Waseda University Master Thesis

A Remote Pair Sightseeing System Supporting Free Viewpoint and Gestural Communication

Minghao CAI
44151588-6

Supervisor: Professor Jiro TANAKA

Master (Engineering)

Interactive Programming
Information Architecture
The Graduate School of Information, Production and Systems

July 2017

ABSTRACT

In this research, we propose a remote pair sightseeing system for a local user and a remote user. It allows both users have independently free viewpoints of the scenery. It supports a gestural communication between the two of users. With the integration of Head-mounted Display and Depth Camera, we allow the local user to perform a gestural interaction with the remote user on top of the remote scene while each user is provided an independent free viewpoint. Through this system, two side of users could get a feeling that they are truly walking outdoor together side by side for a trip. We carried out a preliminary user study to evaluate our design and the performance of the system. A positive feedback has been received.

Keywords: Artificial, augmented, and virtual realities; Virtual sightseeing; Remote communication; Gestural interaction; Panoramic Viewing; Feeling together

ACKNOWLEDGMENTS

I would first like to thank my supervisor Professor Jiro Tanaka. The door to Prof. Tanaka office was always open whenever I ran into a trouble spot or had a question about my research or writing. He steered me in the right direction whenever he thought I needed it.

I would also like to thank the volunteers who joined the experiment of this research. Without their passionate participation and feedback, the evaluation of the system could not have been successfully conducted.

Finally, I wish to express my gratitude to the other members of my laboratory for providing me with unfailing support and continuous encouragement throughout my days of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Contents

Table of contents **i**

List of figures **iii**

List of tables **vii**

1 Research Background **1**
1.1 Remote Communication 1
1.2 Gesture Recognition 2

2 Goal and Approach **5**
2.1 Target Problem 5
2.2 Research Goal 5
2.3 Research Approach 6

3 System Design **8**
3.1 Trip-together Feeling 8
3.2 The 360°Panoramic Browsing 10
3.3 Attention Reminder 12
3.4 Air Gestural Input 14
3.4.1 Tracking 14
3.4.2 Human-skin Hand Model 15
3.5 Gestural Navigation 18
3.5.1 Six Direction Gestures 19
3.5.2 Warning Gestures 25
3.6 Pointing Assistance 27

4 Implementation **29**
4.1 System Hardware Overview 29
4.2 Local Setup 30
4.2.1 Head-mounted Display 31
4.2.2 Depth Camera 32

4.3 Remote Setup . . . 33
4.3.1 Smart Glasses . . . 34
4.3.2 Spherical Camera . . . 36
4.4 Live Panorama of the Remote world . . . 38
4.5 Depth-based Recognition . . . 42

5 Related Work **44**

6 Preliminary Evaluation **46**
6.1 Participants . . . 46
6.2 Method . . . 47
6.3 Conditions . . . 47
6.4 Results . . . 49

7 Conclusion and Future Work **51**
7.1 Conclusion . . . 51
7.2 Future Work . . . 52

Bibliography **53**

List of Figures

1.1 Remote video communication system 2

1.2 Architecture of general vision-based hand gesture recognition 3

1.3 Architecture of general depth-based gesture recognition 4

2.1 The wearable device of the local user (a) is a head-mounted display with a depth camera attached on the front side. The portable setup of the remote user includes a pair of smart glasses and a spherical camera. 7

3.1 A local user (a) remains indoor having an immersive virtual sightseeing with a remote partner (b) who goes outdoor with a portable setup. (c) shows the users feel like they are have a trip together 9

3.2 When the local user looks around, his/her viewpoint turns upward accordingly. The viewpoint is naturally controlled by the head movement just like being personally on the scene. 11

3.3 In (a), the local user turns his head and sees the remote user is making a hand gesture as shown in (b). 12

3.4 When the two users are viewing in the same direction, a joint attention signal would be shown in the center of the local user's view. 13

3.5 The visualization of the remote user's field of vision when a joint attention signal notifying the joint attention moment. 14
3.6 We develop a 3D human-skin hand model associated with the bone structure of the user. 15
3.7 The 3D human-skin hand models presents on top of the scenery. 16
3.8 The local user is making an air gesture (a). He could make a gestural input in the remote scenery with the First-person Perspective. 17
3.9 Visualization example of the remote user's field of vision. The local user's hands present on the left side, superimposing on the physical world. 18
3.10 Calculating the include angle between bones to determine the finger states. 19
3.11 Subgraph (a) shows the physical hand of the local user performing a "Forward Direction" gesture. Subgraph (b) shows the gesture in the local user's view. Subgraph (c) is the visualization of the remote user's field of vision. 20
3.12 Subgraph (a) shows the physical hand of the local user performing a "Back" direction gesture. Subgraph (b) shows the gesture in the local user's view. Subgraph (c) is the visualization of the remote user's field of vision. 21
3.13 Subgraph (a) shows the physical hand of the local user performing a "Leftward" direction gesture. Subgraph (b) shows the gesture in the local user's view. Subgraph (c) is the visualization of the remote user's field of vision. 22
3.14 Subgraph (a) shows the physical hand of the local user performing a "Rightward" direction gesture. Subgraph (b) shows the gesture in the local user's view. Subgraph (c) is the visualization of the remote user's field of vision. 23
3.15 Subgraph (a) shows the physical hand of the local user performing a "Up" direction gesture. Subgraph (b) shows the gesture in the local user's view. Subgraph (c) is the visualization of the remote user's field of vision. 24

3.16 Subgraph (a) shows the physical hand of the local user performing a “Down” direction gesture. Subgraph (b) shows the gesture in the local user’s view. Subgraph (c) is the visualization of the remote user’s field of vision. 25
3.17 Subgraph (a) shows the "OK" Gesture in the physical world. Subgraph (b) shows the local user’s view. Subgraph (c) is the visualization of the remote user’s field of vision. 26
3.18 Subgraph (a) shows the "Wait" Gesture in the physical world. Subgraph (b) shows the local user’s view. Subgraph (c) is the visualization of the remote user’s field of vision. 27
3.19 Subgraph (a) shows the local user is pointing at a statue in the scene. Subgraph (b) is the visualization of the remote user’s field of vision. 28

4.1 Hardware Overview 29
4.2 Local setup overview 30
4.3 Head-mounted display 32
4.4 Depth camera 33
4.5 Remote setup overview 34
4.6 Augmented reality smart glasses 35
4.7 Exchange information with the local side via the Internet 36
4.8 GUI superimposes on the physical world 36
4.9 Spherical camera 37
4.10 The spherical camera is set on the top of a metal rob 38
4.11 Live panoramic video stream 39
4.12 Video stream setting 39
4.13 Dual-fish eyes video 40
4.14 Transparent spherical model 41

4.15 Spherical panoramic view 41

4.16 Depth data 42

4.17 Visualization of depth data of the user's hand 43

6.1 Questionnaire 48

6.2 Questionnaire results 49

List of Tables

4.1 Local Desktop PC 31
4.2 Mobile computer 40
5.1 Comparison between *WithYou* and this system 45

Chapter 1

Research Background

1.1 Remote Communication

Nowadays, with increasingly geographically separated social networks, high-speed Internet and mobile communication techniques make it possible to keep in touch with someone conveniently [22]. Nonetheless, the potential of mobile video communication has yet to be fully exploited. Commercial video communication systems mostly developed for a face-to-face communication which helps little to focus on the other information like body language or the ambient or distant objects. Additionally, although might possible with current technologies, there are few communication platforms offer a way for users to achieve effective gestural communication. When users want to describe the objects or directions in the scene, only using verbal description might be challenging. Such constraints make it difficult for users to get a common perception or feel like staying together.

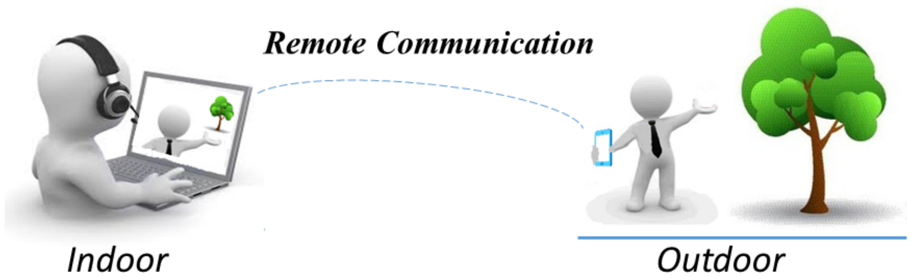


Figure 1.1 Remote video communication system

1.2 Gesture Recognition

The human hand has a complex anatomical structure consisting of many connected parts and joints, involving complex relations between them providing a total of roughly 27 degrees of freedom (DOFs) [30]. User Interface development requires a sound understanding of human hand's anatomical structure in order to determine what kind of postures and gestures are comfortable to make. Although hand postures and gestures are often considered identical, the distinctions between them need to be cleared. Hand posture is a static hand pose without involvement of movements. For example, making a fist and holding it in a certain position is a hand posture. However, a hand gesture is defined as a dynamic movement referring to a sequence of hand postures connected by continuous motions over a short time span, such as waving good-bye. With this composite property of hand gestures, the problem of gesture recognition can be decoupled into two levels- the low level hand posture detection and the high level hand gesture recognition [7].

In traditional vision-based hand gesture recognition system, the capture of the hand images is recorded by 2D camera(s).It limits to the static hand posture capture without involvement of movements.For gestures recognition, estimation from a series of continuous images could not guarantee a stable and high accuracy.

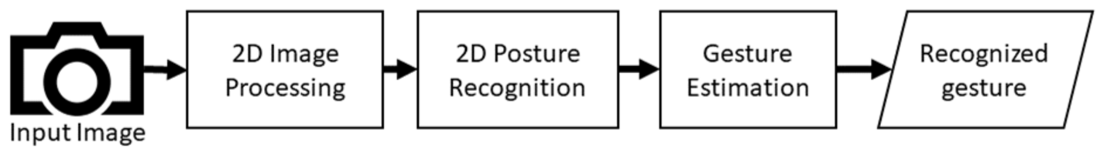


Figure 1.2 Architecture of general vision-based hand gesture recognition

Currently, low-cost depth sensors facilitate a practicable way for gestural recognition, giving a new opportunity for research. The operating principle of the measurement of optical depth sensors could be, in principle, divided into the three mechanisms: Structured Light, Time of Flight and Stereo-Vision. Structured light sensors analyze the deformation of a known pattern onto an unknown surface to determine the three-dimensional shape [25]. The Time of Flight(TOF) 3D cameras are based on the well-known time of flight principle [18]. Stereo Vision cameras consist of two optical 2D cameras with known extrinsic parameters. The concept of determining the depth in the scene is based on searching correspondence points in both 2D images [1].

The general depth-based gesture recognition system (shown in Figure 1.3) encompasses three main steps. Firstly, the samples corresponding to the hand region are extracted from the depth map and further subdivided into palm and fingers samples, some new technique even into finger bone samples.Then, features such as spatial position and directions are extracted. Finally, heuristic algorithms or machine learning techniques are used to recognize the different gestures [6, 16].

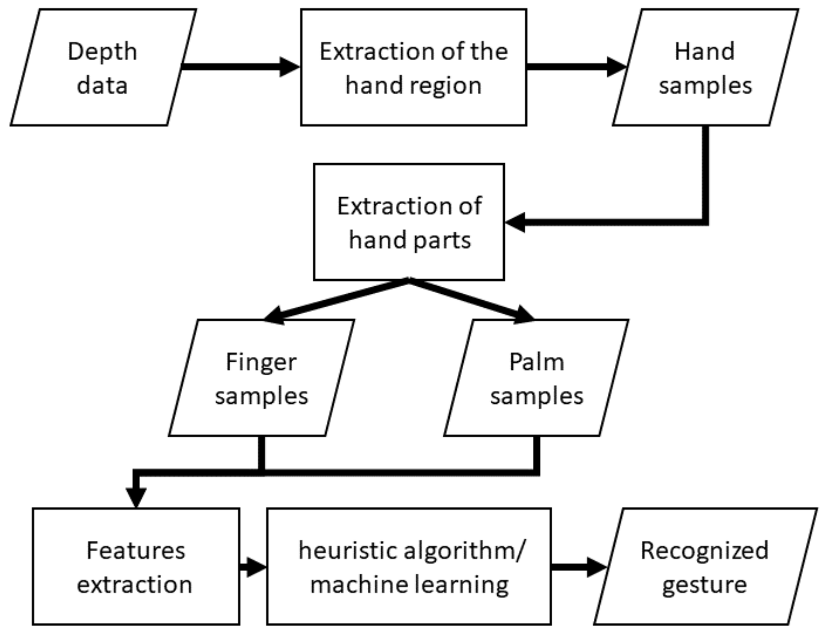


Figure 1.3 Architecture of general depth-based gesture recognition

Depth-based recognition technique detects the gestures from depth data without computing the full pose, by applying heuristic algorithms or machine learning techniques. It allows capturing not only the real-time shape and appearance but also the moving data, rotation and orientation of physical objects. Besides helps realize the expansive possibilities of 3D mobile gestural communication, depth-based recognition technique makes the physical environment a core part of communication. In this research, we use depth-based recognition for the gesture detection with a new generation depth camera.

Chapter 2

Goal and Approach

2.1 Target Problem

The problem we are targeting is helping the users in separated position get a feeling of being together during a mobile communication. Some previous researchers have demonstrated that hand gesture is helpful in remote communication in different approaches [26, 28, 11, 8]. We find that users intend to use hand gestures to describe direction information or point out objects especially in the spatial scene, which might make the conversation smoothly. For example, imagine receiving a video call from your parents who live in distant hometown, asking to buy a local specialty in the market. You might walk around and ask which one they like. Rather than just using some scanty expressions like "that one", "over there", it is a better idea that they could point out something satisfactory directly on the scene, which may make the talk more meaningful.

