Doctorate Thesis

Human Body Mapping and Augmentation for Immersive Telepresence Systems

（没入型テレプレゼンス環境における身体のマッピングと拡張に関する研究）

June. 17, 2015

The University of Tokyo

Graduate School of Interdisciplinary Information Studies

49-127405

Keita Higuchi　樋口　啓太

Adviser Prof. Jun Rekimoto

ABSTRACT

Immersive environments are interactive spaces generated by computers, which integrate different types of information, such as embedded data, user input, and real-world information, in order to (re)construct environments. An immersive environment can provide an optimized workspace for users depending on the desired application. Immersive environments are required to provide not only environment (re)construction, but also interactive viewing and graphics facilities. Immersive environment applications include remote-working features involving, for example, telepresence or telexistence technology. Such technology can allow remote operation and collaboration. For example, telexistence robots utilize a mapping method in which upper body gestures are synchronized between operators and robots. The operators directly manipulate the robot hands using their hand movements. Recently developed techniques have provided a method of motion scaling that alters the mapping ratio of the operator and robot hand motions to allow accurate small-scale manipulation beyond the capability of the human hand. Telepresence technology can offer remote communication and collaboration between people, and telepresence systems can transmit a remote partner's social cues, such as eye contact, gesture direction, and body proximity, to achieve seamless collaboration between users. Well-designed telepresence systems facilitate immersive experiences where users can concentrate on their tasks without requiring prior practice with the system. However, the current technology still has several limitations regarding movement control operability/scalability, and deployability of the systems. In this thesis, we focus on the following three problems that affect two types of telepresence applications:

(A) In remote robot operation, current controlling methods require special skills or long-term training to manipulate robot movements. Rescue and surveillance situations aim to find targets, or create environmental maps using remote robots. An operator needs to control the robot from a local operation room manually, but control methods are still traditional controller (e.g. Joystick). Thus, direct movement controlling is still challenging. (Operability)

(B) When the remote robot operation uses the mapping method for movement controlling, an efficient mapping ratio has not been explained. In addition, vertical movements of the robot are still limited to the operator's height. Workspaces of the operator and the robot are not same dimensions at all time. Thus, the system should realize scalable direct controlling. (Scalability)

(C) In a remote conferring system, it is hard to balance transmitting social cues and constructing a system on a simple setup. Social cues such as eye contact, attention/intention, and facial expression are the key factor for remote collaboration. The system captures human body information using complex equipment (e.g. multiple cameras). The system also represents visualization of the remote partner on special displays. In contrast, the simple setup is also crucial for spreading to office and home environments. (Deployability)

In this thesis, we introduce a body mapping and augmentation method. The body mapping involves an interaction mechanism for correspondence between information regarding the human body and regarding a remote environment, for use in immersive telepresence systems. Body information includes movement, postures, and gestures. This method transforms the behavior of the human body into system outcomes. Moreover, we propose an augmentation method that seamlessly varies in correspondence with the human body, in accordance with specific purposes and situations. The method allows overcoming human physical restrictions (e.g., body lengths). We assume that the application of this mapping and augmentation method to immersive telepresence systems will solve the three previously defined problems. We also apply this method to two types of telepresence systems: robot operation and remote collaboration in immersive environments.

First, we develop a flying telepresence system that synchronizes the movements of an operator with an unmanned aerial vehicle (UAV) in order to address two of these problems (A and B). The system also enables a linear positional mapping that linearly maps the operator's movements onto the motion of the UAV. In addition, the system supports the use of a small hand-held device that controls the altitude of the UAV to overcome the limitations imposed by human height. We perform two user studies to measure the system operability and the effects of the mapping.

The first study indicates that the system allows more direct manipulation of the UAV compared to traditional control methods. The second study compares four mapping ratios (1:1 — 1:4 mappings). The result shows that augmenting mapping ratios can provide the operator with better control capability than the normal mapping (1:1 mapping).

Next, we also develop a remote collaboration system using a digital whiteboard to solve the problem (C). The system provides participants with an immersive collaboration experience, enabled using only an RGBD camera, which is mounted on the side of a large touchscreen display. The system realizes three novel visualizations of a remote partner using human body transformations and a simple setup. A user study shows that the system can provide participants with a quantitatively better ability to identify their remote partners' social cues. In addition, these quantitative capabilities translate qualitatively into a heightened sense of togetherness and a more enjoyable communication experience. A form factor of the system is its suitability for practical and easy installation in homes and offices.

These designs, implementations, and their subsequent evaluations prove that the three stated problems, operability, scalability, and deployability, are solved by our two proposed systems. By developing these systems, we aim to realize scalable remote interactions that connect differently scaled remote and local workspaces.

Acknowledgment

First, I especially would like to thank Prof. Jun Rekimoto, Interfaculty Initiative in Information Studies, The University of Tokyo and Deputy Director, Sony Computer Science Laboratories, Inc. It was because of his kind advice and supports that I could enjoy my Ph.D. career and finish this thesis. He has patiently believed me for the whole my career and he allowed me to research whatever I believe important.

My special thanks also go to the thesis committee members Prof. Ken Sakamura, Prof. Noboru Koshizuka, Prof. Akihiro Nakao, and Prof. Yoichi Satofor accepting to be my thesis committee and giving detailed and fruitful comments on this Ph.D. thesis.

During my twice internships at Microsoft Research, Redmond, USA, I received kind help from my mentors Dr. Zhengyou Zhang, Dr. Yinpeng Chen, Dr. Philip A Chou, and Dr. Zicheng Liu along with other members of the Multimedia Interaction and Communication group. It was my first time to live and research in a foreign country, but I felt really comfortable thanks to them. I believe this experience really broadened my research directions, as well as my way of thinking things internationally.

In my B.Sc. career in Kanazawa Institute of Technology, I was taught the way of academic thinking and presentation by Dr. Ryuichiro Hara. It was a great fortune that I received his advices in the beginning of my research life. I also thank Prof. Taku Takeshima, Associate Prof. Tomohito Yamamoto, and Associate Prof. Hidekazu Kanemitsu to give many valuable advises for my life.

My research and development activities were guided by Dr. Takashi Miyaki, Kensuke Habuka, Dr. Masaki Hiraga, Naoyoshi Kobayashi, Dr. Naotaka Fujii, Dr.

Sohei Wakisaka, Associate Prof. Kouta Minamizawa and Prof. Masahiko Inami. They give me meaningful advises to think my carrier.

In my academic research activities, I was excited working with Dr. Akiyama Soramichi, Katsuya Fujii, Kei Nitta, Michihiko Ueno, Natsumi Asahara, Thammathip Piumsomboon, Shunsuke Takahashi, Takamitsu Hamajo, Tetsuro Shimada, Yohei Yanase, and Dr. Yoshio Ishiguro. They provided many opportunities to discuss and proceed researches together. I thank to Dr. Azusa Kadomura, Shogo Ando, and Dr. Yoichi Ochiai to give many comments for my thesis.

I would love to state my thanks to all members of Rekimoto laboratory. My five years here were really precious, and with people in Rekimoto lab, I did experience many things. Some were enjoyable, some were tough, but all of them are living in me as good memories. My appreciation goes not only to the current members but also to former members and staffs.

I am also grateful to all my friends who are not only in Japan but also in all over the world, not only from academic field but also from anywhere I have been involved. I was gifted many kind friends in Kitahara Childcare Center, Nakajo Elemental School, Najajo Junior High School, Tokamachi-Sogo Senior High School, Kanazawa Institute of Technology, The University of Tokyo, internship at Microsoft Research Redmond, academic conferences I attended, and many meet-ups and events related to my hobbies. Having friends outside of the laboratory allowed me to broaden my sights in the sense both of research and private life.

Lastly but mostly, I want to express my deepest thanks to my family, for believing and supporting me for 27 years of my life. I believe that the education and many chances they give to me are the most valuable things that I have ever received.

table of contents

list of figures

list of tables

Chapter 1

Introduction

In this thesis, we focus on interaction mechanisms in immersive telepresence systems using human body mapping and augmentation. In this section, we first provide the background for our research, discussing immersive environments in humancomputer interaction (HCI) as well as technical components and interaction designs for immersive applications. Then, we state the aims of this thesis.

1.1 Background

1.1.1 Environments for Human-Computer Interaction

Environments for interaction between humans and computers are growing for each purpose such as mathematical calculation, controlling systems, finding information. The Electronic Numerical Integrator And Computer (ENIAC), the first electronic general-purpose computer developed in 1946, could allow program conversion for calculations. Programmers used patch boards and punch cards in ENIAC, though it was not interactive. Interactive environments came to the fore following the 1960s, but were mainly employed on flat and fixed displays. During the same time, however, Ivan E. Sutherland proposed futuristic displays and interactive environments in his historical essay "The Ultimate Display" [9] in 1965. Three years later, he developed the "Sword of Damocles" as the first head-mounted display [6]. It could change the user's perspective of digital objects depending on his/her head movements. This application exemplified the concepts of Virtual

Reality (VR) and Augmented Reality (AR). This work inspired the next generation of research on humancomputer interaction (HCI) environments, including real world-oriented environments and immersive environments.

Real-world oriented interaction provides computer-augmented environments to perform real-world tasks [10]. Research on real-world oriented interaction began in the early 1990s, and included work on AR, mobile interaction, and ubiquitous computing [11]. Wellner developed a projected physical desk system called "Digital Desk" to enhance deskwork experience, which offered features such as copy-and-paste functions, calculator, and typing functions [12]. Rekimoto proposed the first mobile augmented reality device [1]. As shown in Figure 1.1, he conceptualized interaction environments, including graphical user interfaces (GUIs), VR, ubiquitous computing, and AR. Ubiquitous computing and AR are suited to real-world tasks and can apply digital behaviors to the real world.

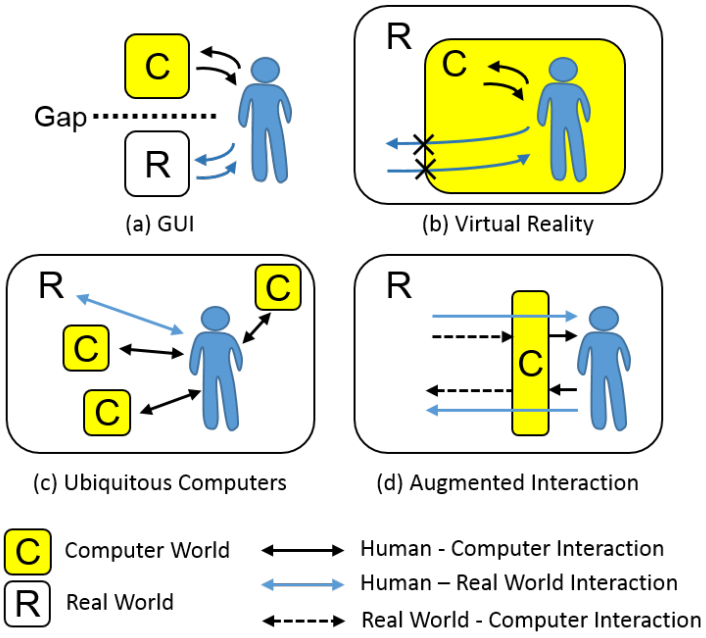

Figure 1.1: Concepts of Interaction Environments (redrew based on Rekimoto's paper[1]) (a) GUI, (b) Virtual Reality, (c) Ubiquitous Computing, (d) Augmented Reality

By contrast, immersive environments are interactive spaces constructed by com-

puters. To construct such environments, the computer integrates varied information, such as embedded data, user inputs, and real-world information. Saied explained immersive environments as "Immersive Telepresence" [13] in that its technologies support the (re)construction of environments, and interactive viewing and graphics. The environments can be optimized according to applications, including remote robot controlling, remote collaboration, big data browsing, training, and gaming. These environments were initially called "Artificial Reality," a name coined by Myron Krueger [14]. He developed "Videoplace" as an artificial reality system that combined a user's live image with computer graphics in the mid-1970s [15]. In Videoplace, users could interact with objects generated by computer graphics by using their bodies. In the 1980s, Tachi proposed telexistence [16], "a fundamental concept that refers to the general technology that allows a human being to experience a real-time sensation of being in a place other than his/her actual location, and to interact with the remote environment, which may be real, virtual, or a combination of both." He worked on the TELESAR system, which is a telexistence surrogate robot controllable using a masterslave system. Operators of the system can control the arms of the remote, human-like robot by moving their arms, and can watch real-time images from the robot's eye-camera using a head-mounted display (HMD) [3].

Figure 1.2 shows the conceptual diagram of the immersive environment along the lines proposed by Rekimoto (Figure 1.1). A user interacts with a computer in a generated environment. The computer can mediate interaction between the user and the real world according to applications. Real-world computer interaction affects the real-world environment and objects based on inputs from the user, and senses real-world information to reconstruct the environment. In a humancomputer interaction loop, the computer represents sensory feedback, such as visual, audio, and haptic. At the same time, the computer receives purposeful/involuntary inputs from the user through gestures, the pressing of buttons, and biological information.

The applications of immersive environments have become increasingly important. In the past, the development of the immersive applications required expensive equipment and special tools, and was thus limited to large-scale projects. In this decade, research has focused on the development of devices and equipment for

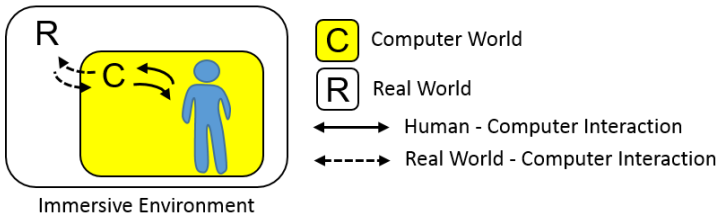

Figure 1.2: Concept of Immersive Environment

immersive applications for personal use, such as the Oculus Rift, Leap Motion, and Microsoft Kinect.

1.1.2 Technical Components in the Immersive Environment

Systems of immersive environments have four technical components from the perspective of user experience.

Representation

Representation offers sensory feedback to support the visual, audio, haptic, palatal, and olfactory senses. Most current applications only include audio and visual displays as general sensory feedback. As shown in Figure 1.3, immersive systems need a display to represent the relevant environment, such as an HMD, large displays, or cave automatic virtual environment (CAVE) systems [17]. Each display has different features, and system developers choose the display depending on the application at hand, the users, and the budget. For example, HMDs can provide an immersive perspective where a user can see the computer-constructed world. It can change perspective based on the user's head rotation and translation. Representation includes physical devices as well as software technology to transform visual information.

Input

The input component is an interface to receive incoming commands from the user. Input includes traditional button/stick devices and information regarding the human body, such as figure, gesture, motion, and position. Figure 1.4 shows

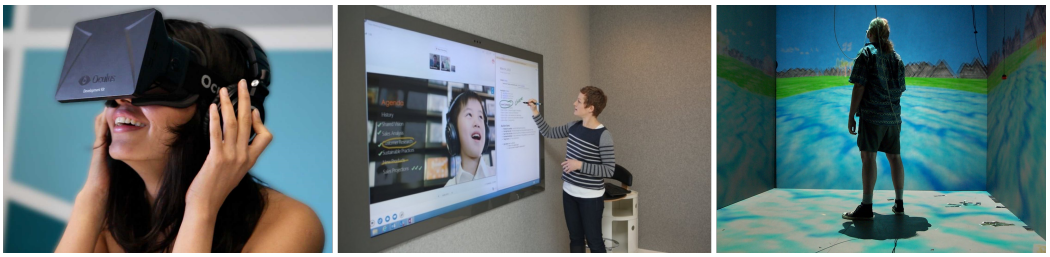

Figure 1.3: Displays for immersive environments

typical input devices and systems, such as the Wii-Remote, Microsoft Kinect, and the motion capture system. Modern devices can sense multiple forms of information. For example, the Wii-Remote can detect button push, hand movement using acceleration sensors, and the position of the controller. Involuntary biological information, such as heart rate, can also be used as input. The computer transforms outputs to the real/artificial world according to the user's input.

Figure 1.4: Input methods for immersive environments

Sensing

Sensing captures characteristics of the real-world environment, including optic, acoustic, and positional information. The devices and systems used for sensing are similar to those used for input, can be on a fixed platform or movable sensors (e.g., robots). Sensing also offers software technologies to reconstruct immersive environments using the captured information. Kanade et al. proposed "Virtualized Reality" [18], which captured and reconstructed real-world environments to provide a free-viewpoint experience, whereby the user can control his/her viewpoint in the reconstructed environment. They developed a prototype system that contained 51 cameras and carried out three-dimensional (3D) reconstruction.

Actuation

Actuation is a function that affects human inputs to the real-world environment. To implement actuation, immersive systems use robotic systems (Figure 1.5), such as ground vehicles and humanoids that can act and move in a real-world environment. A projected avatar [4] can communicate and dictate to a remote worker to realize human-supported collaboration. Immersive applications can substitute robotic systems and avatars for users.

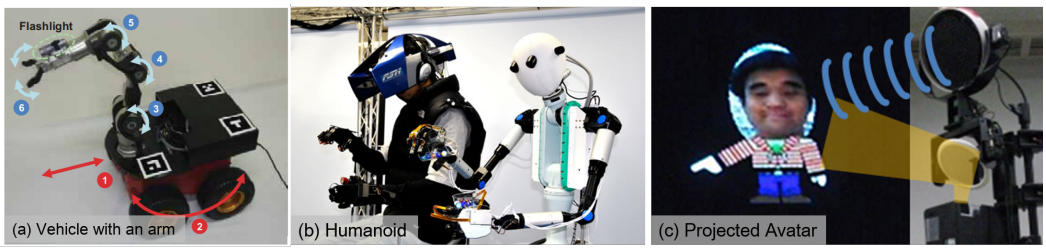

Figure 1.5: Actuators for systems of immersive environments. (A) a ground vehicle with a robotic arm [2], (B) a humanoid system [3], (c) a projected avatar system for remote collaboration [4]

1.1.3 Interaction Design for Immersive Environment

Interaction design describes the behavior of a system, and defines users' actions and their outcomes [5]. In an immersive system, interaction design also introduces technical components based on application types, such as remote operation/collaboration, and VR tasks. Figure 1.6 shows that applications, components, and interaction design affect one another. Thus, Immersive systems need to select interaction mechanisms depending on application and technical components according to the objectives in question.

Immersion/immersive experience is crucial for work productivity and learning efficiency that the experience is subjective impression as realistic experience in the immersive environment. A recent study showed that immersive experience can enhance educational experience in three ways: multiple/changeable perspectives, situated learning, and transfer [19]. The quality of the immersive experience is

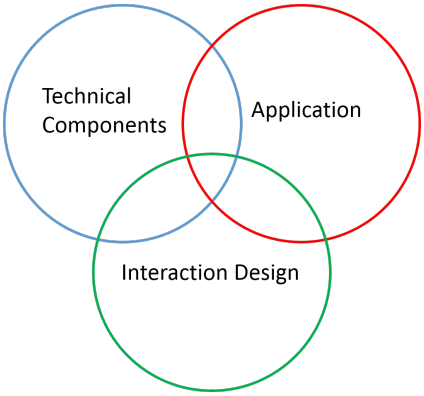

Figure 1.6: Interaction design (Redrawn based on [5])

determined through interaction design strategies, such as input mechanisms and sensory feedback. In an effective immersive experience, the user can focus on tasks and learning without requiring time to familiarize himself/herself with the system. A failed interaction design leads to unaccomplished objectives, or requires long training times. For instance, as shown in Figure 1.5 (A), an operator cannot intuitively manipulate a remote ground robot with a three-jointed arm if the controller only allows multiple buttons and a stick control. With regard to this issue, Hashomoto et al. developed an information over-rayed interface for direct robot manipulation using a touch display [2]. A successful interaction mechanism saves time and enhances immersion experience.

An interaction design theory supports the development of immersive environment applications. As shown in past research, interaction designs already exist for several specialized applications, such as ground vehicle navigation [20], human-supported remote co-working [21], and artistic medium in the artificial world [22]. Poupyrev et al. proposed "the Go-Go interaction technique," which uses non-linear mapping for direct manipulation of the virtual reality environment [8]. Bowman summarized interaction design research as a theory of 3D user interfaces to develop virtual reality systems, and developed complex data visualizations and education applications based on the theory [23]. Interaction design can help to decide technical components and its relationship to configure immersive systems.

1.2 Problem Statement

Immersive environments offer various remote working experiences through telepresence/telexistence systems. Major applications of this kind include remote collaboration and robot operation. In remote robot operation systems, an operator controls a remote robot and experiences remote environments using sensing information from that robot. The computer synchronizes the operator's body movements with those of the robot. On the other hand, in remote collaboration systems, local and remote users collaborate through an immersive environment, which transmits information regarding each user's body movements as well as a shared working space to the other users. These users can work from their local environments, thus, these systems reduce translation costs, working time, and a number of risks. In addition, these systems can introduce digital benefits to enhance the working experience.

However, current remote collaboration and operation systems have several limitations regarding movement control operability/scalability, and deployability of the systems. The sources of these difficulties include problems posed by different workspaces in different locations, communication latency, and user interface problems. In local applications, human body mapping has been introduced to immersive interaction mechanisms to facilitate direct interaction. On the other hand, in immersive telepresence applications, the effectiveness of this approach is still unclear. That is, certain difficulties exist regarding the manner in which the mapping mechanism connects the human body with environmental information and overcomes human physical limitations.

