactions provided by the WithYou system, we observed other important interaction patterns. The outdoor user often picked something up and showed it to the camera (P2); in doing so, the outdoor user stood in front of a product and waited for a short time (P4). Often, the outdoor user stood in front of showcases and remained still for a time, which was interpreted as allowing the indoor user to obtain a stable image and allow them to know where they should focus. The outdoor user also often showed something to the indoor user directly. Outdoor users utilized the camera to show something or to point to somewhere with their finger.

**Table 8** Interaction patterns in evaluation.

Interaction pattern	Times
P1 Outdoor user noticed the direction in which the indoor user was facing and then turned his/her head accordingly.	51
P2 Outdoor user picked something up and showed it to the camera.	36
P3 Outdoor user stood in front of a product and then waited for a while.	26
P4 Outdoor user found a product for the indoor user.	14
P5 Indoor user sent a message to notify the outdoor user of something.	13
P6 Indoor user determined what to buy and outdoor user bought the item.	6
P7 Outdoor user pointed something out using their finger.	4

**Table 9** Comparison of interaction patterns in experiment A and C.

Interaction patterns in experiment A	Interaction patterns in experiment C
Notice the direction in which their partner is facing then focus to the same direction.	Joint attention. Outdoor user noticed the direction in which the indoor user was facing and then turned his/her head accordingly.
Notice their partner standing still somewhere, and look where they are focusing their attention.	Outdoor user stood in front of a product and then waited for a while.
Pick something up and look at it together.	Outdoor user picked something up and showed it to the camera.
Pointing with a finger.	Outdoor user pointed something out using their finger.

Comparing the result of the experiment A and C, we noticed some important interaction patterns from experiment A that were also employed in experiment C, albeit in a slightly different manner. **Table 9** lists these interaction patterns and describes the differences.

We also noticed a number of similar interaction patterns between experiments B and C, where the indoor users lost their desired viewing direction (i.e., where they were looking at via the camera). Although the indoor users can control the camera freely using our system, they still had difficulty in some cases knowing which direction the outdoor user was facing and where the outdoor user was, especially when the outdoor user moved a lot.

## 7.5 Discussion

### Stability of the camera:

In the experiment, the stability of the camera image for indoor users is tolerable. From the questionnaire data in Table 5, the question 8, “Did you feel the remote image was stable?,” scored 3.5 by the indoor users. However, by reviewing subjects’ comments, we found that three of the four subjects said that they thought the remote image displayed to the indoor user was stable. The chest is relatively stable position than the head or the shoulder. As described in Section 5, in the absolute view mode, the camera is controlled to cancel the panning movement of the outdoor user’s body. However, it does not work as an anti-shake image stabilizer. Moreover, in the current implementation, vertical movement of the body is not canceled. Incorporating an image stabilization mechanism will improve the experience of indoor users.

### Other comments from the subjects:

During the street evaluation, we noticed that indoor users frequently used the “pointing with a finger” feature (i.e., using the wireless controller to point out a direction) when they wanted to look at something. Although indoor users can control the remote camera by turning their head, they became fatigued when facing a particular direction for an extended period of time. For this reason, he/she learned to use “pointing with a finger” instead of turning his/her head for this situation.

At the end of evaluation, we received multiple comments that

it may be helpful also to allow the outdoor user to see the indoor user via a camera. The outdoor user had difficulty in reaching the same level of *out together feeling* as the indoor user. Allowing the outdoor and indoor users to look at each other may further enhance the feeling of sharing an activity together.

## 8. Conclusions and Future Work

In this paper, we defined the concept of *out together feeling*, a sensation shared by two people at different locations that feels like they are going out together. As a step toward the achievement of this concept, we have extracted three basic interaction elements from the experiments that we have designed and performed to examine what types of interaction (communication skills) people take when they actually go outside together, and when they go out together virtually via videophone.