2.2 Research Goal

In this study, we aim to propose a prototype of remote pair sightseeing system. It is constructed for two users in separated places: a remote user and a local user. The remote user walks around in the physical environment which would be shared, while the local user would like to have a virtual sightseeing of such shared world. The local user may have expertise related to the environment to help the remote user, or just need the surrounding to be part of the communication. For example, a tourist guide (local user) can offer a private guide for an outdoor visitor (remote user). Or, an elderly person who has mobility problem (local user) may ask someone (remote user) to help buy something in the market. We aim to realize the gestural interaction between the two users during the sightseeing. It simulates the situation that the two users walk side by side in the same physical world chatting with hand gestures.Although the two users might both stay indoors or outdoors, we assume that the local user remains indoors and the remote user goes outside in this research.

2.3 Research Approach

Our system's setup consists of two parts: the wearable device for the local user and the portable setup for the remote user (Figure 2.1). Different from the traditional telepresence system, with the use of spherical camera and head-mounted display (HMD), we allow the local user to access the remote world with a 360°panoramic free viewpoint. The hand gestures of the remote user are provided directly in the capture of the remote scenery for the local user.

For the remote user, we introduce the augmented reality technique. By using a pair of smart glasses, our system presents the 3D air gestures of the local user directly on top of the physical world, which gives an immersive feeling.

Our system uses a depth-based approach to tracking the hands and fingers of the local user. We use a heuristic recognition design requiring no training or calibration and provides a high accuracy. We develop two functions for gestural interaction: (1) Gestural Navigation function, with which the local user uses air gestures to show the spatial direction information which may guide the way for the remote user. (2) Pointing Assistance function helps the local user point out the specific objects directly in the shared scenery.

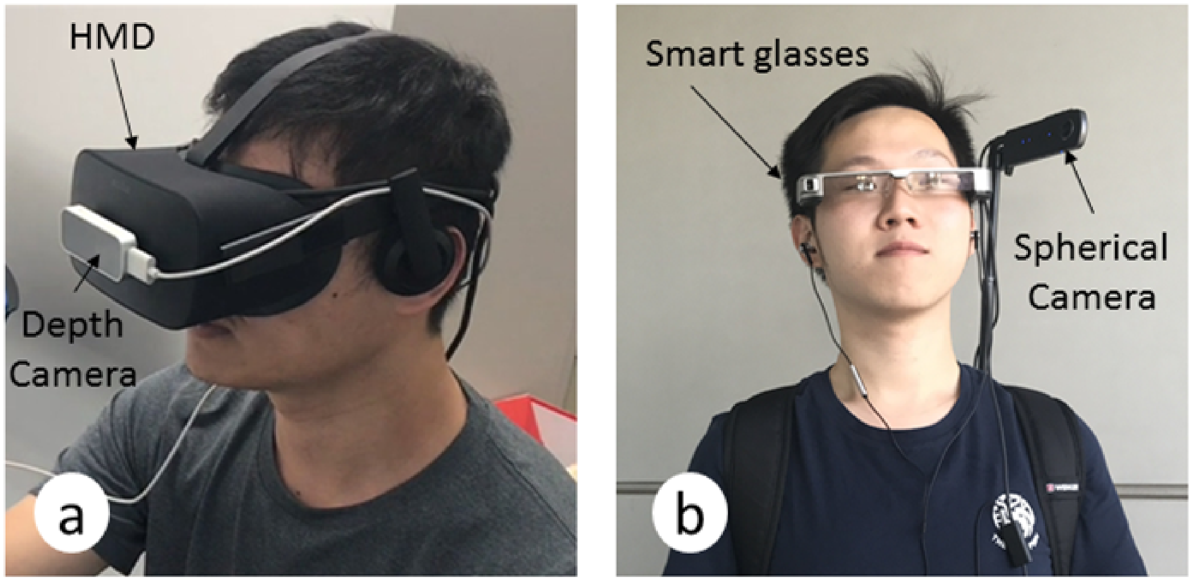


Figure 2.1 The wearable device of the local user (a) is a head-mounted display with a depth camera attached on the front side. The portable setup of the remote user includes a pair of smart glasses and a spherical camera.

Our system has several merits. Firstly, the local user can perform air gestural interaction with the remote user in the same remote physical environment. Secondly, we provide a 360°panoramic capture of the remote real world for sightseeing. With this, the local user could view in whole 360°remote environment freely with no missed information and see the hand gestures performed by the remote user easily, just like truly being there. Thirdly, we support both users having separate independent free viewpoint for sightseeing while each user still could easily tell a joint attention.

Chapter 3

System Design

In this chapter, we introduce Trip-together Feeling and explain our designs to realize such sensation.

3.1 Trip-together Feeling

Trip-together Feeling is a sensation shared by two geographically separated people that they are tripping together in the same place. Although numbers of aspects might be needed to fully realize such sensation, our research focuses on enhancing the human-to-human interaction in the mobile communication by supporting 3D air gestural communication.

We first define three basic elements, which are necessary for users to achieve Trip-together Feeling:

1. Each user could have an independent free viewpoint of the shared environment.

2. Users could see the hand postures of each other directly.

3. Users could join in the same field of vision and complete a smooth gestural communication relevant to the shared environment.

(a) A local user

(b) A remote user

(c) Trip together

Figure 3.1 A local user (a) remains indoor having an immersive virtual sightseeing with a remote partner (b) who goes outdoor with a portable setup. (c) shows the users feel like they are have a trip together .

3.2 The 360°Panoramic Browsing

Our work is a pair sightseeing system that allows the local user to view the remote scenery where the remote user is.

In standard video communication like videophone call, the camera providing a remote view for the local user is carried and controlled by the remote user. In this case, the local user could not choose their own viewpoint conveniently without help from the remote one, just browsing the video more like a bystander. A certain number of different attempts have been researched to solve this restriction [14, 21, 17, 15, 9]. In this work, by using a dual-fish eye spherical camera, we provide a 360°panoramic browsing of surrounding so that the local user could feel personally on the scene. Unlike the normal camera providing a limited angle of capture, our spherical camera could catch the whole 360°panoramic view in both vertical and horizontal simultaneously with no missed information.

Head-mounted display (HMD) usually means stereoscopic 3D displays mounted on user's heads to provide an immersive virtual reality experience. The local user wears an HMD to see in the virtual remote scenery. The viewpoint is controlled by the rotation of HMD which manipulated by the local user's head movement. The local user could freely and naturally control the viewpoint by simply turning the head, just like one truly viewing in the real world, feeling personally on the scene(Figure 3.2).

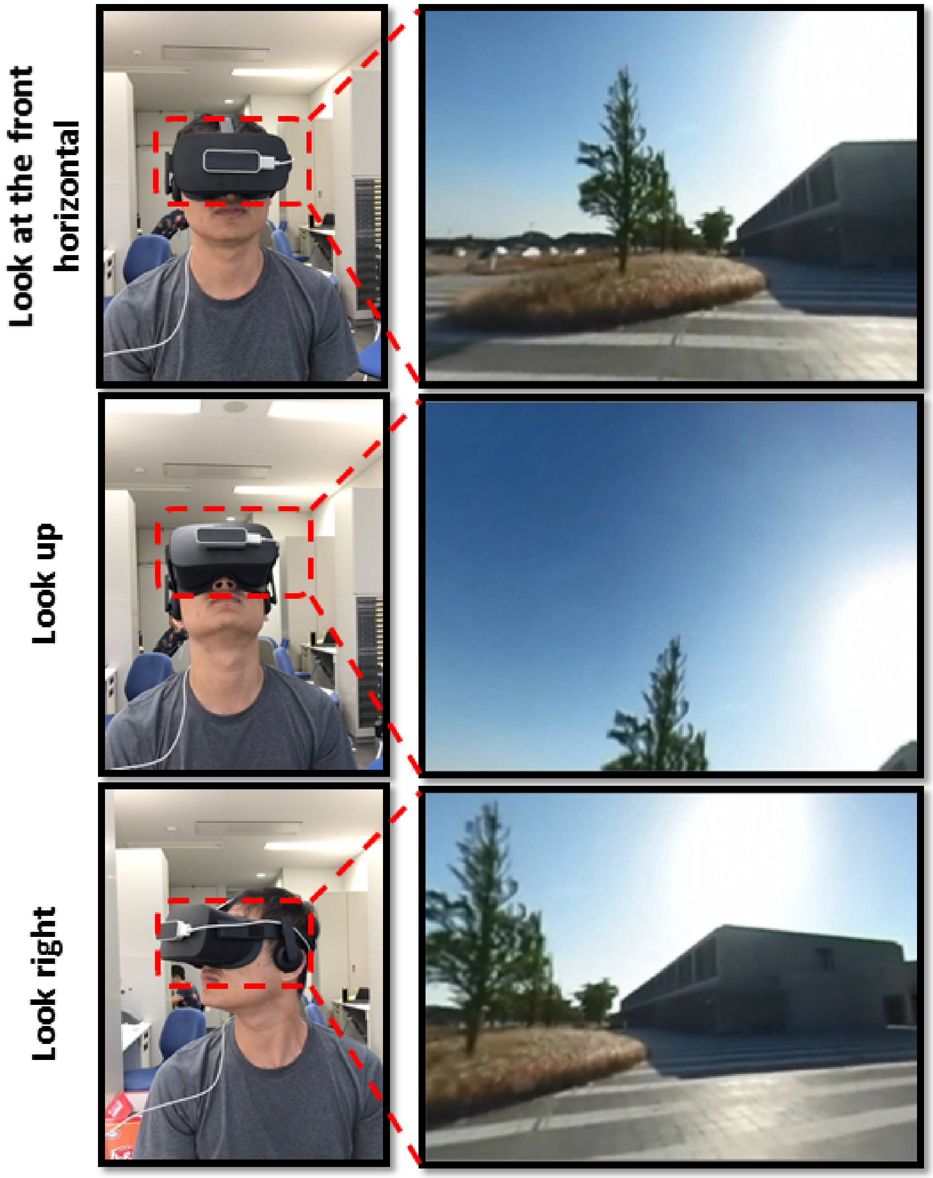


Figure 3.2 When the local user looks around, his/her viewpoint turns upward accordingly. The viewpoint is naturally controlled by the head movement just like being personally on the scene.

This releases the constraint that the local user's viewpoint is restricted by the shooting direction of the camera. The local user has an independent free viewpoint without being influenced or restricted when the remote user seeing around. Consequently, the local and remote users could have separate free viewpoints during the sightseeing.

In addition, such panoramic capture includes the view of remote user's hands. The local user could directly see the hand gestures of the remote user in the remote scenery(Figure 3.3).

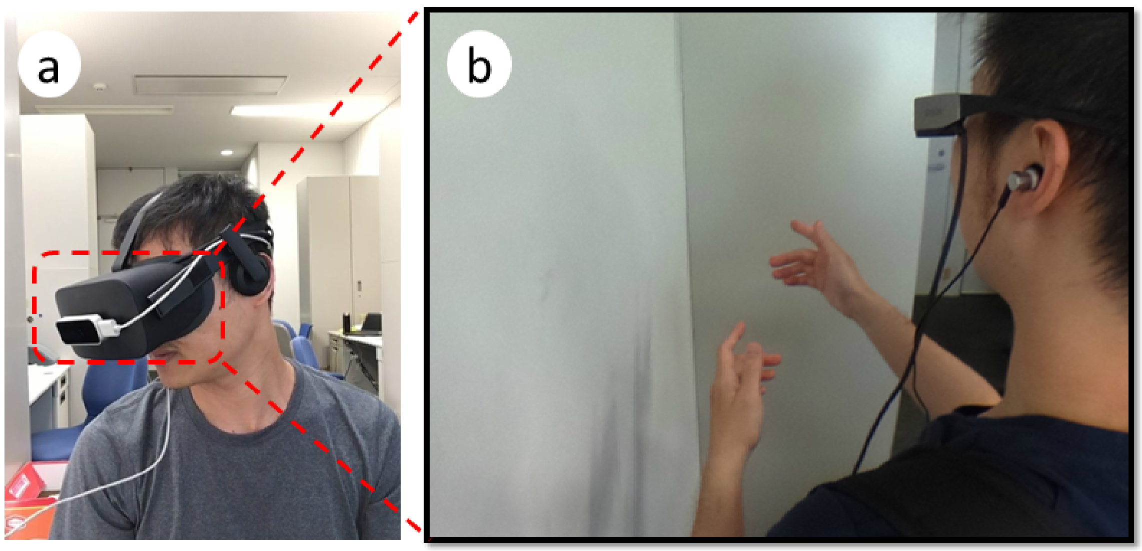


Figure 3.3 In (a), the local user turns his head and sees the remote user is making a hand gesture as shown in (b).

3.3 Attention Reminder

The attention reminder is used to remind both local and remote user a joint attention moment, which means they are viewing in the same direction.

This makes the users easy to know partner's situation while they are viewing independently. It provides both users a common feeling to enhance an experience of tripping with each other. Additionally, by knowing the joint attention moment, the user could keep in the same viewpoint and talk about something in his/her sight or to start a gestural interaction conveniently and achieve a smooth communication.

The system extracts the viewpoint data from local user's HMD and the remote user's smart glasses. By calculating the included angle between the two users' viewpoint in the remote environment, our system gives a signal to remind both users when they are looking at same direction (Figure 3.4 and Figure 3.5). The system notifies the users by showing a "SAME VIEW" signal in the center of both users' the GUI.