This thesis focuses on the following three problems (A, B, and C) regarding two types of immersive telepresence applications:

(A) In remote robot operation, current control methods require special skills or long-term training in order to manipulate robot movements. For example, in rescue and surveillance situations, the aim is to locate targets or to create environmental maps using remote robots. An operator must manually control the robot from a local operation room, but the control methods still involve a traditional controller (e.g., a joystick). A mapping method that syn-

chronizes upper body gestures between the operator and the robot has been introduced to the humanoid telexistence robot. However, direct movement control remains a challenge. (Operability)

(B) An efficient mapping ratio for cases in which remote robot operation utilizes a mapping method for movement control has not been established. In addition, vertical movements of the robot are still limited by the operator's height, although the operator and robot workspaces do not always have the same dimensions. Thus, the system should realize scalable direct control. (Scalability)

(C) In a remote conference system, it is difficult to balance the transmission of social cues and the construction of a system based on a simple setup. Social cues such as eye contact, attention/intention, and facial expression are key factors for remote collaboration. The system must capture information related to the human body using complex equipment (e.g., multiple cameras). The system must also allow visualization of the remote partner on special displays. In contrast, a simple setup is also crucial for the distribution of this technology to office and home environments. (Deployability)

1.3 Thesis Statement

In this thesis, we introduce a body mapping and augmentation method. The body mapping involves an iteration mechanism for correspondence between information regarding the human body and regarding a remote environment, for use in immersive telepresence systems. "Body information" includes movement, postures, and gestures. This method transforms the behavior of the human body into system outcomes. Moreover, we propose an augmentation method that seamlessly varies in correspondence with the human body, in accordance with specific purposes and situations. The method allow overcoming human physical restrictions (e.g., body lengths). We assume that the application of this mapping and augmentation method to immersive telepresence systems will solve the three previously defined problems. We also apply this method to two types of telepresence systems:

robot operation and remote collaboration in immersive environments. The robot operation system uses a positional mapping between the operator and robot positions. The system also supports a small hand-held device that controls the altitude of an unmanned aerial vehicle (UAV), as a means of addressing problems (A and B). The remote collaboration applies a linear 2D/3D visual mapping and transformation to solve the problem (C). We show that these designs, implementations, and their subsequent evaluations prove that the three stated problems, operability, scalability, and deployability, are solved by our two proposed systems.

1.4 Thesis Structure

Chapter 1 provides a general introduction to immersive environments and the problems affecting the development of immersive telepresence applications. Chapter 2 is devoted to a description of work related to mapping method, and also provides a relationship between our method and the proposed two systems. In Chapter 3, we describe the developed flying telepresence robot, including the research background in this area, details of the interaction mechanism, and its evaluation. In Chapter 4, We explain the immersive remote collaboration system known as ImmerseBoard, and discuss related research and remaining challenges. Chapter 5 discusses our findings related to the efficacy of the body mapping augmentation, and shows potential immersive environment applications using body mapping. Finally, in Chapter 6, we conclude the thesis and describe future research related to this study.

Chapter 2

Human Body Mapping in Immersive Telepresence Systems

In this chapter, we describe related work on mapping methods. This thesis aims to address the remaining problems affecting telepresence systems using the body mapping method.

2.1 Related Work

Mapping is an interaction design method that enables correspondence between a user's input and output motions. Mouse devices, proposed by Douglas Carl Engelbart, form the most popular mapping interface because their input speeds accord with Fitts'law in 2D GUI space [24, 25]. The mapping technique for hand-held devices has recorded satisfactory results in video gaming [26]. Using these mapping techniques, users predict outcomes directly from their 2D motion.

Mapping techniques are also useful for 3D motion inputs in VR research. Ware et al. explored a virtual camera manipulation technique using 3D hand motions with a small ball device [27]. Hinckley et al. extended this technique to two-handed manipulation, where users could use both hands to hold the camera and the shooting object [28]. Sturman et al. proposed data grove devices that measure hand shapes so that users can obtain self-hands in VR environments [29]. "VooDoo Dolls" is a two-handed interaction technique that uses data groves to create 3D computer-generated objects [30]. These 3D mapping techniques can provide direct

interaction for applications in complete VR environments.

The concept of telexistence was first proposed by Tachi et al. [16], and refers to the creation of surrogate humanoid robots in the physical world. In such systems, the operator's body movements are synchronized with those of a remote robot (Figure 2.1). TELESAR V [3] is the most recently developed telexistence robot. This robot synchronizes its movements based on information regarding the user's head, upper body, and hand movements, and transmits visual and haptic information to the operator. This complete mapping mechanism can provide suitable interaction capabilities for remote operation. However, complete mapping is limited to the abilities of the human body, such as its speed and movement range.

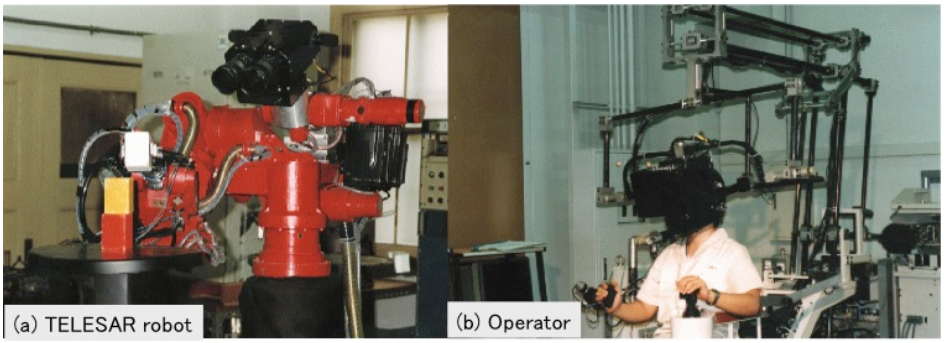

Figure 2.1: Telexistence robot TELESAR [6]

As shown in Figure 2.2, the da Vinci [7] is a surgical support system produced by Intuitive Surgical Inc. This system has introduced a master-slave control system that synchronizes the operator and machine hands. This system has also introduced a motion scaling method that linearly reduces the robot hand motion relative to the operator's hand motion. This method enables accurate manipulation of the robot hand in small spaces. The method proposed in this thesis contributes an augmentation method that seamlessly varies in correspondence with the human body, in accordance with specific purposes and situations for overcoming human physical limitations.

The Go-Go interaction technique proposed by Poupyrev et al. [8, 31]. realized direct manipulation of virtual objects using an operator's hands in a first-person virtual reality (VR) environment. This technique consists of normal and non-linear

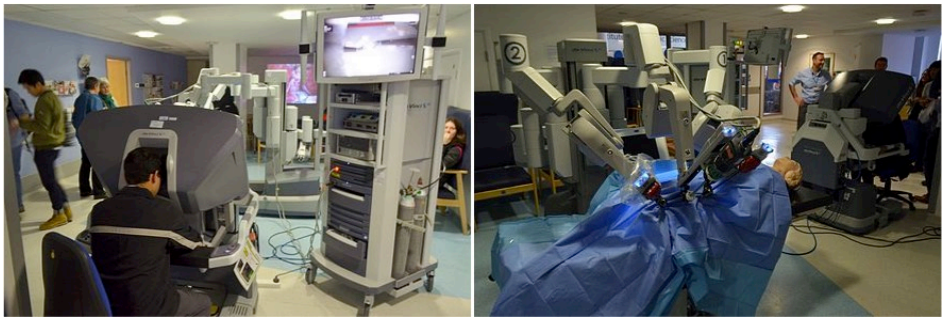

Figure 2.2: The da Vinci Surgical System [7]

mapping, whereby the hands can extend far into the non-linear region (Figure 2.3). Saraji et al. introduced a method of telexistence robot control that combines a real-world view with virtual extended hands to control environmental equipment [32, 33]. This method also significantly affects 3D user-interface fields, but it is specialized for first-person VR tasks.

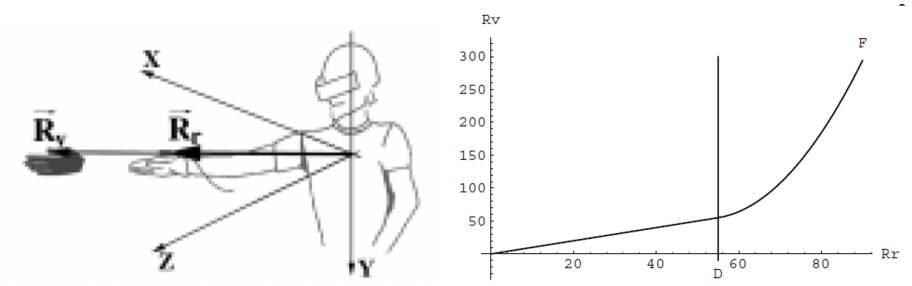

Figure 2.3: The Go Go Interaction Technique [8]

2.2 Proposed Immersive Telepresence Systems

Our challenge here is to deliver the advantages of the devised body mapping and augmentation method to immersive telepresence systems. We focus on telepresence systems that connect remote and local locations through an immersive environment. With these systems, users can remotely perform tasks from their local spaces. We assume that the proposed method can seamlessly transform the behavior of the human body (e.g., movements and shape) to outcomes in the re-

mote space. We propose and develop two types of telepresence systems using body mapping methods, i.e., a flying robot operation system and a process allowing remote collaboration through a digital whiteboard.

2.2.1 FlyingHead

FlyingHead is a telepresence system that remotely connects humans and UAVs. UAVs are teleoperated robots used in various situations, including disaster area inspection and movie content creation. FlyingHead aims to integrate machines with various capabilities (i.e., flying) with virtually augmented human capabilities. The precise manipulation of UAVs typically involves simultaneous control of the motion parameters, which requires a skilled operator. We propose a method of directly relating the motions of the operator's body and head to those of the UAV. The operator's natural movement can be synchronized with the motion of the UAV, facilitating rotation, horizontal and vertical movement, etc. In addition, this system also supports the use of a small hand-held device to control the altitude of the UAV to overcome the limitations imposed by human height.

The development challenges overcome by this system are related to two of the previously defined problems (A and B) regarding remote robot operation using an immersive environment. To examine the efficacy of this system, we first conduct an operability study to evaluate the performance of the head-synchronized control method in comparison with traditional control methods (e.g., using a joystick). A second user study is also performed to compare different mapping ratios, in order to prove the efficacy of the linear body mapping.

2.2.2 ImmerseBoard

ImmerseBoard is a system for remote collaboration through a digital whiteboard that provides participants a 3D immersive experience, enabled only by an RGBD camera (Microsoft Kinect) mounted on the side of a large touch display. Using 3D processing of the depth images, life-sized rendering, and novel visualizations, ImmerseBoard realizes three novel visualization of a remote partner, including Hybrid, Tilt, and Mirror conditions. ImmerseBoard provides participants with

a quantitatively better ability to estimate their remote partners' gaze direction, gesture direction, intention, and level of agreement. Moreover, these quantitative capabilities translate qualitatively into a heightened sense of togetherness and a more enjoyable experience. ImmerseBoard's form factor is suitable for practical and easy installation in homes and offices.

The design challenges overcome in the development of ImmerseBoard constitute a solution of problem (C). This system realizes to satisfy the requirements of both a simple setup and visualizations for the transmission of social cues. ImmerseBoard provides three types of visualization in order to transmit social cues such as eye contact, pointing direction, and body proximity within a single setup. To examine the performance of this system, we first perform qualitative and quantitative studies to evaluate the efficacy of and benefits provided by our three visualizations. We also determine the preferences of our study participants using questionnaires.

Chapter 3

FlyingHead: A Head-Synchronization Mechanism for Flying Telepresence

3.1 Introduction of this chapter

The underlying idea of telexistence and telepresence comes from "Waldo", which is a science fiction story by Robert Heinlein. He proposed a master-slave manipulator system for big scale robot control[34].This master-slave manipulator system of controlling a robot using the human body has been introduced as the research area of telepresence [35, 3]. Remote-operated robots have many applications, such as telecommunication [36] and disaster site inspection [37, 38]. Technologies involved in remote operation are often called telexistence or telepresence.

An unmanned aerial vehicle (UAV) is a flying robot that can move freely through the air and circumvent poor ground conditions such as uneven roads and non-graded areas. When the Tohoku-Pacific Ocean Earthquake occurred, human-controlled UAVs were used to survey the damage at the Fukushima Dai-1 nuclear plant. In a recent study, UAVs were used to capture 3D reconstructed images of indoor and outdoor environments using mounted cameras [39, 40]. Open-hardware UAVs such as MikroKopter [41] and Quaduino [42] have also contributed to projects.

Remote operational UAVs need to introduce localization technology such as GPS sensor, visual odometry, and its fusion. Modern micro-UAV systems [43, 44] provide location based controlling with GUI that its users point some locations to

Figure 3.1: FlyingHead is a telepresence system that remotely connects humans and unmanned aerial vehicles (UAVs). The system synchronizes the operator's head motions with the UAV's movement and it enables the operator to experiences augmented abilities as if he or she become a flying robot.

move the UAV. Location based controlling offers high-level interaction which can conceal small manipulations such as inclination and altitude keeping. The users can only focus on positional controlling. However, when the UAV is located in narrow indoor positions, there are several requirements: (a) the users navigates the UAV interactively in indefinite positions, (b) the users require real-time feedback to understand the remote environment such as obstacle objects and moving humans, (c) the users need a direct controlling mechanism with a robot movement latency.

This chapter addresses the challenge of realizing telepresence using the UAV. "Flying Telepresence" is the term we use for the remote operation of a flying surrogate robot. We propose a head-synchronization mechanism called FlyingHead. FlyingHead synchronizes user head motions with the movements of a flying robot, which can be manipulated by natural motions. FlyingHead aims to provide flying immersive experience, its operators can feel as if they became the flying machine.

3.2 FlyingHead Mechanism

FlyingHead is a head-synchronization mechanism that uses human head motions to control UAV movements. In this method, operators wear a head-mounted display (HMD) and move their own body. The operator's motions can be synchronized with a UAV, movements of the UAV are mapped to the user's kinesthetic imagery. For example, when an operator walks forward, the UAV flies in the same direction. When the operator crouches, the UAV also lowers itself to the ground. When the operator looks right or left, the UAV rotates to the same direction.

FlyingHead synchronizes operator head and UAV movements in 3-axis posotion $(x, y, andz)$ and 1-axis orientation (yaw) (Figure 3.2). FlyingHead is based on the positional control mechanism which provides high level manipulation. The positional control requires setting four parameters at least: pitch (front and back), roll (left and right), yaw (rotation), and altitude. FlyngHead automatically calculates these parameters in real-time based on the operator's head position to synchronize the head and the UAV position. Potentially, FlyingHead can realize 6-DOF synchronization between the head and the UAV. In this chapter, we aim to study values of the head-synchronization mechanism, thus we prototype 3-axis position and 1-axis orientation controlling.

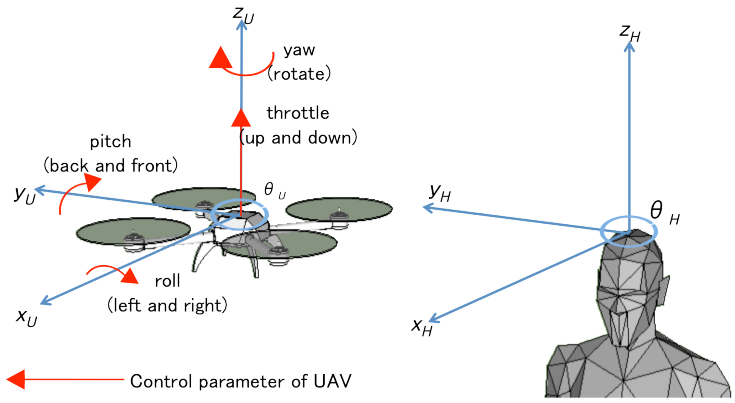

Figure 3.2: This mechanism synchronizes positions and orientations of humans and UAVs. For parameters (pitch, roll, yaw and throttle) are sent to control the UAV.

FlyingHead supports mapping ratio changing to augment head movements lin-

eally. If the system only uses same ratio mapping, the UAV movements are limited to operator's abilities such as speed and range of movements. Thus, the system has a gain parameter G to change mapping ratio. The system defines G based on the head movements $[\Delta X_H, \Delta Y_H, and\ \Delta Z_H]$ and the UAV movements $[\Delta X_U, \Delta Y_U, and\ \Delta Z_U]$.

$$G = (\frac{\Delta X_U}{\Delta X_H}, \frac{\Delta Y_U}{\Delta Y_H}, \frac{\Delta Z_U}{\Delta Z_H}) \tag{3.1}$$

As shown in Figure 3.3, the system decides the UAV moving distance based on the gain parameter. For example, in $G = 2$, the UAV moves 2 m when the operator walks 1 m. As the operator's small movement can be extended to a large movement by the UAV, the operator can control the speed of the UAV and feel less fatigue.

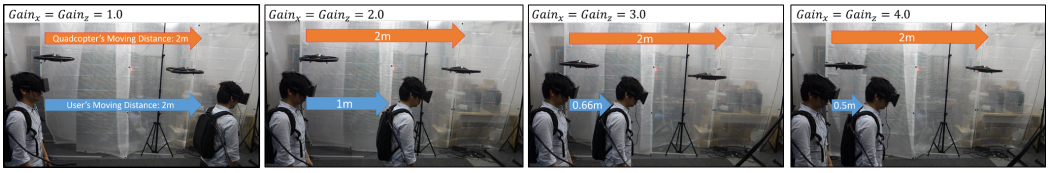

Figure 3.3: FlyingHead has a gain parameter G which changes mapping ratio of movements between head and UAV positions

In addition, FlyingHead also supports an accessional altitude controlling method using a hand-held controller device which are combined with the head-synchronization mechanism. The UAV can fly at higher/lower positions than the operator's head position. Even if the system raises the gain parameter, the operator cannot manipulate the UAV to high altitude positions. Because, moving space of altitude is restricted to the operator's height. To overcome this gap, the system provide the combined interaction mechanism that the operator moves the UAV to high/low positions through the controller device.

3.3 Related Work

A telepresence robot can conduct a wide range of tasks including telecommunications and remote operations. In recent years, telepresence robots have been used

in office environments [45]. Telepresence robots have even attracted attention for use in military applications [46].

3.3.1 Body motion input

In research applications, telepresence robots are manipulated by human body motions. Mancini et al. developed Mascot (Manipulatore Servo Controllato Transistorizzato), which has two stereo cameras and two rudimentary slave hands [47]. From 1983 to 1988, Hightower et al. demonstrated the possibility of remote presence by developing Green Man, which is an anthropomorphic manipulator with arthroarms [48]. Heuring et al. developed a visual telepresence system that slaves a static pan-tilt camera to human head motions [49]. Although flying robots must engage in up-and-down spatial movements, these developed robots are static plane robots incapable of generating differing vertical motions.

3.3.2 UAV operations

Quigley et al. described how devices such as PDAs, joysticks, and voice recognition systems can be used to set UAV control parameters [50]. Giordano et al. developed a situation-aware UAV control system that provides vestibular and visual sensation feedback using a CyberMotion simulator [51]. This system represents UAV motion information within the operator's vestibular system. Naseer et al. developed a person following UAV using gesture recognition technique. [52]. Shan et al. demonstrated a hand gesture-controlled UAV that uses six different gestures to control movements such as takeoffs, landings, climbing, and descending [53]. However, these gestures are essentially just a replacement for the device input, so using them for inputting parallel control parameters of the UAV is difficult. Vries et al. developed a UAV mounted with a head-slave camera [54]. Recent researchers have introduce head-coupled control with modern HMDs such as Oculus Rift [55], and Google Glass [56], but an operator needs to manipulate a hand held device for horizontal and vertical controlling. Hayakawa et al. developed a head-coupled controlling system for quadcopters, but the operator cannot control vertical movements. [57]. We focus kinesthetic imagery for controlling the UAV.

3.3.3 Interactive applications with UAVs

Recently, UAVs are used in many types of field such as entertainment, sports training, and media art. Iwata demonstrated a interactive installation "Floating Eye", which can show out-of-body images from a floating camera [58]. Yoshimoto et al. developed a unmanned blimp system, which has four types of use-case in entertainment computing fields [59]. Okura et al. proposed a augmented reality entertainment system using autopilot airship and omni-directional camera [60]. Previously, our group proposed a autonomous UAV to capture out-body-vision images for entertainment contents and sports training [61, 62]. Graether et al. also presented a jogging support UAV "Joggobot", which accompanies user's jogging to motivate exertion activities [63, 64]. We aim to realize telexistence and telepresence fields with UAVs.

3.4 Prototype System

The prototype system comprises a positioning measurement system, mini UAV, and HMD. Figure 3.4 shows the configuration of the system control using point information. An operator wears an HMD to represent the UAV's camera image, which allows the operator to control successive motions of the UAV.