After that, we have designed and implemented three core interaction methods to achieve them:

- Free viewing for the indoor user and interaction methods for the camera control. Indoor users are able to look around freely by turning his/her head.
- Sharing the focus between both side of users. Both users know where the other user is facing to. They system detects the focusing status of both indoor/outdoor user and notifies them.
- Gestural communication. Both indoor/outdoor users can communicate each other with gestures.

We also have performed a street evaluation of our system. WithYou was evaluated positively by the subjects, and mostly achieved the basic interaction elements to achieve the concept of *out together feeling*.

In future studies, we are planning to implement new functions to enhance the *out together feeling*. For example, in the current implementation, the outdoor user was not able to experience the same level of *out together feeling* as the indoor user. This is attributed to the greater focus on bringing the experience of the outdoor environment to the indoor user. In future studies, to further enhance the *out together feeling*, we plan to implement features that allow the indoor user and outdoor user to look at each other using video cameras.

## References

- [1] Tsumaki, Y., Fujita, Y., Kasai, A., Sato, C., Nenchev, D.N. and Uchiyama, M.: Telecommunicator: A Novel Robot System for Human Communications, *Proc. 11th IEEE International Workshop on Robot and Human Interactive Communication*, pp.35–40 (2002).
- [2] Kashiwabara, T., Osawa, H., Shinozawa, K. and Imai, M.: TEROOS: A wearable avatar to enhance joint activities, *Proc. SIGCHI Conference on Human Factors in Computing Systems, CHI '12*, pp.2001–2004 (2012).
- [3] Ohta, S., Yukioka, T., Yamazaki, K., Yamazaki, A., Kuzuoka, H., Matsuda, H. and Shimazaki, S.: Remote instruction and support using a shared-view system with head mounted display (HMD), *Japan Science and Technology Agency* (in Japanese), pp.1–7 (2000).
- [4] Kuzuoka, H., Oyama, S., Yamazaki, K., Suzuki, K. and Mitsuishi, M.: GestureMan: A Mobile Robot that Embodies a Remote Instructor’s Actions, *Proc. CSCW '00*, pp.155–162 (2000).
- [5] Adalgeirsson, S.O. and Breazeal, C.: MeBot: A robotic platform for socially embodied telepresence, *Proc. 5th ACM/IEEE International Conference on Human-robot Interaction* (2010).
- [6] Michaud, F., Boissy, P., Corriveau, H., Grant, A., Lauria, M., Labonte, D., Cloutier, R., Roux, M.-A., Royer, M.-P. and Iannuzzi, D.: Telepresence robot for home care assistance, *Proc. AAAI* (2006).



- [7] Koizumi, S., Kanda, T., Shiomi, M., Ishiguro, H. and Hagita, N.: Preliminary field trial for teleoperated communication robots in Robot and Human Interactive Communication, *Proc. 15th IEEE International Symposium*, pp.145–150 (2006).

### Editor's Recommendation

The initial version of this paper was reviewed by three reviewers and has received excellent scores; especially, its coolness scores were high. The authors developed a system that enables a new style of collaboration with which a person at a shop and his/her virtual collaborator at remote site work together to make their shopping decisions. This paper shows a direction of new style of collaboration support technologies.

(Chairman of SIGGN Minoru Kobayashi)



**Ching-Tzun Chang** is a Ph.D. candidate in computer science at University of Tsukuba. His research interests include wearable robots, communication Support, and Tele-presence. He received a B.S. in computer science at National Taipei University of Technology and an M.S. in computer science at University of

Tsukuba in 2006 and 2011 respectively.



**Shin Takahashi** is an Associate Professor of Department of Computer Science, University of Tsukuba. His research interests include user interface software and ubiquitous computing. He received his B.Sc., M.Sc., and Ph.D. in information science from The University of Tokyo in 1991, 1993, and 2003. He is a member of

ACM, IPSJ, and JSSST.



**Jiro Tanaka** is a Professor of Department of Computer Science, University of Tsukuba. His research interests include ubiquitous computing, interactive programming, and computer-human 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.