Figure 3.4 When the two users are viewing in the same direction, a joint attention signal would be shown in the center of the local user's view.

Figure 3.5 The visualization of the remote user's field of vision when a joint attention signal notifying the joint attention moment.

3.4 Air Gestural Input

Our system supports an air gestural input. The local user is allowed to perform air gestures as an effective approach to communicating with the remote user.

3.4.1 Tracking

We choose a depth-based approach for the gesture recognition, which allows the local user completed the air gestural input freely without wearing any sensor on hands. A depth camera is attached on the front side of the HMD of the local user to make sure the interactive range covering the user's viewing direction. The depth camera can extract not only the subtle changes of the spatial position and posture but also the rotation and orientation of the user's finger joints.

3.4.2 Human-skin Hand Model

We build a pair of virtual 3D human-skin hand models to realize the gestural input of the local user(Figure 3.6). Each hand model consists of 19 movable components representing to each bone of a hand (14 phalanges of fingers plus 5 metacarpal bones). By match the hand models with the depth data of hands, the system can reappear the hand gestures of the local user in the virtual sightseeing precisely (Figure 3.7). Once the user changes their hand postures or moves their hands, the virtual models change to match the same gestures almost instantaneously.

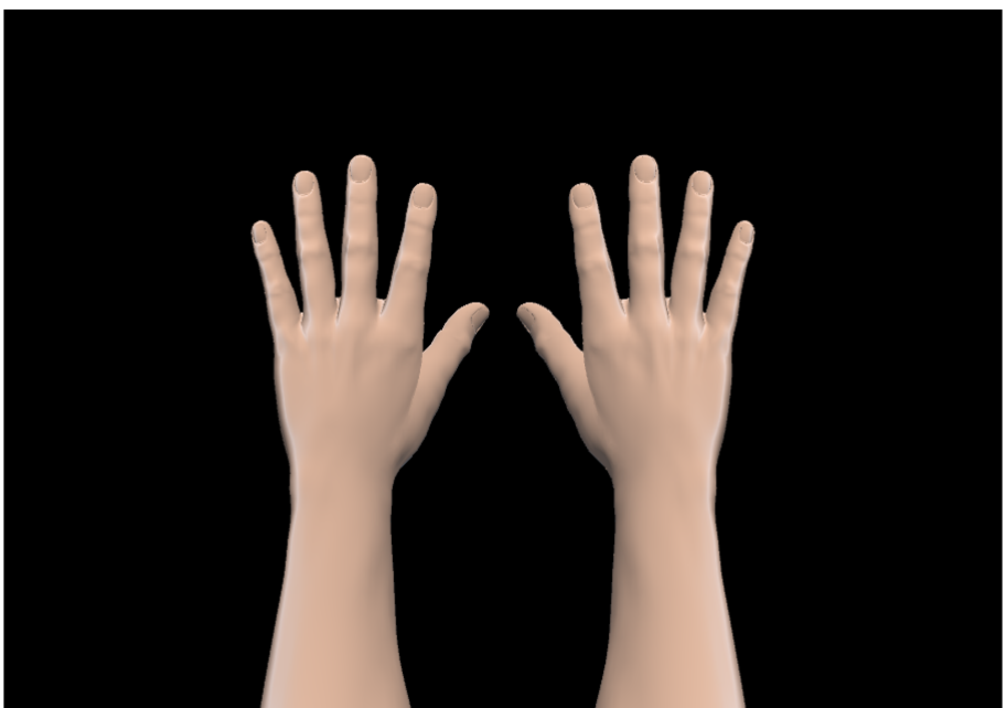


Figure 3.6 We develop a 3D human-skin hand model associated with the bone structure of the user.

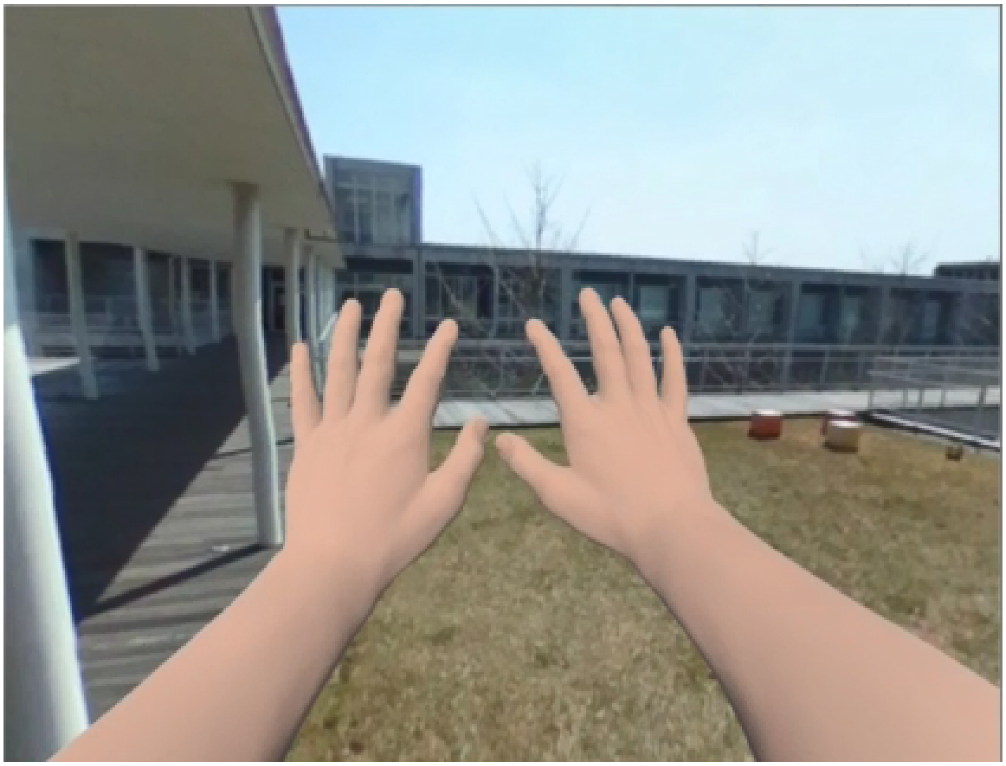


Figure 3.7 The 3D human-skin hand models presents on top of the scenery.

The system presents these human-skin hand models in the local user's facing view with the First-person Perspective (FPP) on top of the remote scenery. The hand models could be activated by simply raising hands in the facing direction. Additionally, the scale of the hand model in the virtual scenery to physical hands is one to one. With the use of the HMD, this design could provide an immersive virtual reality feeling for the local user. Figure 7 shows the example of performing air gestures in the remote scenery.

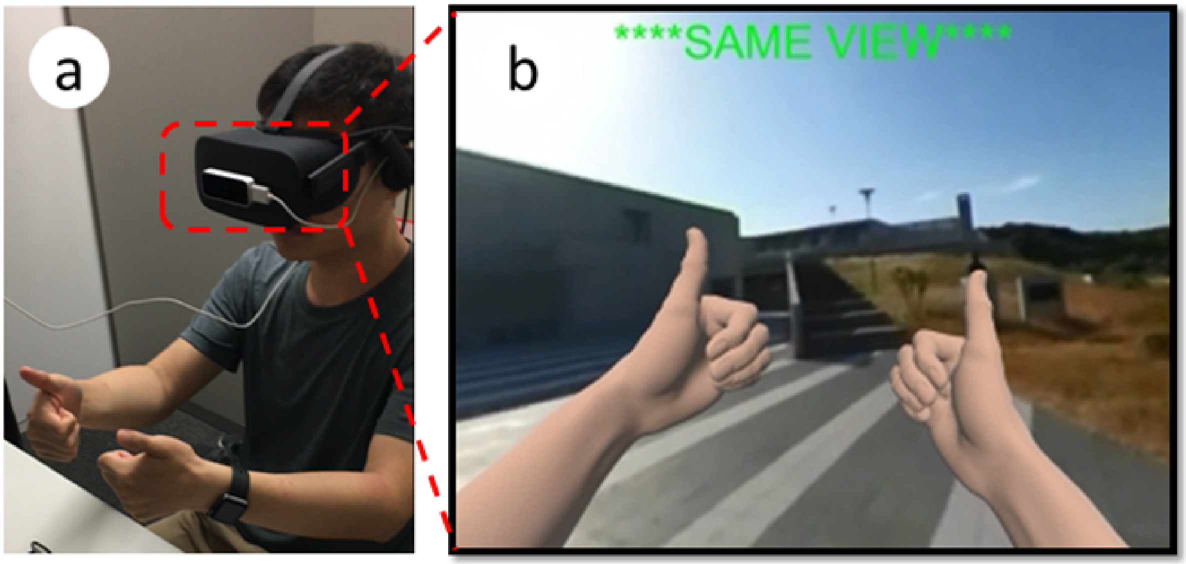


Figure 3.8 The local user is making an air gesture (a). He could make a gestural input in the remote scenery with the First-person Perspective.

These hand models are also sent to the remote user and display on the remote user's smart glasses (Figure 3.9). The hand models are presented on the left side of the field of vision, superimposing in the physical world. Therefore, the remote user could see the gestures of the local user directly while viewing the environment. It is worth to point out that the perspective of the hand models is different with the local user's. For the remote user, it simulates to watch the hand gestures from the side.

Figure 3.9 Visualization example of the remote user's field of vision. The local user's hands present on the left side, superimposing on the physical world.

3.5 Gestural Navigation

Through the air gestural input design, we mentioned above, the local user and remote user could achieve a basic gestural communication. However, since the local user's hand gestures are always presented as long as the depth camera can detect the hands, it is necessary to distinguish the meaningful gestures from those meaningless ones to arouse the remote user's attention. We design a gestural navigation function for the local user to assist the remote user in direction guidance. We develop two groups of navigation gestures: Six Direction Gestures and Warning Gestures. These designed gestures are based on the universal gestures that are common in daily navigation, which makes it easy for users to learn and perform them. When a gesture is detected, a notification signal shows at the lower right corner of both users' GUI.

An important characteristic of our gesture recognition technique is that we calculate the included angle between different finger bones to determine the finger state (Figure 3.10). Previous research has demonstrated that tracking the change of the depth-based bone structure could provide a high accuracy to distinguish different gestures [13, 12]. We calculate the included angle between intermediate bone and proximal bone and the included angle between proximal bone and metacarpal bone after extracting the 3D bone structure. When both angles are smaller than the set thresholds (12°), the finger is fully extended.

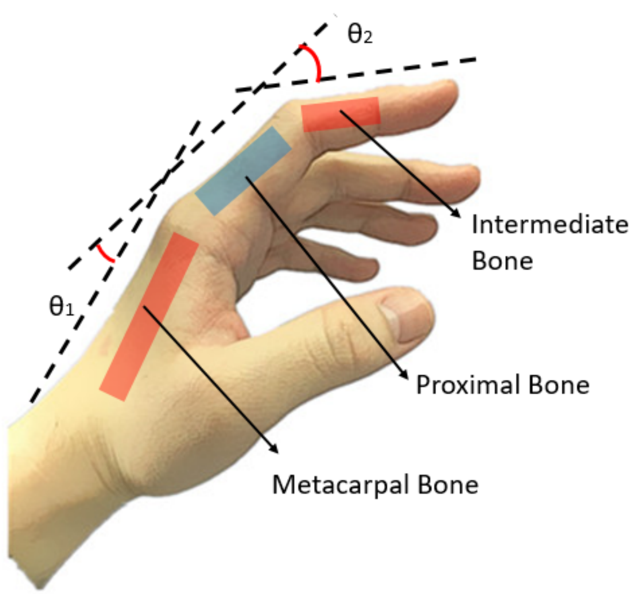
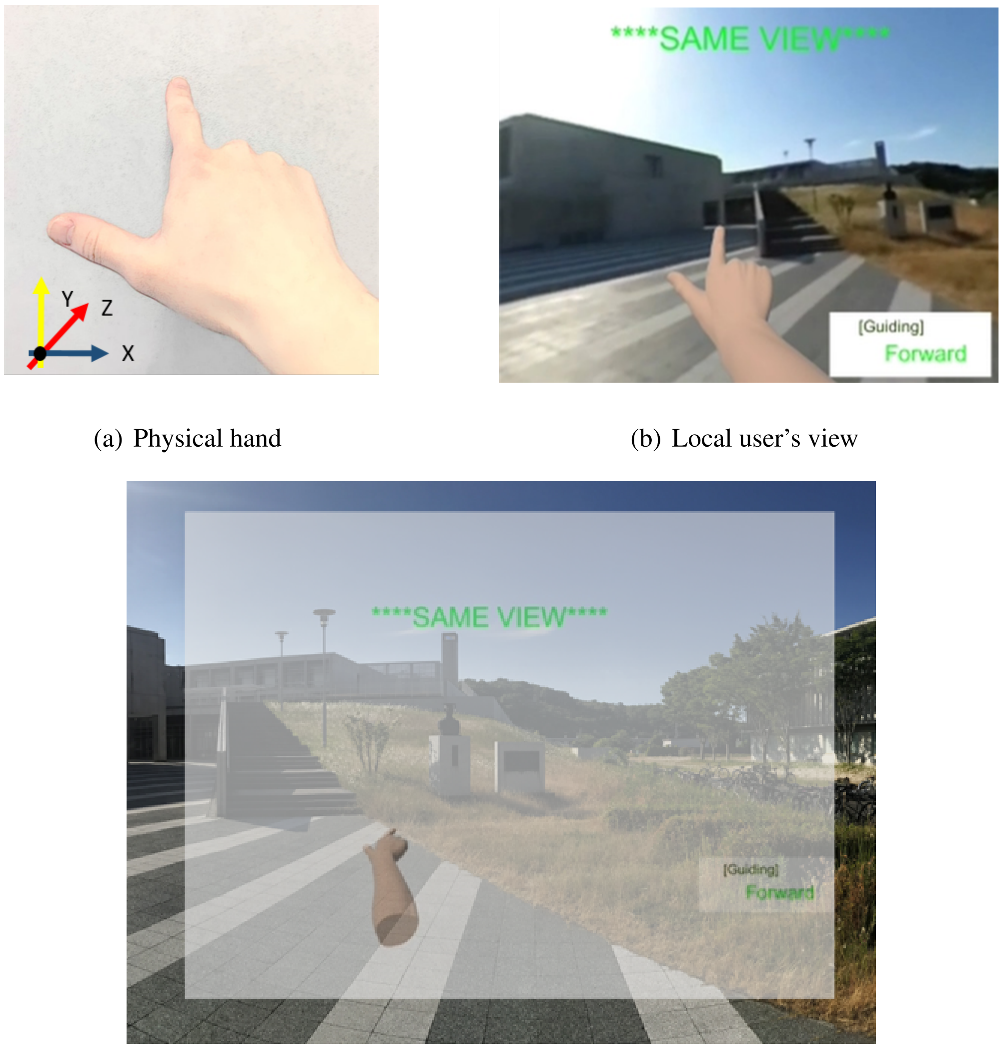


Figure 3.10 Calculating the include angle between bones to determine the finger states.