To synchronize the operator's body motion with that of the UAV, the system requires position information. We used OptiTrack as an optical motion capture system for positional measurements. An OptiTrack S250e IR camera with a high frame rate can capture 120 fps, and motion capture allows the marker's position to be calculated to an accuracy of 1 mm. We captured the marker motions by installing eight cameras in a room divided into human and UAV areas: each was 3.0 m long $\times$ 1.5 m wide.

3.4.1 Micro-Unmanned Aerial Vehicle

We used AR.Drone as the flying telepresence robot: this is a small quadcopter with four blade propellers that can be controlled using Wi-Fi communication. AR.Drone has two cameras: one on the front and the other at the bottom. Fly-

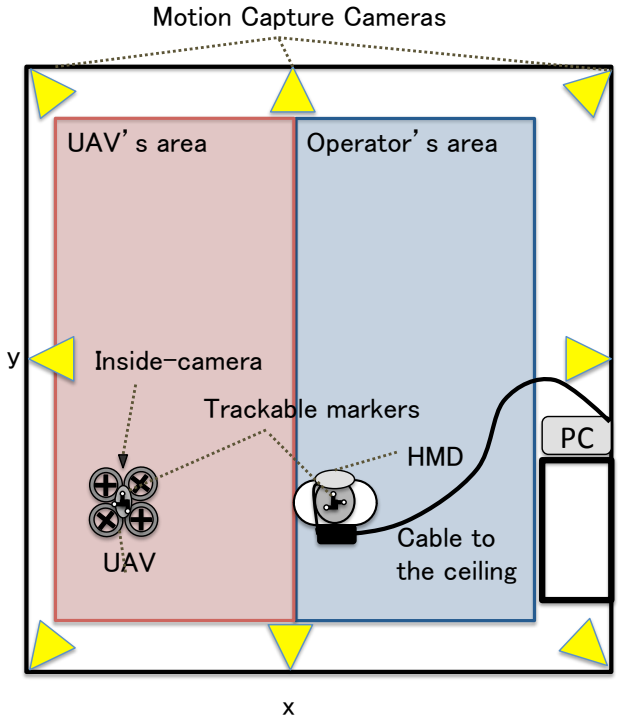

Figure 3.4: System configuration: The prototype system incorporates a position measurement system using eight motion capture cameras, a mini-UAV, and an HMD. The system is capable of several mapping scales.

ingHead uses the front camera for visual feedback.

AR.Drone has four control parameters: $pitch, roll, yaw$, and $throttle$. The $pitch$ controls the forward and backward movements, and the $roll$ controls the right and left movements. When the yaw parameter is changed, AR.Drone rotates on its site, and when the $throttle$ parameter is changed, AR.Drone moves up or down. The system sends the control parameters to AR.Drone once every 30 ms.

3.4.2 Visual Feedback

The operator wears a device with an HMD to represent images captured from the UAV cameras. For the HMD, we adopted a Sony HMZ-T2 or Oculus Rift DK1, which provides high-definition (HD) image quality. The HMD has markers that the system uses to track the operator's body motions. The user determines

the next manipulation of the UAV based on visual feedback from the previous manipulation. The wearable device is connected to 12 m long HDMI and power source cables that are extended to the ceiling. The inner camera of the AR.Drone has a QVGA resolution of 320 × 240 pixels with a capture speed of 30 fps. This camera is located at the front side of the AR.Drone.

3.4.3 Calculation of Control Parameters

The system uses the position information of the operator and UAV measured from the motion capture system. The positioning parameters include the point $[x, y, z]$ and its direction $[\theta]$. The system sets the *pitch* (front and back), *roll* (right and left), *yaw* (rotation), and *throttle* (up and down) parameters.

As shown in Figure 3.2, operator's head and UAV positions has each coordinate. First, the system maps the head position (H) to the a mapped position (M) of UAV coordinate depending on a gain parameter (G).

$$M = (H_x G_x, H_y G_y, H_z G_z, H_\theta) \tag{3.2}$$

The system obtains mapped (M_i) and UAV (U_i) positions at time i ($i = 0...k$), and also calculates difference of positions (D).

$$\mathbf{M}_i = \{x_i, y_i, z_i, \theta_i\} \quad (i = 0..n) \tag{3.3}$$

$$\mathbf{U}_i = \{x_i, y_i, z_i, \theta_i\} \quad (i = 0..n) \tag{3.4}$$

$$\mathbf{D}_i = \mathbf{M}_i - \mathbf{U}_i \tag{3.5}$$

At time i, $pitch_i$, $roll_i$ and yaw_i are calculated based on the following equation.

$$\begin{pmatrix} pitch \\ roll \\ yaw \\ throttle \end{pmatrix} = \begin{pmatrix} cos\ \theta_U & sin\ \theta_U & 0 & 0 \\ sin\ \theta_U & cos\ \theta_U & 0 & 0 \\ 0 & 0 & 1/\pi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_D \\ z_D \\ \theta_D \\ y_D \end{pmatrix} \tag{3.6}$$

The system also estimates the future position (expression 3.7) of the UAV based on the position history for a fast-converging UAV movement. The system trans-

forms the control condition (expression 3.9) so that the future position is greater than the current position (C:constant).

$$\mathbf{F}_{i+1} = \mathbf{U}_i + (\mathbf{U}_i - \mathbf{U}_{i-1})\Delta t \tag{3.7}$$

$$pitch = -pitch \times C \tag{3.8}$$

$$roll = -roll \times C \tag{3.9}$$

3.4.4 Combined interaction mechanisms for Altitude Control

For the combination of devices, the operator uses a combination of body motions for most movements and the control device for altitude control only. Initially, the altitude baseline is the head height of the operator, and the device can switch its baseline height. We adopted a Wii remote controller connected to a PC through Bluetooth. The operator changes the baseline by pressing the remote controller's arrow keys.

3.5 User Study 1

We conducted first user study to review the operability of the FlyingHead mechanism. We designed a shooting task to reveal the potential of remotely operated UAV for searching and inspecting a certain object. In the first study, We compared head-synchronization, hand-synchronization, and a traditional control method, including FlyingHead with the controller device ($G = 1$), FlyingHead with linear mapping ($G = 2$), hand-synchronization mechanism ($G = 2.5$), and the joystick controller. We assigned the same task to all of the controlling mechanisms. The subjects captured four static markers using the inner camera of the UAV with the different control methods.

3.5.1 Task and Environment

We measured the time to task completion. The subjects captured four visible markers using each UAV control mechanism. Figure 3.5(a) shows the experimental

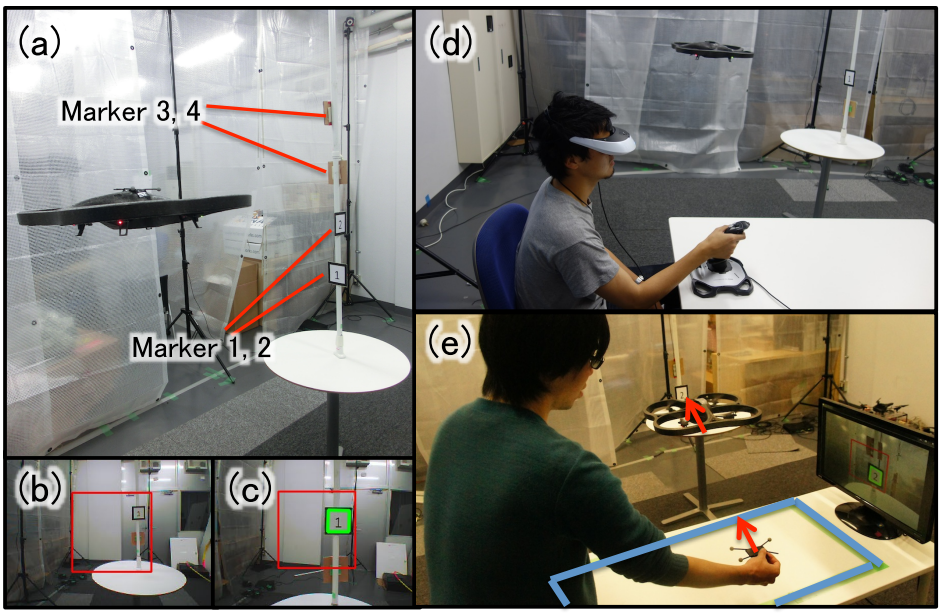

Figure 3.5: Environment of study 1: The subjects captured four visible markers using the each control mechanism. We measured and compared the completion time of task.

environment, which included a pole extending to the ceiling and four 2D markers. The markers were given the numbers 1-4. The subjects captured the markers using the UAV camera in numerical order. We placed the markers on the pole in a counterclockwise fashion at heights of 80-230 cm. When using normal mapping FlyingHead ($G = 1$), the subjects combined body and device control to set the altitude. Figure 3.5(b) shows the image from the inner camera of the UAV. The detection area of the markers is framed in the red square. Figure 3.5(c) shows detection of the marker. The marker is framed by the green square when captured by the operator. Each subject performed three experiment sessions. We preliminarily decided markers positions, which were different in each session. However, the subjects did not know these positions before each session.

We divided group A and B to reduce the number of subjects needed to counterbalance all conditions tested in the study. Group A compared normal mapping FlyingHead with traditional joystick controlling. Group B also compared linear

mapping control mechanisms which are head-couple and hand-synchronization. We instructed each method to subjects for 10 minutes. In user study 1, we adopted Sony HMZ-T2 for visual feedback.

3.5.2 Group A

In Group A, we used FlyingHead ($G = 1$), and joystick controlling. The purpose of this study is to point out which method, controlling with body movement or with the device, is more suitable for searching and capturing tasks. In Group A, subjects were six people between the ages of 23 and 25 and heights of 161-175 cm.

A joystick has one stick and various buttons. The subjects used the joystick which provides positional controlling that the subjects manipulate the UAV's position through four parameters ($pitch, roll, yaw, and throttle$). For the joystick control, the subjects wore an HMD for visual feedback (Figure 3.5(d)).

Figure 3.6 shows to compare the average completion time of every subject for all three sessions. FlyingHead ($G = 1$) produced the fastest times for all three sessions. The average completion time for the three sessions was 40.8 s with FlyingHead and 80.1 s with the joystick method. We conducted a paired ANOVA from the average of each subject, which gave us a $p - value < .01$.

Figure 3.7 shows the UAV trajectory during each session plotted in a 3D-point diagram. For the joystick method, the UAV frequently moved in a rectilinear trajectory. These results suggest that it was difficult to set the parallel control parameters each time with the joystick control. The trajectory of FlyingHead suggests that this mechanism can easily control parallel parameters. In particular, the trajectory during the third session showed a smooth movement.

3.5.3 Group B

In Group B, we compared head-synchronization and hand-synchronization methods when the mapping scale was changed: FlyingHead ($G = 2$), and the hand-synchronization mechanism ($G = 2.5$). The purpose of this study is to reveal if the operator's head is better way to map the trajectory of the UAV or if the operator's hand is more suitable when controlling in the master-slave method that synchro-

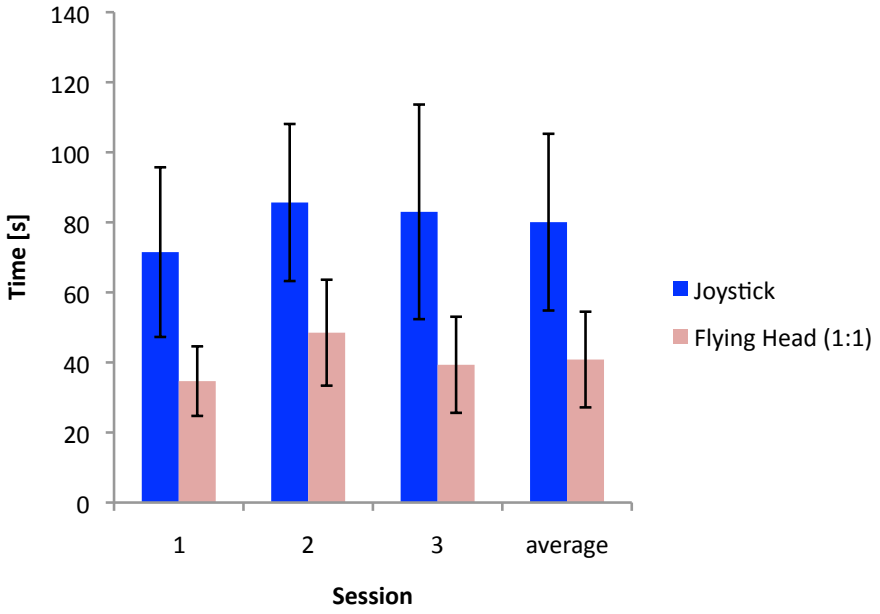

Figure 3.6: The result of Group A: Comparison of the average time required for each subject during three sessions, where the shorter time is the better. FlyingHead ($G = 1$) was faster than the joystick for every session. The average completion time for the three sessions was 40.8 s with FlyingHead ($G = 1$) and 80.1 s with the joystick method. Black lines show standard deviation.

nizes the UAV's movement with the one of operator. In Group B, subjects were six people between the ages of 23 and 30 and heights of 154-180 cm.

In hand synchronization, we set the mapping ration as $G = 2.5$ to reduce arm movements of the subjects. This ratio can be changed depending on areas where UAVs are deployed. In the hand-synchronization, the operator determines the UAV's position and direction by controlling a small dummy of the UAV with his/her hand (Figure 3.5(e)). In this experiment, the UAV moved 2.5 times more than the movement of the dummy UAV. The user sees the view from the display.

Figure 3.8 shows the average completion time of every subjects for all three sessions. FlyingHead ($G = 2$) produced the fastest times for all three sessions. The average completion time for the three sessions was 53.1 s with FlyingHead

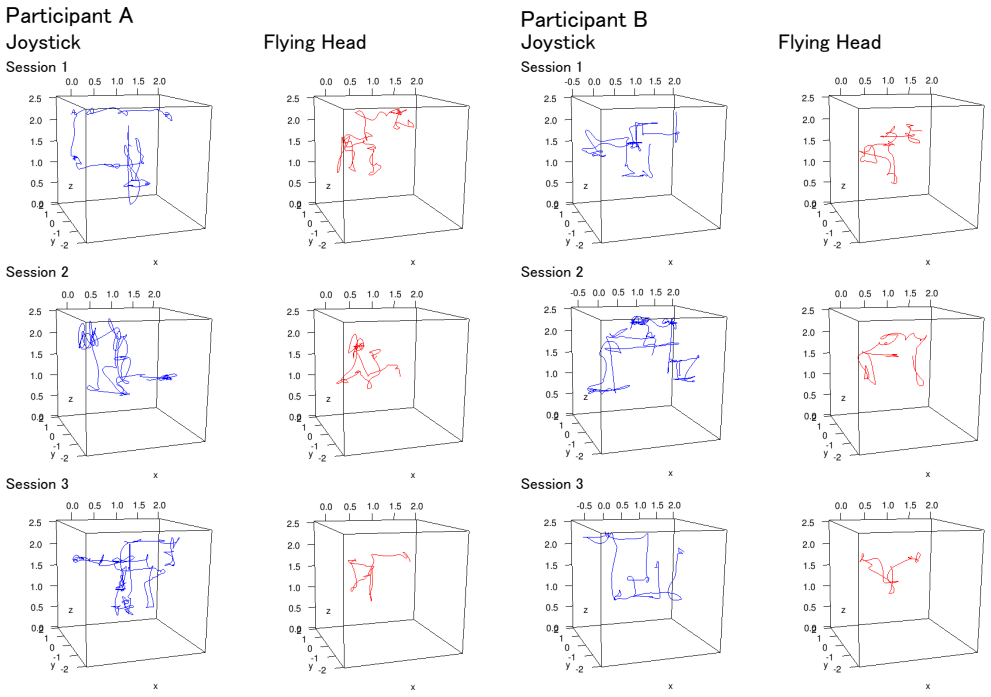

Figure 3.7: Trajectory of the UAV in Group A: where each line is the migration path of the UAV. The red line is trajectory with FlyingHead, and the blue line is trajectory with the joystick.

($G = 2$) and 99.1 s with the hand-synchronization method. We conducted a paired ANOVA from the average of each subject, which gave us a $p - value < .01$. No tracking error was found during the study.

In interview of linear mapping FlyingHead, The subjects said the control method surprised them the first time, but they managed to get used to it soon after they started. They did not feel any sickness in linear mapping FlyingHead.

The hand-synchronization mechanism received worse results than FlyingHead. The reason of this result might be that it is difficult to recognize the UAV's trajectory with the hand-synchronization mechanism and also hard to understand the position difference between the dummy and the UAV. On the other hand, with FlyingHead, as the UAV synchronizes with the head movement, it would be easier to recognize the position difference between the head and the UAV.

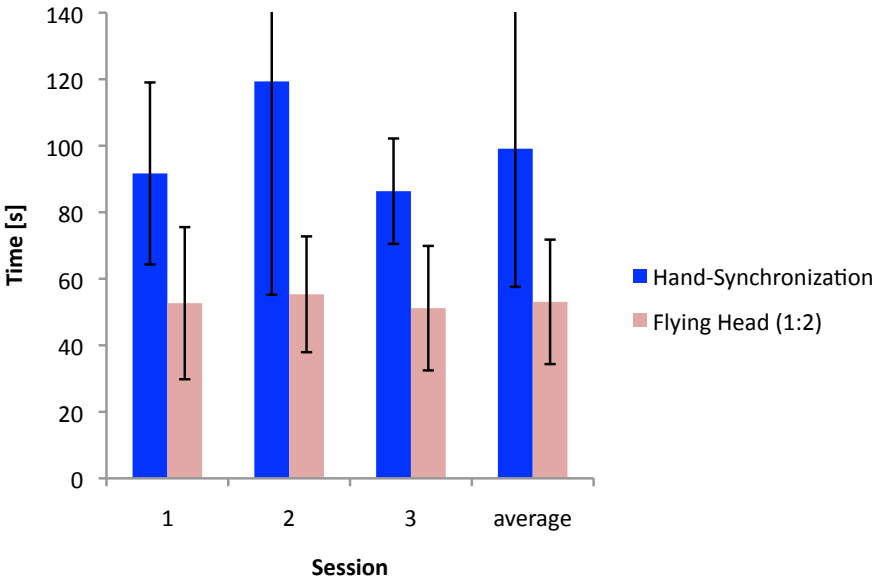

Figure 3.8: The result of Group B: Comparison of the average time required for each subject during three sessions, where the shorter time the better. FlyingHead ($G = 2$) was faster than the hand-synchronization method for every session. The average completion times for the three sessions were 53.1 s with the FlyingHead ($G = 2$) and 99.1 s with the hand-synchronization. Black lines show standard deviation.

3.5.4 Discussion of Study 1

Figure 3.9 shows a total result of the study. FlyingHead mechanisms made significant records compared with joystick and hand-synchronization mechanisms. the normal mapping FlyingHead with a device based altitude control had shorter times than linear mapping ($G = 2$). Potentially, the combined interaction can make beneficial experience in operation tasks.

3.6 User Study 2

We perform second user study to evaluate values of mapping ratio extending that the study compares normal mapping with extended mapping. We focus on user's

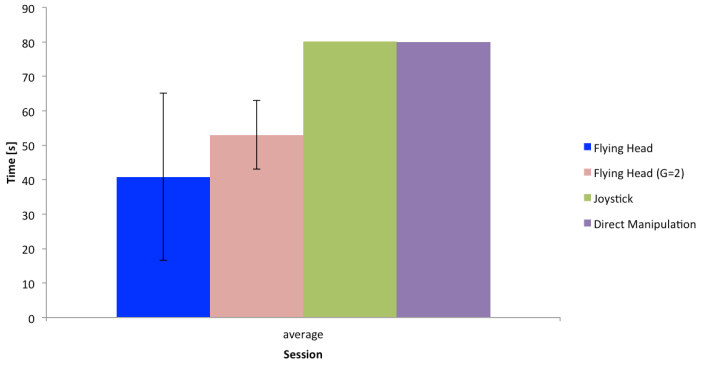

Figure 3.9: The total result of study 1. Black lines show standard deviation.

behaviors on mapping ratio changing, thus we manually change the mapping ratio during the study.

3.6.1 Task and Environment

We plan a moving task that subjects manipulate the UAV using FlyingHead of each mapping ratio. Figure 3.10 shows a task description that the UAV shuttle between the length of 2m in two times. On green points, the subjects make 180-degree turn in the opposite direction. In this study, the G_y parameter is always 1, because the study required only horizontal movements.

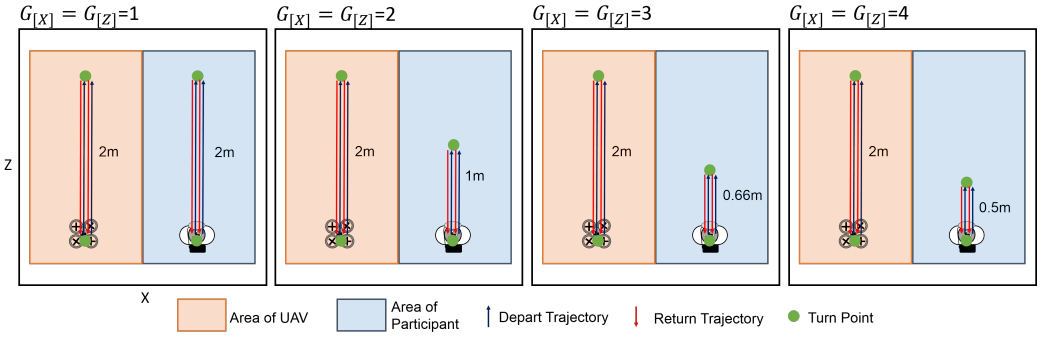

Figure 3.10: Environment of study 2. The subject performed a manipulation task to move the UAV with four gain parameters

When the subjects is close to the green points than 0.25cm, the system shows

a message (”OK!”) on the HMD. The subjects hold a pushing button. If the subjects push the button during showing the message, the system proceeds a next green point. The system also shows a navigation circle on the HMD to indicate an estimated distance between the UAV and the target green point (Figure 3.11 B C).