3.5.1 Six Direction Gestures

Six Direction Gestures are used to help the local user showing the spatial direction.When the system detects index finger and thumb are extended while other fingers are not extended, a “guiding trigger" is activated.The local user could map the index finger’s pointing orientation in the physical world to the spatial direction in the virtual scenery. The system recognizes six direction gestures: “forward”, “back”, “leftward”, “rightward”, “up” and “down”. Finally, a guiding signal presents in the graphical user interface (GUI). Figure 3.11 to Figure 3.16 show the six of the gestures.



(a) Physical hand

(b) Local user's view

(c) Remote user's view

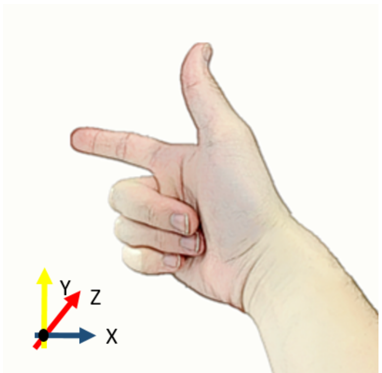
Figure 3.11 Subgraph (a) shows the physical hand of the local user performing a "Forward Direction" gesture. Subgraph (b) shows the gesture in the local user's view. Subgraph (c) is the visualization of the remote user's field of vision.

(a) Physical hand

(b) Local user's view

(c) Remote user's view

Figure 3.12 Subgraph (a) shows the physical hand of the local user performing a "Back" direction gesture. Subgraph (b) shows the gesture in the local user's view. Subgraph (c) is the visualization of the remote user's field of vision.

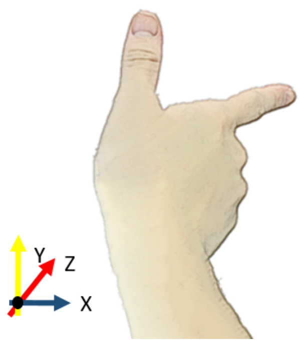


(a) Physical hand

(b) Local user's view

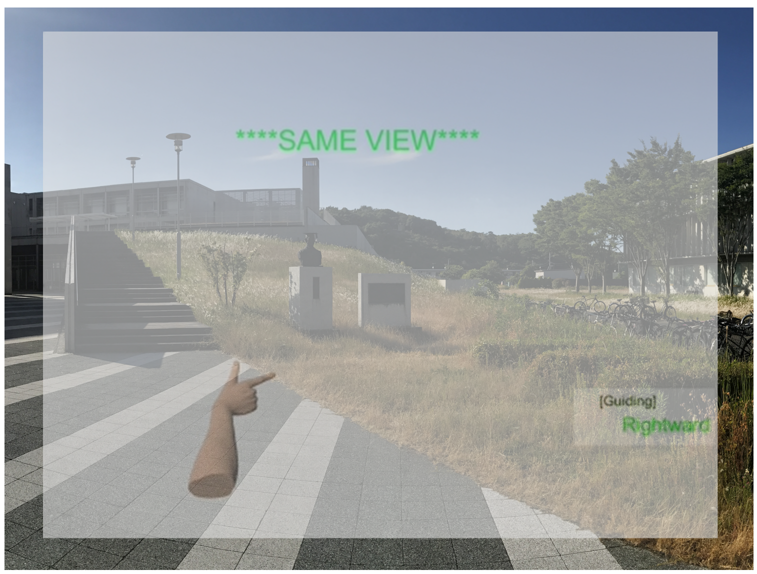
(c) Remote user's view

Figure 3.13 Subgraph (a) shows the physical hand of the local user performing a "Leftward" direction gesture. Subgraph (b) shows the gesture in the local user's view. Subgraph (c) is the visualization of the remote user's field of vision.



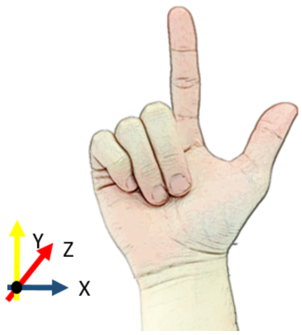
(a) Physical hand

(b) Local user's view



(c) Remote user's view

Figure 3.14 Subgraph (a) shows the physical hand of the local user performing a "Rightward" direction gesture. Subgraph (b) shows the gesture in the local user's view. Subgraph (c) is the visualization of the remote user's field of vision.

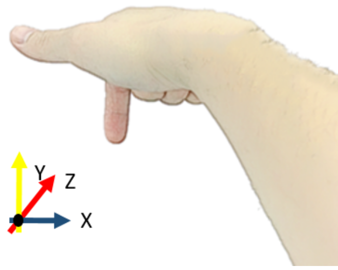


(a) Physical hand

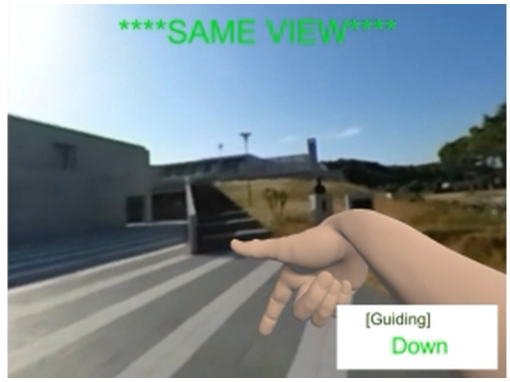
(b) Local user's view

(c) Remote user's view

Figure 3.15 Subgraph (a) shows the physical hand of the local user performing a "Up" direction gesture. Subgraph (b) shows the gesture in the local user's view. Subgraph (c) is the visualization of the remote user's field of vision.



(a) Physical hand



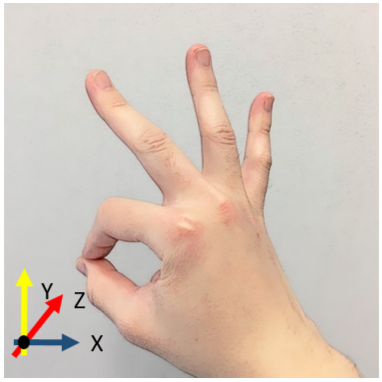
(b) Local user's view

(c) Remote user's view

Figure 3.16 Subgraph (a) shows the physical hand of the local user performing a "Down" direction gesture. Subgraph (b) shows the gesture in the local user's view. Subgraph (c) is the visualization of the remote user's field of vision.

3.5.2 Warning Gestures

Warning Gestures include "OK" Gesture and "Wait" Gesture (Figure 3.17 and Figure 3.18). They are used to help the local user warn the remote user to pause or continue during navigation. When a warning gesture is detected, a warning signal presents to notify the remote user.

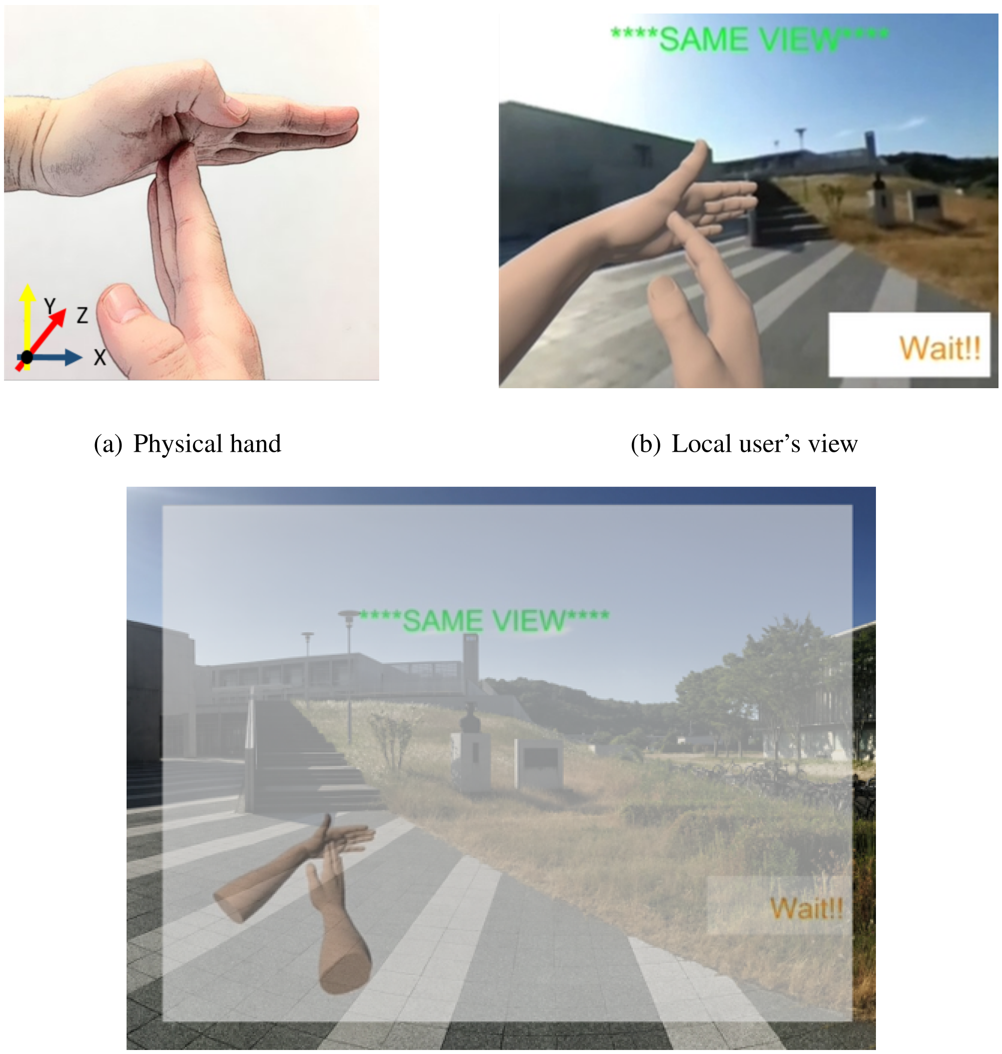


(a) Physical hand

(b) Local user's view

(c) Remote user's view

Figure 3.17 Subgraph (a) shows the "OK" Gesture in the physical world. Subgraph (b) shows the local user's view. Subgraph (c) is the visualization of the remote user's field of vision.



(a) Physical hand

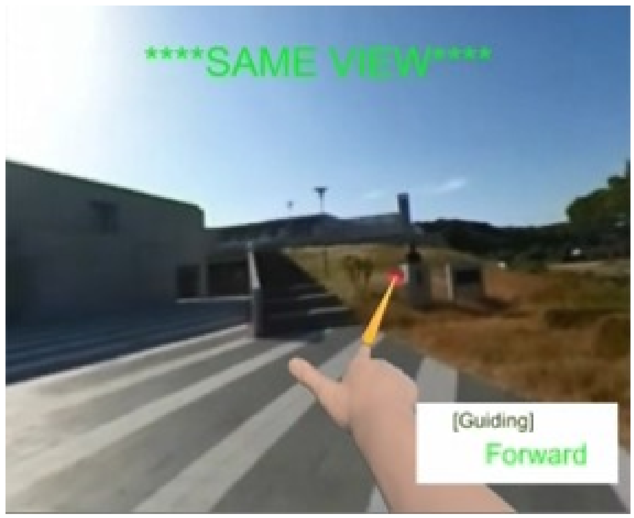
(b) Local user's view

(c) Remote user's view

Figure 3.18 Subgraph (a) shows the "Wait" Gesture in the physical world. Subgraph (b) shows the local user's view. Subgraph (c) is the visualization of the remote user's field of vision.

3.6 Pointing Assistance

The pointing assistance function helps the local user point out specific objects in the field of vision. We develop a tool called “the pointing arrow" to show the precise direction which the user is pointing at. It consists of a yellow stick to highlight the pointing direction and a red cone on the tip to indicate the target object. The “pointing arrow" begins from the tip of the hand model’s index finger and points at the direction of the intermediate bone of index finger (Figure 3.19). Based on the joint attention, the local user could easily show some interesting points in the remote scenery directly to the remote user and create potential conversation topics.



(a) Local user’s view

(b) Remote user’s view

Figure 3.19 Subgraph (a) shows the local user is pointing at a statue in the scene. Subgraph (b) is the visualization of the remote user’s field of vision.