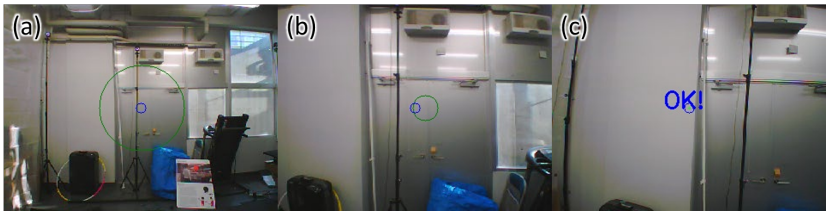

Figure 3.11: Feedback images in the user study 2, (A) and (B) visual feedback with navigation circles, (c) message to push a button.

We measure completion times of the moving task in each mapping ratio. We also recruit six subjects that a half of the subjects has an increasing sequence of the mapping ratio (1, 2, 3, 4), and other half has a decreasing sequence (4, 3, 2, 1). In user study 2, we adopted Oculus Rift DK1 for visual feedback.

3.6.2 Results

Figure 3.12 shows a result of the study includes $G = 1$: 51.8s, $G = 2$: 35.7s, $G = 3$: 35.6s, and $G = 4$: 43.7s in each mapping ratio (average times). We perform the Kruskal-Wallis test which indicates that the average times of $G = 2$ and $G = 3$ are significantly shorter than $G = 1$.

We also perform questionnaire and interview to the subjects. They reply scores (1:disagree to 5:agree) in each mapping ratio. There are five questions as follows.

Q1: Was an operation easy?

Q2: Is your body size lager than usual?

Q3: Did you feel latency?

Q4: Is your sense of visual moving large?

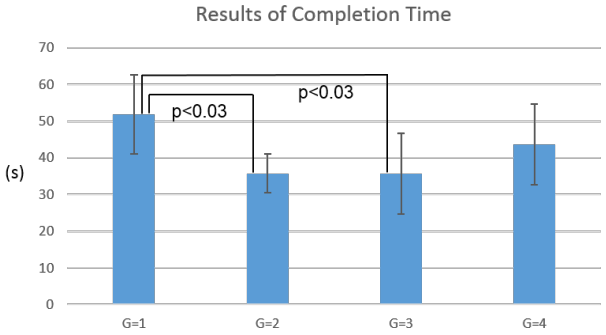

Figure 3.12: Result of the user study 2. Black lines show standard deviation.

Q5: Did you feel a gap between physical and visual moving distance?

Figure 3.13 shows the result of the questionnaire. The result of Q1 means that over 60% subjects mention easy operation in $G = 1, 2, 3$. The result of Q3 indicates that over 60 subjects did not feel latency during the study without $G = 1$. In interview, we ask two questions to the subjects. First question is "what is the most easiest gain for FlyingHead operation?", and second question is "what is the limit gain for FlyingHead operation using your body sensation?".

Table 3.1 shows the result of two questions. four subjects reply that operation with extended mapping ratio ($G \geq 2$) is easier than normal mapping ($G = 1$). All subjects also reply that the limit gain is over than 1.

3.6.3 Discussion of Study 2

In this study, completion times of $G = 2$ and $G = 3$ ware significantly shorter than $G = 1$, and the subjects replied that Gain 2 and 3 are similar operability

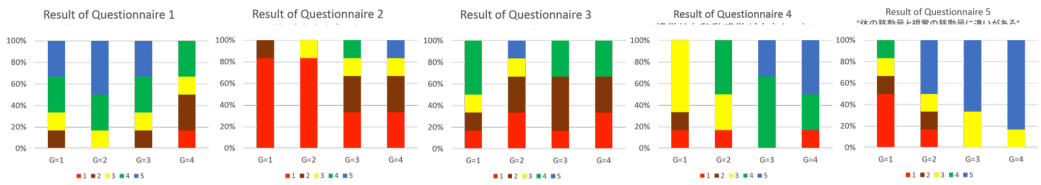

Figure 3.13: Questionnaire Result

Table 3.1: Two question on the interview

subject	what is the most easiest gain for FlyingHead operation? (1, 2, 3, 4)	what is the limit gain for FlyingHead operation using your body sensation? (≥ 1)
A	3	4
B	1	2
C	2	3
D	3	5
E	1	3
F	3	3

with $G = 1$. In the $G = 1$, the system required large movements to the subjects to navigate the UAV. This is an explanation of why the mapping ratio of the $G = 1$ has slower completion time. In $G = 2$ and $G = 3$, the subjects could navigate the UAV to the green points with small body movements. However, in the $G = 4$, a subject's small head movement was heavily reflected on the UAV movement. Figure 3.14 shows trajectories of the subject A which include head (H_i), mapped (M_i), and UAV (U_i) points. a trajectory of gain 4 overshot the green points.

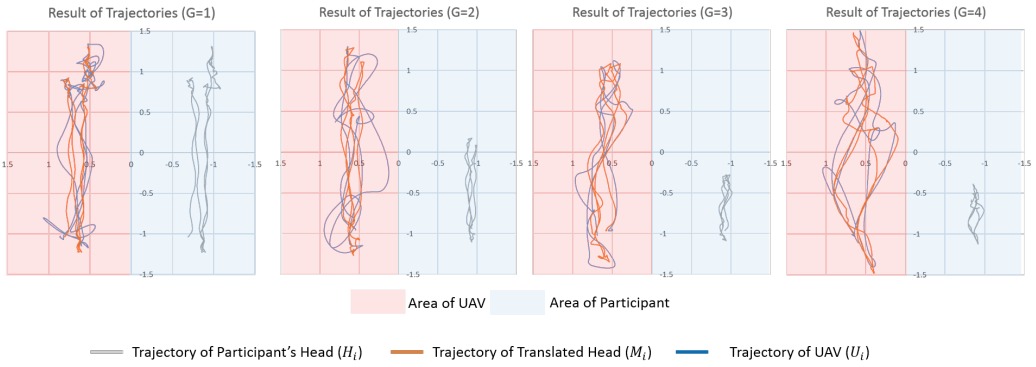

Figure 3.14: Trajectories in the user study 2

3.7 Discussion

In this section, we discuss some plans for future research and applications of flying telepresence

3.7.1 Limitations

In an outdoor environment, FlyingHead cannot use optical motion capture to locate the UAV owing to sunlight or disturbances in the air. We intend to develop a new localization system for outdoor use that will possibly involve the use of GPS, Wi-Fi, or ultra wide-band technology. Due to its accuracy, we feel that the use of an Ubisense ultra wide-band system as a real-time locator may be a valid approach. On the other hand, UAVs can estimate position and orientation using sensor devices such as, GPS, gyro, acceleration sensor, and visual odometry. Although this method is low accuracy, operators may able to easily manipulate UAV because they can be aware of self head trajectory.

3.7.2 Combination with other control methods

In this study, the UAV only flew within ranges commensurate with the distances walked by their operators. However, in some telepresence exercises, the operator and the robot will not move at equal scales, in which case the system should be able to perform distance scaling. For instance, if the operational range of the robot is three times that of the operator, a distance of 1 m walked by the operator would be mapped to a UAV movement of 3 m. We plan to expand FlyingHead system to include such scalability and to measure its usability as well as combine and creatively use additional manipulation methods.

3.7.3 Future Flying Telepresence Applications

Inspection

With this mechanism, the operator can control a UAV as if he or she was a flying robot, this is useful for remote operations such as inspections. A UAV is better able to get into areas inaccessible to people than ground robots. By setting a small

UAV in certain facilities, we can always connect to it for inspections or in the event of an emergency. Tiny helicopters (around 15 cm wide) can currently be purchased at low prices. We can consider scenarios of putting this kind of helicopter in every room of a facility.

Remote Collaboration

Figure 3.15 shows an application that provides instructions to a remote operator using a laser pointer mounted to the UAV as an example of flying telepresence. This function is used by a specialist to provide instructions to a non-specialist situated in a remote location. For example, people in a disaster-affected area may receive instructions to manipulate a certain device from a specialist with the assistance of pointing. In many cases, audio instructions are inadequate to provide instructions at the remote location. Therefore, visual instructions such as pointing are required to increase communication efficiency. For tasks in large indoor areas or those involving the manipulation of large devices, UAVs need to move and provide instructions simultaneously. FlyingHead can realize these tasks because the UAV has hands-free control, and the operator can thus simultaneously point to provide visual instructions.

Teleoperation

Flying telepresence can also be used to facilitate remote operations. For example, UAVs with manipulation equipment can be employed in tasks such as disaster relief or high-altitude construction. However, current UAVs lack free manipulation equipment comparable to the hands of a human operator. NASA has developed Robonaut, which is a telepresence robot for exterior work in outer space [65]. Robonaut has two arms that are synchronized to the operator's hand motions. Lindsay et al. demonstrated the construction of a cubic structure using mini-UAVs with a crane [66]. Figure 3.16 shows a potential future two-armed flying telepresence robot that can be used for teleoperation. The operator can manipulate this flying telepresence robot's hands as if they were his or her own by using motion capture.

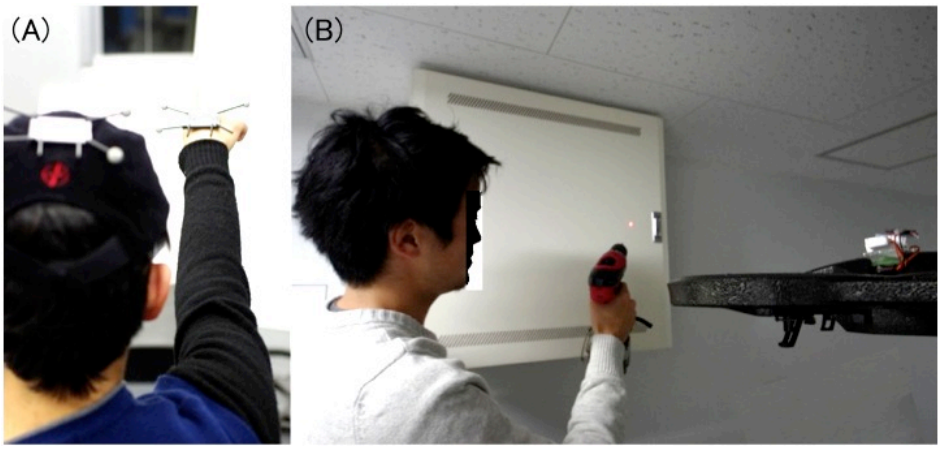

Figure 3.15: Example: a specialist provides instructions to a non-specialist situated in a remote location. (A) The specialist points with fingers. (B) an remote operator gets assistance via a flying telepresence robot

Capturing platform

The VR system can set the location and orientation as a virtual camera using instinctive devices. Ware et al. proposed the hand manipulation of a virtual camera [27]. We believe that FlyingHead can be used to manipulate physical camera systems such as digital movie cameras for motion pictures and game creation for shooting high-realistic movies. FlyingHead can be used in future video content creation systems in which a camera operator would capture the action through the highly effective employment of positioning and orientation. Laviola proposed hands-free camera navigation, which introduce user's head movements in virtual reality environments [67]. We plan to introduce this technique to move wide length fields with FlyingHead.

Entertainment Platform

Flying telepresence may also provide an out-of-body experience or the sensation of leaving one's own body. When we demonstrated a FlyingHead prototype to a large audience (more than 300 people), many subjects noted the novelty of the experience of seeing themselves from outside their bodies. This reflects the abil-

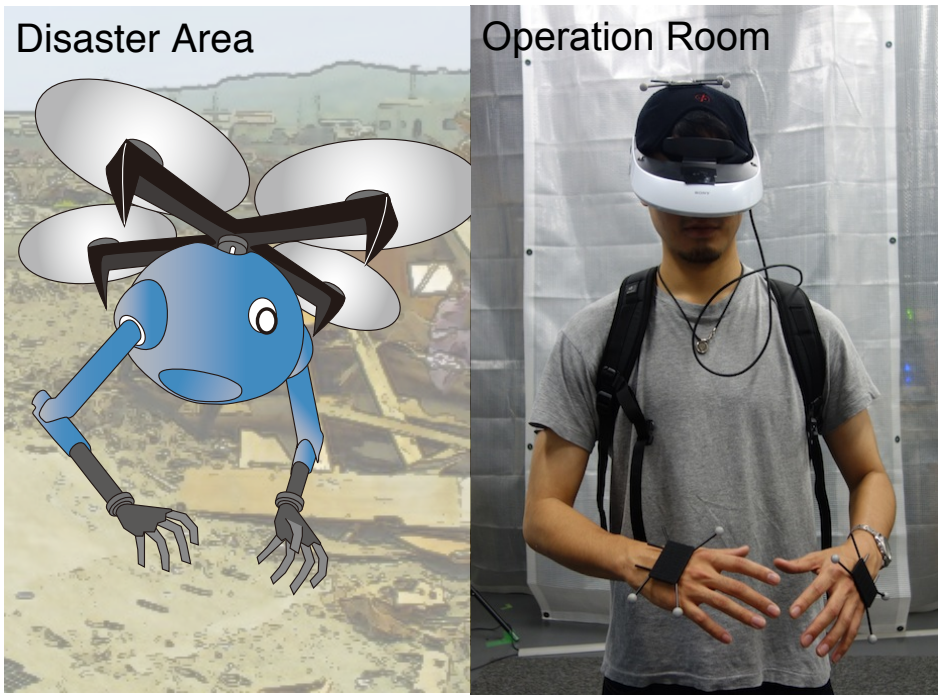

Figure 3.16: An Example of Future flying telepresence robot: The UAV's two-arms are synchronized with operator's hands

ity of flying telepresence operators to observe themselves through UAV cameras. By changing the mapping ratio of the movement, a user can experience an augmented ability. This can be regarded as a new experience and has potential for an entertainment platform.

3.8 Conclusion of this Chapter

Flying telepresence is a term used for the remote operation of a flying surrogate robot so that the operator's "self" seemingly takes control. In this chapter, we propose a control mechanism termed FlyingHead that synchronizes the motions of a human head and a UAV. The operator can manipulate the UAV more intuitively since such manipulations are more in accord with his or her kinesthetic imagery. The results of first study showed that FlyingHead mechanisms recorded better results than joystick and hand-synchronization mechanisms. The results of

second study also shows that its subjects accepted linear mapping of movements on FlyingHead, although an over gain parameter ($G = 4$) recorded worse result than $G = 2$ and $G = 3$. Finally, We discussed additional flying telepresence applications such as capturing platforms, teleoperation, and entertainment platform.

Chapter 4

ImmerseBoard: Immersive Telepresence Experience using a Digital Whiteboard

4.1 Introduction of this chapter

A physical whiteboard can enhance collaboration between people in the same location by allowing them to share their ideas in written form. The existence of the written representations in turn allows the participants to express their relationships to the ideas in physical terms, through pointing, gaze direction, and other forms of gesture. These are important ways, besides the written information itself, that a physical whiteboard enhances collaboration beyond the usual important elements of collaboration between co-located people, such as eye contact, body posture, and proxemics.

When collaborators are remote, a digital whiteboard makes it possible for remote collaborators to share their ideas graphically. Digital whiteboard sharing is a facility found in many modern video conferencing systems. However, it is mostly used to convey information through writing. The ability for the participants to relate with each other and with the writing through pointing, gaze, and other forms of gesture is often lost. Preserving such context, as if the participants were co-located, has been a goal of research on remote collaboration for some time [68][69]. The well-known Clearboard [69] deeply affected the remote collaboration field. It shows the video of the remote participant on a shared workspace as if the participants talk through and draw on a transparent glass window. However, Clearboard

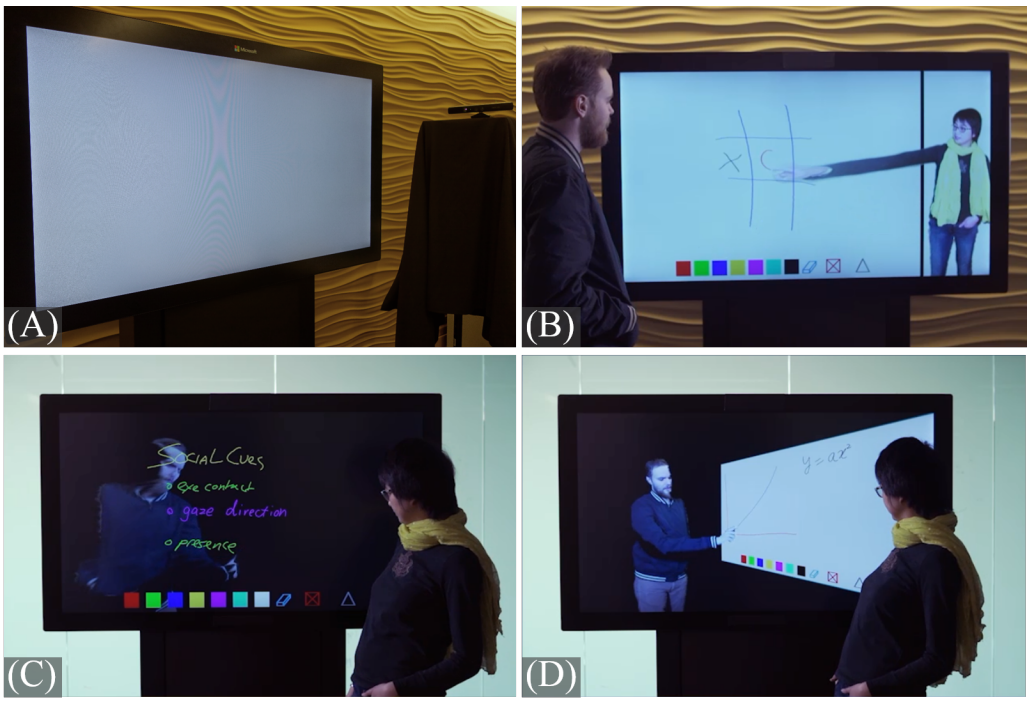

Figure 4.1: ImmerseBoard setup and conditions. (A) Large touch display and a Kinect camera, (B) Hybrid, (C) Mirror, (D) Tilt board.

has several limitations: (a) it requires a rear projector/camera and special screen (either liquid crystal screen switched between transparent and scattering states, or 45 degree tilted projection screen with a polarizing film and half-silvered mirror), whose bulk and cost make large deployment difficult, (b) gaze is correct only when both participants' heads are simultaneously located at the virtual camera positions, (c) collaborating through a glass window is not as familar for users as collaborating in front of a whiteboard, and requires an unexplained image flip, and (d) the writing and remote participant's video are overlapped, which may distract participants. The metaphor that participants talk in front of a whiteboard was discussed in the Clearboard paper and was considered too difficult to implement without using head-mounted displays and special gloves.

In this chapter, we propose ImmerseBoard, which provides an immersive telepresence experience [70] around remote whiteboard collaboration with a simple setup,

using only a large touch display and an RGBD camera. ImmerseBoard preserves the remote participants' physical relation to the whiteboard and to each other, while overcoming ClearBoard's limitations. We design and implement a prototype system, which supports three novel immersive conditions, called Hybrid, Mirror, and Tilt board conditions, shown respectively in Figures 4.1(B)–(D). The Hybrid condition is an augmentation of 2D video conferencing with a whiteboard, extending the remote person's hand out of the video window to reach the location where he or she is writing. The Mirror condition emulates side-by-side collaboration while writing on physical mirror. Though visually similar to ClearBoard, the Mirror explains the flip and extends easily to multiple parties. The Tilt board condition emulates side-by-side collaboration while writing on a physical whiteboard. This has not been possible before without a head-mounted display. Key contributions include the following:

- We use *only an RGBD camera (Microsoft Kinect), mounted on the side of a large touch display*, to enable 3D immersive collaboration in a desirable form factor, practical for home or office use.
- We introduce three new visualization metaphors, including the completely new 2.5D Hybrid and 3D Tilt visualizations, which provide to remote collaborators a sense of the spatial relationships to each other and to their shared writing, thereby preserving varying degrees of gaze direction, gesture direction, intention, and proximity, for immersive collaboration.
- We implement all three visualizations (Hybrid, Tilt, Mirror) in a single system, and allow users to choose their preferred visualization depending on task. To our knowledge, this is the first implementation of any of these visualizations in a simple and practical setup.

We also run a user study to validate the system. In the user study, we design three games that reflect important aspects of real-world four. A total of 32 participants in 16 pairs play these games on ImmerseBoard. The results show that compared to standard video conferencing with a digital whiteboard, ImmerseBoard provides participants with a quantitatively better ability to estimate their remote partners'

eye gaze direction, gesture direction, intention, and level of agreement. Moreover, the participants have a heightened sense of being together and a more enjoyable experience.

4.2 Related Work

ImmerseBoard draws from several fields, including computer supported cooperative work (CSCW) and telepresence. In this section, we explain how ImmerseBoard is related to prior work in these fields.

4.2.1 Large Screen Collaboration

Large displays, and large touch screens in particular, have been used to support collaboration between many people in the same place. Streitz et al. proposed a collaboration workspace, including the DynaWall, which can be jointly operated by two people [71]. Khan et al. proposed a method for showing attention on a large display using a spotlight [72]. Birnholtz et al. evaluated the effectiveness of large screens in negotiation [73]. Other collaboration researchers aimed to enhance co-located collaboration using digital whiteboard systems that use the pen's buttons [74], and handheld-computers [75] . In contrast, we focus on remote collaboration, through a digital whiteboard. Specifically, we focus on generating a sense of presence between the remote participants as well as a seamless collaboration environment.

4.2.2 Remote Collaboration Systems

CSCW researchers aim to realize remote collaboration with an experience similar to local collaboration. Tang and Minneman proposed VideoWhiteboard, which is a remote collaboration system that shows the remote participant's shadow [68]. Apperley et al. also developed a collaboration system that shows shadow information on a large display [76]. However, shadowed facial information does not preserve eye contact. Ishii et al. introduced Clearboard, which shows the video of the remote participant on a shared workspace, as if the participants are looking at each other through a glass wall (on which they can write), approximating eye contact

[69]. Clearboard flipped the remote video to fix the inverse writing problem. Ishii's research deeply affected the remote collaboration field [77]. Roussel designed THE WELL, which introduced the looking down display model [78]. These works all used tabletop computers to show shadow [79] or photographic hands [80] [81] for user's attention. In contrast, ImmerseBoard system extracts a 3D representation of the remote participant in order to reconstruct a more informative representation.

4.2.3 Immersive Human Reconstruction

Raskar et al. introduced immersive telepresence for remote collaboration in an office environment [82]. Various other works on immersive telepresence also involved reconstruction of human images in 2D/3D environments, including 3D human images from stereo or depth cameras [83] [84]. Zhang et al. made realistic human 3D images in real time using a hybrid camera system, consisting of a depth camera, IR cameras, color cameras, and IR laser projectors [85]. Morikawa et al. proposed Hyper Mirror, which mixed images from two places using background subtraction [86]. Several researchers displayed reconstructed humans using tetrahedral displays [87], omni-projection [88], and face-shaped displays [89]. In contrast, ImmerseBoard reconstructs the remote participant as a life-sized human body on a whiteboard in real-time using an RGBD camera for immersive collaboration.

4.2.4 Immersive Telepresence with a Whiteboard

Some prior work in immersive telepresence employs whiteboard collaboration [90]. Kunz et al., in Collaboard, extracted the remote participant from video and used background subtraction for showing attention [91]. Uchihashi et al. proposed a system for mixing remote locations using stereo cameras that can show the touch position of the remote person [92]. Junuzovic et al. created a shared work space on any surface using a camera-projection system [93]. However their IllumiShare system loses eye contact and face-to-face communication because the camera position is behind the user. For eye contact through video, the camera and the display should be at or close to the same position. In our work, we use 3D capture in order to solve the problem of displaying the remote person from the correct point

of view, which solves the eye contact problem if the visual quality is sufficiently high.

Zillner et al. proposed 3D-Board, which can capture and transmit a user's whole body for remote whiteboard collaboration using multi-kinects [94]. Their solution is similar to the Mirror condition in this chapter. However, we provide two additional novel visualizations - Hybrid and Tilt, which are valuable alternatives to Mirror. Our study finds that it is important to provide different conditions for participants to choose from, as their preferences are diverse and task dependent. In addition, there are several differences between our Mirror condition and 3D-Board: (a) 3D-Board is asymmetric. The operator can see the instructor's image, but the operator's image is not sent to the instructor. In contrast, ImmerseBoard provides a symmetric collaboration experience where participants can see each other. (b) 3D-board needs a Kinect away from the board to track the operator's head for motion parallax, while ImmerseBoard uses the Kinect on the side of the display to perform head tracking. (c) 3D-Board uses two Kinects to reconstruct the remote user with a better image quality than the Mirror condition in the ImmerseBoard, which we have left to future work.

4.2.5 Remaining Challenges

There are two major issues in the existing works. The first issue is the tradeoff between the form factor of the system and the level of immersion that it can provide. It is challenging to provide an immersive telepresence experience for whiteboard collaboration in a form factor simple enough for practical installation in homes and offices. The second issue is the difficulty of implementing the whiteboard metaphor (collaborating side by side on a whiteboard), and furthermore providing the glass wall metaphor for users to choose from within the same system. Therefore, we design and implement ImmerseBoard to address these two challenges.

4.3 Two Guiding Metaphors

ImmerseBoard aims to connect remote collaborators as if they were co-located. Two metaphors of physical collaboration guide the design of ImmerseBoard. The

first is the metaphor of side-by-side writing on a physical whiteboard, as shown in Figure 4.2(A). Each participant views the whiteboard and the other participant from his or her personal view, seeing the whiteboard in perspective and seeing the other participant from the side, in front of the whiteboard. The second is the metaphor of side-by-side writing on a physical mirror, as shown in Figure 4.2(B). Each participant sees the image of the other participant reflected in the mirror. The participants write on the mirror. In each metaphor, the participants are able to convey eye contact, eye gaze direction, pointing, hand gestures, body proximity, and other aspects of body language in relation to the other participant as well as to the writing. In addition, there is shared space in front of the writing surface for physical interaction and manipulation. The whiteboard metaphor is discussed in the Clearboard paper, but is considered hard to implement without using head mounted display. The mirror metaphor is similar to the glass wall metaphor in ClearBoard. In this chapter, we implement both metaphors with a much simpler setup, i.e., just setting a Kinect on the side of large touch display (see next section), allowing users to choose their prefered metaphor.

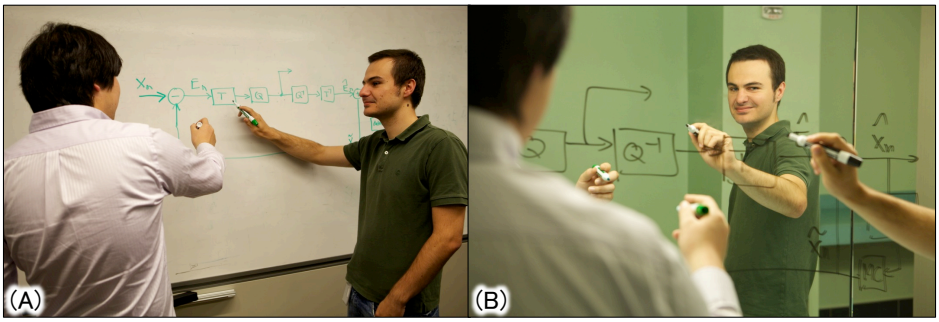

Figure 4.2: Metaphors: Side-by-side writing (A) on a whiteboard, (B) on a mirror.

4.4 System

To emulate the above metaphors, we built ImmerseBoard around a large touch screen and a color plus depth (RGBD) camera, as shown in Figure 4.1(A). In our prototypes, the touch screen is a 55-inch Microsoft Perceptive Pixel (PPI) display, and the camera is a Microsoft Kinect camera. The PPI board is a multi-touch

screen that can be used with either pens or fingers. We built two prototypes, called *Left* and *Right*, respectively configured with the PPI board to the left and right of the Kinect camera. (See Figure 4.3.) In this setup, users can move freely in the capture range of the Kinect camera (0.4-4.5 meters, 70° FOV), which allows them to roam up to 2 meters away from the board at its center. The remote user is rendered on the display close to the Kinect, so that the local user naturally stays within the capture range of and faces the Kinect in order to look at the image of the remote user and to write on the board. In the event that the local user faces the board directly, there may be some minor but not critical occlusions. One hand may be occluded by the torso when the hand is not active, but will be observable when the hand is writing or pointing.

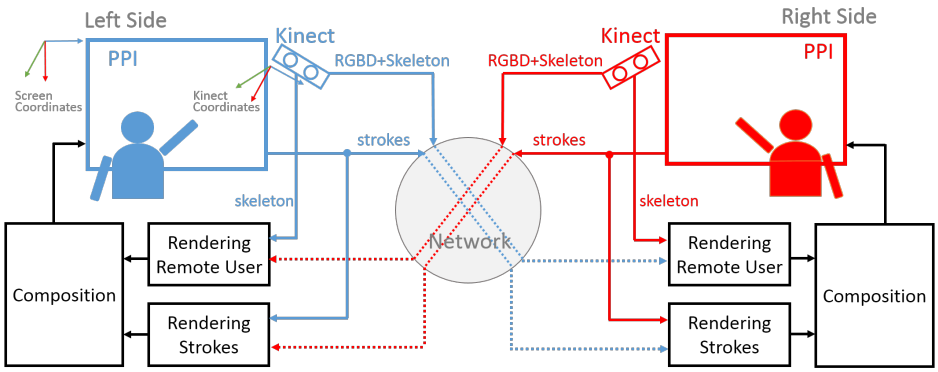

Figure 4.3: Left and right ImmerseBoard prototypes.

The ImmerseBoards transmit to each other stroke data (position and color), color video data, depth video data, and skeleton data. The color and depth data allows us to extract an image and 3D point cloud of the participant without the background, while the skeleton data allows us to track the positions of the limbs of the participant. The depth data and skeleton data are expressed in the coordinate system of the capturing camera. In order to understand the pose of the participant in relation to the board, we transform the data from the camera's coordinate system into the board's coordinate system. This requires prior calibration of the pose of the camera with respect to the board.

We implemented a simple calibration system, which allows a user to tap four

points in the corners of the PPI. When the user taps a point, the system records his 3D hand position from the skeleton information. From these four 3D positions, the system calculates a transformation matrix relating the coordinate systems of the camera and the board.

Once the data are transformed into the board's coordinate system, it can be processed and rendered with different visualizations (to be described in the next section). We use C++ and OpenGL for 2D/3D video processing and rendering, and use TCP for data communication.n.

4.5 ImmerseBoard Conditions

ImmerseBoard supports several visualizations, or *conditions*. The first condition emulates the metaphor of participants writing shoulder-to-shoulder on a physical whiteboard. The second condition emulates the metaphor of the participants writing shoulder-to-shoulder on a mirror. Both of these conditions use 3D capture and rendering of the remote participants. The third condition is a hybrid between a standard 2D video conference and a 3D writing experience. A fourth condition is simply a standard 2D video conference with standard digital whiteboard. We now explain (in reverse order) the conditions and their implementations.

4.5.1 Video Condition

We begin with a standard video condition, in which the left or right side of the display is reserved for standard 2D video, leaving the bulk of the screen as a shared writing surface. The video is captured by the color camera in the Kinect, and displayed on the same side of the PPI as the camera, so that the eye gaze discrepancy is about 15 degrees. The display is large enough to show the upper body of the remote participant, life-sized. The video is processed so that the background is removed and the participant is framed properly regardless of where he is standing.

4.5.2 Hybrid Condition

The Hybrid condition is a hybrid of the above Video condition and a 3D experience. In the Hybrid condition, the remote participant's hand is able to reach out of the video window to gesture, point, or touch the board when writing, as shown in Figure 4.1(B). From the remote participant's hand position, the local participant is often able to understand the remote participant's intention as well as his attention.

ImmerseBoard implements the Hybrid condition using 3D depth and skeleton information from Kinect to guide 2D color video processing, as shown in Figure 4.4. The Kinect determines foreground (person) and background pixels. Each foreground pixel has a 3D coordinate. ImmerseBoard uses these 3D coordinates to segment pixels into body parts according to the pixels' 3D proximity to bones in the skeleton. The foreground pixels are framed within the video window of the display such that the upper body pixels are displayed. (This is the same as in the Video condition.) When the reaching arm is close to the PPI board, the system redraw arm and hand pixels by (a) moving the hand pixels to the appropriate location (orthogonal projection of the hand on the PPI board), and (b) stretching the image of the arm to seamlessly connect the upper body to the hand using texture mapping and deformation. The hand and upper body images are not stretched. Aside from the stretched arm, the foreground image is identical to that coming from the color camera. Thus image quality and eye gaze discrepancy are the same as in the Video condition.

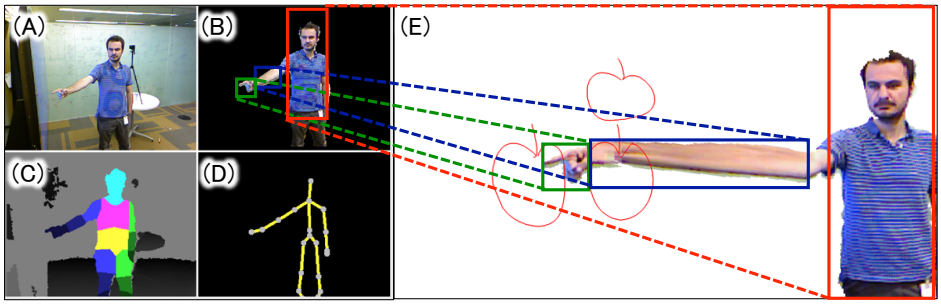

Figure 4.4: Video processing in Hybrid condition: (A) Source RGB image, (B) Extracted human image, (C) Segmentation, (D) Skeleton, (E) Result.

Aside from the stretched arm, the foreground image is identical to that coming from the color camera. Thus image quality and eye gaze discrepancy are the same as in the Video condition.

4.5.3 Mirror Condition

The Mirror condition, shown in Figure 4.1(C), is an emulation of the mirror metaphor. The remote participant's full upper body is seen life-sized, conveying body posture, body proximity, gesture direction, pointing direction, and eye gaze direction, in relation both to the board and to the local participant. Both participants are able to write on the entire surface, and see each other in any part of the surface, as if it were a large mirror.

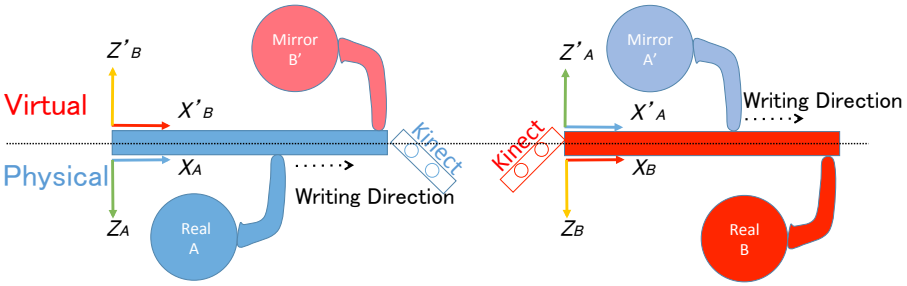

Figure 4.5: Mirror Condition: The system flips the z-axis in both sides

ImmerseBoard implements the Mirror condition by transforming the 3D colored point cloud from the Kinect coordinate system to the PPI coordinate system, and then flipping the z-axis (z to $-z$). The remote participant's point cloud is rendered using a 3D polygonal mesh. The viewpoint from which the remote participant is rendered onto the display can either be fixed at a default position, or for maximum accuracy, can track the head of the observer.

When head tracking is used at both sides, the relative geometry between the participants is precise, and eye contact is possible if the video quality is sufficiently high. Moreover, head tracking allows either participant to move to look around either the figures on the board or around the remote participant, as shown in Figure 4.6. However, the side of the remote participant not seen by his Kinect camera cannot be rendered, leading to a significant loss of perceived visual quality.

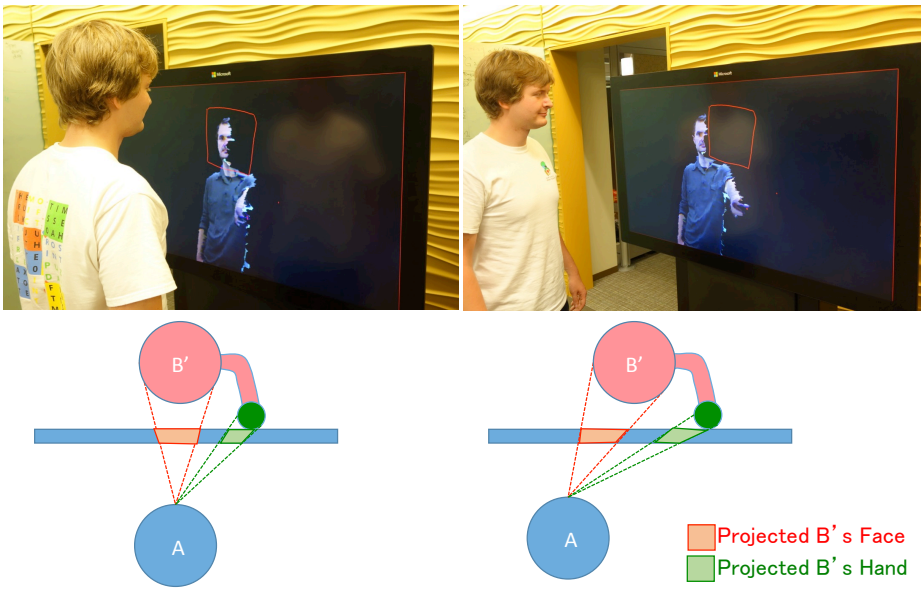

Figure 4.6: Mirror Condition with Head Tracking: The system can change perspective based on the user's head position.

Adding a second Kinect camera on the other side of the PPI board would solve the problem.

4.5.4 Tilt Board Condition

The Tilt board condition, shown in Figure 4.1(D), is an emulation of the metaphor of side-by-side writing on a physical whiteboard. As in the Mirror condition, the remote participant's full upper body is seen life-sized, conveying body posture, body proximity, gesture direction, pointing direction, and eye gaze direction, in relation both to the board and to the local participant. However, to fit the remote participant's image on the display, the image of the rectangular drawing surface is tilted back by 45 degrees (which is adjustable) and rendered in perspective. That is, the drawing surface is now virtual. Participants are able to write on the virtual drawing surface, by writing onto its projection on the physical PPI surface. At the same time, they can see each other as if they were side by side.

If writing onto the projection of a tilted virtual surface becomes awkward, optionally the tilted virtual surface can be rectified so that it coincides with the physical

surface. When the board is rectified, the remote participant is no longer visible. Thus, typically, a user will use the tilted board to watch the remote participant present, and will use the rectified board to write detailed sketches. The tilting and rectification are visualizations for the benefit of the local user only, and can be done independently on either side.

To reduce the gaze divergence between the participants, the remote participant's image should be placed as close as possible to the Kinect camera. Thus, the direction of the tilt is different for the left and right boards, as shown in Figures 4.7 (A) and (C), respectively. For the left board, the Kinect camera is located on the right, and the virtual board is tilted to the left (Figure 4.7A). For the right board, the Kinect camera is located on the left, and the virtual board is tilted to the right (Figure 4.7C). As byproduct, this increases the overlap of the remote participant seen by the local participant and captured by the remote Kinect camera, resulting higher image quality, compared to the Mirror condition.

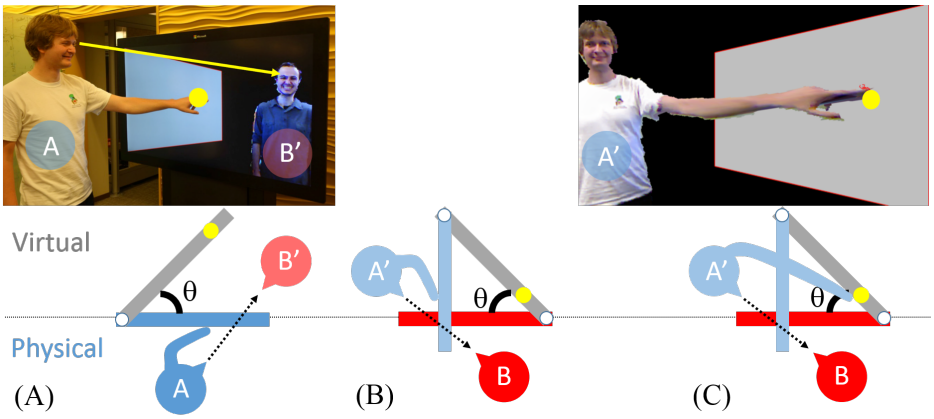

Figure 4.7: Tilt Board Condition: (A) The user touches the projection of the tilted board and looks at the remote person's face on the physical display. (B) The remote user's touch position would be incorrect if the system directly reconstructs the physical environment using the virtual board as the reference. (C) The system extends the remote participant's arm to correct the touch point.

However, when the remote participant writes on a tilted board, he is actually writing on the image of the tilted virtual surface projected onto the physical surface

of the PPI. Therefore, if the system directly reconstructs the physical environment (i.e., rotating the remote participant such that the virtual boards from both sides align) and changes only the viewpoint, the remote participant has correct eye gaze direction but points at the wrong place as shown in Figure 4.7B. Figure 4.7C shows that the correct touch point can be realized by extending the remote participant's arm to reach the correct position in the virtual environment.