Chapter 4

Implementation

4.1 System Hardware Overview

Our system's hardwares include two parts: the local user side and the remote user side. Figure 4.1 shows the system hardware and information overview.

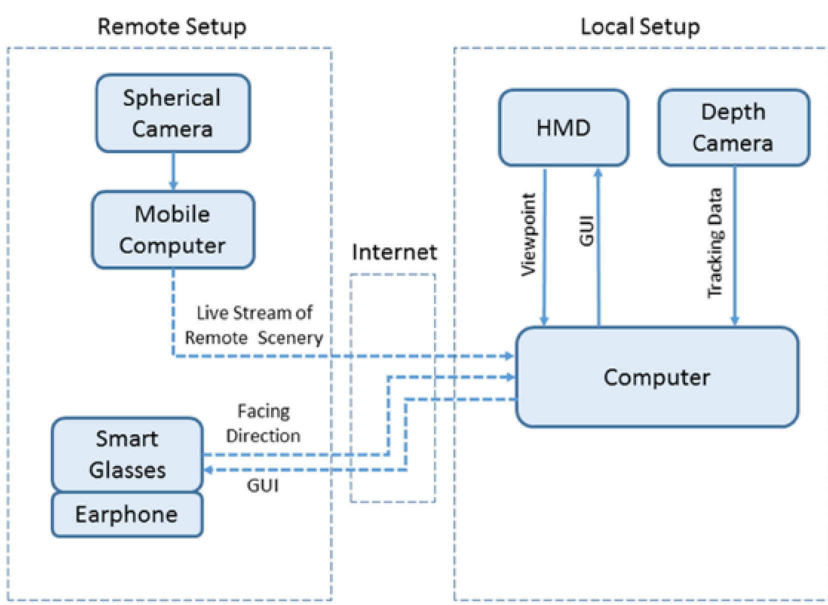


Figure 4.1 Hardware Overview

4.2 Local Setup

Figure 4.2 shows the overview of the local setup. It includes the wearable devices and a desktop PC. The local user sits at the table to use our system.

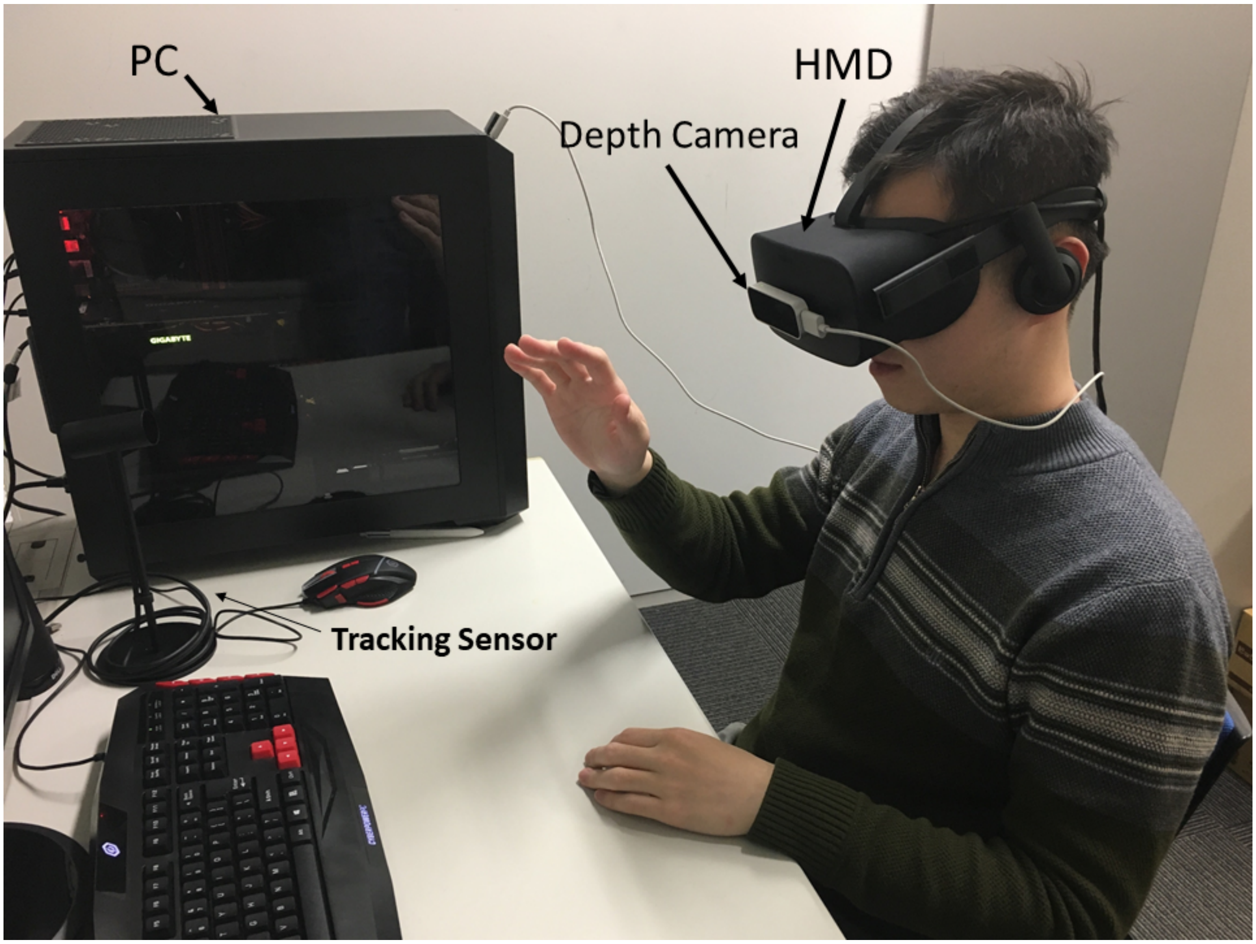


Figure 4.2 Local setup overview

The desktop PC placed on the local user side is used to analyze data and engine the core system. Unity 3D is used to render and process the incoming data from both remote and local side as well as to generate GUI for both users.Table 4.1 shows the information of the PC we used.

Table 4.1 Local Desktop PC

Desktop PC	
Operating System	Microsoft Windows 10
Graphics Card	AMD Radeon RX480
CPU	Intel Core i5-6402P 2.80GHz 2.80GHz
RAM	8.00 GB
Internet Connection	Wi-Fi 802.11n

4.2.1 Head-mounted Display

HMD is commonly used for video gaming but until recently, consumer adoption of HMD technology has not taken off due to high costs, bulky setups, as well as poor framerate and tracking performance [5, 23, 27].

In this research, the local user uses an new generation head-mounted display - Oculus Rift cv1 headset. It uses a pair of low persistence OLED screens, one for each eye, providing a 110° field of view [20]. Its combination of the high refresh rate, global refresh and low persistence offer the user experiences none of the motion blurring or judder that is experienced on a regular monitor. It supports a full 6 degree of freedom rotational and positional tracking of the head movement which is precise, low-latency, and sub-millimeter accurate. This tracking is performed by a point tracking sensor placed on the desk. The headset also gets an integrated headphones, which provide real-time 3D audio effects [19].

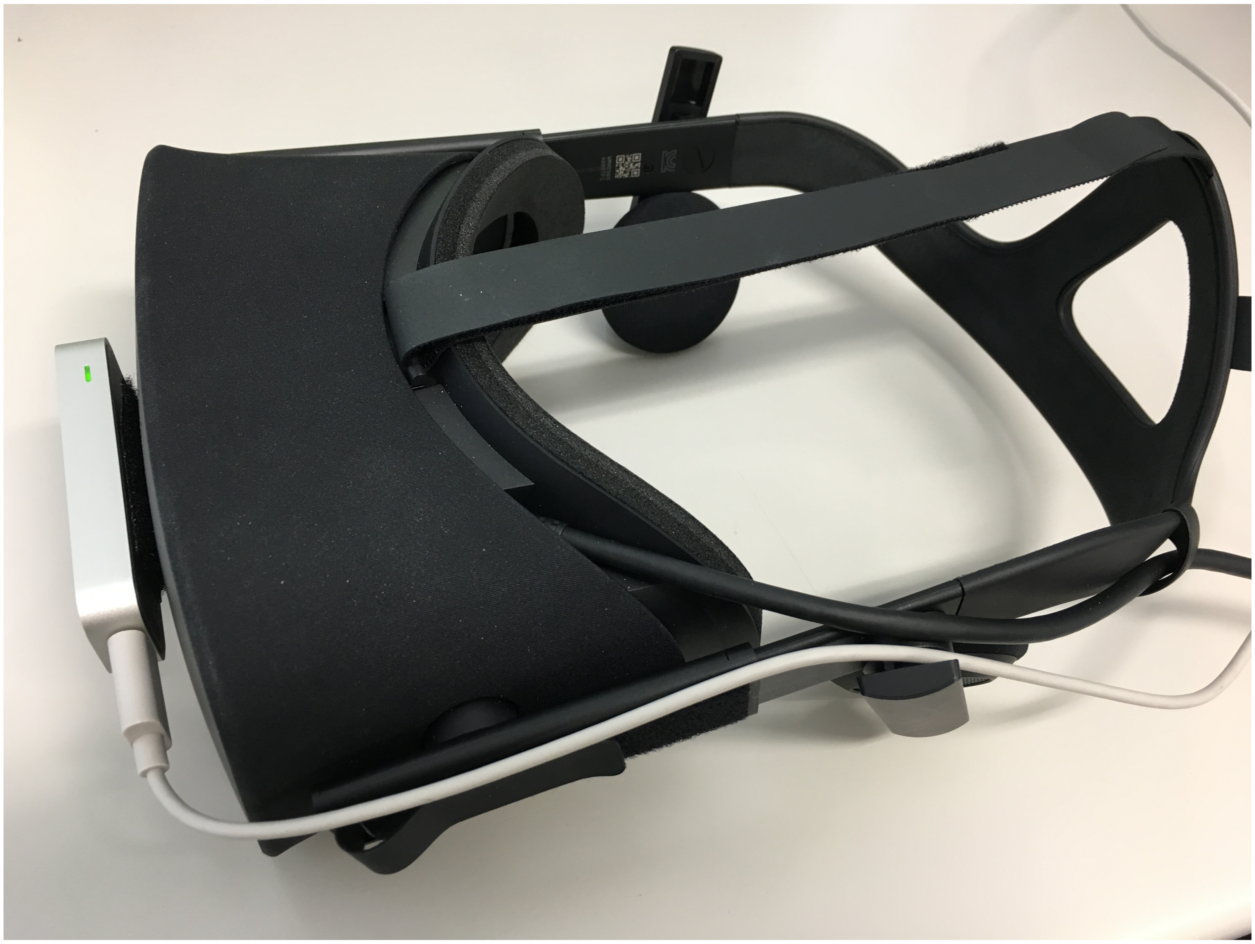


Figure 4.3 Head-mounted display

4.2.2 Depth Camera

To realize the gestural recognition, we choose a new generation depth camera – Leap Motion which introduces a new gesture and position tracking system with sub-millimeter accuracy (an about 0.7 millimeters overall average accuracy with 8 cubic feet interactive range [29]).This above-surface sensor is discussed for use in realistic stereo 3D interaction systems. Consisting of three IR (Infrared Light) emitters and two IR cameras, the Leap Motion can be categorized into optical tracking systems based on Stereo Vision.

The Leap Motion is light enough (only about 45g) to make sure it is comfortable for users to wear. The effective range of the Leap Motion extends from approximately 3 to 60 centimeters above the device like an inverted pyramid. We attach it to the front side of the HMD (shown in Figure 4.4(b)).

(a) Leap Motion (b) Attached to the HMD

Figure 4.4 Depth camera

4.3 Remote Setup

Figure 4.5 shows the overview of the remote setup. It is a integrated wearable device consisting of a augmented reality smart glasses, a spherical camera and a mobile computer.

Figure 4.5 Remote setup overview

4.3.1 Smart Glasses

The remote user wears an augmented reality smart glasses-EPSON Moverio BT-300 which is light and compact enough (only 69 g) but supports an HD binocular displays. It packs with a motion-tracking sensor to detect the user's facing direction and a wireless module to exchange information with the local side via the Internet(Figure 4.7). It presents a semitransparent display on top of the physical world while allows the user to view the physical world clearly (Figure 4.8). It provides an audio output with an earphone.

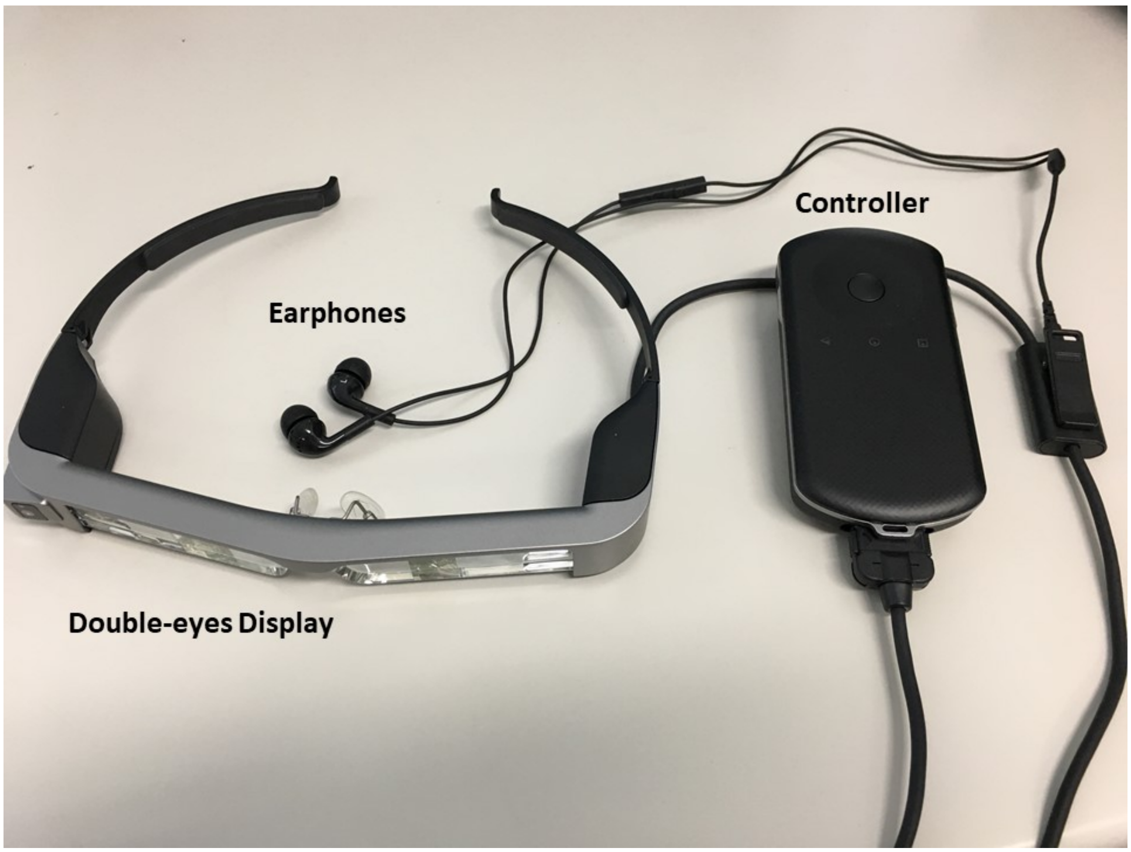


Figure 4.6 Augmented reality smart glasses

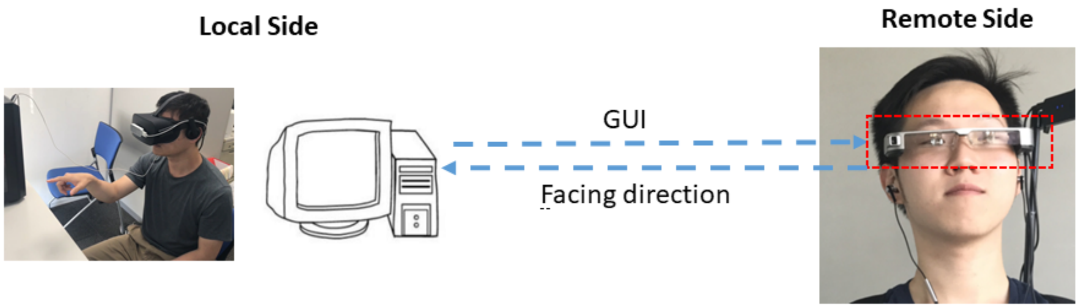


Figure 4.7 Exchange information with the local side via the Internet

Figure 4.8 GUI superimposes on the physical world

4.3.2 Spherical Camera

Early 360-degree imaging systems used a catadioptric camera [24], which combines lens (dioptric) and mirror (catoptric), to record 360-degree contents. However, due to the inherent lens+mirror arrangement, the captured field of view is typically limited to less than 360x180 degrees, and some of the catadioptric systems are not compact [10]. On the downside, these highend 360-degree cameras are bulky and extremely expensive, even with the decreasing cost of image sensors, and are out of reach for most of the regular users. To bring the immersive experience, in this research, we choose a 360°spherical camera - RICOH THETA S (Figure 4.9). This compact camera uses only two fisheye lenses whose field of view is close to 195 degrees each. The images generated by the two fisheye lenses have very limited overlapping field of views but can, however, be stitched together to produce a full spherical 360x180 panorama.

(a) front (b) side view (c) back

Figure 4.9 Spherical camera

The spherical camera is set on the top of a metal rod carried by the remote user (Figure 4.10). We choose this place so that the local user could see the hand gestures of the remote user (see Figure 3.3).

Figure 4.10 The spherical camera is set on the top of a metal rob

4.4 Live Panorama of the Remote world

Figure 4.11 shows the live panoramic video stream from the spherical camera to the local user's HMD. The spherical camera is connected to a mobile computer over USB (1280x720 15fps) to generate a live stream to send the live video data to the desktop PC on the local user side with Real Time Messaging Protocol (RTMP). The streaming uses H.264 software encoder at a 2500 video bit rate and 64 audio bit rate (Figure 4.12 shows the live stream setting). Table 4.2 shows the information of the mobile computer we used.

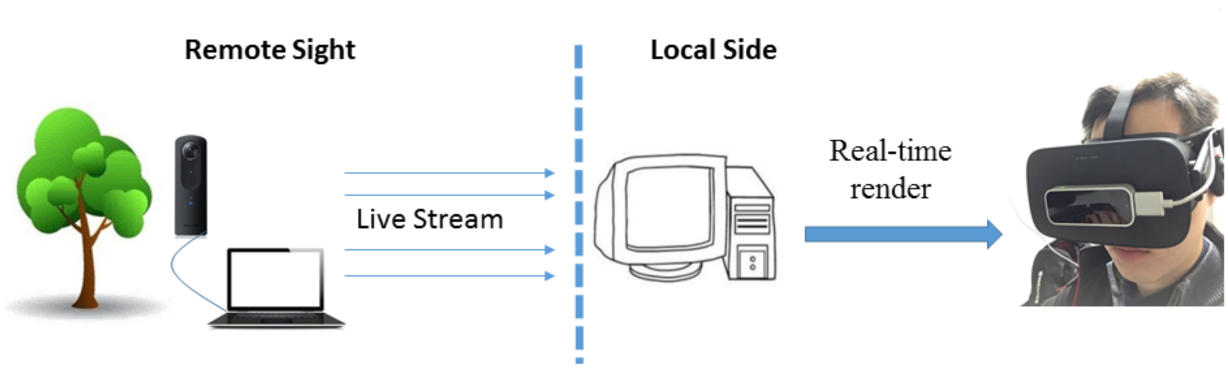


Figure 4.11 Live panoramic video stream

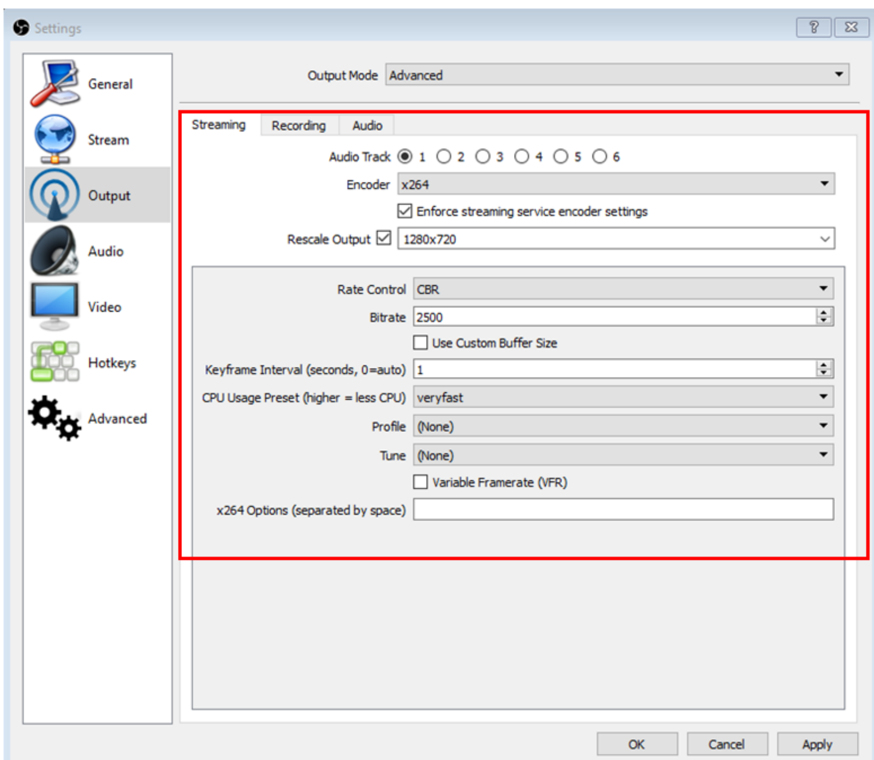


Figure 4.12 Video stream setting

Table 4.2 Mobile computer

Mobile computer	
Operating System	Microsoft Windows 10
Graphics Card	Intel(R)HD Graphics 520
CPU	Intel Core i7-6500 2.50GHz 2.59GHz
RAM	8.00 GB
Internet Connection	Wi-Fi 802.11n
Battery	23.2Wh

As we mentioned aboved, each lens of the spherical camera gets about 195-degree field of view, so the raw video data is dual-fish eyes video (Figure 4.13).

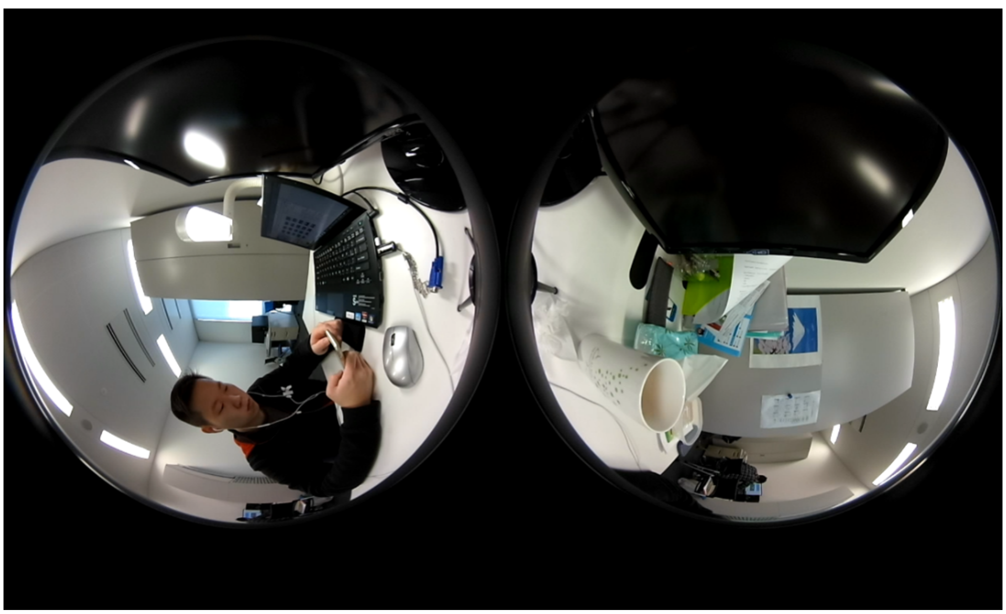


Figure 4.13 Dual-fish eyes video

A real-time switching should be completed on the desktop computer to provide the browsing with free viewpoint. The switching method is operated with Unity 3D. First, we build a transparent spherical model as the render texture(Figure 4.14). The model is constructed by 2 overlapping half spheres with the surface set to face inward. Then each side of dual-fish video is drawn on each half sphere. After adjustment we get a whole 360°spherical view(Figure 4.15).