To extend the remote participant's arm, the system calculates an appropriate hand position in the virtual environment. For example, if the participant is touching the physical board, this corresponds to a position on the virtual board (Figure 4.8 (A)). The hand is moved to this position in the virtual environment. However, if only the hand is moved to this position, it would be disconnected from the body (Figure 4.8 (B)). Thus, the system uses a coefficient α to interpolate the positions for points on the hand ($\alpha = 1.0$), arm ($0.0 < \alpha < 1.0$) and shoulder ($\alpha = 0.0$). The system also uses a coefficient β, based on the hand skeleton position in PPI coordinate system, to perform the interpolation only near the board. The system has two thresholds: $min(= 5cm)$ and $max(= 20cm)$. If the participant's hand is closer than min, β is 1.0. If it is further than max, β is 0.0. Otherwise, β is determined linearly ($0 < \beta < 1.0$). The system transforms each point on the hand, arm, or shoulder to a point $P_t = P_h(1 - \alpha\beta) + P_p(\alpha\beta)$, where P_h is the original point and P_p is the projected point.

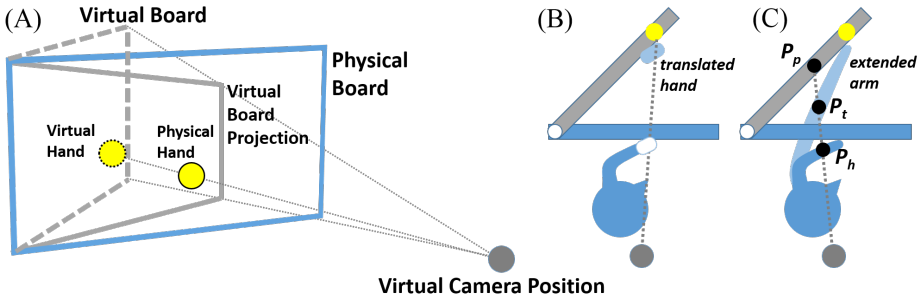

Figure 4.8: Tilt Board Geometry. (A) Projection of physical hand position on virtual board, (B) Hand translation towards virtual board, (C) Arm extension that preserves the proper hand-board relationship and arm-torso connection.

The major limitation of the Tilt board is the shape imprecision due to the perspective. It also causes fewer pixels to use on the side close to the remote user's image. Our system provides a remedy by allowing a user to rectify the board if needed.

4.5.5 Color Palette

ImmerseBoard provides a color palette with drawing colors and an eraser. Three types of color pallet menus are supported: fixed, side-slide, and pop-up (Figure 4.9). The fixed color palette is always on the bottom of the screen. The side-slide appears when the user's hand is close to the left or right side of the screen. The pop-up color palette is triggered by the non-dominant hand when it stays close to the board.

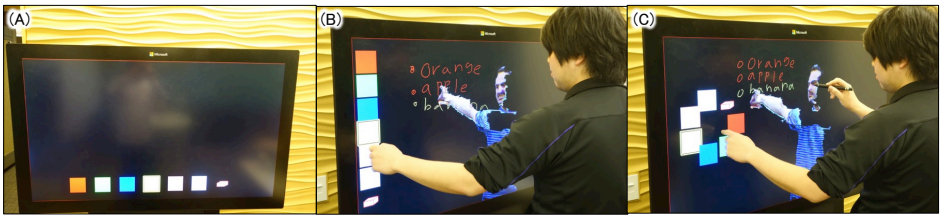

Figure 4.9: Color Pallet Menu Types: (A) Fixed, (B) Side-slide, and (C) Pop-Up

4.6 User Study

We performed a user study to compare the three conditions (video, hybrid, mirror, and tilt board), using both objective and subjective measures (the latter based on user feedback) to analyze key elements of the immersive experience such as gesture direction, intention, and eye gaze direction.

4.6.1 Participants and Studies

We recruited 32 subjects (5 female and 27 male) between the ages of 19 and 66 (mean 36). All subjects were right-handed information workers, with normal

vision, hearing and movement ability. They all had video conferencing experience as part of their work.

The participants were partitioned into two disjoint studies (1 and 2) in order to answer three questions: (a) Is reference, e.g., pointing, useful for remote collaboration? (b) Does 3D rendering provide a better experience than 2D? (c) Which 3D condition (Mirror or Tilt) is more effective? Study 1 was formed to answer the first two questions and Study 2 for the third. This partition reduces the number of subjects needed to counterbalance all conditions tested in the studies.

Study 1 had 12 subjects, working in 6 sessions. Each session had a pair of subjects to act as remote collaborators (or partners). Study 1 compared 2D visualizations (Video, Hybrid) and a 3D visualization (Mirror). The Video and Hybrid conditions have good image quality as they are generated from the RGB source, but do not have exact eye contact nor do they preserve the positional relationship between the user and the board. The Mirror condition preserves eye direction and relationship to the board, but its image quality is relatively poor, due to rendering the remote participant from a viewpoint much different from that of the camera. We distributed the six sessions evenly over the six possible sequences of three conditions.

Study 2 had 20 subjects, working in 10 sessions. The study compared Mirror and Tilt conditions as well as a variation of the Mirror with headtracking and a variation of the Tilt with optional board rotation. Half the sessions evaluated first Mirror, then Tilt. The other half evaluated first Tilt, then Mirror.

4.6.2 Setting

The study session took place in a room with two ImmerseBoards (one left prototype and one right prototype). The two ImmerseBoards were separated by a curtain, thus the two subjects could see each other only through the ImmerseBoard. Subjects could write on the board using either fingers or stylus. They could talk to each other directly. We did not capture and transmit audio, since we focused on visual experience.

4.6.3 Procedure

At the beginning of each session, the two subjects filled out background questionnaires on their prior experience with video conferencing. Then they performed one subjective task (teaching) and three objective tasks (gaze estimation, symbol matching, and negotiation) to evaluate each condition. The four tasks will be discussed in detail in the next section. Each condition was introduced at the beginning in terms of the appropriate metaphor in Figure 4.2 (whiteboard or mirror). We did not observe any difficulties understanding the metaphors. For Study 2 , we also evaluated a variation of each condition (i.e. the Mirror with head-tracking and the Tilt board with optional rotation). Participants experienced the variation immediately after its base condition and were asked to perform only the teaching task in the interest of time. After each condition or variation, the subjects filled out a brief questionnaire on the condition or its variation. At the end of each session, the subjects filled out an overall questionnaire to compare the conditions. We also debriefed each subject with an interview.

4.6.4 Task Design

We designed the four tasks to be realistic, to be fun, and to reveal the strengths and weaknesses of the different conditions on aspects important to real-world collaboration. The first task is a subjective but practical task: teaching. Teaching is a important and broadly representative use case for remote collaboration systems [95]. The remaining tasks are games with measurable objectives. The subjects are instructed to play each game to maximize (or minimize) its objective. We used the game outcomes to evaluate aspects of each condition.

Subjective Task — Teaching

As shown in Figure 4.10(A), one subject plays the role of Teacher, and the other Student. The subjects are free to decide among themselves who will be Teacher, and what the Teacher will teach. We suggested teaching the rules of a card game, board game, or sports game, but many other topics came up during the user study. The subjects had about 3-5 minutes to teach and learn, just enough time to get

some experience and understanding of the condition. In addition, the teaching task evaluates how well the condition is able to convey social cues about the level of agreement and understanding through questionnaire and interview.

Objective Game 1 — Gaze Estimation

The first game (Figure 4.10B) evaluates the accuracy with which a participant can estimate the eye gaze direction of the remote participant in each condition. It is well known that eye contact and eye gaze direction are important elements of communication [96] [89]. This is an asymmetric game with a leader and a follower, so the game is played twice in each condition, to give each subject a chance to play both roles. The players are shown an eight by eight grid of cells on their shared surface. However, on the leader's side, one of the cells is colored red, at random. The red cell is not visible on the follower's side. The leader is prompted to look at the red cell in a natural way. The follower observes his partner via the visual condition in effect, and tries to guess which cell his partner is looking at. The follower clicks on the estimated cell, and then is shown the correct answer. Before the follower clicks a button to move on to the next trial, the system blanks the follower's visual condition and shows the leader a new prompt. This gives the leader time to move to the new prompt without the follower seeing the leader's direction of movement. There are 16 trials per side. We record the estimation accuracy.

Objective Game 2 — Symbol Matching

The second game (Figure 4.11A) evaluates the ability of a participant to follow his partner's gestures as cues of attention and intention. Again, this is an asymmetric game with a leader and a follower, so the game is played twice in each condition, to give each subject a chance to play both roles. The players are shown on their shared surface ten pairs of symbols, randomly permuted on a four by five grid. The symbols come in five shapes (circle, square, diamond, up and down triangles) and two colors (red and blue). The leader taps a colored shape to make it disappear. The follower is instructed to tap the corresponding colored shape,

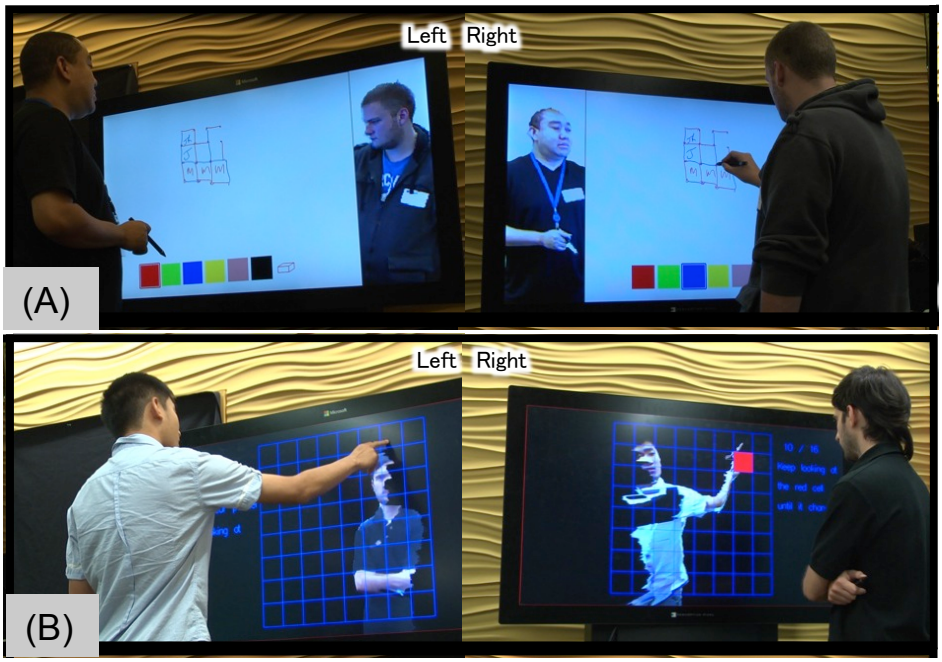

Figure 4.10: Games for User Study. (A) Teaching game in Video condition. (B) Gaze estimation game in Mirror condition. The left participant guesses where the right participant is looking.

as quickly as possible, to make it disappear. If he taps the wrong shape, nothing happens. When all shapes are gone, the game ends. We recorded the follower's response time.

Objective Game 3 — Negotiation

The final game (Figure 4.11B) evaluates how well the condition is able to convey social cues about the level of agreement and understanding, by measuring the ability for participants to negotiate and resolve conflict [97]. This game abstracts real-world scenarios where there is a limited budget, but interested parties have different, often hidden, priorities. This is a symmetric game, so it is played only once for each condition. The players are shown a four by four grid of tiles on their shared surface. The tiles have numbers on them, but they are different numbers for each player. The numbers represent the private values of the tiles to the players.

The players must agree on four of the tiles to select. The players are instructed to maximize their own total value of the tiles selected. The game is designed so that the numbers are often in conflict, i.e., a tile that is valuable to one player is likely to have little value to the other. The players are not allowed to verbalize the numbers, but otherwise are free to negotiate, for example, to say that a tile is very bad, or very good. In addition to any visuals available in the tested condition, the system shows a red circle on the board wherever it is touched. (Without this pointing aid, it was too difficult for the players to reference a tile in the Video condition.) There is a two-minute time limit on the game. If there is no agreement in the time limit, the players get no points. We recorded the scores and time to reach agreement. Before each condition, the positions of the tiles are randomly permuted.

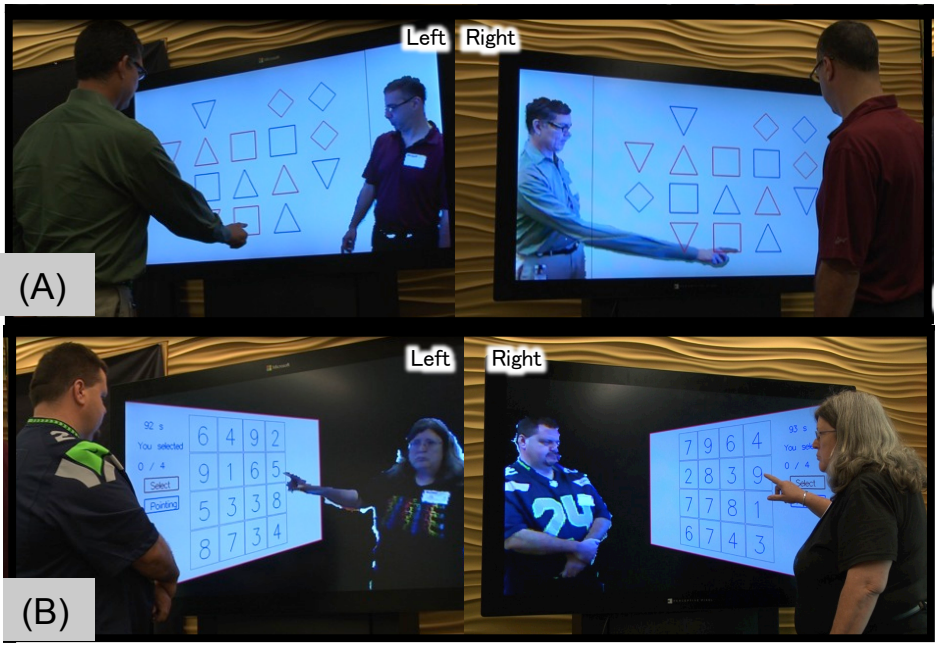

Figure 4.11: Games for User Study. (A) Symbol matching game in Hybrid condition. The right participant looks at where the left participant is about to touch. (B) Negotiation game in Tilt Board condition. The left participant looks at where the right participant is pointing.

4.7 Study Results

In this section, we will demonstrate and discuss the user study results. We will show the quantitative evaluation for the two games (gaze estimation, symbol matching) as well as the results from questionnaire and interview.

4.7.1 Result of Gaze Estimation Game

Figure 4.12 shows the average shift, or bias, from the cell that the follower clicked to the cell that the leader was looking at over four conditions in two studies. The bias is calculated per block that includes 4 cells over all subjects, since each block had only one cell selected in each game.

Figure 4.13 shows the mean and standard deviation of the horizontal and vertical errors for all four conditions. The horizontal (or vertical) error is defined as the absolute distance from the cell which the follower clicked to the cell that the leader was looking at, along the horizontal (or vertical) direction, in units of the number of cells. A Repeated Measures ANOVA revealed a significant main effect in Study 1 on both horizontal ($F_{2,22} = 9.12, p = .001$) and vertical error ($F_{2,22} = 14.75, p < .001$). Post-hoc pairwise comparison (with bonferonni corrections) revealed that (a) Video had significantly more horizontal error than Hybrid ($p = .046$) and Mirror ($p = .003$), and (b) Mirror had significantly more vertical error than Video ($p = .001$) and Hybrid ($p = .011$).

For Study 2 , the paired Student's t-test revealed that the Mirror was significantly better than Tilt on horizontal error ($p = .027$) and significantly worse on vertical error ($p = .033$).

4.7.2 Discussion of Gaze Estimation Game

In Study 1 (top row of Figure 4.12), the Video and Hybrid conditions have large horizontal bias and small vertical bias. The Mirror condition is opposite - small horizontal bias and large vertical bias. For the Video and Hybrid conditions, the horizontal and vertical biases are not equal, because the person-board relationship is preserved vertically, but not horizontally. Hence, it is difficult for the follower to estimate the horizontal gaze direction. Some subjects talked about this in the

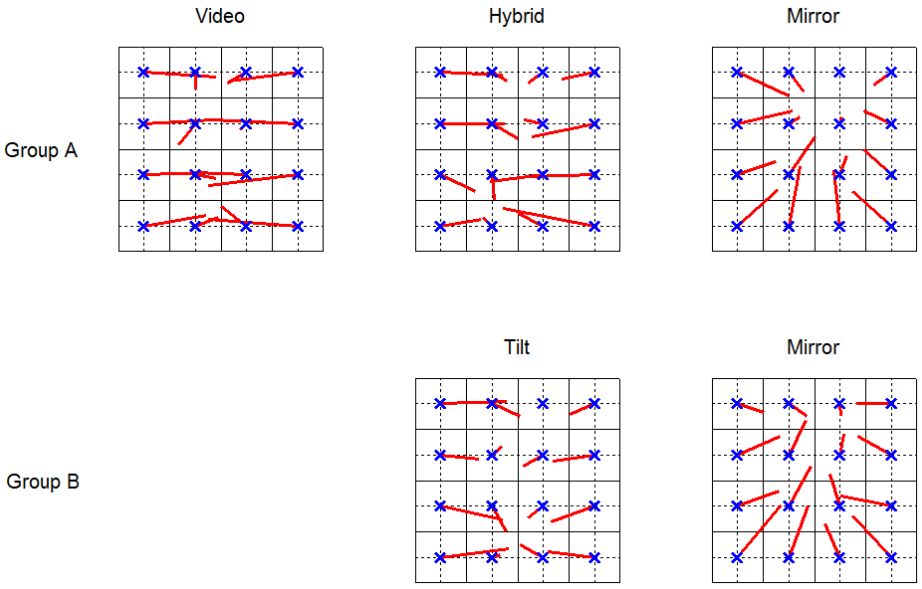

Figure 4.12: Gaze Estimation: Bias over different locations. The blue cross is the leader's true gaze direction and the red line indicates the follower's gaze estimation bias.

interview - *"It was pretty easy to tell up and down but was harder to pick up the column for the Video and Hybrid conditions."* For the Mirror condition, the person-board relationship is preserved both vertically and horizontally. As expected, the bias in each direction has similar magnitude, and the horizontal bias is less than in the Video and Hybrid conditions. However, it is surprising that the Mirror condition has significantly more vertical bias than the Video and Hybrid conditions. This is likely due to the poor video quality in the Mirror condition, especially around the eye area, such that the eye ball direction cannot be seen as well. Hence the participants rely more on the head direction to estimate the gaze direction. Thus, if the highlighted cell is directly in front of the partner (eye balls at neutral position), the bias is small. The bias increases when the highlight cell moves toward the boundaries, since the eye ball movement (other than head movement) contributes more to the eye gaze [98]. This can be confirmed in Figure 4.12. The blocks on the top-middle have small bias because that is the average head position of the remote partner. The blocks on the left/right/down have larger bias and the

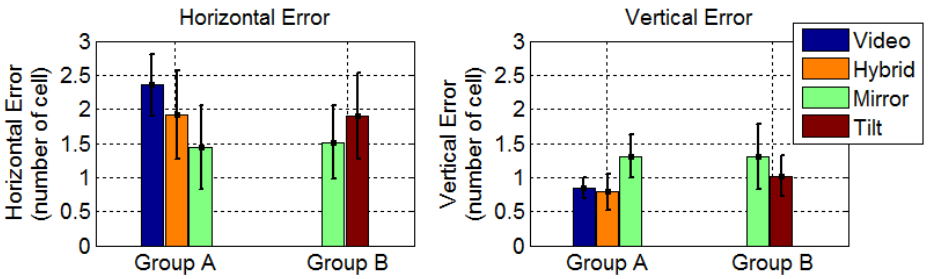

Figure 4.13: Horizontal and Vertical Error of gaze estimation game.

bias increases as the cell moves further.

Unexpectedly, the Hybrid condition is better than the Video condition on horizontal gaze bias in spite of having the same video quality. We conjecture that the leader may have learned implicitly from the reference (i.e., follower's pointing) to convey better his gaze direction to the follower.

In Study 2 (bottom row of Figure 4.12), the Mirror condition has a pattern similar to that of Study 1 . Like the Mirror condition, the Tilt board condition preserves both the horizontal and vertical person-board relationship, but from a different perspective. Since the Tilt board provides a side view, the camera position for rendering is relatively close to the Kinect camera position for capturing. Thus the Tilt board has a better video quality than the Mirror condition. In consequence, the Tilt Board has less vertical bias than the Mirror condition. However, the horizontal bias for the Tilt Board is more than for the Mirror. This is because it is more difficult to estimate horizontal eye gaze direction from a side view than from a frontal view, which the Mirror condition has.

4.7.3 Result of Symbol Matching Game

Figure 4.14 shows the follower's average response time in the symbol matching game. The response time is defined as the difference between the time the leader selects a symbol and the time the follower clicks the correctly matched symbol.