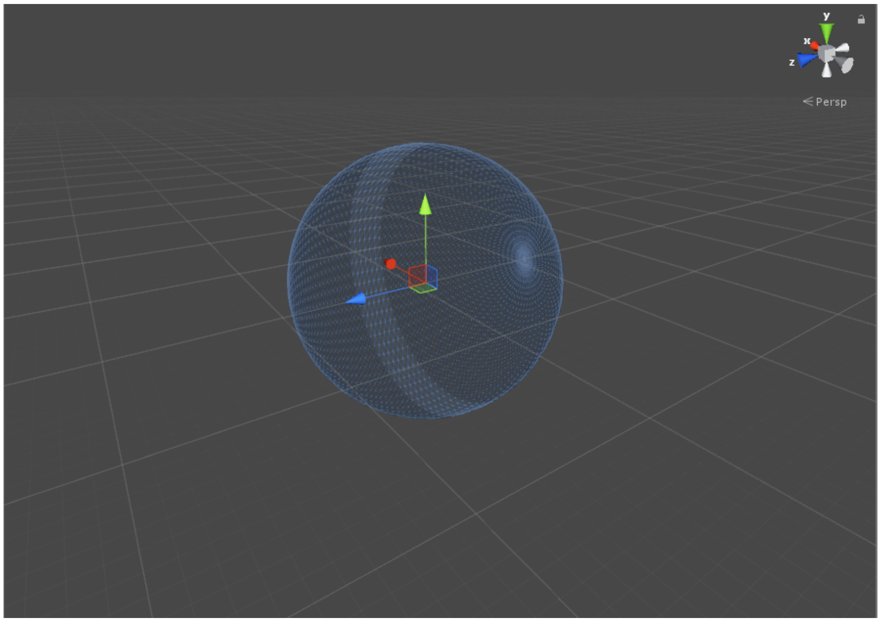


Figure 4.14 Transparent spherical model

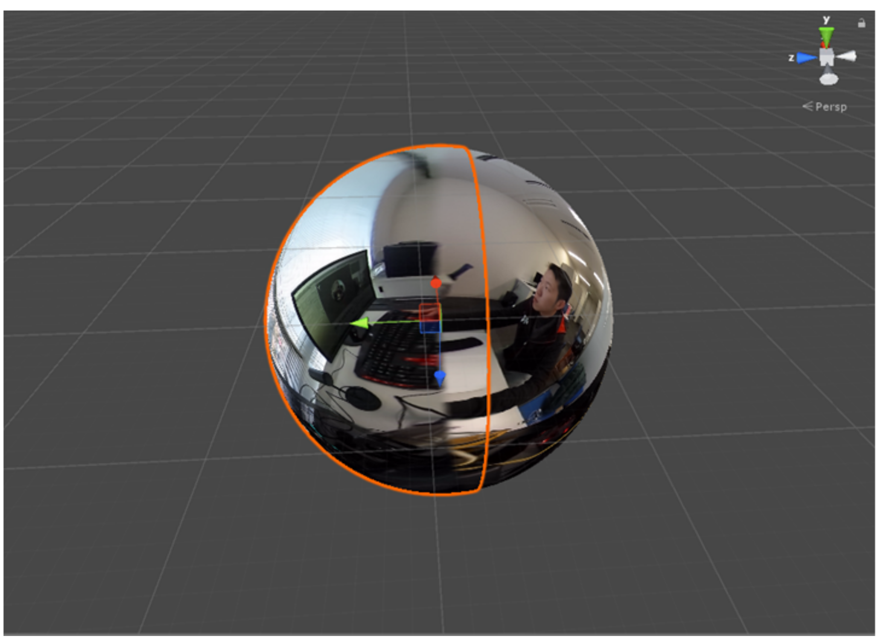
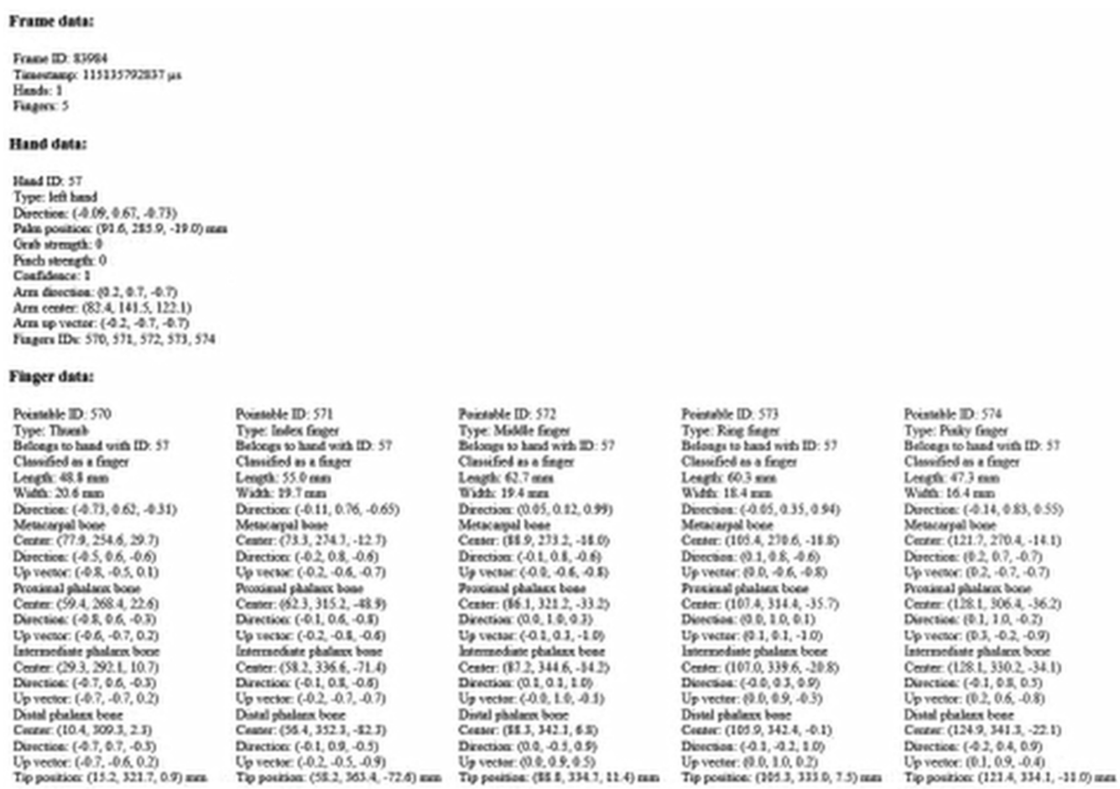


Figure 4.15 Spherical panoramic view

4.5 Depth-based Recognition

In this research,to drive the depth camera, we use Leap Motion SDK 3.2.0 and Leap Motion's Core Assets 4.2.0. We extract the depth data of fingers and palm from the Leap motion. Each finger includes 4 bones: Metacarpal bone, Proximal phalanx bone, Intermediate phalanx bone and Distal phalanx bone. Figure 4.16 shows one frame of the depth data extracted from the depth camera. Figure 4.17 is a reconstruction model used the depth data of the user's hand, without any texture.



Frame data:

Frame ID: 83984
Timestamp: 115135792837 µs
Hands: 1
Fingers: 5

Hand data:

Hand ID: 57
Type: left hand
Direction: (-0.09, 0.67, -0.73)
Palm position: (91.6, 285.9, -19.0) mm
Grab strength: 0
Pinch strength: 0
Confidence: 1
Arm direction: (0.2, 0.7, -0.7)
Arm center: (82.4, 141.5, 122.1)
Arm up vector: (-0.2, -0.7, -0.7)
Fingers IDs: 570, 571, 572, 573, 574

Finger data:

Pointable ID: 570	Pointable ID: 571	Pointable ID: 572	Pointable ID: 573	Pointable ID: 574
Type: Thumb	Type: Index finger	Type: Middle finger	Type: Ring finger	Type: Pinky finger
Belongs to hand with ID: 57	Belongs to hand with ID: 57	Belongs to hand with ID: 57	Belongs to hand with ID: 57	Belongs to hand with ID: 57
Classified as a finger	Classified as a finger	Classified as a finger	Classified as a finger	Classified as a finger
Length: 48.8 mm	Length: 55.0 mm	Length: 62.7 mm	Length: 60.3 mm	Length: 47.3 mm
Width: 20.6 mm	Width: 19.7 mm	Width: 19.4 mm	Width: 18.4 mm	Width: 16.4 mm
Direction: (-0.73, 0.62, -0.31)	Direction: (-0.11, 0.76, -0.65)	Direction: (0.05, 0.12, 0.99)	Direction: (-0.05, 0.35, 0.94)	Direction: (-0.14, 0.83, 0.55)
Metacarpal bone	Metacarpal bone	Metacarpal bone	Metacarpal bone	Metacarpal bone
Center: (77.9, 254.6, 29.7)	Center: (73.3, 274.3, -12.3)	Center: (85.9, 273.2, -18.0)	Center: (105.4, 270.6, -18.8)	Center: (121.7, 270.4, -14.1)
Direction: (-0.5, 0.6, -0.6)	Direction: (-0.2, 0.8, -0.6)	Direction: (-0.1, 0.8, -0.6)	Direction: (0.1, 0.8, -0.6)	Direction: (0.2, 0.7, -0.7)
Up vector: (-0.8, -0.5, 0.1)	Up vector: (-0.2, -0.6, -0.7)	Up vectar: (-0.0, -0.6, -0.8)	Up vecter: (0.0, -0.6, -0.8)	Up vector: (0.2, -0.7, -0.7)
Proximal phalanx bone	Proximal phalanx bone	Proximal phalanx bone	Proximal phalanx bone	Proximal phalanx bone
Center: (59.4, 268.4, 22.6)	Center: (62.3, 315.2, -48.9)	Center: (86.1, 321.2, -33.2)	Center: (107.4, 314.4, -35.7)	Center: (128.1, 306.4, -36.2)
Direction: (-0.8, 0.6, -0.3)	Direction: (-0.1, 0.6, -0.8)	Direction: (0.0, 1.0, 0.3)	Direction: (0.0, 1.0, 0.1)	Direction: (0.1, 1.0, -0.2)
Up vector: (-0.6, -0.7, 0.2)	Up vector: (-0.2, -0.8, -0.6)	Up vectar: (-0.1, 0.3, -1.0)	Up vector: (0.1, 0.1, -1.0)	Up vector: (0.3, -0.2, -0.9)
Intermediate phalanx bone	Intermediate phalanx bone	Intermediate phalanx bone	Intermediate phalanx bone	Intermediate phalanx bone
Center: (29.3, 292.1, 10.7)	Center: (58.2, 336.6, -71.4)	Center: (87.2, 344.6, -34.2)	Center: (107.0, 339.6, -20.8)	Center: (128.1, 330.2, -34.1)
Direction: (-0.7, 0.6, -0.3)	Direction: (-0.1, 0.8, -0.6)	Direction: (0.1, 0.1, 1.0)	Direction: (-0.0, 0.3, 0.9)	Direction: (-0.1, 0.8, 0.5)
Up vector: (-0.7, -0.7, 0.2)	Up vector: (-0.2, -0.7, -0.7)	Up vector: (-0.0, 1.0, -0.1)	Up vecter: (0.0, 0.9, -0.3)	Up vector: (0.2, 0.6, -0.8)
Distal phalanx bone	Distal phalanx bone	Distal phalanx bone	Distal phalanx bone	Distal phalanx bone
Center: (10.4, 309.3, 2.3)	Center: (56.4, 352.3, -82.3)	Center: (88.3, 342.1, 6.8)	Center: (105.9, 342.4, -0.1)	Center: (124.9, 341.3, -22.1)
Direction: (-0.7, 0.7, -0.5)	Direction: (-0.1, 0.9, -0.5)	Direction: (0.0, -0.5, 0.9)	Direction: (-0.1, -0.2, 1.0)	Direction: (-0.2, 0.4, 0.9)
Up vector: (-0.7, -0.6, 0.2)	Up vector: (-0.2, -0.5, -0.9)	Up vector: (0.0, 0.9, 0.5)	Up vecter: (0.0, 1.0, 0.2)	Up vector: (0.1, 0.9, -0.4)
Tip position: (15.2, 321.7, 0.9) mm	Tip position: (58.2, 363.4, -72.6) mm	Tip position: (88.8, 334.7, 11.4) mm	Tip position: (105.3, 333.0, 7.5) mm	Tip position: (121.4, 334.1, -11.0) mm

Figure 4.16 Depth data

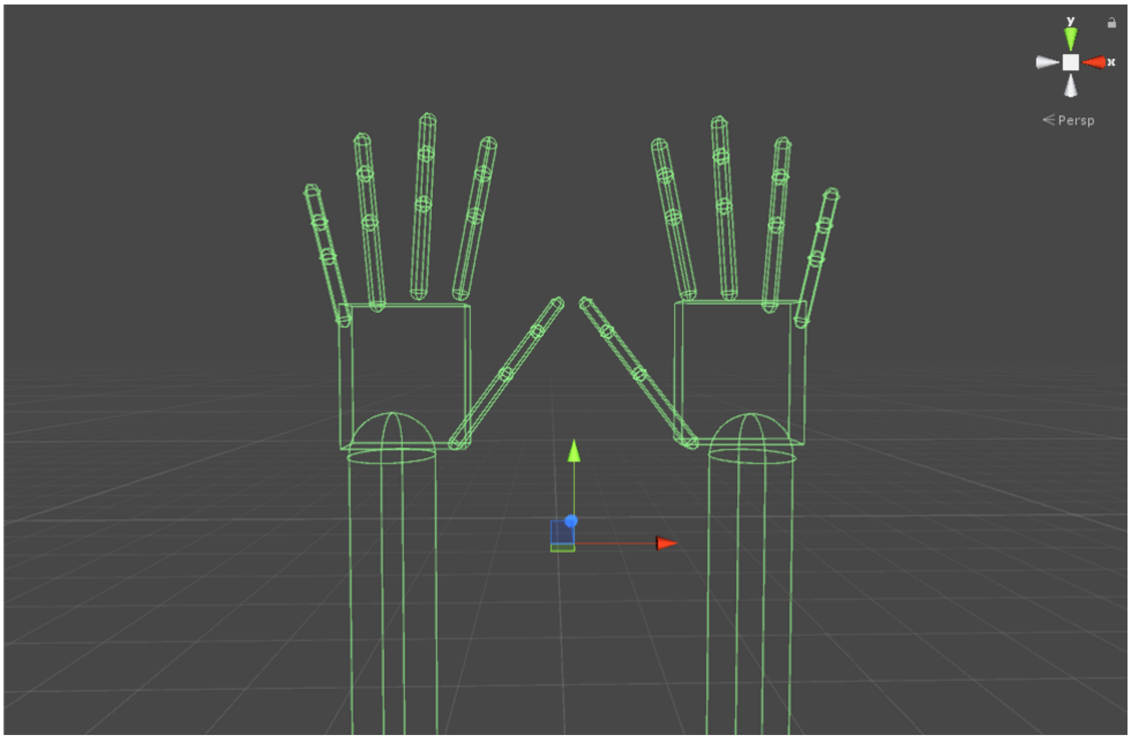


Figure 4.17 Visualization of depth data of the user's hand

Chapter 5

Related Work

Our work is closely related to the previous research called "*WithYou*", a remote communication prototype which aims to help the two users feel they go out together to some extent [4, 2, 3]. *WithYou* defines three elements to get an out together feeling: (1) Enabling both users to freely control the viewing direction onto the outside environment. (2) Users could know the viewing direction of the other one. (3) Gesture communication could support a smooth communication without audio. In this work, the indoor user turns the head to control the rotation of a pan-and-tilt camera carried by the outdoor user so as to get a different viewing direction of the outdoor surrounding. The system shares users' focus direction in horizontal and distinguishes the focus status of users to create a joint attention. Although it mentions the importance of gestural communication, the *WithYou* system just realizes a rough gestural instruction by shaking or tapping the wireless controllers held in the users' hand.

Comparing with *WithYou*, our system has some advantages in following several aspects. First, our system provides an indeed 360°panoramic viewing for the local user while *WithYou* has a blind angle nearly 100°in vertical. Second, we develop a way to allow the real air gestural interaction between the two users. The users could perform gestures naturally without any wearable sensor on hands. What's more, we provide a portable augmented reality setup for the remote user, which allows the remote user to immersive in the gestures communication. The Table 5.1 summarizes the main differences between *WithYou* and our system.

Table 5.1 Comparison between *WithYou* and this system

WithYou	This system
Two pan-and-tilt cameras with a blind angle are used to catch the outdoor view.	Spherical camera providing truly 360° panoramic capture of a remote world.
Wireless controller for the outdoor user to make an instruction.	Panoramic capture provides a direct view of the remote user's hand,gestures.
Indoor user shanks or taps a wireless controller for a rough instruction.	A reconstructed human-skin hand model of the local user presents on top of the remote world.
	The local user uses free air gestures to perform two functions of gestural interaction.
The outdoor user uses a mono LCD display for a single eye to present GUI.	An augmented reality smart glass helps the remote user to get an immersive experience in the gestural communication.
The outdoor setup is a complicated assembly device mounted on the outdoor user's neck.	The remote user wears a pair of portable smart glasses and camera which are light and convenient.

Chapter 6

Preliminary Evaluation

We conducted a user experiment to evaluate the system performance. We wanted to test whether the users could use our system to achieve an effective gestural interaction with our designed functions. Our target was to show whether our designs are reasonable enough and whether such gestural interaction with panoramic browsing could be used in the context of remote sightseeing to provide a *Trip-together Feeling*.

6.1 Participants

We recruited 8 participants, ranging in age from 23 to 27. They included 2 females and 8 males. They were divided into 4 groups, two in each group. In each group, one of the participants (remote user) went outside, and the other one (local user) remained in a room.The study took approximately 35 minutes.

6.2 Method

The task was that the local user instructed the remote user to buy a snack in the supermarket through our system. The remote user might walk around freely and communicate with the local user. The local user was asked to decide what to buy. Before taking the experiment, the participants were asked to practice using the system for about 15 minutes. After that, each group had 20 minutes to accomplish the task.

After finishing the work, every participant filled a questionnaire (Figure 6.1). They needed to answer the following 6 questions by grading from 1 to 5(1=very negative, 5=very positive).

1. Did you feel the Attention Reminder function was useful in your sightseeing?
2. Did you feel the gestural input was helpful?
3. Did you feel the Gestural Navigation function was helpful?
4. Did you feel the Pointing Assistance function was useful?
5. Did you think such gesture communication was easy to use during sightseeing?
6. Did you think such gesture communication was easy to use during sightseeing?

6.3 Conditions

All groups took the experiments during the day time. When experiment began, all remote users started from the main entrance of the supermarket.

Questionnaire

Welcome to this very important questionnaire with which we researchers want to evaluate our designs. Thank you for filling it all out!

About you

1. **Are you the local user or the remote user in the test?**
   □ The local user
   □ The remote user

2. **How old are you?** I am ________ years old.
3. **Male or female?**
   □ Male
   □ Female

Questions

Each question is graded from 1 to 5.

4. **Did you feel the Attention Reminder function was useful in your sightseeing?**
   almost useless □—□—□—□—□ very useful

5. **Did you feel the gestural input was helpful?**
   almost helpless □—□—□—□—□ very helpful

6. **Did you feel the Gestural Navigation function was helpful?**
   almost helpless □—□—□—□—□ very helpful

7. **Did you feel the Pointing Assistance function was useful?**
   almost useless □—□—□—□—□ very useful

8. **Did you think such gesture communication was easy to use during sightseeing?**
   quite hard □—□—□—□—□ very easy

9. **Did you feel you were walking with your partner together?**
   horrible □—□—□—□—□ fantastic

Researcher Remarks

10. No.________ group.

Figure 6.1 Questionnaire

6.4 Results

In our user experiment, all groups completed the task within the stipulated time. After collecting the questionnaire results from the participants (4 remote users and 4 local users), we calculated the average scores of each question from the participants, divided into two categories: the remote user and the local user (Figure 6.2).

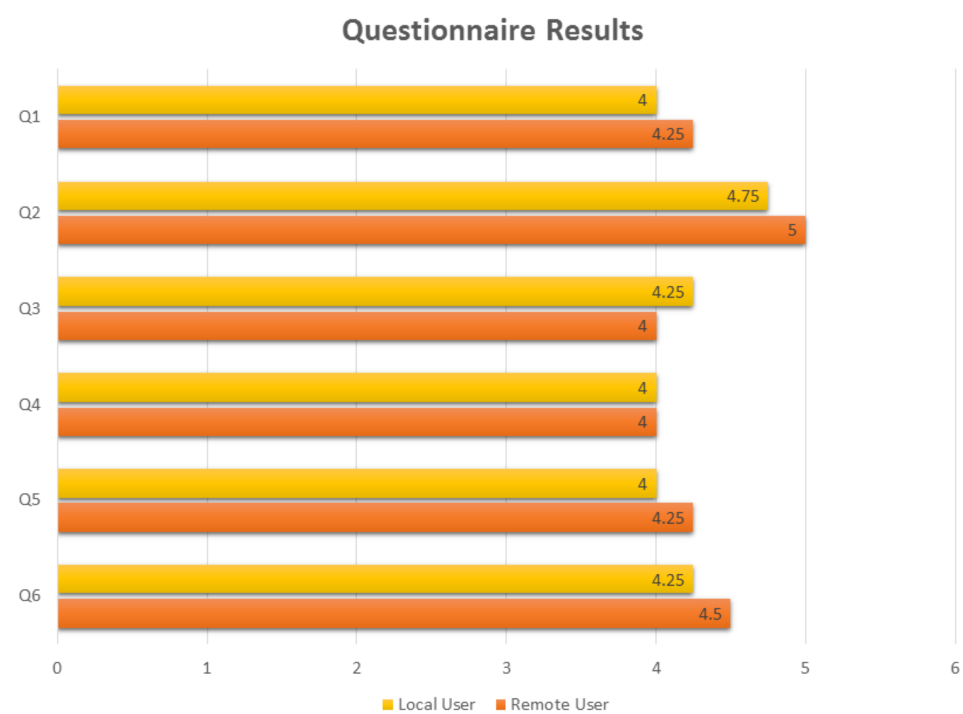


Figure 6.2 Questionnaire results

Question 1 to 4 are regarding the practicability of the four main designs. In each question, the average scores of both local user and remote user are higher than 4 points, which prove that our designs are reasonable and practical. Results of question 1 indicate that each user thought to provide a joint attention was constructive while both users had separated free viewpoint. For question 2, the results show that supporting an air gestural input on the remote scenery is helpful and effective for both local and remote user. Our two functions of gestural interaction did enhance the communication between the two users.

Question 4 and 5 are used to judge the overall performance. Question 4 regards the ease of use of our system. The results suggest that the user generally found the gestural communication is easy to carry out and effortless on our prototype. Question 5 prove that by supporting effective gestural communication on top of the shared world, our prototype could provide a Trip-together Feeling. In the post-task interviews, all the participants commented that they would found feature of *Trip Together* to be useful in the remote sightseeing. When asked about the experience performing gestural communication, the remote users considered that it was intuitive and distinct to see the human-skin hands of the local user in the field of vision, while the local users responded that they could feel personally on the scene to some extent. Some of our participants even played a “rock-paper-scissors" game through our system.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this work, we propose our prototype system for a remote pair sightseeing between a remote user and a local user in the remote pair sightseeing. By providing separated independent free viewpoint and air gestural input on top of the remote scene, we realize an intuitive air gestural communication between the two users. It simulates the local user is tripping together side by side with the remote user.

Our system received a positive feedback from the preliminary experiment. It indicates that the users could perform an effective gestural communication in the mobile pair sightseeing using our system and experience *Trip-together Feeling* to some extent. Although in this paper we test the system in a joint shopping scenario, it also suitable for other possible application like travel guide and cooperative work.

7.2 Future Work

In the future work, we plan to further improve this framework. For example, in the current implementation, some users point out the discomfort caused by camera shake in the moving situation. We may adopt a more stable design of setup to enhance the user experience. In the future studies, we intend to implement new features that presenting an avatar of the local user in the remote scenery to enhance *Trip-together Feeling*.

Bibliography

[1] Ambrosch, K., and Kubinger, W. Accurate hardware-based stereo vision. *Computer Vision and Image Understanding 114*, 11 (2010), 1303–1316.

[2] Chang, C.-T., Takahashi, S., and Tanaka, J. Analyzing interactions between a pair out together real and virtual. *Proc. collabTech'12* (2012), 100–105.

[3] Chang, C.-T., Takahashi, S., and Tanaka, J. Withyou-a communication system to provide out together feeling. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, ACM (2012), 320–323.

[4] Chang, C.-T., Takahashi, S., and Tanaka, J. A remote communication system to provide "out together feeling". *Journal of Information Processing 22*, 1 (2014), 76–87.

[5] Cobb, S. V., Nichols, S., Ramsey, A., and Wilson, J. R. Virtual reality-induced symptoms and effects (vrise). *Presence: teleoperators and virtual environments 8*, 2 (1999), 169–186.

[6] Dominio, F., Donadeo, M., Marin, G., Zanuttigh, P., and Cortelazzo, G. M. Hand gesture recognition with depth data. In *Proceedings of the 4th ACM/IEEE international workshop on Analysis and retrieval of tracked events and motion in imagery stream*, ACM (2013), 9–16.

[7] Garg, P., Aggarwal, N., and Sofat, S. Vision based hand gesture recognition. *World Academy of Science, Engineering and Technology 49*, 1 (2009), 972–977.

[8] Gauglitz, S., Nuernberger, B., Turk, M., and Höllerer, T. In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, ACM (2014), 197–205.

[9] Gurevich, P., Lanir, J., Cohen, B., and Stone, R. Teleadvisor: a versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2012), 619–622.

[10] Ho, T., and Budagavi, M. Dual-fisheye lens stitching for 360-degree imaging. In *Proc. of the 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'17)* (2017).

[11] Hunter, S. E., Maes, P., Tang, A., Inkpen, K. M., and Hessey, S. M. Waazam!: supporting creative play at a distance in customized video environments. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM (2014), 1197–1206.

[12] Karam, H., and Tanaka, J. Two-handed interactive menu: An application of asymmetric bimanual gestures and depth based selection techniques. In *International Conference on Human Interface and the Management of Information*, Springer (2014), 187–198.

[13] Karam, H., and Tanaka, J. Finger click detection using a depth camera. *Procedia Manufacturing 3* (2015), 5381–5388.

[14] Kasahara, S., and Rekimoto, J. Jackin: integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of the 5th Augmented Human International Conference*, ACM (2014), 46.

[15] Kashiwabara, T., Osawa, H., Shinozawa, K., and Imai, M. Teroos: a wearable avatar to enhance joint activities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2012), 2001–2004.

[16] Keskin, C., Kıraç, F., Kara, Y. E., and Akarun, L. Real time hand pose estimation using depth sensors. In *Consumer Depth Cameras for Computer Vision*. Springer, 2013, 119–137.

[17] Koizumi, S., Kanda, T., Shiomi, M., Ishiguro, H., and Hagita, N. Preliminary field trial for teleoperated communication robots. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, IEEE (2006), 145–150.

[18] Kolb, A., Barth, E., Koch, R., and Larsen, R. Time-of-flight sensors in computer graphics. In *Eurographics (STARs)* (2009), 119–134.

[19] The oculus rift, oculus touch, and vr games at e3. https://www3.oculus.com/en-us/blog/the-oculus-rift-oculus-touch-and-vr-games-at-e3/.

[20] Powering the rift. https://www3.oculus.com/en-us/blog/powering-the-rift/.

[21] 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). *Nihon Kyukyu Igakukai Zasshi 11*, 1 (2000), 1–7.

[22] Raffle, H., Ballagas, R., Revelle, G., Horii, H., Follmer, S., Go, J., Reardon, E., Mori, K., Kaye, J., and Spasojevic, M. Family story play: reading with young children (and elmo) over a distance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2010), 1583–1592.

[23] Santos, B. S., Dias, P., Pimentel, A., Baggerman, J.-W., Ferreira, C., Silva, S., and Madeira, J. Head-mounted display versus desktop for 3d navigation in virtual reality: a user study. *Multimedia Tools and Applications 41*, 1 (2009), 161.

[24] Scaramuzza, D., Martinelli, A., and Siegwart, R. A flexible technique for accurate omnidirectional camera calibration and structure from motion. In *Computer Vision Systems, 2006 ICVS'06. IEEE International Conference on*, IEEE (2006), 45–45.

[25] Silberman, N., and Fergus, R. Indoor scene segmentation using a structured light sensor. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, IEEE (2011), 601–608.

[26] Sodhi, R. S., Jones, B. R., Forsyth, D., Bailey, B. P., and Maciocci, G. Bethere: 3d mobile collaboration with spatial input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2013), 179–188.

[27] Tan, C. T., Leong, T. W., Shen, S., Dubravs, C., and Si, C. Exploring gameplay experiences on the oculus rift. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, ACM (2015), 253–263.

[28] Tecchia, F., Alem, L., and Huang, W. 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 (2012), 323–328.

[29] Weichert, F., Bachmann, D., Rudak, B., and Fisseler, D. Analysis of the accuracy and robustness of the leap motion controller. *Sensors 13*, 5 (2013), 6380–6393.

[30] Wu, Y., and Huang, T. S. Hand modeling, analysis and recognition. *IEEE Signal Processing Magazine 18*, 3 (2001), 51–60.