A Repeated Measures ANOVA revealed a significant main effect on the response time in Study 1 , $F_{2,22} = 7.99, p = .002$. Post-hoc pairwise comparison (with bonferonni corrections) revealed that Mirror was significantly faster than Video

($p = .005$). The Hybrid is between the Video and the Mirror. For Study 2 , the Mirror and the Tilt are very similar.

4.7.4 Discussion of Symbol Matching Game

In Study 1 , the Mirror condition has a better capability than the Video condition to transmit leader's gesture direction, intention and attention. In Study 2 , the Tilt board condition has a nearly equal capability with the Mirror condition. Thus, the Mirror and Tilt board conditions significantly (and Hybrid slightly) outperform the Video condition because rendering the leader's arm in the former conditions helps the follower anticipate the symbol that the leader is about to touch. We observed that many users understood the challenges of the Video condition and prepared themselves by concentration. Some other users had to scan through all shapes to identify the missing symbol. The subjects also discussed this in the interview - *"In the Video condition, you do not know which symbol is about to disappear, but when you can see where the hand was in Hybrid and Mirror conditions, you can anticipate which symbol will be selected."* Hence, touch and pointing positions are quite important for the immersive telepresence visualization.

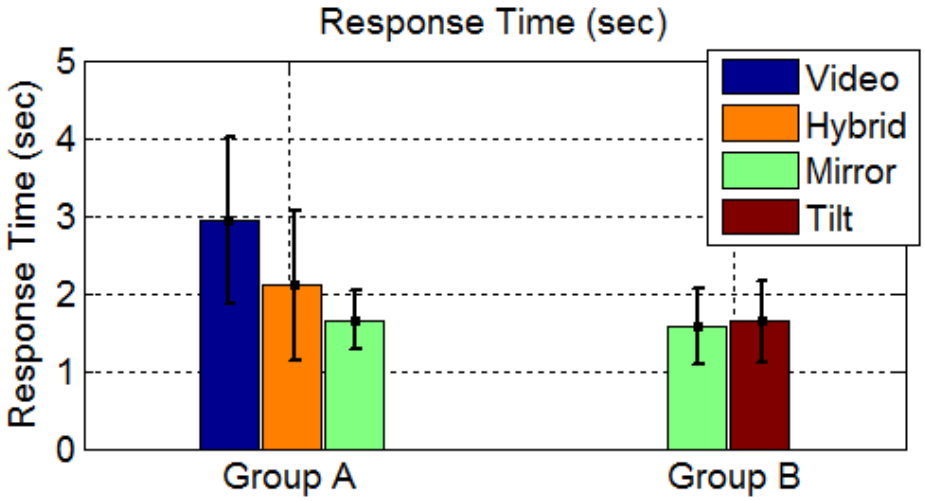

Figure 4.14: Response time of symbol matching game.

4.7.5 Result and Discussion of Negotiation Game

In Group A, the subjects spend more time negotiating in the Mirror condition than in the Video and Hybrid conditions (see Figure 4.15). As a result, the Mirror condition gains more scores in negotiation with a bigger summation of the selected numbers and less difference between the two participants. A Repeated Measures ANOVA revealed a significant main effect on the difference between two participants in Group A, $F_{2,22} = 7.70, p = .009$. Post-hoc pairwise comparison (with Bonferronni corrections) revealed that Mirror has significantly less difference between two partners than Video ($p = .03$). One possible reason is that the Mirror condition brings the two participants closer than the Video and Hybrid conditions. Face-to-face communication in the Mirror condition may encourage more negotiation since it is more like a real negotiation in the physical world.

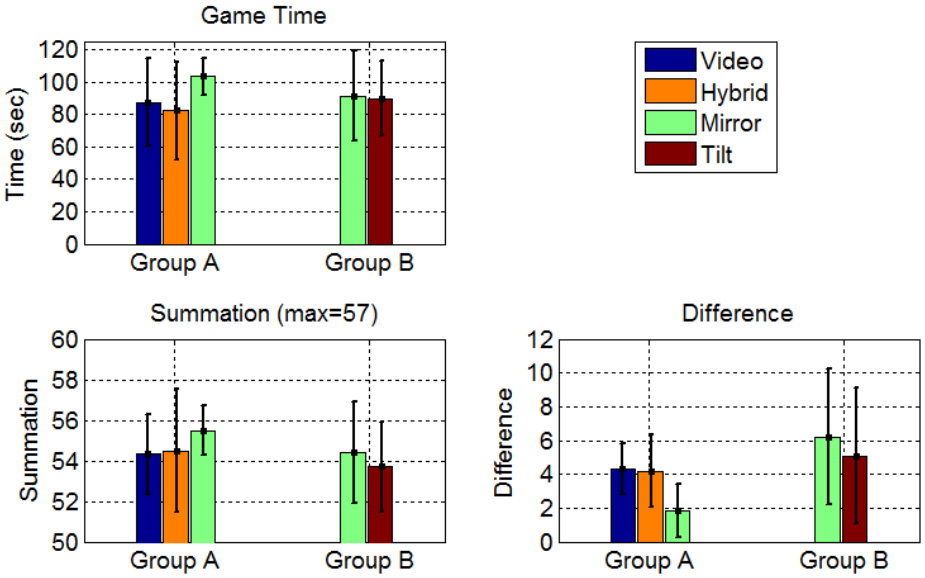

Figure 4.15: Negotiation. Top-Left: Negotiation time, Bottom-Left: average summation of scores from two participants, Bottom-Right: average difference of scores between the two participants.

Table 4.1: Result of Friedman and pairwise Wilcoxon Signed Ranks Tests on the participant's ranking in Study 1 . "*" indicates the significance.

Questions	Friedman Tests		Wilcoxon: Video/Hybrid		Wilcoxon: Video/Mirror	
	$\chi_{12,2}$	p	Z	p	Z	p
being together	10.17	.006*	−3.28	.001*	−2.43	.015*
enjoy experience	8.00	.018*	−2.81	.005*	−2.077	.038
video was useful	12.50	.002*	−3.22	.001*	−2.67	.008*
video quality	6.50	.039*	−1.90	.058	−.58	.564
eye contact	3.50	.174	−1.73	.083	−1.57	.117
ideas conveyed	15.17	.001*	−3.18	.001*	−2.91	.004*
where to look	3.50	.174	−1.73	.083	−1.57	.117
where to touch	18.00	.001*	−3.15	.002*	−3.15	.002*
read agreement	7.17	.028*	−2.80	.005*	−1.71	.087

4.7.6 Questionnaire

The participants are asked to rank the conditions with respect to the questions in Figure 4.16 and Table 4.1. In Study 1 , the subjects ranked the three conditions (Video, Hybrid, Mirror) from the worst to the best. In the Study 2 , the subjects pick the better condition from either Mirror and Tilt Board.

Figure 4.16 shows the ranking results. In Study 1 , most subjects prefer either Hybrid or Mirror for all questions except the question about video quality. Table 4.1 shows the statistic test results. Significant main effects were revealed for seven questions (except the *"eye contact"* and *"see where the partner was looking"*). Within these seven questions, pairwise comparison revealed that both Hybrid and Mirror conditions were ranked significantly better than the Video for four questions (*"being together," "video was useful," "ideas were conveyed," "see where the partner was about to touch"*) and the Hybrid condition was also ranked significantly better than the Video for another two questions (*"enjoy experience,"*, *"read partner's agreement level"*). The main effect ($p < .05$ for Friedman test) and pairwise significance ($p < .0167$ for Wilcoxon signed ranks test) is marked by "*". There is

no significance between the Hybrid and Mirror for any question.

In Study 2 , we observe the diversity of the subjects' preferences over all questions except *"video quality"*, on which the Tilt outperforms the Mirror significantly ($\chi_{20,1} = 9.80, p = .002$). Other than that question, Mirror and Tilt are very close.

Subjects in Study 2 also rated if the variations, i.e., head tracking for the Mirror and optional board rotation for the Tilt board, improved the condition on a 7 point scale (disagree=1, agree=7). Subjects mildly agreed on these two variations (head tracking 4.6/7 and board rotation 4.95/7).

Figure 4.16: Questionnaire: Ranking Results.

4.7.7 Feedback

ImmerseBoard, including Hybrid, Mirror and Tilt Board conditions, received very positive feedback from the subjects. All subjects were excited about using

ImmerseBoard because *"this is very cool, new experience that could be very useful in my profession," "this is impressive for remote collaboration," "we were naturally trying to help each other."*

The subjects also explained why they enjoyed ImmerseBoard, such as *"It was a good experience to see the video of my partner, to see his reaction, where he is looking and what he is doing," "I was able to predict where my partner is about to touch, and this would be great especially for co-workers."*. They also like the simple setup - *"The setup is so simple that I can easily fit this to my office."*

We also received negative feedback mostly on the video quality, particularly for the Mirror condition, like *"Video quality of partner was not that great, particularly face and eyes," "The video quality was not great. If it were better the experience would be much improved. It was difficult to see the eyes with current video quality."*

For the Hybrid condition, the participants liked the arm extension as the reference as they said, *"The hybrid condition is my favorite because I can see a clear cut of her and her arm, and I know where she is about to touch," "The video was decent, and you can see where your partner was writing and the part he was pointing."* The negative feedback included *"It is hard to tell where my partner was looking at horizontally"* and *"The extended arm sometimes made me distracted especially when the hand moves fast."*

For the Mirror condition, the subjects enjoyed the fluid experience, especially for rapid interactions such as brainstorming and collaborations - *"I definitely like the Mirror condition for the collaboration purpose, I was standing right with the partner and interacting closely," "It feels like both of us were physically there," "It was easy to see where my partner was looking and pointing and it was a little more precise."* Not surprisingly, the subjects did not like the video quality - *"I did not like the video quality in the Mirror, I really want to see where my partner's eyes were going."* Also, some participants were concerned about the overlapping of the remote user and the writing, *"My partner's body was blocked by what we drew on the board."* Actually, the overlapping causes difficulty seeing the writing if participant's outer clothes and the writing are same color.

For the Tilt condition, subjects felt that it was natural and realistic especially for the teaching scenarios - *"I like the Tilt board so much because it is more real-*

istic, we normally work on one side of the board," "It was more natural especially for teaching or presentation." However, some subjects were concerned that the perspective introduces imprecision as they commented - *"Compared to the Mirror, the Tilt board was more imprecise to see where my partner was pointing and where she was going to select," "Because of the perspective, the shapes look skewed as it should not be, and it made more difficulty to mentally register which shapes they were."*

Finally, the subjects discussed their preferences based on applications. In general, the subjects preferred the Hybrid and Tilt board conditions for teaching and presentation, while preferring the Mirror for close interaction and collaborations. (e.g. brainstorming).

4.8 Discussion

We now summarize what we learned from the study. First, participants quickly got used to the ImmerseBoard and preferred the three immersive conditions (Hybrid, Mirror, Tilt) over the Video condition since the immersive conditions provided them better ability to estimate their remote partners' eye gaze direction, gesture direction, and intention, making the remote collaboration more natural. Second, the participants enjoyed the 3D immersion (Mirror and Tilt) as it is more natural and realistic, and would show even more preference if the video quality were improved. Third, it is important to provide different conditions for participants to choose from, as they have diverse preferences and their preferences are also task-dependent. Finally, the participants felt the setup or form factor is so simple that they could envision using it in their office or home.

4.9 Conclusion of the Chapter

We introduced ImmerseBoard, which combines a large touch display with an RGBD camera to give remote collaborators an immersive experience as if they were writing side by side on a physical whiteboard or mirror. In addition to designing and implementing the system, we conducted a detailed user study involving

32 subjects performing several tasks. The tasks included a teaching task, as well as tasks to assess awareness of eye gaze direction and gestural attention/intention, all reflecting important aspects of real-world collaboration. Subjects were quantitatively better at estimating their remote partners' gesture direction and intention, and level of agreement, which translated qualitatively into a heightened sense of being together and a more enjoyable experience. However, the results also revealed limitations due to the 3D image quality from Kinect. In the future, we plan to improve the ImmerseBoard in several directions. First and foremost, we will improve the image quality by using high resolution sensors on both sides of the PPI screen. In addition, we will apply human body models for 3D reconstruction. We also plan to investigate the integration of three conditions (Hybrid, Mirror and Tilt) to provide the best experience for users based on applications and user's preference.

Chapter 5

Discussion

In this chapter, we discuss the knowledge that we obtained from the design, construction, and application of our proposed systems.

5.1 Scalable Remote Interaction

In this thesis, we developed two types of immersive telepresence systems using human body mapping and augmentation methods. We realized scalable remote interactions that connect differently scaled workspaces in remote and local locations. In the case of FlyingHead, the system allows operators to manipulate UAVs in large areas from the same operation room using changes in the mapping ratio. This system also offers small mapping capabilities, in which the operator movements are linearly reduced to correspond with the UAV movements on a different scale. ImmerseBoard facilitates the visualization of participants within wide virtual workspaces using body-shape mapping. ImmerseBoard can also transmit three visualizations to various display sizes. Thus, this system can support remote collaboration between differently sized displays.

In FlyingHead, we used linear mapping so that the system could linearly alter the mapping ratio between the operator and the moving aerial vehicle. In the second study, the extended mapping ratio exhibited better performance than normal mapping for a given movement task , although the performance of the mapping with the greatest difference in scale ($G = 4$) was lower than that of the other linear

mappings. All participants in our experiment reported the ability to control the aerial robot with linear mapping using the head-coupled system. We believe that linear mapping helps users to comprehend the movement distance of the robot. If the system uses non-linear mapping, participants can misinterpret the relationship between their head movements and the corresponding robot movements. This is because the physical operation of the robot involves delays in the robot's movement. Thus, the system must maintain a clear correspondence between the movement inputs and their outcomes.

The ImmerseBoard system transforms body shapes on Hybrid and Tilt conditions for cases in which the remote participant's arm is visually extended in a pointing/writing position. A user study involving the symbol-matching game showed that visualization of the extended hand helped users to understand the intentions of the remote participant. The extended hand also resolved the problem of over-wrapping strokes and body. Through the use of the Hybrid and Tilt conditions, participants naturally interacted with their remote partners in spite of the deformed body. The system seamlessly mapped the physical hand (local) to the visual hand (remote) from the perspectives of the participants.

5.2 Combined interaction

The two proposed systems designed using our developed method allow the use of device/button control mechanisms as a form of augmentation. This helps to overcome the limitations of body-based interactions. FlyingHead uses a device to manipulate the altitude of the aerial robot and, in the first user study, the shortest completion time was obtained using this mechanism. In ImmerseBoard, the Tilt condition also featured a rotation button that allowed adjustment of a tilted whiteboard so that it became a suitable writing medium. In particular, the subjects expressed favorable responses to the button used in the user-study teaching game (Group B). We assume that a key factor in designing combined interaction technology is seamless connection between inputs from the body and other inputs, such that users can concurrently use both or immediately change the input mode.

5.3 Further Applications

In this section, we discuss potential applications and platforms on which the devised the method and the knowledge obtained from the research presented in this thesis can be deployed.

Visual Human Body Mapping for Artificial Reality Environments

We plan to introduce our visual body mapping method to third-person artificial reality systems. ImmerseBoard has demonstrated that visualization of extended arms can convey human attention and intention to a remote environment. Further, the Go-Go Interaction [8] has used visual arm extension from the first-person perspective. We plan to expand visual body extension technology to artificial reality systems from third-person perspectives.

Figure 5.1 (a) shows body extension in which the user's arms are visually augmented in the artificial 3D environment. A self-human avatar is placed in virtual reality environments in order to provide an immersive working experience to users. Typically, a self-avatar's motions are constrained by human physical limitations, such as the operator's reach, jumping ability, and body height. We will design an interaction technique using a human-body geometry transformation in artificial environments that will provide augmented abilities, including extended arms, significantly enhanced jumping ability, and changes in body scale. This technique can potentially be used in a variety of applications such as gaming, user interfaces, and collaboration.

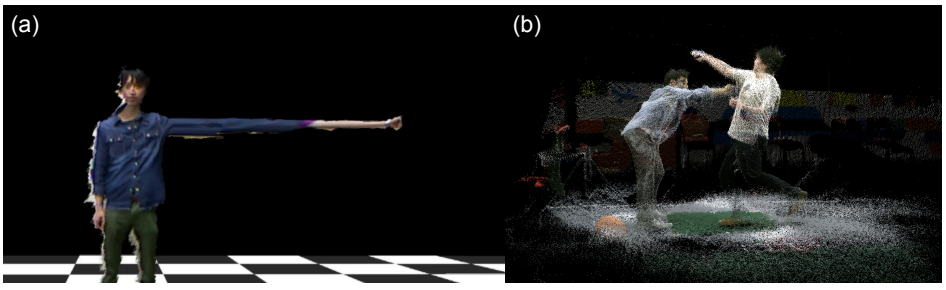

Figure 5.1: Further applications and platforms

We also plan to build a full-body capturing system for use with these artificial environments, as shown in Figure 5.1 (b). This system will capture the human body from several positions using cameras. Full-body capturing will permit free viewpoints without occlusions. Previous studies have presented markerless full-body capturing systems that utilize a large number of RGB cameras, but these systems require a long image processing time [99, 100]. We will focus on real-time capturing for interactive applications. Takeuchi et al. have developed PRIMA, a system that enables real-time 3D capturing using multiple depth cameras [101]. We plan to extend this real-time 3D capturing technique to improve the capturing area and geometric accuracy, and will also incorporate our body extension techniques.

Internal information Mapping

The body mapping and augmentation method can facilitate the monitoring of not only apparent information, but also unconscious physical information (e.g., biological information). The research presented in this thesis uses information related to the human body, including its position, rotation, and profile, to allow remote information correspondence. This study has succeeded in enabling scalable remote interactions in two immersive telepresence systems. We now envision that human body mapping can be used to permit correspondence with internal information, in order to improve creativity, motivation, and communication.

As an initial prototype, we have developed an exertion gaming system that uses biological information to motivate exercise activities through the immersive experience [102]. Many VR environments have recently adopted motion sensors to measure user exercise. Motion-sensor-based games can aid users in becoming more active and to improve their health. Biological data points, such as heart rate, and body temperature, vary with the amount of exercise taken. This system uses real-time biological information feedback to synchronize the behavior of the user and the virtual avatar in the game. For instance, the color of the avatar varies in accordance with the user's biological information. Further, we have designed a control mechanism that combines body motions and a hand-held controller device to provide consistency between the operability and exercise intensity.

Chapter 6

Conclusion

This thesis has addressed three problems affecting two types of telepresence applications, specifically, tele-robot operation and remote collaboration. To overcome these problems, we have proposed immersive telepresence systems using a human body mapping and augmentation interaction method. The first problem involved providing users with the ability to directly manipulate remote robot movements using body-movement mapping. The second challenge was to overcome the limitations of normal human movements via synchronization using both linear mapping and augmentation of the remote robot operation. The third problem was related to remote collaboration, which was overcome via a digital whiteboard system that provides human visualizations in order to transmit social cues (e.g., intention/attention) using a simple setup.

We developed a flying telepresence system called FlyingHead, which synchronizes operator movements with those of a flying robot, as a means of addressing the first and second problems described above. The operator's natural movement can be synchronized with the UAV motion, such as rotation and horizontal and vertical movements. The system enables a linear positional mapping through which the operator movements are linearly augmented in the motion of the flying robot. The system also supports the use of a small hand-held device to control the altitude of the UAV. We performed two user studies to measure the operability of this system and the efficacy of the mapping. The first study indicated that FlyingHead allows direct manipulation of the flying robot. In the second study, four mapping ratios

were compared and it was found that appropriate augmenting mapping ratios can offer superior control capabilities to the operator than normal mapping techniques.

To overcome the third issue, we developed the ImmerseBoard system for remote collaboration through a digital whiteboard. This system provides participants with a 3D immersive experience, which is simply enabled using only an RGBD camera mounted on the side of a large touchscreen display. ImmerseBoard realizes three novel visualizations of a remote partner on this simple setup, based on body-shape transformations. ImmerseBoard provides participants with a quantitatively better ability to estimate their remote partners' social cues. In addition, quantitative capabilities translate qualitatively into a heightened sense of togetherness and a more enjoyable experience. ImmerseBoard's form factor is suitable for practical and easy installation in homes and offices.

By developing two telepresence systems, we realized scalable remote interactions that connect differently scaled workspaces in remote and local locations. To expand on the knowledge obtained in this study, our proposed future work in this field includes refinement of the two telepresence systems discussed in the chapters of this thesis, as well as exploration of other applications of immersive telepresence systems.

Appendix A: Publications on the Thesis

Main Publications of the Thesis

Journal Paper

1. 樋口啓太, 暦本純一：移動感覚の拡張が可能なフライングテレプレゼンスプラットフォーム, 日本バーチャルリアリティ学会 論文誌, vo.19, no.3, pp 397-404, 2014. ISSN:1344011X

Refereed International Conference Proceedings

2. Keita Higuchi, Katsuya Fujii, Jun Rekimoto Flying Head: A Head-Synchronization Mechanism for Flying Telepresence,The 23rd h IEEE International Conference on Artificial Reality and Telexistence (ICAT 2013), pp.28-34, December 11-13, 2013, Tokyo, Japan. DOI:10.1109/ICAT.2013.6728902

3. Keita Higuchi, Yinpeng Chen, Philip A Chou, Zhengyou Zhang, Zicheng Liu, ImmerseBoard: Immersive Telepresence Experience using a Digital Whiteboard, the SIGCHI Conference on Human Factors in Computing Systems 2015 (CHI 2015), April 18-23, Seoul, Korea, 2015. DOI:10.1145/2702123.2702160

Related Publications

4. Keita Higuchi, Michihiko Ueno, Jun Rekimoto, Scarecrow: Avatar Representation using Biological Information Feedback, The 2014 IEEE International Conference on Cyber, Physical and Social Computing (CPSCom 2014), September 1-3, 2014. doi:10.1109/iThings.2014.66

5. Keita Higuchi, Jun Rekimoto Flying Head: A Head Motion Synchronization Mechanism for Unmanned Aerial Vehicle Control,CHI 2013 Extended Abstracts (alt.chi), pp.2029-2038, April 27May 2, 2013, Paris, France. DOI:10.1145/2468356.2468721

6. Keita Higuchi, Jun Rekimoto Flying Head: Head-synchronized Unmanned Aerial Vehicle Control for Flying Telepresence,Siggraph Asia 2012 Emerging Technologies, Article No. 12:1-2, November 28 December 1, 2012, Singapore. DOI:10.1145/2407707.2407719

7. Keita Higuchi, Tetsuro Shimada and Jun Rekimoto, Flying Sports Assistant: External Visual Imagery Representation for Sports Training, The 2nd International conference on Augmented Human (AH 2011), Article No.7:1-4, March 12-14, Odaiba, Tokyo, Japan. DOI:10.1145/2582051.2582064

Reference

[1] Jun Rekimoto and Katashi Nagao. The world through the computer: Computer augmented interaction with real world environments. In *Proceedings of the 8th annual ACM symposium on User interface and software technology*, pages 29–36. ACM, 1995.

[2] Sunao Hashimoto, Akihiko Ishida, Masahiko Inami, and Takeo Igarashi. Touchme: An augmented reality based remote robot manipulation. In *The 21st International Conference on Artificial Reality and Telexistence, Proceedings of ICAT2011*, 2011.

[3] Charith Lasantha Fernando, Masahiro Furukawa, Tadatoshi Kurogi, Kyo Hirota, Sho Kamuro, Katsunari Sato, Kouta Minamizawa, and Susumu Tachi. Telesar v: Telexistence surrogate anthropomorphic robot. In *ACM SIGGRAPH 2012 Emerging Technologies*, SIGGRAPH '12, pages 23:1–23:1, New York, NY, USA, 2012. ACM.

[4] Kentaro Ishii, Yuji Taniguchi, Hirotaka Osawa, Kazuhiro Nakadai, and Michita Imai. Merging viewpoints of user and avatar in automatic control of avatar-mediated communication.

[5] Alan Cooper, Robert Reimann, David Cronin, and Christopher Noessel. *About Face: The essentials of interaction design 3rd Edition.* John Wiley & Sons, 2014.

[6] Telexistence Robot "Telesar". http://tachilab.org/modules/projects/telesar.html.

[7] da Vinci Surgical System. http://www.intuitivesurgical.com/.

[8] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. The go-go interaction technique: non-linear mapping for direct manipulation in vr. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, pages 79–80. ACM, 1996.

[9] Ivan E. Sutherland. The ultimate display. In *Proceedings of the IFIP Congress*, pages 506–508, 1965.

[10] Pierre Wellner, Wendy Mackay, and Rich Gold. Back to the real world. *Commun. ACM*, 36(7):24–26, July 1993.

[11] M. Weiser. Ubiquitous computing. *Computer*, 26(10):71–72, October 1993.

[12] Pierre Wellner, Wendy Mackay, and Rich Gold. Back to the real world. *Commun. ACM*, 36(7):24–26, July 1993.

[13] Saied Moezzi. Immersive telepresence. *IEEE Multimedia*, 4(1):17, 1997.

[14] Myron W Krueger. *Artificial reality II*, volume 10. Addison-Wesley Reading (Ma), 1991.

[15] Myron W Krueger, Thomas Gionfriddo, and Katrin Hinrichsen. Videoplace —an artificial reality. In *ACM SIGCHI Bulletin*, volume 16, pages 35–40. ACM, 1985.

[16] S. Tachi. Real-time remote robotics-toward networked telexistence. *Computer Graphics and Applications, IEEE*, 18(6):6–9, 1998.

[17] Carolina Cruz-Neira, Daniel J Sandin, and Thomas A DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the cave. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 135–142. ACM, 1993.

[18] T. Kanade, P. Rander, and P.J. Narayanan. Virtualized reality: constructing virtual worlds from real scenes. *MultiMedia, IEEE*, 4(1):34–47, Jan 1997.

[19] Chris Dede. Immersive interfaces for engagement and learning. *science*, 323(5910):66–69, 2009.

[20] D. Saakes, V. Choudhary, D. Sakamoto, M. Inami, and T. Lgarashi. A teleoperating interface for ground vehicles using autonomous flying cameras. In *Artificial Reality and Telexistence (ICAT), 2013 23rd International Conference on*, Dec 2013.

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

[22] Daniel F Keefe, Daniel Acevedo Feliz, Tomer Moscovich, David H Laidlaw, and Joseph J LaViola Jr. Cavepainting: a fully immersive 3d artistic medium and interactive experience. In *Proceedings of the 2001 symposium on Interactive 3D graphics*, pages 85–93. ACM, 2001.

[23] Doug A Bowman, Ernst Kruijff, Joseph J LaViola Jr, and Ivan Poupyrev. An introduction to 3-d user interface design. *Presence: Teleoperators and virtual environments*, 10(1):96–108, 2001.

[24] I Scott MacKenzie and William Buxton. Extending fitts' law to two-dimensional tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 219–226. ACM, 1992.

[25] I. Scott MacKenzie. Fitts' law as a research and design tool in human-computer interaction. *Hum.-Comput. Interact.*, 7(1):91–139, March 1992.

[26] Mitchell W. McEwan, Alethea L. Blackler, Daniel M. Johnson, and Peta A. Wyeth. Natural mapping and intuitive interaction in videogames. In *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-human Interaction in Play*, CHI PLAY '14, pages 191–200, New York, NY, USA, 2014. ACM.

[27] C. Ware and S. Osborne. Exploration and virtual camera control in virtual three dimensional environments. In *ACM SIGGRAPH Computer Graphics*, volume 24, pages 175–183. ACM, 1990.

[28] Ken Hinckley, Randy Pausch, Dennis Proffitt, and Neal F. Kassell. Two-handed virtual manipulation. *ACM Trans. Comput.-Hum. Interact.*, 5(3):260–302, September 1998.

[29] D. J. Sturman, D. Zeltzer, and S. Pieper. Hands-on interaction with virtual environments. In *Proceedings of the 2Nd Annual ACM SIGGRAPH Symposium on User Interface Software and Technology*, UIST '89, pages 19–24, New York, NY, USA, 1989. ACM.

[30] Jeffrey S Pierce, Brian C Stearns, and Randy Pausch. Voodoo dolls: seamless interaction at multiple scales in virtual environments. In *Proceedings of the 1999 symposium on Interactive 3D graphics*, pages 141–145. ACM, 1999.

[31] Ivan Poupyrev, T Ichikawa, S Weghorst, and M Billinghurst. Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. In *Computer Graphics Forum*, volume 17, pages 41–52. Wiley Online Library, 1998.

[32] Mhd Yamen Saraiji, Charith Lasantha Fernando, Yusuke Mizushina, Youichi Kamiyama, Kouta Minamizawa, and Susumu Tachi. Enforced telexistence: Teleoperating using photorealistic virtual body and haptic feedback. In *SIGGRAPH Asia 2014 Emerging Technologies*, SA '14, pages 7:1–7:2, 2014.

[33] MHD Yamen Saraiji, Charith Lasantha Fernando, Kouta Minamizawa, and Susumu Tachi. Mutual hand representation for telexistence robots using projected virtual hands. In *Proceedings of the 6th Augmented Human International Conference*, AH '15, pages 221–222, 2015.

[34] R. Heinlein. *Waldo.* Astounding Science Fiction, August 1942.

[35] Maeda. Arvin Agah. Eimei, Oyama. Taro and Tachi. Susumu. Robots for telexistence and telepresence: from science fiction to reality. In *ICAT 2004*, 2004.

[36] S.W. Lee. Automatic gesture recognition for intelligent human-robot inter-

action. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 645–650. Ieee, 2006.

[37] T. Kamegawa, T. Yamasaki, H. Igarashi, and F. Matsuno. Development of the snake-like rescue robot. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 5, pages 5081–5086. IEEE, 2004.

[38] J. Scholtz, J. Young, J.L. Drury, and H.A. Yanco. Evaluation of human-robot interaction awareness in search and rescue. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 3, pages 2327–2332. IEEE, 2004.

[39] A.S. Huang, A. Bachrach, P. Henry, M. Krainin, D. Maturana, D. Fox, and N. Roy. Visual odometry and mapping for autonomous flight using an rgb-d camera. In *Int. Symposium on Robotics Research (ISRR),(Flagstaff, Arizona, USA)*, 2011.

[40] A. Wendel, M. Maurer, G. Graber, T. Pock, and H. Bischof. Dense reconstruction on-the-fly. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1450 –1457, june 2012.

[41] MikroKopter. http://www.mikrokopter.de/.

[42] quaduino. http://code.google.com/p/quaduino-ng/.

[43] PHANTOM 2 Visun Plus. http://www.dji.com/ja/product/phantom-2-vision-plus.

[44] Bebap Drone. http://www.parrot.com/products/bebop-drone/.

[45] VGo. http://www.vgocom.com/.

[46] MITRE. http://www.mitre.org/.

[47] C. Mancini and F. Roncaglia. Il servomeccanismo elettronico mascot i del cnen. 32(6):379–392, 1963.

[48] D.C. Smith J.D. Hightower. Teleoperator technology development. In *Proc. of the 12th Meeting of the United States-Japan Cooperative Program in Natural Resource*, 1983.

[49] JJ Heuring and DW Murray. Visual head tracking and slaving for visual telepresence. In *Robotics and Automation, 1996*, volume 4, pages 2908–2914. IEEE, 1996.

[50] M. Quigley, M.A. Goodrich, and R.W. Beard. Semi-autonomous human-uav interfaces for fixed-wing mini-uavs. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 3, pages 2457–2462. IEEE, 2004.

[51] P.R. Giordano, H. Deusch, J. Lächele, and HH Bülthoff. Visual-vestibular feedback for enhanced situational awareness in teleoperation of uavs. In *Proc. of the AHS 66th Annual Forum and Technology Display*, 2010.

[52] T. Naseer, J. Sturm, and D. Cremers. FollowMe: Person following and gesture recognition with a quadrocopter. In *Proc. of the Int. Conf. on Intelligent Robot Systems (IROS)*, 2013.

[53] Ehud Sharlin Wai Shan (Florence) Ng. Collocated interaction with flying robots. In *RO-MAN, 2011 IEEE*, pages 143–149. IEEE, 2011.

[54] Pieter Vries, Sjoerd C.; Padmos. Steering a simulated unmanned aerial vehicle using a head-slaved camera and hmd: effects of hmd quality, visible vehicle references, and extended stereo cueing. In *Proc. SPIE*, volume 3362, pages 80–91.

[55] Corey Pittman and Joseph J. LaViola, Jr. Exploring head tracked head mounted displays for first person robot teleoperation. In *Proceedings of the 19th International Conference on Intelligent User Interfaces*, IUI '14, pages 323–328, 2014.

[56] J.M. Teixeira, R. Ferreira, M. Santos, and V. Teichrieb. Teleoperation using

google glass and ar, drone for structural inspection. In *Virtual and Augmented Reality (SVR), 2014 XVI Symposium on*, pages 28–36, May 2014.

[57] Hirohiko Hayakawa, Charith Lasantha Fernando, MHD Yamen Saraiji, Kouta Minamizawa, and Susumu Tachi. Telexistence drone: Design of a flight telexistence system for immersive aerial sports experience. In *Proceedings of the 6th Augmented Human International Conference*, AH '15, pages 171–172, 2015.

[58] Hiroo Iwata. Art and technology in interface devices. In *Proceedings of the ACM symposium on Virtual reality software and technology*, VRST '05, pages 1–7, New York, NY, USA, 2005. ACM.

[59] Hideki Yoshimoto, Kazuhiro Jo, and Koichi Hori. Designing interactive blimps as puppets. *Entertainment Computing ICEC 2009*, 5709:204–209, 2009.

[60] F. Okura, M. Kanbara, and N. Yokoya. Augmented telepresence using autopilot airship and omni-directional camera. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*, pages 259–260, 2010.

[61] K. Higuchi, Y. Ishiguro, and J. Rekimoto. Flying eyes: free-space content creation using autonomous aerial vehicles. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, pages 561–570. ACM, 2011.

[62] K. Higuchi, T. Shimada, and J. Rekimoto. Flying sports assistant: external visual imagery representation for sports training. In *Proceedings of the 2nd Augmented Human International Conference*, page 7. ACM, 2011.

[63] Eberhard Graether and Florian Mueller. Joggobot: a flying robot as jogging companion. In *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts*, CHI EA '12, pages 1063–1066. ACM, 2012.

[64] Florian 'Floyd' Mueller and Matthew Muirhead. Jogging with a quadcopter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2023–2032, 2015.

[65] W. Bluethmann, R. Ambrose, M. Diftler, S. Askew, E. Huber, M. Goza, F. Rehnmark, C. Lovchik, and D. Magruder. Robonaut: A robot designed to work with humans in space. *Autonomous Robots*, 14(2):179–197, 2003.

[66] Q. Lindsey, D. Mellinger, and V. Kumar. Construction of cubic structures with quadrotor teams. *Proc. Robotics: Science & Systems VII*, 2011.

[67] Joseph J. LaViola, Jr., Daniel Acevedo Feliz, Daniel F. Keefe, and Robert C. Zeleznik. Hands-free multi-scale navigation in virtual environments. In *Proceedings of the 2001 symposium on Interactive 3D graphics*, I3D '01, pages 9–15, New York, NY, USA, 2001. ACM.

[68] John C. Tang and Scott Minneman. Videowhiteboard: video shadows to support remote collaboration. CHI '91, pages 315–322. ACM.

[69] Hiroshi Ishii and Minoru Kobayashi. Clearboard: a seamless medium for shared drawing and conversation with eye contact. CHI '92, pages 525–532. ACM.

[70] S. Moezzi. Immersive telepresence. *MultiMedia, IEEE*, 4(1):17–17, 1997.

[71] Norbert A. Streitz, Jörg Geissler, Torsten Holmer, Shin'ichi Konomi, Christian Müller-Tomfelde, Wolfgang Reischl, Petra Rexroth, Peter Seitz, and Ralf Steinmetz. i-land: an interactive landscape for creativity and innovation. CHI '99, pages 120–127.

[72] Azam Khan, Justin Matejka, George Fitzmaurice, and Gordon Kurtenbach. Spotlight: directing users' attention on large displays. CHI '05, pages 791–798. ACM.

[73] Jeremy P. Birnholtz, Tovi Grossman, Clarissa Mak, and Ravin Balakrishnan. An exploratory study of input configuration and group process in a negotiation task using a large display. CHI '07, pages 91–100. ACM.

[74] Scott Elrod, Richard Bruce, Rich Gold, David Goldberg, Frank Halasz, William Janssen, David Lee, Kim McCall, Elin Pedersen, Ken Pier, John Tang, and Brent Welch. Liveboard: a large interactive display supporting group meetings, presentations, and remote collaboration. CHI '92, pages 599–607. ACM.

[75] Jun Rekimoto. A multiple device approach for supporting whiteboard-based interactions. CHI '98, pages 344–351. ACM, 1998.

[76] Mark Apperley, Laurie McLeod, Masood Masoodian, Lance Paine, Malcolm Phillips, Bill Rogers, and Kirsten Thomson. Use of video shadow for small group interaction awareness on a large interactive display surface. AUIC '03, pages 81–90.

[77] Kar-Han Tan, I. Robinson, R. Samadani, Bowon Lee, D. Gelb, A. Vorbau, B. Culbertson, and J. Apostolopoulos. Connectboard: A remote collaboration system that supports gaze-aware interaction and sharing. In *MMSP '09*, pages 1–6.

[78] Nicolas Roussel. Experiences in the design of the well, a group communication device for teleconviviality. MULTIMEDIA '02, pages 146–152. ACM.

[79] Anthony Tang, Michel Pahud, Kori Inkpen, Hrvoje Benko, John C. Tang, and Bill Buxton. Three's company: understanding communication channels in three-way distributed collaboration. CSCW '10, pages 271–280. ACM.

[80] Aaron M. Genest, Carl Gutwin, Anthony Tang, Michael Kalyn, and Zenja Ivkovic. Kinectarms: a toolkit for capturing and displaying arm embodiments in distributed tabletop groupware. CSCW '13, pages 157–166. ACM.

[81] Andre Doucette, Carl Gutwin, Regan L. Mandryk, Miguel Nacenta, and Sunny Sharma. Sometimes when we touch: how arm embodiments change reaching and collaboration on digital tables. CSCW '13, pages 193–202. ACM.

[82] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. The office of the future: a unified approach to image-based modeling and spatially immersive displays. SIGGRAPH '98, pages 179–188.

[83] J. Mulligan, X. Zabulis, N. Kelshikar, and K. Daniilidis. Stereo-based environment scanning for immersive telepresence. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(3):304–320, 2004.

[84] Shujie Liu, Philip A. Chou, Cha Zhang, Zhengyou Zhang, and Chang Wen Chen. Virtual view reconstruction using temporal information. *IEEE ICME 2012*, pages 115–120.

[85] Cha Zhang, Qin Cai, P.A. Chou, Zhengyou Zhang, and R. Martin-Brualla. Viewport: A distributed, immersive teleconferencing system with infrared dot pattern. *MultiMedia, IEEE*, 20(1):17–27, 2013.

[86] Osamu Morikawa and Takanori Maesako. Hypermirror: toward pleasant-to-use video mediated communication system. CSCW '98, pages 149–158. ACM.

[87] Norman P. Jouppi, Subu Iyer, Stan Thomas, and April Slayden. Bireality: mutually-immersive telepresence. MULTIMEDIA '04, pages 860–867. ACM.

[88] Kibum Kim, John Bolton, Audrey Girouard, Jeremy Cooperstock, and Roel Vertegaal. Telehuman: effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. CHI '12, pages 2531–2540. ACM.

[89] Kana Misawa, Yoshio Ishiguro, and Jun Rekimoto. Livemask: a telepresence surrogate system with a face-shaped screen for supporting nonverbal communication. AVI '12, pages 394–397. ACM.

[90] Belle L. Tseng, Z.-Y. Shae, Wing Ho Leung, and Tsuhan Chen. Immersive whiteboards in a networked collaborative environment. In *IEEE Multimedia and Expo*, 2001.

[91] Thomas Nescher and Andreas Kunz. An interactive whiteboard for immersive telecollaboration. *The Visual Computer*, 27(4):311–320, 2011.

[92] Shingo Uchihashi and Tsutomu Tanzawa. Mixing remote locations using shared screen as virtual stage. MM '11, pages 1265–1268. ACM.

[93] Sasa Junuzovic, Kori Inkpen, Tom Blank, and Anoop Gupta. Illumishare: sharing any surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1919–1928.

[94] Jakob Zillner, Christoph Rhemann, Shahram Izadi, and Michael Haller. 3d-board: A whole-body remote collaborative whiteboard. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 471–479, 2014.

[95] E. Ilana Diamant, Susan R. Fussell, and Fen-Ly Lo. Collaborating across cultural and technological boundaries: team culture and information use in a map navigation task. IWIC '09, pages 175–184. ACM.

[96] Adam Kendon. Some functions of gaze-direction in social interaction. *Acta psychologica*, 26:22–63, 1967.

[97] Wei Dong and Wai-Tat Fu. One piece at a time: why video-based communication is better for negotiation and conflict resolution. CSCW '12, pages 167–176.

[98] Gottlob I. Proudlock FA, Shekhar H. Coordination of eye and head movements during reading. *Invest Ophthalmol Vis Sci.*, 44(7):2991–2998, 2003.

[99] T. Brox, B. Rosenhahn, J. Gall, and D. Cremers. Combined region- and motion-based 3d tracking of rigid and articulated objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):402–415, 2009.

[100] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. Youmove: Enhancing movement training with an augmented reality mirror. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 311–320, 2013.

[101] Toshiki Takeuchi, Totaro Nakashima, Kunihiro Nishimura, and Michitaka Hirose. Prima: Parallel reality-based interactive motion area. In *ACM SIGGRAPH 2011 Posters*, page 80. ACM, 2011.

[102] Keita Higuchi, Michihiko Ueno, and Jun Rekimoto. Scarecrow: Avatar representation using biological information feedback. In *Internet of Things (iThings), 2014 IEEE International Conference on, and Green Computing and Communications (GreenCom), IEEE and Cyber, Physical and Social Computing (CPSCom), IEEE*, pages 352–359. IEEE, 2014.